# Modeling Call Degradation and Admission Policies for Multimedia Traffic in Wireless Networks

**Sajal K. Das and Sanjoy K. Sen**

Center for Research in Wireless Computing (CReW)
Department of Computer Sciences
University of North Texas
Denton, TX 76203-1366

**Kalyan Basu**

Wireless Networks
NORTEL
Richardson, TX 75083

## What is QoS?

- Guaranteeing desirable/predictable service for multimedia traffic from network resources such as processor time, buffer space, network bandwidth etc.
- QoS is specified by $\langle bandwidth, delay, reliability \rangle$.

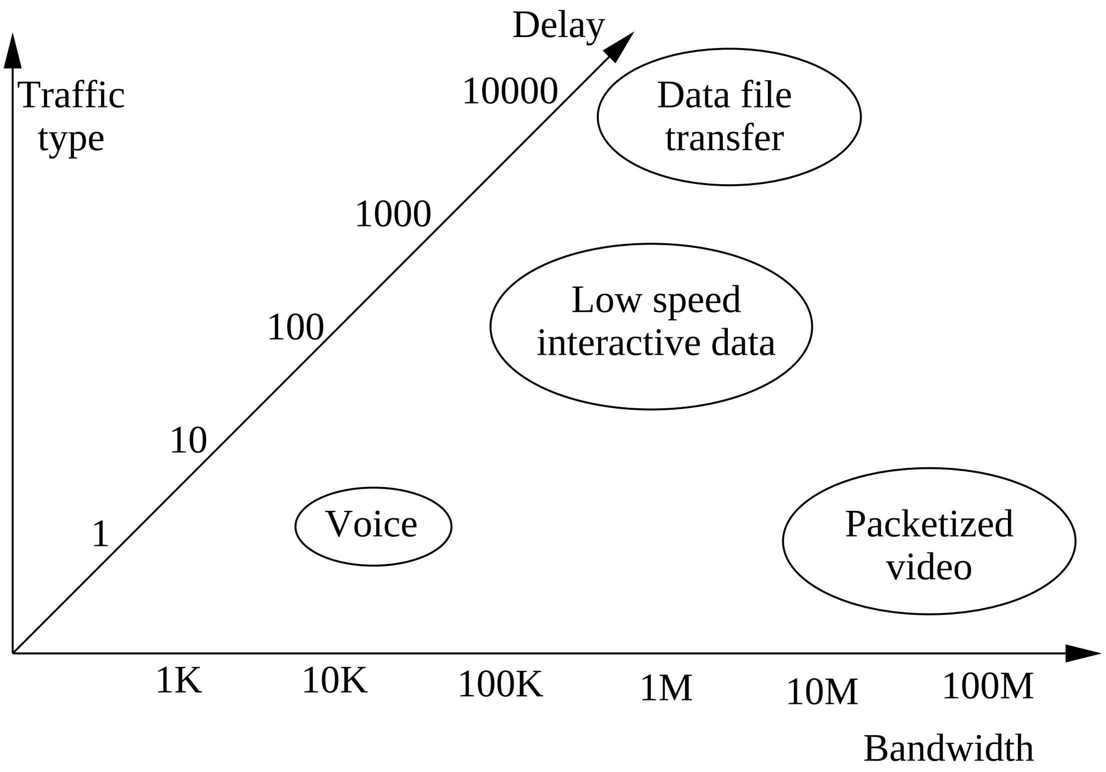
## Difference between Wireless and Wireline networks

- *link characteristics:* low bandwidth and high error rates
- *mobility:* changing UNI (User Network Interface)

## Motivations:

• Limited bandwidth spectrum is the main issue in designing multiple access techniques.

• Improve system *capacity* and maintain QoS in the third generation wireless multimedia networks.




**Multimedia traffic:** Voice, Video, Data

**Real-time traffic:** Voice, video

**Non-real-time traffic:** TCP/IP packets, text data

Existing wireless data networks – CDPD, GPRS, HSCSD – utilize unused voice capacity to support non-real-time data.

### Oliveira, Kim and Suda [1996]

- For real-time traffic, users to be allocated request bandwidth in the current cell and all of its neighbors.

- Call not admitted if bandwidth can't be reserved.

- **Drawbacks:**

  — Redundant reservation of bandwidth.

  — Poor call admission in overloaded systems.

### Sur and Singh [1996]

- Multicasts incoming messages to the group of surrounding cells.

  — QoS parameters: *graceful degradation of service* and *guarantee of seamless service.*

  — A new transport sublayer proposed, implementing a selective data discard function with user-supplied *loss profiles*.

- **Drawback:**

  Redundant multicasting consumes resources like network bandwidth and buffer.

**Acampora and Naghshineh [1995]**

- A *virtual connection tree* (cluster of base stations) is used for call routing, call admission and control.
- A set of virtual connections reserved for each admitted user.
- Call admission based on adaptive resource sharing among real-time and non-real-time traffic, the former having preemptive priority over the latter.
- Simple analytical model proposed to capture various performance metrics like call blocking and dropping probabilities.
- **Drawbacks:**

  — Over-reserves virtual connections.

  — Leftover bandwidth equally shared among non-real-time traffic.

  — Buffering of non-real-time traffic not considered.

# Our Past Work

## Das, Sen and Jayaram [ICUPC 1997]

- Efficient *call admission* and *call control* algorithms for real-time and non-real-time multimedia traffic.

- *Optimistic* bandwidth reservation in certain *destination cells* for only *departing* users. Bandwidth reservation through *channel borrowing.*

- Real-time call terminations lead to unutilized bandwidth "holes" in the spectrum. Moving the "holes" on one side of the spectrum to provide contiguous free bandwidth is called *bandwidth compaction.*

- Leads to efficient spectrum utilization in systems (e.g. AMPS) requiring allocation of contiguous channels for high bandwidth multimedia traffic.

  Simulation shows 24% improvement of bandwidth utilization.

- **Drawbacks:**

  — Buffering of non-real-time traffic not considered.

  — No analysis incorporating multi-rate traffic in a mixture of real-time and non-real-time traffic.

# Our Current Work

- Develop a framework for *bandwidth degradation.*

  Trade-off between QoS degradation and the *net revenue earned.*

- Admission priority of certain types of calls over others.

  Buffer size estimation for non-real-time traffic using a Markov modulated Poisson process (MMPP) model.

- Detailed simulation experiments to validate our schemes and models.

  (Compression techniques to improve bandwidth utilization not considered here.)

# Graceful Degradation



- Wireless capacity can be increased by degradation of some or all of the existing services.

- Wireless systems use *channels*, e.g. $\langle$time slot, carrier frequency$\rangle$ tuple in TDMA systems or a fixed block of radio frequency bandwidth in FDMA systems.

- Multiple channels are required to service multimedia traffic like video, text and images. HSCSD can use upto 8 GSM 9.6 Kbps channels to transmit high bandwidth video.

- *Degradation* implies reducing the bandwidth allocated to ongoing communications to allow more new calls.

# Modeling Bandwidth Degradation **Real-Time traffic**

- $K$ classes of user traffic

  Class $i$ call uses $i$ channels under normal operation.

- *Degraded mode:* class $i$ call uses $i-1$ channels, releasing one to a common pool.

- Trade-off between *total carried traffic* (positive revenue) and *channel degradation* (negative revenue).

- *Bandwidth degradation policy*:

$$D_{bw} = \{y_2, y_3, \ldots, y_K\}$$

  where $y_i$ of the ongoing Class $i$ calls are to be degraded, $2 \leq i \leq K$.

- Number of Class $i$ calls without degradation

$$X_i = \frac{\alpha_i C}{\Sigma_{j=1}^{K} j\alpha_j} \tag{1}$$

- Let

  $t_{c,i}(X_i; y_2, y_3, \ldots, y_K) \equiv$ total number of admitted Class $i$ calls in the system

  $C_{t,i} \equiv$ revenue generated by each admitted call

  $t_{d,i}(y_2, y_3, \ldots, y_K) \equiv$ number of degraded Class $i$ calls

- **Revenue earned** due to carried traffic

$$\Phi_t = \sum_{i=1}^{K} C_{t,i} t_{c,i}(X_i; y_2, y_3, \ldots, y_K) \tag{2}$$

- **Revenue loss** due to degradation of calls

$$\Phi_d = \sum_{i=2}^{K} \frac{C_{d,i}}{2}(t_{d,i}(y_2, y_3, \ldots, y_K))^2 \tag{3}$$

- **Effective (or Net) revenue** earned by the system from the bandwidth degradation policy $D_{bw}$

$$\Phi = \Phi_t - \Phi_d \tag{4}$$

- An **optimal policy** $D^*_{bw}$ maximizes $\Phi$.

- Using $\Sigma_{j=2}^{K} y_j$ released channels, the number of admitted Class $i$ calls

  is $\frac{\alpha_i \sum_{j=2}^{K} y_j}{\alpha_1 + \sum_{j=2}^{K}(j-1)\alpha_j}$.

$$t_{c,i}(X_i; y_2, y_3, \ldots, y_K) = X_i + \frac{\alpha_i \Sigma_{j=2}^{K} y_j}{\alpha_1 + \Sigma_{j=2}^{K}(j-1)\alpha_j} \tag{5}$$

$$t_{d,i}(y_2, y_3, \ldots, y_K) = y_i + \frac{\alpha_i \Sigma_{j=2}^{K} y_j}{\alpha_1 + \Sigma_{j=2}^{K}(j-1)\alpha_j} \tag{6}$$

- Complete form of revenue cost function

$$\Phi = \sum_{i=1}^{K} \{X_i + \frac{\alpha_i \Sigma_{j=2}^{K} y_j}{\alpha_1 + \Sigma_{j=2}^{K}(j-1)\alpha_j}\} C_{t,i} - \sum_{i=2}^{K} \{y_i + \frac{\alpha_i \Sigma_{j=2}^{K} y_j}{\alpha_1 + \Sigma_{j=2}^{K}(j-1)\alpha_j}\}^2 \frac{C_{d,i}}{2} \tag{7}$$

  where $y_i \leq X_i$ for $2 \leq i \leq K$.

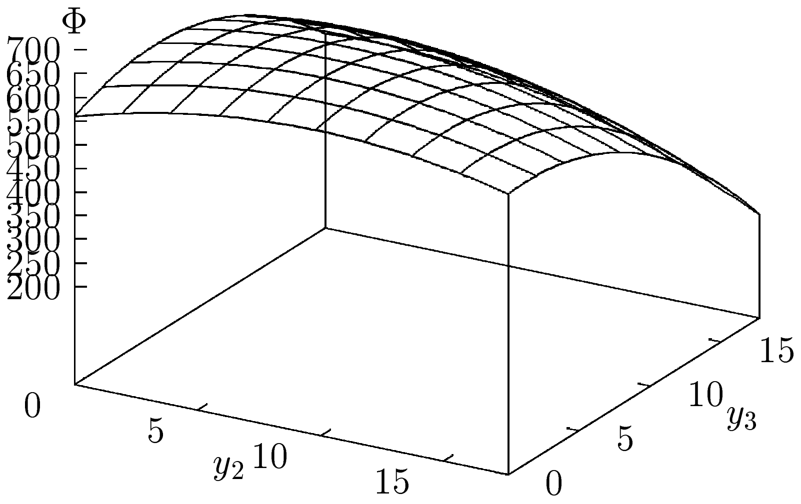
- $(K-1)$-dimensional constrained optimization problem. Can be solved by Langrange's multiplier method.

- The maxima is the surface point $(y_2, y_3)$ given as

$$y_3 = \frac{(\alpha_1 C_{t,1} + \alpha_2 C_{t,2} + \alpha_3 C_{t,3})/A}{\{C_{d,2}(1+\frac{\alpha_2}{A})^2 + C_{d,3}(\frac{\alpha_3}{A})^2\}\Gamma + \{\frac{\alpha_2 C_{d,2}}{A}(1+\frac{\alpha_2}{A}) + \frac{\alpha_3 C_{d,3}}{A}(1+\frac{\alpha_3}{A})\}} \tag{8}$$

$$A = \alpha_1 + \alpha_2 + \alpha_3, \quad y_2 = \Gamma y_3, \quad \Gamma = \frac{C_{d,3}(1+\frac{\alpha_3}{A}) - C_{d,2}\frac{\alpha_2}{A}}{C_{d,2}(1+\frac{\alpha_2}{A}) - C_{d,3}\frac{\alpha_3}{A}}. \tag{9}$$

- Plot of $\Phi$ for $C = 50, \alpha_1 = 0.5, \alpha_2 = 0.3, \alpha_3 = 0.2$, and $C_{d,2} = 1$ and $C_{d,3} = 2$ units




- Maxima is $(8, 3)$. Hence, degradation of 8 out of 9 Class 2 calls and 3 out of 6 Class 3 calls maximizes earned revenue.

# Call Degradation & Admission Control

- Assume a batch arrival scenario for the users, represented as a demand vector $\mathcal{D} = \{n_1, n_2, \cdots, n_K\}$ for the $K$ classes of traffic.

- Define Admission vector $\mathcal{A} = \{a_1, a_2, \cdots, a_K\}$
  $a_i = \#$ of admitted calls of Class $i$.

- Degraded admission vector $\Gamma = \{\gamma_2, \gamma_3, \cdots, \gamma_K\}$
  $\gamma_i =$ fraction of incoming calls of Class $i$ admitted in degraded mode.

- **Problem :**

  Compute a solution vector $\mathcal{S} = \{\mathcal{A}, D_{bw}, \Gamma\}$ maximizing the reward function

$$\Phi = \sum_{i=1}^{K} a_i C_{t,i} - \sum_{i=2}^{K} (y_i + \gamma_i a_i)^2 \frac{C_{d,i}}{2} \tag{10}$$

  satisfying the constraints

$$a_1 + \Sigma_{i=2}^{K}\{(i-1)\gamma_i a_i + i(1-\gamma_i)a_i\} \leq C - \Sigma_{i=1}^{K} iX_i + \Sigma_{i=2}^{K} y_i$$

$$y_i \leq X_i, \quad a_i \leq n_i, \quad \gamma_i \leq 1$$

- A $(3K-2)$-dimensional constrained optimization problem.
  Solution vector $\mathcal{S}$ is a combination of *allocation* and *degradation* policies.

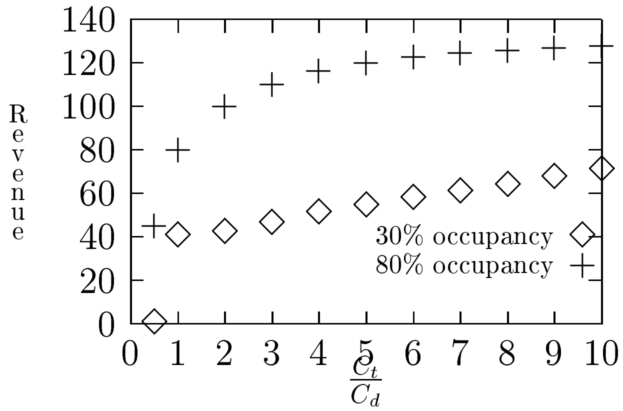
- 5 classes of user traffic considered.
- # of full-duplex radio channels supported by any cell is $C = 50$.
- A static call mix of various types of traffic assumed:

  50% voice calls

  20% data calls

  Remaining calls equally distributed among other classes.
- Experiment with various values of the ratio $\frac{C_t}{C_d}$ for different classes.

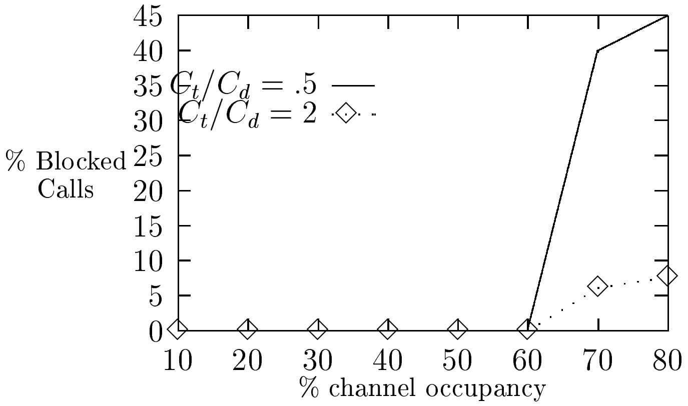
**Net revenue earned with various ratios of $\frac{C_t}{C_d}$**




- $\frac{C_t}{C_d}$ ratio varied from 0.5 to 10. Net revenue earned is very sensitive to this ratio for $C_t \leq C_d$.

- Revenue is higher for a highly loaded system with 80% occupancy than a moderate 30% loaded system due to the static call mix assumption.

## Percentage of blocked calls with various channel occupancy




- Upto 60% channel occupancy, call blockade is insignificant.

- About 45% call blockade for $\frac{C_t}{C_d} = 0.5$, which decreases to 7% when $\frac{C_t}{C_d} = 2$.

- Number of channels in base station is $C$.
  Both real-time and non-real-time traffic use a single channel.

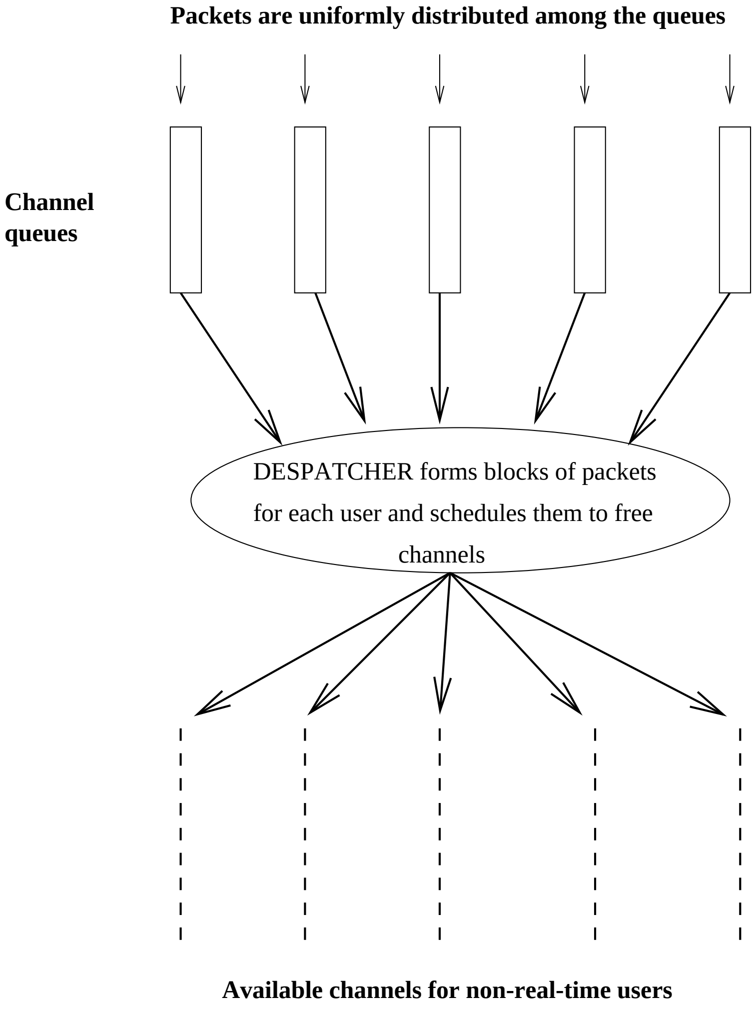
- Real-time packets are never buffered and always allocated a channel if one is available or being used by non-real-time traffic.

- Each channel has a queue used only by the arriving non-real-time packets being serviced by that channel.
  If the channel is servicing a real-time call, its queue remains empty.

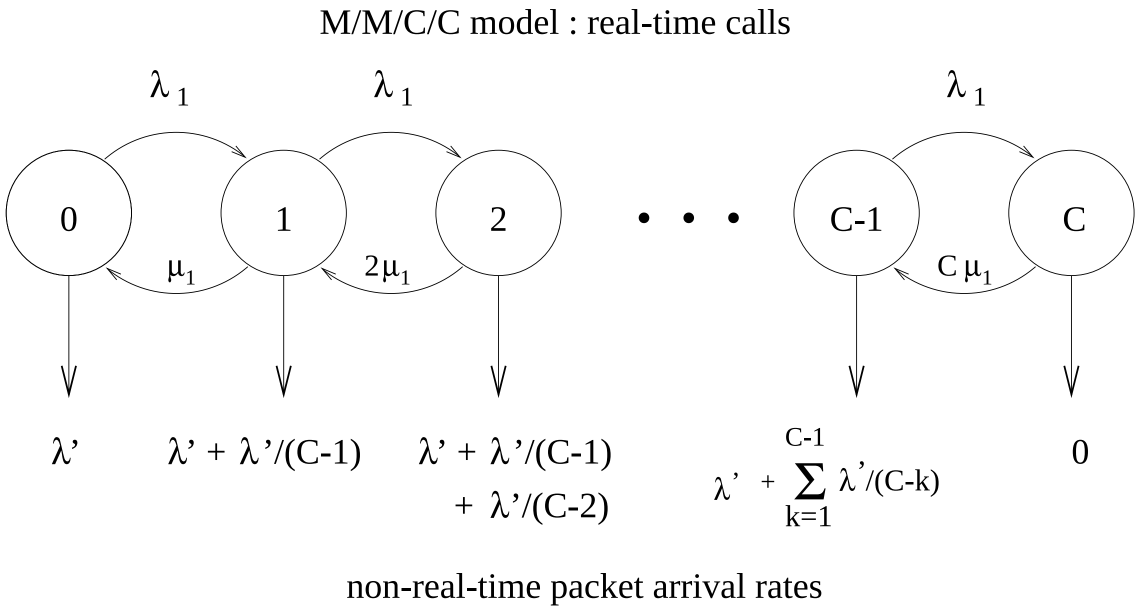
- When a non-real-time call is preempted from a channel, the system distributes the incoming packets with equal probability among the queues servicing other ongoing calls of similar type.

- QoS metric for non-real-time traffic is the *average queue length.*

# An Implementation of Traffic Degradation



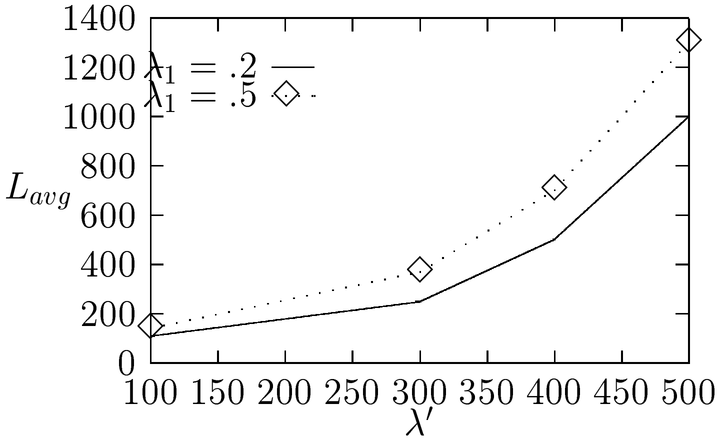

- A *dispatcher process* constructs packet blocks arriving in the queues for each user and dispatches them to the channel currently being used by the user.
- Synchronous dispatch of blocks.
- User's receipt of data block can be arbitrarily delayed due to absence of sufficient data or to provide a fair service to all users. User is instructed to disconnect temporarily to allow other waiting non-real-time users.




- Call arrival and departure processes for real-time traffic are Poisson with mean rates $\lambda_1$ and $\mu_1$. Follows $M/M/C/C$ queuing discipline.
- Packet arrival and departure rates for non-real-time traffic are Poisson distributed with mean $\lambda'$ and $\mu'$, respectively.
- For $i$ real-time calls, the packet arrival rate in an active queue is

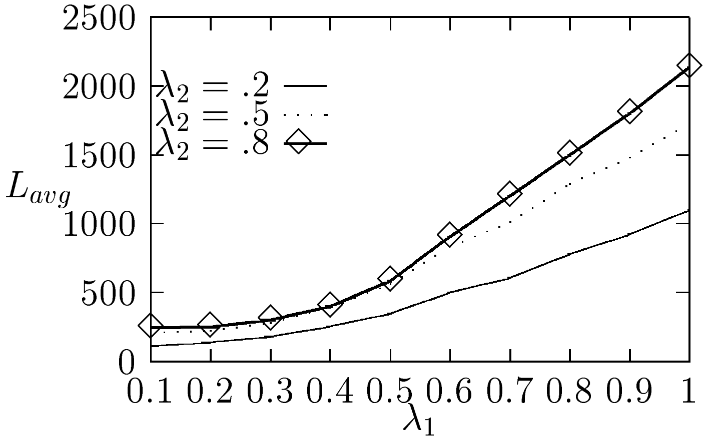
$$\lambda' + \sum_{k=1}^{i} \frac{\lambda'}{C-k},$$

  leading to a *Markov Modulated Poisson Process* (MMPP).
- Two dimensional state space $\langle i, j \rangle$, where $i \equiv$ number of non-real-time packets in a queue and $j \equiv$ number of ongoing real-time calls.
- Solution using **matrix-geometric** method.

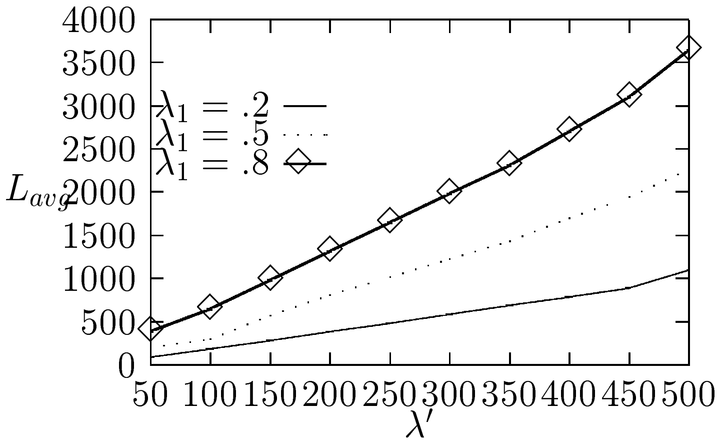



**Average queue length, $L_{avg}$, versus $\lambda'$ (non-real-time packet arrival rate) with $\mu_1 = 0.005$ and $C = 50$**

**Average queue length vs. real-time call arrival rate**




**Average queue length vs. non-real-time packet arrival rate**



## Conclusions

- Bandwidth degradation and call admission strategies are proposed based on modeling QoS degradation of real-time and non-real-time traffic in a multimedia wireless network.

- Two orthogonal QoS parameters are identified, – *total carried traffic* and *bandwidth degradation* – and a revenue function formulated.

- Simulation results demonstrate that the system performance is highly sensitive to proper choice of the cost of degradation and the revenue earned per admitted call.

- A novel channel sharing scheme for non-real-time users is proposed and analyzed using a Markov modulated Poisson process (MMPP) based queuing model.

- **Future Work :**

  Study the impact of header *compression* and *coding* techniques (developed to utilize the limited bandwidth spectrum more efficiently) on our algorithm and its effect on the $\frac{C_t}{C_d}$ ratio.

- **CReW Home Page URL**

  **http://www.cs.unt.edu/home/joy/crew.html**