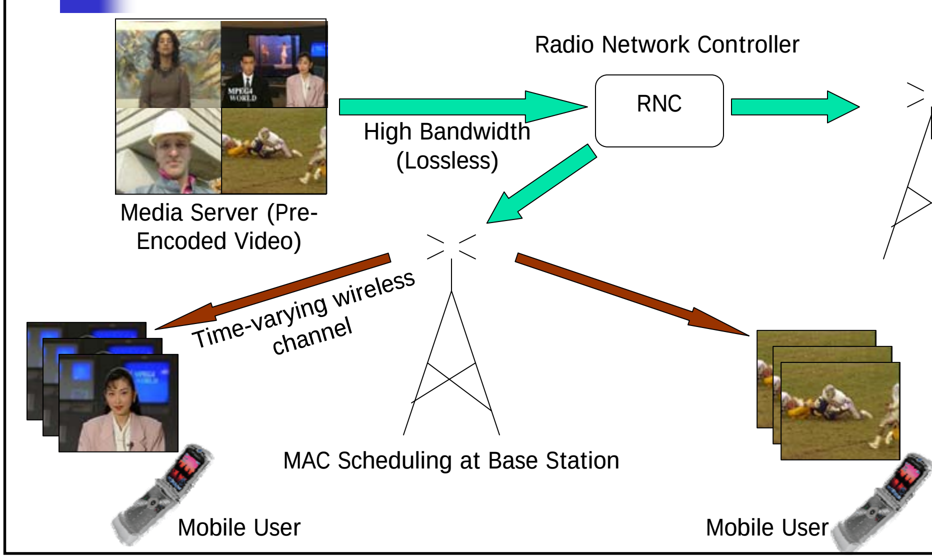
CASE III

Downlink Packet Scheduling and Resource Allocation for Multiuser Video Transmission Over Wireless Networks

Downlink Packet Scheduling and Resource Allocation for Multiuser Video Transmission Over Wireless Networks



Scheduling and Resource Allocation

- General Problem Definition
  - Transmit multiple pre-encoded video sequences
  - To multiple users
  - Over 3G/4G wireless networks
  - Find optimal distribution of resources at base station
  - Such that each user receives a reasonable quality of service
- Limited Resources (limits achievable data rates)
  - Transmission power
  - Number of spreading codes (Bandwidth)
- Quality of Service Measures
  - End-to-end distortion of video sequence
  - Transmission delay (stringent requirement for real-time applications)

Advancements in Wireless Networks

High Speed Downlink Packet Access (HSDPA)

- CDMA/TDM
  - 5 MHz bandwidth, 2ms time slots
- Fast Scheduling at Base Station
  - MAC scheduling at 2ms time slot
  - Dynamically adapt to channel conditions
- Adaptive Modulation and Coding
  - QPSK and 16 QAM
  - Rate 1/3 Turbo codes, variable with *puncturing* and *repetition*
- Fast Hybrid ARQ
  - Chase Combining
  - Incremental Redundancy

IEEE 802.16e (Mobile WiMAX)

- OFDMA/TDM for Multiple Access
  - 5 -10 MHz bandwidth,
- Fast Scheduling at Base Station
  - MAC scheduling at 5ms time slots
- Adaptive Modulation and Coding
  - QPSK, 16 QAM, 64 QAM
  - Variable rate convolutional turbo codes
- Fast Hybrid ARQ

Downlink Packet Scheduling

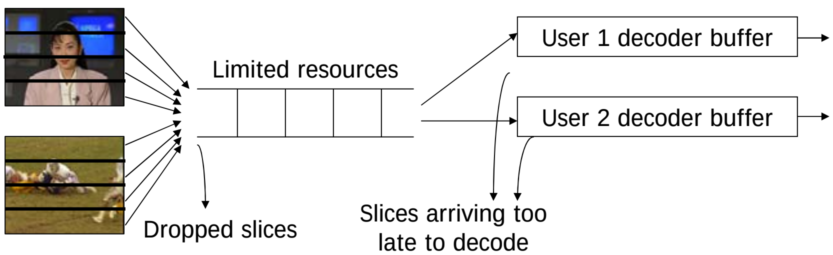
- "Dumb" Method
  - Round Robin
- State of the Art
  - Basic Idea: Allocate resources to users with better channel quality
  - Maximum Throughput Methods
  - Proportionally Fair Methods
    - Fairness criterion based on current average throughput
  - Gradient Based Scheduling
    - Maximize rate to users that will gain the most, subject to channel conditions
- Scheduling for Streaming Video
  - Minimize queue length (delay of head-of-line packet)
  - Current work does not consider rate-distortion trade-offs for individual video packets

Downlink Scheduling- Background

- Channel Independent
  - Round robin scheduling for TDM (e.g., 2G GSM networks)
- State of the Art – Channel Dependent
  - Exploit multiuser diversity
    - Maximize throughput
  - Maintain fairness across users
    - Proportionally fair
    - Gradient-based scheduling techniques
- Scheduling for Streaming Video
  - Queue-length / delay based QoS metrics
  - Current techniques are content independent
- **Proposed Technique**
  - Content-aware packet scheduling for wireless video streaming

Video Transmission Assumptions

- Frames are split into independently decodable slices
- Video will be viewed in real-time (slices from the current frame must be received by the decoder before it finishes decoding the previous frame)
- Achievable data rates may not be sufficient to transmit every slice of every frame to all the users within the real-time constraints



Key Ideas

- Define importance of a packet based on distortion reduction achieved by transmitting the packet
- Order video packets by importance
- Generate distortion-based utility function
- Consider error concealment at the decoder which conceals losses due to dropped packets
  - Simple concealment : Copy MB from the same position in previous frame
  - Complex concealment (described later) adds dependencies between slices/packets

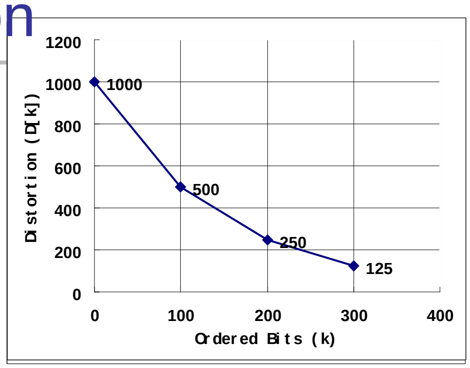
Formulation

- Key Idea
  - Order slices by amount of distortion reduction
  - Define Slice Utility in terms of distortion reduction
- Assumptions
  - Channel states may change by an order of magnitude over one frame's duration
    - Feedback is available every 2ms in HSDPA
    - Optimization performed every time slot (2ms)
  - Simple error concealment
    - Copying MB from the same position in previous frame
    - Complex error concealment adds dependencies between slices
      - complicates slice ordering
    - Will eventually include complex error concealment

Utility Function

- Utility Per User Based on Slice Ordering
  - $D[k]$:= Distortion given k slices are received
  - $D[M]$ = Minimum distortion

    where, M = total # of slices



Distortion vs bits after slice ordering

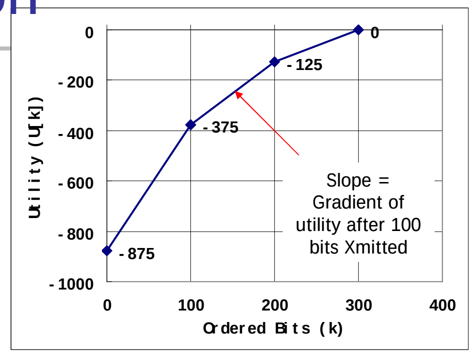
Utility Function

- Utility Per User Based on Slice Ordering
  - $D[k]$:= Distortion given k slices are received
  - $D[M]$ = Minimum distortion

    where, M = total # of slices

$$U[k] = \left(D[M] - D[k]\right)$$

$$u[k] = U'[k] = \frac{D[k] - D[k+1]}{B_{k+1}}$$



Utility vs bits after slice ordering

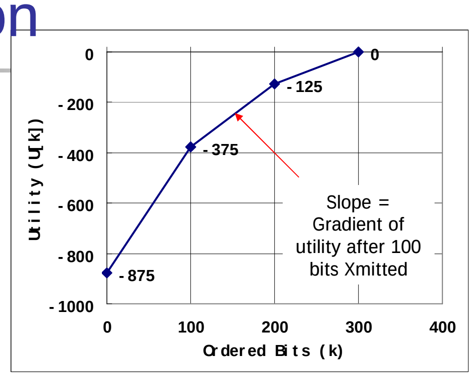
Utility Function

- Utility Per User Based on Slice Ordering
  - $D[k]$:= Distortion given k slices are received
  - $D[M]$ = Minimum distortion

    where, M = total # of slices

$$U[k] = \left(D[M] - D[k]\right)$$

$$u[k] = U'[k] = \frac{D[k] - D[k+1]}{B_{k+1}}$$



Utility vs bits after slice ordering

- Gradient Based Scheduling – Maximizes rates per user weighted by gradient of utility function

$$\max \sum_{i=1}^{K} w_i u_i \cdot r_i$$

K : # of users
r_i : rate per user
w_i : fairness weighting

HSDPA System Constraints

- K users
- Maximum transmission power: → $\sum_{i=1}^{K} p_i \le P$
  - p_i := power per user
- Number n_i of codes per user: → $n_i \le N_i$
- Total number of spreading codes: → $\sum_{i=1}^{K} n_i \le N$
- Achievable rates: $r_i = n_i \Gamma(\xi_i \cdot SINR_i)$
  - $SINR_i := \frac{p_i}{n_i} e_i$
  - Where e_i := channel state (SINR per unit power)
  - Γ := Shannon capacity assuming Gaussian noise channel
  - $\xi_i \in (0,1]$:= gap from capacity

Problem Definition

- Maximize sum of user rates weighted by utility gradients:

$$V^* = \max_{\mathbf{n},\mathbf{p}} \sum_{i=1}^{K} w_i u_i \cdot n_i B \log\left(1 + \frac{p_i e_i}{n_i}\right)$$

subject to : $\sum_{i=1}^{K} n_i \le N$

and $\sum_{i=1}^{K} p_i \le P$

additional constraint : $n_i \le N_i$

P.V. Pahalawatta, et al, **"Content-Aware Resource Allocation and Packet Scheduling for Video Transmission Over Wireless Networks,"** *IEEE Journal on Selected Areas In Communications, Special Issue on Cross-Layer Optimization for Wireless Multimedia Communications*, vol. 25, issue 4, 749-759, May 2007.

Solution

Define Lagrangian

$$L(\mathbf{n},\mathbf{p},\lambda,\mu)=\sum_{i=1}^{K} w_i u_i \cdot n_i B \log\left(1+\frac{p_i e_i}{n_i}\right)+\lambda\left(P-\sum_{i=1}^{K} p_i\right)+\mu\left(N-\sum_{i=1}^{K} n_i\right)$$

Define Dual Function

$$L(\lambda,\mu)=\max_{(\mathbf{n},\mathbf{p})\in C'} L(\mathbf{n},\mathbf{p},\lambda,\mu)$$

Solve Dual Problem

$$L^{*}=\min_{(\lambda,\mu)\ge 0} L(\lambda,\mu)$$

$$L^{*}=\min_{\lambda\ge 0} L(\lambda) \quad \text{where,} \quad L(\lambda)=\min_{\mu\ge 0} L(\lambda,\mu)$$

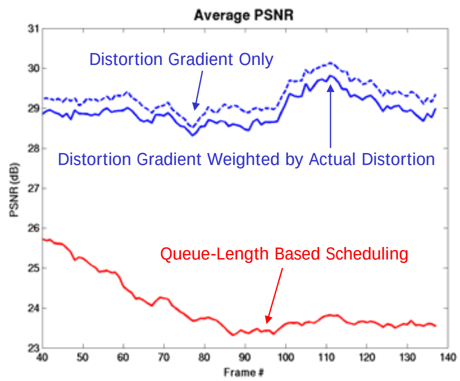
Can find λ using bisection search

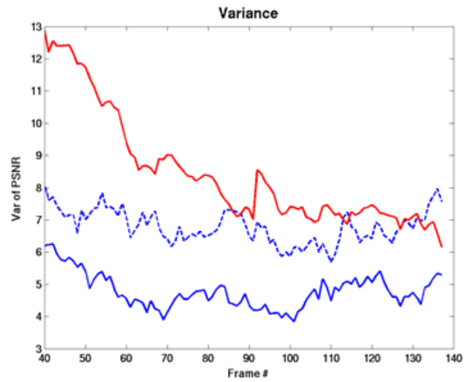
R. Agrawal, V. Subramanian, and R. Berry, "Joint Scheduling and Resource Allocation in CDMA Systems," *IEEE Trans. on Information Theory*

Simulation I : Similar Fading Stats

- HSDPA System
  - # of Users (K) = 12
  - Total # of spreading codes (N) = 15
  - Max codes per user (N_i) = 5
  - Total Power (P) = 9.9W
  - Max SINR = 1.59 dB
  - Simulated uncorrelated Raleigh fading channels, with same average channel conditions, for each user
- Video Source
  - H.264 encoded at 256kbps
  - QCIF (176x144) {foreman, mother and daughter, carphone, news, silent, news, 0…139, 150…289}
  - 11 MBs per slice

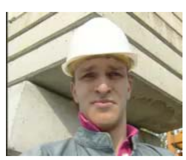
Simulation I Results



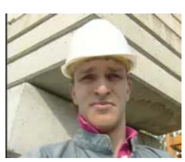


Demo 1

- 6 users
- P = 7W



Weighted Distortion Gradient Metric

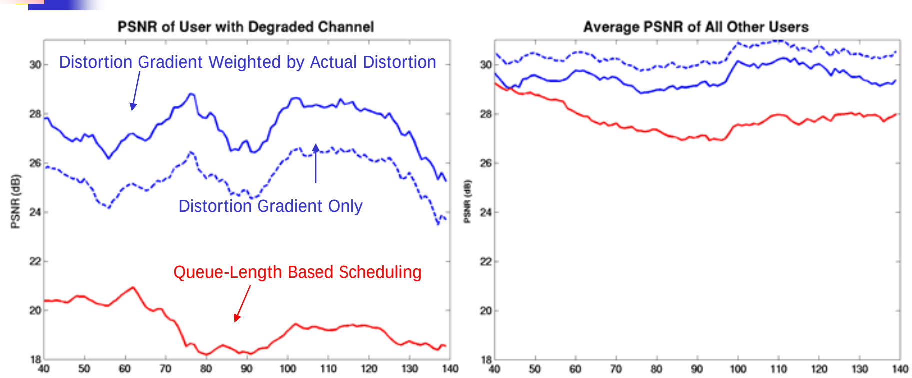


Queue-based Utility Metric

Simulation II : One Degraded User

- HSDPA System
  - # of Users (K) = 10
  - Total # of spreading codes (N) = 15
  - Max codes per user (N_i) = 5
  - Total Power (P) = 9.9W
  - Max SINR = 1.59
  - Simulated uncorrelated Raleigh fading channels, with one user having lower average channel quality
- Video Source
  - H.264 encoded at 256kbps
  - QCIF (176x144) {foreman, mother and daughter, carphone, news, silent, 0...139, 150...289}
  - 11 MBs per slice

Simulation II Results



- The weighted distortion gradient tends to be fair to all users while not significantly sacrificing overall quality

Demo 2

Reception Quality for User with Degraded Channel

Distortion-based Utility

Queue length-based Utility

Additional Considerations

- Complex error concealment
  - Incremental gain not additive
  - Myopic solution
  - Smoothed utility gradient
- Use of a scalable bitstream
- Random Packet Losses

E. Maani, et al, **"Resource Allocation for Downlink Multiuser Video Transmission Over Wireless Lossy Networks,"** *IEEE Trans. Image Processing, Sept. 2008*.

Simulations

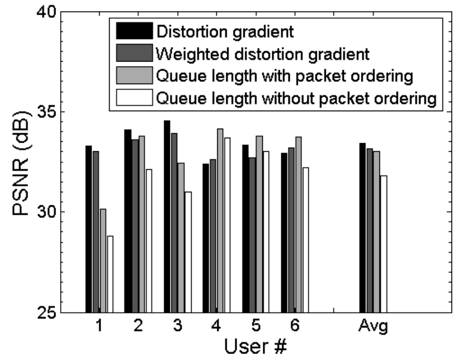
- Channel Statistics
  - Realistic HSDPA channel traces obtained from simulators at Motorola assuming users are 0.8 km from base station traveling at 30 kmph
  - Results averaged over 5 channel realizations, 100 frames per sequence
- System Parameters
  - # of users (K) = 6
  - Total # of spreading codes (N) = 15
  - Max codes per user (N_i) = 5
  - Total power (P) = 10W
  - Max SINR ($\hat{s}_i$ = 1.8dB)
- Video Source (6 sequences)
  - H.264 encoded with average PSNR of 35dB for each sequence
  - QCIF (176x144) {foreman, mother and daughter, carphone, news, silent}
  - 11 MBs per slice

Results: Comparison of Metrics

$$\max_{(\mathbf{n},\mathbf{p})\in C} \sum_{i=1}^{K} w_i u_i \cdot r_i(\mathbf{n},\mathbf{p})$$

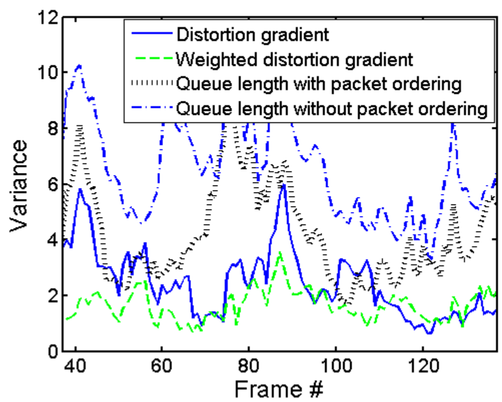
- Distortion Gradient
  - Proposed technique (w_i = 1)
- Weighted Distortion Gradient
  - Proposed technique (w_i = current distortion of sequence)
- Queue Length
  - Conventional content-independent technique
  - Replace $w_i u_i$ with # of bits remaining in frame for user i
  - Equivalent to utility function of form ($-Q_i^2$), found in conventional buffer management techniques
- Ordered Queue
  - Proposed packet ordering technique without content-dependent resource allocation
  - Same as Q but first, order the slices in each frame by distortion gradient

Results – Average PSNR



1: Foreman
2: Mother and daughter
3: Car phone
4: News
5: Silent
6: Hall Monitor

Results – Variance of PSNR



- The ordered queue method performs just as well as the distortion gradient methods on average but causes higher variability in quality across users

Demo

WDG 33.1dB | Ordered Queue 30.0dB | Queue 28.6dB

32.6dB | 34.0dB | 33.6dB

Formulation with Random Pkt Losses

- Order Video Packets by Importance
  - Based on expected reduction in distortion achieved by transmitting the packet
- Allocate Resources to Packets/Users with Larger Expected Distortion Gradients
- Consider Error Concealment at the Decoder
- Channel Model
  - Packet loss rates can be modeled as a function of estimated channel quality and allocated resources

Expected Distortion Calculation

$$E[D_n] = \frac{1}{N}\sum_{i=1}^{N} E\left\{\left(f_n^i - \tilde{f}_n^i\right)^2\right\}$$

- Expected distortion of frame n → $E[D_n]$
- Number of pixels → N
- Actual value of i th pixel → f_n^i (Known)
- Decoded value of i th pixel → $\tilde{f}_n^i$ (Random due to channel loss)

- Need first and second order moments of decoded pixel value
  - Recursive Formulation (ROPE)

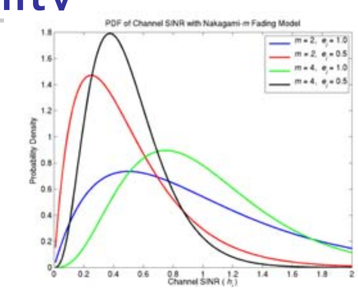
$$E\left\{\tilde{f}_n^i\right\} = \sum_{\alpha} p(\alpha)\left(E\left\{\tilde{f}_{n-1}^{i(\alpha)}\right\} + residual(\alpha)\right)$$

- Probability of event α → $p(\alpha)$
- Expected value of reference pixel → $E\left\{\tilde{f}_{n-1}^{i(\alpha)}\right\}$

Outage Probability

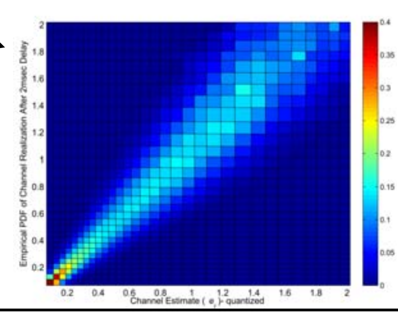
- Nakagami-m Fading Model
  - Mean at channel estimate (e_i)
  - Cumulative cdf at h_i

$$\phi_{e_i}(h_i) = \frac{\gamma\left(m, \frac{mh_i}{e_i}\right)}{\Gamma(m)}$$

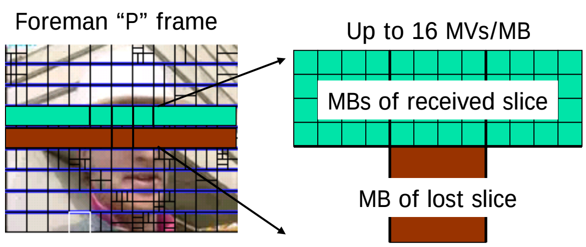


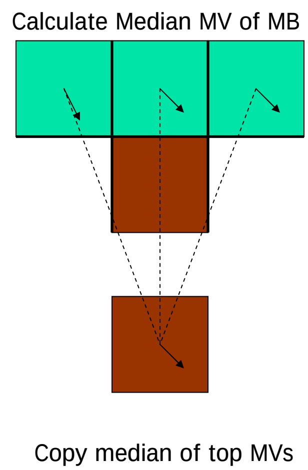
- Use Empirical Distribution
- Packet Loss Rate

$$\varepsilon_i = \phi_{e_i}\left(\frac{n_i}{p_i} 2^{\frac{r_i}{n_i B}-1}\right)$$



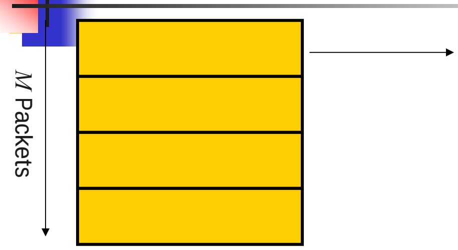
Error Concealment Scheme





- Use most likely MV from top MBs
  - If top MBs are available, use median MV of top MBs, else use zero MV
- Error concealment only depends on top slice and previous frame

Packet Ordering Within User



$$\begin{Bmatrix} \mu_m = 0 \\ \mu_m = 1 \end{Bmatrix},$$

Source Modes: SEND or SKIP

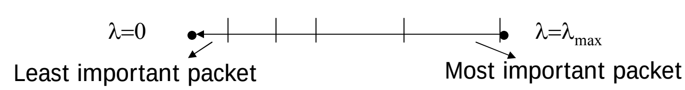
$$\varepsilon_m$$

Packet Loss Probability

$$L(\boldsymbol{\mu},\boldsymbol{\varepsilon};\lambda) = \sum_{m=1}^{M} E\{D_m(\mu_m,\varepsilon_m,\mu_{m-1},\varepsilon_{m-1})\} + \lambda R(\mu_m)$$

$$\boldsymbol{\mu}^*(\lambda,\boldsymbol{\varepsilon}) = \arg\min_{\boldsymbol{\mu}\in\{0,1\}^M} L(\boldsymbol{\mu},\boldsymbol{\varepsilon};\lambda) \longrightarrow$$

Complexity of 4 x M



Resource Allocation Across Users

At each time slot,

$$\min_{\mathbf{n},\mathbf{p},\mathbf{r}} \sum_{i=1}^{K} ED_i\{r_i, \varepsilon_i(n_i, p_i, r_i, e_i)\}$$

subject to,

$$0 \le \sum_{i=1}^{K} p_i \le P, \quad 0 \le \sum_{i=1}^{K} n_i \le N, \quad 0 \le n_i \le N_i, \quad 0 \le r_i \le R_{\max_i}$$

K : Number of users; ε_i : Probability of loss; e_i : Channel estimate

ED_i : Expected distortion of user i's frame

- Difficult to solve because no simple analytical form exists for ED_i

Solution – Step I

- First Step – Fix Probability of Loss
  - Can write r_i as function of parameters n_i, p_i

$$r_i = n_i B \log_2\left(1 + \frac{p_i \phi_{e_i}^{-1}(\varepsilon_i)}{n_i}\right)$$

→ Inverse cumulative cdf of channel SINR conditioned on estimate e_i

- Gradient-Based Solution for Fixed ε_i

$$\max_{\mathbf{n},\mathbf{p}} \sum_{i=1}^{K} -\frac{\partial ED_i}{\partial r_i} \cdot r_i(n_i, p_i)$$

Maximize sum of rates weighted by expected distortion gradients

Solution – Step II

- Second Step - Fix Resource Allocation (n_i^*, p_i^*)

$$\varepsilon_i = \phi_{e_i}\left(\frac{n_i^*}{p_i^*} 2^{\frac{r_i}{n_i^* B} - 1}\right)$$

- One-Dimensional Search Over Individual User Rates

$$\min_{0 \le r_i \le R_{\max_i}} ED_i\{r_i\} \qquad \forall i$$

  - Can solve using fast line search method

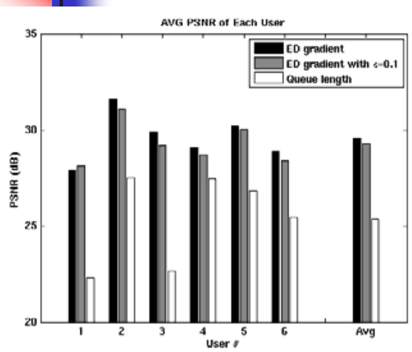
Implementation Issues

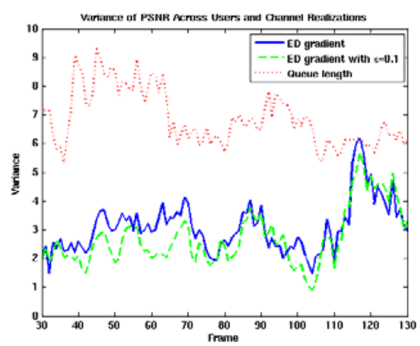
- Acknowledgements
  - Assume ACK/NACK feedback is available for each transmission (MAC layer feedback delay 10msec)
- Packet Fragmentation
  - Application layer packets (i.e., video slices) may be too large to transmit within one timeslot and must be fragmented
  - Assume all fragments of a slice must be received in the correct order, in order to decode the slice
  - Loss probability of a fragmented slice
    - Pr(Any transmitted and not yet ACK/NACK'ed fragment is lost)

Simulations: Comparison of Metrics

- Distortion Gradient
  - Proposed technique
- Distortion Gradient with Fixed Probability of Loss
  - Proposed ordering technique but fix probability of loss
- Queue Length
  - Conventional content-independent technique (no packet ordering)
  - Fix probability of loss, and replace gradient of expected distortion with queue length

Results

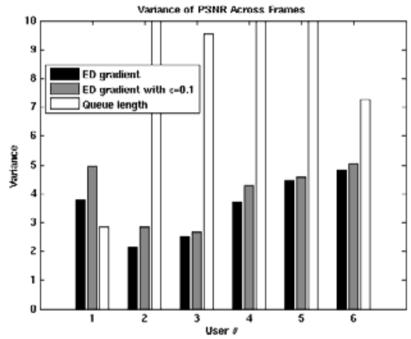




1: Foreman
2: Mother and daughter
3: Car phone
4: News
5: Hall Monitor
6: Silent

- Packet ordering with expected distortion gradients performs significantly better than queue length method in terms of average and variance of PSNR

Results



1: Foreman
2: Mother and daughter
3: Car phone
4: News
5: Hall Monitor
6: Silent

- Variance of PSNR across frames within each user is higher when probability of loss is fixed

Demo

Distortion gradient	Distortion gradient with fixed loss	Queue length with fixed loss
30.7 dB	30.0 dB	23.7 dB

30.3 dB	29.7 dB	28.6 dB