Rate Control and Dimensioning of Infrastructure-based Wireless Networks

Ravi Mazumdar
Dept. of Electrical and Computer Engineering
University of Waterloo
Waterloo, Canada N2L 3G1

Joint Work with Aditya Karnik and Catherine Rosenberg

Communication Networks and Rate Allocation

- Allocation of bandwidths to network applications sharing a communication infrastructure is a fundamental problem in networking
- Owing to the dominance of data traffic, applications which can tolerate variable rate of data transfer and delay but are sensitive to loss are of particular importance
- This elastic (and decentralized) nature of applications raises two basic questions
  1. What constitutes a good and acceptable distribution of the network capacity to its users?
  2. How can applications determine such an allocation themselves in a distributed fashion?

Rate Allocation in Wired Networks

- In the context of wired networks these questions have been resolved
- Network capacity is fixed and representable in the form of individual link capacities
- Hence the problem can be treated as an economic problem of fair and efficient allocation of commodities to consumers
- A fair and efficient allocation then may be seen as the one which
  - maximizes aggregate utility under the link capacity constraints

$$\max \sum_{s=1}^{N} U_s(x_s) \ \text{subj.to} \ \sum_{s \in S_l} x_s \le c_l, \ \ l = 1, \dots, L$$

  - is the Nash Bargaining solution

$$\max \Pi_{s=1}^{N} (x_s - x_s^0) \ \text{subj.to} \ \sum_{s \in S_l} x_s \le c_l, \ \ l = 1, \dots, L$$

Rate Allocation in Wired Networks

- Either of these solution concepts can be realized in a distributed fashion
  - in utility maximization there exists price p_l^* for each link l such that

$$x_s^* = \arg\max_{x_s \geq 0} U_s(x_s) - \left(\sum_{l \in L_s} p_l^* \right) x_s, \quad s = 1, \ldots, N$$

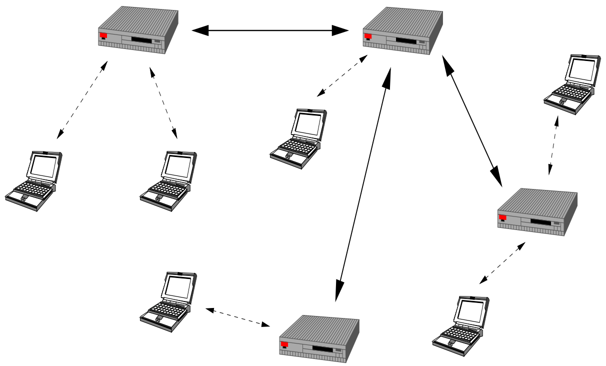
$$p_l^* \left(c_l - \sum_{s \in S_l} x_s^* \right) = 0, \quad l = 1, \ldots, L$$

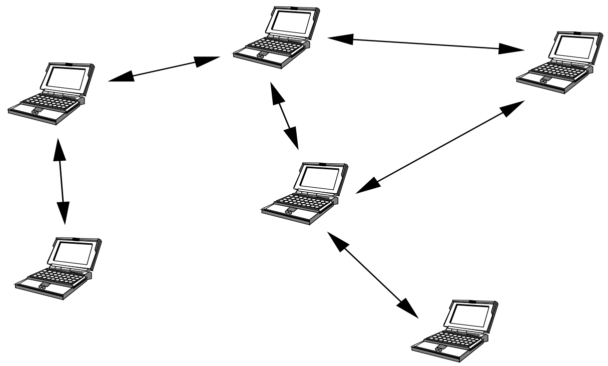
- Optimal rates and prices can be obtained iteratively
- Prices can be interpreted as congestion signals from the network links
- Interestingly, TCP congestion control has been identified as a Primal-dual algorithm maximizing certain utility functions

Rate Allocation in Wireless Networks

- The problem is yet to be completely settled in the case of wireless networks
- Very concept of capacity is not straightforward for wireless links in a network
  - capacities cannot be assigned to individual links in networks which employ randomized MAC protocols
  - the problem cannot be simply cast as that of dividing individual link capacities in general
- Radio transmission parameters can be modified on the fly
  - network capacity can be dimensioned dynamically rather than taking it as fixed and given
  - rate allocation is tied to optimization of the network itself
  - a precise characterization of network capacity in terms of protocol parameters is difficult to obtain

Rate Allocation in Wireless Networks





- Infrastructure-based networks (e.g., IEEE 802.16 Broadband Wireless),
  - networking functionality is built into the network devices whereas the applications act as the end users
  - the network and the users can be seen as separate economic entities
- Wireless ad hoc networks (e.g., IEEE 802.11 ad hoc mode)
  - the user devices not only act as application endpoints but also as routers
  - the users themselves constitute the network and can be seen as competing economic entities

Implications

- New questions open up for wireless networks

  3. How to formulate the constraints on data rates?
  4. How to dimension the network in conjunction with the allocation of rates?
  5. How can the networking devices determine an optimal configuration of parameters in a distributed fashion?

- We will be concerned with questions 3 and 4 (along with 1 and 2)
- We address these issues in the case of infrastructure-based networks
  - users can be seen as consumers, data rates as produced commodities, network as a producer and protocol parameters as factors of production
  - network dimensioning in response to user rate requirements and vice versa is not a cross-layer optimization issue

Formulating the Rate Constraints

- The capacity of a communication link is the maximum rate at which data units can be transported across it successfully
- Unlike a wired link, the capacity of a wireless link is not invariant to what transpires on the other wireless links in the network
- The interference problem necessitates some sort of coordination among the wireless links
- How this coordination is achieved is the crux in formulation of rate constraints

Scheduled Networks

- If the coordination is achieved through a pre-computed conflict-free link transmission schedule, then it is possible to assign each link a capacity
- Capacity is simply the link capacity in isolation multiplied by fraction of time allocated to the link for transmission in the schedule
- User rate constraints can be split into individual link constraints as in the wired networks
- Link capacities are parameterized by the transmission schedule and PHY parameters

Unscheduled Networks

- If the transmissions are coordinated in response to the traffic, for example, with a randomized MAC
  - transmission attempts and successes on a link depend on the traffic on the other links, how they respond to their traffic and their perception of the channel, and which PHY parameters they employ
  - capacities cannot be defined for all the links in the network simultaneously
- Data rates can only be constrained jointly through the network stability region
- It is the stability region that is parameterized by the network parameters

Formulations for Infrastructure-based Networks

- Users: N elastic
  - user r perceives utility of its rate x_r through function $U_r(.)$
- Network: M nodes, L wireless links
  - Q configurable parameters denoted by $\alpha = (\alpha_1, \ldots, \alpha_Q)$
  - under configuration α, $C(\alpha)$ denotes the stability region of the network
  - $h(\alpha)$ denotes the operating (or production) cost for the network
- Certain technical assumptions
  - $U_r(.)$ is power utility of second kind with $U_r(0) = 0$, $x_r \in [m_r, M_r]$
  - $\mathcal{A}$, the parameter space, is compact subset of R_+^Q
  - $C(\alpha)$ is a compact convex subset of R_+^N
  - C as correspondence is upper hemi-continuous

Welfare Maximization

- Network may be deployed as community (mesh) network
- Joint optimization of network and allocation of rates to users is captured by welfare maximization
- Social Welfare Maximization (SWM):

$$\begin{aligned} \max \quad & \textstyle\sum_{r=1}^{N} U_r(x_r) - h(\alpha) \\ & \mathbf{x} \in C(\alpha) \cap (\times_r [m_r, M_r]) \\ & \alpha \in \mathcal{A} \end{aligned}$$

- SWM has an optimal solution

Network Monopoly

- Network may be deployed by a service provider to provide backbone service
- Situation is best modeled as monopoly
- Monopoly Problem (MP): Given the rate price vector $\mathbf{p} = (p_1, \ldots, p_N)$, user r maximizes surplus

$$\max_{x_r \in [m_r, M_r]} U_r(x_r) - p_r x_r$$

Network sets α and $\mathbf{p}$ to maximize its revenue keeping the network stable

$$\begin{aligned} \max \quad & \textstyle\sum_{i=1}^{N} p_r x_r - h(\alpha) \\ & \mathbf{x} \in C(\alpha) \cap \mathbf{x}^*(\mathbf{p}) \\ & \alpha \in \mathcal{A} \\ & \mathbf{p} \in R_+^N \end{aligned}$$

$x_r^*(\mathbf{p}) = \arg\max_{x \in [m_r, M_r]} U_r(x) - p_r x$ and $\mathbf{x}^* = \Pi_r x_r^*$

- MP has an optimal solution

Scheduled Networks

- Motivated by the IEEE 802.16 standard for broadband wireless networks
  - fractions of a transmission frame can be allocated to network devices
  - modulation schemes and transmission power can be chosen adaptively
  - data can be fragmented and packed to a fine scale; justifies treatment of data as fluid
- Transmission power, modulation and transmission time allocations will constitute the network parameters
- Welfare maximization and monopoly will have special structure owing to parameterized link constraints
  - we also consider an interesting variation of monopoly called the bilevel optimization

Scheduled Networks: Parameters

- BER requirement on each link translates into an SINR threshold β
- Powers $\mathbf{P} \in \mathcal{P}$ determine the set of feasible modulation schemes $\mathcal{Z}(\mathbf{P})$
- $\mathbf{P}$ and $\mathbf{z} \in \mathcal{Z}(\mathbf{P})$ determine the conflicting set $D_l(\mathbf{z}, \mathbf{P})$ for link l
  - each $D \in \mathcal{D}_l(\mathbf{z}, \mathbf{P})$ is a set of links
  - at least one link from each D must be silent for success on link l
  - generalization of the popular contention graph model
- $D_l(\mathbf{z}, \mathbf{P})$s dictate the set of non-conflicting transmission schedules $\Theta(\mathbf{z}, \mathbf{P})$
  - schedules are constructed over $[0, 1]$; $\theta = (t_1^s, t_1^f, \ldots, t_L^s, t_L^f)$

$$\Theta(\mathbf{z}, \mathbf{P}) = \{\theta | 0 \leq t_l^s \leq t_l^f \leq 1, (t_l^s, t_l^f) \cap \cap_{l' \in D} (t_{l'}^s, t_{l'}^f) = \Phi, D \in \mathcal{D}_l(\mathbf{z}, \mathbf{P}), l \in \mathcal{L}\}$$

- For $\mathbf{P}$, $\mathbf{z} \in \mathcal{Z}(\mathbf{P})$ and $\theta \in \Theta(\mathbf{z}, \mathbf{P})$ capacity of link l is $c_l(z_l)(t_l^f - t_l^s)$

Non-conflicting schedules

- To simplify discussion assume for each link l, each $D \in \mathcal{D}_l(\mathbf{z}, \mathbf{P})$ is singleton; constraints can be represented by a contention graph
- Conflicting sets specify multiple contention graphs: for each link l, select one interferer from each $D \in \mathcal{D}_l(\mathbf{z}, \mathbf{P})$
- Non-conflicting transmission times $(t_l^f - t_l^s) = \delta_l$ can be assigned to links l if
  - contention graph is perfect and its maximum weighted clique has weight at most 1; weight of vertex l in the contention graph equals δ_l
  - contention graph is not perfect and a multiple bin packing problem has a feasible solution; one bin for each vertex l in the maximum weighted clique having capacity δ_l

Scheduled Networks: Welfare Maximization

- (SWM) translates into the following

$$
\begin{aligned}
\max \quad & \textstyle\sum_{r=1}^{N} U_r(x_r) - h(\mathbf{z}, \mathbf{P}, \theta) \\
& \textstyle\sum_{r \in \mathcal{N}_l} x_r \le c_l(z_l, \mathbf{P}, \theta) \qquad l = 1, \ldots, L \\
& x_r \in [m_r, M_r] \qquad r = 1, \ldots, N \\
& \theta \in \Theta(\mathbf{z}, \mathbf{P}) \\
& \mathbf{z} \in \mathcal{Z}(\mathbf{P}), \mathbf{P} \in \mathcal{P}
\end{aligned}
$$

- It is NP-hard to check if (SWM) has a feasible solution
  - (SWM) will have feasible solution if for some $\mathbf{z}$ and $\mathbf{P}$, there exists a non-conflicting schedule assigning capacity $\sum_{r \in \mathcal{N}_l} m_r$ to link l
  - by earlier result this requires finding the maximum weighted clique in the contention graph; if perfect, clique weight $\sum_l \frac{\sum_{r \in \mathcal{N}_l} m_r}{c_l(z_l)}$ must be at most 1

Welfare Maximization: Fixed $\mathbf{z}$ and $\mathbf{P}$

- Can (SWM) be distributed?
- Constraint set may not be convex; $Z_l(.)$ finite, Θ not convex-valued
- Modulation scheme and transmission power are fixed

$$
\begin{array}{lll}
\max & \sum_{r=1}^{N} U_r(x_r) - h(\theta) & \\
& \sum_{r\in\mathcal{N}_l} x_r \leq c_l(t_l^f - t_l^s) & l = 1,\ldots,L \\
& x_r \in [0, M_r] & r = 1,\ldots,N \\
& \theta \in \tilde{\Theta}(\mathbf{z},\mathbf{P}) &
\end{array}
$$

- $\tilde{\Theta}$ is set of schedules preserving order of start-times of conflicting links
  - $\tilde{\Theta}$ is a convex subset of Θ; proper in general
- SWM can now be decomposed in network and user problems

Welfare Maximization: Fixed $\mathbf{z}$ and $\mathbf{P}$

- There exist prices $p_l^* \geq 0,\ l = 1, \ldots, L$ such that

  (i) Each user $r = 1, \ldots, N$ maximizes its surplus, i.e.,

$$x_r^* = \arg \max_{x_r \in [0, M_r]} U_r(x_r) - \left(\sum_{l \in \mathcal{L}_r} p_l^* \right) x_r$$

  (ii) The network maximizes its profit, i.e.,

$$\theta^* = \arg \max_{\theta \in \tilde{\Theta}(\mathbf{z}, \mathbf{P})} \sum_{l=1}^{L} p_l^* c_l(t_l^f - t_l^s) - h(\theta)$$

  (iii) Supply equals demand at positive price, i.e.,

$$p_l^* \left(c_l(t_l^{f*} - t_l^{s*}) - \sum_{r \in \mathcal{N}_l} x_r^* \right) = 0$$

Welfare Maximization: Link Scheduling Problem

- Assume that $h(.) = 0$ for all $\theta \in \Theta(\mathbf{z}, \mathbf{P})$.
- The optimal solution is that if a link transmits it transmits for the full duration
- Thus the problem is combinatorial
- Let $u_l = 1$ if link l transmits and 0 otherwise

$$
\begin{aligned}
&\max \textstyle\sum_{l=1}^{L} p_l^* u_l \\
&u_l \Pi_{l' \in D} u_{l'} = 0 \qquad D \in \mathcal{D}_l(\mathbf{z}, \mathbf{P}),\ l = 1, .., L \\
&\quad u_l \in \{0, 1\} \qquad l = 1, .., L
\end{aligned}
$$

- Generalization of the weighted maximal independent set problem
- Hence NP-hard

Welfare Maximization: Exact Scheduling Algorithm

- Let $\mathbf{u}^*$ denote the optimal solution the above problem

- The idea is to consider for some $0 < \epsilon < 1$,

$$
\begin{aligned}
&\max \textstyle\sum_{l=1}^{L} p_l^* u_l \\
&u_l \Pi_{l' \in D} u_{l'} \leq \epsilon \qquad D \in \mathcal{D}_l(\mathbf{z}, \mathbf{P}),\ l = 1, .., L \\
&\quad 0 \leq u_l \leq 1 \qquad l = 1, .., L
\end{aligned}
$$

- If $\tilde{\mathbf{u}}$ denotes the optimal solution then

  (i) MFCQ holds at $\tilde{\mathbf{u}}$

  (ii) $\tilde{u}_l \in \{\epsilon, 1\}$ for each l

  (iii) $\mathbf{u}^* = 0 \Leftrightarrow \tilde{\mathbf{u}} = \epsilon$, $\mathbf{u}^* = 1 \Leftrightarrow \tilde{\mathbf{u}} = 1$

- The above problem is a geometric program

Network Monopoly

- Owing to the special structure of rate constraints, price setting behavior of the network can be modeled in two ways
- Profit monopoly: network sets prices with sole aim of maximizing its profit
- Bilevel optimization: network sets prices with dual aim of maximizing its profit while maximizing the aggregate user utility
  - thus the network chooses the profit maximizing Lagrange multiplier of the aggregate utility maximization problem

Scheduled Networks Monopoly: Fixed $\mathbf{z}$ and $\mathbf{P}$

- In a network setting it is befitting to consider per link pricing instead of per user
- Assume $x_r \geq 0$ and $U_r(x_r) = w_r \frac{(1+x_r)^{(1-\nu_r)}-1}{1-\nu_r}$ with $\nu_r > 1$
- Demand $x_r^*(p)$ is strictly convex and profit $px_r^*(p)$ strictly concave
- Denote by v_l, $l = 1, \ldots, L$ the link prices

$$
\begin{aligned}
\max \qquad & \textstyle\sum_{i=1}^{N} p_r x_r^*(p_r) - h(\theta) & \\
& \textstyle\sum_{r\in\mathcal{N}_l} x_r(p_r) \leq c_l(t_l^f - t_l^s) & l = 1, \ldots, L \\
& p_r = \textstyle\sum_{l\in\mathcal{L}_r} v_l & r = 1, \ldots, N \\
& \theta \in \tilde{\Theta}, \mathbf{v} \in R_+^L &
\end{aligned}
$$

- An optimal solution exists for this concave program
- Even if not known functionally, the network can compute demand function at a specific price by measuring user data rate

Scheduled Networks: Bilevel Optimization

- Users choose rates by maximizing aggregate utility given the network parameters

$$
\begin{array}{cc}
\max \sum_{r=1}^{N} U_r(x_r) & \\
\sum_{r \in \mathcal{N}_l} x_r \leq c_l(z_l, \mathbf{P}, \theta) & l = 1, \ldots, L \\
x_r \in [0, M_r] & r = 1, \ldots, N
\end{array}
$$

- Network configures parameters so as to maximize the revenue generated from the prices the users pay

$$
\begin{array}{c}
\max \sum_{l=1}^{L} \lambda_l(\mathbf{z}, \mathbf{P}, \theta) c_l(z_l, \mathbf{P}, \theta) - h(\mathbf{z}, \mathbf{P}, \theta) \\
\theta \in \Theta(\mathbf{z}, \mathbf{P}) \\
\mathbf{z} \in \mathcal{Z}(\mathbf{P}), \mathbf{P} \in \mathcal{P}
\end{array}
$$

Scheduled Networks: Bilevel Optimization

- Lagrange multipliers for utility maximization exist but need not be unique
  - uniqueness is strong condition requiring LICQ or Strict-MFCQ
  - the set of Lagrange multipliers is closed and convex
- User response is a multifunction; we consider the cooperative model
- Onus is on the network to set-up prices so that it maximizes revenue while allowing users to maximize their aggregate utility

$$\max \sum_{l=1}^{L} \lambda_l c_l(z_l, \mathbf{P}, \theta) - h(\mathbf{z}, \mathbf{P}, \theta)$$
$$\lambda \in \bar{\lambda}(\mathbf{z}, \mathbf{P}, \theta)$$
$$\theta \in \Theta(\mathbf{z}, \mathbf{P})$$
$$\mathbf{z} \in \mathcal{Z}(\mathbf{P}), \mathbf{P} \in \mathcal{P}$$

- Optimal solution exists under certain conditions
- Maximizes welfare if $\mathbf{z}$ and $\mathbf{P}$ are fixed

Unscheduled Networks

- Harder to handle since the network can influence rate constraints only through the stability region

  Conjecture 1 $\{\mathbf{x} | \sum_{r \in \mathcal{N}_l} x_r < c_l(\alpha),\ l = 1, \ldots, L\} \subset C(\alpha)$

- $c_l(\alpha)$ denotes the saturation throughput of link l
  - maximum rate at which transmissions can be made successfully on the link given that all the other nodes in the network always have data to transmit
- Aggregate rate constraint $C(\alpha)$ may be split into (sub-optimal) link constraints
- Difficulties
  - $c_l(.)$s need not be concave functions of α
  - functional form of $c_l(.)$ would not be known in general

A Perspective on Formulations

- Problems are well-posed
- Social Welfare Maximization
  - yields a Pareto optimal solution
  - is hard to decompose into network and user problems
  - may not be able to optimize over all the available parameters
- Monopoly and Bilevel optimization
  - are naturally separated into user and network problems
  - rate control is unaltered; based on price feedback
  - network problem has a higher complexity (IEEE802.16 Mesh provides for a centralized decision making by a base-station)
  - may have an actual economic connotation in deploying hot-spots

Summary

- Problem of allocation of rates to users is inherently tied to the optimization of the wireless network
- Formulation of rate constraints is determined by the orchestration of transmissions on links in the network
- In so-called scheduled networks, links can be assigned capacities parameterized by transmission power, modulation schemes and time allocations
- Welfare maximization and monopoly capture the interactions through which the rate allocations and network optimization can occur
- In this work we have not focused on distributed algorithms, but we sought formulations that can be decomposed
- Derivation of explicit distributed algorithms is on-going work