

异构蜂窝网络中 QoS 感知的负载平衡方案设计

钱叶旺^{1,3} 周天清^{2,3} 杨绿溪³

(1. 池州学院机电工程学院, 池州, 247000; 2. 华东交通大学信息工程学院, 南昌, 330013; 3. 东南大学信息科学与工程学院, 南京, 211189)

摘要: 鉴于异构蜂窝网络(Heterogeneous cellular network, HCN)的自身特性, 传统最强信号接入(小区选择)方式已不再适合, 新型小区选择方案急需引入。不同于传统接入方案, 新型方案应具备平衡各类基站负载的能力。为反映用户资源消耗水平与实现平衡网络负载的目的, 设计了一个服务质量(Quality of service, QoS)感知的负载平衡方案。该方案以资源消耗量作为基站负载, 且同时拟合了用户服务质量需求。最终, 该方案被规划为网络加权效益最大化问题。从规划问题的形式来看, 该问题为非线性、混合整数优化问题, 因此求解其最优解富于挑战性(尤其针对大规模问题)。为解决该问题, 尝试利用对偶分解法设计了一个分布式算法。仿真结果表明, 相比于最强信号接入和区域拓展接入, 提出的接入方案具有更高的负载平衡水平与更低的呼叫阻塞概率。

关键词: 异构蜂窝网; 负载平衡; 小区选择; 干扰管理

中图分类号: TN911 文献标志码: A

Design of QoS-Aware User Association for Load Balancing in Heterogeneous Cellular Networks

Qian Yewang^{1,3}, Zhou Tianqing^{2,3}, Yang Luxi³

(1. Mechanical and Electronic Engineering Department, Chizhou University, Chizhou, 247000, China; 2. School of Information Engineering, East China Jiaotong University, Nanchang, 330013, China; 3. School of Information Science and Engineering, Southeast University, Nanjing, 211189, China)

Abstract: Considering that the conventional associations may not be appropriate due to the special characteristics of heterogeneous cellular networks (HCNs), some novel associations that can balance the network loads should be designed. To reflect the resource consumption levels of users and balance loads, a QoS-aware user association is proposed for load balancing in HCNs. In the association, the load is defined as the amount of consumed resources, which can reflect the resource consumption levels of users. Finally, this scheme is formulated as a network-wide weighted utility maximization problem. That is a nonlinear mixed-integer optimization problem, and obtaining its optimal solutions is very challenging (especially for large-scale problems). To solve the problem, a distributed algorithm is designed using dual decomposition. Numerical results show that, compared with the best power association and range expansion association, the proposed scheme has higher load balancing levels and lower call blocking probabilities.

Key words: heterogeneous cellular networks; load balancing; user association; interference management

引 言

为消除传统蜂窝网络中的覆盖漏洞、提升热点区域网络吞吐量^[1],传统宏蜂窝网络中引入了诸多异构元素(如微基站、皮蜂窝、飞蜂窝和中继等),形成了异构蜂窝网络(Heterogeneous cellular network, HCN)^[2]。尽管 HCN 可带来极大频谱效益,但其用户接入(小区选择)问题富于挑战性且往往影响系统性能。由于 HCN 中各类基站具有明显不同的发射功率,一些仅聚焦于用户接收信号强度的接入方案(如最大信干噪比接入,最大可达速率接入、最近基站接入与最佳信道质量接入等)已然不适用。当这些方案被应用于 HCN 时,即使负载均匀分布于网络覆盖区域,各类基站间的负载水平仍表现为极度不平衡。具体地,绝大部分用户选择高功率基站,少数用户选择低功率基站^[3]。显然,为充分开发新型网络架构的潜力,具有转载能力小区选择方案的设计有必要且意义重大。注意到,转载能力是指将传统最强信号接入中超载基站的用户转入轻负荷基站的能力,而这些转入其他基站的用户则被称为转载用户。

目前,就网络负载定义而言,已有的负载均衡方案可分为如下 3 类。

第 1 类,定义接入用户数为基站负载。在 HCN 中,绝大多数负载均衡方案基于此定义。偏离因子法(区域拓展)接入^[4]作为通用负载均衡方案,它为低功率基站添加一个偏离/补偿进而理论上提升其发射功率。在此理论功率基础上,该方案再依据最强信号接入准则执行小区选择,从而达到拓展小基站(低功率基站)覆盖面积的目的。虽然该法实行简单,但由于闭式最优偏离因子/补偿往往难以捕捉,因此无法保证最佳系统性能。为避免搜索最佳参数,其他负载均衡方案渐已融入 HCN 中^[5-7],如最小路径损耗接入^[7]等。

第 2 类,定义小区数据流为基站负载。根据业务数据流的分布情况,Son 等^[8]联合用户接入与基站开关操作设计了一个折衷能耗与时延的接入方案,并为其设计了两类贪婪算法。Siomina 等^[9]认为小区间负载具有耦合性,并于此基础上尝试最大化网络负载均衡指数,从而达到平衡小区间数据流的目的。

第 3 类,定义用户消耗资源量(物理资源块数)为基站负载。根据用户资源消耗情况,Wang 等^[10]设计了一个小区选择方案以联合优化负载均衡水平与系统容量。此外,Wang 等^[11]对其优化问题中衡量负载均衡水平的标准加以简化,使其原问题更易处理。基于 Wang 等的工作,Li 等^[12]进一步联合考虑小区内负载平衡与小区间负载平衡。不难发现,尽管这些方案具有较好的思路,但在实际应用中却难以确定模型中的最佳网络参数,如优化问题中所涉及的权重等。

在上述 3 类定义中,第 1 类定义有利于小区选择方案的设计,但有时可能因无法反映小区数据流量(或资源消耗量)的变化而导致接入方案的性能偏差;第 2 类定义能反映网络数据流量的变化,但因系统模型过于复杂而使得接入算法的设计富于挑战性;第 3 类定义能很好地捕捉用户服务质量(Quality of service, QoS)需求,但要在严格保证用户 QoS 需求下设计有效的接入算法同样富于挑战性。目前,基于第 2,3 类定义的小区选择方案多数采用集中算法,而这类算法复杂度往往偏高。此外,在第 2,3 类定义下,针对 HCN 系统的负载均衡方案更是凤毛麟角。因此,在 HCN 中,如何设计针对第 2,3 类定义的负载均衡方案,并为其开发简单、有效的接入算法仍是一个富有意义的话题^[13]。

鉴于此,本文基于第 3 类负载定义,提出了一类新型 QoS 感知的负载均衡方案。该方案能折衷系统容量与负载,即适当地降低用户可达速率以尽量平衡基站负载。此处,“降低用户可达速率”是指由于用户选择信道质量略差的轻负荷基站而导致可达速率下降。不同于 Zhou 等的工作^[14],本文采用动态步长,在此基础上给出了相应的收敛性与复杂度分析,并给出了更为充分的仿真分析(仿真参数及指标有所变动)。

1 系统模型与问题描述

异构蜂窝网络包含规则与不规则的两种布置。在规则布置中,宏基站(Macro base station, MBS)按

传统蜂窝结构布置;在不规则布置中,宏基站则被随机撒点。鉴于接入规则本身特性,无论采用何种结构,所设计接入方案较其他方案的优越性不变。不失一般性,仅考虑由 MBS 与皮蜂窝(Pico base station, PBS)组成的两层异构蜂窝网络^[1],其中 MBS 依传统蜂窝结构布置,而用户与 PBS 则被随机撒入宏小区,其拓扑结构如图 1 所示。

定义包含 MBS 与 PBS 的基站集为 N , 用户集为 K , 其基分别为 $N = |N|$ 和 $K = |K|$ 。根据长期演进(Long term evolution, LTE)技术关于物理层的描述^[15], 相邻 12 个子载波组成一个物理资源块(Physical resource block, PRB), 该资源块具有 180 kHz 的带宽, 且为可分配给用户的最小单元。那么, 用户 $k \in K$ 接收来自基站 $n \in N$ 的信噪比(Signal-to-interference-plus-noise-ratio, SINR)为

$$\text{SINR}_{nk} = \frac{p_n g_{nk}}{\sum_{j \in N/(n)} p_j g_{jk} + W_{\text{prb}} N_0} \quad (1)$$

式中: p_n 为基站 n 在单个 PRB 上的发射功率; g_{nk} 代表基站 n 与用户 k 之间的信道增益; W_{prb} 为 PRB 的带宽; N_0 为噪声功率谱密度。注意到, 本文假设每个基站对所用频带进行等功率分配。

当给定功率 p_n 时, 用户 k 接收来自基站 n 的可达速率(单位为 Kbps)为 $r_{nk} = W_{\text{prb}} \log_2(1 + \text{SINR}_{nk})$, 且 $r = \{r_{nk}, \forall n \in N, \forall k \in K\}$ 。

在规划问题之前, 需要根据用户实际速率需求计算接入每个基站时所消耗的资源量。具体地, 若用户 k 接入基站 n , 则资源消耗量为 $s_{nk} = \frac{d_k}{r_{nk}}$, 其中 d_k 为用户 k 的实际速率需求。当了解用户资源消耗情况后, 为便于方案设计, 给出了一些相关定义。

定义 1: 基站的有效负载定义为该基站服务用户所消耗资源块数。数学上, 基站 n 的负载定义为 $y_n = \sum_{k \in K} x_{nk} s_{nk}$, 其中 x_{nk} 为接入指示变量。 $x_{nk} = 1$ 表示用户 k 接入(选择)基站 n , $x_{nk} = 0$ 则表示用户 k 未接入基站 n 。

定义 2: 用户负载效益定义为用户可达速率与基站有效负载之比。数学上, 当基站 n 的有效负载为 y_n 时, 用户 k 接入基站 n 时所能获取的负载效益为 $E_{nk} = \frac{r_{nk}}{y_n}$ 。用户负载效益既能反映用户信道质量, 又能反映基站负载水平。

为突显用户资源消耗水平与实现负载平衡, 尝试设计了一类 QoS 感知的小区选择方案。代替最强信号接入方案中的准则, 该方案以资源消耗量作为接入指标。基于此, 设计了一个最大化网络加权效益的负载平衡方案, 该方案可规划为

$$\begin{aligned} & \max_x \sum_{n \in N} \sum_{k \in K} x_{nk} s_{nk} U_{nk}(E_{nk}) \\ & \text{s. t. } C_1: \sum_{k \in K} x_{nk} s_{nk} \leq S_n \quad \forall n \in N \\ & C_2: \sum_{n \in N} x_{nk} = 1 \quad \forall k \in K \\ & C_3: x_{nk} \in \{0, 1\} \quad \forall n \in N, \forall k \in K \end{aligned} \quad (2)$$

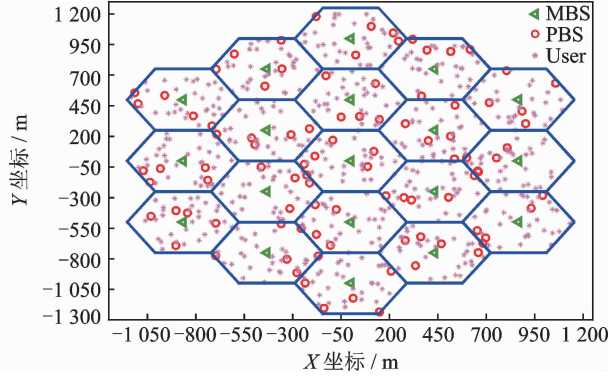


图 1 两层异构蜂窝网络

Fig. 1 Two-tier HCNs

式中: $\mathbf{x} = \{x_{nk}, \forall n \in N, \forall k \in K\}$; $U_{nk}(\cdot)$ 为用户 k 接入基站 n 时所获得的效益; S_n 为基站 n 的资源总量; 约束 C_1 表明服务用户消耗资源量不能超过基站资源总量; 约束 C_2 表明用户仅能选择单一基站。在该式中, 效用函数 $U_{nk}(\cdot)$ 为连续可微、单调递增且严格凹的函数。

为保证用户公平性, 引入对数效用函数, 问题(2)则可转化为

$$\begin{aligned} \max_{\mathbf{x}} \sum_{n \in N} \sum_{k \in K} x_{nk} s_{nk} \left\{ \log r_{nk} - \log \sum_{j \in K} x_{nj} s_{nj} \right\} \\ \text{s. t. } C_1, C_2, C_3 \end{aligned} \quad (3)$$

观察问题(3)不难发现, 当 r_{nk} 增加时, $\log r_{nk} - \log \sum_{j \in K} x_{nj} s_{nj}$ 增加, 而 s_{nk} 则下降。那就意味着, 接入问题(3)并非最大可达速率接入, 而是具有转载能力的接入方案。

利用 $\sum_{k \in K} x_{nk} s_{nk} = y_n$, 问题(3)可进一步转化为

$$\begin{aligned} \text{Max}_{\mathbf{x}, \mathbf{y}} F(\mathbf{x}, \mathbf{y}) &= \sum_{n \in \{N\}} \sum_{k \in \{K\}} x_{nk} s_{nk} \log r_{nk} - \sum_{n \in \{N\}} y_n \log y_n \\ \text{s. t. } C_2, C_3 \\ C_4: \sum_{k \in \{K\}} x_{nk} s_{nk} &= y_n \quad \forall n \in \{N\} \\ C_5: 0 < y_n &\leq S_n \quad \forall n \in \{N\} \end{aligned} \quad (4)$$

其中 $\mathbf{y} = \{y_n, \forall n \in N\}$ 。容易发现, 当接入指示变量 \mathbf{x} 从整数域 $\{0, 1\}$ 松弛到分数域 $[0, 1]$ 时, 问题(4)则为凸优化问题。

2 算法设计与分析

为解决凸形式的问题(4), 梯度下降法不失为有效手段。为搜索其最优解, 需收集全局网络信息, 那么需涉及集中式算法。鉴于集中式算法复杂度偏高, 这里采用有效分布式算法设计。不同于集中式算法, 分布式算法往往无需基站间的协作处理。

2.1 分布式算法设计

观察问题(4)不难发现, 约束 C_4 具有耦合性。为便于问题的求解, 去耦(Decoupling)操作不可或缺。为此, 引入了一个对偶变量 $\mu = \{\mu_n, n \in N\}$ 对该约束进行松弛。于是, 关于此约束的拉格朗日函数为

$$L(\mathbf{x}, \mathbf{y}, \mu) = \sum_{n \in N} \sum_{k \in K} x_{nk} s_{nk} \log r_{nk} - \sum_{n \in N} y_n \log y_n + \sum_{n \in N} \mu_n \left(y_n - \sum_{k \in K} x_{nk} s_{nk} \right) \quad (5)$$

不难得出, 问题(4)的对偶函数为

$$G(\mu) = \begin{cases} \text{Max}_{\mathbf{x}, \mathbf{y}} L(\mathbf{x}, \mathbf{y}, \mu) \\ \text{s. t. } C_2, C_3, C_5 \end{cases} \quad (6)$$

且其对偶问题为

$$\text{Min}_{\mu} G(\mu) \quad (7)$$

鉴于问题(7)为非耦合问题, 主变量 \mathbf{x} 和 \mathbf{y} 可以分开求解。根据对偶分解法的思想^[16], 问题(7)可分解为

$$H(\mu) = \begin{cases} \text{Max}_{\mathbf{x}} \sum_{n \in N} \sum_{k \in K} x_{nk} s_{nk} \{ \log r_{nk} - \mu_n \} \\ \text{s. t. } C_2, C_3 \end{cases} \quad (8)$$

$$I(\mu) = \begin{cases} \text{Max}_{\mathbf{y}} \sum_{n \in N} y_n \{ \mu_n - \log y_n \} \\ \text{s. t. } C_5 \end{cases} \quad (9)$$

为获取对偶问题(7)的最优解,梯度下降法^[17]不失为一种有效手段。具体地,拉格朗日乘子 μ 沿着对偶函数(6)的负梯度方向 $(-\nabla G(\mu))$ 进行迭代更新。为获此梯度,主问题的最优解应事先获知,即求解问题(8,9)。鉴于问题(8,9)的形式,可对其采用并行处理方式。

至于问题(8),它可简化成

$$n^* = \arg \operatorname{Max}_{n \in N} \{s_{nk} (\log r_{nk} - \mu_n)\} \quad \forall k \in K \quad (10)$$

该规则意味着:任意用户 $k \in K$ 选择某基站以最大化其自身效益 $\bar{\omega}_{nk} (\log r_{nk} - \mu_n)$ 。显然,该过程发生于用户端,详细描述可参考算法 1,其中 t 表示第 t 次迭代。

算法 1 用户 k 端算法

- (1) If $t=0$
- (2) 根据来自所有基站的导频信号估计 r 。
- (3) Else
- (4) 用户 k 接收来自所有基站的广播信息 μ , 并根据规则(10)选择基站 n^* 。
- (5) 反馈接入信息 $x_{n^*k}=1$ 给基站 n^* 。
- (6) End If

至于问题(9),可根据其 Karush-Kuhn-Tucker (KKT)^[17] 条件获得基站最优负载,即

$$y'_n = \operatorname{Min} \{ \exp(\mu'_n - 1), S_n \} \quad \forall n \in N \quad (11)$$

式中 $\exp(z)$ 为关于 z 的指数函数。

当最优 x 和 y 给定时,拉格朗日乘子则可以梯度下降法进行更新,即

$$\mu_n^{t+1} = \mu'_n - \xi^t \left(y'_n - \sum_{k \in K} s_{nk} x_{nk}^t \right) \quad \forall n \in N \quad (12)$$

式中 ξ^t 为第 t 次迭代时用于更新乘子 μ_n 的步长,该步长遵循 Bertsekas 规则^[17]。具体地

$$\xi^t = \underline{\omega}^t (G(\mu^t) - \tilde{G}^t) / \|G(\mu^t)\|^2 \quad 0 < \underline{\omega} \leq \omega^t \leq \bar{\omega} < 2 \quad (13)$$

式中 $\underline{\omega}$ 和 $\bar{\omega}$ 为标量; \tilde{G}^t 为对偶问题(7)在第 t 次迭代时最优值 G^* 的一个估计量,即

$$\tilde{G}^t = \operatorname{Min}_{i=0, \dots, t} G(\mu^i) - \epsilon^t \quad (14)$$

通过式(13,14)更新乘子 μ 后, ϵ 执行如下更新

$$\epsilon^{t+1} = \begin{cases} \alpha \epsilon^t & G(\mu^{t+1}) \leq \tilde{G}^t \\ \operatorname{Max} \{ \beta \epsilon^t, \underline{\epsilon} \} & \text{其他} \end{cases} \quad (15)$$

式中 $\alpha \geq 1$, $\beta < 1$ 且 $\underline{\epsilon}$ 为 ϵ 的正下界。

为获取动态步长 ξ^t , \tilde{G}^t 的计算是关键。从式(14)来看,它需要前 t 次迭代的对偶函数值 $G(\mu^i)$, $i=0, \dots, t$ 。在寻找 ξ^t 的过程中,所涉及的系列标量可在给定值域内依算法收敛速度与精度进行合理设置。

不难发现,乘子 μ 的更新过程在于搜索对偶问题(7)的最佳估计 \tilde{G}^t , 该估计量比以往每次迭代目标值 $G(\mu)$ 都小 ϵ^t , 即使得 \tilde{G}^t 满足式(14)。在第 t 次迭代中,当最佳估计达到时,需要增加 ϵ^t ($\alpha > 1$) 或维持不变 ($\alpha = 1$); 当最佳估计未能获得时,需要降低 ϵ^t 。

算法 2 基站 n 端算法

- (1) If $t=0$
- (2) 初始化步长 ξ^t 和 μ^t 。
- (3) Else
- (4) 对任意用户 $k \in K$, 接收来自该用户的反馈信息 $x'_{nk}=1$ 。
- (5) 根据式(11)计算负载 y_n^{t+1} , 并利用规则(12)更新乘子 μ_n^{t+1} 。

(6) 广播信息 μ_n^{t+1} 至所有用户。

(7) End If

从 y 与 μ 的更新规则来看,不难发现,它们发生于基站端,即组成了基站端算法,其详细描述可参考算法 2。

从规则(12)来看,容易发现,它有些特别意义。在该规则中,乘子 μ_n 可被视为所有用户与基站 n 之间的消息,并可解释为后者的服务费用(价格),此费用依赖于负载分布。对于基站 n ,如果 $\sum_{k \in K} s_{nk} x_{nk}$ 被视为服务需求量,而 y_n 代表服务供应量,那么上述费用将折衷供需。具体地,如果需求 $\sum_{k \in K} s_{nk} x_{nk}$ 高于供应 y_n ,则费用将上涨,反之则反。在小区选择过程中,超载基站将提升其服务费用以减少资源消耗,而轻负荷基站则降低其服务费用以提升资源消耗。前者可阻止一些用户接入,后者则吸引更多用户,从而达到负载均衡的目的。

2.2 算法收敛性与复杂度分析

完整的小区选择算法应交替执行用户与基站端算法,其收敛性可通过下面的理论加以证明。

定理 1 如果存在 $\epsilon > 0$ 且 $G^* > -\infty$, 则

$$\inf_t G(\mu^t) \leq G^* + \epsilon \quad (16)$$

式中 G^* 定义为对偶问题(7)的最优值。

证明:函数 $G(\mu)$ 的一阶导为

$$\frac{\partial G}{\partial \mu_n}(\mu) = y_n(\mu) - \sum_{k \in K} s_{nk} x_{nk}(\mu) \quad (17)$$

在实际网络系统中,撒入宏小区的用户通常有限,于是 $\sum_{k \in K} s_{nk} x_{nk}$ 和 y_n 应为有界函数,进而使得 ∂G 是有界函数。显然,问题(7)满足文献[14]中推论 6.3.6 的必要条件。于是,定理 1 可通过该推论加以证明。

接下来,针对所设计的算法进行复杂度分析。在式(12)中,乘子 μ 的调整完全由基站分布式实现,且只需非常少的局部信息。在每次迭代中,用户与基站端算法的计算复杂度分别为 $O(N)$ 和 $O((t+1)NK)$ 。后者计算复杂度主要在于 \tilde{G}^t 的计算。当采用静态步长时,基站端算法的计算复杂度为 $O(K)$ 。然而,同样在静态步长下,集中式算法每次迭代的计算复杂度为 $O(NK)$,远高于分布式算法的计算复杂度。

为突出所设计算法的优越性,仍调查了算法实现时的信息交互情况。在分布式算法中,任意基站 n 广播其自身的服务费用 μ_n 给所有用户,而任意用户 k 仅反馈其自身接入信息 $x_{nk} = 1$ 给期望接入的基站 n^* 。显然,在每次迭代中,分布式算法的交互信息量为 $N+K$ 。不同于分布式算法,集中式算法的交互信息量与 NK 成比例关系。

3 仿真结果与分析

假定 MBS 间的距离为 1 000 m,每个宏小区布置 5 个 PBS,MBS 和 PBS 发射功率分别为 46 与 30 dBm,系统带宽为 10 MHz,且噪声功率谱密度为 -174 dBm/Hz。此外,MBS 与 PBS 的路径损耗分别为 $128.1 + 37.6 \log d_{nk}$ 和 $140.7 + 36.7 \log d_{nk}$ [19-21],其中 d_{nk} 为用户 k 与基站 n 间的距离,单位为 km。阴影衰弱则取为标准差为 8 dB 的对数正态阴影。

为突出所设计方案(QoS-aware association, QAA)的优越性,仿真中引入了其他通用方案作比较。这些方案包括最强功率接入(Best power association, BPA) [5] 和有偏信号强度接入(Biasing channel gain

association, BCGA)^[18]。前者属于传统最强信号接入,后者则属于负载均衡接入方案。因此,方案 QAA 和 BCGA 具有转载能力,而 BPA 却不具备。此外,方案 QAA 具有 QoS 感知能力,而方案 BPA 和 BCGA 却不具备。

为探索 QoS 需求对接入性能的影响,仿真中引入了两类调度方案:(1) 最大实际速率优先(Maximal practical rate first, MPRF)方案,即每个基站从接入用户队列中按实际速率从大到小的顺序调度用户;(2) 最大可达速率优先(Maximal achievable rate first, MARF)方案,即每个基站从接入用户队列中按可达速率从大到小的顺序调度用户。通过引入调度算法,能更进一步、更为明显地展示方案 QAA 所带来的负载均衡增益。

为测量网络负载均衡水平,引入了 Jain 公平性指数,即

$$J = \left(\sum_{n \in N} \kappa_n \right)^2 / \left(N \sum_{n \in N} \kappa_n^2 \right) \quad (18)$$

式中: $\sum_{k \in K} z_{nk} s_{nk} = \kappa_n$ 为基站 n 的负载(资源消耗量); z_{nk} 为调度指示(若基站 n 调度用户 k , 则 $z_{nk} = 1$, 否则该指示为 0)。鉴于该指数能很好地反映网络负载均衡水平,它亦可称之为负载均衡指数(Load balancing index, LBI)。由于小区选择过程中的负载均衡水平直接影响用户调度过程中的负载均衡水平,且用户调度时需考虑基站资源限制,因此结合调度算法分析负载分布更为实际。此外,由于方案 QAA 在资源限制下实现负载均衡,因此其负载均衡水平更为真实,结合调度算法更能体现方案 QAA 的优势,且这一优势并不依赖于调度算法。

针对不同调度方案,图 2 展示了各类接入方案的网络负载均衡水平随用户数(Number of users, NU)的变化情况,其中 NU 是指宏小区内的用户数。随着 NU 的增加,各基站有更多的机会选择用户,从而有更多的机会实现基站间负载均衡。因此,LBI 随 NU 的增加而提高。在方案 BPA 中,绝大多数用户接入至高功率基站 MBS 而使得该类基站资源消耗极大,且少部分用户选择低功率基站 PBS 而使得该类基站资源消耗偏少。因此,相比于其他负载均衡方案,方案 BPA 具有最低 LBI。不同于方案 BPA,方案 BCGA 通过为低功率基站添加一个偏离来拓展其覆盖范围,因此迫使方案 BPA 中超载基站 MBS 的用户转入邻近的轻负荷基站 PBS。通过进一步平衡方案 BPA 中的负载,方案 BCGA 达到了较之更高的 LBI。从方案 QAA 的接入规则来看,用户并非接入至拥有最佳信道质量的基站(高功率基站),而是选择折衷资源消耗和可达速率的基站,即表现为转载能力。通过平衡基站间的负载,方案 QAA 达到了最高的 LBI。尽管方案 BCGA 可通过调整偏离因子改善基站负载水平,但调整过程很难保证最优。换言之,若偏离因子过小,则超载基站减压不明显;若偏离因子过大,则原信道质量较差的基站反成超载基站。鉴于用户调度严格受限于基站资源,无论何种调度方案,调度资源均不能超过基站本身所有。因此,两类调度方案下 LBI 可能几乎相同。

针对不同调度方案,图 3 展示了各类接入方案的网络负载均衡水平随用户实际速率需求(Practical rate, PR)变化的情况。如图 3 所示,随着用户 PR 的增加,方案 BPA 和 BCGA 的负载均衡水平开始增加而后下降。在低速率域,随着 PR 的增加,调度资源越发接近各类基站资源供应量,因此负载均衡水平上升。当 PR 增加到一定量时,再度增加速率需求将导致更多资源片断(资源块)无法满足用户需求。由于高功率基站用户往往具有高可达速率,因此高功率基站调度资源块数通常多于低功率基站,且这一差距随 PR 的增加而增大。因此,方案 BPA 和 BCGA 的 LBI 会出现下降趋势。不同于方案 BPA 和 BCGA,方案 QAA 起始时 LBI 就很高,即各类基站资源消耗趋近平衡。因此,方案 QAA 的 LBI 不存在上长趋势。但同其他方案一样,当增加 PR 时,方案 QAA 的 LBI 应呈现出下降趋势。同时注意到,起始 LBI 越高,LBI 增长过程越短。其原因在于,原本高的 LBI 缩短了调度资源趋近饱和的过程。

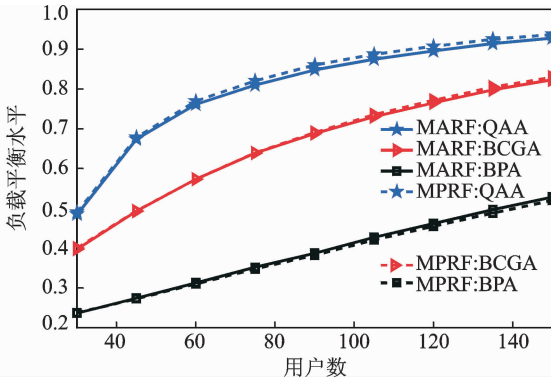


图2 用户数对网络负载均衡水平的影响
Fig. 2 Impact of NU on LBI

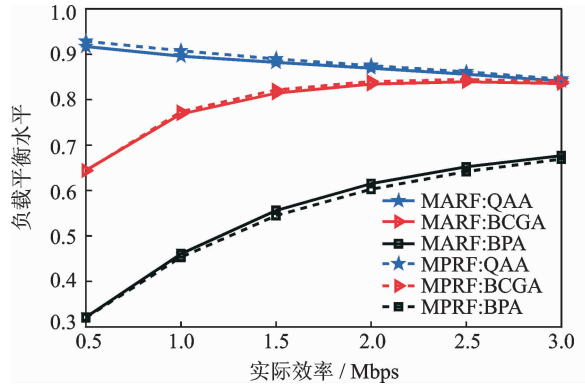


图3 用户实际速率需求对网络负载均衡水平的影响
Fig. 3 Impact of user's PR on LBI

为突显负载均衡增益,引入了呼叫阻塞概率(Call blocking probability,CBP),即 $P = 1 - \frac{u}{K}$, 其中 u 为从基站接入队列(由选择该基站的所有用户构成的队列)中所调度的用户数。显然,呼叫阻塞主要是由于资源不足而导致用户无法调度。

针对不同调度方案,图4展示了各类接入方案的用户CBP随NU的变化情况。正如之前所述,在方案BPA中,绝大多数用户接入至MBS,仅有少许用户选择PBS。显然,在方案BPA中,很多接入至MBS的用户由于基站没有足够资源而无法被调度,而PBS由于接入用户过少而使得其资源无法被充分利用,从而造成了资源的极大浪费。因此,相比于其他具有转载能力的方案,方案BPA应具有更高的CBP。尽管方案QAA以损耗资源(提升资源消耗)来迫使基站间实现负载均衡,但它通过平衡负载实现网络资源的充分利用、提升用户满意度。因此,同方案BPA相比,方案QAA具有更低的CBP。此外,根据接入规则,不难发现,方案BPA未考虑用户QoS需求和资源限制,但方案QAA却有所涉及。因此,方案QAA获得了较方案BPA更低的CBP。尽管方案BCGA可通过调整偏离因子而使得CBP下降,但调整过程很难保证最优,且转载过程无法保证转载用户满足基站资源限制。换言之,若偏离因子过小,则因超载基站减压不明显而使得许多用户未能被调度;若偏离因子过大,则因原信道质量较差的基站反成超载基站而使得更多用户不能被调度。此外,在有限资源条件下,相比于调度方案MPRF,方案MARF可能调度更多用户,从而使得调度方案MARF可能拥有较方案MPRF更低的CBP。

针对不同调度方案,图5展示了各类接入方案的用户CBP随用户PR的变化情况。随着用户PR的

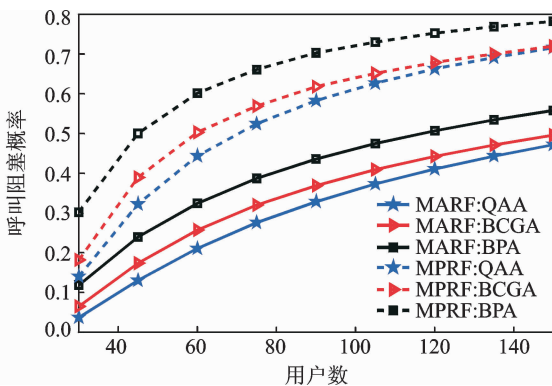


图4 用户数对呼叫阻塞概率的影响
Fig. 4 Impact of NU on CBP

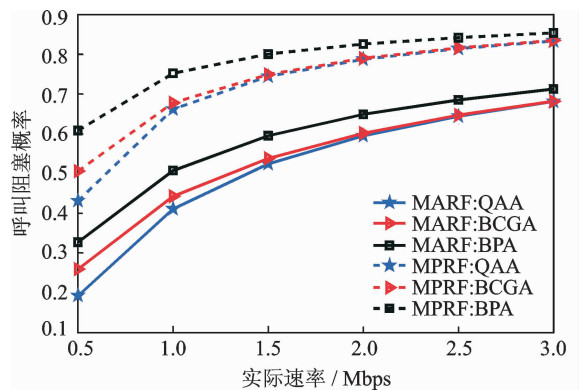


图5 用户实际速率需求对呼叫阻塞概率的影响
Fig. 5 Impact of user's PR on CBP

增加,为支持用户正常通信,基站不得不分配更多的资源给用户。然而,每个基站的资源毕竟有限,因此用户 CBP 随用户实际需求的增加而提高。此外,随着用户 PR 的提升,资源需求量逐渐增加,可接入用户数越来越少。当用户 PR 足够大后,各类方案接入用户数可能就非常接近,即少数几个高可达速率用户被调度。因此,各方案的 CBP 越接近。

针对方案 QAA 的接入算法,图 6 描绘了其收敛情况。根据仿真结果,不难发现,所设计的接入算法具有较高的收敛速率,非常适用于实际应用。

4 结束语

针对 HCN 系统,设计了一类 QoS 感知的小区选择方案以实现网络负载平衡。在该方案中,基站资源消耗量被视为其负载。同时,该方案还融入了用户服务质量需求,可反映用户资源消耗水平。最终,该方案被规划为网络范围的加权效益最大化问题。鉴于所规划问题为非线性、混合整数优化问题,其最优解的获取富于挑战性,尝试利用对偶分解法开发了一个有效的分布式算法。仿真结果表明,同其他接入方案相比,所设计的接入方案具有较高的负载平衡水平和更低的用户呼叫阻塞概率。此外,数值仿真仍显示:所设计的分布式算法具有较快的收敛速率,非常适合实际应用场景。未来研究工作可包括动态接入算法的开发、联合功控方案的设计和推广至上行系统。

参考文献:

- [1] Andrews J, Singh S, Ye Q, et al. An overview of load balancing in HetNets: Old myths and open problems[J]. *IEEE Wireless Communications*, 2014, 21(2): 18-25.
- [2] 王禄生, 王文浩, 汪晏如, 等. 基于 CRE 与 ABS 的异构蜂窝干扰协调方案[J]. *数据采集与处理*, 2016, 31(3): 473-481. Wang Lusheng, Wang Wenhao, Wang Yanru, et al. Interference coordination for heterogeneous cells based on CRE and ABS [J]. *Journal of Data Acquisition and Processing*, 2016, 31(3): 473-481.
- [3] 林臻, 燕雪峰. 一种具有负载平衡特性的容错技术[J]. *数据采集与处理*, 2012, 27(S1): 111-115. Lin Zhen, Yan Xuefeng. Fault-tolerant technology for load balancing [J]. *Journal of Data Acquisition and Processing*, 2012, 27(S1): 111-115.
- [4] Tang H, Peng J, Hong P, et al. Offloading performance of range expansion in picocell networks: A stochastic geometry analysis[J]. *IEEE Wireless Communications Letters*, 2013, 2(5): 511-514.
- [5] Fooladivanda D, Rosenberg C. Joint resource allocation and user association for heterogeneous wireless cellular networks[J]. *IEEE Transactions on Wireless Communications*, 2013, 12(1): 248-257.
- [6] Ye Q, Rong B, Chen Y, et al. User association for load balancing in heterogeneous cellular networks[J]. *IEEE Transactions on Wireless Communications*, 2013, 12(6): 2706-2716.
- [7] Khandekar A, Bhushan N, Ji T, et al. LTE-advanced: Heterogeneous networks[C]// *Wireless Conference*. Lucca, Italy: IEEE, 2010: 978-982.
- [8] Son K, Kim H, Yi Y, et al. Base station operation and user association mechanisms for energy-delay tradeoffs in green cellular networks[J]. *IEEE Journal on Selected Areas in Communications*, 2011, 29(8): 1525-1536.
- [9] Siomina I, Yuan D. Load balancing in heterogeneous LTE: Range optimization via cell offset and load-coupling characterization[C]// *IEEE International Conference on Communications*. 2012 (ICC 2012). [S. l.]: IEEE, 2012: 1357-1361.
- [10] Wang H, Ding L, Wu P, et al. Dynamic load balancing in 3GPP LTE multi-cell networks with heterogeneous services[C]// *5th International ICST Conference on Communications and Networking in China 2010*. Beijing, China: IEEE, 2010: 1-5.
- [11] Wang H, Ding L, Wu P, et al. QoS-aware load balancing in 3GPP long term evolution multi-cell networks[C]// *IEEE Inter-*

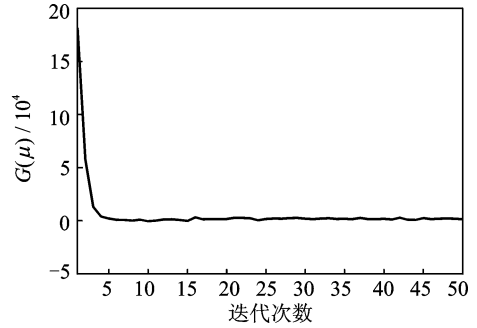


图 6 QAA 算法的收敛性

Fig. 6 Convergence of algorithm QAA

national Conference on Communications. Kyoto, Japan: IEEE, 2011:1-5.

- [12] Li Z, Wang H, Pan Z, et al. Dynamic load balancing in 3GPP LTE multi-cell fractional frequency reuse networks[C]//IEEE Vehicular Technology Conference. Quebec, Canada: IEEE, 2012:1-5.
- [13] 周一青,潘振岗,翟国伟,等. 第五代移动通信系统 5G 标准化展望与关键技术研究[J]. 数据采集与处理, 2015, 30(4): 714-724.
- Zhou Yiqing, Pan Zhengang, Zhai Guowei, et al. Standardization and key technologies for future fifth generation of mobile communication systems [J]. Journal of Data Acquisition and Processing, 2015, 30(4): 714-724.
- [14] Zhou T, Huang Y, Fan L, et al. Load-aware user association with quality of service support in heterogeneous cellular networks[J]. IET Communications, 2015, 9(4):494-500.
- [15] 3GPP, TS 36.201: LTE physical layer: general description (release 9). v9.1.0[EB/OL]. <http://www.3gpp.org>, 2010-03-02.
- [16] Xiao L, Johansson M, Boyd S. Simultaneous routing and resource allocation via dual decomposition[J]. IEEE Transactions on Communications, 2004, 52(7):1136-1144.
- [17] Bertsekas D. Convex optimization Theory[M]. [S. l.]: Athena Scientific, 2009.
- [18] Singh S, Andrews J. Joint resource partitioning and offloading in heterogeneous cellular networks[J]. IEEE Transactions on Wireless Communications, 2014, 13(2):888-901.
- [19] Zhou T, Jiang N, Liu Z, et al. Joint cell activation and selection for green communications in ultra-dense heterogeneous networks[J]. IEEE Access, 2018, 6:1894-1904.
- [20] Zhou T, Liu Z, Zhao J, et al. Joint user association and power control for load balancing in downlink heterogeneous cellular networks[J]. IEEE Transactions on Vehicular Technology, 2018, 67(3):2582-2593.
- [21] Zhou T, Liu Z, Qin D, et al. User association with maximizing weighted sum energy efficiency for massive MIMO-enabled heterogeneous cellular networks[J]. IEEE Communications Letters, 2017, 21(10):2250-2253.

作者简介:



钱叶旺(1971-),男,副教授,研究方向:MIMO 通信信号处理、异构蜂窝网中的负载平衡技术和能效通信技术等,Email: yewang_qian@163.com.



周天清(1983-),男,博士,研究方向:4G/5G 中的小区选择、干扰管理、能效优化,Email: zhoutian930@163.com.



杨绿溪(1964-),男,教授、博士生导师,研究方向:移动通信空时信号处理、协作通信、绿色通信、大规模多输入多输出通信、网络编码和无线局域网通信技术,Email: lxyang@seu.edu.cn.

(编辑:陈珺)