Lecture 00
Course Overview

Introduction

- Resource allocation in communication networks and computing systems is a pressing research topic that has huge applications.
- The purpose of resource allocation is to intelligently assign the limited available resources among terminals/clients in an efficient way to satisfy end users' service requirements.
- It is imperative to develop advanced resource allocation techniques for ensuring the optimal performance of these systems and networks.
- In recent years, many tools including ***optimization theory***, ***control theory***, ***game theory***, and ***auction theory*** have been employed to model and solve a variety of practical resource allocation problems.

Part I
Resource allocation in networks via convex programming

- Many network resource allocation problems can be formulated as a constrained maximization of some utility function or minimization of some disutility function. In the first part of the course, we use convex programming as a toolbox to frame allocation problems.

- **Convex optimization** refers to the minimization of ***a convex objective function subject to convex constraints***. There are at least three benefits obtained from framing resource allocation problems as a convex program:

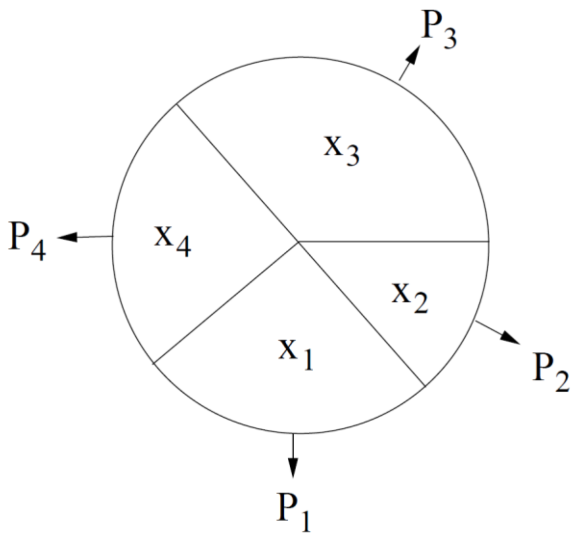
  - Convex optimization techniques are important because a **local optimum is also a global optimum in a convex problem** and a rigorous optimality condition and a duality theory exist to verify the optimal solution.

  - **Powerful numerical algorithms** exist to solve for the optimal solution of convex problems.

  - The third is on **decomposability structures**, which may lead to distributed (and often iterative) algorithms that converge to the global optimum. Distributed solutions are particularly attractive in large-scale networks where a centralized solution is infeasible, nonscalable, too costly, or too fragile.

A Simple Example

- Let us consider a simple example of resource allocation.
  - Suppose that a central authority has a divisible good of size *C* that is to be divided among N different users, as illustrated below.

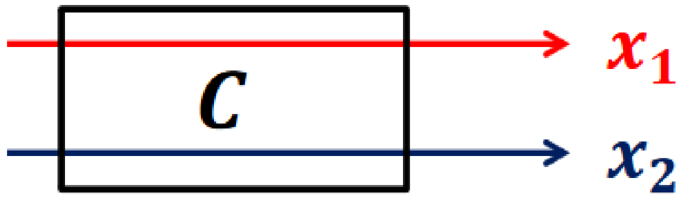


- For example, suppose that the government has determined that a fixed quantity of ground water may be pumped in a certain region and would like to allocate quotas that may be pumped by different farms.
- One way of performing the allocation is to simply divide the resource into *N* parts and allocate *C/N* to each player.
- But such a scheme does not take into account the fact that ***each player might value the good differently***.

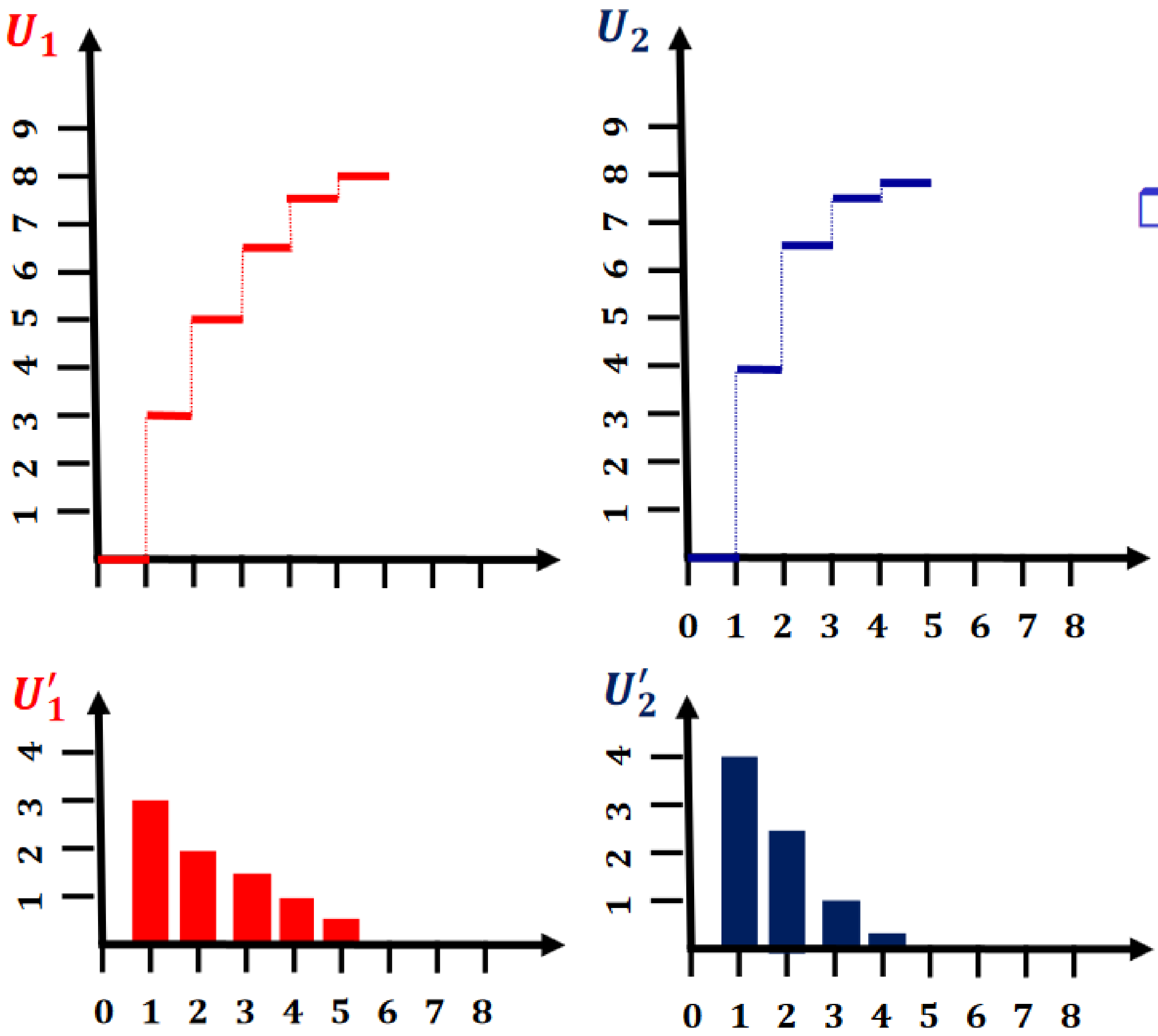
Utility

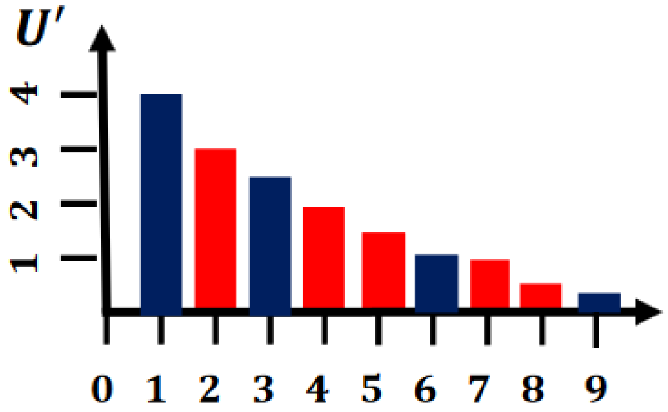
- **We refer to the value or "utility" obtained from an allocation x as U(x).**
  - This utility is measured in any denomination common to all the players such as dollars.
  - So a farm growing rice would have a higher utility for water than a farm growing wheat.
- **What would be the properties of a utility function?**
  - ➢We would expect that it would be ***increasing in the amount of resource*** obtained.
    - ❖More water would imply that a larger quantity of land could be irrigated leading to a larger harvest.
    - ❖We might also expect that a ***law of diminishing returns*** applies.
      - ▪ In our example of water resources, the return obtained by increasing the quota from 10 units to 20 units would make a large difference in the crop obtained, but an increase from 100 units to 110 units would not make such a significant difference.
      - ▪ Such a law of diminishing returns is modeled by specifying that the ***utility function is a strictly concave function*** since the second derivative of a strictly concave function is negative.
        - ✓Thus, the first derivative (which is the rate at which the function increases) decreases.

Law of Diminishing Returns



- Assign resource unit by unit





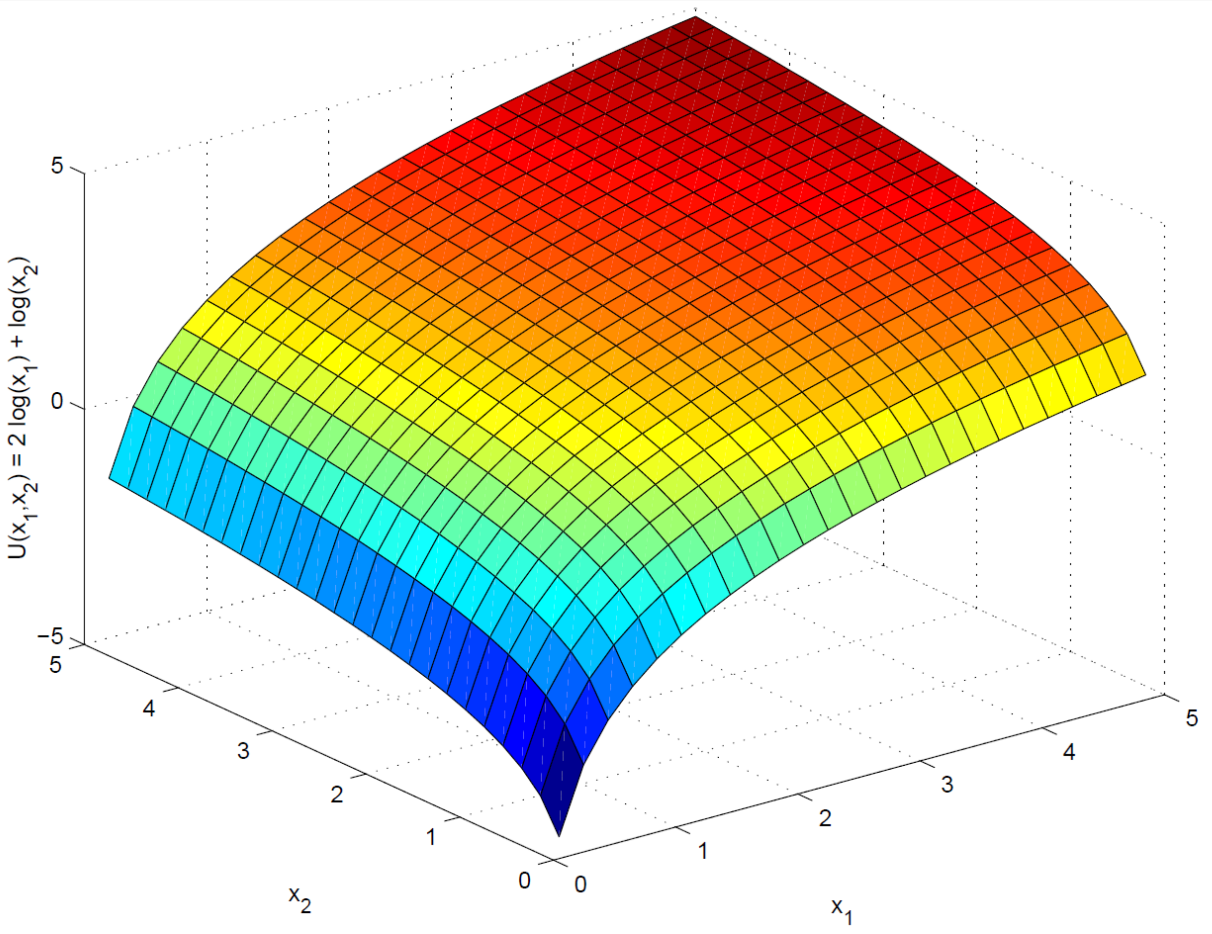
System-Wide Objective (Social Welfare)

- The objective of the authority would be to maximize the “system wide” utility.
  - One commonly used measure of system-wide utility is the sum of the utilities of all the players.
  - Since the utility of each player can be thought of the happiness that he/she obtains, the objective of the central authority can be likened to maximizing the total happiness in the system, subject to resource constraints.

System-Wide Objective (Social Welfare)

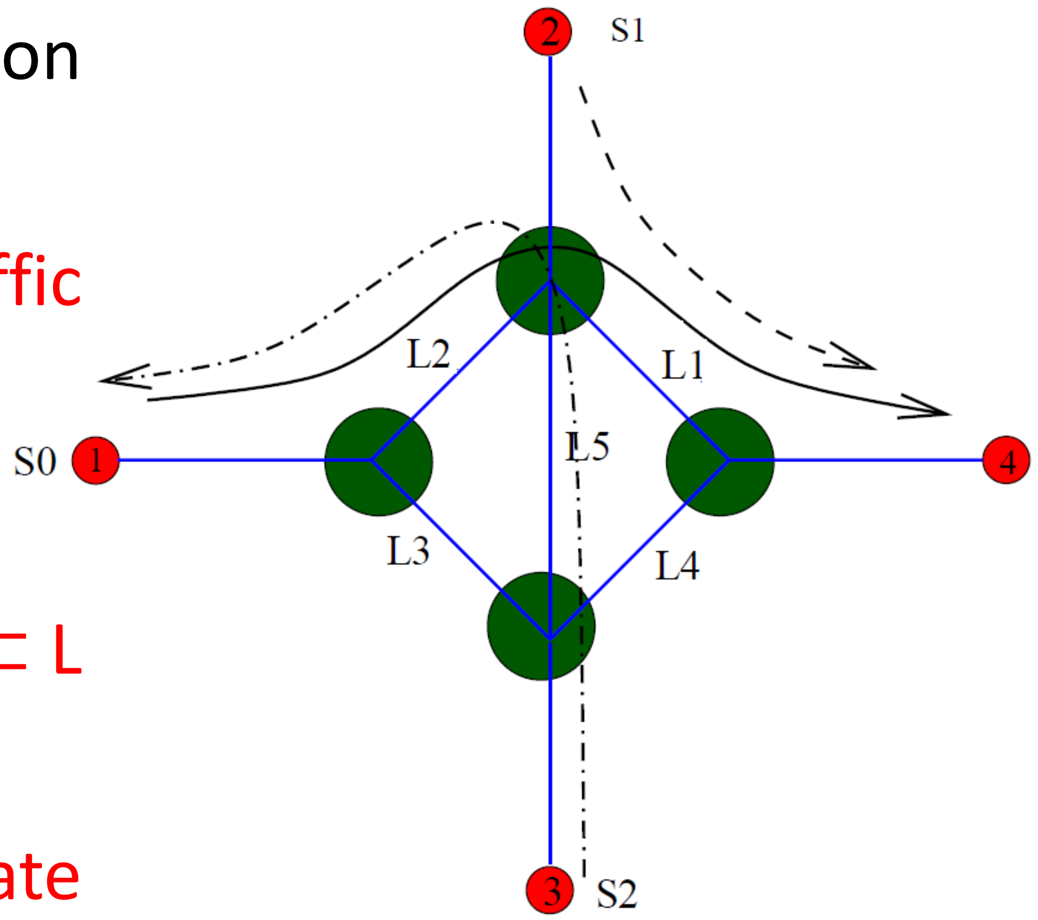
- We illustrate the sum utility with the utility functions of players 1 and 2, who receive an allocation x_1 and x_2 respectively being $2\log(x_1)$ and $\log(x_2)$.

- The marginals in the figure are strictly concave, which results in the sum being strictly concave as well.



Network Utility Maximization

- We now consider the analog of the resource allocation problem in a communication network such as the Internet.
- Suppose we have a network with a set of traffic sources S and a set of links L.
- Each link $l \in L$ has a finite fixed capacity c_l.
- Each source in S is associated with a route $r \subset L$ along which it transmits at some rate x_r.
- Note that we can use the index r to indicate both a route and the source that sends traffic along that route and we will follow this notation.



Utility Maximization in Networks

- The utility that the source obtains from transmitting data on route r at rate x_r is denoted by $U_r(x_r)$.

- We assume that the utility function is continuously differentiable, non-decreasing and strictly concave.

- As mentioned before, the concavity assumption follows from the diminishing returns idea:

  - A person downloading a file would feel the effect of a rate increase from 1 kbps to 100 kbps much more than an increase from 1 Mbps to 1.1 Mbps although the increase is the same in both cases.

Remark: A similar example holds for the problem of dividing a pie of fixed amount C among two individuals, one poor and one rich. The poor will valuate a given portion x of the resource differently than the rich one. That means the different users to which the pie is to be allocated have different utility functions. That is, $U_1(x) \neq U_2(x), \forall x$.

- It is straightforward to write down the problem as an optimization problem of the form:

$$\max_{x_r} \sum_{r \in \mathcal{S}} U_r(x_r)$$

subject to the constraints

$$\sum_{r: l \in r} x_r \leq c_l, \ \forall l \in \mathcal{L},$$

$$x_r \geq 0, \ \forall r \in \mathcal{S}.$$

- The above inequalities state that the capacity constraints of the links cannot be violated and that each source must be allocated a nonnegative rate of transmission.
- It is well known that **a strictly concave function has a unique maximum over a closed and bounded set**.
- In the above problem, the utility function is strictly concave, and the constraint set is closed (since we can have aggregate rate on a link equal to its capacity) and bounded (since the capacity of every link is finite).
- In addition, the constraint set for the utility maximization problem is convex which allows us to use the ***method of Lagrange multipliers*** and the **Karush-Kuhn-Tucker (KKT)** theorem.

Part II

- <u>**Resource Allocation via Economic Mechanism Design**</u>
  - **Centralized**
    - ➢ A central network planner solicits private information from non-cooperative or strategic networking agents, and they can potentially lie about it!
    - ➢ We use economic mechanism design <u>to incentivize the agents to tell the truth</u>!
      - **VCG mechanism**
      - **Kelly Mechanism**
  - **Distributed**
    - ➢ We assume that the users cooperate, we ignore incentive issues which may occur in networks with non-cooperative users.
    - ➢ Our objective is to coordinate the decisions of all users to optimize the overall performance, which is measured in terms of the total network utility.
    - ➢ Can be formalized through the theory of ***network externalities***
      - There are a range of approaches used to try to manage externalities through appropriate incentives, from ***pricing*** to ***regulation***.

Example: Centralized Resource Allocation for Non-Cooperative Networking Agents

- Consider a network planner who is interested in allocating resources to users with the **goal of maximizing the sum of the user's utilities**.
- The network planner can do this **only if he knows the utility functions of all the users** in the network, or if there is an incentive for the users to reveal their utility functions truthfully.
- We will discuss **incentive mechanisms** which make it profitable for the users to reveal their true utilities to the central network planner.

Example: Centralized Resource Allocation for Non-Cooperative Networking Agents

- Consider the problem of dividing a fixed resource among different users.
- Consider a network planner who wants to solve the utility maximization considered in previous slides, where each user is associated with a route:

$$\max_{x_r} \sum_{r \in \mathcal{S}} U_r(x_r)$$

subject to the constraints

$$\sum_{r: l \in r} x_r \leq c_l, \ \forall l \in \mathcal{L},$$

$$x_r \geq 0, \ \forall r \in \mathcal{S}.$$

Here, x_r is the rate allocated to user r, who has a utility function given by U_r and c_l is the capacity of link l.

Example: Centralized Resource Allocation for Non-Cooperative Networking Agents

Suppose that the network planner asks each user to reveal their utilities and user r reveals its utility function as $\tilde{U}_r(x_r)$, which may or may not be the same as $U_r(x_r)$.

Users may choose to lie about their utility function to get a higher rate than they would get by revealing their true utility function.

the network solves the maximization problem

$$\max_{x \geq 0} \sum_r \tilde{U}_r(x_r)$$

subject to

$$\sum_{r:l \in r} x_r \leq c_l, \forall l$$

and allocates the resulting optimal solution $\tilde{x}_r$ to user r.

What is a proper incentive mechanism?

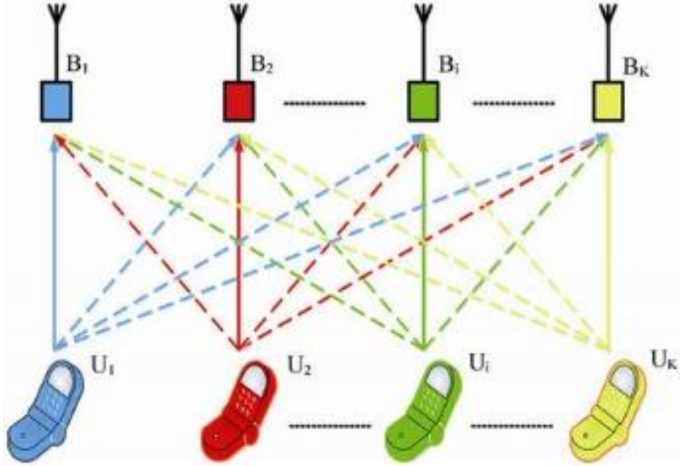
In return for allocating this rate to user r, the network charges a certain price p_r.

Distributed Resource Allocation Based on Network Externalities

- Informally, ***externalities*** are effects that one agent's actions have on the utility received by another.
- These effects can be
  - ***negative*** (e.g., resource congestion, or the interference one user in a wireless network causes to others) or
  - ***positive*** (e.g., the performance improvement a user in a peer-to-peer network experiences due to the participation of other users).
- From our perspective, externalities prove challenging because they often lead to distortions of economic outcomes away from efficiency.
- In response, there are a range of approaches used to try to manage externalities through appropriate incentives, from pricing to regulation.

Example: Interference externalities

Consider a wireless network where a set of R users share a common frequency band. Each user corresponds to a unique transmitter/receiver pair.



User r sets a power level P_r and transmits over the entire band treating all interference as noise.

The performance of each user r is given by a utility $u_r(\gamma_r)$ that is an increasing function of their signal-to-interference plus noise ratio (SINR) γ_r, which is given by

$$\gamma_r = \frac{h_{rr} P_r}{n_0 + \sum_{s \neq r} h_{sr} P_s},$$

where h_{sr} denotes the channel gain from transmitter s to receiver r, and n_0 is the noise power.

Example: Interference externalities

$$\gamma_r = \frac{h_{rr} P_r}{n_0 + \sum_{s \neq r} h_{sr} P_s},$$

Suppose that each user can select a power P_r from the interval $[0, P_{\max}]$ to maximize their own utility.

Since γ_r is increasing in P_r, it follows that each user would transmit at the maximum power $P_{\max}$, which can result in a total utility that is much smaller than the maximum achievable utility.

The reason for this outcome is that while increasing an individual's power increases their own SINR, it decreases the SINR of every other user, i.e., this is a negative externality.

Each player does not account for this externality in making her own decision;

a common expression for this result is that players are not "internalizing the externality."

How to internalize the externalities?

For distributed interference compensation, we will propose a protocol in which the users **exchange price signals** that indicate the negative externality of the received interference.

Part III

- **Resource Allocation in Stochastic Networks**

  - In this part, we consider the analysis and control of ***stochastic networks***, that is, networks with random events, time variation, and uncertainty.

  - Our focus is on communication and queueing systems.

  - Example applications include wireless mesh networks with opportunistic scheduling, cognitive radio networks, ad-hoc mobile networks, internets with peer-to-peer communication, and sensor networks with joint compression and transmission.

  - The techniques are also applicable to stochastic systems that arise in operations research, economics, transportation, and smart-grid energy distribution.

  - These problems can be formulated as problems that optimize the ***long-run performance of the system***; e.g., time averages of certain quantities subject to time average constraints on other quantities, and they can be solved with a common mathematical framework that is intimately connected to queueing theory.

  - We approach these problems using two types of mathematical toolboxes:
    - ***Markov devision processes (MDP)***
    - ***Lyapunov optimization (drift + penalty method)***

Example: opportunistic scheduling problem

Here we provide a simple wireless example to illustrate how the theory for optimizing time averages can be used.

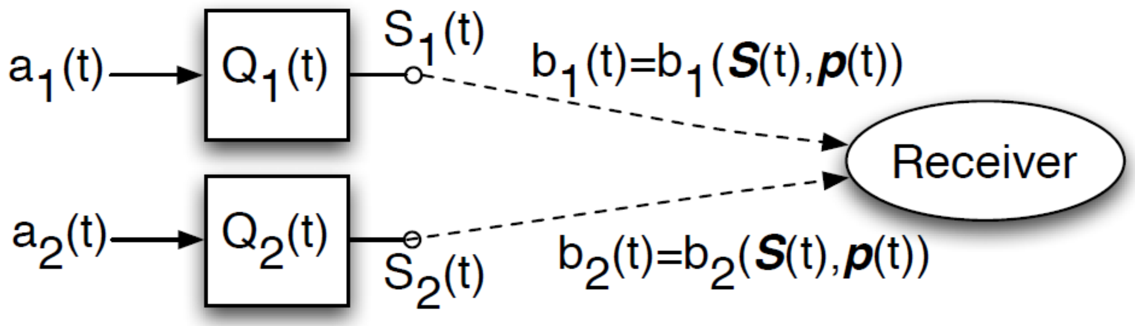
Consider a 2-user wireless uplink that operates in slotted time $t \in \{0, 1, 2, \ldots\}$.

Every slot new data randomly arrives to each user for transmission to a common receiver.

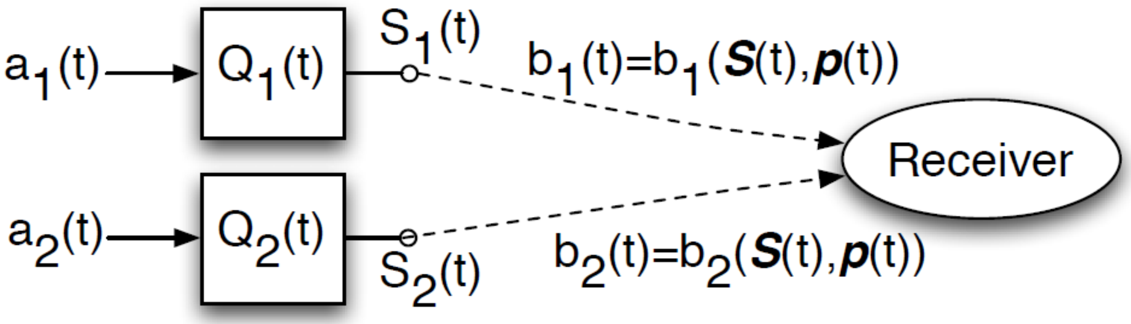
Let $(a_1(t), a_2(t))$ be the vector of new arrivals on slot t, in units of bits.

The data is stored in queues $Q_1(t)$ and $Q_2(t)$ to await transmission

We assume the receiver coordinates network decisions every slot.



Example: opportunistic scheduling problem



Channel conditions are assumed to be constant for the duration of a slot, but they can change from slot to slot.

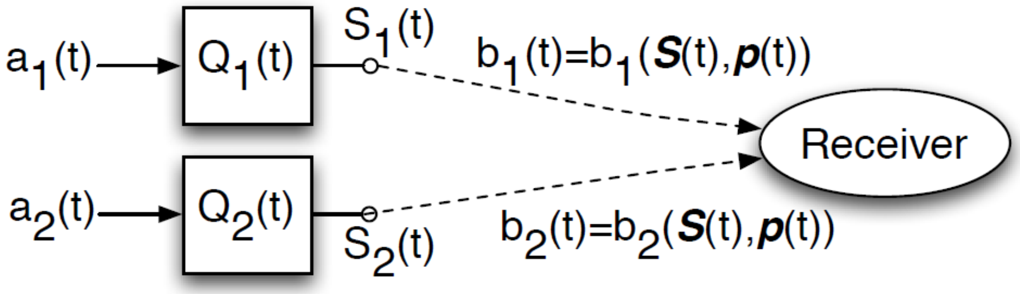
Let $\boldsymbol{S}(t) = (S_1(t), S_2(t))$ denote the channel conditions between users and the receiver on slot t.

The channel conditions represent any information that affects the channel on slot t, such as fading coefficients and/or noise ratios.

We assume the network controller can observe $\boldsymbol{S}(t)$ at the beginning of each slot t before making a transmission decision.

This channel-aware scheduling is called *opportunistic scheduling*.

Example: opportunistic scheduling problem



Every slot t, the network controller observes the current $\boldsymbol{S}(t)$ and chooses a *power allocation vector* $\mathbf{p}(t) = (p_1(t), p_2(t))$ within some set $\mathcal{P}$ of possible power allocations.

This decision, together with the current $\boldsymbol{S}(t)$, determines the *transmission rate vector* $(b_1(t), b_2(t))$ for slot t, where $b_k(t)$ represents the transmission rate (in bits/slot) from user $k \in \{1, 2\}$ to the receiver on slot t.

Specifically, we have general transmission rate functions $\hat{b}_k(\boldsymbol{p}(t), \boldsymbol{S}(t))$:

$$b_1(t) = \hat{b}_1(\boldsymbol{p}(t), \boldsymbol{S}(t)) \ , \ b_2(t) = \hat{b}_2(\boldsymbol{p}(t), \boldsymbol{S}(t))$$

The precise form of these functions depends on the modulation and coding strategies used for transmission.

The queueing dynamics are then:

$$Q_k(t+1) = \max[Q_k(t) - \hat{b}_k(\boldsymbol{p}(t), \boldsymbol{S}(t)), 0] + a_k(t) \quad \forall k \in \{1, 2\}, \forall t \in \{0, 1, 2, \ldots\}$$

Example: opportunistic scheduling problem

Several types of optimization problems can be considered for this simple system.

EXAMPLE PROBLEM 1: MINIMIZING TIME AVERAGE POWER SUBJECT TO STABILITY

Let $\overline{p}_k$ be the time average power expenditure of user k under a particular power allocation algorithm (for $k \in \{1, 2\}$):

$$\overline{p}_k \triangleq \lim_{t\to\infty} \frac{1}{t} \sum_{\tau=0}^{t-1} p_k(\tau)$$

The problem of designing an algorithm to minimize time average power expenditure subject to queue stability can be written mathematically as:

$$\begin{array}{lll} \text{Minimize:} & & \overline{p}_1 + \overline{p}_2 \\ \text{Subject to:} & 1) & \text{Queues } Q_k(t) \text{ are stable } \forall k \in \{1, 2\} \\ & 2) & \boldsymbol{p}(t) \in \mathcal{P} \;\; \forall t \in \{0, 1, 2, \ldots\} \end{array}$$

Our theory will allow the design of a simple algorithm that makes decisions $\boldsymbol{p}(t) \in \mathcal{P}$ every slot t, without requiring a-priori knowledge of the probabilities associated with the arrival and channel processes $\boldsymbol{a}(t)$ and $\boldsymbol{S}(t)$.

Example: opportunistic scheduling problem

EXAMPLE PROBLEM 2: MAXIMIZING THROUGHPUT SUBJECT TO TIME AVERAGE POWER CONSTRAINTS

Consider the same system, but now assume the arrival process $\boldsymbol{a}(t) = (a_1(t), a_2(t))$ can be *controlled* by a flow control mechanism.

We thus have two decision vectors: $\boldsymbol{p}(t)$ (the power allocation vector) and $\boldsymbol{a}(t)$ (the data admission vector).

The admission vector $\boldsymbol{a}(t)$ is chosen within some set $\mathcal{A}$ every slot t.

Let $\overline{a}_k$ be the time average admission rate (in bits/slot) for user k, which is the same as the time average throughput of user k if its queue is stable.

We have the following problem of maximizing a weighted sum of throughput subject to average power constraints:

$$
\begin{array}{lll}
\text{Maximize:} & & w_1\overline{a}_1 + w_2\overline{a}_2 \\
\text{Subject to:} & 1) & \overline{p}_k \leq p_{k,av} \ \ \forall k \in \{1, 2\} \\
 & 2) & \text{Queues } Q_k(t) \text{ are stable } \ \forall k \in \{1, 2\} \\
 & 3) & \boldsymbol{p}(t) \in \mathcal{P} \ \ \forall t \in \{0, 1, 2, \ldots\} \\
 & 4) & \boldsymbol{a}(t) \in \mathcal{A} \ \ \forall t \in \{0, 1, 2, \ldots\}
\end{array}
$$

Example: opportunistic scheduling problem

EXAMPLE PROBLEM 2: MAXIMIZING THROUGHPUT SUBJECT TO TIME AVERAGE POWER CONSTRAINTS

We have the following problem of maximizing a weighted sum of throughput subject to average power constraints:

$$\begin{array}{lll}
\text{Maximize:} & & w_1\overline{a}_1 + w_2\overline{a}_2 \\
\text{Subject to:} & 1) & \overline{p}_k \leq p_{k,av} \ \ \forall k \in \{1, 2\} \\
 & 2) & \text{Queues } Q_k(t) \text{ are stable } \ \forall k \in \{1, 2\} \\
 & 3) & \boldsymbol{p}(t) \in \mathcal{P} \ \ \forall t \in \{0, 1, 2, \ldots\} \\
 & 4) & \boldsymbol{a}(t) \in \mathcal{A} \ \ \forall t \in \{0, 1, 2, \ldots\}
\end{array}$$

where w_1, w_2 are given positive weights that define the relative importance of user 1 traffic and user 2 traffic, and

$p_{1,av}$, $p_{2,av}$ are given constants that represent desired average power constraints for each user.

Course Outline

- Part I: Convex optimization theory: a primer
  - Convex sets and convex functions
  - Optimization basics
  - Canonical problem forms
  - Gradient descent
  - Sub-gradients and Sub-gradient method
  - Duality and KKT conditions
- Part II: Economic mechanism design: a primer
  - Mechanism design for enforcing truth-telling (Auctions, VCG, Kelly, etc.)
  - Pricing and regulation mechanisms for internalizing network externalities
- Part III: Stochastic control and optimization: a primer
  - Markov decision processes (MDP)
  - Lyapunov optimization

References

- Refs for PART I:

  - R. Srikant and L. Ying, Communication Networks: An Optimization, Control, and Stochastic Networks Perspective. Cambridge University Press, 2013.
  - Boyd, Stephen P.; Vandenberghe, Lieven (2004). Convex Optimization. Cambridge University Press. ISBN 978-0-521-83378-3

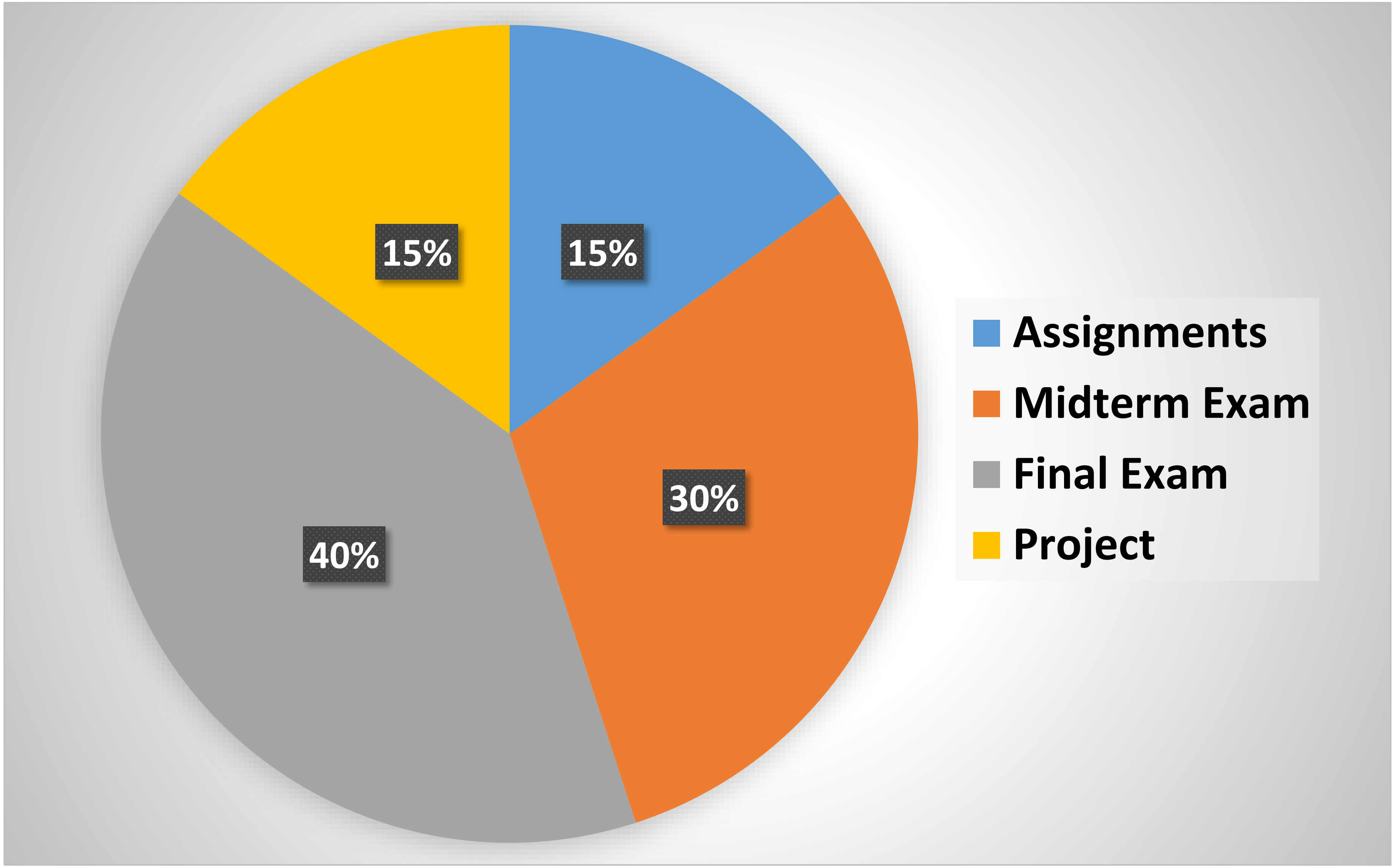
- Refs for PART II:

  - R. A. Berry and R. Johari, "Economic Modeling in Networking: A Primer", *Foundations and Trends in Networking*, Vol. 6, No. 3, pp. 165-286, 2013.
  - S. Shakkottai and R. Srikant, "Network Optimization and Control", *Foundations and Trends in Networking*, Vol. 2, No. 3, pp 271-379, 2008.

- Refs for PART III:

  - Sutton RS, Barto AG. Reinforcement Learning: An Introduction. Cambridge, Mass: A Bradford Book; 1998. 322 p.
    - The draft for the second edition is available for free: https://webdocs.cs.ualberta.ca/~sutton/book/the-book.html
  - M. J. Neely, "Stochastic Network Optimization with Application to Communication and Queueing Systems," Synthesis Lectures on Communication Networks, Morgan & Claypool Publishers, 2010.

Grading Scheme



Course Logistics

- Visit the following URL:

  http://www.edmodo.com

and,

- Get registered via the following “join code”:

  c87cmz

E-mail: iust.courses@gmail.com