System Level Optimization in Wireless Networks with Uncertain Customer Arrival Rates

Sungho Yun and Constantine Caramanis

***Abstract*— We consider a system-level approach to interference management in a cellular broadband system operating in an interference-limited and highly dynamic regime, as put forth in [1]. Here, base stations in neighboring cells (partially) coordinate their transmission schedules in an attempt to avoid simultaneous transmission to their mutual cell edge. Limits on communication overhead and use of the backhaul require base station coordination to occur at a slower time scale than the arriving customers. Depending on the overhead restrictions, the slower time scale could be on the scale of minutes or even hours. Thus base stations coordinate using only the *statistics* of customer arrival, while they serve users based on the actual realizations. The central challenge is to properly structure coordination decisions at the slow time scale, as these subsequently restrict the actions of each base station until the next coordination period.**

A further challenge comes from the fact that over longer coordination intervals, the statistics of the arriving customers, e.g., the load, may themselves vary or be only approximately known. Indeed, we show through simulation that while the approach of [1] is effective for a broad range of arriving load, performance rapidly degrades as the variation of the arriving load from the nominal (or assured) arriving load grows. We show this is true even when the variations are neutral, namely when the aggregate load is fixed, but there are local variations.

In this paper we show that a two-stage robust optimization framework is a natural way to model two time-scale decision problems. We provide tractable formulations for the base-station coordination problem, and show that our formulation is robust to fluctuations (uncertainties) in the arriving load. This tolerance to load fluctuation also serves to reduce the need for frequent re-optimization across base stations, thus helping minimize the communication overhead required for system level interference reduction. Building in robustness to load variation comes at the potential cost of somewhat degraded performance when variations happen to be very small. Our robust optimization formulations are flexible, allowing us to control the conservatism of the solution. Our simulations show that we can build in robustness without significant degradation of nominal performance.

I. INTRODUCTION

In down-link cellular systems with small cells and full frequency re-use, base station service rates are coupled by inter-cell interference, thus jeopardizing their performance. Thus minimizing the effect of inter-cell interference is paramount. Rather than physical-level approaches, we consider here a *system level* algorithmic approach, that aims to determine an optimal scheduling policy of the whole system simultaneously, to minimize joint transmissions of base stations to their mutual boundary. The naive implementation of such a scheme would require full knowledge of the distribution of arriving customers over the entire system at all time, requiring prohibitively high inter-base station communication overhead. To overcome this difficulty, Rengarajan and de Veciana [1] propose a novel framework in which customers are aggregated into classes, and then base stations coordinate at this coarse level, jointly optimizing a transmission schedule using only the statistics of customer (class) arrivals, and the offered load, allowing base-station coordination to occur at a much slower time scale than customer arrival.

This work was partially supported by NSF Grants EFRI-0735905, CNS-0721532 and a grant from DTRA. S. Yun and C. Caramanis are with the Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin TX, and part of the Wireless Communications and Networking Group (WNCG). e-mail: {shyun,caramanis}@mail.utexas.edu.

When the offered load is known at the time of coordination, the techniques proposed in [1] have been shown to decrease file transfer delay significantly at all load levels, while also increasing the uniformity in the coverage. As we demonstrate in the sequel, this exact knowledge of the offered load at each time seems to be crucial, as performance quickly degrades as uncertainty in the offered load increases.

In this paper we develop tools from robust and stochastic optimization to tackle this problem of load uncertainty. Our first contribution is in showing that multi-stage (for our purposes here two-stage) robust and stochastic optimization are the right optimization framework for considering distributed decision-making with coordination and uncertainty at different time scales. We consider uncertainty in the load, as well as uncertainty in its distribution among base stations. We formulate tractable (in particular, convex) optimization formulations for robust coordination. In extensive simulation experiments, we show that our robust and stochastic optimization formulations successfully immunize our solutions to variations in the load. At the same time, we investigate the so-called price of robustness, or conservatism, by considering the degradation of the nominal performance (how well does the robust solution do when there are no variations?). Our simulation results indicate that we can build in robustness without significant degradation in our nominal performance.

The structure of this paper is as follows. In Section II-A, we motivate the robust optimization counterpart of the system-level network optimization problem [1]. In Section II-B and Section II-C, we propose approaches that attempt to handle different uncertain models of arrival rates and show the simulation results and remarks. Finally, Section III concludes this paper.

II. Uncertain Arrival Rates

Rengarajan and de Veciana [1] propose a novel solution to enhance wireless broadband capacity in the case that transmissions in the neighboring cells generate interference and the base stations are coordinated. To take advantage of the diversity in users' sensitivity to interference from the neighboring cells, they first group customers in each base station into several classes according to their interference sensitivities and system loads. Then base stations jointly optimize a transmission policy determining when and which class of customers each base station would serve.

When the loads are known exactly, simulations reveal that the solution of [1] demonstrates remarkable improvements over a simple baseline no-coordination solution. Yet these gains deteriorate when the offered loads change over time at a faster scale that the base stations can re-optimize. In this paper, by using robust and stochastic optimization techniques ([3], [4], [6] and [2]), we propose an approach to make the solution robust to the changes of the offered loads.

We let N denote the number of base stations, with base station b serving K_b user classes. Arrivals to class k of base station b are Poisson with rate λ_{bk} and mean file size $\overline{F}_{bk}$, thus the offered load is $\rho_{bk} = \lambda_{bk}\overline{F}_{bk}$.

A. Two-stage Optimization

In [1], base stations coordinate, choosing joint power-and-class transmission schedules. To set the stage for our two-stage formulation, we consider separately the two different elements of the transmission profile: a power profile and a class profile. The power profile represents the transmit power level for each base station; the class profile represents the class that each base station will serve. Since the interference level seen by a base station depends only on the power level of its neighbors, and not which classes they might be serving, fixing a power profile is sufficient to specify interference to each base station. Therefore we can make the class profiles to be functions of realizations of actual offered loads and determine them after the actual loads are realized. Hence, the two-stage optimization setting becomes natural: before the actual offered loads for each class in each base station become known, base stations coordinate and decide upon the power profile schedule; next base stations decide on the class profile schedule after the offered loads become known without further communication with other base stations.

It is this two-stage formulation that allows us to consider robustness to uncertainty in the offered load. We consider two different uncertainty models for the variation in the offered loads, exploiting the strengths of robust and stochastic optimization, respectively. The first model is a *fixed total arrival rate* model: while the actual arrival rates of user classes fluctuate in each base station, the sum of fluctuations is always zero. We handle this using the robust optimization paradigm. So that rates fluctuate arbitrarily *in a given uncertainty set*, and the solution must be robust to all allowed variations. The second model is a *fixed arrival rates ratio* model in which the total arrival rate fluctuates while the fraction of each arrival rate for each base station and class is fixed. We use a stochastic optimization approach. The arrival rate varies according to a stochastic process, and the solution minimizes the expected customer delay. In both models, we deterministically enforce stability constraints. We believe that many other models, and combinations of the ones we treat here, can be approached using the methods we present in the subsequent sections of this paper.

We illustrate the main two-stage optimization formulation using the robust model. The stochastic model is considered in detail in Section II-C. Let $\vec{\lambda}$ denote the (unknown) offered load, varying in an uncertainty set Z. The decision variables $\{\alpha_l\}_{l=1}^{L}$ represent the joint decisions on power profile coordination, with each $l = 1, \ldots, L$ denoting a different joint power profile. The second stage decisions are given by $p_{bk}^{l}(\vec{\lambda})$. Their explicit dependence on $\vec{\lambda}$ indicates that they are second-stage decisions, made after the uncertainty realization. We write them as general functions here for clarity of exposition. To solve the optimization, we must restrict the class of functions, so as to maintain tractability (in particular, convexity) of the problem. In the next two sections (for both stochastic and robust formulations), we restrict to *affine functions of the uncertainty*. We obtain the following robust optimization problem:

$$\begin{array}{lllll} \min_{\vec{\alpha}, p(\vec{\lambda})} & f(\vec{\alpha}, \vec{p}(\vec{\lambda})) & & \\ \text{s.t.} & \rho_{bk} \le R_{bk}(\vec{\alpha}, \vec{p}(\vec{\lambda})) & \forall b,k & \forall \vec{\lambda} \in Z \\ & \sum_{k=1}^{K_b} p_{bk}^{l}(\vec{\lambda}) \le \alpha_l & \forall b,l & \forall \vec{\lambda} \in Z \\ & \sum_{l=1}^{L} \alpha_l \le 1 & & \\ & p_{bk}^{l}(\vec{\lambda}) \ge 0 & \forall b,k,l & \forall \vec{\lambda} \in Z \\ & \alpha_l \ge 0 & \forall l & \end{array}, \quad (1)$$

where $R_{bk}(\vec{\alpha}, \vec{p}(\vec{\lambda}))$ is the capacity allocated to class k in base station b by schedule $\vec{\alpha}, \vec{p}(\vec{\lambda})$. If base stations use processor sharing to serve users within each class, R_{bk} is given by a harmonic mean (see [1]). Optimizing using the harmonic mean is difficult, as it is nonconvex, and for this reason, we use an arithmetic mean approximation. Minimizing $f = \sum_{l=1}^{L} \alpha_l$ corresponds to a capacity optimizing schedule, while $f = \sum_{b=1}^{N} \sum_{k=1}^{K_b} \frac{\frac{\rho_{bk}}{R_{bk}(\vec{\alpha}, \vec{p}(\vec{\lambda}))}}{1 - \frac{\rho_{bk}}{R_{bk}(\vec{\alpha}, \vec{p}(\vec{\lambda}))}}$ minimizes the total average delay. System stability is enforced in the first constraint.

We note that with a singleton uncertainty set, we recover the original single-stage formulation in [1].

B. Uncertain Arrival Rates Ratio Model with Fixed Total Arrival Rate

1) Assumptions and an Adjustable Counterpart: In this section we consider a *fixed total arrival rate* model. If load fluctuations are independent across classes, by LLN results, fixed total arrival rate holds in the limit of many classes. While we do not address it in this paper, we note that it is equally easy to treat the case where the sum of fluctuations is small but not necessarily zero.

Let us define an uncertainty set, Z_b for base station b

$$Z_b = \left\{ \begin{array}{l} \vec{\lambda}_b = (\lambda_{b1}, \cdots, \lambda_{bK_b}) | \\ \quad \forall k, \lambda_{bk} \in [(1-\theta)\lambda^*_{bk}, (1+\theta)\lambda^*_{bk}], \\ \quad \sum_{k=1}^{K_b} \lambda_{bk} = \sum_{k=1}^{K_b} \lambda^*_{bk} \equiv \lambda^*_b \end{array} \right\}, \quad (2)$$

where λ^*_{bk} is the nominal arrival rate of user class k in base station b. The system designer adjusts the parameter θ to balance the conservatism and robustness of the solution. Since we enforce feasibility for all realizations in the uncertainty set, larger θ results in a more robust and conservative solution. For $\theta = 0$, we recover the nominal optimal solution. For each base station b, $\vec{p}$ depends on the arrival rates in its cell. Thus second stage decision variables tune themselves according to the offered loads, and no additional communication overhead is required.

2) Affinely Adjustable Robust Counterpart (AARC): We must restrict the structure of the functions representing the second stage decisions in (1), in order to be able to solve the resulting optimization problem. Restricting $p^{\,l}_{bk}$ to be an affine function of $\vec{\lambda}_b$, as in [4], [7], we have $p^{\,l}_{bk}(\vec{\lambda}_b) = \pi^{\,l}_{bk0} + \sum_{m=1}^{K_b} \pi^{\,l}_{bkm}\lambda_{bm}$. Using the linear capacity optimizing objective, we obtain a linear two-stage robust optimization problem:

$$\begin{array}{ll} \min_{\vec{\alpha},\vec{\pi}} & \sum_{l=1}^{L} \alpha_l \\ \text{s.t.} & \rho_{bk} \le \sum_{l=1}^{L} (\pi^{\,l}_{bk0} + \sum_{m=1}^{K_b} \pi^{\,l}_{bkm}\lambda_{bm}) E_I[R^l_I | b,k] \\ & \qquad \forall b,k\ , \forall \vec{\lambda}_b \in Z_b \\ & \sum_{k=1}^{K_b} (\pi^{\,l}_{bk0} + \sum_{m=1}^{K_b} \pi^{\,l}_{bkm}\lambda_{bm}) \le \alpha_l \\ & \qquad \forall b,l\ , \forall \vec{\lambda}_b \in Z_b \\ & \sum_{l=1}^{L} \alpha_l \le 1 \\ & \pi^{\,l}_{bk0} + \sum_{m=1}^{K_b} \pi^{\,l}_{bkm}\lambda_{bm} \ge 0 \\ & \qquad \forall b,k,l\ , \forall \vec{\lambda}_b \in Z_b \\ & \alpha_l \ge 0 \qquad \forall l \end{array} \quad (3)$$

Theorem 1: The AARC (3) is equivalent to the following linear optimization problem.

$$\begin{array}{lll} \min_{\vec{\alpha},\vec{\pi},\vec{\beta},\vec{\zeta},\vec{\mu},\vec{\sigma},\vec{\nu}} & F & \\ \text{s.t.} & \beta^{\,l}_{bk0} - \beta^{\,l}_{bkm} \le \pi^{\,l}_{bkm} \le \beta^{\,l}_{bk0} + \beta^{\,l}_{bkm} & \forall b,k,l,m \\ & \zeta_{bkm} = \sum_{l=1}^{L} \pi^{\,l}_{bkm} E[R^l_I | b,k] & \forall b,k,m,k \ne m \\ & \zeta_{bkm} = \sum_{l=1}^{L} \pi^{\,l}_{bkm} E[R^l_I | b,k] - \overline{F}_{bk} & \\ & & \forall b,k,m,k = m \\ & \mu_{bk0} - \mu_{bkm} \le \zeta_{bkm} \le \mu_{bk0} + \mu_{bkm} & \forall b,k,m \\ & \sigma^{\,l}_{bm} = \sum_{k=1}^{K_b} \pi^{\,l}_{bkm} & \forall b,l,m \\ & \nu^{\,l}_{b0} - \nu^{\,l}_{bm} \le \sigma^{\,l}_{bm} \le \nu^{\,l}_{b0} + \nu^{\,l}_{bm} & \forall b,l,m \\ & \sum_{l=1}^{L} \pi^{\,l}_{bk0} E_I[R^l_I | b,k] & \\ & \quad + \sum_{m=1}^{K_b} \zeta_{bkm}\lambda^*_{bm} & \\ & \quad - \theta \sum_{m=1}^{K_b} \mu_{bkm}\lambda^*_{bm} \ge 0 & \forall b,k \\ & \sum_{k=1}^{K_b} \pi^{\,l}_{bk0} + \sum_{m=1}^{K_b} \sigma^{\,l}_{bm}\lambda^*_{bm} & \\ & \quad + \theta \sum_{m=1}^{K_b} \nu^{\,l}_{bm}\lambda^*_{bm} \le \alpha_l & \forall b,l \\ & \pi^{\,l}_{bk0} + \sum_{m=1}^{K_b} \pi^{\,l}_{bkm}\lambda^*_{bm} & \\ & \quad - \theta \sum_{m=1}^{K_b} \beta^{\,l}_{bkm}\lambda^*_{bm} \ge 0 & \forall b,k,l \\ & \beta^l_{bkm} \ge 0 & \forall b,k,m,l,m \ne 0 \\ & \zeta_{bkm} \ge 0 & \forall b,k,m,m \ne 0 \\ & \sigma^l_{bm} \ge 0 & \forall b,m,l,m \ne 0 \\ & \sum_{l=1}^{L} \alpha_l \le F & \\ & \sum_{l=1}^{L} \alpha_l \le 1 & \\ & \alpha_l \ge 0 & \forall l \end{array} \quad (4)$$

Proof: Consider the 4^{th} constraint of (3) :

$$\pi^{\,l}_{bk0} + \sum_{m=1}^{K_b} \pi^{\,l}_{bkm}\lambda_{bm} \ge 0 \ \ \forall b,k,l, \ \ \forall \vec{\lambda}_b \in Z_b.$$

This constraint holds if and only if the optimal value in the following problem

$$\begin{array}{lll} \min_{\vec{\lambda}_b} & \pi^{\,l}_{bk0} + \sum_{m=1}^{K_b} \pi^{\,l}_{bkm}\lambda_{bm} & \\ \text{s.t.} & \lambda_{bm} \ge (1-\theta)\lambda^*_{bm} & \forall m \in [1,\cdots,K_b] \\ & \lambda_{bm} \le (1+\theta)\lambda^*_{bm} & \forall m \in [1,\cdots,K_b] \\ & \sum_{m=1}^{K_b} \lambda_{bm} = \lambda^*_b & \end{array} \quad (5)$$

is nonnegative.

By the Duality Theorem of linear programming, the optimal value is nonnegative if and only if the corresponding dual problem

$$\begin{array}{lll} \max_{\vec{\gamma},\vec{\delta},\xi} & \sum_{m=1}^{K_b} (1-\theta)\lambda^*_{bm}\gamma_m & \\ & \quad - \sum_{m=1}^{K_b} (1+\theta)\lambda^*_{bm}\delta_m + \lambda^*_b\xi + \pi^{\,l}_{bko} & \\ \text{s.t.} & \gamma_m - \delta_m + \xi = \pi^{\,l}_{bkm} & \forall m \in [1,\cdots,K_b] \\ & \vec{\gamma} \ge 0 & \\ & \vec{\delta} \ge 0 & \end{array} \quad (6)$$

has a nonnegative optimal value., i.e., $\exists \vec{\gamma}$, $\vec{\delta}$, and ξ s.t.

$$\begin{array}{ll} \sum_{m=1}^{K_b} (1-\theta)\lambda^*_{bm}\gamma_m & \\ \quad - \sum_{m=1}^{K_b} (1+\theta)\lambda^*_{bm}\delta_m + \lambda^*_b\xi + \pi^{\,l}_{bko} \ge 0 & \\ \gamma_m - \delta_m + \xi = \pi^{\,l}_{bkm} & \forall m \in [1,\cdots,K_b] \\ \vec{\gamma} \ge 0 & \\ \vec{\delta} \ge 0 & \end{array} \quad (7)$$

Let $\beta^{\,l}_{bkm} = \gamma_m + \delta_m$ and $\beta^{\,l}_{bk0} = \xi$. Then (7) is equivalent to

$$\begin{array}{l} \sum_{m=1}^{K_b} \lambda^*_{bm}(\pi^{\,l}_{bkm} - \beta^{\,l}_{bk0}) \\ \quad - \theta \sum_{m=1}^{K_b} \lambda^*_{bm}\beta^{\,l}_{bkm} + \lambda^*_b\beta^{\,l}_{bk0} + \pi^{\,l}_{bko} \ge 0 \\ 2\gamma_m = \beta^{\,l}_{bkm} - \beta^{\,l}_{bk0} + \pi^{\,l}_{bkm} \ge 0 \\ 2\delta_m = \beta^{\,l}_{bkm} - \beta^{\,l}_{bk0} - \pi^{\,l}_{bkm} \ge 0 \end{array} \quad (8)$$

which is equivalent to

$$\begin{array}{l} \pi^{\,l}_{bk0} + \sum_{m=1}^{K_b} \pi^{\,l}_{bkm}\lambda^*_{bm} - \theta \sum_{m=1}^{K_b} \beta^{\,l}_{bkm}\lambda^*_{bm} \ge 0 \\ \beta^{\,l}_{bk0} - \beta^{\,l}_{bkm} \le \pi^{\,l}_{bkm} \le \beta^{\,l}_{bk0} - \beta^{\,l}_{bkm} \\ \beta^{\,l}_{bkm} \ge 0 \end{array} \quad (9)$$

Similar arguments can be applied for the rest of the constraints in (3). ■

Solving the resulting LP, we obtain an affine policy for determining the class profile schedule for each base station, as a function of its local offered load variation.

3) Simulations and Results: We evaluate the performance of our affine policy using the same simulation model of Rengarajan and de Veciana [1]. We consider three base stations facing each other in a hexagonal layout with radius 250m. A carrier frequency of 1GHz and a bandwidth of 10MHz are assumed. The base stations are assumed to be able to transmit at three different power levels: 0, 5 and 10W. The mean file size is 2MB.

We used different total arrival rates from 0.5 to 2.2 and different protection levels, θ from 0% to 40%. We used 100,000 customer samples to estimate the harmonic formula capacities and the mean delay. For each pair of total arrival rate and protection level, we randomly picked λ_{bk}'s in their bounds and computed the estimated delay 1,000 times to get average performance under the proposed uncertainties. Out of 1,000 experiments, we counted the number of cases that the system became unstable and computed the average mean delay under the proposed uncertainties. As shown in Fig. 1(a), at higher loads and higher uncertainty levels the number of unstable experiments is larger with the solution of the nominal optimization problem, i.e., $\theta = 0$. On the other hand, as shown in Fig. 1(b), with the solution of the AARC, the number of unstable experiments remains at zero until very high load, 2.0/sec. Note that even without fluctuations of arrival rates, the system is unstable around that point.

Fig. 2(a) and Fig. 2(b) show the average mean delay with the solution of nominal optimization problem and the solution of the AARC respectively, at each load and uncertainty level. Note that we assumed arbitrarily large delays for the unstable experiments, since we cannot compute the mean delay of unstable systems. To make this assumption reasonable, we assumed that the delay of an unstable experiment is as large as the maximum delay over all delays of stable experiments. At lower loads, even with high uncertainty levels the nominal solutions slightly outperform the AARC solutions. However, at higher loads, while the AARC solutions give acceptable low average mean delays, the nominal solutions give extremely high average mean delays even if the system is stable.

As we typically do not precisely know the uncertainty level in reality, we must balance the tradeoff between building in protection to uncertainty, and the loss of performance in the nominal setting, i.e., the cost of over-protection. To compare these factors, we pick two protection levels, 20% and 40%, and we consider the performance of these three solutions in different uncertainty regimes. Figure 3(a) shows the load at which stability breaks down for each solution, under large (40%) uncertainty. The nominal solution becomes unstable at a much lower average arrival rate, than the 20%-protection and 40%-protection solutions. Interestingly, the 20%-protection solution remains stable for very heavy loads – essentially its stability performance is equivalent to the 40%-protection solution. Figures 3(b,c,d) show the average mean delay (from simulation) of our three solutions. Figure 3(b) shows the delay curves when there is *no uncertainty* in arrival rates, i.e., the simulations are generated according to the known offered load. This shows the price of robustness. Indeed, as expected, the nominal solution outperforms both robust solutions giving lower delay – but the difference becomes pronounced only at very high loads. Meanwhile, Figures 3(c) and (d) illustrate the relatively quick deterioration of the nominal solution under 20% and 40% uncertainty in the offered load. These results illustrate that the 20%-protection solution appears to have low price of robustness, i.e., performance comparable to the nominal solution in the no-uncertainty regime, and yet captures most of the robustness properties of even the 40%-protection solution, outperforming the latter, except when both the load and the uncertainty level are high.

C. Uncertain Total Arrival Rate Model with Fixed Arrival Rates Ratio

1) Assumptions and an Adjustable Counterpart: In this section we consider a stochastic uncertainty model. We assume the arrival process is Poisson with rate λ and users arrive uniformly over the entire area. Thus for each class k and base station b, the fraction of arrival rate is fixed regardless of the change of the total arrival rate. We assume the total arrival rate changes according to a Markov process with drift towards the nominal rate. We discretize this process, approximating it via a Discrete Markov chain.

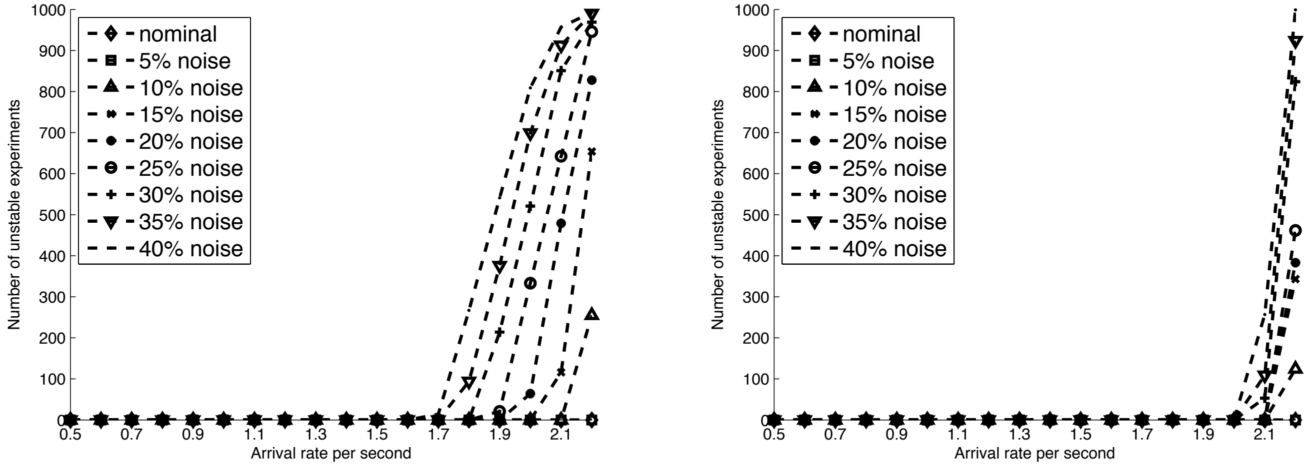
We let Δ represent the quantized unit step the arrival rate can move, with probabilities $\bar{p}$ and $\bar{q}$ representing the drift away from and towards the nominal rate, respectively. The following is the transition matrix of the Markov chain we use:

$$\begin{aligned}
&\forall i \geq 0,\\
&Pr(\lambda(t+1) = \lambda^*(1+(i+1)\Delta)|\lambda(t) = \lambda^*(1+i\Delta)) = \bar{p}\\
&Pr(\lambda(t+1) = \lambda^*(1+i\Delta)|\lambda(t) = \lambda^*(1+(i+1)\Delta)) = \bar{q}\\
\\
&\forall i \leq 0,\\
&Pr(\lambda(t+1) = \lambda^*(1+i\Delta)|\lambda(t) = \lambda^*(1+(i-1)\Delta)) = \bar{q}\\
&Pr(\lambda(t+1) = \lambda^*(1+(i-1)\Delta)|\lambda(t) = \lambda^*(1+i\Delta)) = \bar{p},
\end{aligned} \tag{10}$$

Although we know the full distribution of the arrival rate, it is hard to compute the expected mean delay over the infinitely supported distribution of the uncertainty. To overcome this difficulty, we accept some error probability ϵ and truncate the distribution of the arrival rate into a finitely supported distribution. Let n be a number s.t. $Pr(\lambda \in [\lambda^*(1-n\Delta), \lambda^*(1+n\Delta)]) \geq 1-\epsilon$.

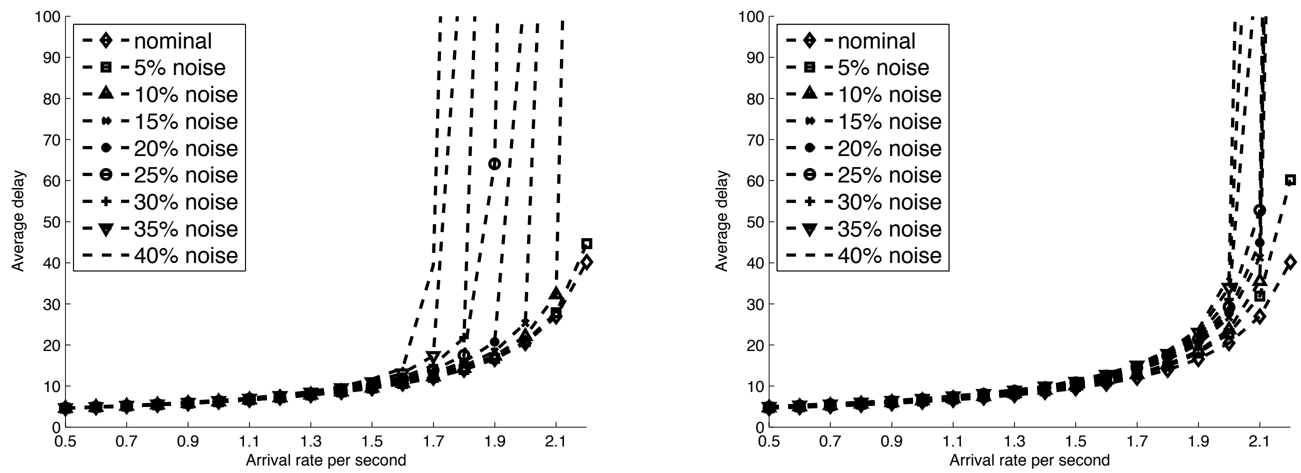
2) Affinely Adjustable Stochastic Counterpart of Objective: Computational experiments in [1] reveal that the delay minimizing form yields improved delay performance over the capacity optimizing form. In the stochastic formulation, the objective is an expected value over a discretely supported distribution, and hence becomes a sum of weighted variations. If the original objective is convex, then so is its expectation. Exploiting this fact and taking advantage of the knowledge of the distribution, we use the delay-minimizing objective rather than the capacity-optimizing objective we use in the robust formulation.

Then the objective of the stochastic optimization problem is as follows.



(a) (b)

Fig. 1. Number of unstable experiments out of 1000 simulations against different actual uncertainty levels ranging from 0% to 40%: (a) the nominal solution, (b) each AARC solution runs against its predicted uncertainty level



(a) (b)

Fig. 2. Average file transfer delay against different actual uncertainty levels ranging from 0% to 40%: (a) the nominal solution, (b) each AARC solution runs against its predicted uncertainty level

$$
\begin{aligned}
&E_\lambda\left[\sum_{b=1}^{N}\sum_{k=1}^{K_b}\frac{\frac{\rho_{bk}}{R_{bk}(\vec{p}_{bk}(\lambda))}}{1-\frac{\rho_{bk}}{R_{bk}(\vec{p}_{bk}(\lambda))}}\right]\\
&=\sum_{i\in\{-n,n\}} Pr(\lambda=\lambda^*(1+i\Delta))\\
&\qquad\times\left(\frac{1}{\lambda}\sum_{b=1}^{N}\sum_{k=1}^{K_b}\frac{\frac{\rho_{bk}}{R_{bk}(\vec{p}_{bk}(\lambda))}}{1-\frac{\rho_{bk}}{R_{bk}(\vec{p}_{bk}(\lambda))}}\right)
\end{aligned}
$$

3) Affinely Adjustable Robust Counterpart of Constraints: In the truncated distribution, the support of the total arrival rate, λ, is $[\lambda^*(1-n\Delta), \lambda^*(1+n\Delta)]$. Next, we restrict p_{bk}^{l} to be an affine function of λ, i.e., $p_{bk}^{l}=\pi_{bk0}^{l}+\pi_{bk1}^{l}\lambda$. In order to make the solution feasible for every realization of the arrival rate, the constraints must remain feasible for all arrival rates in the support of the truncated distribution, which requires the similar robust optimization techniques we used in Section II-B. Let $\theta=n\Delta$ and $Y=[\lambda^*(1-\theta),\lambda^*(1+\theta)]$. Then as we discussed in Section II-B.1, θ represents the conservatism of the solution and Y is the support of the uncertainty set. We can control θ by adjusting the truncation error ϵ. If we allow smaller truncation error, then the support of the uncertainty set gets larger, i.e., n gets larger, hence the conservatism level θ becomes higher. Let $\tilde{\lambda}_{bk}$ be the fraction of arrival rate of user class k in base station b. Then $\lambda_{bk}=\tilde{\lambda}_{bk}\lambda$, and the resulting optimization problem is as follows.

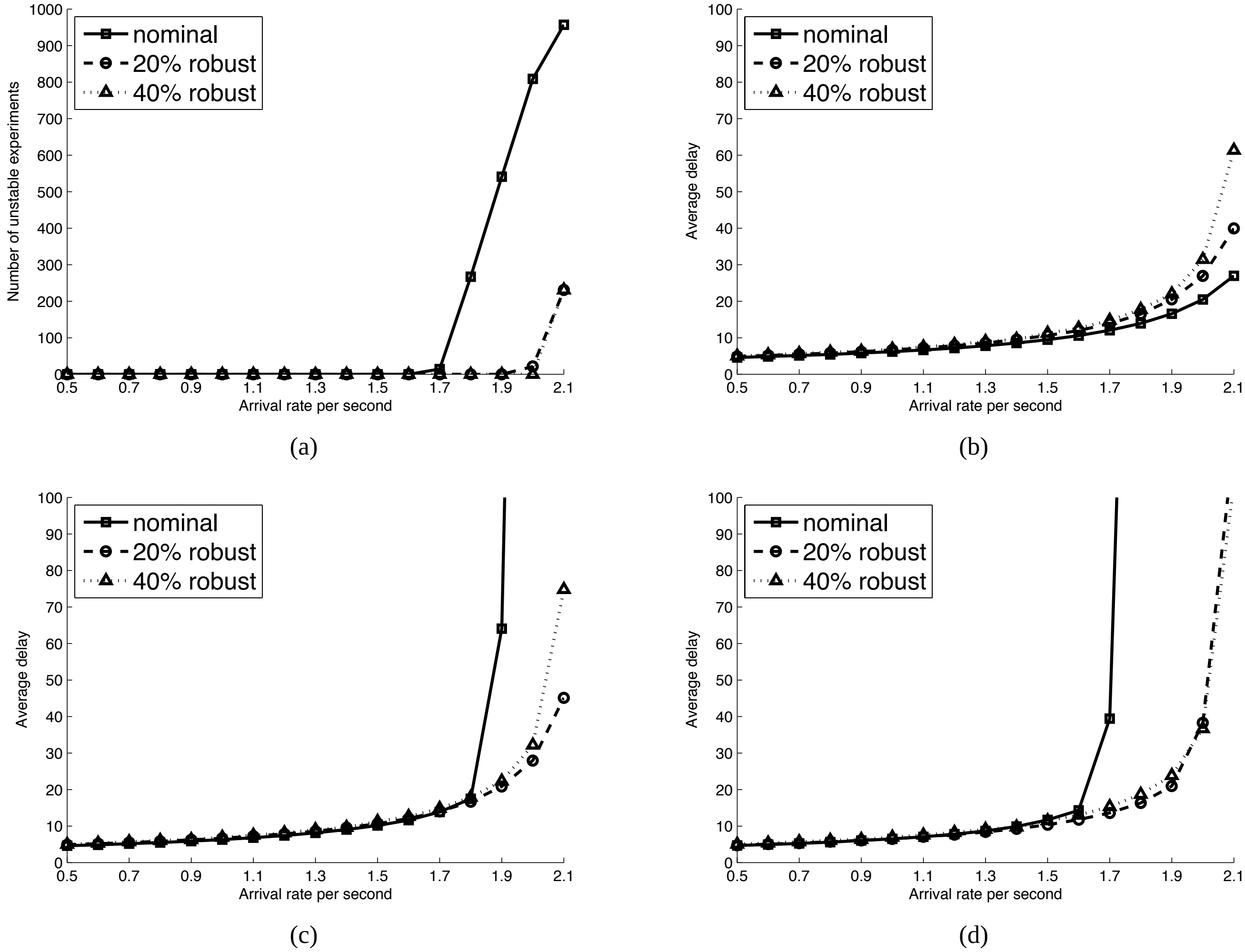


Fig. 3. (a) Number of unstable experiments out of 1000 simulations against 40% uncertainty level of offered loads, (b), (c) and (d) Average file transfer delay : (b) at the nominal offered loads, (c) against 20% uncertainty level of offered loads, (d) against 40% uncertainty level of offered loads

$$\begin{array}{lll}\min_{\vec{\alpha},\vec{\pi},\vec{\beta},\vec{\zeta},\vec{\mu},\vec{\sigma},\vec{\nu}} & \sum_{i\in\{-n,n\}} Pr(\lambda=\lambda^*(1+i\Delta)) & \\ & \times\left(\frac{1}{\lambda}\sum_{b=1}^{N}\sum_{k=1}^{K_b}\frac{\frac{\rho_{bk}}{R_{bk}(\vec{p}_{bk}(\lambda))}}{1-\frac{\rho_{bk}}{R_{bk}(\vec{p}_{bk}(\lambda))}}\right) & \\ \text{s.t.} & \tilde{\lambda}_{bk}\lambda\overline{F}_{bk}\le\sum_{l=1}^{L}(\pi_{bk0}^{\ l}+\pi_{bk1}^{\ l}\lambda)E_I[R_I^l|b,k] & \\ & & \forall b,k\ \ \forall\lambda\in Y \\ & \sum_{k=1}^{K_b}(\pi_{bk0}^{\ l}+\pi_{bk1}^{\ l}\lambda)\le\alpha_l & \forall b,l\ \ \forall\lambda\in Y \\ & \sum_{l=1}^{L}\alpha_l\le 1 & \\ & \pi_{bk0}^{\ l}+\pi_{bk1}^{\ l}\lambda\ge 0 & \forall b,k,l\ \ \forall\lambda\in Y \\ & \alpha_l\ge 0 & \forall l\end{array}, \tag{11}$$

Since we use an arithmetic mean approximation for R_{bk}, and since $\vec{p}_{bk}(\lambda)$ is linear in $\vec{\pi}$, R_{bk} is linear in $\vec{\pi}$. The term $\frac{\frac{\rho_{bk}}{R_{bk}(\vec{p}_{bk}(\lambda))}}{1-\frac{\rho_{bk}}{R_{bk}(\vec{p}_{bk}(\lambda))}}$ in the objective function is convex in $\vec{\pi}$ if $\frac{\rho_{bk}}{R_{bk}(\vec{p}_{bk}(\lambda))}\le 1$ ([1]) and indeed we enforce this for every λ in the support of its truncated distribution. Moreover, the infinite constraints ("$\forall\lambda\in Y$") can be transformed into a finite collection of linear constraints, again by employing a duality argument. Therefore this problem is a convex optimization problem with linear constraints.

4) Simulations and Results: Again, we evaluate the performance of our optimization using the same simulation model of [1]. We used different nominal arrival rates ranging from 0.5 to 2.2 and different truncation error, ϵ ranging from 1% to 20%. As we see in Section II-C.3, *lower truncation error means higher protection level*. We used 100,000 customer samples to estimate the harmonic formula capacities and the mean delay. At each simulation, we chose an arrival rate randomly from the stationary distribution of our Markov chain model. For each pair of nominal arrival rate and error probability, we ran 1,000 experiments. We counted the number of cases that the system is unstable and computed the average mean delay under the Markov chain model. We used the transition matrix (10) with $\bar{p}=1/3$, $\bar{q}=2/3$, and $\Delta=6\%$.

As shown in Fig. 4(a), even under nominal arrival rates, our uncertainty-protected solutions work comparably to the nominal solution, hence the price of robustness is very low in this model.

Fig. 5(a) shows the number of unstable experiments for each solution and Fig. 5(b) shows the average delay. At higher loads, the number of unstable experiments is larger with the nominal solution. But, even the 20% truncation error model shows better results than the nominal solution in terms of stability and optimality. This is because the distribution of the uncertainty is concentrated around its mean. Even the truncated distribution with 20% error captures the original distribution well. The original distribution of the Markov chain with transition matrix (10) and the truncated distributions under different truncation errors are shown in Fig. 4(b). On the other hand, the constraints with less than 20% truncation error models (i.e., bigger θ) are overly protective, and the conservativeness of the solution results in infeasibility at higher loads.

III. DISCUSSION, CONCLUSION, AND FUTURE WORK

We proposed two different approaches that attempt to make the solution of the system level coordination optimization problem robust to the variations of offered loads under different models of uncertain data. In the case that each offered load fluctuates individually but the sum of variations is zero, we used two-stage robust optimization with affine second-stage decisions, obtaining tractable optimization formulations to obtain solutions robust to variations in the offered load. We also considered variation in the total arrival rate. There, we combined the stochastic optimization and the robust optimization paradigms, again obtaining solutions that remain stable under heavy loads, and get better average performance. In our simulation results, we have shown that nominal solutions are vulnerable to the fluctuations of the offered loads while properly tuned robust solutions capture the best of both worlds: resilience to uncertainty, with good performance even under the nominal setting.

Resilience to load variation could potentially help reduce coordination and hence communication requirements, without severely compromising the performance. Understanding the tradeoffs involved, between the benefits and costs of more frequent coordination is a key step towards understanding the viability of implementation of such a system-level optimization approach to interference mitigation, and is a topic of future work.

An issue we have not addressed here, and the subject of future work, is to treat the stochastic variation in the customer arrival process. Particularly in the low-load regime, this could result in empty customer classes, allowing base stations to (briefly) turn off, thus increasing the rates observed by customers of neighboring base stations. Optimizing coordination schemes to take advantage of this effect is a natural domain for multi-stage optimization models, although some considerable challenges stand in the way of immediate extensions of the methods presented here.

REFERENCES

[1] Rengarajan, B., de Veciana, G., *Network Architecture and Abstractions for Environment and Traffic Aware System-Level Coordination of Wireless Networks: The Downlink Case*, IEEE Infocom, pp 1-9, (2008) To appear.

[2] Birge, J.R., Louveaux, F., *Introduction to stochastic program*, Springer, New York, (1997).

[3] Ben-Tal, A., Nemirovski, A., *Robust solutions of linear programming problems contaminated with uncertain data*, Math. Progr., 88, pp. 411-424, (2000).

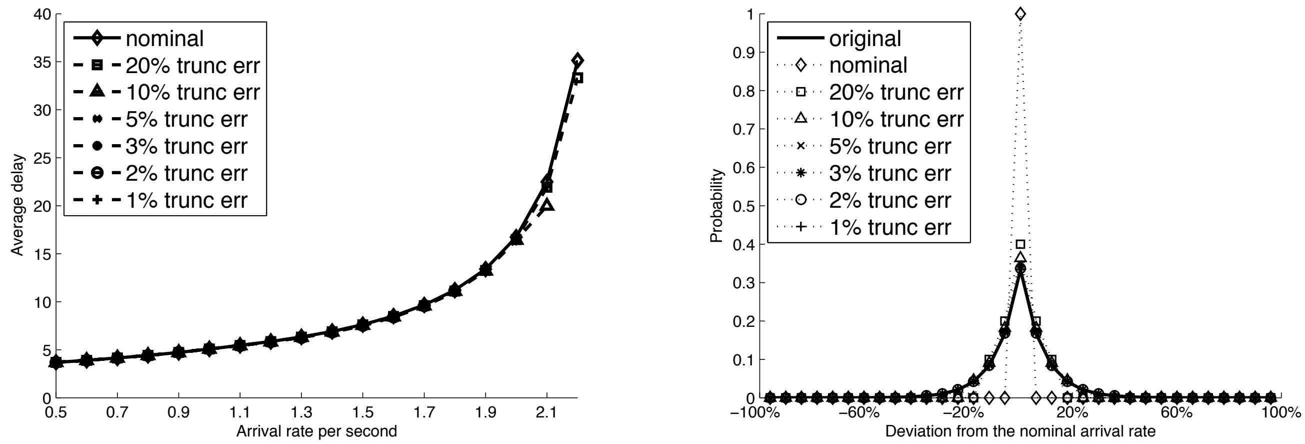
[4] Ben-Tal, A., Goryashko, A., Guslitzer, E., Nemirovski, A., *Adjustable robust solutions of uncertain linear programs*, Math. Progr., 99, pp. 351-376, (2004).

[5] Ben-Tal, A., Boyd, S., Nemirovski, A., *Extending scope of robust optimization: Comprehensive robust counterparts of uncertain problems*, Math. Progr., 107, pp. 63-89, (2006).

[6] Bertsimas, D., Sim, M., *Price of robustness*, Oper. Res., 52(1), pp. 35-53, (2004).

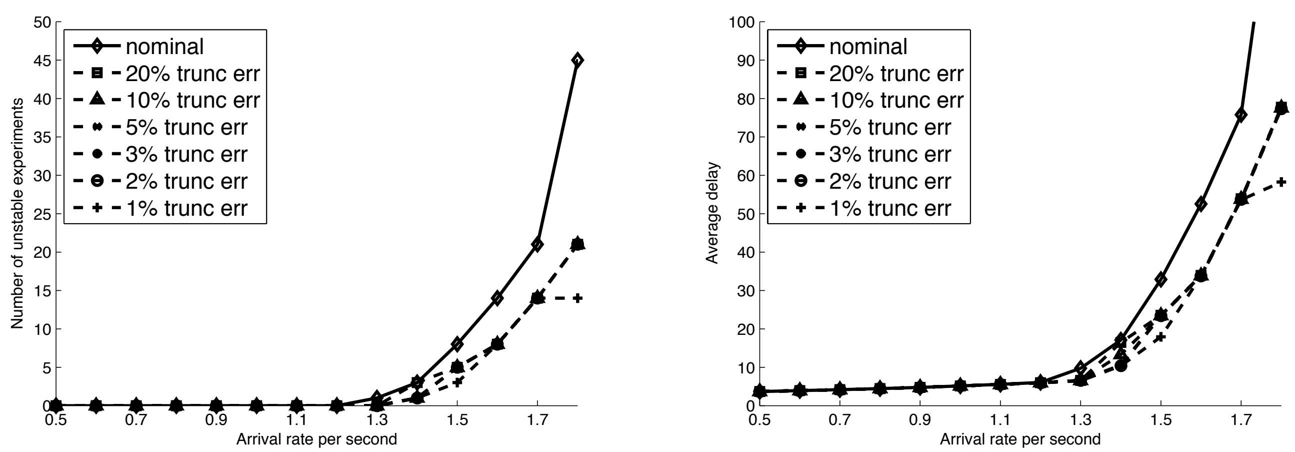
[7] Chen, X., Sim, M., Sun, P., Zhang, J., *A linear-decision based approximation approach to stochastic programming*, Oper. Res., (2007).

[8] Yun, S., *System Level Optimization in Wireless Networks with Uncertain Customer Arrival Rates*, Masters Degree Report., University of Texas at Austin, (2008).



(a) (b)

Fig. 4. (a) Number of unstable experiments out of 1000 simulations at the nominal arrival rates, (b) Distribution of the uncertain arrival rate: Original distribution of the Markov chain v.s. Truncated distributions under different truncation errors



(a) (b)

Fig. 5. against the Discrete Markov chain uncertainty model of arrival rates with transition matrix (10) : (a) Number of unstable experiments, (b) Average file transfer delay