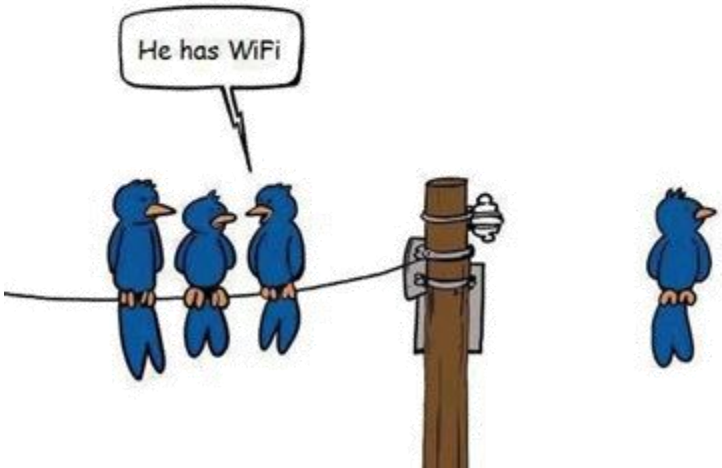
COM-405: Mobile Networks

Lecture 4.1: Scheduling

Haitham Hassanieh

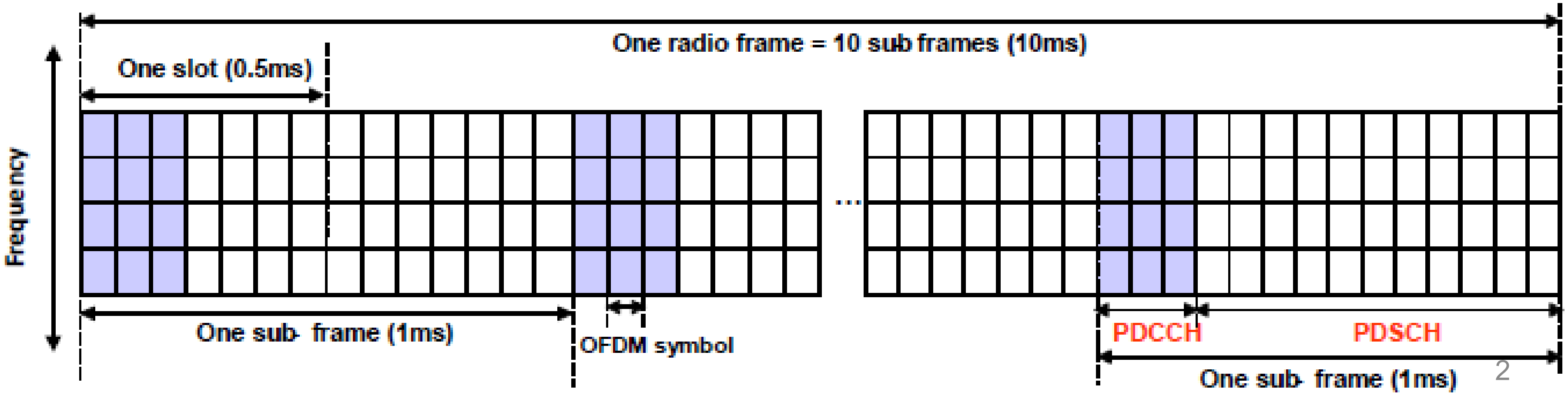
slides shamelessly stolen from Prof. JP Hubaux with minor modifications



Laboratory of SEnsing
& Networking Systems

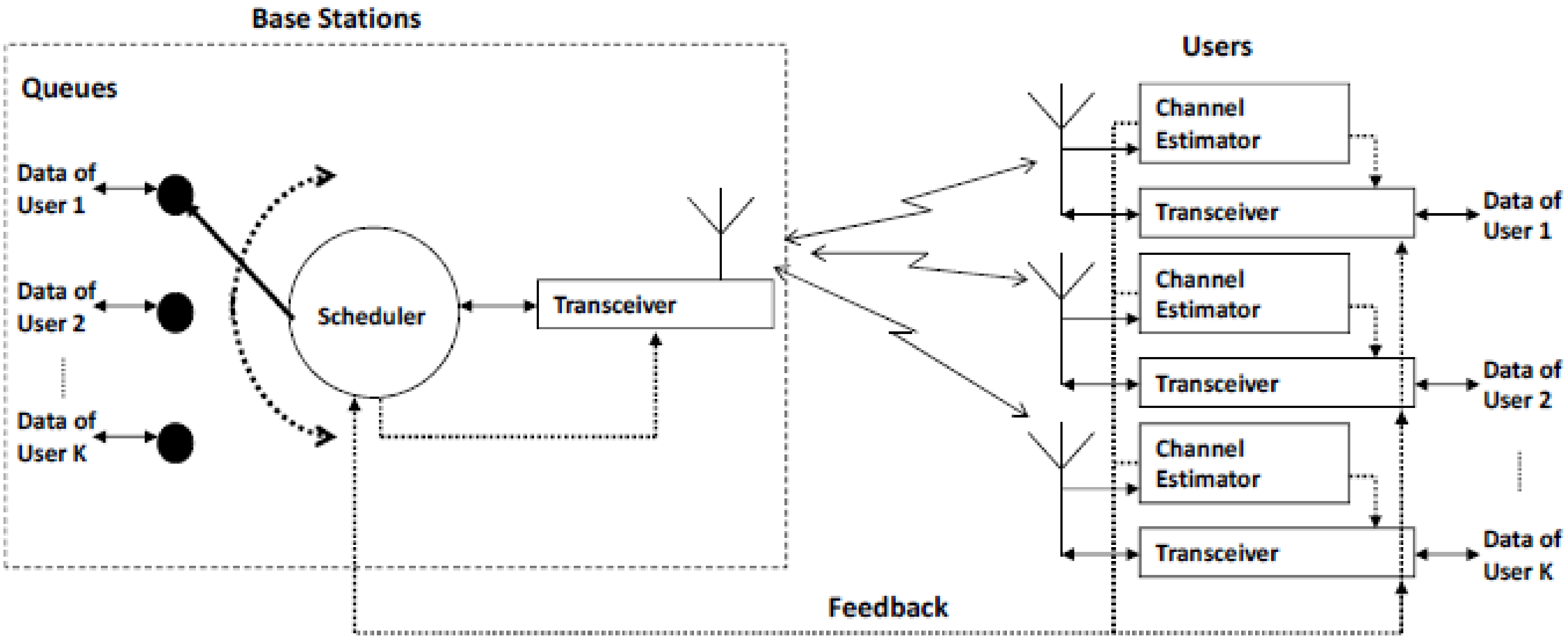
A Reservation-based Protocol

- Physical Downlink Control Channel (PDCCH): conveys control information for each user.
- Physical Downlink Shared Channel (PDSCH): multiplex the data of all terminals:
  - Each user will transmit on a unique set of Orthogonal Frequency Division Multiplexing (OFDM) symbols and frequency blocks.
- Reservation phase: PDCCH. Data phase: PDSCH.



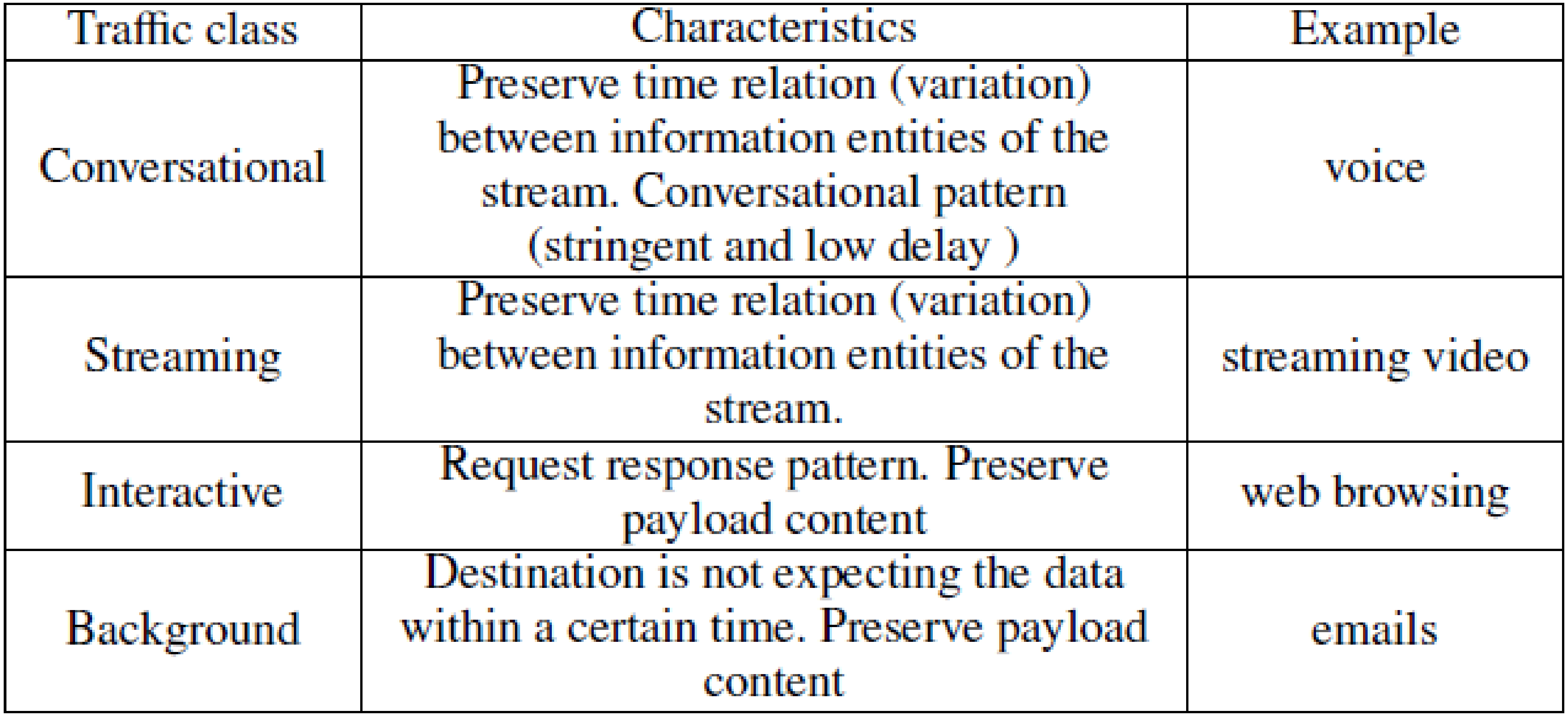
Role of Scheduling

- Reservation phase:
  - Estimate channel → Feedback → Scheduling



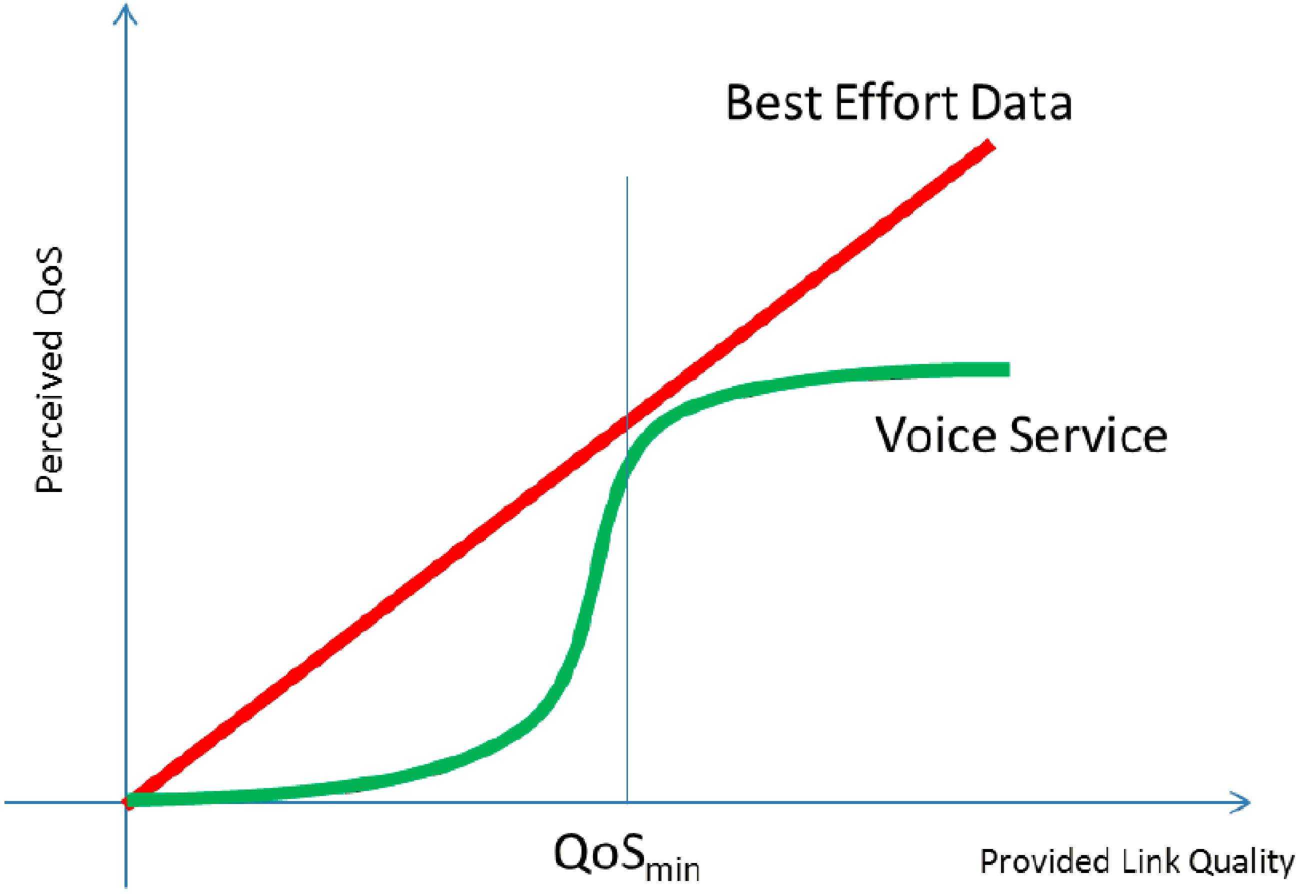
Issues in Wireless Scheduling

- Need to support mixed classes of traffic with different characteristics:



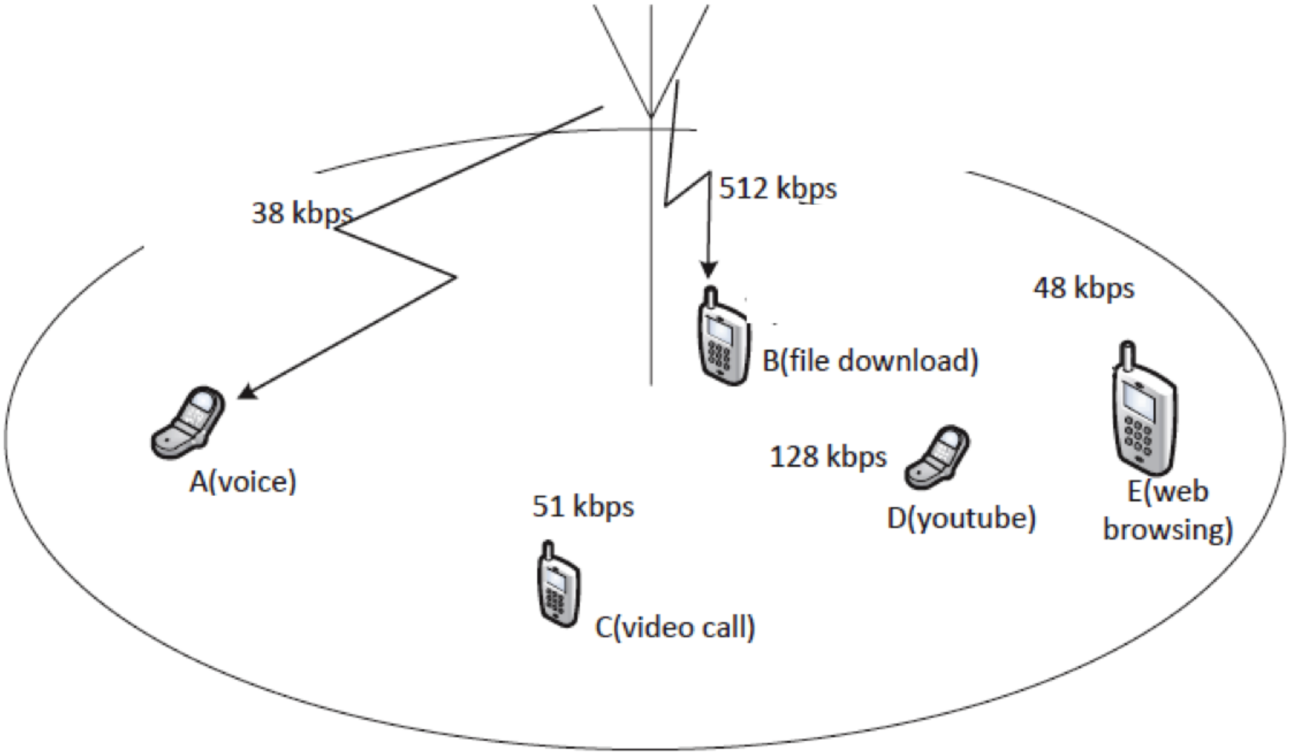
Traffic class	Characteristics	Example
Conversational	Preserve time relation (variation) between information entities of the stream. Conversational pattern (stringent and low delay)	voice
Streaming	Preserve time relation (variation) between information entities of the stream.	streaming video
Interactive	Request response pattern. Preserve payload content	web browsing
Background	Destination is not expecting the data within a certain time. Preserve payload content	emails

Perceived QoS



Issues in Wireless Scheduling

- Need to be flexible in allocating network resources
  - Achieved using scheduling algorithms



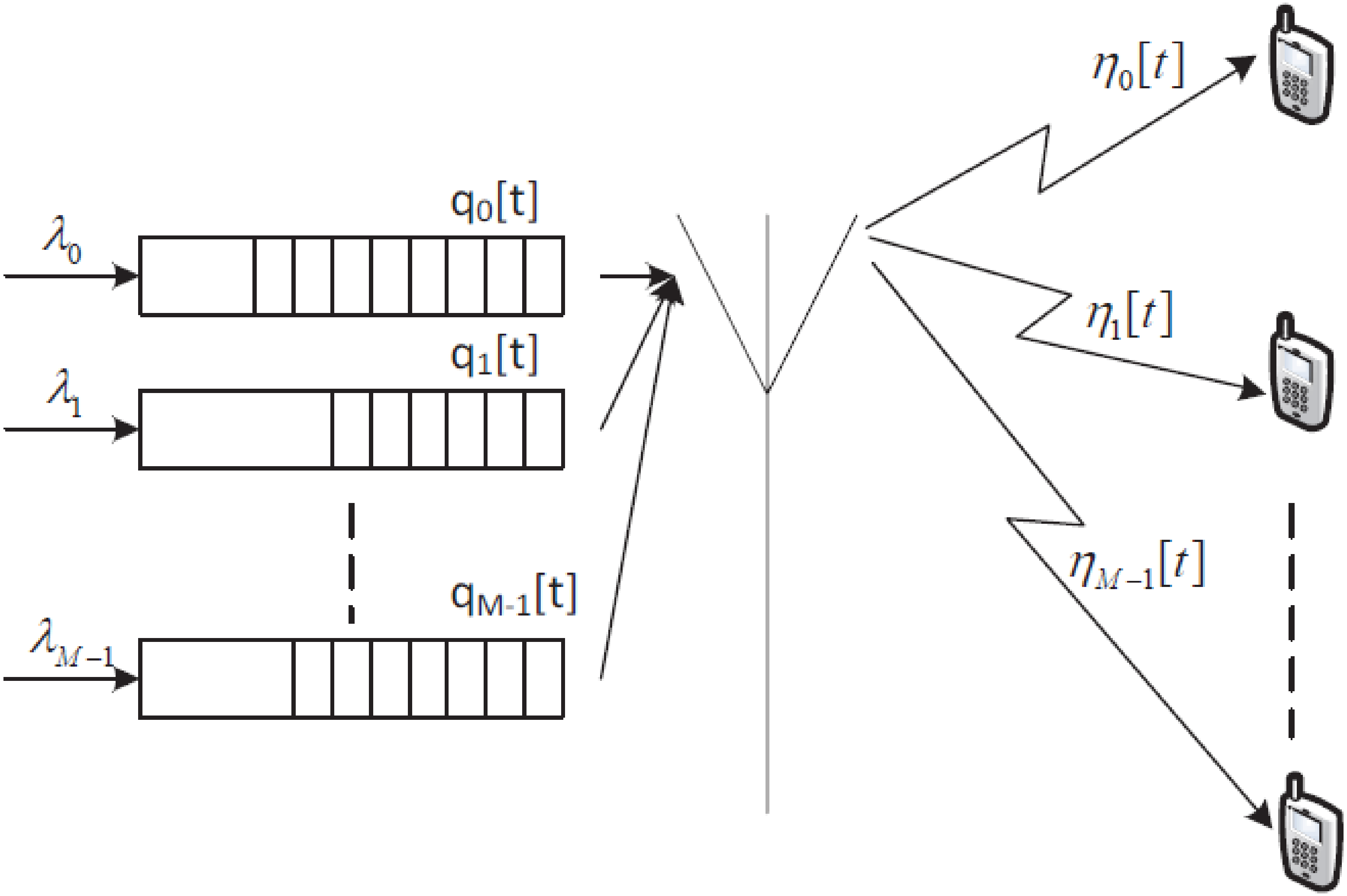
Note: In modern cellular networks, bitrates can be much higher than those in the picture

Issues in Wireless Scheduling

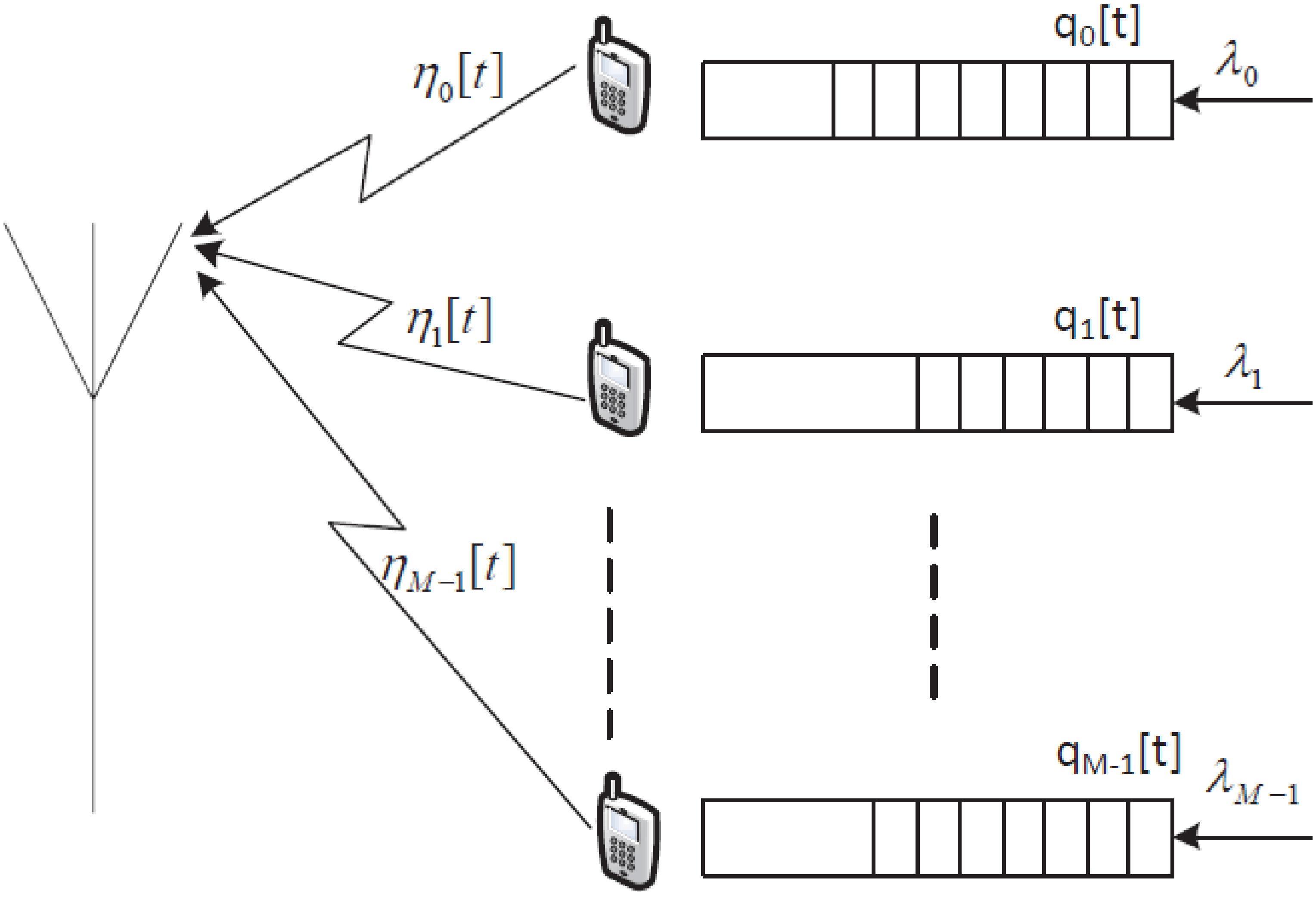
- Channel variation
  - Shadowing, fading, noise, interference, and user mobility
  - Unstable, error-prone, and hard to predict
  - Capacity of each link varies significantly in different time periods and locations

- Even if the scheduler knows QoS requirements
  - Difficult to estimate the amount of resources needed
  - An adaptive procedure is needed to assure QoS, taking into account the requirements and channel variations

Downlink Scheduling



Uplink Scheduling



Queue Modeling

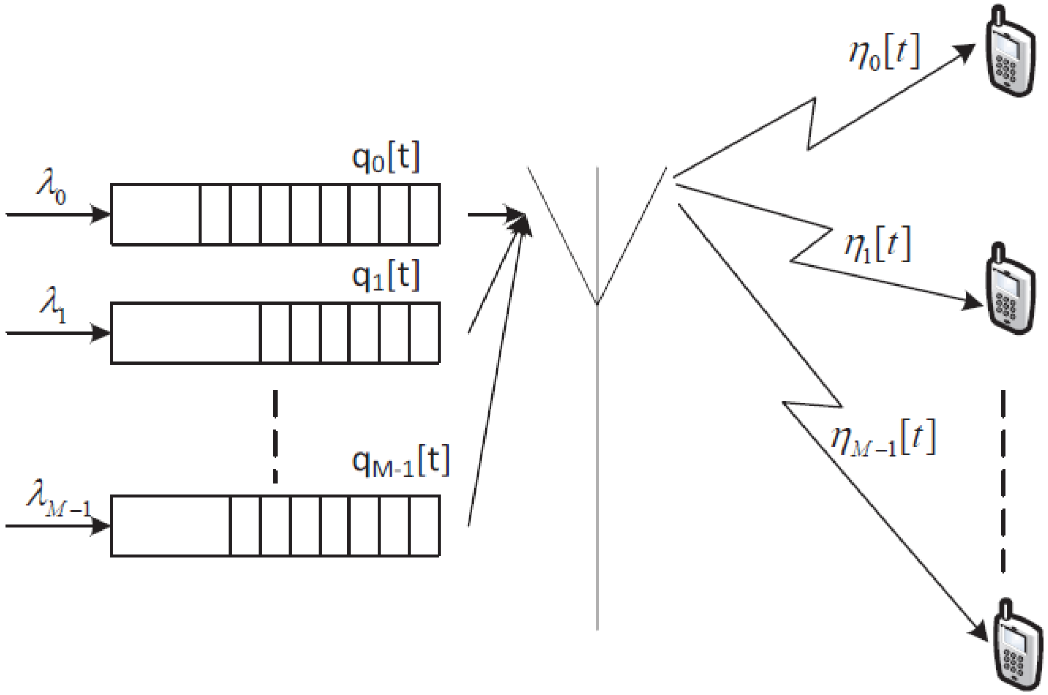
- At timeslot t, the queue of user i changes as follows:

$$q_i[t+1] = q_i[t] + \delta_i[t] - \eta_i[t]$$

Number of bits in the queue

Number of new bits arriving

Number of bits scheduled to transmit, determined by the scheduling algorithm



Round-Robin (RR) Scheduling

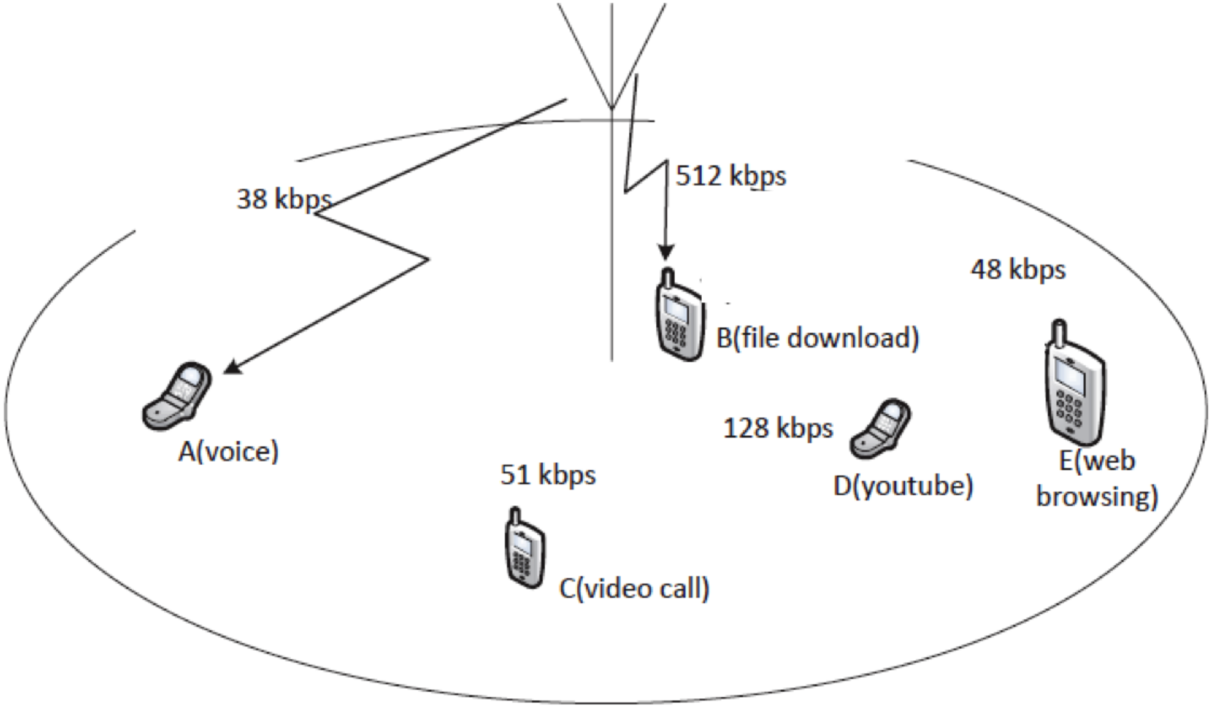
- Users are scheduled in a round robin, i.e. cyclic order
- i[t]: user scheduled at time t. RR scheduler:

$$i[t+1]=i[t]+1 \pmod{M-1}$$

- The algorithm is fair: all users are given the same amount of time resources

Round-Robin Scheduling

- Performance
  - All users are allocated the same amount of network resources
  - What is the throughput of all users in the following network?



Max Throughput Scheduling

- Objective: maximize total network throughput
- If user i is scheduled, the expected data rate is:

$$\widehat{r_i}[t] = \frac{\hat{\eta}_i[t]}{T}$$

Expected number of bits that can be successfully delivered

Slot length

- The total expected network throughput is

$$\hat{r}[t] = \sum_{i=0}^{M-1} \widehat{r_i}[t] I(i)$$

I(i): Scheduling indicator:
1 scheduled, 0 otherwise.

Max Throughput Scheduling

- Schedule the user with the highest expected data rate
  - one way of estimating $\widehat{r_i}[t]$ is:

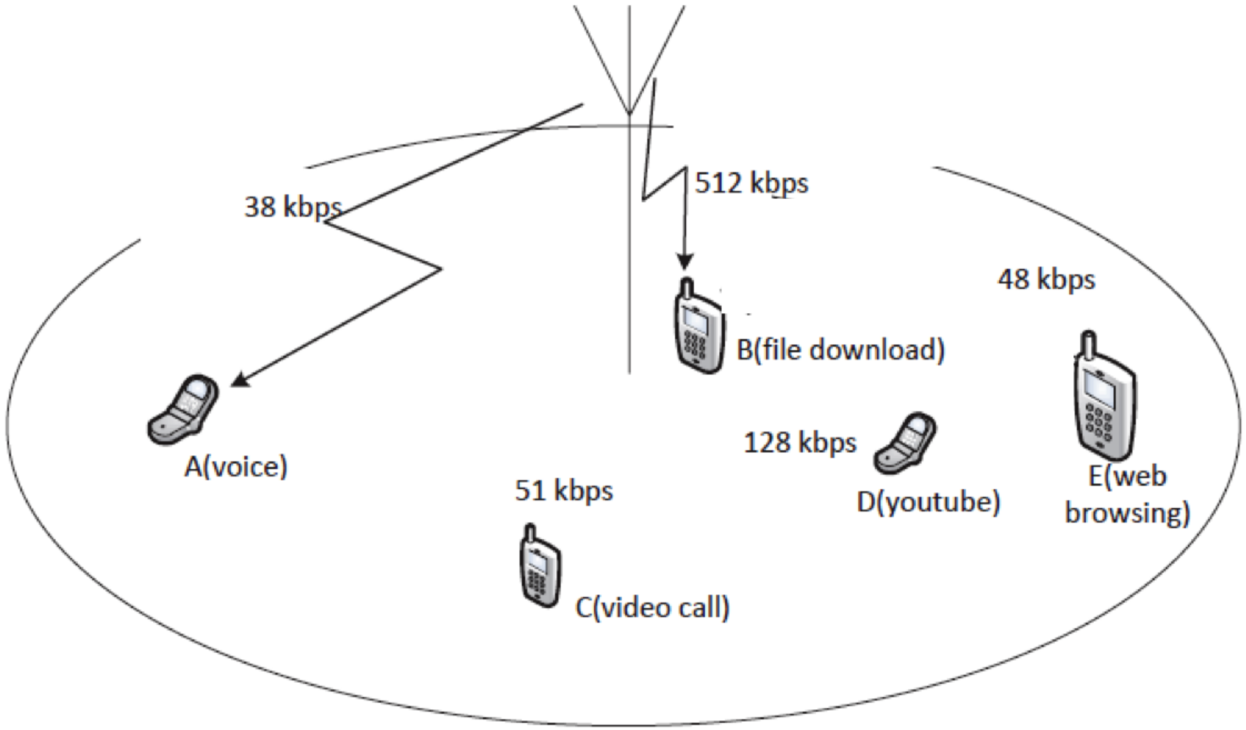
$$\widehat{r_i}[t] = B \log_2\left(1 + \frac{\mathrm{SINR}_i[t]}{\theta}\right)$$

where:

- B is the frequency bandwidth
- $\mathrm{SINR}_i[t]$ is the signal-to-interference-plus-noise ratio (SINR) at time t given the allocated power.
- θ is the SINR gap that defines the gap between the channel capacity and a practical coding and modulation scheme

Max Throughput Scheduling

- Main drawbacks
  - Unfairness
  - Coverage limitation
  - Most users may never be served



Proportional Fair Scheduling

- PF scheduling: balance the competing interests of network throughput and minimum service level
- Objective: maximize

$$\sum_{i=0}^{M-1} \ln S_i$$

- S_i: long-run throughput for user i can be predicted using

$$\widehat{S}_i[t] = \left(1 - \frac{1}{\tau}\right) S_i[t-1] + \frac{1}{\tau}\, \widehat{r}_i[t] I(i)$$

where $\tau \gg 1$ is a constant defined by the scheduler.

→ schedule the user with the highest $\frac{\widehat{r}_i[t]}{S_i[t-1]}$

Proportional Fair Scheduling

- Meet the proportional fairness criterion:

Assuming s_i is the optimal value, for any other feasible value, the sum of proportional changes is non-positive, i.e.

$$\sum_i \frac{\hat{s}_i - s_i}{s_i} \leq 0$$

- Intuition: when the scheduling result is already proportionally fair, if scheduling is changed s.t. the throughput of a user is increased by a percentage, the cumulative decrease of the throughputs of other users will be higher
- In other words, any attempt of improvement somewhere will generate a higher damage elsewhere

Proportional Fair Scheduling

- Why proportional fair

Assume $\delta_i \ll s_i$

$$\sum_{i=0}^{M-1} \ln(s_i+\delta_i) = \sum_{i=0}^{M-1} \ln(s_i) + \sum_{i=0}^{M-1} \ln\left(1+\frac{\delta_i}{s_i}\right)$$

$$\approx \sum_{i=0}^{M-1} \ln(s_i) + \sum_{i=0}^{M-1} \frac{\delta_i}{s_i}$$

but $\sum_i \frac{\delta_i}{s_i} \leq 0$

$$\leq \sum_{i=0}^{M-1} \ln(s_i)$$

Max-Min Scheduling

- Objective: maximize the minimum user throughput

$$\max \min_{i} S_i$$

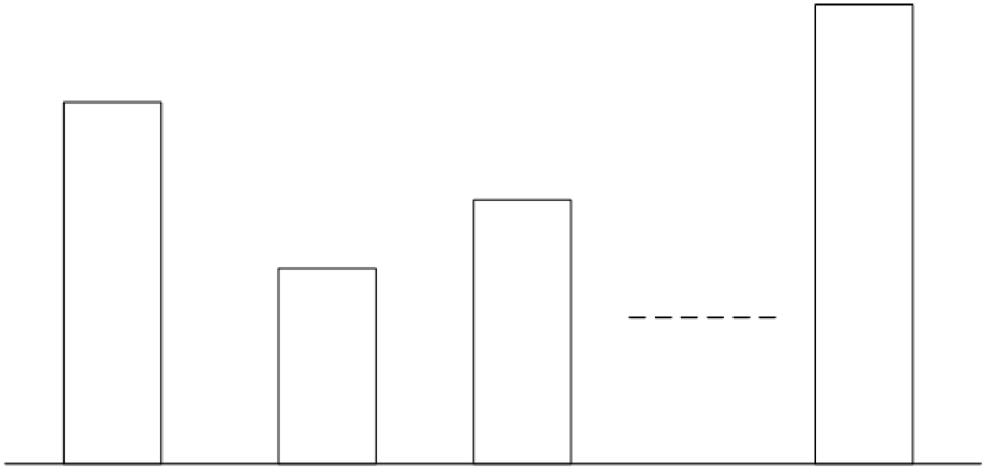
- A scheduling result is max-min fair if and only if a further increase of throughput of one user will result in the decrease of a user with a smaller throughput

$$\widehat{S}_i[t] = \left(1 - \frac{1}{\tau}\right) S_i[t-1] + \frac{1}{\tau} \hat{r}_i[t] I(i)$$

- Schedule the user with the minimum $(1 - \frac{1}{\tau}) S_i[t-1]$, i.e. the one with the smallest throughput at time t-1.

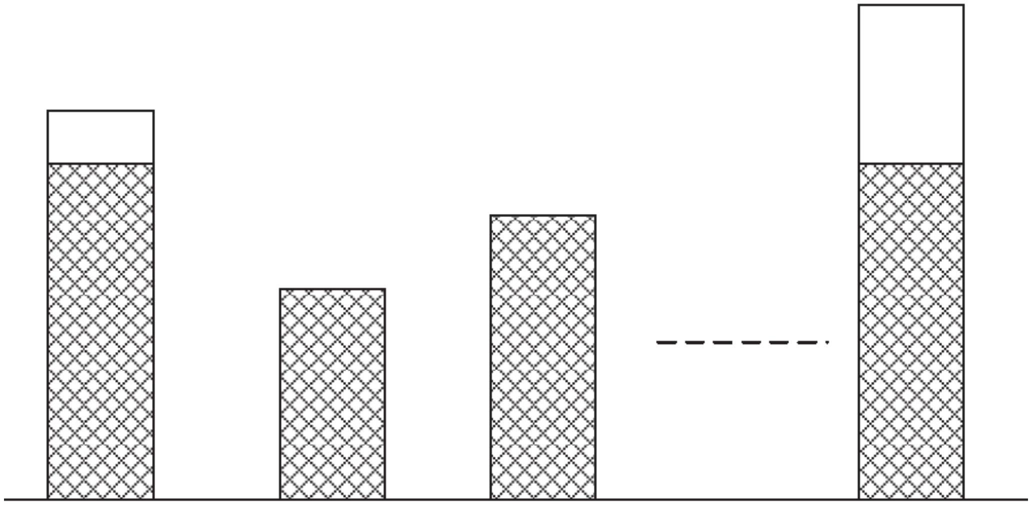
Max-Min Scheduling

- M empty cylindrical buckets (users), all with the same radius but different heights.
- Allocate water

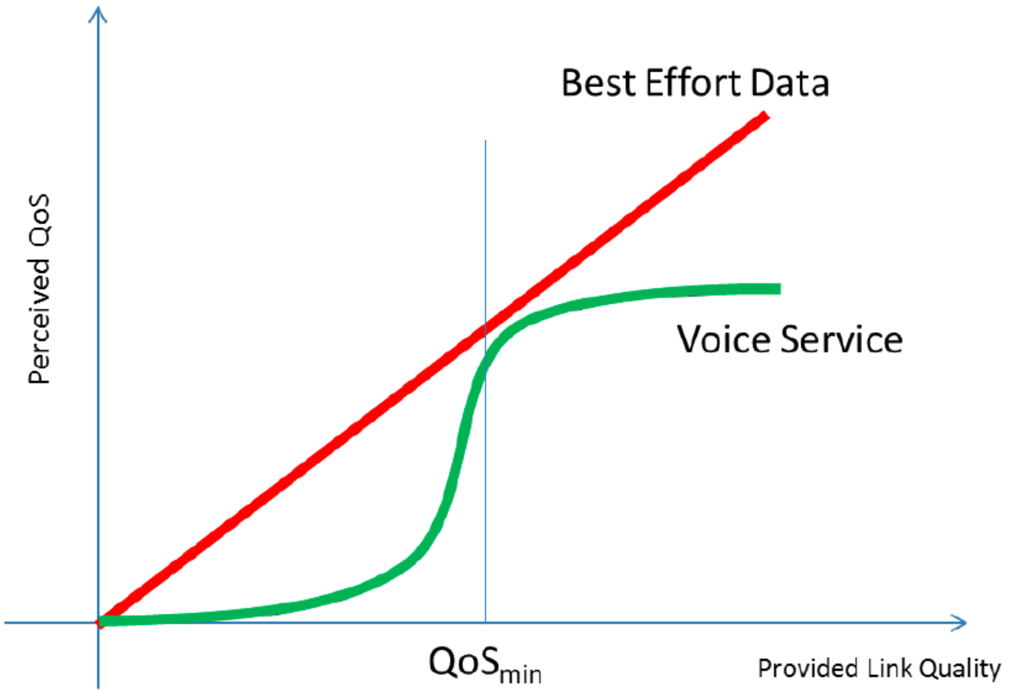


Max-Min Scheduling

- M empty cylindrical buckets (users), all with the same radius but different heights.
- Allocate water



Can these schedulers deal with QoS?



Max Utility Scheduling

- Previous schedulers do not consider QoS
- Utility-based scheduling
  - Utility quantifies the satisfaction of each user given the allocated resources
  - Model the QoS perception of users
  - Objective: maximize the sum utility of all users, i.e. total network satisfaction
- Utility functions: model how user perceives services

Max Utility Scheduling

$$\max \sum_{i=0}^{M-1} U_i(S_i)$$

- Different utility functions can be designed.
- In addition to QoS modeling, different utility functions can be designed to reflect fairness and efficiency.
- Max-throughput (highest efficiency):

$$U(S) = S$$

- Proportional fair:

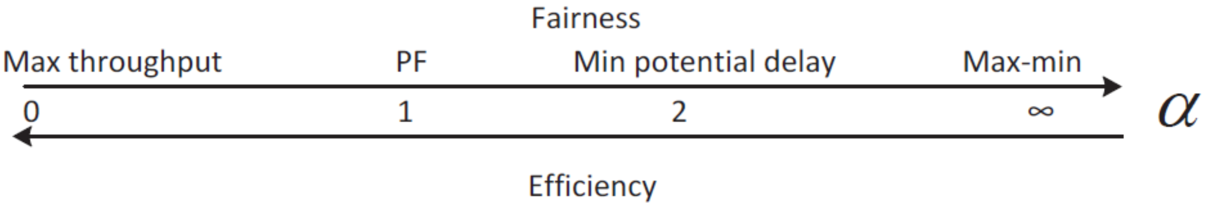
$$U(S) = \ln(S)$$

Max Utility Scheduling - Alpha Fair Utility

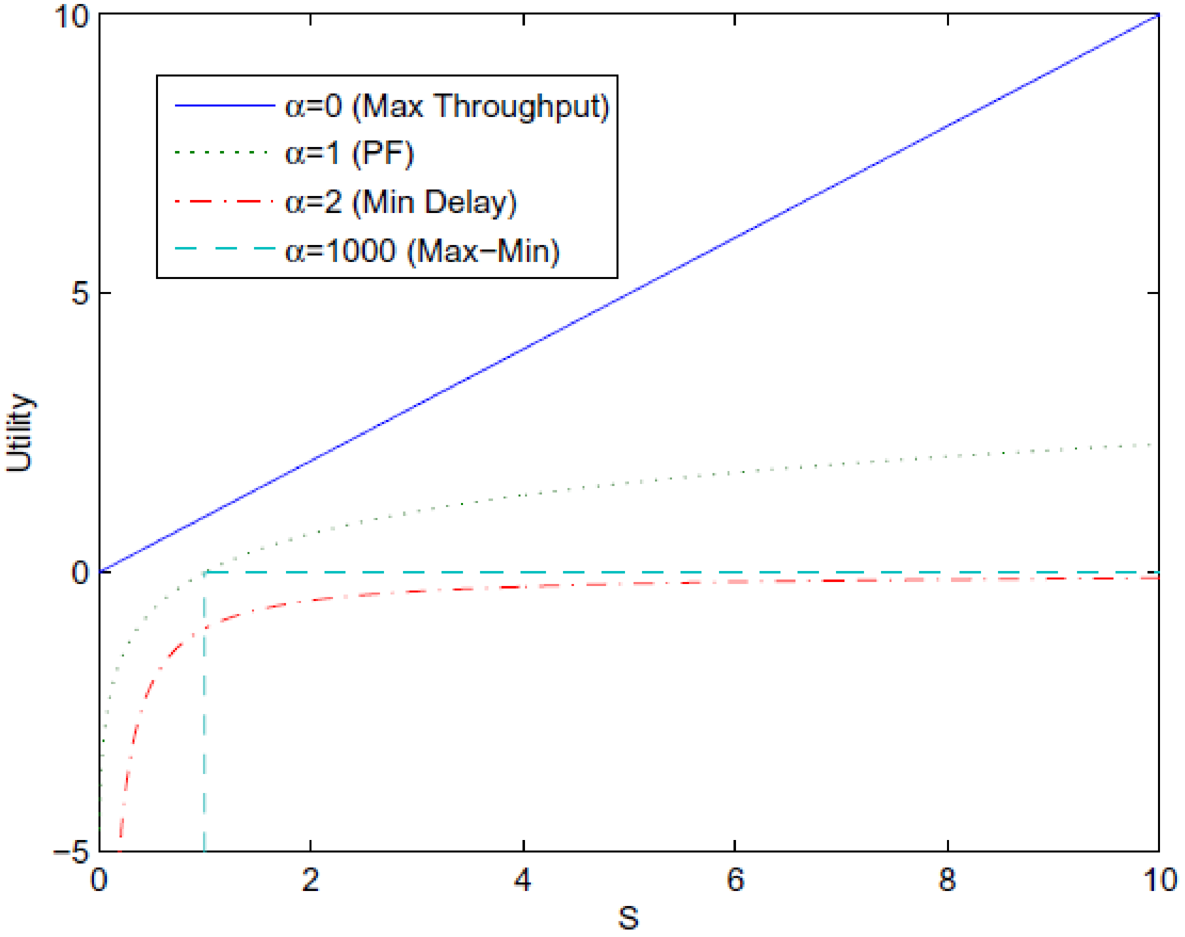
- More generic definition: α fair scheduling

$$U_\alpha(S) = \begin{cases} \frac{S^{1-\alpha}}{1-\alpha} & \alpha \geq 0 \text{ and } \neq 1 \\ \ln(S) & \alpha = 1. \end{cases}$$

- α measures how fair the scheduling result is
  - 0: Max throughput;
  - 1: Proportional fair;
  - 2: equivalent to minimizing $\sum_{i=0}^{M-1} \frac{1}{S_i}$
    - Minimize the total potential delay
  - Infinity: Most fair, max-min scheduler.



Alpha Fair Utility

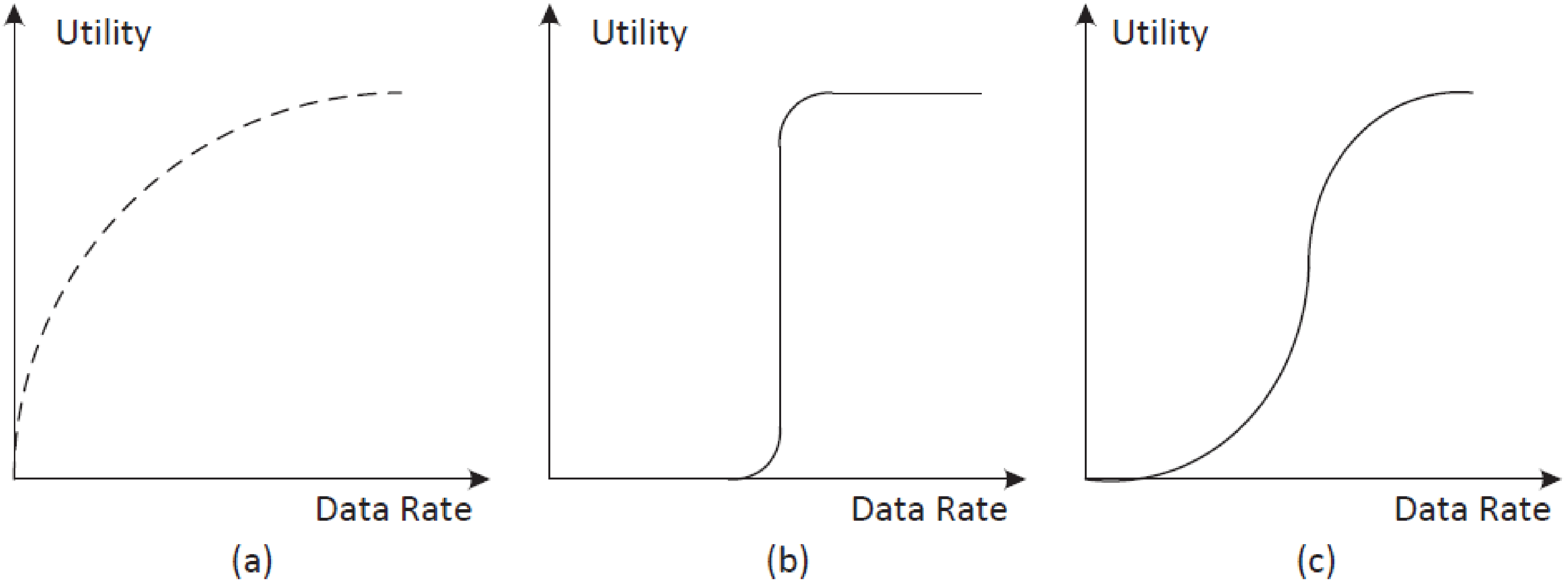


PF: proportional fairness

Resource

Utility Functions with QoS Consideration

- Determined based on traffic characteristics



- (a) best effort; (b) real time with tight delay requirement; (c) real time with loose delay requirement.