论 文

LTE 网络中具备 QoS 保障的动态负载均衡算法

王浩^{①②*}, 李知航^①, 潘志文^①, 尤肖虎^①, 吴平^②

① 东南大学信息学院移动通信国家重点实验室, 南京 210096

② Department of Engineering Sciences, Uppsala University, Uppsala 75237, Sweden

* 通信作者. E-mail: hao.wang@seu.edu.cn

收稿日期: 2011-10-19; 接受日期: 2012-04-19

中瑞国际合作项目 (批准号: 2008DFA12090)、国家重点基础研究发展计划(批准号: 2012CB316004)、国家科技重大专项 (批准号: 2011ZX03003-002-02)、江苏省 2011 年度普通高校研究生科研创新计划(批准号: CXLX-0116) 和江苏省“六大人才高峰”资助项目

摘要 研究了 3GPP LTE 网络中考虑不同服务质量 (quality-of-service, QoS) 要求的动态负载均衡算法. 小区间的负载不均衡对于不同 QoS 要求用户有不同负面影响. 对于有保障速率要求的用户, 负载不均衡会导致较高的新呼叫阻塞率; 而对于那些没有速率要求的用户, 负载不均衡会导致繁忙小区中边缘用户过于恶化的吞吐量. 全网中这两类用户的负载均衡问题紧密耦合, 难以用一个统一的目标函数来分析研究. 因此提出一个相应的多目标优化问题, 其目标函数分别是针对全网内有 QoS 要求用户的负载均衡指示函数和针对全网内没有 QoS 要求用户的总效用函数, 限制条件为实际小区物理资源和用户 QoS 要求. 对该问题的复杂度进行了分析后, 提出了一个实时的低复杂度低开销的分布式负载均衡算法结构, 包括 QoS 保障的混合调度, QoS 感知的负载均衡切换和呼叫准入控制. 最后, 系统级仿真结果显示提出的全新的负载均衡算法结构达到了较好的负载均衡效果, 可以显著降低有 QoS 要求用户的新呼叫阻塞率, 同时以略微损失全网无 QoS 要求用户总吞吐量的代价大幅提升繁忙小区中边缘用户的实际吞吐量.

关键词 负载均衡 3GPP LTE 服务质量 (QoS) 多目标优化 负载均衡指示函数 总效用函数

1 引言

3GPP LTE 系统由于采用了 OFDM, MIMO 等先进的物理层传输技术、扁平化的网络结构和全 IP 系统架构, 可以支持更大的系统带宽、更高的峰值速率和更短的传输延时, 并得到更高的频谱效率^[1,2]. 但由于业务的突发性, 用户移动的随机性, 它仍然存在类似于之前 2G 的 GSM 网络和 3G 的 WCDMA 网络中的负载不均衡问题^[3~5], 即在某一时段内, 系统中某些小区的用户数或需求的服务量远大于周边邻区, 导致这些小区内的用户体验较差或是小区的服务拒绝率较高, 而周边邻区内用户体验过高或是小区资源没有被完全利用的情况. 针对蜂窝移动通信网中存在的负载不均衡问题, 已经有相当多的研究. 在诸如 2G 的 GSM 等电路交换网络中, 每个服务用户被分配一条专用的信道, 小区的负载程度可以体现为剩余信道的多少, 则可以通过“信道借用”^[6,7]和“呼叫转移”^[8,9]等方式平衡不同小区间的剩余信道数, 从而达到负载均衡的目的. Tonguz 等提出了一个全面综合的数学模型^[10], 可

引用格式: 王浩, 李知航, 潘志文, 等. LTE 网络中具备 QoS 保障的动态负载均衡算法. 中国科学: 信息科学, 2012, 42: 674-686, doi: 10.1360/112011-1119

用于理论分析电路交换网络中大部分负载均衡方案. 在包交换网络中, 如 3G 的 WCDMA 网络和下一代 LTE 或 WiMAX 网络, 对单一用户已不存在专用信道的概念, 小区通过合适的机会调度算法来利用其内的时频资源, 并同时给较多的用户提供服务^[1]. 而与之对应的负载均衡研究大多被建模为小区与用户之间依据不同性能指标动态配对的优化问题^[11~15]. 然而, 以上大多数研究在进行负载均衡算法设计时, 考虑的用户类型多为没有任何服务质量 (quality of service, QoS) 要求的用户, 考虑的性能指标都为与吞吐量相关的效用函数, 负载均衡的效果体现在效用函数的提升上. 但是, 实际的 LTE 系统提供服务的对象为具有不同 QoS 要求的不同类型用户, 它力求保证每个用户的 QoS 要求并尽可能接纳更多的用户. 因此, 3GPP LTE 网络中针对不同 QoS 要求用户的负载均衡切换算法的设计, 需要考虑这些用户进行负载均衡切换时产生的不同影响.

本文从一个全新的角度来研究 3GPP LTE 网络中的动态负载均衡算法, 其特点在于区分不同类别用户的不同 QoS 需求, 在严格保证用户 QoS 需求的同时最大化网络收益. 对于 LTE 网络中具有 QoS 要求的用户, 最主要的两类 QoS 指标为保障速率要求和最大 (或平均) 延时要求^[16]. 由于后者数学上很难分析并可以由前者隐含地保证^[17], 因此本文采用保障速率要求作为有 QoS 要求用户的性能指标. 而对于没有 QoS 要求的用户, 则用尽力而为业务用户来建模, 即该类用户没有速率保障, 每个用户实际可得速率取决于系统的可用资源及采用的调度算法. 以上保障速率需求和尽力而为两类业务用户基本囊括了 LTE 系统中实际的所有业务类型, 且便于进行数学上的分析. 之后将该问题建模为一个多目标优化问题, 其目标函数分别是针对全网内有 QoS 要求用户的负载均衡指示函数和针对全网内没有 QoS 要求用户的总效用函数, 限制条件为实际小区物理资源和用户 QoS 要求. 接着对该问题的复杂度进行了分析, 并提出了一个实时的低复杂度低开销的负载均衡算法结构, 包括 QoS 保障的混合调度, QoS 感知的负载均衡切换和新呼叫准入控制. QoS 保障的混合调度以合理的开销严格保证用户的 QoS 需求, 同时最大化无 QoS 要求用户的总体收益, 即总效用函数. QoS 感知的负载均衡切换, 对于有 QoS 要求的用户, 可以平衡不同小区间的硬负载; 对于没有 QoS 要求的用户, 可以提升繁忙小区中边缘用户的吞吐量. 新呼叫准入控制则用以严格保证新接入用户的 QoS 需求. 最后, 充分的系统级仿真结果显示本文提出的全新的负载均衡算法结构可以显著提高针对 QoS 要求用户的负载均衡指示函数, 从而降低该类用户的新呼叫阻塞率, 并提供严格的 QoS 保证; 同时也提升了针对没有 QoS 要求用户的总效用函数, 从而提升了繁忙小区中边缘用户的实际可得吞吐量.

本文结构如下: 第 2 节详述了参考的系统模型; 第 3 节对问题进行构造, 提出相关的多目标优化问题并对其复杂度和开销进行了分析; 第 4 节提出了实时的低复杂度低开销的分布式算法结构; 第 5 节给出仿真场景设置及结果; 第 6 节对全文进行了总结.

2 系统模型

2.1 网络模型

本文提出的负载均衡算法适用于任意的网络拓扑结构, 为便于表达, 实际参考的网络模型为图 1 所示的 3GPP LTE 下行多蜂窝网络. 其中两类不同用户分别为有 QoS 需求的常量比特率 (constant bit rate, CBR) 用户和没有 QoS 需求的尽力而为 (best effort, BE) 用户. 为叙述方便, 分别称这两类用户为 CBR 用户和 BE 用户. 仍如图 1 所示, 此处有 7 个小区, 每个小区由一个中心基站 (enhanced nodeB, eNodeB) 来控制. 频域上每 12 个相邻的子载波组成一个物理资源块 (physical resource block, PRB), 这是一个子帧 (1 ms) 内可以分给一个用户的最小频域单元. 假设小区 1 为繁忙小区, 其边缘

图 1 网络模型

Figure 1 Network Model

CBR 用户 a 和 BE 用户 b 可分别切换至 5, 6 和 2, 3 两个相邻小区. 此时将哪个用户切换至哪个目标小区最优, 是下面解决的问题.

使用 $\mathcal{N}$, $\mathcal{L}$, $\mathcal{K}$, $\mathcal{C}$, $\mathcal{B}$ 来分别表示小区集合, 每个小区中的 PRB 集合、总用户集合、CBR 用户集合和 BE 用户集合. 很显然有 $\mathcal{K} = \mathcal{C} \cup \mathcal{B}$, 即总用户集合为 CBR 用户集和 BE 用户集的并集. 定义如下的小区、用户匹配指示变量 $I_{i,k}(t)$, 当用户 k 在时刻 t 由小区 i 提供服务, 该变量为 1, 否则为 0. 本文中所有的时刻 t 代表某个进行负载均衡切换判决的时刻, 而任意两个时刻 t 和 $t+1$ 之间的间隔为 1 个负载均衡周期, 其长度远大于一个子帧. 在每一时刻 t 进行负载均衡算法的判决依据来源于上一负载均衡周期 $(t-1, t]$, 其判决结果作用于下一负载均衡周期 $(t, t+1]$.

2.2 链路模型

通过导频检测每个用户可以得到每个可测小区的每个 PRB 上的瞬时信号强度, 即在某个子帧 τ 上, 用户 $k \in \mathcal{K}$ 在小区 $i \in \mathcal{N}$ 中的第 $l \in \mathcal{L}$ 个 PRB 上的瞬时接收信干噪比 $\text{SINR}_{i,l,k}(\tau)$ 为

$$\text{SINR}_{i,l,k}(\tau) = \frac{g_{i,l,k}(\tau)p_{i,l}(\tau)}{N_o + \sum_{j \in \mathcal{N}, j \neq i} g_{j,l,k}(\tau)p_{j,l}(\tau)}, \quad (1)$$

其中 N_o 为每个 PRB 上相等的加性高斯白噪声功率. $g_{i,l,k}(\tau)$ 和 $p_{i,l}(\tau)$ 分别表示在时刻 τ , 用户 k 在基站 i 的第 l 个 PRB 上的瞬时信道增益和基站 i 在第 l 个 PRB 上的发射功率, 则 $g_{i,l,k}(\tau)p_{i,l}(\tau)$ 表示时刻 τ 用户 k 在小区 i 的第 l 个 PRB 上接收到的瞬时信号强度. 则在时刻 t , 用户 k 在小区 i 中的平均带宽效率 $e_{i,k}(t)$ 可按如下方式进行计算

$$e_{i,k}(t) = \frac{1}{|\mathcal{L}| \cdot N_\tau} \left\{ \sum_{\tau \in (t-1, t]} \sum_{l \in \mathcal{L}} \log_2[1 + \text{SINR}_{i,l,k}(\tau)] \right\}. \quad (2)$$

上式中 $|\mathcal{L}|$ 表示 PRB 集合 $\mathcal{L}$ 的势, 即 PRB 集合内元素的个数; N_τ 表示一个负载均衡周期内子帧的个数, 如负载均衡周期为 1 s, 则每周期内包含 1000 个子帧. 各用户将其在每个负载均衡周期内

统计的平均带宽效率 $e_{i,k}(t)$ 上报给其服务 eNodeB. 对于有 QoS 需求的 CBR 用户 k , 其实际可得速率取决于服务基站分配的时频资源和它的信道状态, 即平均带宽效率 $e_{i,k}(t)$. 使用 $w_{i,k}(t)$ 表示在时刻 t 小区 i 分配给 CBR 用户 k 的时频资源, 则用户 k 的实际可得速率为 $R_{i,k}(t) = w_{i,k}(t)e_{i,k}(t)$.

2.3 针对全网内所有 CBR 用户的负载均衡指示函数

分别使用 $s_i(t)$, $s_i^c(t)$ 和 $s_i^b(t)$ 来表示时刻 t 小区 i 中的总资源, 所有 CBR 用户占用的资源及所有 BE 用户占用的资源. 定义小区 i 的硬负载为其内所有 CBR 用户占用资源数与小区总资源数之比, 如下所示

$$\rho_i(t) = \frac{s_i^c(t)}{s_i(t)} = \frac{\sum_{k \in \mathcal{C}} I_{i,k}(t) w_{i,k}(t)}{s_i(t)}. \quad (3)$$

在 LTE 网络中, 同类型小区一般有相等的时频资源, 所以用 s 代替 $s_i(t)$ 来表示小区总资源. 为评估全网内所有小区硬负载的均衡程度, 使用如下的 Jain 公平指数^[18]

$$\xi(t) = \frac{(\sum_{i \in \mathcal{N}} \rho_i(t))^2}{|\mathcal{N}| (\sum_{i \in \mathcal{N}} \rho_i^2(t))}. \quad (4)$$

上式中 $|\mathcal{N}|$ 表示小区集合 $\mathcal{N}$ 的势, 即小区集合内元素的个数. 从该定义式可以看出, 全网内对于所有 CBR 用户的负载均衡指示函数 $\xi(t)$ 的取值范围位于 $[1/|\mathcal{N}|, 1]$. 并且一个较大的 $\xi(t)$ 意味着小区间更均衡的硬负载分布. 当小区间达到完全均衡的硬负载分布时, 即每个小区有相同的硬负载时, 负载均衡指示函数 $\xi(t)$ 取值为 1. 对于全网内所有 CBR 用户而言, 负载均衡的目标就是在每个时刻 t 最大化该负载均衡指示函数 $\xi(t)$.

2.4 针对全网内 BE 用户的总效用函数

使用 $R_{i,m}(t)$ 来表示在时刻 t , BE 用户 m 在小区 i 内的可得吞吐量. 并用 $U_m(R_{i,m}(t))$ 来表示与该吞吐量对应的效用函数, 该效用函数一般为凹函数^[19,20], 则全网内所有 BE 用户的总效用函数为

$$\Psi(t) = \sum_{i \in \mathcal{N}} \sum_{m \in \mathcal{B}} U_m(I_{i,m}(t) R_{i,m}(t)). \quad (5)$$

则对于全网内所有 BE 用户而言, 负载均衡的目的就是最大化该总效用函数.

3 问题构造

本节提出了针对以上网络及相应目标函数的优化问题, 即在每个负载均衡周期 t , 同时最大化负载均衡指示函数 $\xi(t)$ 和总效用函数 $\Psi(t)$. 由于这两个变量都取决于小区和用户之间的匹配关系 $I_{i,k}(t)$, 则该问题可由如下的包含限制条件的多目标优化问题来表述

$$\max [\xi(t), \Psi(t)]^T \quad (6)$$

$$\text{s.t.} \quad \sum_{k \in \mathcal{K}} I_{i,k}(t) w_{i,k}(t) \leq s, \quad \forall i \in \mathcal{N}, \quad (7)$$

$$\sum_{i \in \mathcal{N}} I_{i,k}(t) = 1, \quad \forall k \in \mathcal{K}, \quad (8)$$

$$\sum_{i \in \mathcal{N}} I_{i,c}(t) R_{i,c}(t) \geq \theta_c, \quad \forall c \in \mathcal{C}, \quad (9)$$

式 (6) 表明了问题目标是同时最大化 $\xi(t)$ 和 $\Psi(t)$ 这两个目标函数, 是多目标优化问题的标准定义方式. 限制条件 (7) 表明一个小区中所有用户占用的资源之和不能够超过小区总资源 s , 限制条件 (8) 表明每个用户在某时刻 t 有且只有一个小区为之提供服务, 限制条件 (9) 说明每个 CBR 用户 c 的保障速率需求 θ_c 在其服务小区内需要被严格满足.

对于一个多目标优化问题, 最直观的方法是将多目标整合为单目标, 如将多个目标进行线性加权求和得到一个单一的目标函数^[21,22], 再对该目标函数进行优化. 在本文中, 由于这个两个目标函数值有不同的量级, 因此很难设计针对不同目标的加权系数并评估由此带来的对网络性能的影响. 而对于最新的基于进化算法的多目标优化算法^[23,24], 需要一个中心控制单元参与计算, 并需要得到所有用户的 QoS 和带宽效率信息, 这将给系统带来很大的开销. 除此之外, 控制单元需要较多的计算时间来达到 Pareto 最优解. 并且, 3GPP LTE 网络采用扁平的网络结构, 并没有类似 UMTS 系统中 RNC 的中心控制单元, 切换判决需要由每个 eNodeB 独自做出, 并要足够迅速来跟踪网络的动态变化, 同时用户状态信息交互带来的开销需要尽可能的少. 因此, 负载均衡切换必须是分布式, 实时且低复杂度低开销的. 下面给出一个满足上述所有要求的启发式算法结构.

4 算法结构

本节提出一个用以解决上述多目标优化问题的算法结构. 由于该算法结构按负载均衡周期进行, 为叙述方便, 接下来的分析省略了时间符号 t .

4.1 QoS 保证的混合调度

实际系统中, 高优先级用户的 QoS 需求将被优先且严格保证. 因此, 该算法结构首先将资源分配给小区内所有 CBR 用户来满足它们的 QoS 需求. 由于 BE 用户没有优先级的区别, 较好的调度算法应当同时兼顾效率与公平, 即在最大化吞吐量的同时保证用户间资源分配的公平性, 所以将剩余资源平均分配给小区内所有 BE 用户.

由于系统采用合适的机会调度算法给小区内所有用户提供服务, 因此需要考虑由于调度产生的多用户分集增益. 对于小区 i 中速率需求为 θ_c 的 CBR 用户 c , 给其分配的时频资源 $w_{i,c}$ 为

$$w_{i,c} = \left\lceil \frac{\theta_c}{e_{i,c} G_{i,c}} \right\rceil. \quad (10)$$

上式中 $e_{i,c}$ 为 CBR 用户 c 在小区 i 中由式 (2) 定义的平均带宽效率. $G_{i,c}$ 为 CBR 用户 c 在小区 i 中由于调度产生的多用户分集增益, 其与系统采用的调度方案及参与竞争的用户数相关, $\lceil \cdot \rceil$ 为上取整操作. 则小区 i 中所有 CBR 用户占用的资源, 即小区 i 的硬负载为 $s_i^c = \sum_{c \in C} I_{i,c} w_{i,c}$, 小区 i 中留给所有 BE 用户的剩余资源为 $s_i^b = s - s_i^c$.

由于小区内所有 BE 用户公平分配剩余资源 s_i^b , 任一用户 b 可得资源 $w_{i,b} = S_i^b / Y_i^b$, Y_i^b 为此时小区 i 中所有 BE 用户的数目. 则对于小区 i 中 BE 用户 b , 其实际可得速率为

$$R_{i,b} = e_{i,b} w_{i,b} G_{i,b}. \quad (11)$$

上式中 $e_{i,b}$ 为 BE 用户 b 在小区 i 中由式 (2) 定义的平均带宽效率. $G_{i,b}$ 为 BE 用户 b 在小区 i 中的多用户分集增益.

下面给出 $G_{i,c}$ 和 $G_{i,b}$ 的计算方式. 对于任一用户 k , 当其为 CBR 用户时, 初始分配资源 $w_{i,k} = \lceil \theta_c / e_{i,c} \rceil$, 再将其依次带入 (12), (13) 和 (10) 3 式中迭代计算, 得到最终的稳定值; 当其为 BE 用户时, $w_{i,k} = s_i^b / Y_i^b$.

已知 $w_{i,k}$ 后, 就可以知道对于任一用户 k , 小区中的竞争用户数 Num_k 为

$$\text{Num}_k = \left\lceil \frac{s}{w_{i,k}} \right\rceil. \quad (12)$$

上式中 $\lceil \cdot \rceil$ 为上取整操作. 由于基于用户自身累积速率分布的调度方法被证明在用户固定可得资源数时效率最高^[25], 采用该方法为系统调度方法. 并当假设所有用户经历 Rayleigh 快衰落且速率与信干噪比呈线性对应关系时, $G_{i,k}$ 为^[25]

$$G_{i,k} = \sum_{z=1}^{\text{Num}_k} \frac{1}{z}. \quad (13)$$

对于小区中所有 BE 用户, 由于每一用户的可得资源相等, 所以它们有相同的多用户分集增益.

4.2 CBR 用户切换条件

对于小区 i 中的 CBR 用户 c , 出于负载均衡的目的将其切换至某个相邻小区 j 应当提升负载均衡指示函数 ξ . 用 $\xi_{i,c}$ 和 $\xi_{j,c}$ 来分别表示负载均衡切换前和切换后的负载均衡指示函数, 很显然应满足 $\xi_{i,c} < \xi_{j,c}$. 由于将 CBR 用户 c 由其源小区 i 切换至目标小区 j 不会影响这两个小区中其他 CBR 用户的负载, 所以只需考虑切换用户 c 对负载均衡指示函数 ξ 带来的影响. 由于负载均衡切换选择用户大多为小区边缘用户, 它们在源小区和目标小区中所占负载大致相当, 所以可假设 $\xi_{i,c}$ 和 $\xi_{j,c}$ 的分子相同, 则 $\xi_{i,c} < \xi_{j,c}$ 结合 (4) 式可得

$$\rho_i^2 + \rho_j^2 > \left(\rho_i - \frac{w_{i,c}}{s} \right)^2 + \left(\rho_j + \frac{w_{j,c}}{s} \right)^2 \Rightarrow \frac{w_{i,c}(2s_i^c - w_{i,c})}{w_{j,c}(2s_j^c + w_{j,c})} > 1. \quad (14)$$

定义 $\mu_{i,j,c}^{\text{cbr}} = w_{i,c}(2s_i^c - w_{i,c}) / w_{j,c}(2s_j^c + w_{j,c})$ 为将 CBR 用户 c 由其源小区 i 切换至目标小区 j 带来的负载均衡切换增益, 该增益值应当大于 1, 否则此切换会带来负载均衡指示函数下降的负增益. 而如果很多 CBR 用户同时进行负载均衡切换, 可能导致较多的乒乓切换. 因此, 每负载均衡周期内每个小区只选择一个最优的用户, 即可得到最大负载均衡切换增益的用户进行切换, 如下所示

$$c^* = \arg \max_{c \in C, I_{i,c}=1} \mu_{i,j,c}^{\text{cbr}}. \quad (15)$$

我们引入门限 $\text{thr}_{\text{cbr}} (> 1)$ 来控制 CBR 用户的负载均衡切换并降低可能的乒乓切换. 只有满足 $\mu_{i,j,c^*}^{\text{cbr}} > \text{thr}_{\text{cbr}}$ 时, 用户 c^* 才可进行负载均衡切换.

4.3 BE 用户切换条件

对于小区 i 中的 BE 用户 b , 出于负载均衡的目的将其切换至其目标小区 j 应当提升 (5) 式中的总效用函数 Ψ . 用 $\Psi_{i,b}$ 和 $\Psi_{j,b}$ 来分别表示负载均衡切换前和切换后的总效用函数, 则有 $\Psi_{i,b} < \Psi_{j,b}$. 当假设全网内所有 BE 用户的效用函数为相同以 $a (> 1)$ 为底的对数函数时, $\Psi_{i,b}$ 和 $\Psi_{j,b}$ 分别为

$$\Psi_{i,b} = \log_a \left\{ \left[\left(\prod_{m=1,2,\dots,Y_i^b} e_{i,m} \right) \cdot (w_{i,\text{BE}} G_{i,\text{BE}})^{Y_i^b} \right] \left[\left(\prod_{n=1,2,\dots,Y_j^b} e_{j,n} \right) \cdot (w_{j,\text{BE}} G_{j,\text{BE}})^{Y_j^b} \right] \right\}, \quad (16)$$

$$\Psi_{j,b} = \log_a \left\{ \left[\left(\prod_{m=1,2,\dots,Y_i^b, m \neq b} e_{i,m} \right) \cdot (w'_{i, \text{BE}} G'_{i, \text{BE}})^{Y_i^b - 1} \right] \left[\left(\prod_{n=1,2,\dots,Y_j^b} e_{j,n} \right) \cdot e_{j,b} \cdot (w'_{j, \text{BE}} G'_{j, \text{BE}})^{Y_j^b + 1} \right] \right\}. \quad (17)$$

上述两式中 m, n 分别表示小区 i, j 中 BE 用户. 由于小区中每个 BE 用户的可得资源及其相应的多用户分集增益相等, 分别用 $w_{i, \text{BE}}$ 和 $G_{i, \text{BE}}$ 来表示切换前小区 i 中每个 BE 用户的可得资源和多用户分集增益, $w'_{i, \text{BE}}$ 和 $G'_{i, \text{BE}}$ 来表示切换后小区 i 中每个 BE 用户的可得资源和多用户分集增益. 此时有 $w_{i, \text{BE}} = s_i^b / Y_i^b$, $w'_{i, \text{BE}} = s_i^b / (Y_i^b - 1)$, $w_{j, \text{BE}} = s_j^b / Y_j^b$, $w'_{j, \text{BE}} = s_j^b / (Y_j^b + 1)$, 相应的 $G_{i, \text{BE}}$, $G'_{i, \text{BE}}$, $G_{j, \text{BE}}$, $G'_{j, \text{BE}}$ 分别由 $w_{i, \text{BE}}$, $w'_{i, \text{BE}}$, $w_{j, \text{BE}}$, $w'_{j, \text{BE}}$ 根据 (12) 和 (13) 两式决定. 为简化计算复杂度, 当小区中用户较多时, 根据多用户分集增益的计算方式我们可近似认为 $G'_{i, \text{BE}}$ 等于 $G_{i, \text{BE}}$; 同理, $G'_{j, \text{BE}}$ 等于 $G_{j, \text{BE}}$.

则 $\Psi_{i,b} < \Psi_{j,b}$ 结合 (16) 和 (17) 两式及以上说明, 可简化推得

$$\begin{aligned} e_{i,b} (w_{i, \text{BE}} G_{i, \text{BE}})^{Y_i^b} (w_{j, \text{BE}} G_{j, \text{BE}})^{Y_j^b} &< e_{j,b} (w'_{i, \text{BE}} G_{i, \text{BE}})^{Y_i^b - 1} (w'_{j, \text{BE}} G_{j, \text{BE}})^{Y_j^b + 1} \\ \Rightarrow e_{i,b} w_{i, \text{BE}} G_{i, \text{BE}} \left(\frac{Y_i^b - 1}{Y_i^b} \right)^{Y_i^b - 1} &< e_{j,b} w'_{j, \text{BE}} G_{j, \text{BE}} \left(\frac{Y_j^b}{Y_j^b + 1} \right)^{Y_j^b}. \end{aligned} \quad (18)$$

当小区 i 和 j 中 BE 用户较多时, 上式中 $(\frac{Y_i^b - 1}{Y_i^b})^{Y_i^b - 1}$ 和 $(\frac{Y_j^b}{Y_j^b + 1})^{Y_j^b}$ 都可以近似为相同的常数 e^{-1} , 则 $\Psi_{i,b} < \Psi_{j,b}$ 最终演化为

$$e_{i,b} w_{i, \text{BE}} G_{i, \text{BE}} < e_{j,b} w'_{j, \text{BE}} G_{j, \text{BE}}, \quad (19)$$

式 (19) 说明当小区中有较多的 BE 用户时, 切换用户 b 不会对相应的两个小区中其他 BE 用户的总效用函数产生显著影响, 负载均衡切换增益只取决于 BE 用户 b 自身的吞吐量增益. 与 CBR 用户的负载均衡切换相似, 定义 $\mu_{i,j,b}^{\text{be}} = e_{j,b} w'_{j, \text{BE}} G_{j, \text{BE}} / e_{i,b} w_{i, \text{BE}} G_{i, \text{BE}}$ 为将 BE 用户 b 由其源小区 i 切换至目标小区 j 带来的负载均衡切换增益, 该增益值同样应当大于 1. 同 CBR 用户类似, 每负载均衡周期内每小区只选择一个最优用户, 即可得到最大负载均衡切换增益的用户, 如下所示

$$b^* = \arg \max_{b \in B, I_{i,b}=1} \mu_{i,j,b}^{\text{be}}. \quad (20)$$

同样引入门限 $\text{thr}_{\text{be}} (> 1)$ 来控制 BE 用户的负载均衡切换并降低可能的乒乓切换. 只有满足 $\mu_{i,j,b^*}^{\text{be}} > \text{thr}_{\text{be}}$ 时, 用户 b^* 可进行负载均衡切换.

4.4 新呼叫准入控制

对于新产生的 CBR 业务用户 c , 如其要求接入的小区 i 有足够多的时频资源来满足它的 QoS 要求, 则用户 c 才可被接入小区 i , 如下所示

$$s^{\text{cmax}} - s_i^c > w_{i,c}. \quad (21)$$

上式中 s^{cmax} 为在每个负载均衡周期中, 为保障一定的公平性, 每个小区可以给其内所有 CBR 用户分配资源的最大百分比. 如没有此限制, 小区中所有资源将被 CBR 用户占用.

对于所有新产生的 BE 业务用户, 在接入时没有限制.

5 数值仿真与结果

本节评估了同时变化小区 1 中两类业务用户达到率时, 对不同业务类型组合进行负载均衡切换时对于系统性能产生的影响. 评估的系统性能指标为对于全网 CBR 用户的负载均衡指示函数, CBR 用户新呼叫阻塞率, 对于全网 BE 用户的总效用函数, 繁忙小区中 BE 用户最低 5% 吞吐量和全网 BE 用户总吞吐量.

5.1 仿真参数设置

本文参考的仿真网络为图 1 所示的 7 蜂窝小区, 相邻 eNodeB 间距为 130 m. 所有 eNodeB 有相等的最大传输功率 38 dBm, 相同的 10 MHz 带宽, 这与 3GPP 推荐的仿真场景相似 [26]. 并采用了 wrap-around 技术来避免边界效应 [27]. 将每小区中 CBR 用户占用资源最大百分比 s^{cmax} 设置为小区总资源的 80% 来保证 BE 用户可以得到小区中部分资源. 对于全网内所有 BE 用户, 采用相同的效用函数 $U_m(\cdot) = \log_{10}(\cdot)$.

为了得到实际的仿真结果, 本文采用了动态模型来评估前述的算法结构. 网络中任一小区 i 分别以到达率为 λ_i^c 和 λ_i^b 的 Poisson 分布来动态产生新的 CBR 和 BE 用户, 它们的位置在小区内是均匀分布的. 所有用户的呼叫保持时间满足均值为 100 s 的指数分布, 则每个小区中 CBR 和 BE 用户的平均数目取决于达到率和保持时间. 全网内所有 CBR 用户的保障速率要求在 64~256 kbps 间均匀分布. 为区分不同小区间的负载, 设置小区 1 为繁忙小区, 其中的 CBR 和 BE 用户有相等的可变到达率, 而将其他 2~7 小区中的 CBR 和 BE 用户到达率都设置为固定的 0.2 用户/秒. 负载均衡周期的选择需要折中考虑信令开销和算法性能增益 (通常情况下该周期越短, 则算法性能越好, 但同时信令开销也越大), 本文选择负载均衡周期为 1 s 来进行实际仿真. 对于 thr_{cbr} 和 thr_{be} 这两个负载均衡切换切换门限的选择, 通常该门限值越小, 则越容易进行负载均衡切换, 但同时由于切换带来的开销 (如信令交互和延时) 会较多, 产生乒乓切换的概率会较大, 故选择两类切换门限都为 1.5 举例仿真.

5.2 仿真结果

为表述方便, 接下来用 N/A, BE/LB, CBR/LB 和 CBR+BE/LB 来分别表示两类用户都不进行负载均衡切换 ($\text{thr}_{\text{be}} = \text{thr}_{\text{cbr}} = +\infty$), 只对全网内 BE 用户进行负载均衡切换 ($\text{thr}_{\text{be}} = 1.5$, $\text{thr}_{\text{cbr}} = +\infty$), 只对全网内 CBR 用户进行负载均衡切换 ($\text{thr}_{\text{be}} = +\infty$, $\text{thr}_{\text{cbr}} = 1.5$) 和对两类用户都进行负载均衡切换 ($\text{thr}_{\text{be}} = \text{thr}_{\text{cbr}} = 1.5$) 这 4 种负载均衡切换场景.

图 2 显示的是随着小区 1 中两类用户到达率提升, 4 种场景下负载均衡指示函数 ξ 的变化情况. 从图中可以发现, 随着 CBR 用户到达率的提升, N/A 场景下负载均衡指示函数 ξ 逐步下降. 这是由于随着小区 1 中 CBR 用户的增多, 使得其相对于其他小区的硬负载越来越高, 导致了全网硬负载分布愈加的不均衡. 从图中还可以发现, 只对 BE 用户进行负载均衡切换的场景 BE/LB 与不进行负载均衡切换的 N/A 场景, 其曲线是重合的, 这说明 BE 用户的负载均衡切换操作对于全网 CBR 用户相关的负载均衡指示函数 ξ 没有影响. 相类似, CBR/LB 与 CBR+BE/LB 也是重合的, 说明只有 CBR 用户的负载均衡切换操作才会对该性能指标产生影响. 通过计算得出, 对于 CBR 用户进行负载均衡切换操作可以使得负载均衡指示函数 ξ 平均上升 6.76%.

图 2 小区 1 不同到达率下的负载均衡指示函数 ξ
Figure 2 Load balance index ξ with various arrival rates of cell 1

图 3 小区 1 不同到达率下的 CBR 用户新呼叫阻塞率
Figure 3 New call blocking rate of CBR users with various arrival rates of cell 1

图 3 显示了小区 1 中两类用户不同到达率时, 4 种负载均衡切换场景下的 CBR 用户的新呼叫阻塞率. 这张图显示了图 2 中负载均衡指示函数提升带来的增益. 首先, 随着用户到达率的上升, 4 种场景下的 CBR 用户新呼叫阻塞率都随之上升. 与图 2 类似, N/A 和 BE/LB 的曲线重合, 且由于都不进行 CBR 用户的负载均衡切换, 导致该类用户的新呼叫阻塞率较高. 而 CBR/LB 和 CBR+BE/LB 由于对 CBR 用户进行负载均衡切换操作, 将繁忙小区 1 中的部分边缘 CBR 用户切换至相邻低负载小区, 其 CBR 用户新呼叫阻塞率比前两者平均低 74.28%, 在较低到达率时 (0.3, 0.35 用户/秒), 几乎可以完全避免新 CBR 用户的阻塞情况.

图 4 显示了不同到达率情况下, 4 种负载均衡切换场景下与全网 BE 用户相关的总效用函数的变化情况. 首先, 随着用户到达率的提高, 总效用函数 ψ 也随之上升, 这是由于随着 BE 用户数的增多, 位于小区中心的 BE 用户数也较多, 对于总效用函数有较大的增益; 同时, 用户数增多, 对应的多用户分集增益也随之上升, 使得每个用户的效用随之增大. 我们可以发现 N/A 在每种到达率情况下都是最低的, 由于它没有进行任何用户的负载均衡切换. 而 BE/LB 大于 N/A, 这完全符合 BE 用户负载均衡切换的目标, 即增大总效用函数. 同时, CBR/LB 也大于 N/A, 这是因为虽然只对 CBR 用户进行负载均衡切换, 但会使得剩余分配给 BE 用户的资源在小区间得到一定程度的平衡, 从而对 BE 用户的负载均衡产生正面的影响, 一定程度上相当于对 BE 用户进行了负载均衡切换操作. 最终, CBR+BE/LB 在每种到达率的情况下都是最高的, 说明对两类用户都进行负载均衡切换时, 对于 BE 用户的负载均衡效果最好, 即总效用函数最高.

图 5 显示了不同负载均衡切换场景下, 小区 1 中 BE 用户最低 5% 吞吐量随用户到达率的变化情况. BE 用户最低 5% 吞吐量是指全小区所有 BE 用户中, 吞吐量由小到大累加至总吞吐量 5% 时的用户吞吐量之和, 这个指标可以很好地表征边缘 BE 用户的性能. 该图中的 4 条曲线都随着用户到达率的升高而降低, 这是由于小区 1 中到达了较多的 CBR 用户, 它们需要占用更多的资源, 留给 BE 用户的资源越来越少, 其边缘用户可以分到的资源也进一步减少, 故其最低 5% 吞吐量随之下降. 可以发现, N/A 在所有到达率情况下都是最低的, 这是由于其不进行任何负载均衡切换操作. 在所有到达率情况下, BE/LB 平均高于 N/A 情况 6.21%, 这就是 BE 用户负载均衡切换的效果, 即在增加总效用函数的同时, 提升了繁忙小区中边缘 BE 用户的吞吐量. 还可以发现, 只进行 CBR 用户负载均衡切换的 CBR/LB 也平均高于 N/A 9.35%, 这是由于 CBR 用户的负载均衡切换, 实质上使得各小区间剩余给

图 4 小区 1 不同到达率下的 BE 用户总效用函数 ψ

Figure 4 Total utility function ψ of BE users with various arrival rates of cell 1

图 5 小区 1 不同到达率下的小区 1 中 BE 用户最低 5% 吞吐量

Figure 5 Lowest 5% throughput of BE users in cell 1 with various arrival rates of cell 1

图 6 小区 1 不同到达率下的全网 BE 用户总吞吐量

Figure 6 Total throughput of BE users with various arrival rates of cell 1

BE 用户的资源更加均衡, 从另一个方面达到一定程度上 BE 用户的负载均衡, 这与图 4 中 CBR/LB 的总效用函数高于 N/A 也是对应的. 而 CBR+BE/LB 在所有到达率情况下都是最高的, 由于这两类用户的负载均衡切换都可以对 BE 用户的负载均衡产生正面的影响.

图 6 显示了不同到达率情况下, 4 种不同负载均衡切换场景下全网 BE 用户总吞吐量的变化情况. 可以看到 N/A 在所有到达率情况下其总吞吐量都是最高的. 而 BE/LB, CBR/LB 和 CBR+BE/LB 分别平均低于 N/A 时 0.61%, 1.58% 和 1.91%, 是以上所有正增益的代价.

综合以上结果, 可以发现 BE/LB 相对于 N/A, 以牺牲 BE 用户总吞吐量为代价, 提升了总效用函数和繁忙小区中边缘 BE 用户的性能, 这与参考文献 [11] 中的结论一致, 同时对 CBR 用户的负载均衡没有任何影响. CBR/LB 相对于 N/A, 提高了全网负载均衡指示函数, 并降低了 CBR 用户新呼叫阻塞率, 对 BE 用户也有部分的负载均衡效果, 而 CBR+BE/LB 在 CBR/LB 的基础上, 进一步对 BE 用户进行了负载均衡, 达到两类用户同时负载均衡的最优效果.

6 结论

本文研究了 LTE 系统中考虑不同 QoS 业务要求的负载均衡算法. 首先将问题建模为一个多目标优化问题, 接下来分析了它的复杂度和开销, 提出了一个实时的低复杂度低开销的分布式解决方案, 包括 QoS 保障的混合调度, QoS 感知的负载均衡切换和新呼叫准入控制. 之后, 评估了不同负载均衡方案对于网络性能的影响. 仿真结果显示对于两类用户的负载均衡算法可以显著降低 CBR 用户的新呼叫阻塞率, 并以总吞吐量略微降低为代价大幅提升繁忙小区中边缘 BE 用户的吞吐量.

参考文献

- 1 Shen J, Suo S, Quan H, et al. 3GPP Long Term Evolution: Principle and System Design. Beijing: POSTS & TELECOM PRESS, 2008. 1–31 [沈嘉, 索世强, 全海洋, 等. 3GPP 长期演进 (LTE) 技术原理与系统设计. 北京: 人民邮电出版社, 2008. 1–31]
- 2 3GPP. Evolved Universal Terrestrial Radio Access (E-UTRA); LTE physical layer; General description (Release 10). 3GPP TS 36.201. 2010
- 3 3GPP. Self-configuring and self-optimizing network (SON) use cases and solutions(v9.3.1). 3GPP TR 36.902. 2011
- 4 Tolli A, Barbancho I, Gomez J, et al. Intra-system load balancing between adjacent GSM cells. In: Proceedings of VTC. Jeju Island: IEEE, 2003. 393–397
- 5 Zhu H, Buot T, Nagaike R, et al. Load balancing in WCDMA systems by adjusting pilot power. In: Proceedings of WPMC. Hawaii: IEEE, 2002. 936–940
- 6 Jiang H, Rappaport S S. CBWL: A new channel assignment and sharing method for cellular communication systems. IEEE Trans Veh Technol, 1994, 43: 313–322
- 7 Kahwa T J, Georganas N D. A hybrid channel assignment scheme in large-scale, cellular-structured mobile communication systems. IEEE Trans Commun, 1978, 4: 432–438
- 8 Das S K, Sen S K, Jayaram R. A novel load balancing scheme for the tele-traffic hot spot problem in cellular networks. Wirel Netw, 1998, 4: 325–340
- 9 Eklundh B. Channel utilization and blocking probability in a cellular mobile telephone system with directed retry. IEEE Trans Commun, 1986, 34: 329–337
- 10 Tonguz O K, Yanmaz E. The mathematical theory of dynamic load balancing in cellular networks. IEEE Trans Mob Comput, 2008, 7: 1504–1518
- 11 Son K, Chong S, Veciana G. Dynamic association for load balancing and interference avoidance in multi-cell networks. IEEE Trans Wirel Commun, 2009, 8: 3566–3576
- 12 Wang H, Ding L, Wu P, et al. Dynamic load balancing and throughput optimization in 3gpp lte networks. In: Proceedings of IWCMC. Caen: IEEE, 2010. 939–943
- 13 Sang A, Wang X, Madhian M, et al. Coordinated load balancing, handoff/cell-site selection, and scheduling in multi-cell packet data systems. Wirel Netw, 2008, 14: 103–120
- 14 Das S, Viswanathan H, Rittenhouse G. Dynamic load balancing through coordinated scheduling in packet data systems. In: Proceedings of IEEE INFOCOM, San Francisco, 2003. 786–796
- 15 Bu T, Li L, Ramjee R. Generalized proportional fair scheduling in third generation wireless data networks. In: Proceedings of IEEE INFOCOM, Barcelona, 2006. 1–12
- 16 3GPP. Technical Specification Group Services and System Aspects; Policy and charging control architecture (Release 11). 3GPP TS 23.203. 2012
- 17 Wang X, Giannakis G B, Marques A G. A unified approach to QoS-guaranteed scheduling for channel-adaptive wireless networks. Proc IEEE, 2007, 95: 2410–2431
- 18 Jain R, Chiu D, Hawe W. A quantitative measure of fairness and discrimination for resource allocation in shared systems. Digital Equipment Corp. Technical Report DEC-TR-301, 1984
- 19 Gao N, Wang X. Optimal subcarrier-chunk scheduling for wireless OFDMA systems. IEEE Trans Wirel Commun, 2011, 10: 2116–2123

- 20 Chiang M, Low S, Calderbank A, et al. Layering as optimization decomposition. *Proc IEEE*, 2007, 95: 255–312
- 21 Erfania T, Utyuzhnikova S V. Directed search domain: a method for even generation of the Pareto frontier in multi-objective optimization. *Eng Optimiz*, 2011, 43: 467–484
- 22 Mayorga M E, Subramanian R. Factoring environmental concerns in supply chain decision making. *Int J Busin Syst Res*, 2010, 4: 469–487
- 23 Zhou A, Qu B, Li H, et al. Multiobjective evolutionary algorithms: A survey of the state of the art. *Swarm Evol Comput* 2011, 1: 32–49
- 24 Hirsch C, Shukla P K, Schmeck H. Variable Preference Modeling Using Multi-Objective Evolutionary Algorithms. *Comput Sci*, 2011, 6576: 91–105
- 25 Wang H, Ding L, Pan Z, et al. QoS Guaranteed Call Admission Control with Opportunistic Scheduling. In: *Proceedings of IEEE GLOBECOM*, Houston, 2011. 1–5
- 26 3GPP. Physical layer aspects for evolved Universal Terrestrial Radio Access (UTRA)(Release 7). 3GPP TR 25.814. 2006
- 27 Kim S J, Wang X. Hierarchical approach to interference mitigation in multi-cell downlink orthogonal frequency-division multiple-access networks with low feedback. *IET Commun*, 2011, 5: 660–666

QoS guaranteed dynamic load balancing algorithm in 3GPP LTE networks

WANG Hao^{1,2*}, LI ZhiHang¹, PAN ZhiWen¹, YOU XiaoHu¹ & WU Ping²

¹ *National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China;*

² *Department of Engineering Sciences, Uppsala University, Uppsala 75237, Sweden*

*E-mail: hao_wang@seu.edu.cn

Abstract In this paper, dynamic load balancing with different quality-of-service (QoS) requirements is investigated in 3GPP long-term evolution (LTE) networks. Load unbalancing among neighboring cells often has different negative impacts on users with different QoS demands. For users with minimum rate requirements, the load unbalancing results in high new call blocking rate, while for users with no rate guarantees, the throughput of boundary users in an overloaded cell often significantly decreases. Furthermore, the different load unbalancing problems are coupled with each other so that it is difficult to analyze the problem in a uniform manner. To deal with this issue, a multi-objective optimization problem is proposed. The objectives in the problem are load balancing index of users with QoS requirements and the total utility function of users without QoS requirements, and the constraints are physical resource limits and QoS demands. Then the complexity of the problem is analyzed, and a real-time distributed algorithm framework with low complexity/overhead is proposed, which includes a QoS guaranteed hybrid scheduling scheme, handover of users with and without QoS requirements, and a call admission control algorithm. Extensive simulations are conducted and the results show that the proposed algorithm framework leads to significantly better load balancing so as to yield the decrement of new call blocking rate of users with QoS requirements, and the increment of throughput of boundary users with only a bit degradation of total throughput of users without QoS requirements.

Keywords load balancing, 3GPP LTE, quality of service (QoS), multi-objective optimization, load balance index, total utility function

WANG Hao was born in 1983. He received the B.S. and M.S. degree in communication and information system from Nanjing University of Science and Technology in 2006 and 2008, respectively. Currently, he is working towards his Ph.D. degree in School of Information Science and Engineering at Southeast University. His research interests include inter-cell optimization, load balancing, opportunistic scheduling, coordinated multi-point transmission and energy efficiency.

LI ZhiHang was born in 1987. He received the B.S. degree in telecommunications engineering from Xidian University in 2009, and the M.S. degree in School of Information Science and Engineering at Southeast University in 2011, respectively. Currently, he is working towards his Ph.D. degree in School of Information Science and Engineering at Southeast University. His research interests include load balancing, opportunistic scheduling and interference coordination.

YOU XiaoHu was born in 1962. He received his M.S. degree and Ph.D. degree from Southeast University, Nanjing, in electrical engineering in 1985 and 1988, respectively. Since 1990, he has been working with National Mobile Communications Research Laboratory at Southeast University, where he held the rank of Professor. His research interests include mobile communication systems, signal processing and its applications. He was the Premier Foundation Investigator of the China National Science Foundation. From 1999 to 2002, he was the Principal Expert of the C3G Project. From 2001 to 2006, he was the Principal Expert of the National 863 Future Project. Dr. You now is the Chairman of IEEE Nanjing Section. He was selected as IEEE Fellow due to his contributions to development of mobile communications in China.

WU Ping was born in 1959. He received B.S. degree in electrical and electronic engineering in Nanjing Aeronautical & Astronomical University, M.S. and Ph.D. degrees in electrical and electronic engineering in Xi'an Jiaotong University, in 1982, 1987 and 1991, respectively. After receiving Ph.D. degree, he worked as a postdoctor in Fraunhofer Institute for Biomedical Engineering, St. Ingbert, Germany for two years (1992–1994). Since Oct. 1994 he has been with the Division of Signals and Systems in Department of Engineering Sciences, Uppsala University, Sweden, and is an Associate Professor. His current research interests include wireless mesh and sensor networks, OFDM and MIMO OFDM, cross-layer design and optimization, network coding, and signal processing with applications in telecommunications.