Network Management & Optimal Design Laboratory

Probabilistic Short-term Delay and Throughput Requirements of Multimedia Services in High Throughput Wireless Networks

Symeon Papavassiliou and Timotheos Kastrinogiannis

School of Electrical and Computer Engineering
National Technical University of Athens

Introduction

- Future wireless networks will be characterized by:
  - Higher transmission data rate
  - Integration of services
  - Flexibility and QoS Provisioning
- Scheduling in HDR systems plays a key role in achieving above goals
  - Schedule users' transmission and allocate available resource with efficiency and fairness
- Transmission scheduling over a wireless environment is very challenging due to
  - Users' time-varying channel condition that will eventually result in time-varying service quality and, due to
  - Different physical layer technologies.

Channel-Aware Scheduling

- Time-varying channel characteristic **_multiuser diversity_** can be "positively" exploited to improve overall system performance.
  - **Principle**: In a wireless system with multiple users having independent time-varying channels, there exists one with instantaneous signal to interference and noise ratio larger than others.
  - **Opportunistic Scheduling**: Overall resource allocation can be maximized by providing service at any time only to the user/users with the highest instantaneous channel quality.

Services' Basic Characteristics

- *Non Real Time Services*
  - Long term satisfaction of QoS requirements.
  - Smooth tuning procedure responses when violations occur – Perfectly aligned with the intrinsic nature of an opportunistic scheduler.
  - High Throughput Performance.

- *Real Time Services*
  - QoS requirements satisfied within short-time intervals.
  - QoS requirements violation demands the scheduler's response within small time-intervals.
  - Upper and lower bounded rate demands.

Non Real Time Services' QoS Requirements Classification

- *Minimum performance guaranties per user individually*
  - Target throughput rates
  - Lower – Maxim bounds for users' average transmission rates
- *Users' Fairness Issues (GPS - Generalized Processor Sharing models)*
  - Long-term access time fairness
  - Long-term throughput fairness

Real-Time Services QoS requirements (I)

- *Strict delay constraints*
  - Every user i receives $\varphi_i M$ time slots with every non-overlapping M successive slot time intervals.
- *Probabilistic short-term delay constraints*

$$\Pr[d_i \geq T_i] < g_{STD,i}$$

where:

d_i denotes user's i r.v. that represents his delay in terms of successive time slots at which he has no access at the system's resources,

T_i denotes user's i delay threshold,

$g_{STD,i}$ denotes user's i probabilistic delay constraint threshold.

Real-Time Services QoS requirements (II)

- *Probabilistic short-term throughput constraints*

$$\Pr[b_i < B_i(T_{B_i})] < g_{STT,i}$$

where:

- T_{Bi} denotes a RT user's *i* observation time interval in terms of slots,
- $B_i(T_{Bi})$ denotes user's *i* predefined data units threshold,
- b_i *denotes* his received amount of data within a specific time interval from slot $t- T_{Bi}+1$ to slot *t,*
- $g_{STT,i}$ denotes user's *i* probabilistic short-term throughput constraint threshold.

Objectives

- To provide analytical framework regarding users' both throughput and access time properties under basic opportunistic schedulers, in order
  - To identify the main system regulating parameters that affect the users' performance and
  - To determine at what extent opportunistic scheduling policies can be used, adopted and/or modified for the support of real-time traffic services in addition to non-real-time services under a unified and integrated framework.

A framework for computing Real Time services QoS properties

if a user's probability of being selected at any time instant under a specific scenario is **fixed** then

- RT user's short-term delay properties:

$$\Pr\left[d_i \geq T\right] = \left[1 - \Pr\left[i_Q^*(t) = i\right]\right]^T \quad (1)$$

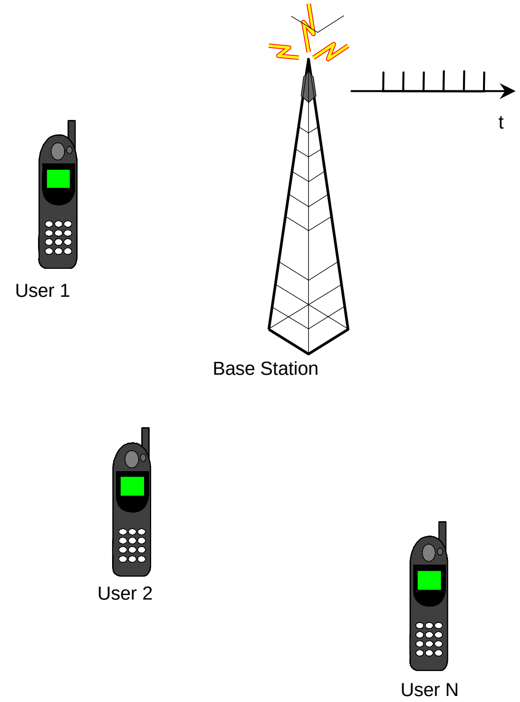
- RT user's short-term throughput properties:

$$\Pr[b_i < B_i(k, \overline{R}_{Q,i})]_{T_{B_i}} = \sum_{k'=0}^{k-1} \binom{T_{B_i}}{k'} (\Pr\left[i_Q^*(t) = i\right])^{k'} (1 - \Pr\left[i_Q^*(t) = i\right])^{T_{B_i} - k'} \quad (2)$$

where

$$B_i(k, \overline{R}_{Q,i}) = k \cdot \overline{R}_{Q,i} \cdot T_s$$

System Model



- Single cell DS-CDMA over TDMA system
  - *W* spreading bandwidth
  - No power control
  - Base Station transmission power P_{max}

- User's *i* received SINR

$$\rho_i(t) = G_i(t)\frac{P_{\max}}{I_i + Wn_0}$$

G_i the corresponding channel gain,
I_i the intercell interference
n_0 is the one-side power spectrum density of AWGN

- User's *i* achievable transmission rate (linear relationship)

$$r_i(t) = \rho_i(t) \cdot W / \gamma_i$$

- User's Rayleigh fading channels, independently distributed
  - fixed expected values $E[\rho_i(t)] = \bar{\rho}_i$
  - density functions $f_{\rho_i}(\rho_i) = e^{-\rho_i/\bar{\rho}_i} \Big/ \bar{\rho}_i$ (3)

Basic Opportunistic Schedulers (I)

A. Greedy Opportunistic Scheduler (MAX)

$$i_M^*(t) = \arg\max_{i \in S}\{\rho_i(t)\}$$

B. Proportional Fair Algorithm (PF)

$$i_{PF}^*(t) = \arg\max_{i \in S}\{\rho_i(t) / \rho_i^*(t)\}$$

$$\rho_i^*(t+1) = (1-\beta)\rho_i^*(t) + \beta\rho_i^{'}(t)$$

$$\rho_i^{'}(t) = \begin{cases} \rho_i(t) & \text{if user i is currently receiving service} \\ 0 & \text{otherwise} \end{cases}$$

C. Relatively Best Algorithm (RB)

$$i_{RB}^*(t) = \arg\max_{i \in S}\{\left(\rho_i(t) - \bar{\rho}_i(t)\right) / c_i\}$$

where c_i are positive control parameters.

Basic Opportunistic Schedulers (II)

Two types of users:

- N_g (S_g) users with "good" channel conditions
  - Mean achieved SINR $\bar{\rho}_g$
- N_b (S_b) users with "bad" channel conditions
  - Mean achieved SINR $\bar{\rho}_b$

Consider: $\bar{\rho}_g = a \cdot \bar{\rho}_b$ **where** $a \geq 1$

Observations

- *Max Algorithm*
  - Under a specific users' scenario a RT user's access probability is fixed and depends only on:
    - The number of the users in the system
    - The ratio of good to bad users
    - The ratio of good to bad users' average channel condition
- *PF Algorithm*
  - A user's probability of being selected at a time instant depends only on the number of active users in the system.
  - Algorithm's property of providing long-term access time fairness to all users
- *RB Algorithm*
  - Under a specific users' scenario a RT user's access probability is fixed and depends only on:
    - The number of the users in the system
    - The ratio of good to bad users
    - The ratio of good to bad users' average channel condition

Numerical Results (Scheduling Scenarios)

Number of RT continuously backlogged users: 7

Good to bad users' SINR ratio: 2

- Scheduling Scenario 1 (SC1)

$$\bar{\rho}_{g,sc1} = 3dB \ \ and \ \bar{\rho}_{b,sc1} = 0dB$$

- Scheduling Scenario 2 (SC2)

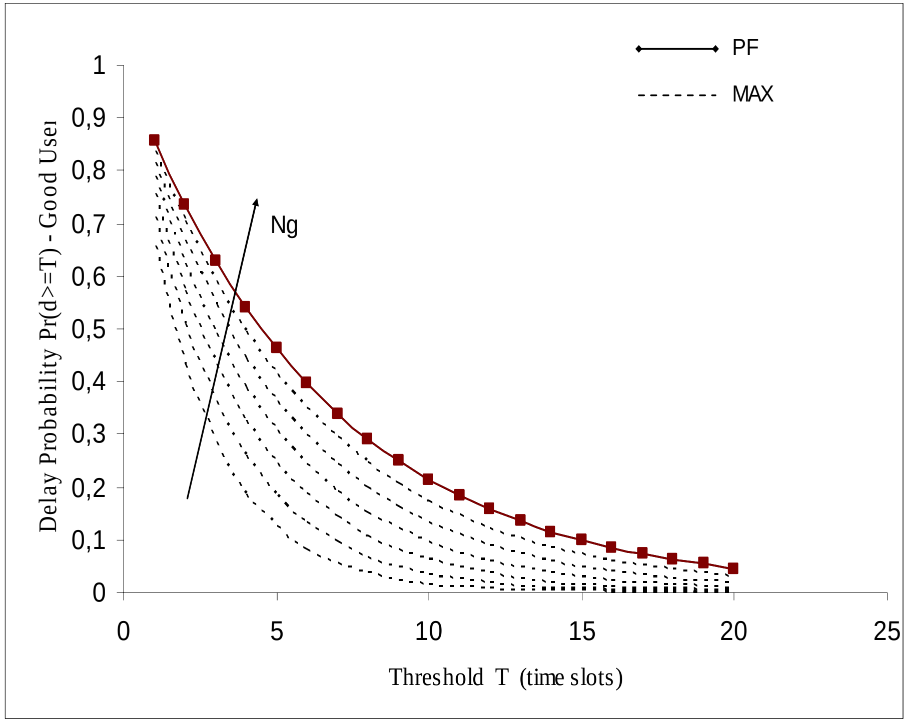
$$\bar{\rho}_{g,sc2} = 6dB \ \ and \ \bar{\rho}_{b,sc2} = 3dB$$

Numerical Results (Short-Term Delay Properties) (I)

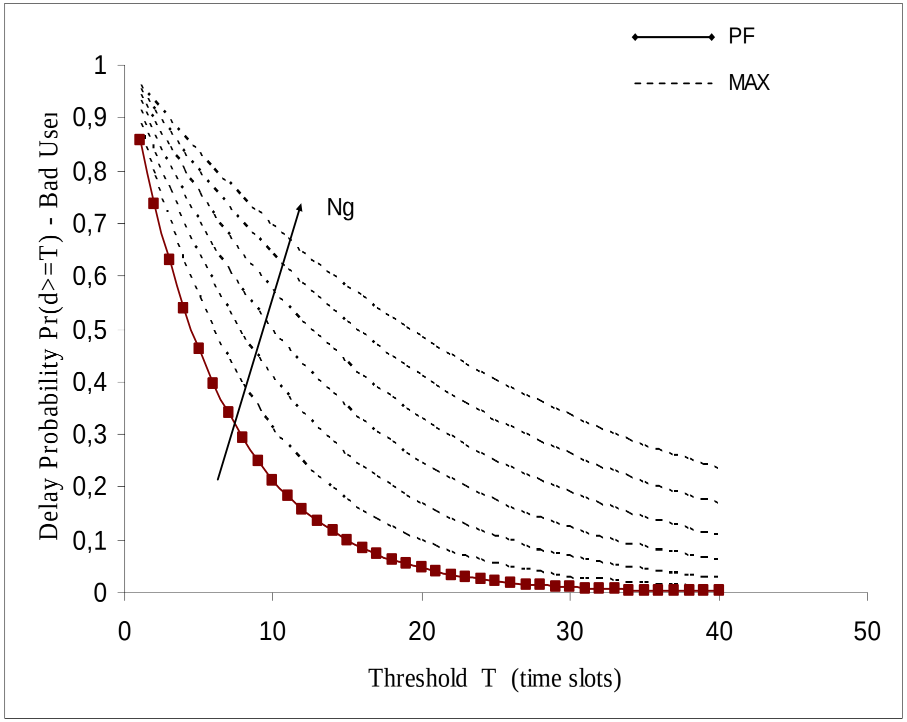
Scheduling Scenario 1

PF vs MAX

Good users case



Bad users case



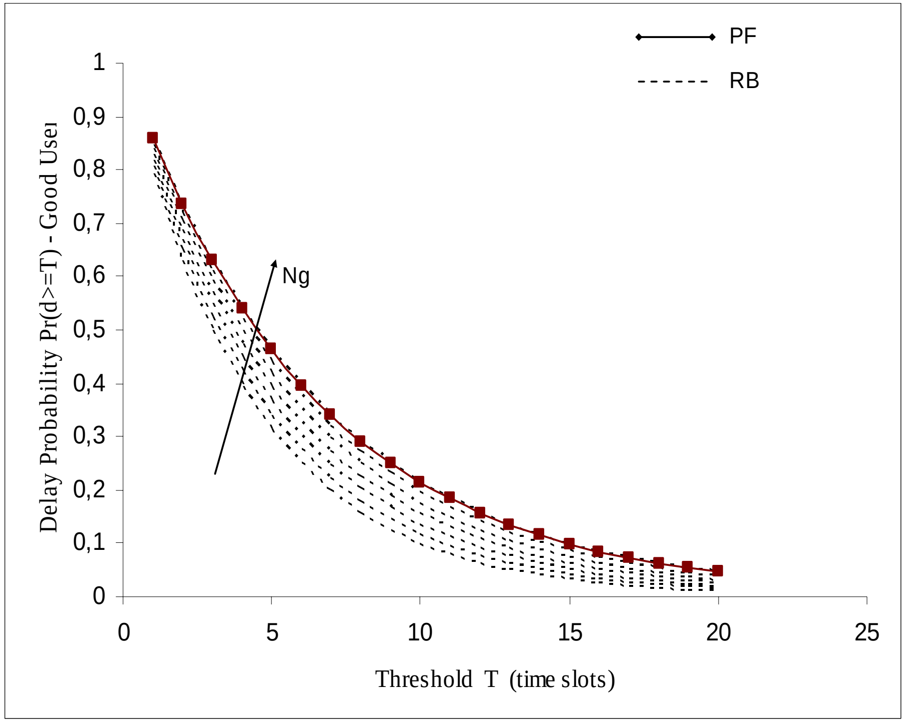
Consider threshold $\Pr_{Q,i}[d \geq T] \leq 0.001$

Numerical Results (Short-Term Delay Properties) (II)

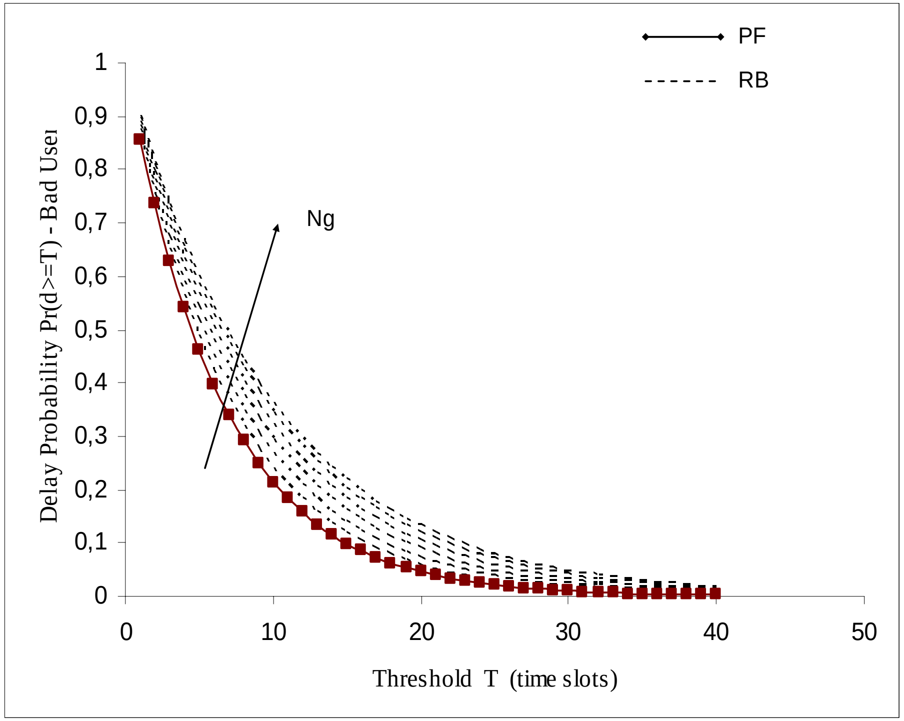
Scheduling Scenario 1

PF vs RB

Good users case



Bad users case



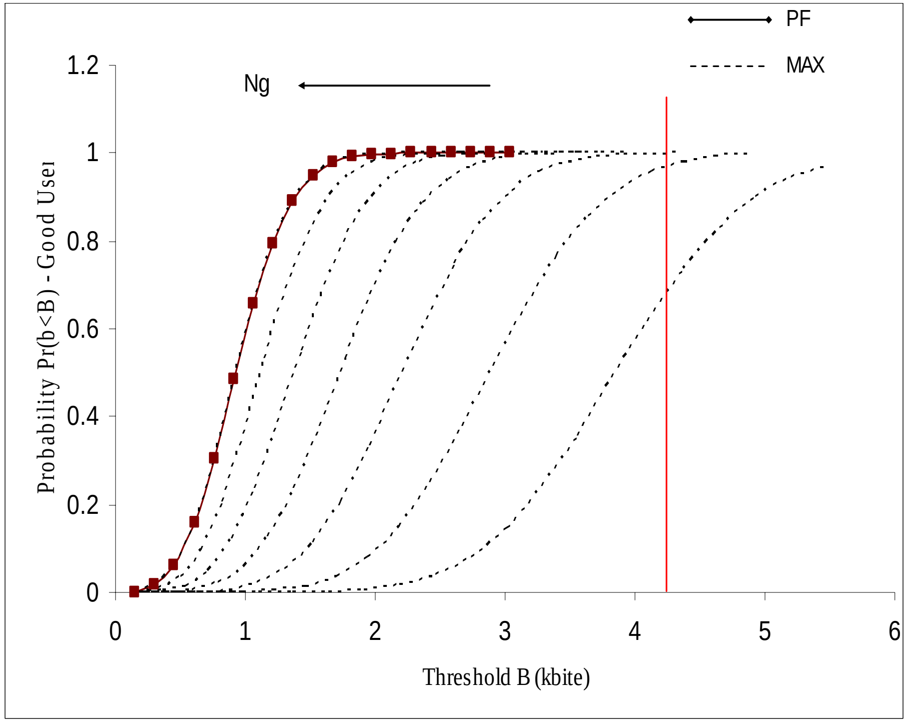
Consider threshold $\Pr_{Q,i}[d \geq T] \leq 0.001$

Numerical Results (Short-Term Throughput Properties) (I)

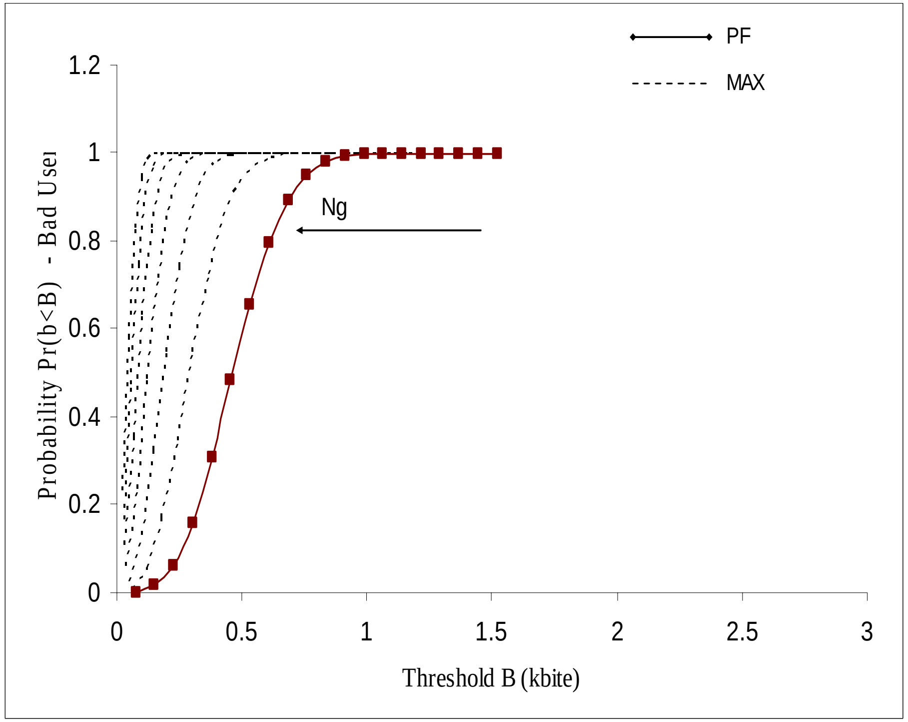
Scheduling Scenario 1

PF vs MAX

Good users case



Bad users case



Vertical red line indicates the considered threshold

$$\Pr_{Q,i}[b_i < B_i]_{T_{B_i}=40} < 0{,}001$$

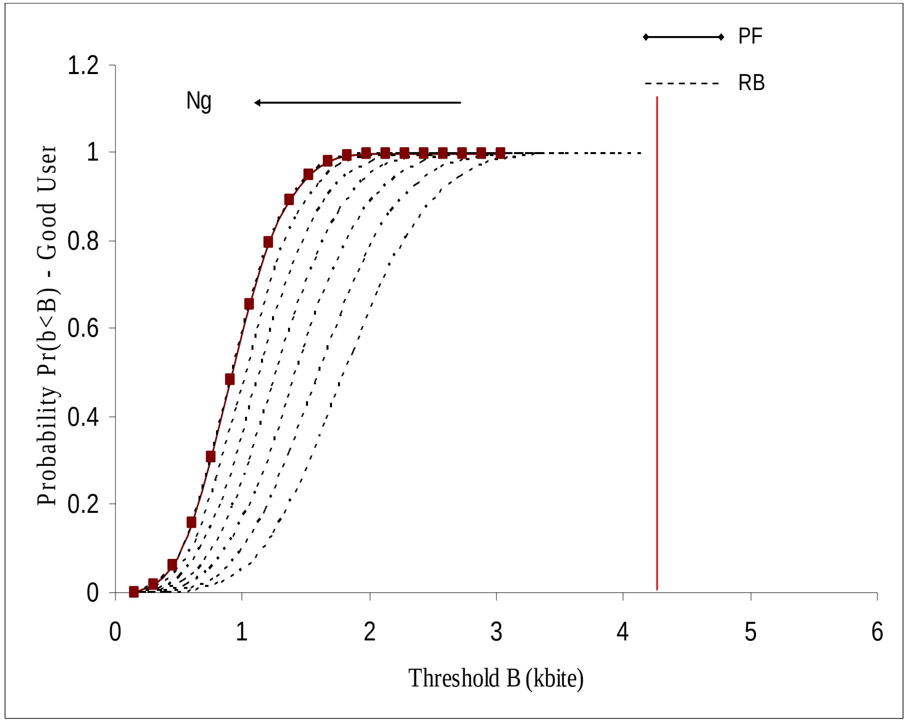
Numerical Results (Short-Term Throughput Properties) (II)

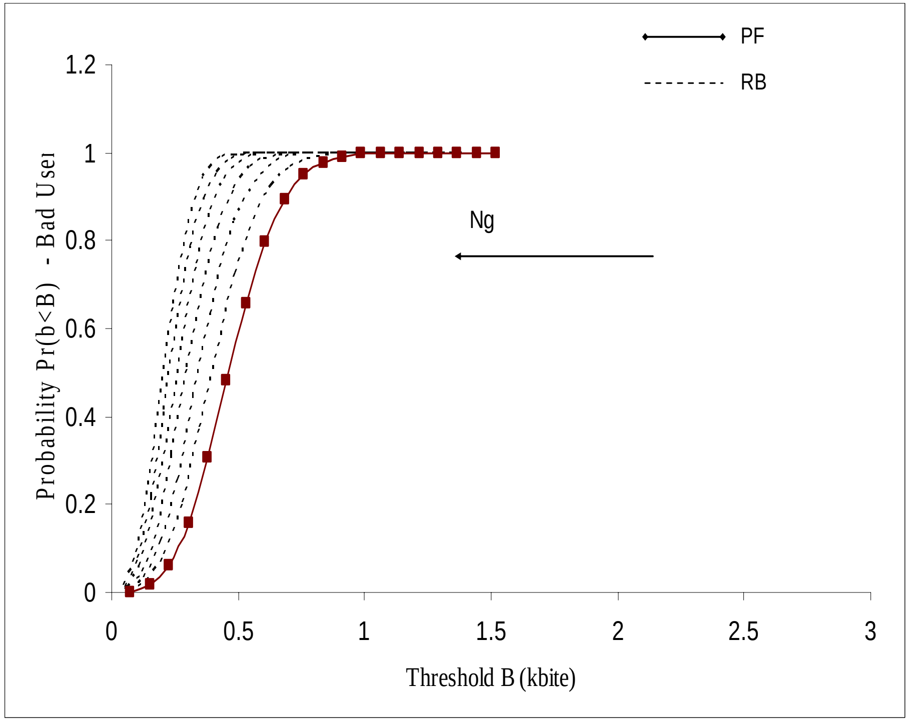
Scheduling Scenario 1

PF vs RB

Good users case

Bad users case





Vertical red line indicates the considered threshold

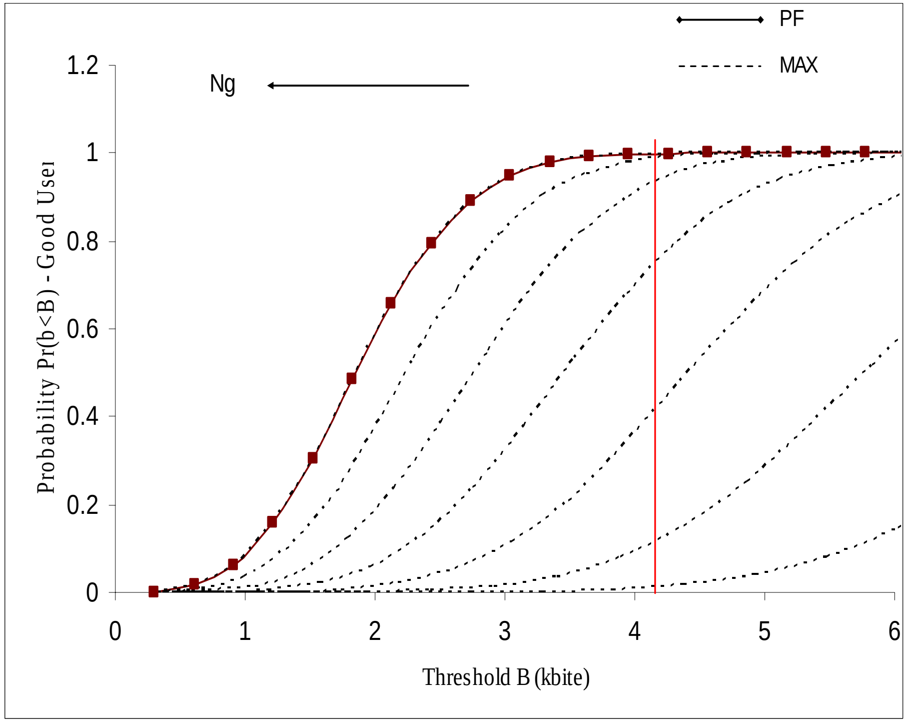
$$\Pr_{Q,i}[b_i < B_i]_{T_{B_i}=40} < 0{,}001$$

Numerical Results (Short-Term Throughput Properties) (III)

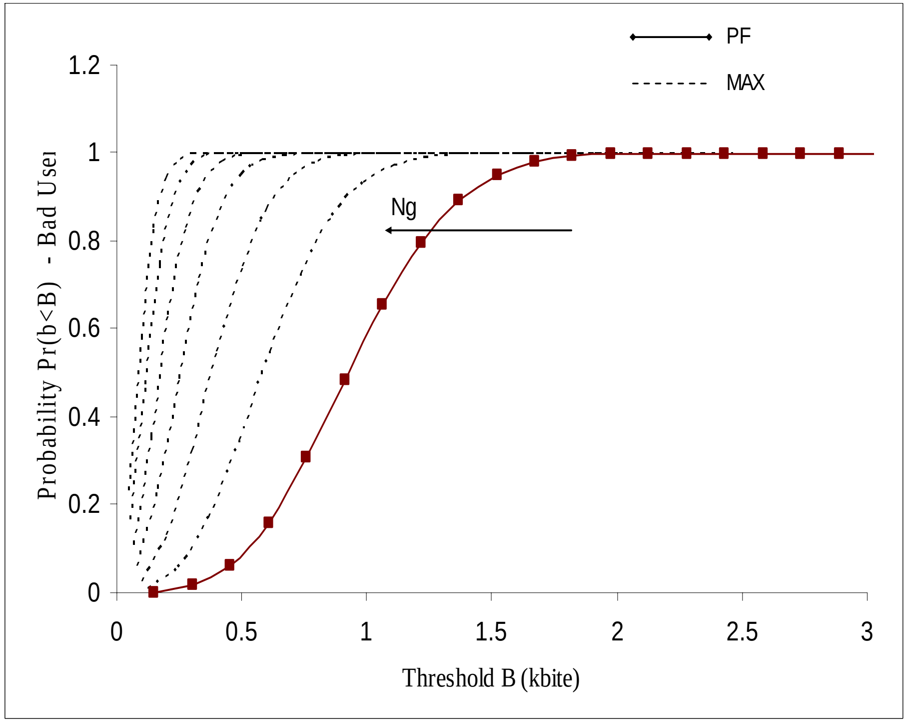
Scheduling Scenario 2

PF vs MAX

Good users case



Bad users case



Vertical red line indicates the considered threshold

$$\Pr_{Q,i}[b_i < B_i]_{T_{B_i}=40} < 0,001$$

Key Observations

- Real-time users' short-term QoS requirements are not satisfied.
  - Even if their number in the system is small

- Under opportunistic scheduling policies with long-term objectives their QoS properties depend on:
  - Their number in the system
  - Their average channels conditions relationship

- "Disappointing" results even if the system's ***throughput performance is optimal due to:***

Inherent nature of opportunistic scheduling which requires only the "best" users to access the system (in consequent slots)

Real-Time Service QoS requirements (frequent slot access)

Future Directions and Conclusions

A myopic scheduling approach of serving only one user per time slot may not be the most appropriate.

1. Under TDMA based opportunistic scheduling techniques (which are satisfying RT users QoS requirements).
   - NRT users “required” high throughput performance is not achieved
   - RT users restrain unnecessarily large amounts of the system’s resources

Real-Time services require Fixed Data Rates

2. Pure TDMA based opportunistic scheduling is **not optimal** under **heterogeneous** wireless environment (Shroff et all)

Common utility based framework

- The idea of allocating at each time slot system's resources to more than one users, seems to be more attractive.

Optimization problems' formulation

Multiple services QoS requirements reflection

} Under a common utility based framework

Users' utility function reflect their degree of satisfaction with respect to their services' QoS requirements fulfillment.

Problem's Formulation (I)

optimization problem – basic constraints

$$\max_{\overline{R},\overline{P}} \sum_{i=1}^{N} U_i^*(\overline{R},\overline{P})$$

$$s.t. \quad \sum_{i=1}^{N} P_i \leq P_{max}$$

$$0 \leq P_i \leq P_{max} \quad i = 1, 2, \ldots, N$$

$$0 \leq R_i \leq R_i^{max} \quad i = 1, 2, \ldots, N$$

Where:

$U_i^*(\overline{R},\overline{P})$ user's *i* utility function

P_{max} base station's maxim achievable downlink transmission power

R_i^{max} user's *i maximum achievable transmission rate*

Problem's Formulation (II)

User's utilities

$$U_i^*(\overline{R},\overline{P},t)=U_{TL,i}^*(\overline{R},\overline{P},t)\cdot U_{SL,i}^*(\overline{R},\overline{P},t)\cdot U_{PE,i}^*(\overline{R},\overline{P},t)$$

- Users' transmission level performance indicators
  - maximum data rates
  - transmission scheme
  - transmission environment
- Users' service level performance indicators
  - Achieved Transmission Rates
  - Average receive throughput
  - Non-Real-Time & Real-Time users' QoS requirements satisfaction
- Pre-emption (real-time observations):
  - Monitoring user's performance within small time intervals
  - Predicting their expected short-term throughput performance

Advantages

- Through users' utility functions we can
  - Efficiently satisfy RT services Short-term Throughput requirements
    - Optimize their satisfaction (Pre-emption)
    - Release the rest of system's excess resources
  - Satisfy NRT users' high throughput demands
  - Trigger scheduler's QoS satisfaction mechanism independently for each type of users.

Greedy Opportunistic Algorithm

$$i_M^*(t) = \arg\max_{i \in S}\{\rho_i(t)\}$$

User's *i* probability to be served at a random instant (normalized fraction of access time):

$$\Pr[i_M^*(t) = i] = \int_0^\infty \prod_{j=1, j \neq i}^{N} (1 - e^{-\rho_i/\bar{\rho}_j}) \frac{e^{-\rho_i/\bar{\rho}_i}}{\bar{\rho}_i} d\rho_i$$

Asymptotic transmission rate:

$$\bar{R}_{M,i} = \frac{W}{\gamma_i} \int_0^\infty \rho_i \prod_{j=1, j \neq i}^{N} (1 - e^{-\rho_i/\bar{\rho}_j}) \frac{e^{-\rho_i/\bar{\rho}_i}}{\bar{\rho}_i} d\rho_i$$

Real Time (Good) User's QoS Properties

$$\Pr[i_M^*(t) = i]_{i \in S_g} = \sum_{k=0}^{N_g - 1} \sum_{t=0}^{N - N_g} \binom{N_g - 1}{k} \binom{N - N_g}{t} \frac{(-1)^{k+t}}{k + at + 1}$$

$$\bar{R}_{M, i \in S_g} = \sum_{k=0}^{N_g - 1} \sum_{t=0}^{N - N_g} \binom{N_g - 1}{k} \binom{N - N_g}{t} \frac{(-1)^{k+t}}{(k + at + 1)^2} \frac{W \bar{\rho}_g}{\gamma_i}$$

Proportional Fair Algorithm

$$i_{PF}^{*}(t) = \arg\max_{i \in S}\{\rho_i(t) / \rho_i^{*}(t)\}$$

Where $\rho_i^{*}(t+1) = (1-\beta)\rho_i^{*}(t) + \beta\rho_i^{'}(t)$ and $\rho_i^{'}(t) = \begin{cases} \rho_i(t) & \text{if user } i \text{ is currently receiving service} \\ 0 & \text{otherwise} \end{cases}$

__Under Rayleigh Fading Channels__, user's *i* asymptotic received throughput

$$\bar{R}_{PF,i} = \frac{G(N)}{N}\frac{W}{\gamma_i}\bar{\rho}_i \quad \text{where} \quad G(N) = \sum_{j=1}^{N} 1/j$$

In *system's steady state the policy turns into:* $i_{PF}^{*}(t) = \arg\max_{i \in S}\{\frac{\rho_i(t)}{G(N)\bar{\rho}_i(t)/N}\}$

User's i probability to be served

at a random instant:

$$\Pr[i_{PF}^{*}(t) = i] = \int_0^{\infty} \prod_{j=1, j\neq i}^{N} F_{\rho_j}\left(\frac{\bar{\rho}_j}{\bar{\rho}_i}\rho_i\right) \frac{e^{-\rho_i/\bar{\rho}_i}}{\bar{\rho}_i} d\rho_i$$

Real Time (Good) User's

$$\Pr[i_{PF}^{*}(t) = i] = \sum_{k=1}^{N}\binom{N}{k}\frac{(-1)^{k+1}}{N} = \frac{1}{\mathrm{N}} \quad \forall \mathrm{i} \in S$$

Relatively Best Algorithm

$$i_{RB}^{*}(t) = \arg\max_{i \in S}\{\rho_i(t) - \bar{\rho}_i(t) \big/ c_i\}$$

User's *i* probability to be served at a random instant (normalized fraction of access time):

$$\Pr[i_{RB}^{*}(t)=i] = \int_{\rho_{0,i}}^{\infty} \prod_{j=1,\, j\neq i}^{N} \left(1 - e^{-\left((c_j/c_i)(\rho_i - \bar{\rho}_i) + \bar{\rho}_j\right)/\bar{\rho}_j}\right) \frac{e^{-\rho_i/\bar{\rho}_i}}{\bar{\rho}_i} d\rho_i \qquad \rho_{0,i} = \max_{i\neq j}\{0, \bar{\rho}_i - c_i/c_j\, \bar{\rho}_j\}$$

Asymptotic transmission rate:

$$\bar{R}_{RB,i} = \frac{W}{\gamma_i} \int_{\rho_{0,i}}^{\infty} \rho_i \prod_{j=1,\, j\neq i}^{N} \left(1 - e^{-\left((c_j/c_i)(\rho_i - \bar{\rho}_i) + \bar{\rho}_j\right)/\bar{\rho}_j}\right) \frac{e^{-\rho_i/\bar{\rho}_i}}{\bar{\rho}_i} d\rho_i$$

Real Time (Good) User's QoS Properties

$$\Pr[i_{RB}^{*}(t)=i]_{i\in S_g} = \sum_{k=0}^{N_g-1} \sum_{t=0}^{N-N_g} \binom{N_g-1}{k} \binom{N-N_g}{t} \frac{(-1)^{k+t} e^{\left(1/\alpha - 1\right)(k+1)}}{k+\alpha t+1}$$

$$\bar{R}_{RB,i\in S_g} = \sum_{k=0}^{N_g-1} \sum_{t=0}^{N-N_g} \binom{N_g-1}{k} \binom{N-N_g}{t} \frac{(-1)^{k+t} e^{(1+k)\left(1/\alpha - 1\right)}}{k+\alpha t+1} \left(1 - \frac{1}{\alpha} + \frac{1}{1+k+\alpha t}\right) \frac{W\bar{\rho}_g}{\gamma_i}$$