Proportional Fairness for Overlapping Cells in Wireless Networks

Bin Bin Chen
Department of Computer Science
National University of Singapore
Email: chenbinb@comp.nus.edu.sg

Mun Choon Chan
Department of Computer Science
National University of Singapore
Email: chanmc@comp.nus.edu.sg

Abstract— **Coordination of cellular base stations (BS) with overlapping coverage enables joint optimization of radio resource allocation in a multiple cell environment. This paper extends existing Proportional Fairness model for wired networks and single wireless cell, to the context of multiple (probably heterogeneous) wireless cells with overlapping coverage. The proposed fair allocation achieves both global Pareto optimality and inter-cell fairness (load balance). However, the ideal allocation is not practical as it requires a mobile station (MS) be simultaneously associated with multiple BSs. Instead, we use a simple GLS (Greedy Logarithmic Sum) scheme, which associates each new arrival MS with only one BS, to approximate the optimal allocation. Simulation result shows that GLS performs close to optimal scheme in a wide range of network settings.**

I. INTRODUCTION

Overlapping coverage of cellular base stations(BS) is a common phenomenon in wireless communication systems. For a particular radio access network, neighboring cells or sectors overlap with each other. In addition, deployment and interoperation of a wide array of wireless access networks, ranging from 3G network, WiMax, to 802.11 hotspots, opens the opportunity of overlapping coverage from BSs using heterogeneous radio access technologies. When neighboring (heterogeneous) BSs can coordinate with each other, and (multi-radio) MSs can flexibly associate with any one or even multiple of nearby BSs, it is important to study the radio resource allocation model for such multi-cell environment.

This paper extends existing proportional fairness framework for wired networks [4] and single wireless cell [3], to the context of multiple (probably heterogeneous) wireless cells with overlapping coverage. The proposed fair allocation achieves both global Pareto optimality and inter-cell fairness (load balance). However, the ideal allocation requires a multi-radio MS be simultaneously associated with multiple BSs. Instead, we show that a simple GLS (Greedy Logarithmic Sum) scheme achieves performance close to the optimal allocation while associating each MS to only one BS.

The rest of the paper is organized as follow. In Section II, we review related works. Section III presents the ideal proportional fairness model for multiple cells with overlapping coverage. Section IV describes the GLS scheme and some common heuristics. Performance of above schemes is evaluated in Section V. Finally, we conclude in Section VI.

II. RELATED WORKS

The most common understanding of fairness in networking literature is *max-min fairness*, which allocates rates as equal as possible [1]. Although max-min fairness is Pareto-optimal, it has been criticized to allocate too much resource to resource-inefficient requests, thus not suitable for wireless communication systems which are characterized by scarce radio resource. When some MSs in a cell use a lower bit rate than the others, the performance of all MSs is considerably degraded to almost the worst one [2]. Compared to max-min fairness, proportional fairness [4] strikes a good balance between overall system throughput and user fairness.

Definition(Proportional Fairness): A scheduling P is *proportionally fair* if and only if, for any feasible scheduling S, it satisfies:

$$\sum_{i \in I} \frac{R_i^{(S)} - R_i^{(P)}}{R_i^{(P)}} \leq 0 \qquad (1)$$

where I is the user set, and $R_i^{(P)}$ and $R_i^{(S)}$ are the average rates of user i allocated by scheduler P and S respectively. Proportional fairness favors resource-efficient requests more than max-min fairness, by allowing large sharing to increase further, if the sum of the proportional changes is non-negative.

[4] shows that proportional fair scheduling scheme in (1) is also the unique solution for a social utility optimization problem:

$$P = arg \max_S \sum_{i \in I} U_i(R_i^{(S)}) \qquad (2)$$

where the utility functions $U_i(\cdot)$ are logarithmic functions of the allocated bandwidth. Traffic that exhibits such utility function is often denoted as *elastic traffic*.

The utility maximization interpretation used in [4] suffers from the disadvantages that user utilities are only known in some qualitative sense. Instead, [6] considers bandwidth directly rather than abstract utility functions using Nash bargaining framework from cooperative game theory [7]. It is shown that proportional fairness is an *Nash Bargaining Solution (NBS)* out of all Pareto optimal points.

Proportional Fairness is implemented by opportunistic scheduling in Qualcomm's High Data Rate (HDR) system [3]. Instead of simultaneously serving multiple MSs and stabilizing every MS's received SINR through power control, HDR BS

transmits packet data to one single MS every Transmit Time Interval (TTI) using the maximum power. and scheduling priority is given to MS i^* with the highest data rate request relative to its received data rate averaged over a time window, according the equation:

$$i^* = arg \max_{i \in I} \frac{r_i}{R_i} \qquad (3)$$

where r_i is the instantaneous data rate. All MSs inform BS instantaneous Channel Quality Indication (CQI) through an uplink data rate request every TTI. R_i is the average data rate. On one hand, by considering instantaneous rate r_i, PF scheduler allocates resources opportunistically to the MS whose radio channel is experiencing a constructive multipath fade to achieve user diversity gain. On the other hand, by monitoring average throughput R_i, BS will raise the priority of the MS which has not been transmitted for some time. [3] shows that all MSs asymptotically get transmitted with the same fraction of time under some symmetric assumptions on their fading process. Most of successive research so far consider single cell case.

For multi-cell setting, [5] extends HDR opportunistic scheduling algorithm to multi-carrier system. Multi-carrier can be viewed as a special case of overlapping environment. The proposed algorithm is an opportunistic scheduling scheme asymptotically approaching the ideal proportional inter-carrier fairness allocation. After this paper is submitted, we notice the publication of [8], which takes a similar approach as us. [8] focuses on 3G networks, thus they model user diversity gain and single-BS association constraint in their optimization problem, which thus becomes NP-hard and inapproximable. Instead, we consider general wireless networks, which may consist of heterogeneous radio access technologies (thus opportunistic scheduling may not be implemented). Ideally we allow multi-BS association so that the problem can be processed using standard convex optimization techniques. In our simulation, we focus on the impact of different overlapping settings.

III. PROPORTIONAL FAIRNESS FOR OVERLAPPING CELLS

Let us first briefly review Kelly's proportional fairness model. In a network with a set J of resources, and C_j is the finite capacity of resource j, for $j \in J$. A route r is a non-empty subject of J, and R is the set of possible routes. $A_{jr} = 1$ if $j \in r$, i.e., resource j lies on route r, and $A_{jr} = 0$ otherwise. This defines a 0-1 matrix $A = (A_{jr}, j \in J, r \in R)$. A source-sink i can simultaneously use multiple paths, thus it is defined as a non-empty subset of R. I is the set of possible source-sinks. $H_{ir} = 1$ if $r \in i$, i.e., route r serves the source-sink i, and $H_{ir} = 0$ otherwise. This defines a 0-1 matrix $H = (H_{ir}, i \in I, r \in R)$. For simplicity and WLOG, each $r \in R$ maps to a single $i \in I$ such that $H_{ir} = 1$, i.e. every route r serves only a single source-sink i. Suppose that if a rate x_i is allocated to the source-sink i then this has utility $U_i(x_i)$ to i, so that the aggregate utility of rates $x = (x_i, i \in I)$ is $\sum_{i \in I} U_i(x_i)$. A flow pattern $y = (y_r, r \in R)$ supports the rates $x = (x_i, i \in S)$ if $Hy = x$,

so that the flows over all routes serving the source-sink i sum to the rate x_i. A flow pattern y is feasible if $y \geq 0$ and $Ay \leq C$, so that the flow over routes through resource j sum to not more than the capacity of resource j. To find the proportional fairness rates is to solve the optimization problem below by setting $U_i(x_i) = \ln(x_i)$.

ProportionalFair(U,H,A,C):

$$maximize \sum_{i \in I} U_i(x_i) \quad s.t. \quad Hy = x, Ay \leq C, y \geq 0 \quad (4)$$

The objective function is differentiable and concave and the feasible region is convex and compact; hence a unique solution x exists and can be calculated numerically with the maturing convex programming techniques (y is not necessary unique).

In wired networks, resource is characterized directly in terms of bandwidth (such as a router's interface forwarding speed), and a resource can serve all routes passing it with same efficiency. In contrast, wireless resource is characterized by, for example, time slot or frequency spectrum. To support per unit flow from different MSs, a single BS needs to use different amount of radio resource because of the location dependent and time varying channel condition. Thus, *the definition of matrix A in Kelly's model should be generalized from a 0-1 matrix to a matrix with arbitrary non-negative entries.* A_{jr} is set to the amount of radio resource BS j used to support per unit flow from route r. If MS is under poor channel condition, more radio resource is required to support per unit flow, i.e., A_{jr} is larger, while MS with good channel condition has smaller A_{jr}. If resource j doesn't lie on route r, still set $A_{jr} = 0$. A_{jr} is subjected to change because of mobility and channel variation, which makes resource allocation decision difficult. In practice, an estimation of time-averaging channel condition can be used for resource allocation decision, which can be updated dynamically when new estimation is available.

Above modification of matrix A doesn't change the nature of *ProportionalFair(U,H,A,C)*, hence standard convex programming techniques can still apply just as in Kelly's original model. To further simplify the formalization, we assume all sources are located in wired networks and all sinks are MSs. Only the last hop (BS to MS) of every route is bottleneck, thus the capacity constraint inequalities for resources in wired networks can be removed from *ProportionalFair(U,H,A,C)* without changing the feasible solution region. Given m BSs, n MSs, and L links between them, each BS corresponds to a resource, each MS corresponds to a sink-source, and each link corresponds to a route. Thus A is a $m \times L$ matrix, and H is a $n \times L$ matrix.

Figure 1 shows a simple example. There are two BSs $[M, N]$ with overlapping coverage, four MSs $[a, b, c, d]$, and 6 links $[(a, M), (a, N), (b, M), (b, N), (c, M), (d, N)]$. The weight of link is the amount of resource required to support per unit flow in this link (in comparison, wired network can be viewed as a special case of uniform weight). Thus:
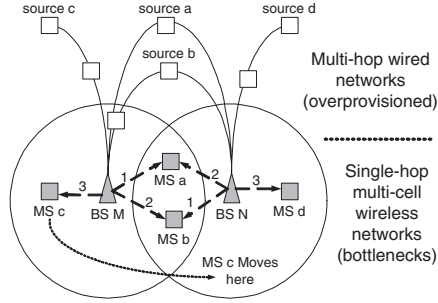
Fig. 1. A simple example

$$A = \begin{bmatrix} 1 & 0 & 2 & 0 & 3 & 0 \\ 0 & 2 & 0 & 1 & 0 & 3 \end{bmatrix}, H = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Both M and N have capacity 1. The optimal solution is: $x = [1/2, 1/2, 1/6, 1/6]'$, $y = [1/2, 0, 0, 1/2, 1/6, 1/6]'$. As the load is symmetrically distributed, it's observed that MS a is served totally over route (a, M), and MS b is served over route (b, N), both of which have better channel condition than their alternatives. *The solution of ProportionalFair(U,H,A,C) is global Pareto optimal.*

By considering proportional fairness in a global sense (among all MSs), the solution of *ProportionalFair(U,H,A,C)* also achieves inter-cell load balance. For example, if MS c moves from area covered only by BS M to area covered only by BS N, the traffic load becomes asymmetric, and BS N is more congested than BS M. Route (c, M) is replaced by (c, N) with weight still 3, then $A = \begin{bmatrix} 1 & 0 & 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 1 & 3 & 3 \end{bmatrix}$, while H is unchanged. The optimal solution becomes $x = [3/4, 3/8, 1/8, 1/8]'$, $y = [3/4, 0, 1/8, 1/4, 1/8, 1/8]'$. Notice that in this setting MS b is also served partially by BS M, because BS N is more congested than BS M. *The optimal solution adaptively distributes load from congested BS to other BSs.*

IV. GLS (Greedy Logarithmic Sum) scheme

The optimal solution for *ProportionalFair(U,H,A,C)* requires MSs to be simultaneously assigned to multiple BSs, which is hard to implement in practice. For example, software defined radio can only associate with one BS at a time. The complexity and instability of maintaining multiple association, as well as the increased power consumption also prohibit this approach.

In this section, we study resource allocation schemes which associate each MS with only one BS. Under this constraint, to maximize the logarithmic sum of all MSs' data rate, every single BS should maximize the logarithmic sum of data rate over MSs which have been assigned to it. Thus, every BS should allocate radio resource proportionally fair as in HDR. The remaining problem is to decide for each MS with which BS it should associate. As *ProportionalFair(U,H,A,C)* is to maximize the logarithmic sum of data rate over all MSs, a natural heuristic is to greedily associate arrival MS with the BS

which is selected so that the logarithmic sum of all MSs' data rate after the assignment is greedily maximized. We name this scheme as *GLS (Greedy Logarithmic Sum) scheme*. Formally *GLS* scheme assigns new arrival MS i to BS $j^* \in J_i$:

$$j^* = arg \max_{j \in J_i}(\ln \frac{R_{ij}}{|I_j|+1} + \sum_{n \in I_j}(\ln \frac{R_{nj}}{|I_j|+1} - \ln \frac{R_{nj}}{|I_j|}))$$

$$= arg \max_{j \in J_i}(\ln \frac{R_{ij}}{|I_j|+1} - |I_j| * \ln(1 + \frac{1}{|I_j|})) \quad (5)$$

$$= arg \max_{j \in J_i}(\ln R_{ij} - \ln \frac{(|I_j|+1)^{|I_j|+1}}{|I_j|^{|I_j|}}) \quad (6)$$

where J_i is the set of BSs covering MS i, I_j is the set of MSs associated to BS j before MS i arrives, and $|I_j|$ is its cardinality. R_{nj} is the *full data rate* of the wireless link between MS n and BS j. When $|I_j| = 0$, the second terms in both (5) and (6) are defined as zero.

The first term in (5) is the *allocated data rate* of MS i if it is associated with BS j, taking into account both channel condition (R_{ij}) and congestion ($|I_j|$). The second term in (5) takes fairness into consideration by accounting for the decrease in allocation in existing MSs, for which the only information required is the number of existing MSs. A selfish MS may choose to associate with BS $\bar{j} = arg \max_{j \in J_i} \ln \frac{R_{ij}}{|I_j|+1}$, i.e., greedily maximizes its individual allocated data rate regardless of the social utility. We denote this as *Non-Cooperative* scheme against *GLS* scheme in which MSs are cooperative. The difference between social utility of *Non-Cooperative* and *GLS* scheme after the decision is upper-bounded by the second term in (5), which itself is upper bounded by 1.

Equation (6) separates R_{ij} and $|I_j|$. Two common practices, *Best-Signal* scheme and *Least-Population* scheme, choose the first and second term as association criteria respectively. For *Best-Signal* scheme, MS i is greedily associated with BS $\hat{j} = arg \max_{j \in J_i} R_{ij}$. *Least-Population (LP)* scheme, instead, selects BS $\tilde{j} = arg \min_{j \in S_i} |I_j|$.

V. Simulation

A. Setup

Our simulation is based on a 10×4 *torus* topology where two *grids* of BSs are generated, and each grid consists of 5×2 BSs, as shown in Figure 2. Grid-A of BSs are centered at location (x, y), where both x and y are even integers. Their coverage areas are marked by dash lines. Instead, both coordinates of grid-B BSs' center are odd integers. Their coverage areas are circled by solid lines. Every BS covers a disk with $radius = \sqrt{2}$. The percentage of area covered by different number of overlapping BSs is calculated in Table I.

TABLE I

COVERAGE PROBABILITY

Number of BSs	1	2	3	4
Grid-A or Grid-B separately	42.92%	57.08%	0	0
Grid-A and Grid-B together	0	17.36%	51.12%	31.52%

MSs arrive following a Poisson distribution, and their sojourn time follows an exponential distribution, which is
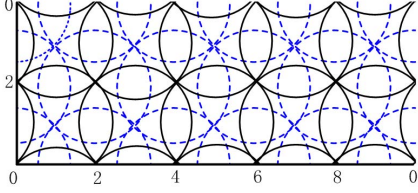
Fig. 2.   BSs topology

assumed not affected by MS's allocated bandwidth for simplicity. Load $\rho = \frac{E[|ActiveMS|]}{|BS|}$ is defined as the average number of active MSs in the system divided by the number of BSs (20). MS arrives randomly in the torus, and the set J_i of BSs covering MS i are determined from MS i's location. To introduce asymmetry in traffic distribution, The 4×2 rectangular with vertices $(0,0)$ and $(4,2)$ (occupying 20% of the whole torus) is selected as hot area. Among all MSs generated, $\kappa \geq 20\%$ are generated in this area. For BS $j \in J_i$, the full rate R_{ij} between MS i and BS j is a random variable depending on varying channel condition. For simplicity, we assume:

(1) R_{ij} follows a uniform distribution $U[R_{min}, R_{max}]$ with mean $\mu = \frac{R_{min}+R_{max}}{2}$ and standard deviation $\sigma = \frac{\sqrt{3}*(R_{max}-R_{min})}{6} = \frac{\sqrt{3}*(\mu-R_{min})}{3} \leq \frac{\sqrt{3}}{3}\mu$.

(2) Grid-A and grid-B have different distribution parameters, denoted as (μ^A, σ^A) and (μ^B, σ^B) respectively.

(3) In the coverage area of a specified BS j, the distribution of R_{ij} does not depend on location of MS i. For BS $j \in J_i$ and $k \in J_i$, R_{ij} and R_{ik} are *independent*. In real situation, *path loss*, for example, introduces location dependent rate distribution and results in *negative inter-BS rate covariance* in many topologies, which may stochastically increase the difference between R_{ij} and R_{ik}, thus reduces the effective coverage area of overlapping cells for rate-aware algorithms, while further exposing the ignorance of rate-unaware algorithms. Results of path loss simulations reveal this trend and are not shown here to save space.

(4) Once MS i arrived, the BSs set J_i covering it and all R_{ij} for $j \in J_i$ are fixed through MS i's lifetime. R_{ij} can be interpreted as the average full rate from BS j through MS i's lifetime. To model the measure and prediction error, we use another random variable R'_{ij} as the input to algorithms below. $R'_{ij} = \alpha R_{ij} + (1-\alpha)u$ where u is an i.i.d random variable of R_{ij}, and $\alpha \in [0,1]$ tunes accuracy of R'_{ij}. It is easy to prove that $\alpha = \frac{cov(R_{ij}, R'_{ij})}{\sigma(R'_{ij})\sigma(R_{ij})}$. Thus α is the statistical correlation of R'_{ij} and R_{ij}. When $\alpha = 0$, R'_{ij} is statistically independent of R_{ij}. When $\alpha = 1$, R'_{ij} is perfect estimation.

All parameters are summarized in Table II. Performance of Coordinated Proportional Fairness (*Ideal*) scheme (calculated using convex optimization), *GLS* scheme, *Best-Signal* scheme, *Least-Population* scheme, and *Non-Cooperative* scheme are compared.

B. Numerical Results

Figure 3 plots the cumulative distribution function (with arithmetic and geometric mean in parentheses) of per-flow throughput under standard setting in Table II. Among all schemes, *Best-Signal* scheme has the highest arithmetic mean of throughput, because it greedily assigns each MS to the BS providing the best channel condition. However, *Best-Signal*'s geometric mean of throughput is lower than the ideal fairness scheme, because of MSs associated with BSs which though offer the best channel condition, but too crowded to allocate enough radio resource to each of them. As shown in Figure 3, there are 22% of the flows in *Best-Signal* have throughput lower than $150kbps$, compared to 8% in ideal fairness.

GLS and *Non-Cooperative* scheme approximate ideal fairness very well, because they consider both R_{ij} and $|I_j|$. *Non-cooperative network performs similar to cooperative network*, because the second term in (5) is upper bounded. This phenomenon exists for almost all simulations we conducted. *Least-Population* performs poor in both arithmetic and geometric mean of throughput, implying that traditional load balancing technique is not applicable to wireless data networks.

We also simulate *Independent-Grid* setting. In this setting, there is no inter-operation between Grid-A and Grid-B, and every MS must pre-define their type. Type-A MS can only associate with Grid-A BSs, while Type-B MS can only associate with Grid-B BSs. Load of type-A MSs ρ^A and type-B MSs ρ^B are both set to $\rho/2$. It is shown that *GLS* scheme under *Independent-Grid* performs worse than the standard inter-operation setting, because of the smaller overlapping probability. This suggests the benefits of overlapping coverage, and also indicates the potential benefits of inter-operation among (heterogeneous) radio access networks.

Since geometric mean of per-flow throughput embodies both the overall system resource efficiency and fairness among flows, following experiments compare geometric mean.

Figure 4 shows the impact of asymmetric traffic distribution. The figure shows that *GLS* performs close to *ideal* scheme even under highly asymmetric traffic distribution. Such robustness is expected for *GLS* since its objective function also takes into account BS utilization. In comparison, performance of *Best-Signal* deteriorates with increasing traffic asymmetry.

Figure 5 shows the impact of load ρ. The performance gap between *GLS* and ideal fairness is getting smaller with in-
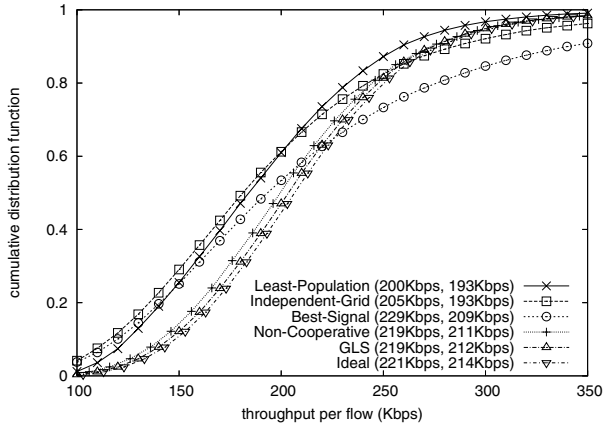
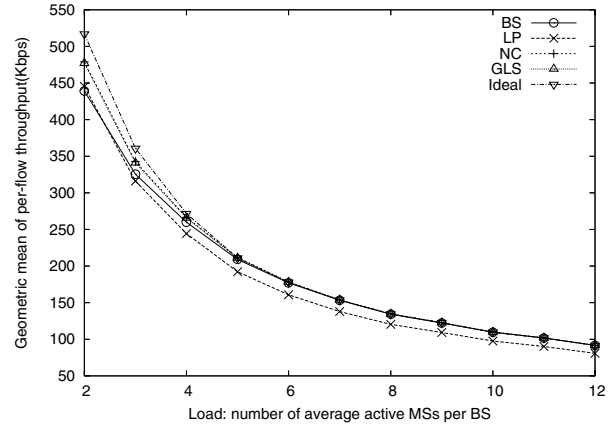Fig. 3. Cumulative distribution function of per-flow throughput

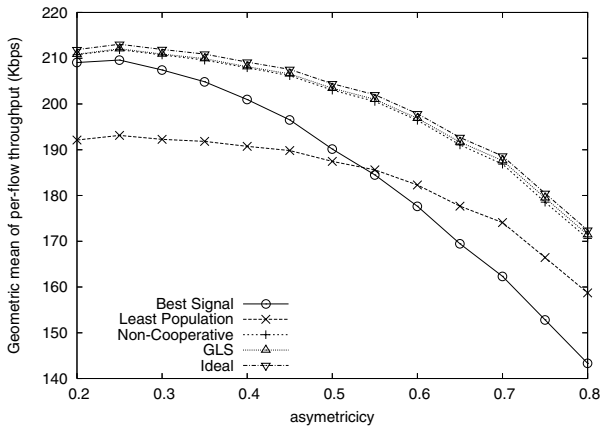Fig. 5. Geometric mean of throughput vs load

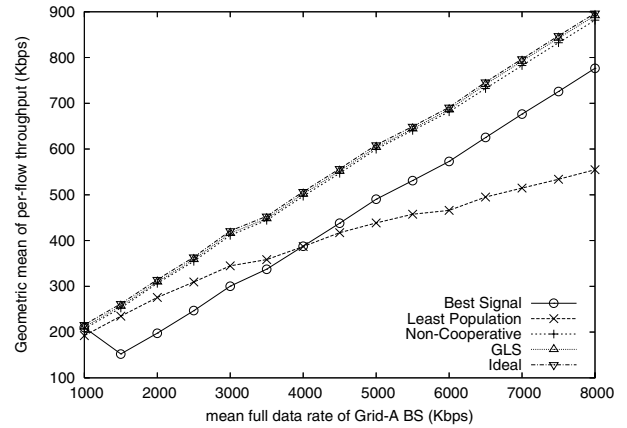Fig. 4. Geometric mean of throughput vs traffic distribution

Fig. 6. Geometric mean of throughput vs networks heterogeneity

creased traffic intensity. This is because *GLS* allocates resource on a per MS basis. Hence, the larger the traffic load, the finer the relative granularity of *GLS*, also the more chance *GLS* is able to recompute or perform load balancing. In fact, the only case that we observe obvious difference between ideal scheme and *GLS* scheme is when the average number of MS per BS is very small (e.g.≤ 3)

Figure 6 examines network heterogeneity by tuning μ^A from $1Mbps$ to $8Mbps$. the relative performance gap Notice that there is a strange drop of performance of *Best-Signal* when μ^A increases from $1Mbps$ to $2Mbps$, because with this setting *Best-Signal* scheme *shadows* the relatively lower rate network (Grid-B BSs) from associating MSs, thus accumulates all load to Grid-A BSs. In comparison, other schemes can smoothly exploit the increased capacity.

Rate estimation accuracy α and rate standard deviation σ are also interesting. When $\sigma = 0$ or $\alpha = 0$ with $\mu^A = \mu^B$, *GLS* and *Non-Cooperative* degenerate to *Least-Population* scheme. When σ or α increase, relative performance of *Least-Population* scheme decreases, while relative performance of *Best-Signal* scheme increases. Figures are omitted to save space.

VI. CONCLUSION

In this paper, a general proportional fairness model is formalized for elastic traffic in the context of multiple wireless cells with overlapping coverage. And a simple GLS scheme is proposed which approximates ideal fairness scheme closely under a wide range of simulation settings.

REFERENCES

[1] D. Bertsekas and R. Gallager, *Data Networks (Second edition)*, Prentice Hall, Englewood Cliffs, New Jersey, 1992.
[2] M. Heusse, F. Rousseau, G. B. Sabbatel, A. Duda,"Performance Anomaly of 802.11b," in *IEEE INFOCOM*, San Francisco, CA, USA, March 2003.
[3] J. Holtzman, "CDMA forward link waterfilling power control," *Proc. VTC2000-Spring*, Tokyo, Japan, pp. 1663–1667, May 2000.
[4] F. Kelly, A. Maulloo and D. Tan, "Rate control in communication networks: shadow prices, proportional fairness and stability," *J. of the Operational Research Society*, vol. 49, no. 3, pp. 237–252, March 1998.
[5] H. Kim and Y. Han, "A Proportional Fair Scheduling for multicarrier transmission systems," *IEEE Communications Letters*, vol. 9, no. 3, pp. 210-212, March, 2005.
[6] R. R. Mazumdar, H. Yaiche and C. Rosenberg, "A game theoretic framework for bandwidth allocation and pricing in broadband networks," *IEEE/ACM Trans. on Networking*, vol. 8, no. 5, pp. 667–677, 2000.
[7] J. F. Nash, Jr, "The bargaining problem," *Econometrica*, Vol. 18, No. 2, pp. 155–162, Apr.,1950.
[8] B. Tian, L. Li and R. Ramjee, "Generalized Proportional Fair Scheduling in Third Generation Wireless Data Networks," in *IEEE INFOCOM*, Barcelona, Spain, April 2006.