Cross-Layer Design of Multi-hop Wireless Networks: A Loose Coupling Perspective

Ness B. Shroff
Center for Wireless Systems and Applications (CWSA)
School of Electrical and Computer Engineering
Purdue University, West Lafayette
E-mail: shroff@ecn.purdue.edu
URL: www.ece.purdue.edu/~shroff

PURDUE UNIVERSITY

Joint work with Xiaojun Lin

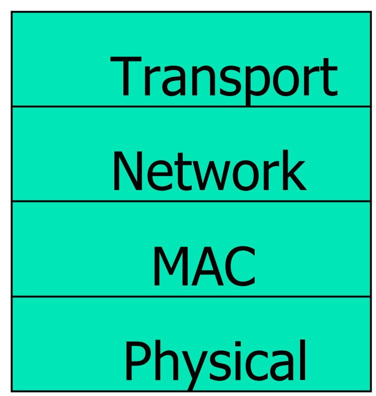
CWSA

Outline

- Cross-Layer Design and "Loose Coupling"
- Focus on congestion control and scheduling problem
  - Model and formal formulation
  - Optimal solution
- Difficulties with optimal solution
- Impact of imperfect scheduling
  - Static system
  - Dynamic system
- Ongoing work and open problems

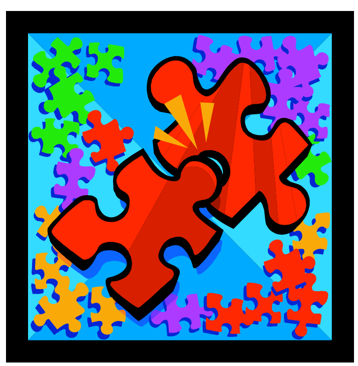
Cross-Layer Design

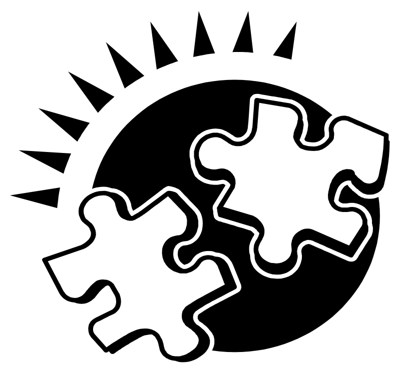
- Layered architecture offers simplicity and modularity
- Optimizing within layers has reached the point of diminishing returns.
- Future applications that will fuel the growth of wireless require orders of magnitude increase in performance.
- **Thesis:** To satisfy the increasing demand for new wireless services, a *cross-layer perspective* needs to be taken to obtain significant improvements in wireless spectrum efficiency



The Cross-layer Dilemma: Efficiency vs. Modularity

- Cross-Layer design needed to improve *efficiency*
- Layers are coupled
  - Potential loss of *modularity*
  - Could lead to complex and fragile overall design





Cross-Layer with "Loose Coupling"

- **Loose coupling idea:**
  - Minimal interaction between layers
  - Imperfect measurements or decision at one layer should not affect the entire system
  - Overall cross-layer solution must ensure both **efficiency** and **modularity**
- Appropriately designed cross-layer solutions do exhibit a layered structure with minimal but crucial interaction between the layers.

The Cross-Layer Congestion-Control and Scheduling Problem

- **Congestion control**: Determines end-to-end rate at which users should transmit
  - Maximize capacity and avoid excessive congestion
  - Improve fairness of the service to different users
- **Scheduling**: Everything in MAC and Physical layer, e.g., power control, link scheduling, adaptive modulation and coding
- **Goal:** To determine the maximum end-to-end rate at which users should transmit and at the same time find the associated "scheduling policy" that stabilizes the system --- a cross-layer problem

Transport	Congestion control (e.g. TCP)
Network	
MAC	Scheduling
Physical	

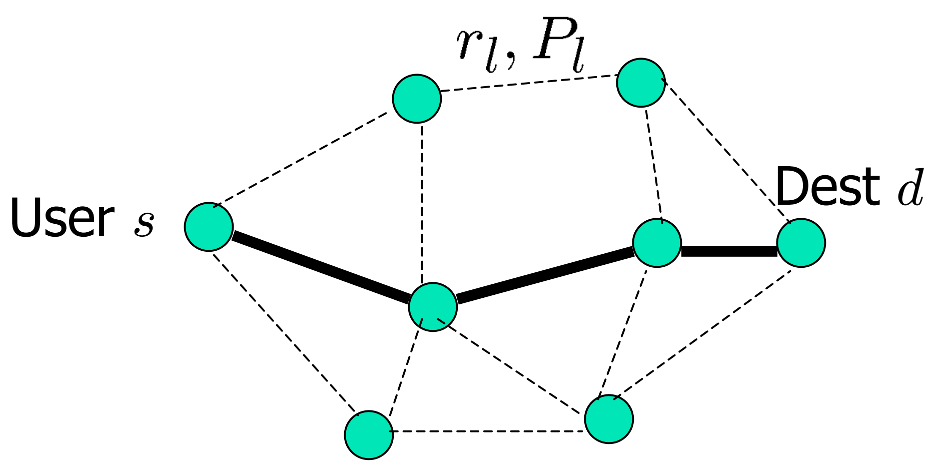
- For simplicity, we assume that ***routing is fixed***. Results can be readily extended to incorporate multi-path routing.

Related Work

- Congestion control in wireline networks [Kelly, Kunniyur and Srikant, Mazumdar et al., Lapsy and Low, Towsley, Qiu and Shroff, and many others]
- Simple rate-power functions
  - Rate of a link is a function of its own power assignment [Xiao *et. al.* 2002]
  - Mapped to convex problems (the high-SINR case) [Johansson *et. al.* 2003, Chiang 2004]
  - The node-exclusive interference model [Sarkar & Tassiulas 2003, Yi & Shakkottai 2004, Paschalidis *et. al.* 2005]
  - The clique-based interference model [Xue *et. al.* 2003, Chen *et. al.*, 2005]
- Offline cross-layer solution
  - Column-Generation Approach [Johansson & Xiao 2004]
- On-Line Centralized cross-layer solutions [Lin & Shroff 2004, Neely *et al.* 2005, Eryilmaz and Srikant, 2005, Paschalidis *et al.* 2005, Chiang *et al.* 06]
  - Scheduling is still the bottleneck!
- Cross-Layer solutions with imperfect scheduling and distributed solution [Lin & Shroff 2005]
- More recent work on distributed scheduling [Wu and Srikant, Charpokar et al.] , joint congestion control and scheduling [Eryilmaz and Srikant, Bui et al.] , complexity of scheduling [Sharma, Mazumdar and Shroff], random access solutions [Lin and Rasool], [Joo and Shroff]

The Network Model

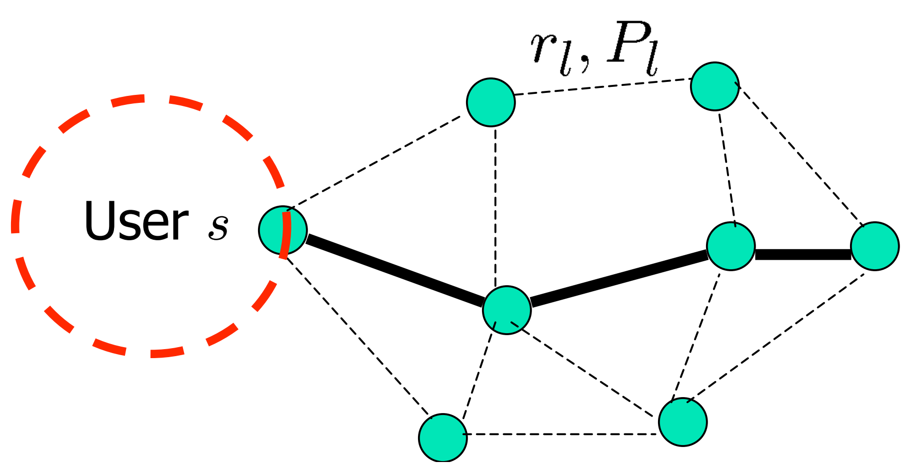
- **A multihop wireless network serving multiple users**
- N nodes and L Links
  - A link corresponds to a transmitter-receiver pair
- S users:
  - Each user transmits from a source node to a destination node
  - The path of each user s could traverse multiple wireless links
  - H: routing matrix



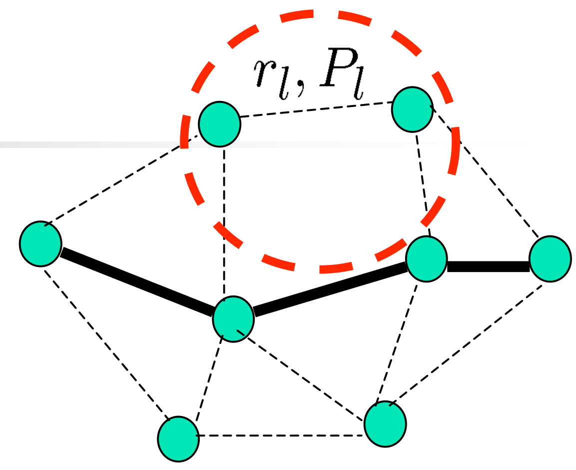
$$H_s^l = \begin{cases} 1 & \text{if path of user } s \text{ traverses link } l, \\ 0 & \text{otherwise} \end{cases}$$

The User Model

- $U_s(x_s)$: utility of user s if its end-to-end rate is x_s (measures the level of satisfaction of the user).
  - $U_s(\cdot)$: strictly concave, non-decreasing
    - "Principle of diminishing return"
    - Fairness
  - M_s: the maximum data rate

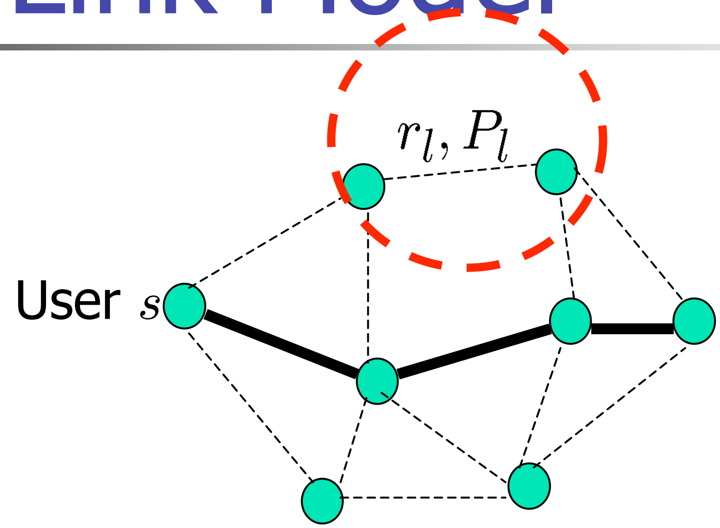


The Link Model



- P_l: power assignment on link l
  r_l: data rate on link l
- Shared nature of the wireless medium
  - The data rate on link l depends on the interference due to power assignments on other links.
  - Assume (for now) no channel variations due to fading, etc.
  - Hence, the link capacity $\vec{r} = [r_1, ..., r_L]$ is a function of the global power assignment $\vec{P} = [P_1, ..., P_L]$

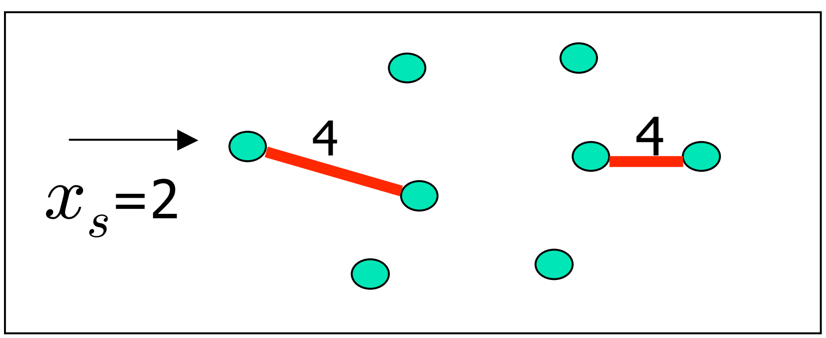
The Link Model

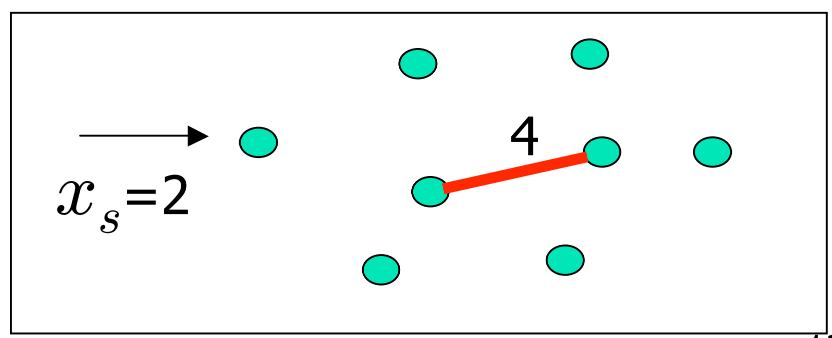


- Each link uses the appropriate modulation and coding scheme to achieve data rate $\vec{r} = g(\vec{P})$
- $\vec{r} = g(\vec{P}), \vec{P} \in \Pi$: the rate-power function
- $\vec{P} \in \Pi$: feasible power assignments

Link Scheduling

- Interleaving different schedules over time will typically increase capacity
- $\vec{P}(t)$ or $\vec{r}(t)$: the *schedule* at time t
- *Scheduling policy*
  - pick $\vec{P}(t)$ or $\vec{r}(t)$ at each time





The Capacity Region Λ

- The set of end-to-end rates that the network can support

- The *capacity region* Λ is given by [Neely 03, Cruz & Santhanam 03]

Rate-power function

$$\Lambda = \left\{ \vec{x} \,\middle|\, [\sum_{s=1}^{S} H_s^l x_s] \in \mathbf{Convex\_Hull}(g(\Pi)) \right\}$$

The sum rate at each link

The set of feasible power assignment

The Cross-Layer Congestion Control and Scheduling Problem

A Cross-Layer Problem:

- Find the user rate vector $\vec{x} \in \Lambda$ that maximizes the total system utility, i.e.,

$$\max_{0 \leq x_s \leq M_s, s=1,\ldots,S} \quad \sum_{s=1}^{S} U_s(x_s)$$
$$\text{subject to} \quad [x_s] \in \Lambda$$

  - An *end-to-end* problem
- Find the associated scheduling policy that stabilizes the system (i.e., keeps all queues finite)
  - A *link-by-link* problem

The Optimal Cross-Layer Solution

- $q^l(t)$: the queue length of link l at time t *(price)*
- Congestion control component *(max. net utility)*

$$x_s(t) = \underset{0 \le x_s \le M_s}{\operatorname{argmax}} \left[U_s(x_s) - x_s \sum_{l=1}^{L} H_s^l q^l(t) \right]$$

- Scheduling component *(max. value of data)*

$$\vec{r}(t) = \underset{\vec{r}=g(\vec{P}), \vec{P} \in \Pi}{\operatorname{argmax}} \sum_{l=1}^{L} q^l(t) r_l.$$

- Two components are coupled by the queue length *(difference between demand & supply)*

$$q^l(t+1) = \left[q^l(t) + \alpha_l \left(\sum_{s=1}^{S} H_s^l x_s(t) - r_l(t) \right) \right]^+.$$

The Optimal Cross-Layer Solution

Theorem: For any $\epsilon > 0$, there exists a set of stepsizes α_l such that for any initial queue lengths there exists a time T_0 such that for all $t \geq T_0$,

$$||\vec{x}(t) - \vec{x}^*|| < \epsilon.$$

Further, the queue length is bounded over all time t.

- Above Theorem shows that our cross layer solution converges to the optimal rate allocation provided the chosen stepsizes are sufficiently small
- **Proof techniques**: optimization of *non-differentiable* functions, convex analysis

Extension: Dealing with Channel Variations

- **K: channel state, with stationary distribution π_K**
  - The rate-power function: $\vec{r} = u(\vec{P}, K), \vec{P} \in \Pi$
  - The capacity region:

$$\Lambda = \left\{ \vec{x} \,\middle|\, [\sum_{s=1}^{S} H_s^l x_s] \in \sum_{K} \pi_K \mathbf{Convex\_Hull}(u(\Pi, K)) \right\}$$

- **Only the scheduling component needs to change slightly:**

$$\vec{r}(t) = \operatorname*{argmax}_{\vec{r}=g(\vec{P},K(t)),\vec{P}\in\Pi} \sum_{l=1}^{L} q^l(t) r_l.$$

- **Does not require prior knowledge of the stationary distribution of the channel**

Comments on the Optimal Cross-Layer Solution

- Achieves the full capacity region Λ
- Exhibits an aspect of *loose-coupling* property

Hard to solve even in a centralized fashion

$$\vec{r}(t) = \underset{\vec{r} \in g(\Pi)}{\operatorname{argmax}} \sum_{l=1}^{L} q^{l}(t) r_{l}$$

Not convex

Congestion Control

Queue Length

Scheduling (MAC/Physical)

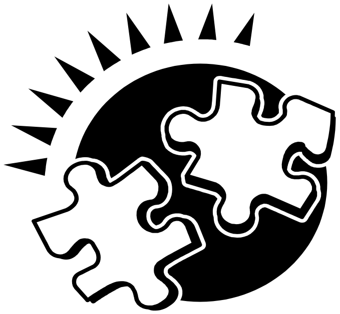
Distributed

Centralized

To obtain **simple** and potentially **fully distributed** solutions → Imperfect scheduling

Loose-Coupling Revisited

- Problem: Will our cross-layer solution break down if the scheduling component is **imperfect**?
  - Will it get stuck into local sub-optimal solutions?
  - Will it lead to excessive inefficiency?



Imperfect Scheduling Policies S_γ

- S_γ policies:

$$\boxed{\sum_{l=1}^{L} r_l(t) q^l(t)} \geq \gamma \boxed{\max_{\vec{r} \in g(\Pi)} \sum_{l=1}^{L} r_l q^l(t),} \quad 0 < \gamma < 1$$

  - Compute a schedule *r(t) that achieves a queue-weighted rate sum of at least* γ times the optimal.

$\gamma = 0$ ⟵ easier — harder ⟶ $\gamma = 1$

The Impact of Imperfect Scheduling

- One naturally hopes: If we were to use an S_γ scheduling policy that the data rate allocation of each user would be around γ times the optimal rate allocation (γ reduced problem)
- Not true! The rates of some users can be significantly worse
  - Weak fairness property: Rates cannot be arbitrarily worse.
- Question: does such sub-optimality in static system matter when considering the more *realistic dynamic* case?

Connection Between the Static Setting and the Dynamic Setting

- Question: How much of Λ can we utilize?
- Previous results on stability for wireline networks [Bonald & Mossoulie 01, De Veciana *et al.* 01, Fayolle *et al.* 01, Ye 03]
  - Fairness $\Rightarrow$ largest stability region (largest set of offered loads & maintain finite queues) = Λ
  - Unfairness $\Rightarrow$ significantly reduced stability region.

Fairness	$\Longleftrightarrow$	Stability
Static Setting		Dynamic Setting

- Fairness is not just an aesthetic property but also carries a strong performance implication
- *Is weak fairness enough?*

Main Result

Best we can hope to achieve

[Theorem]

If
$$\max_l \alpha_l \le \frac{1}{T\bar{S}\bar{L}} \frac{2^\beta - 1}{16} \min_s \frac{w_s}{\rho_s M_s^\beta},$$

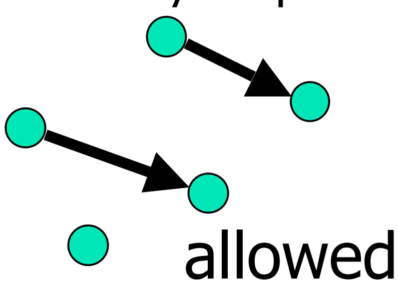
then **the stability region of the system is no smaller than γΛ**, where

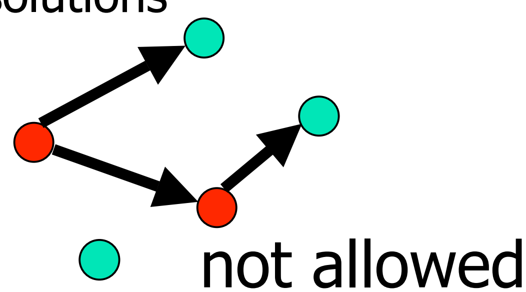
- $\bar{\mathcal{S}} = \max_l \sum_{s=1}^{S} H_s^l$ denotes the maximum number of classes going through any link, and
- $\bar{\mathcal{L}} = \max_s \sum_{l=1}^{L} H_s^l$ denotes the maximum number of links used by any class.

Local sub-optimality in the static setting does not matter ⇒ **Loose-coupling**

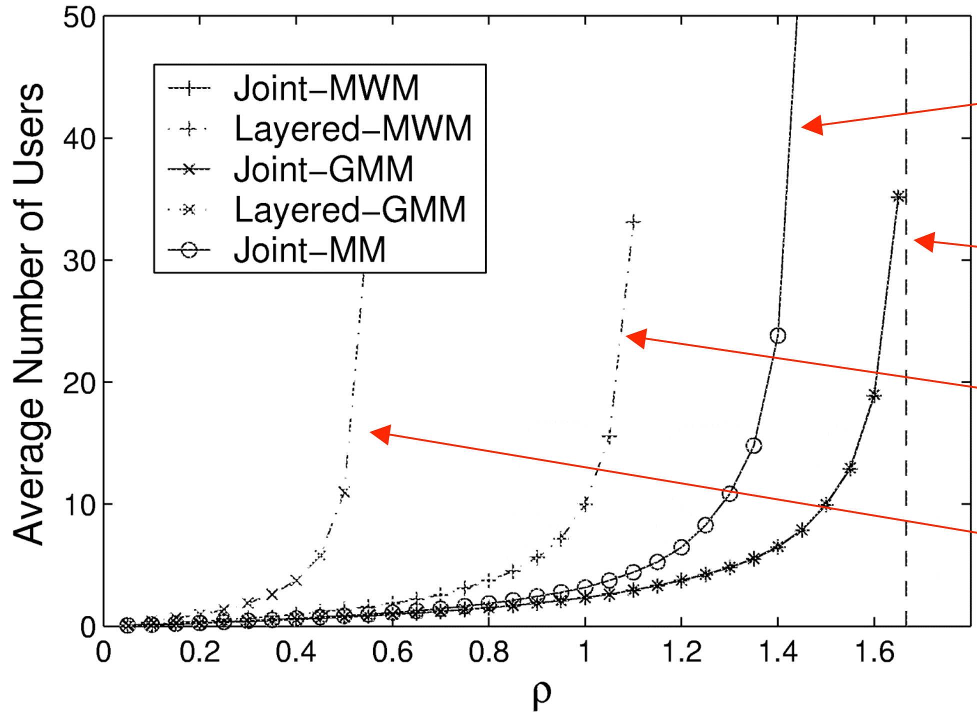
Example Scenario: Node Exclusive Model

- Thus far: Results applicable to general interference models
- Focus on **Node Exclusive Model** [Sarkar & Tassiulas 2003, Yi & Shakkottai 2004]
  - Each node can communicate with one other node at any given time
  - The data rate of each *active* link is fixed at some c_l
  - Applicable for Blue-tooth networks and approximates FH-CDMA
  - Provides insights on distributed algorithms for other models.
- First to develop a fully distributed algorithm **Maximal Matching** that
  - Provably achieves a stability region of at least $\Lambda/2$
  - Empirically, achieves much better performance
  - Significantly outperforms layered solutions





Performance Comparison



Cross-Layer solution With MM (distributed)

Cross-Layer solution With MWM (optimal) or GMM (Greedy $S_{1/2}$)

Layered solution With MWM (optimal)

Layered solution With GMM

GMM: Greedy Maximal Matching MM: Maximal Matching (Distributed)

MWM: Maximum-Weighted Matching (Optimal)

- **The $\Lambda/2$ guarantee is in fact quite conservative**
- **Cross-Layer (Imperfect)>> Layered (Perfect)**

Recent Related Works

- [Wu and Srikant, INFOCOM 2006]
  - 2-hop interference model
  - Prove that "greedy scheduling" (maximal matching) achieves a throughput within a factor of N_ϵ of the optimal, where

$$N_\varepsilon = \max_{(i,j)\in E} d(i) + d(j) - 1$$

- [Chaporkar Sarkar, and Kar, Allerton 2005]
  - Bi-directional equal power model and a general interference model
  - Prove that "maximal scheduling" achieves a throughput within a factor of K_N of the optimal, where K_N is the maximum number of non-conflicting links that can interfere with any given link in the network
- [Bui, Eryilmaz, and Srikant INFOCOM 2006]
  - Asynchronous congestion control and scheduling under node-exclusive interference model
  - Algorithm that supports at least 1/3 of the maximum achievable throughput

Recent Related Works

- [Sharma, Mazumdar, and Shroff, FAWN 2006]
  - Studied a family of K-hop interference model (links within K hops cannot simultaneously transmit)
  - Hardness and approximability of scheduling: K>1, problem is NP-hard and not approximable within a large factor.
  - PTAS solutions for disk (geometric) graphs
    - PTAS guarantees performance within $1+\varepsilon$ factor for any ε greater than zero.

PURDUE
UNIVERSITY

Ongoing/Future Work

- Developing *distributed solutions* for more general *interference models* with *provable performance bounds*
  - Use of maximal scheduling results in low γ.
  - Need to improve performance by sharing local queue length information.
- Developing cross-layer solutions for
  - Random access MAC [Lin and Rasool, Joo and Shroff]
  - Multi-carrier OFDM types of systems
  - Minimal feedback (e.g., binary feedback as in TCP)
- Experimentation on Purdue Mesh Network (with Profs. Hu and Lin: Mesh@Purdue)

PURDUE UNIVERSITY

CWSA

Open Problems

- Tightness of throughput-loss bounds
  - Bounds on loss of throughput are based on worse-case analysis
  - Simulations suggest that average performance could be quite good
  - Open Problem: characterizing the average perceived performance?
- Incorporating the effects of delay in the feedback for general interference models
- Determining the performance limits of distributed algorithms.
  - Study the tradeoffs between performance and overhead
  - Development of constant/low overhead solutions
- Cross-Layer design with fairness under session-level dynamics
- Non-concave utility functions
  - Inelastic traffic
  - Non-convexity appears in both the rate-power function and the objective function.
- Impact of mobility on overall solution

Concluding Remarks

- **Potential:** Cross-layer gains are *multiplicative*

- **Key to Success:**
  - Cross-layer solutions should be *loosely coupled* across the layers such that *high performance* gains are achieved without a significant loss of *modularity*.

Thank you!

URL: http://www.ece.purdue.edu/~shroff

- Xiaojun Lin, Ness B. Shroff, R. Srikant, " A Tutorial on Cross-Layer Optimization in Wireless Networks, " to appear in the IEEE Journal on Selected Areas in Communications on ``Non-Linear Optimization of Communication Systems'', 2006.

- Gaurav Sharma, Ravi R. Mazumdar, and Ness B. Shroff, "Maximum Weighted Matching with Interference Constraints," in *IEEE FAWN* 2006.

- Gaurav Sharma, Ravi R. Mazumdar, and Ness B. Shroff, "On the Complexity of Scheduling in Wireless Networks

- Xiaojun Lin and Ness B. Shroff, "The Impact of Imperfect Scheduling on Cross-Layer Rate Control in Multihop Wireless Networks," in *IEEE INFOCOM,* Miami, Florida, 2005.

- Xiaojun Lin and Ness B. Shroff, "The Impact of Imperfect Scheduling on Cross-Layer Rate Control in Multihop Wireless Networks," to appear in *IEEE/ACM Tran. On Networking, 2006.*

- Xiaojun Lin and Ness B. Shroff, "Joint Rate Control and Scheduling in Multihop Wireless Networks," in *43rd IEEE Conference on Decision and Control,* Paradise Island, Bahamas, December 2004.

- Xiaojun Lin and Ness B. Shroff, "On the Stability Region of Congestion Control," in *42nd Annual Allerton Conference on Communication, Control and Computing,* Monticello, Illinois, September 2004.