Data Mining and Coordination to Avoid Interference in Wireless Networks

Balaji Rengarajan
Dept. of ECE, UT Austin

Gustavo de Veciana
Dept. of ECE, UT Austin

Abstract—**This paper gives an overview of a system-level framework to mitigate interference using coarse grained coordination of transmissions across base stations. Our approach is based on collecting and mining measured data capturing a user population's diversity in sensitivity to interference. Measurements of user's channel gains are clustered and aggregated into a finite set of traffic classes, which abstract both the traffic and environmental character of the system load. These in turn serve as coarse grain variables that can be exchanged among base stations and used in optimizing coordinated schedules. Unfortunately, the clustering and optimization steps are intimately related, eventually impacting system performance – i.e., they should ideally be carried out jointly. Our work on this problem points to some insights on how clustering might be suitably carried out, and approaches towards optimizing coordinated schedules. We present a subset of our extensive system-level simulations, which show reductions in file transfer delay ranging from 20–80%, depending on the traffic loads, as compared to a simple baseline not unlike those in the field today. In particular, we explore the benefits of traffic and environment aware coordination for systems subject to clustered traffic loads, e.g., due to spatial hotspots. Overall we believe this approach appears to achieve reasonable gains in performance, but perhaps more importantly achieves substantially more uniform coverage while reducing average power consumption by up to 45%.** [1]

I. Introduction

Devising practical wireless systems that effectively cope with inter-cell interference as well as spatial heterogeneity in traffic loads and the environment while maximizing spectral efficiency may be one of the most important problems engineers face in realizing the vision of ubiquitous broadband wireless. If this vision is to be successfully realized, users should be able to expect the same 'deterministic' service they can expect from a wireline network, independent of location and at reasonable power cost. Let us consider two straightforward approaches to overcoming this problem.

First, the traditional approach of mitigating inter-cell interference in a cellular network by partitioning resources, e.g., frequency, so that concurrent transmissions can be realized with minimal interference. For example, with a frequency reuse 1/3, the overall bandwidth would be partitioned into three bands, and allocated to cells, in a manner that minimizes the number of neighboring cells that share the same band. This solution is relatively simple and indeed effective at reducing interference seen by users while enhancing coverage areas for each base station. However, this comes at a high cost. Indeed, a single cell/sector can now only use a fraction, e.g., 1/3, of the available bandwidth resulting in a linear reduction in system capacity. Since capacity depends linearly in the bandwidth, and logarithmically in the signal-to-noise plus interference (SINR), overall a simple reuse pattern achieves a more homogeneous performance, but at the cost of reduced overall capacity and reduced individual user's peak capacity. It is desirable to engineer systems that use all the available spectrum in every cell, achieving very large network capacities, provided inter-cell interference is effectively managed. Recently there has been intensive work towards devising improved reuse schemes that can partially remedy this picture. We will return to this below.

A second simple approach is to consider increasing the number of base station/access points. This can drastically decrease the distance between users and their base stations, resulting in a drastic increase in capacity and reduction in transmission energy requirements. However, this comes at a significant increase in infrastructure and management costs. Moreover, in such deployments the proportion of users whose capacity is limited by interference from their neighbors grows. In other words, considering a fixed population density and a typical base station's point of view, the number of users that are close, say within 100m would not change, yet the neighboring base stations would be closer resulting in higher interference. Furthermore, the coverage area per base station would presumably decrease, meaning that the capacity provided by each individual base station would be shared by a smaller number of users. This has the undesirable side effect of reducing the network's capability to statistically multiplex bursty traffic – particularly desirable for data traffic, so as to share the cost of the additional infrastructure. As a result, one might expect low average but bursty utilization profiles for individual base stations. So, under this approach, not only would the capacity of a larger number of users be interference limited, but the interference might be quite dynamic. Again, managing inter-cell interference will remain essential to fully realizing the potential of the costly broadband wireless infrastructure.

Related and prior work. Let us consider some of the related work and recent proposals towards addressing this challenge, and identify some of the areas for improvement. Most approaches for mitigating the effects of inter-cell interference have been studied in the context of a static user population.

[1]This work was supported by a gift from the Intel Research Council and the NSF Award CNS-0721532.

Centralized joint user scheduling schemes, requiring large amounts of information to be conveyed to a centralized scheduler, are presented in [1], [2]. The centralized scheduler also has to solve a highly complex optimization problem based on the queue and channel states of all the users in the network to make scheduling decisions. Alternatively, static schemes using different reuse factors over different time periods to protect vulnerable users have been considered, see e.g., [3]–[6]. A quasi-static scheme based on a similar principle is presented in [7]. The above schemes only considered base stations that either transmit at maximum power, or are turned off. They also do not take into consideration the impact of using adaptive modulation and coding schemes. A power-control based interference management scheme is proposed in [8]: users are served using one of two sets of carriers that use different power levels. A different approach that varies transmit power across time at a slow pace so as to improve performance is proposed in [9]. The users then track the varying channel conditions and this information is used by the base station to effectively schedule transmissions.

The focus of these schemes is to ensure that all users perceive acceptable signal to interference ratios. However, this metric does not fully describe the performance experienced by best effort users. The characteristics of the user population being served do not influence the power control policy, leaving scope for further improvement. In a realistic scenario, data requests from users are generated at random times, and the users leave when their service requirements have been met. Such load dynamics also translate to time varying interference seen by users, and further impact the performance of schemes designed to mitigate inter-cell interference.

Potential capacity gains from inter-cell coordination in a dynamic setting were characterized in [11], and the results confirm that significant gains can be obtained through inter-cell coordination in an interference limited system. For a practical system, the delay performance experienced by users at typical system loads is an important consideration. The static capacity-optimal schedule developed in [11] is not a practicable solution for a system at light to moderate loads. Also, the system model considered in [11] is idealized, and would in reality be prohibitively complex in terms of the communication, and computation overhead required.

Contributions: In this paper, we propose a system-level framework whose primary goal is to mitigate the impact of inter-cell interference through coordination of transmissions. The approach is driven by two key objectives that differentiate it from previous work with a view on both effectiveness and practicality. The first is that coordination of transmissions should be environment- and traffic-aware. Our expectation, is that system performance can be substantially improved relative to the state-of-the-art, as well as solutions proposed in the research literature. The second objective is to minimize the information that needs to be exchanged among base stations, and thus reduce the demands on the backhaul. This excludes approaches that require fine grain information on the time varying channels of individual users to be exchanged among base stations to enable joint scheduling and coordination of transmissions across base stations.

Traffic and environment awareness are achieved by measuring the average traffic loads and characteristics of the user population a system supports. If the ‘average’ traffic loads are fairly stable over reasonable periods of time one should ideally be able to optimize a coordination scheme to the specific spatial traffic load, propagation environment, and factor the sensitivity to interference seen in the field. As we will see, thousands of measurements per second are already performed in the system, in order to adapt modulation and coding on a per user basis and trigger handoffs for mobile terminals. However, in order to limit the communication overheads, and the complexity of optimizing coordination, we seek to extract from the measured data the salient characteristics of the load – i.e., coarse grained information that still captures the key characteristics of the traffic. This may be viewed as a data-mining/aggregation task over fairly large (but local to a base station) measured data sets to define coarse grain classes of user loads. These in turn are used in formulating and optimizing the coordination schedule across base stations. The key idea is to take advantage of the diversity in users’ sensitivity to interference originating from the adjoining cells – this is not new. The novelty of our proposal lies in the development of new abstractions, a network architecture, and associated optimizations that make this practical, and efficient.

The contributions of this paper are as follows. First, we explore various tradeoffs in defining the coarse grain user classes from measured spatial loads with a view on optimizing system performance. We shall explore performance-complexity tradeoffs in the number of classes, variability in individual characteristics of users within a class, and balancing the sizes i.e., offered loads among user classes. Second, we will discuss various approaches to optimizing coordinated schedules. Third, in this paper we evaluate how spatial clustering in traffic loads impacts the types of gains one can hope to get from an approach that is traffic and environment aware. While conceptually similar, coordination of downlink versus uplink transmissions have fairly different characteristics. In this paper we shall report results for the downlink case, which illustrate the significant gains that can be achieved in terms of delay performance, power consumption at the transmitter, and substantially enhanced spatially homogeneous service to users. We note at the outset that the performance gains e.g., 50%, depending on the load that we report can not be viewed as radical improvements. However in our view the substantially improved spatial homogeneity in performance and reduced average power consumption, are big wins. The first makes substantial headway towards providing more deterministic service, while the second, suggests that additional performance gains could be achieved if one compared coordinated system to a baseline constrained to the same average power expenditures.

The rest of this paper is organized as follows. In Section II we introduce our system model for optimizing coordinated schedules, assuming traffic has already been aggregated into user classes. In Section III we explore the challenges in moving from measured data to coarse grain user classes, and the impact this has on performance. In Section IV,V, we consider optimizing coordination schedules. Section VI explores the impact of spatial clustering in traffic loads, e.g., due to hotspots, on system performance gains. Finally, Section VII concludes the paper.

II. System Model

Due to the nature of path loss, it is typically only transmissions in the neighboring cells that cause interference. As such, we propose to split large networks into a collection of independent coordinated clusters, where cells/sectors which are tightly coupled through mutual interference are grouped together. Let N denote the number of neighboring base stations/sectors being coordinated, indexed by $b = 1, \ldots, N$. For simplicity, each user is assumed to be served by a single fixed base station, i.e., no mobility for now. For each user i, we let $\vec{h}_i = (h_i^b | b = 1, \ldots, N)$ be the (average) channel gain vector, where h_i^b is the gain from base station b to user i, which is measured by each user and fed back to the serving base station. Fig. 1 depicts the measurements made by each user when coordinating three facing sectors in a hexagonal layout of base stations. This is the canonical example we will consider throughout this paper. Thus if measurements are made

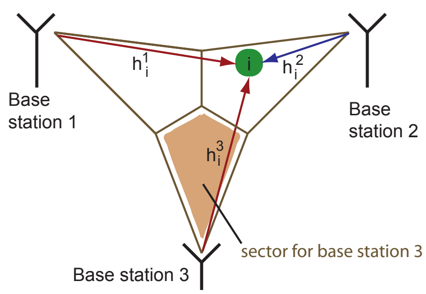


Fig. 1. An example scenario for coordination.

for a period of time, each base station would have a sample of gain vectors associated with the user population it sees.

A. Traffic Loads and System Dynamics

In this paper, we adopt a dynamic flow-level model where user requests follow a Poisson point process which may be spatially inhomogeneous over an area of interest A. In particular we assume a spatial arrival intensity function $\lambda(x), x \in A$ giving an overall rate $\int_A \lambda(x)dx = \lambda$ requests/sec. Users leave the system when the associated data transfer on the downlink is completed. The actual system load depends strongly on the user locations, since interference from neighboring base stations, and thus the channel capacity are location dependent. In addition to homogeneous load distributions, we shall explore various types of inhomogeneities, particularly with respect to spatial clustering, e.g., may be induced by hotspots where users tend to congregate. We will be interested in the impact that such inhomogeneities have on traffic aware coordination.

B. Traffic Classification and Aggregation

Users at each base station/sector b are classified into one of K_b *user classes* based on their measured channel gain vectors. We will discuss how these classes are defined in the next section. For now, recall that they are intended to abstract key characteristics of the load distribution and the propagation environment for purposes of optimizing a coordinated schedule. Thus, after classification, each class $k = 1, \ldots, K^b$ associated with base station/sector b sees an arrival process which is Poisson, with rate denoted by λ_{bk}. Note that since user classes correspond to spatial aggregates, the channel and and interference characteristics seen by users in a given class may be quite different depending on the number of classes and the classification mechanism. We will return to this in the next section.

Base stations have a file to transmit for each user request, with mean file size $\overline{F}_{bk}$ bits. Define $\rho_{bk} = \lambda_{bk}\overline{F}_{bk}$ to be the mean traffic load (bits/second) arriving at class k in base station b. Let $\vec{\rho} = (\rho_{bk} : b = 1, \ldots, N, k = 1, \ldots, K^b)$ denote the expected traffic load vector. Fig. 2 illustrates a scenario with two base stations, and two classes per base station. The classes may have different mean traffic loads, capturing in part the spatial distribution of traffic supported by the system.

C. Coordination and Service Model

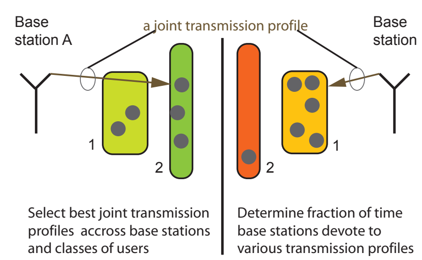


Fig. 2. Illustration of a joint transmission profile.

A *joint transmission profile* represents one of the various modes in which the network can be operated. As illustrated in Fig. 2, it specifies a power profile, i.e., the transmit power level for each base station, and the associated user classes to be jointly served. Note that this is *not* a specification of which user to serve, only a restriction on the transmit power to be used at the base station and a 'recommended' class that might be beneficially served. Base stations can independently devise complementary dynamic user/packet scheduling policies to serve their users. For simplicity, in this paper, we assume that base stations use processor sharing scheduling (or an approximation thereof) to serve the active users in a class.

The base stations are assumed to be able to transmit at one of P discrete power levels, including 0, corresponding to no transmission. The N-dimensional column vectors $\vec{p}^i$ and $\vec{c}^j$ specify the power levels and classes to be served by the base stations under power profile i and class combination j. The b^{th} component of these vectors, p_b^i and c_b^j, specify the transmit power to be used by base station b and the class to be served.

The number of different power profiles is denoted by $U = P^N$, the number of class combinations by $V = \prod_{b=1}^{N} K_b$, and thus the number of joint transmission profiles is $L = UV$. Let $\mathcal{P} := \{\vec{p}^1, \ldots, \vec{p}^U\}$ and $\mathcal{C} := \{\vec{c}^1, \ldots, \vec{c}^V\}$ denote the sets of admissible joint power profiles and class combinations respectively for the N base stations. Thus, each joint transmission profile l where $l = 1, \ldots, L$ is two vectors: $\vec{p}(l) = \vec{p}^i \in \mathcal{P}$ and $\vec{c}(l) = \vec{c}^j \in \mathcal{C}$.

A joint transmission schedule corresponds to the fractions of time $\vec{\alpha} = (\alpha_l : l = 1, \ldots, L)$ for which the network uses each transmission profile. In general, this schedule will be picked to optimize a chosen load dependent performance measure, $f(\vec{\rho}, \vec{\alpha})$, through an optimization of the form:

Problem 2.1: A generic optimization problem to determine a coordination schedule:

$$\min_{\vec{\alpha}} f(\vec{\rho}, \vec{\alpha})$$

such that

$$\begin{aligned} \rho_{bk} &\leq R_{bk}(\vec{\alpha}), \forall b, k, &(1)\\ \sum_{l=1}^{L} \alpha_l &\leq 1, &(2)\\ \alpha_l &\geq 0, \, l = 1, \ldots, L. &(3) \end{aligned}$$

Here, $R_{bk}(\vec{\alpha})$ denotes the capacity allocated to class k at base station b by the schedule $\vec{\alpha}$. Eq. (1) constrains the rate allocation across classes to be one that stabilizes the network. Eqs. (2), and (3) ensure that the coordination schedule is valid. In the sequel, we will describe different methods to determine joint transmission schedules, and use extensive simulations to compare their performance. The following section describes the simulation model in detail.

D. Simulation Model

In the simulations, we consider three facing sectors in a hexagonal layout of base stations with cell radius 250m. Users associate themselves to the geographically closest base station. A carrier frequency of 1GHz, and a bandwidth of 10MHz are assumed. The maximum transmit power is restricted to 10W. The base stations are assumed to be able to transmit at three different power levels: 0, 5, and 10W. Additive white Gaussian noise with power -55dBm is assumed. We consider a log distance path loss model [12], with path loss exponent 2. Shadowing, and fading are not considered in these preliminary results, but the addition of shadowing does not fundamentally change the characteristics of our measurement driven scheme, as noted in Sec. III-A.

File sizes are assumed to be log normally distributed, with mean 2MB. The data rate at which users are served is calculated based on the perceived SINR using Shannon's capacity with rates quantized to 0, 1, 2, 5, 10, 20, and 30Mbps. The mean user perceived delay is estimated within a relative error of 1%, at a confidence level of 95%. In sections IV, and VI, we consider Poisson arrivals that are distributed uniformly within the simulated area, and in Sec. VI, we study the impact of non-homogeneous load distributions.

III. From Measurements to User Classes

The most important element of our proposed framework is our notion of user classes. By contrast with related work, our notion of user classes and class loads aggregate users (locations) that share similar (but not necessarily identical) sensitivity to interference from neighboring base stations. They enable base stations to measure, aggregate, and share coarse grained information about the traffic loads they support. They also drive our system-level optimization, e.g., Problem 2.1, which has a number of constraints and decision variables which respectively grow linearly and polynomially (of degree N) in the number of classes. As the number of user classes is increased, their fidelity in capturing the characteristics of the user population increases. However, communication overheads, and the computational complexity associated with the proposed coordination scheme also grow. Moreover, perhaps surprisingly, we will see that it is not necessarily the case that higher fidelity leads to better coordination performance. Therefore, it is advantageous to use a relatively small number of classes. However, in this case, there may be large disparities among transmission of users in the same class. In order to solve Problem 2.1, one must properly capture the capacities $R_{bk}(\vec{\alpha})$ that are allocated to user classes under different coordination schedules parametrized by $\vec{\alpha}$. As will be seen in this section, this is not a simple problem, yet good approximations that make the optimization problem convex can be found to make this tractable.

A. Measurement and Clustering of Users into Classes

Consider monitoring a user population sharing a wireless system during a period of time. As shown in Fig. 1, a simple way to capture the environmental conditions is to measure the average channel gains between users and neighboring base stations – this is already done in practice to facilitate handoffs. A user i sharing a similar gain vector, $\vec{h}_i$, with other users has a similar susceptibility to interference from neighboring base stations. Yet, in an interference limited regime, Shannon's capacity formula suggests that users' transmission rates will vary as the logarithm of the ratio of the received signal power to interference. Thus, for each measured user, let us define a logarithmically distorted gain vector $\vec{g}_i = (g_i^b | b = 1, \ldots, N)$, where $g_i^b = \log(h_i^b)$. Users sharing similar log-gain vectors will share similar transmission rates under the various power profiles.

In this paper, a k-means clustering algorithm [13], [14] is used to cluster measured log-gain vectors into a fixed number of user classes. Specifically, the algorithm partitions users associated with base station b into K_b clusters with centroids $\vec{g}^*_{bk}, k = 1, \ldots, K_b$, such that the mean Euclidean distance between the log-gain vectors and the centroids is minimized. Given a clustering, and the resulting centroid vectors, future users can be classified based on which centroid its log-distorted gain vector is closest to. With classes defined, estimating the average loads for each class under a given spatial traffic load is a simple task.

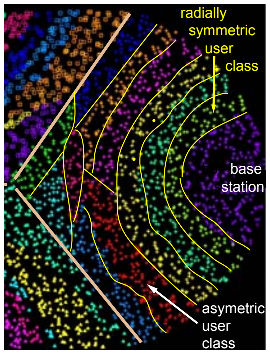


Fig. 3. An example of class definitions.

Fig. 3 exhibits a clustering for a sector in our example scenario where three neighboring base stations are to be coordinated. Note that in practice, due to shadowing and real environment obstructions, user classes will not result in the 'smooth' structure or spatial locality exhibited in this example. In fact, they would instead reflect the character of the environment as well as the typical locations where a user population tends to dwell.

B. Estimating Class Rates

Let the random variable X denote the location of a typical arrival, since the arrivals are non-homogeneous Poisson Point Processes, its density would be $f_X(x) = \lambda(x)/\lambda$ for $x \in A$. Let $b(x)$, and $k(x)$ be user x's base station and class respectively. These reflect a location dependent the base station association and classification policies which are for simplicity not load dependent. Finally, let R_x^l be the maximum rate at which user x can be served under profile l, *assuming* all base stations are active. Note that R_x^l is zero, if a class other than $k(x)$ is served by base station $b(x)$ under profile l. The following proposition captures the capacity seen by downlink queues in the system – see [15] for details.

Proposition 3.1: Consider the downlink queue associated with class k at base station b. It sees an offered load of ρ_{bk} bits/sec., and time varying capacity that depends on $\vec{\alpha}$. Suppose the rate at which base stations switch among profiles is fast compared to the time scale of the user dynamics, and the base station uses processor sharing to serve users in each class, then the queue is stable if $u_{bk} = \frac{\rho_{bk}}{R_{bk}^H(\vec{\alpha})} \leq 1$, where

$$R_{bk}^H(\vec{\alpha}) = \left(\mathbf{E}\left[\frac{1}{\sum_{l=1}^L \alpha_l R_X^l} \;\middle|\; b(X) = b,\, k(X) = k \right] \right)^{-1}.$$

Further, when the queue is stable, the mean number of active users associated with the class is given by $\frac{u_{bk}}{1-u_b k}$.

Note that $R_{bk}^H(\vec{\alpha})$ is the harmonic mean of the average transmission rates seen by users in class k at base station b. It captures the capacity allocated to the user class under schedule $\vec{\alpha}$. Unfortunately, estimating this for each $\vec{\alpha}$ requires knowledge of the complete distribution of users versus simple descriptive statistics, e.g., means and variances, which would reduce both communication and computational overheads.

The arithmetic and geometric mean of the average transmission rate perceived by users given by

$$\begin{aligned} R_{bk}^A(\vec{\alpha}) &= \sum_{l=1}^L \alpha_l \mathbf{E}[R_X^l \mid b(X) = b,\, k(X) = k], \\ R_{bk}^G(\vec{\alpha}) &= \exp(E[\log(\sum_{l=1}^L \alpha_l R_X^l) \mid b(X) = b,\, k(X) = k]), \end{aligned}$$

respectively are two alternatives for estimating class capacity. Note that the arithmetic mean is simple to compute: it depends only on the mean rates observed by users in the class under each profile, and is linear in $\vec{\alpha}$. However, it can be shown that $R_{bk}^H(\vec{\alpha}) \leq R_{bk}^G(\vec{\alpha}) \leq R_{bk}^A(\vec{\alpha})$, whence the geometric mean is the better estimate for the harmonic mean [16]. Unfortunately, the geometric mean is also burdensome to compute, making it impractical.

One can obtain approximations for the geometric mean based on moments which can be fairly accurate [17], [18]. In [15] we propose such an approximation as follows. Let Σ_{bk} be the covariance matrix of the transmission rates to the users in class k in base station b, $\Sigma_{bk}(l,m) = \mathbf{Cov}[R_X^l,\, R_X^m \mid b(X) = b,\, k(X) = k]$. The rate allocated to class k in base station b is approximated as

$$R_{bk}^{GA}(\vec{\alpha}) = R_{bk}^A(\vec{\alpha}) - \frac{\vec{\alpha}^T \Sigma_{bk} \vec{\alpha}}{c_{bk}}, \tag{4}$$

where, c_{bk} is a positive constant corresponding to a preliminary estimate for the class capacity. Note, for a given $vec\alpha$ the capacity allocated to all classes can be estimated with the coordinating base stations exchanging only the class means, and covariances of the transmission rates under the different profiles. Also it can be shown [15] that $R_{bk}^{GA}(\vec{\alpha})$ is convex giving a stability constraint (1) which is convex.

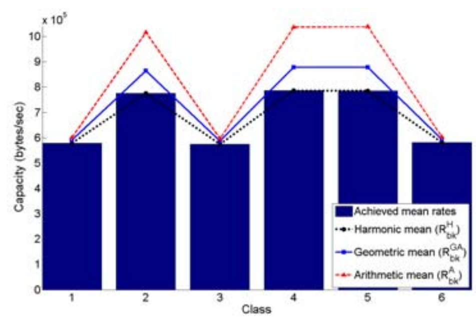


Fig. 4. Comparing the different estimates for class capacity.

Fig. 4 the actual class throughputs and our estimates, for a given coordination schedule, i.e., $\vec{\alpha}$, for our three sector scenario. Users are partitioned into two classes per sector. In this scenario, the classes with higher load correspond to larger proportion of the uniformly distributed users resulting in larger intra-class variance in users' rates. As can be seen, the arithmetic and geometric mean approximations are optimistic particularly for classes with higher variability. Indeed the arithmetic mean overestimates the capacity allocated to the classes by up to 20% compared to the geometric mean estimate. Below we will use the more accurate geometric mean approximation, which provides marginal improvements in performance.

IV. Static Coordination

The key element of base station coordination for downlink transmission is the joint selection of a coordinated schedule. Modeling and optimizing the performance of a set of spatially coupled (through interference) queues is fairly difficult, see [19]–[21], so in this paper we shall consider only a simplified model where base stations user class queues are decoupled M/GI/1 processor sharing queues. Tractable heuristics to capture this coupling better are proposed in [15].

A. Maximizing Capacity

The optimization problem obtained by setting $f(\vec{\rho}, \vec{\alpha}) = \sum_{l=1}^{L} \alpha_l$ in Problem 2.1 has a linear objective function, and convex constraints under the geometric approximation $R_{bk}(\vec{\alpha}) = R_{bk}^{GA}(\vec{\alpha})$, where c_{bk} can be appropriately approximated based on a further linear approximation. The resultant optimal schedule is one that stabilizes the network whenever possible, for any load distribution proportional to $\vec{\rho}$, see [15] for details.

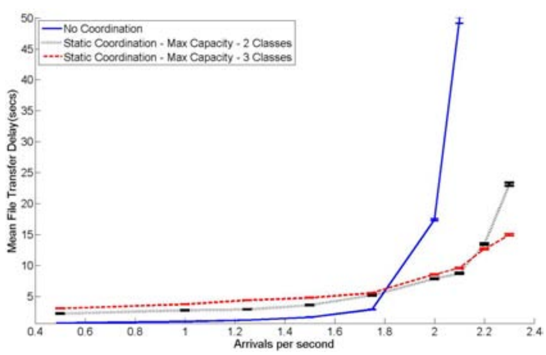


Fig. 5. Delay performance under the capacity maximizing schedule.

The graph in Fig. 5 shows average downlink file transfer delays vs. offered load under three schemes: uncoordinated transmissions at the maximum power, and two static approximations with two and three user classes per base station. At higher loads, coordination performs extremely well, improving delay performance over the scheme with no coordination by over 80%. However, this is not uniformly the case, and at very low loads, the coordination scheme increases mean delays by around 50% compared to the non-coordinated scheme. Under low loads, coordinating across base stations to mitigate interference is less of a concern because the probability that neighboring base stations are simultaneously transmitting is low. Therefore, one might as well allow base stations to transmit at higher power without coordination. Also, since we are using a static schedule, the probability that there are no active users in the class scheduled at a base station is high at low loads. This leads to the base station unnecessarily wasting time while users wait their turn to get served. This is also the reason for the coordination scheme using two classes outperforming the scheme with three classes until the offered load is high enough. A larger number of classes results in base stations wasting more time when using a static schedule, as the scope for statistical multiplexing is further reduced. Splitting the load and the resources into independent small chunks results in reduced capacity for sharing, and incurrs a statistical multiplexing loss. At low loads, the gains from reduced interference levels resulting from careful coordination across base stations are not sufficient to compensate for this statistical multiplexing loss.

B. Minimizing Delay

Setting $f(\vec{\rho}, \vec{\alpha}) = \sum_{b=1}^{N} \sum_{k=1}^{K_b} \frac{\rho_{bk}}{R_{bk}(\vec{\alpha}) - \rho_{bk}}$ in Problem 2.1 yield a convex optimization problem under the geometric approximation. The solution to this problem is one that minimizes the mean user perceived file transfer delay under our simplified model. This solution is different from the capacity maximizing schedule when there are variations in the load offered by different classes, see [15] for details. Fig. 6 shows the delay performance of the static delay optimal schedule which achieves better delay performance than the capacity maximizing schedule by accounting for the variations in load and multiplexing capability across classes. However, delay performance remains poor compared to the non-coordinated case at low loads due to the statistical multiplexing loss incurred by dividing the load.

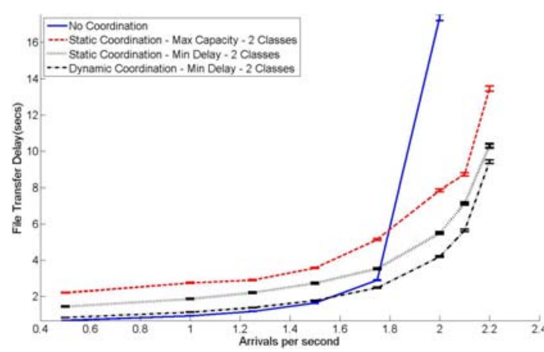


Fig. 6. Average file transfer delays under delay-minimizing schedules.

V. Dynamic Inter-Class Scheduling

Note that, in downlink transmissions, the capacity perceived by users in neighboring base stations is independent of the user/class that a base station serves and depends only on the transmit power levels used by the various base stations. Thus, when there are no active users in the class picked by the static schedule, the base station can dynamically pick an alternate class to serve without adversely affecting any of the cooperating base stations, i.e., without increasing the interference levels perceived by users. This class can be chosen by the base station based on different criteria, such as maximizing transmission rates, or serving the class with the largest number of active users. We refer to this as inter-class scheduling.

The dynamic scheduling strategy that we adopted is to serve all active users associated with a base station according to a processor sharing mechanism when the scheduled class has no active users. We found in our simulations that the delay performance of this strategy compared favorably to other policies. Note that this strategy allocates a proportionally larger rate to those classes that have a large number of active users. When the traffic offered by all classes share similar characteristics, the optimized static schedule balances the expected number of active users in each class. Thus, this dynamic scheduling strategy attempts to align the available capacity to the particular instantiation of the offered load. In Fig. 6, we show results for coordination along with dynamic inter class scheduling.

As can be seen in Fig. 6, dynamic scheduling significantly improves user's average delay performance, especially at light to moderate loads where mean delays are reduced by up to 40% as compared to the static scheme. At very low loads, it is still true that a scheme that transmits at maximum power without any coordination outperforms the coordination scheme. Attempting to coordinate transmissions at low loads results in base stations needlessly using a lower power, thus transmitting at a lower rate even when the neighboring base stations are idle. Since the probability of simultaneous transmissions occurring is minimal at low loads, coordinating is not worthwhile.

VI. The Importance of being Traffic Aware

In a real-world wireless network, the traffic load is unlikely to be spatially homogeneous and may exhibit significant variations over time. For example, at different times of the day, one might see concentrations in different regions, e.g., coffee houses, lunch spots, public transportation, or associated with congestion patterns, etc. We explore the potential gains from coordination in such a scenario. In particular, we are interested in understanding the degree to which optimizing for a particular load is critical. For example, if no information is available, a natural choice is to optimize for a uniform distribution of users. We shall evaluate the performance of the dynamic scheme proposed previously when it is optimized in this fashion, versus an optimization that is traffic-aware.

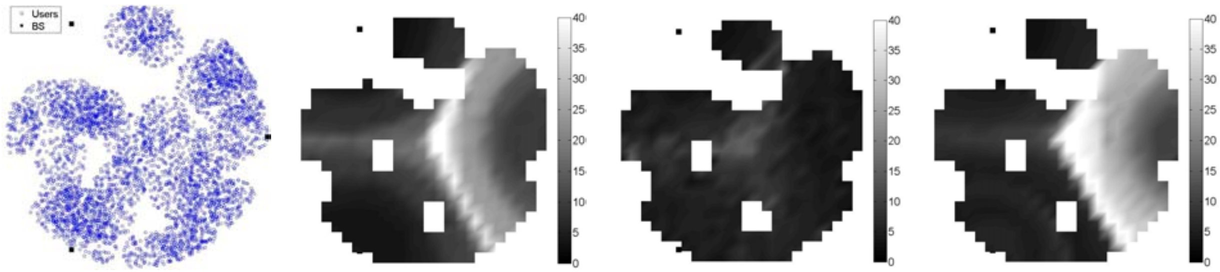
Our clustered traffic model is as follows. User locations are constrained to a subset of the simulated area determined by the realization of a Boolean germ-grain model [22]. The grains of the Boolean model are discs of fixed radius, while the germs are distributed uniformly within the simulated area. The probability that an arrival's location falls in any of the discs is equal. The density of users at various points within the cell depends on the number of grains covering it. The density of users in areas covered by multiple grains is high, resembling a hotspot. Fig. 7a exhibits a realization of the spatial load with 70 germs, and discs with radius equal to one fifth the radius of the cell are used. Note that there are regions within the cell with sparse user densities, and others where users tend to cluster. As the number of germs increases, the arrivals process converges to a homogeneous Poisson process. A small number of germs represents a user population that is highly clustered, with large variations in user densities within the coverage area.

In our simulations, the number of germs is varied from 10 to 10,000 to simulate various degrees of clustering in the spatial load. For each case, we investigate the performance in twenty different realizations for the Boolean model. As explained previously, the actual load on the system is highly dependent on the spatial characteristics of the traffic. In roughly evaluate performance under vastly different spatial loads, we normalize the overall arrival rate so that the actual loads are comparable. Specifically, we choose the arrival rate that results in the base stations being 95% utilized when the base stations, assuming all base stations transmit at maximum power all the time even if they have nothing to send. This operating point is computed using the harmonic mean, as described is Sec. III-B.

Fig. 8 depicts the reduction in delay achieved by the two schemes compared to the non-coordinated case. It is clear that when the actual traffic being served is highly clustered (small number of germs), the traffic-independent coordination scheme performs much worse. In fact the average delays experienced by users are more than doubled vs to no coordination. As the number of germs is increased, and the spatial distribution of users approaches the uniform distribution, the traffic-independent scheme performs better than the non-coordinated one, and eventually catches up to the traffic-aware scheme. The reduction in delay achieved by the traffic aware scheme appears independent of clustering in the loads. Note however, that our normalization is imperfect, and in fact the measured loads were lower for scenarios subject to clustered loads. Thus we conjecture that subject to the same system load the gain achieved by the traffic aware will increase if the spatial load exhibits higher random clustering.

Fig. 9 plots the variance across the scenarios under the traffic-aware scheme and the case where no coordination is used. This variance is induced by the sensitivity to inter-cell interference, and because different locations are affected very differently by interference. A non-coordinated system that serves a varying, non-homogeneous spatial distribution of users is prone to excessive variations in user perceived performance, and can experience very poor delay performance during time periods when it has to support a user population that is "poorly situated". The traffic aware coordination scheme is successful in shielding users from varying spatial loads, and achieves relatively homogeneous performance independent of where a user population lies. Finally, Figs. 7b,7c, 7d demonstrate the spatial homogeneity achieved by the traffic-aware scheme even when the spatial load is non homogeneous. This decoupling of performance from both the variable spatial distribution of load, and the location of the users is a significant benefit.

Additionally, we find that the average power expended by the base station is substantially reduced when coordination is used, e.g., 45% when the arrival rate is 2 users per second. This suggests a reduction in cooling costs at the base station, and also indicates that we can further improve delay performance if the base stations are allowed to transmit at higher peak power levels.



(a) A realization of a clustered user population (b) No coordination (c) Traffic-aware coordination (d) Traffic-independent coordination

Fig. 7. A scenario with non-homogeneous spatial load

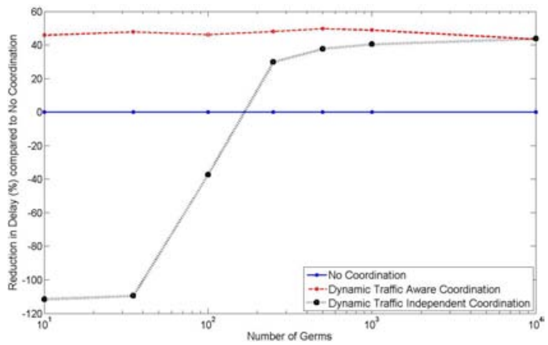


Fig. 8. Percentage reduction in average file transfer delays.

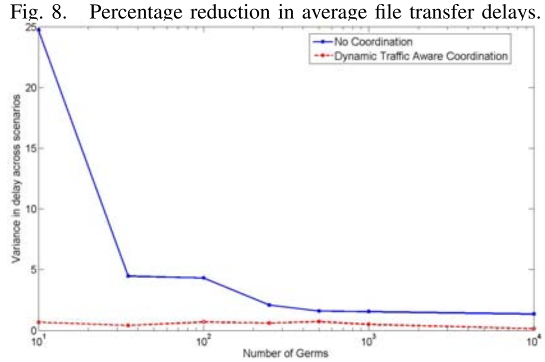


Fig. 9. Variance across scenarios in the average file transfer delay achieved by the traffic aware, and non-coordinated scheme.

VII. Conclusion and Future Work

This paper proposes a low complexity, traffic-aware system-level approach to interference mitigation that improves performance perceived by best-effort users on the downlink without requiring high channel measurement and estimation, communication, and computational overheads. Our preliminary results suggest substantial gains in performance, spatial homogeneity of service and reduced power expenditures. Our ongoing work, aims to set this approach on a solid foundation while addressing practical requirements in implementing these techniques, see [15].

References

[1] N. Kahale and P. E. Wright, "Dynamic global packet routing in wireless networks," in *IEEE INFOCOM*, vol. 3, Apr. 1997, pp. 1414–1421.

[2] S. Das, H. Viswanathan, and G. Rittenhouse, "Dynamic load balancing through coordinated scheduling in packet data systems," in *IEEE INFOCOM*, vol. 1, 2003, pp. 786–796.

[3] T. K. Fong, P. S. Henry, K. K. Leung, X. Qiu, and N. K. Shankaranarayanan, "Radio resource allocation in fixed broadband wireless networks," *IEEE Trans. Commun.*, vol. 46, no. 6, pp. 806–818, Jun. 1998.

[4] K. K. Leung and A. Srivastava, "Dynamic allocation of downlink and uplink resource for broadband services in fixed wireless networks," *IEEE J. Select. Areas Commun.*, vol. 17, no. 5, pp. 990–1006, May 1999.

[5] X. Qiu and K. Chawla, "Resource assignment in a fixed broadband wireless system," *IEEE Communications Letters*, vol. 1, no. 4, pp. 108–110, Jul. 1997.

[6] A. Ghasemi and E. S. Sousa, "Distributed intercell coordination through time reuse partitioning in downlink CDMA," in *IEEE Wireless Communications and Networking Conference*, vol. 4, Mar. 2004, pp. 1992–1997.

[7] K. Chawla and X. Qiu, "Quasi-static resource allocation with interference avoidance for fixed wireless systems," *IEEE J. Select. Areas Commun.*, vol. 17, no. 3, pp. 493–504, Mar. 1999.

[8] J. Li, N. B. Shroff, and E. K. P. Chong, "A static power control scheme for wireless cellular networks," in *IEEE INFOCOM*, vol. 2, 1999, pp. 932–939.

[9] X. Wu, A. Das, J. Li, and R. Laroia, "Fractional power reuse in cellular networks," in *Proceedings of the 44th Allerton Conference on Communication, Control, and Computing*, September 2006.

[10] S. Borst, "User-level performance of channel-aware scheduling in wireless data networks," in *INFOCOM 2003*, vol. 1, March-April 2003, pp. 321 – 331.

[11] T. Bonald, S. Borst, and A. Proutiere, "Inter-cell scheduling in wireless data networks," in *European Wireless Conference*, 2005.

[12] T. S. Rappaport, *Wireless Communications: Principles and Practice,2/E*. Prentice Hall PTR, 2002.

[13] M. Anderberg, *Cluster Analysis for Applications*. Academic Press, 1973.

[14] A. Jain and R. Dubes, *Algorithms for Clustering Data*. Prentice Hall, 1988.

[15] B. Rengarajan and G. de Veciana, "Network architecture and abstractions for environment and traffic aware system-level coordination of wireless networks: The downlink case," in *Proc. IEEE INFOCOM*, April 2008, submitted.

[16] G. Hardy, J. E. Littlewood, and G. Polya, *Inequalities*. Cambridge, 1997.

[17] W. E. Young and R. H. Trent, "Geometric mean approximations of individual security and portfolio performance," *The Journal of Financial and Quantitative Analysis*, vol. 4, no. 2, pp. 179–199, Jun. 1969.

[18] W. H. Jean and B. P. Helms, "Geometric mean approximations," *The Journal of Financial and Quantitative Analysis*, vol. 18, no. 3, pp. 287–293, Sep. 1983.

[19] S. Borst, O. Boxma, and P. Jelenkovic, "Coupled processors with regularly varying service times," in *IEEE INFOCOM 2000*, vol. 1, 2000, p. 157164.

[20] S. Borst, O. Boxma, and M. van Uitert, "The asymptotic workload behavior of two coupled queues," *Queueing Systems*, vol. 43, no. 1-2, pp. 81–102, January 2003.

[21] G. Fayolle and R. Lasnogorodski, "Two coupled processors: The reduction to a riemann–hilbert problem," *Wahrscheinlichkeitstheorie*, no. 3, pp. 1–27, Jan. 1979.

[22] D. Stoyan, W. Kendall, and J. Mecke, *Stochastic Geometry and its Applications*. J. Wiley & Sons, Chichester, 1995.