Project	**IEEE 802.20 Working Group on Mobile Broadband Wireless Access** <http://grouper.ieee.org/groups/802/20/>
Title	**Equivalent Circuit Rate: A User-based Performance Metric for Shared Packet Access Networks**
Date Submitted	**2003-07-16**
Source(s)	N. K. Shankaranarayanan AT&T Labs-Research 200 Laurel Avenue South Middletown, NJ 07748 V: (732) 420-9029 email: shankar@research.att.com Vincent Park Flarion Technologies 135 Route 202/206 Bedminster, NJ V: (908) 997-2000 email: park@flarion.com
Re:	802.20 Session#3
Abstract	This document describes a user-based performance metric called Equivalent Circuit Rate that offers some advantages in evaluating MBWA systems. We present an analytical model, simulation results, and issues involved in using this for MBWA systems.
Purpose	For information.
Notice	This document has been prepared to assist the IEEE 802.20 Working Group. It is offered as a basis for discussion and is not binding on the contributing individual(s) or organization(s). The material in this document is subject to change in form and content after further study. The contributor(s) reserve(s) the right to add, amend or withdraw material contained herein.
Release	The contributor grants a free, irrevocable license to the IEEE to incorporate material contained in this contribution, and any modifications thereof, in the creation of an IEEE Standards publication; to copyright in the IEEE's name any IEEE Standards publication even though it may include portions of this contribution; and at the IEEE's sole discretion to permit others to reproduce in whole or in part the resulting IEEE Standards publication. The contributor also acknowledges and accepts that this contribution may be made public by IEEE 802.20.
Patent Policy	The contributor is familiar with IEEE patent policy, as outlined in Section 6.3 of the IEEE-SA Standards Board Operations Manual <http://standards.ieee.org/guides/opman/sect6.html#6.3> and in *Understanding Patent Issues During IEEE Standards Development* <http://standards.ieee.org/board/pat/guide.html>.

Equivalent Circuit Rate: A User-based Performance Metric for Shared Packet Access Networks

IEEE 802.20 Standards Meeting

July 20-25, 2003

N. K. Shankaranarayanan
shankar@research.att.com

AT&T Labs-Research

Middletown, NJ

Vincent Park
park@flarion.com

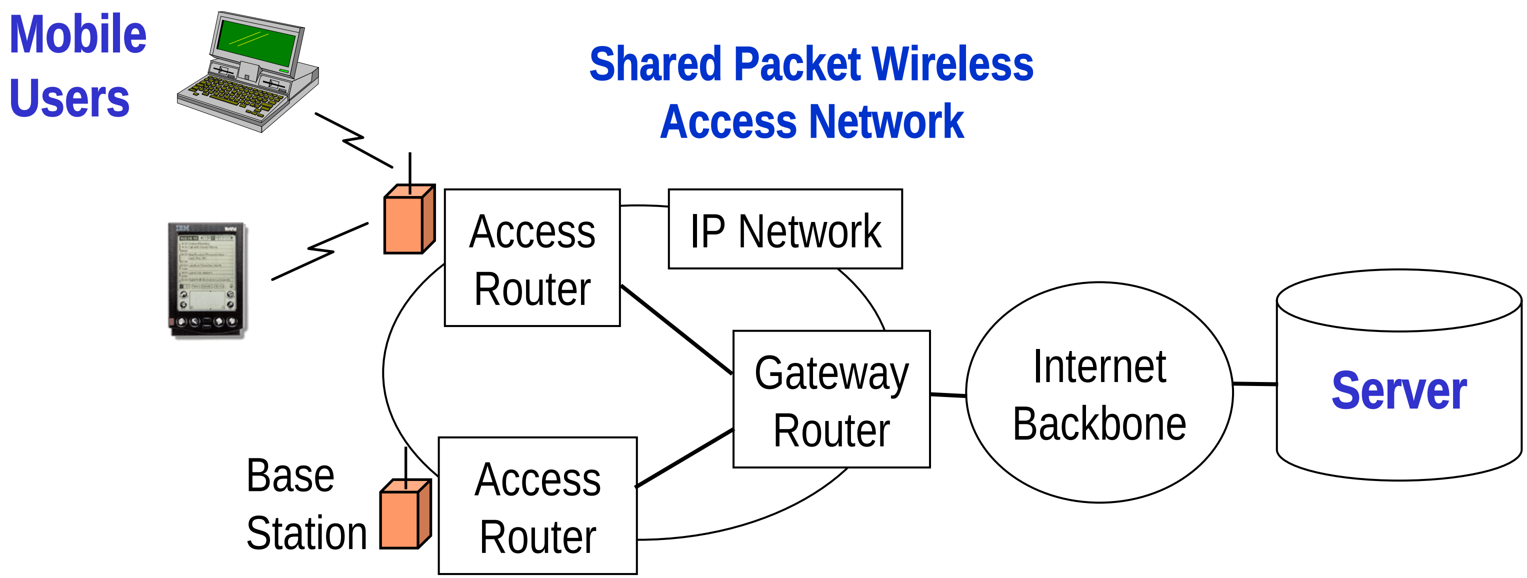
Flarion Technologies

Bedminster, NJ

Talk Outline

- ECR (Equivalent Circuit Rate) measure
- Analytical model for interactive traffic
- Shared-TDMA downlink simulation
- ECR used for MBWA

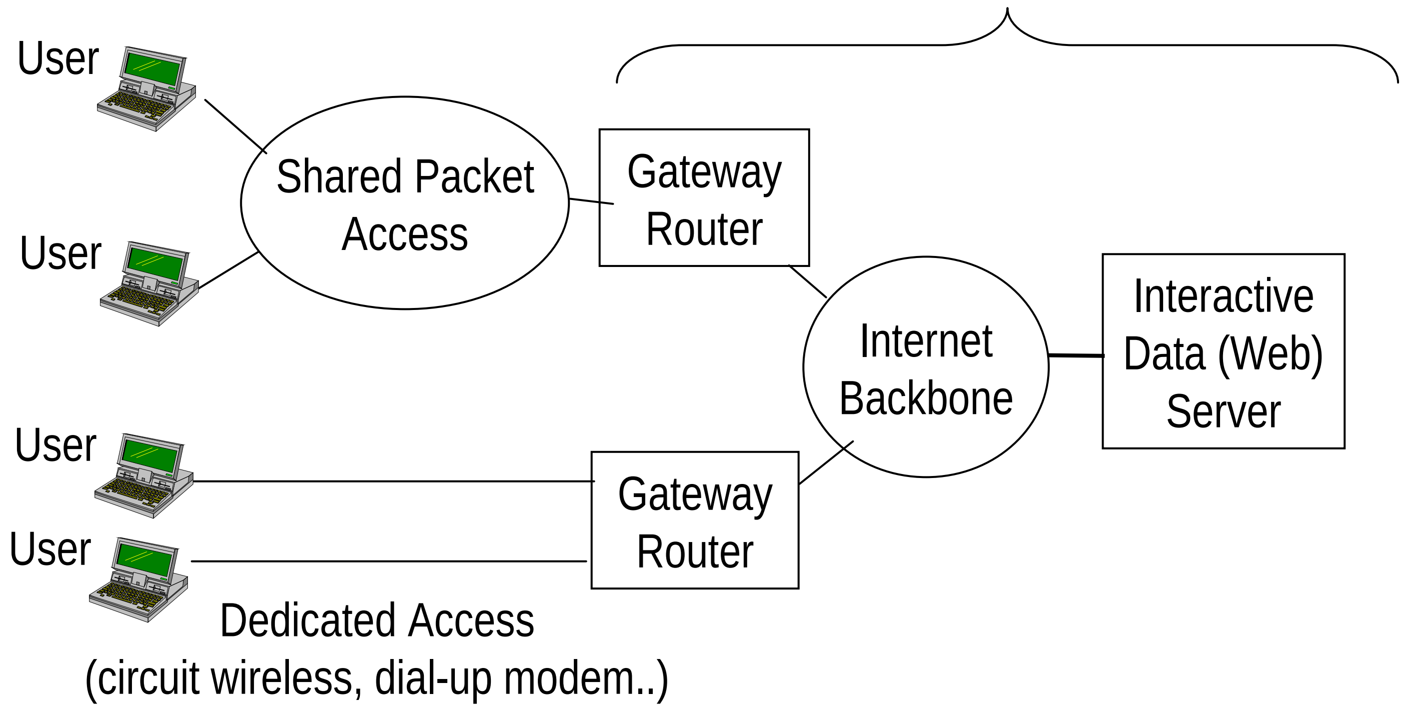
Wireless Data Network



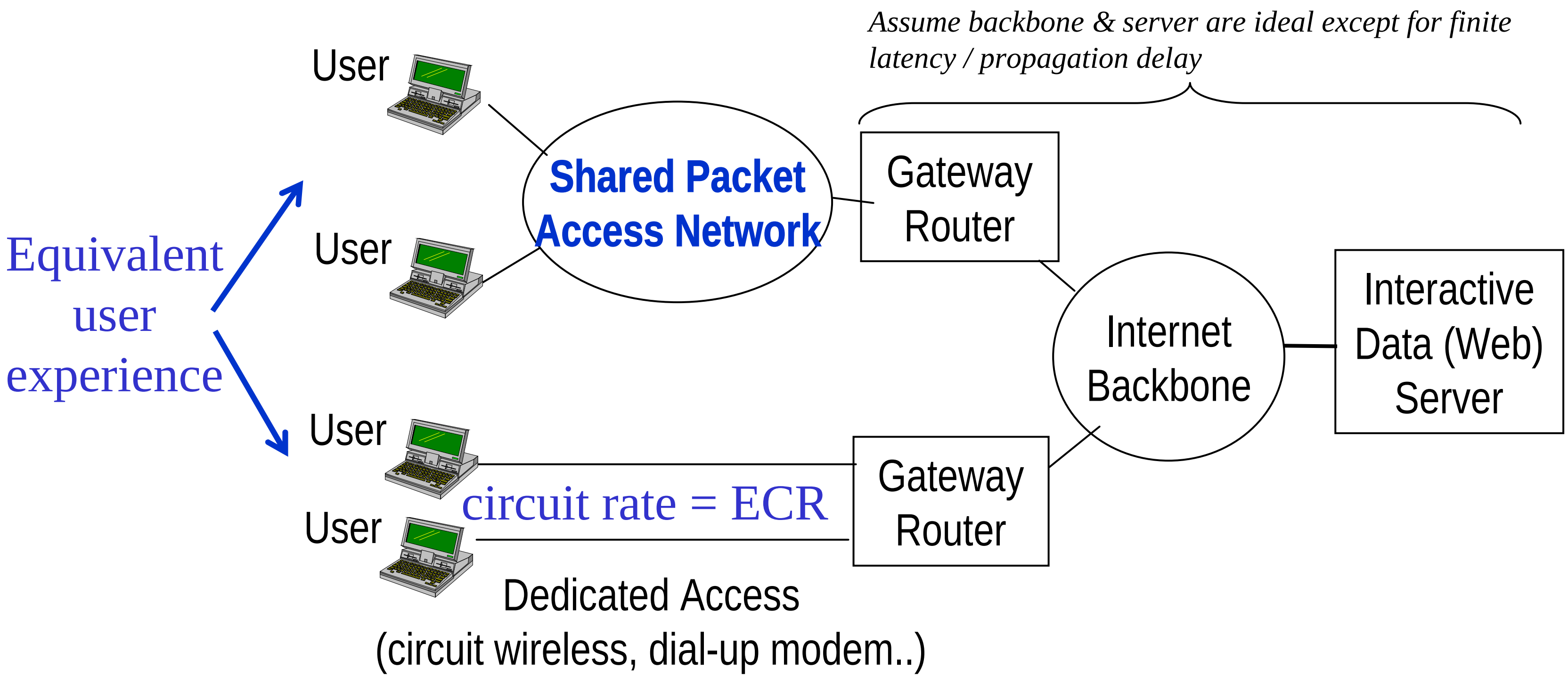
- We want to characterize performance & capacity
  - Quantify user experience
  - Determine number of users
- We will take a user-centric view

Shared v. Dedicated Access

Assume backbone & server are ideal except for finite latency / propagation delay



Equivalent Circuit Rate (ECR)



Equivalent Circuit Rate (ECR) of a shared packet access network:

- Dedicated rate required to get same user performance
- Useful and intuitive measure for users, planners, marketing
- Equivalence based on statistics of application (e.g. mean or percentile of web page delay)

ECR References

1) N. K. Shankaranarayanan, Zhimei Jiang, and Partho Mishra, "Performance of a shared packet wireless network with interactive data users," Mobile Networks and Applications, Vol 8, June 2003, pp. 279-293.
Full paper: complete analytical model and wireless simulations

2) N. K. Shankaranarayanan, Anupam Rastogi, and Zhimei Jiang, "Performance of a wireless data network with mixed interactive user workloads
Proc. ICC 2002. Vol 2, May 2002, pp. 887 -890
Mixed web & dataphone traffic workloads

3) N. K. Shankaranarayanan, Zhimei Jiang, and Partho Mishra, "User-perceived performance of Web-browsing and interactive data in HFC cable access networks," Proc. ICC 2001. Vol 4, June 2001, pp. 1264 -1268
Cable modem scenario

ECR Model: Clarifications

- Shared v. dedicated equivalence - very general
- Equivalence can be based on:
  - Mean delay
  - 90th percentile delay
  - etc
- Analytical model
  - Interactive traffic model
  - Mean delay
- Simulation work (shared TDMA wireless system):
  - Downlink simulation
  - Web browsing traffic model
  - Mean page delay, 90th percentile page delay, etc.

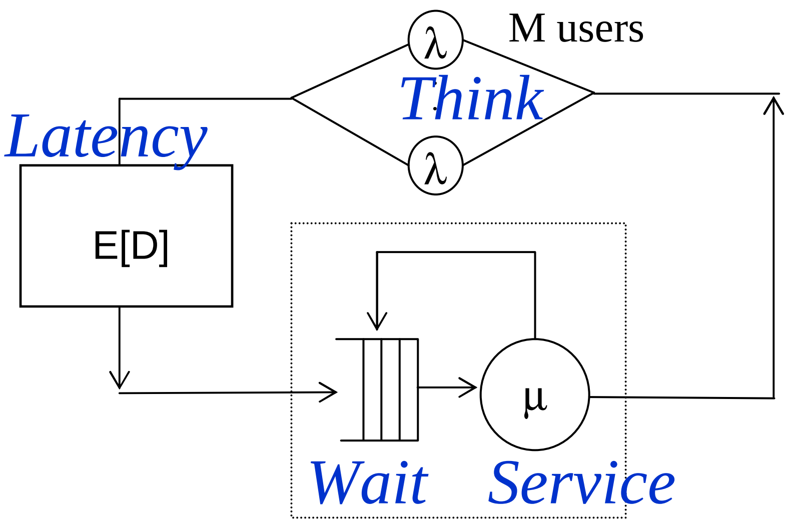
Talk Outline

- ECR (Equivalent Circuit Rate) measure
- Analytical model for interactive traffic
- Shared-TDMA downlink simulation
- ECR used for MBWA

Analytical Model for Interactive Traffic

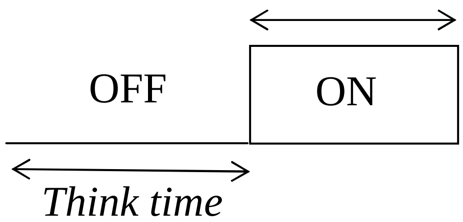
- Closed queueing network with finite population of ON/OFF sources
- OFF state independent of ON & Delay state => interactive traffic model
- Model each page (response) as one job for the server
- Per-user workload decreases as number of users increases

Closed Queueing Model



Interactive user workload model

Delay = Wait_Time + Service_Time + Latency

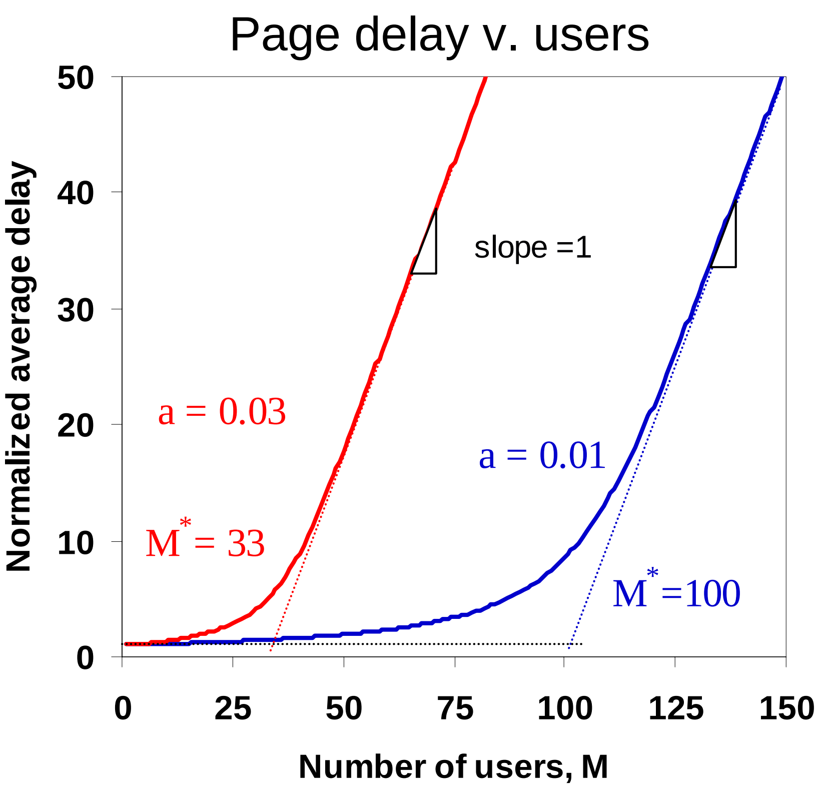


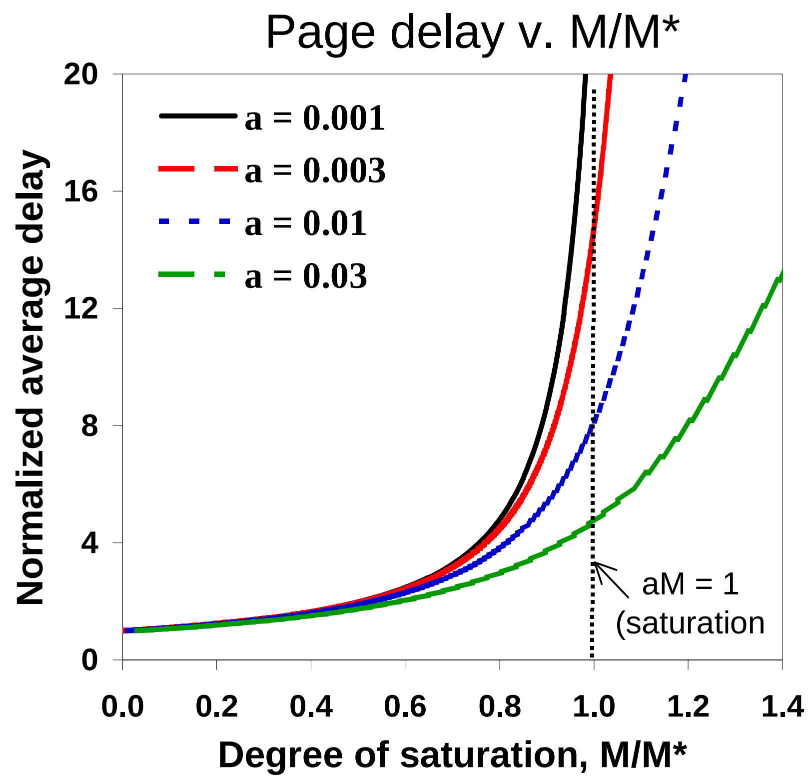
- Average OFF (think) time = $1/\lambda$
- Average service time = $1/\mu$
- Average latency = E[D]

Per-user offered load $a = \dfrac{1/\mu}{1/\lambda + 1/\mu + E[D]} \sim \dfrac{\text{average user rate}}{\text{shared network rate}}$

Page delay v. users

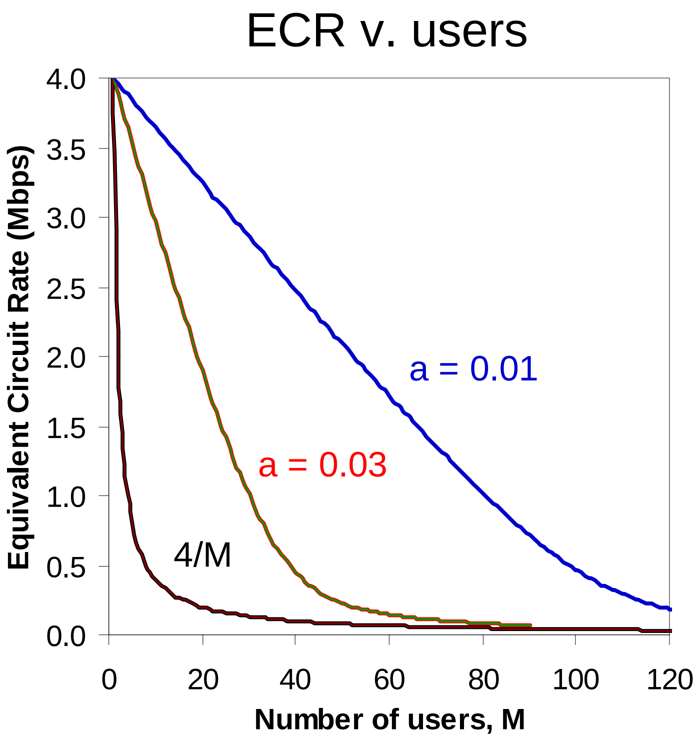
- Average delay - normalized to minimum (single-user case)
- Saturation phenomenon: Linear rise in delay beyond M*=1/a
- Network sizing ~ 0.7M* = 0.7/*a* for wide range of *a*

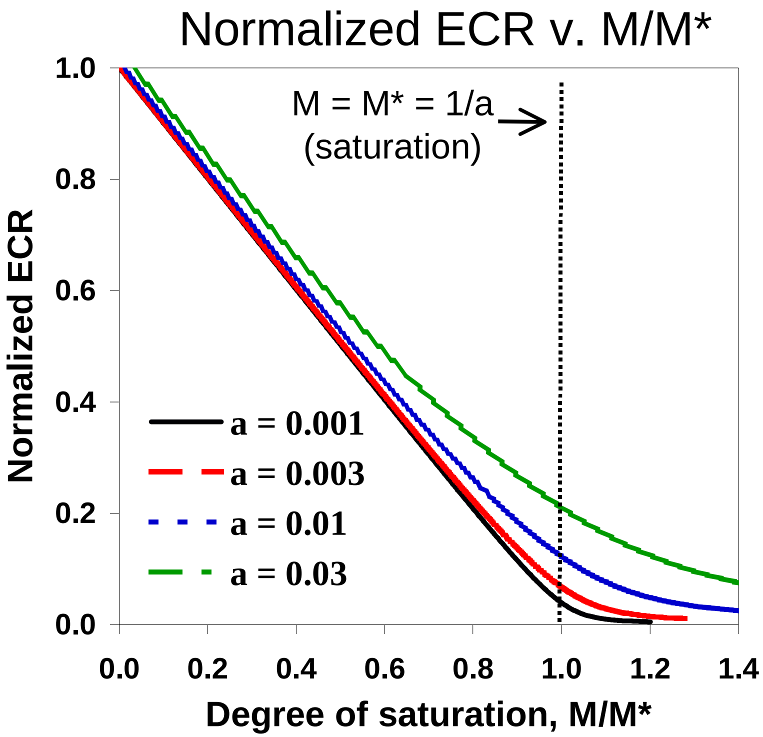




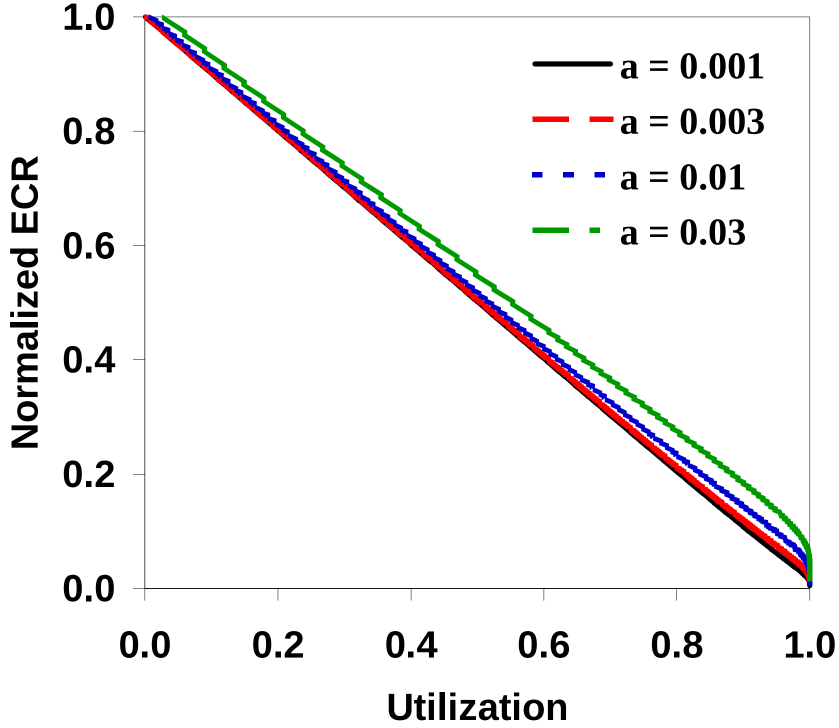
Equivalent Circuit Rate (ECR) v. users

- Assume network rate of 4.0 Mbps shared by M users
- ECR = rate required for same average page delay
- ECR >= 4.0/M Mbps => multiplexing gain





Normalized ECR v. Link utilization

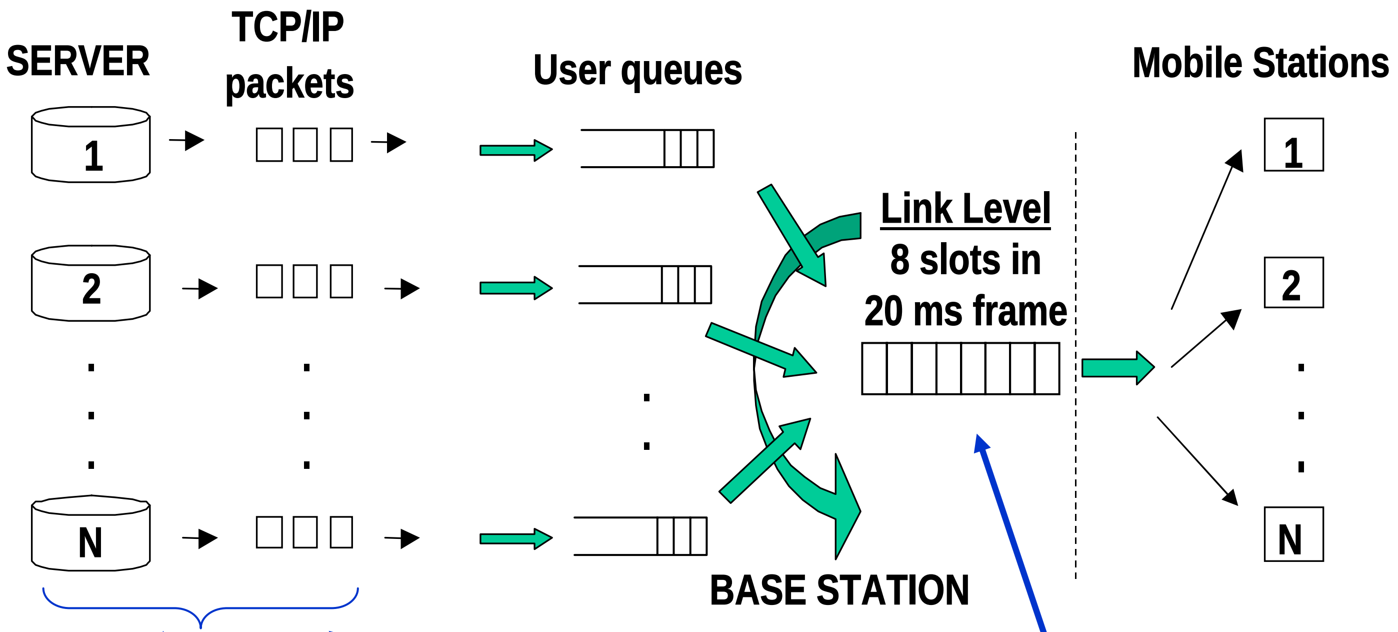


- Normalized ECR ~ 1- *Utilization* for wide range of a !
- ECR ~ *Shared Rate**(1-*Utilization*)
- Universal relationship – quite general
- Intuitive on hindsight

Talk Outline

- ECR (Equivalent Circuit Rate) measure
- Analytical model for interactive traffic
- Shared-TDMA downlink simulation
- ECR used for MBWA

Simulation Model



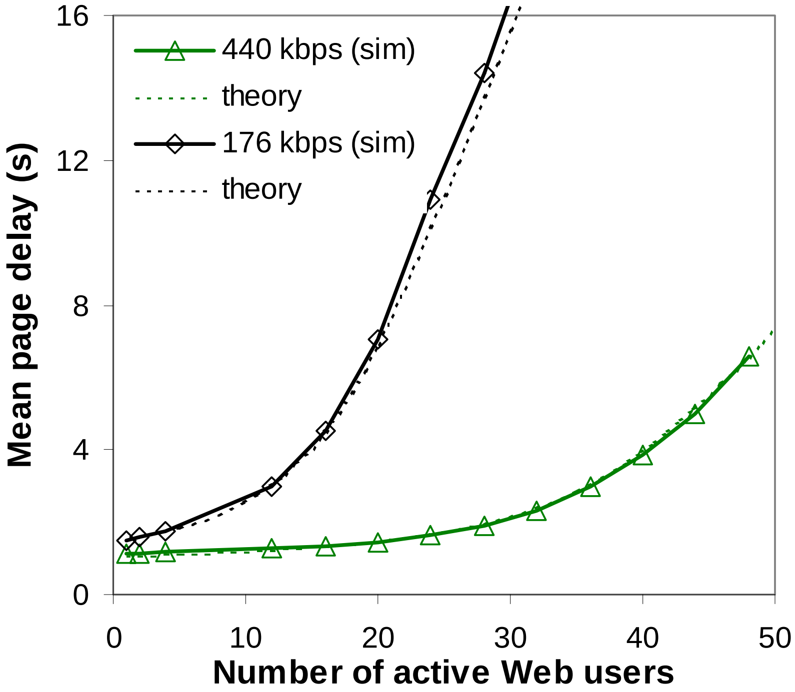
- Web workload
- FreeBSD TCP code

Backbone network round-trip latency: 2 ms to 200 ms

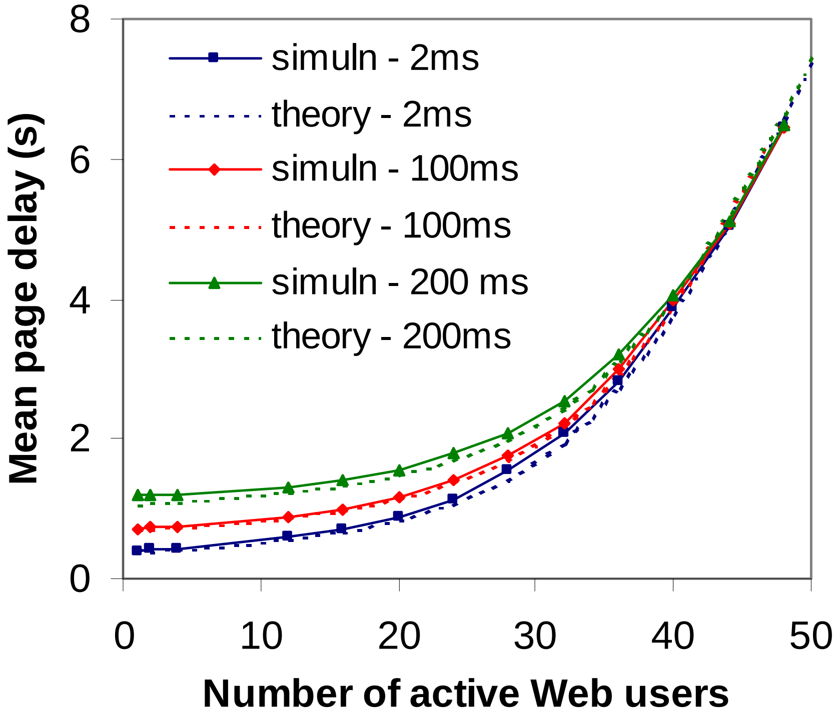
Shared channel rate: 176 to 440 kbps (constant rate per simulation)
Block error rate: 0 to 40%

Page delay: Analysis and Simulation

- Different network rates
- 200 ms latency

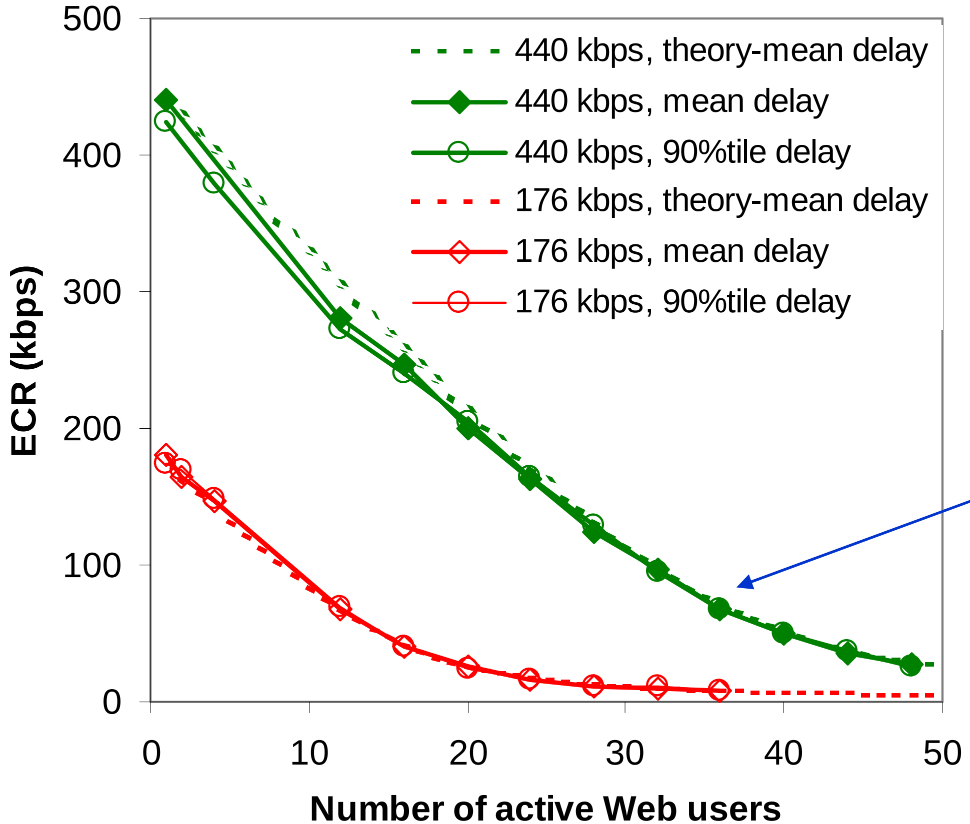


- Different latencies
- 440 kbps network rate



ECR: Mean & percentile page delay

- ECR result is same when calculated using two methods to capture user experience: average & 90th percentile page delay
- Good match with analytical results (ECR based on average page delay)

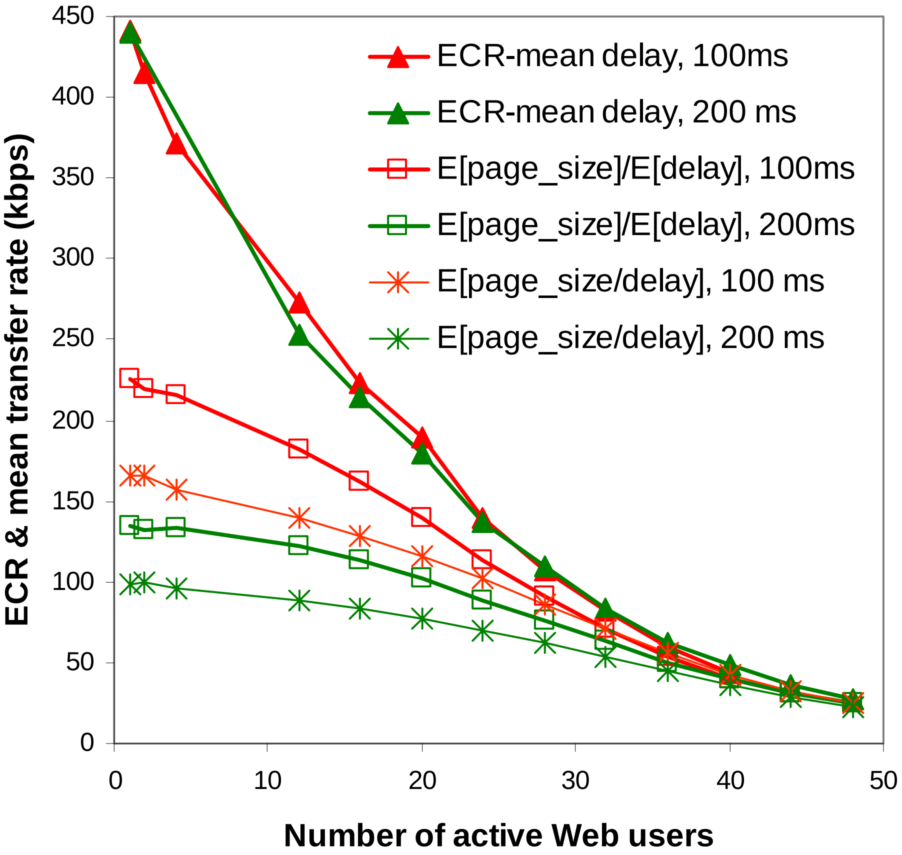


- 200 ms latency
- No errors
- 9-13 kbps user rate

Network can be sized based on ECR target

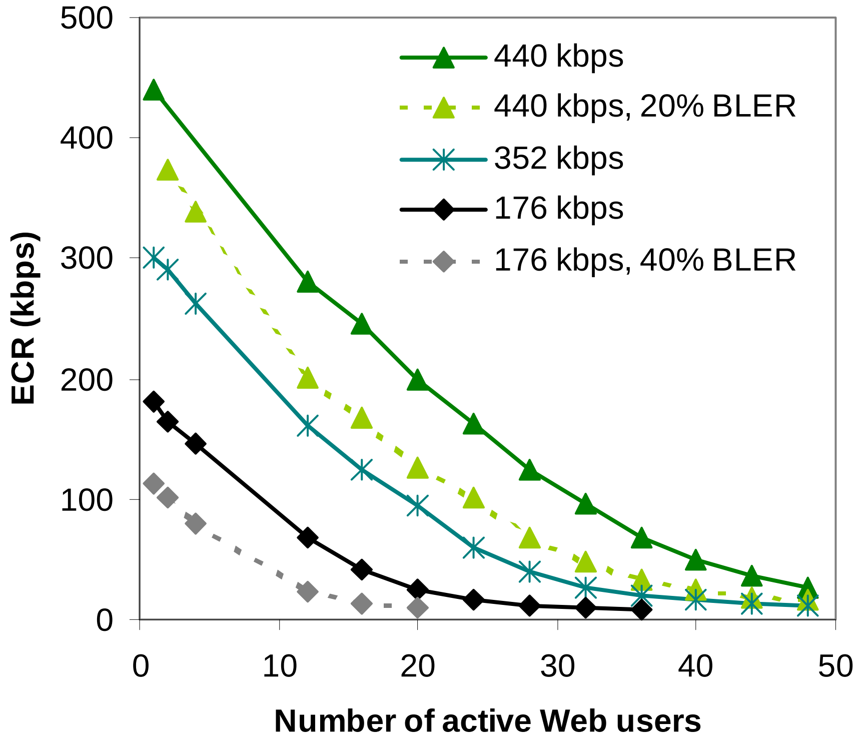
ECR and other end-to-end rate measures

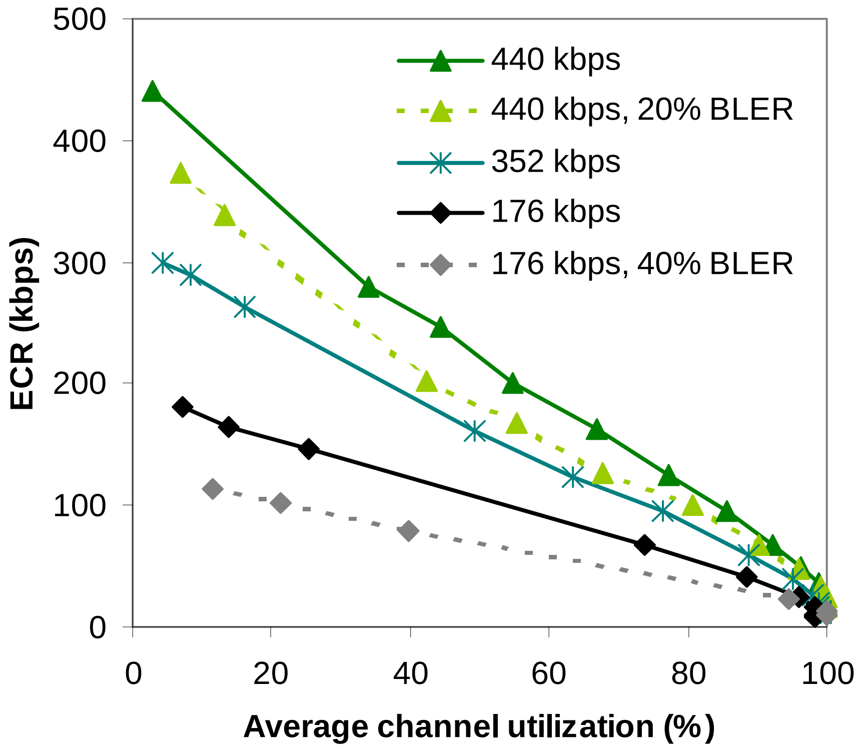
- ECR uses end-to-end delay to characterize only the access portion
- ECR reflects effective access network bandwidth available to user
- ECR insensitive to delay, but other measures are not
- Differences decrease at high loads



Different radio channel conditions

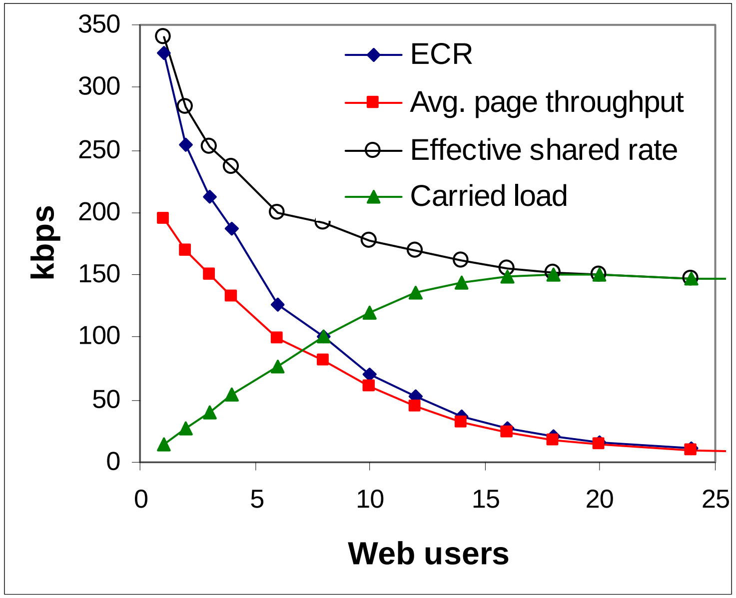
- 176 to 440 kbps with different error rates
- Includes link-level retransmission
- Linear ECR v. utilization curves

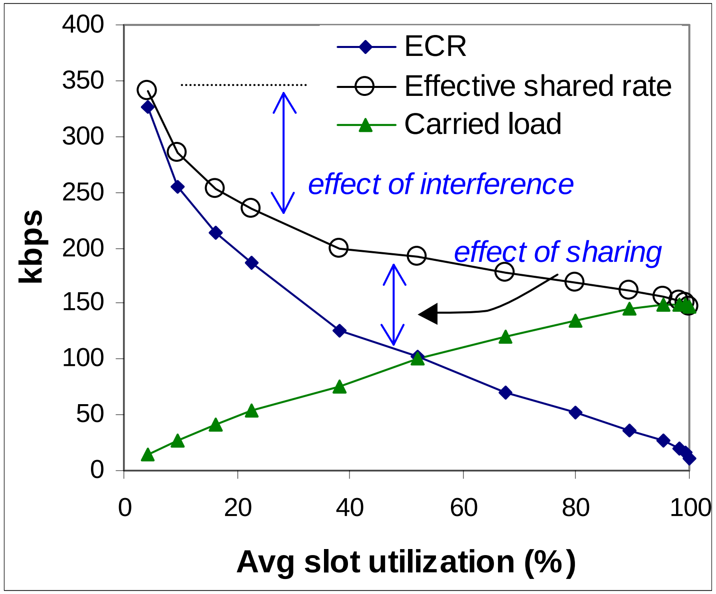


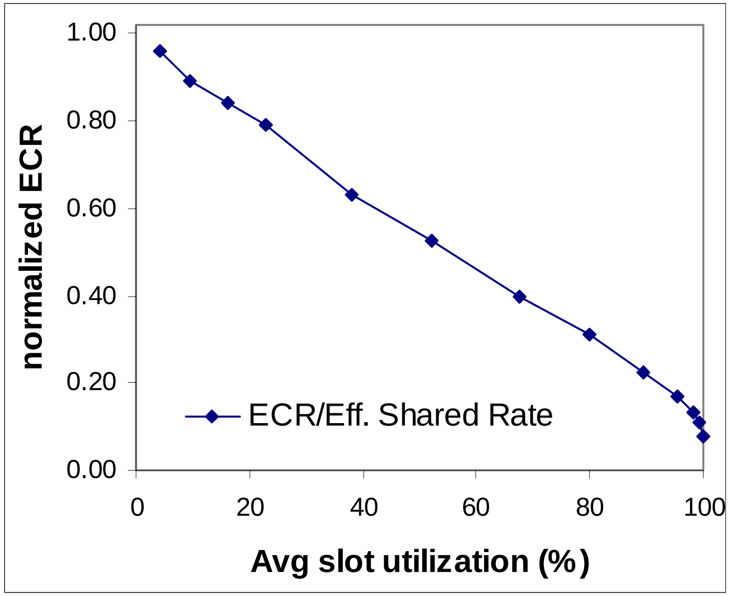


Markov channel model

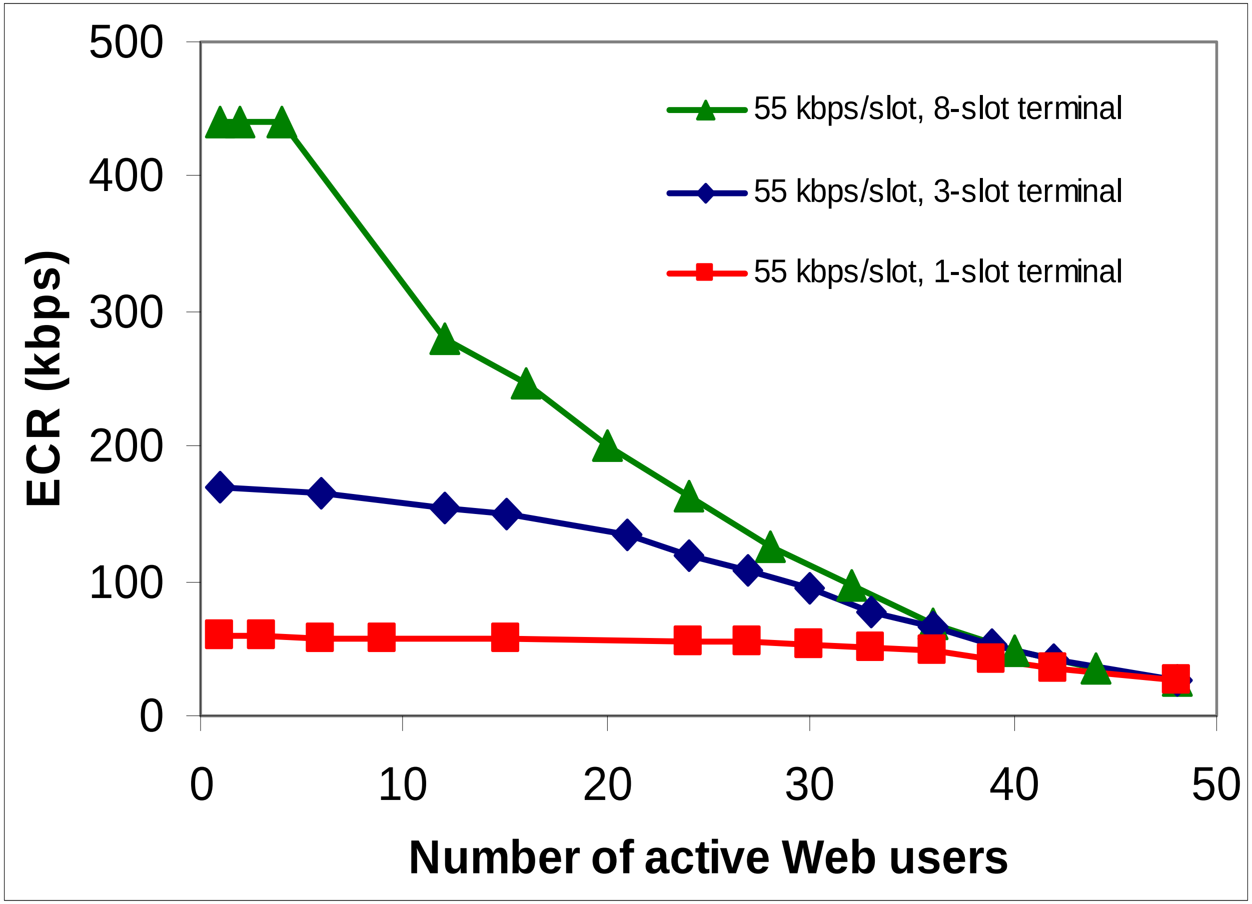
- Time-varying
- Degradation with load (interference)
- ECR decrease with load
  - channel degradation
  - sharing



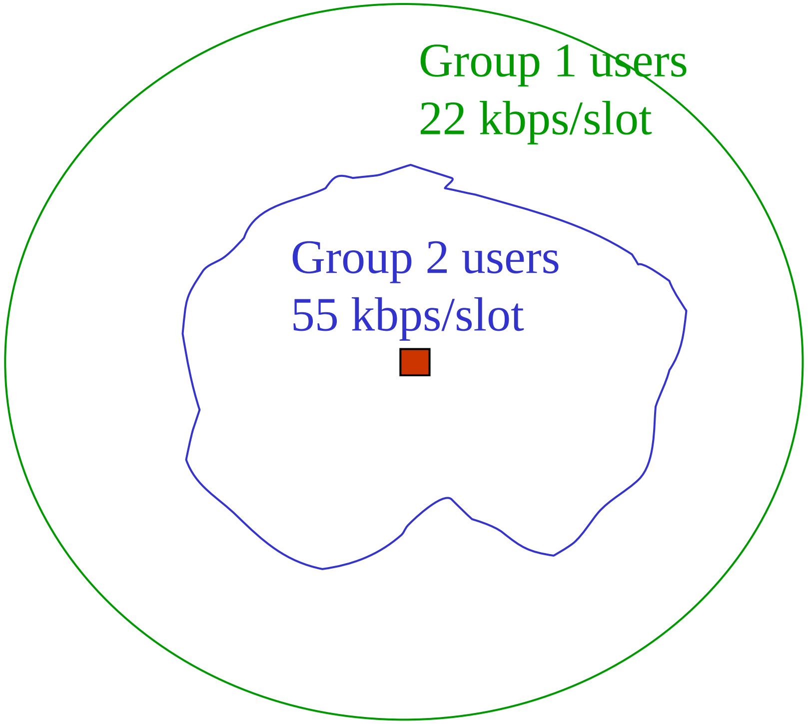




ECR for different user modem rate limits



Mixed channel conditions



- Users in Group 1:
  - Farther from base
  - Low rate
- Users in Group 2:
  - Closer to base
  - High rate

Uniform Case Results:

Only low rate:

$a_1 = .067$, $M_1^* = 1/a = 15$

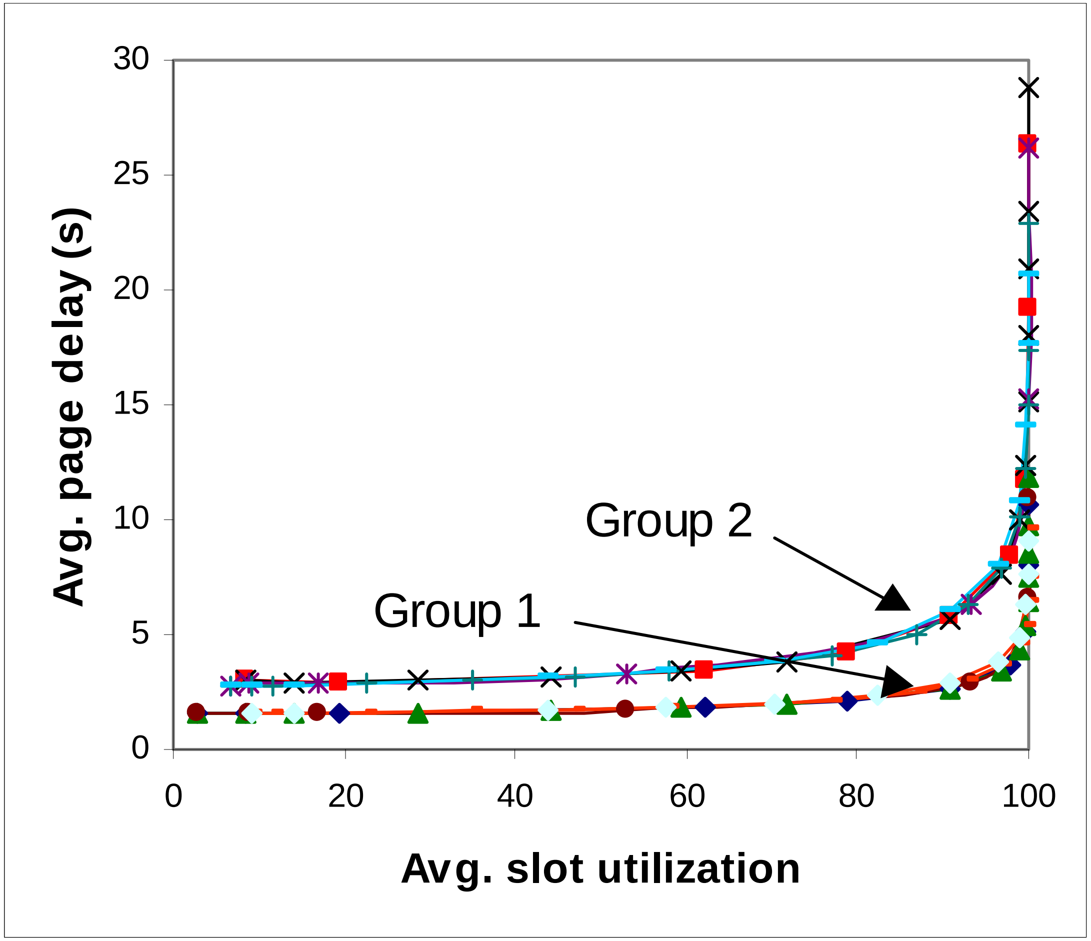
Only high rate:

$a_2 = .028$, $M_2^* = 1/a = 35$

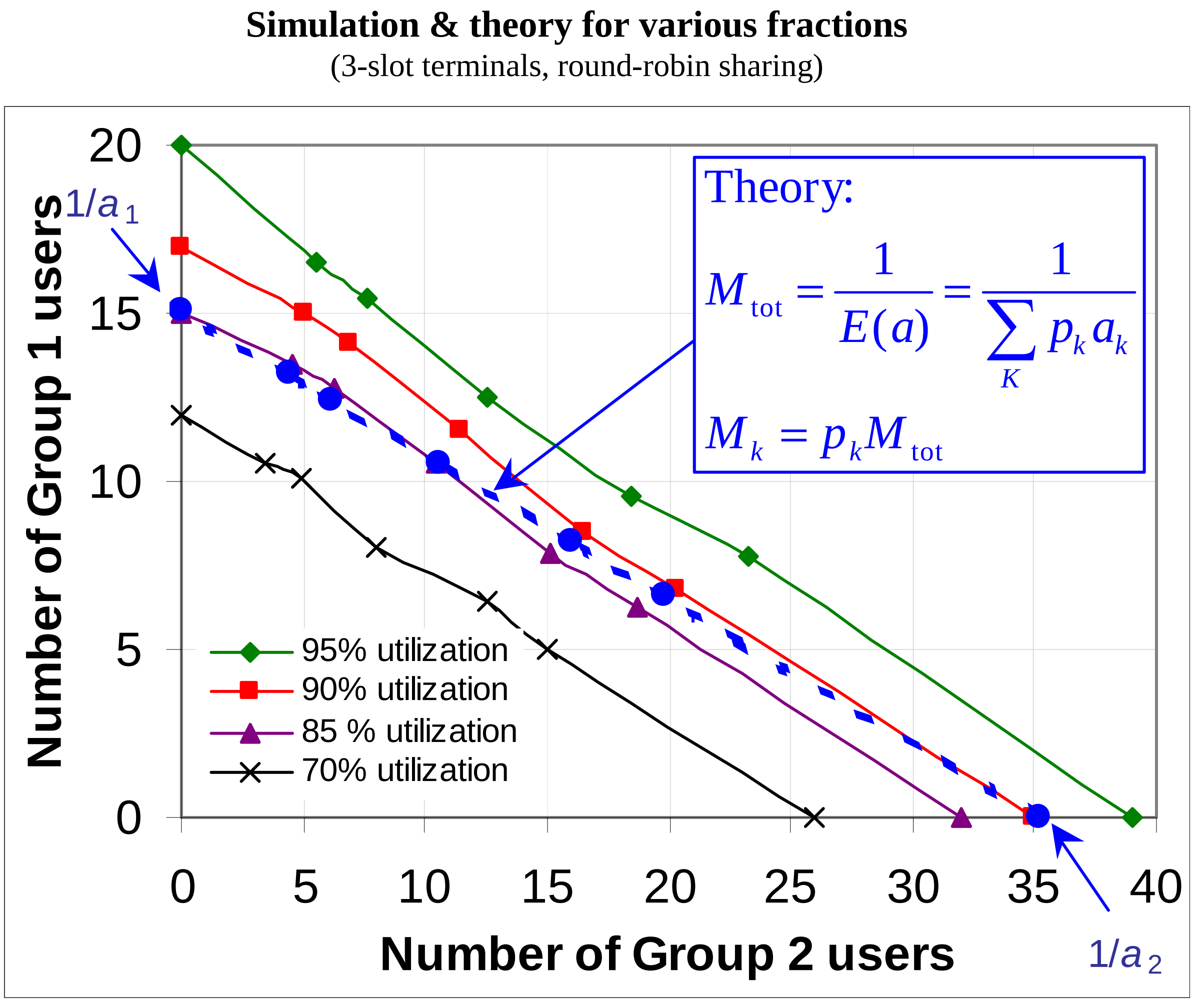
Non-uniform Case:

???

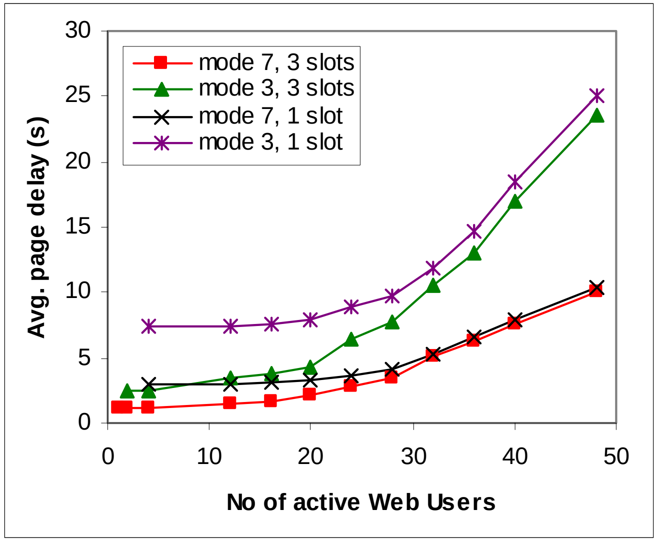
Delay v. utilization for mix of channel conditions: insensitive to fraction of mix

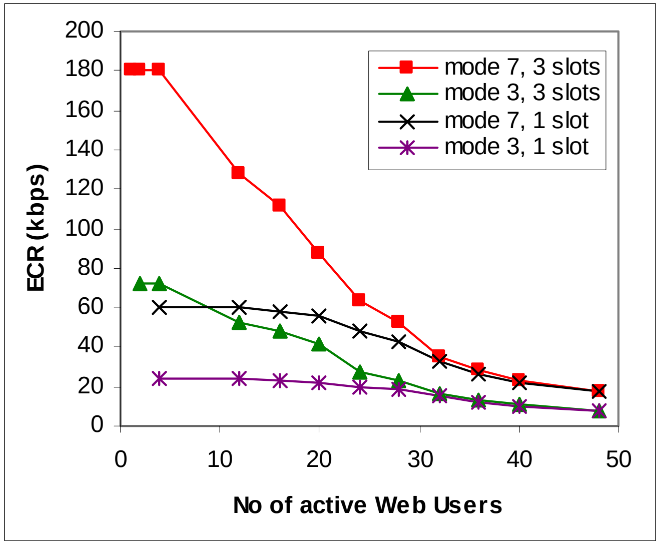


Number of users for mixed channel rates

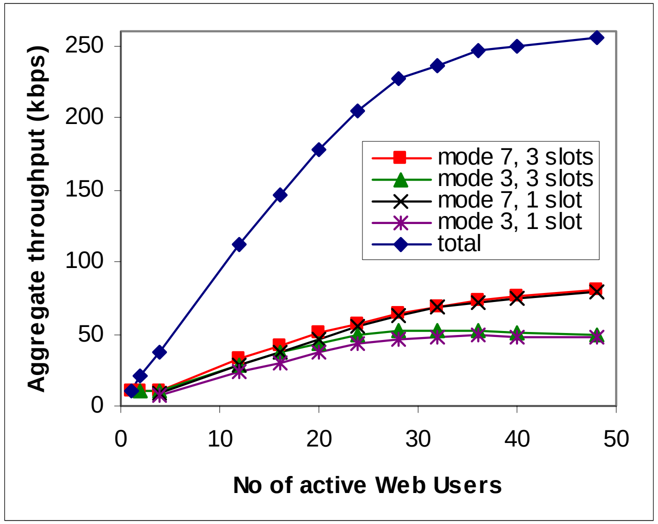


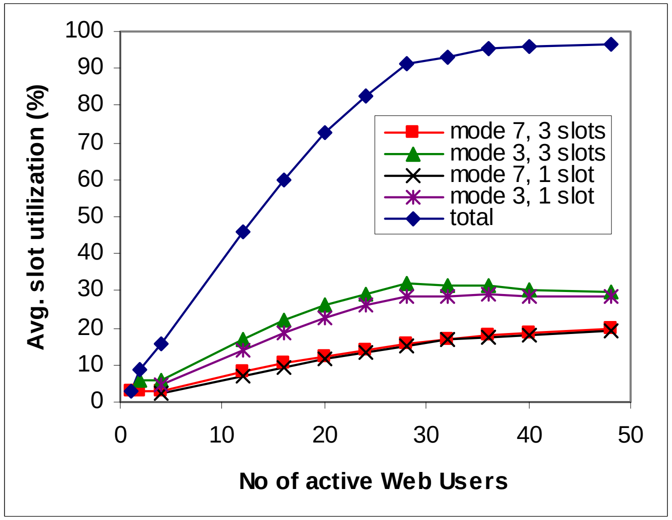
4 groups:hi/lo channel rates & hi/lo modem rate limits



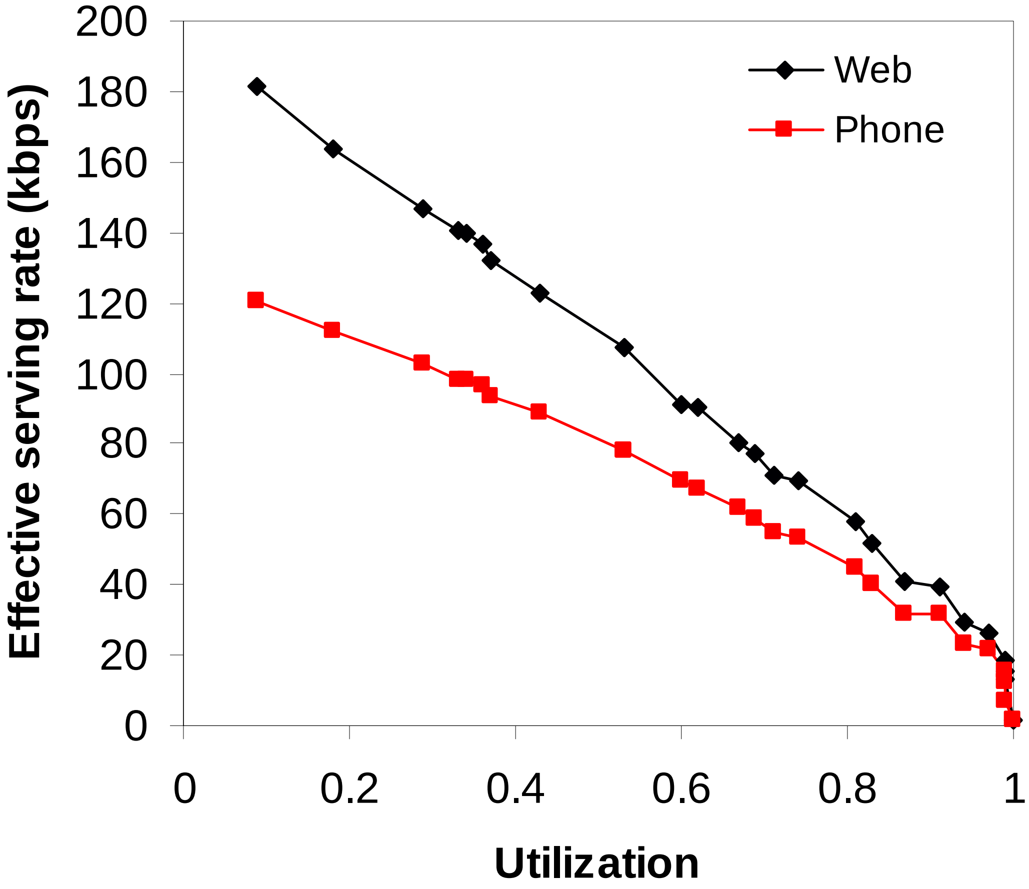


4 equal groups:
55 kbps, 3-slot
55 kbps, 1-slot
22 kbps, 3-slot
22 kbps, 1-slot





Mix of traffic loads:
Web and dataphone (WAP-like) users



- **Wireless channel:** Time-varying TDMA channel with mean rate of 220 kbps
- **Web user:** average rate ~ 9 kb/s
- **Dataphone:** average rate ~ 0.65 kbps
- Small size of dataphone (WAP-like) payload resulted in inefficient use of TDMA slots

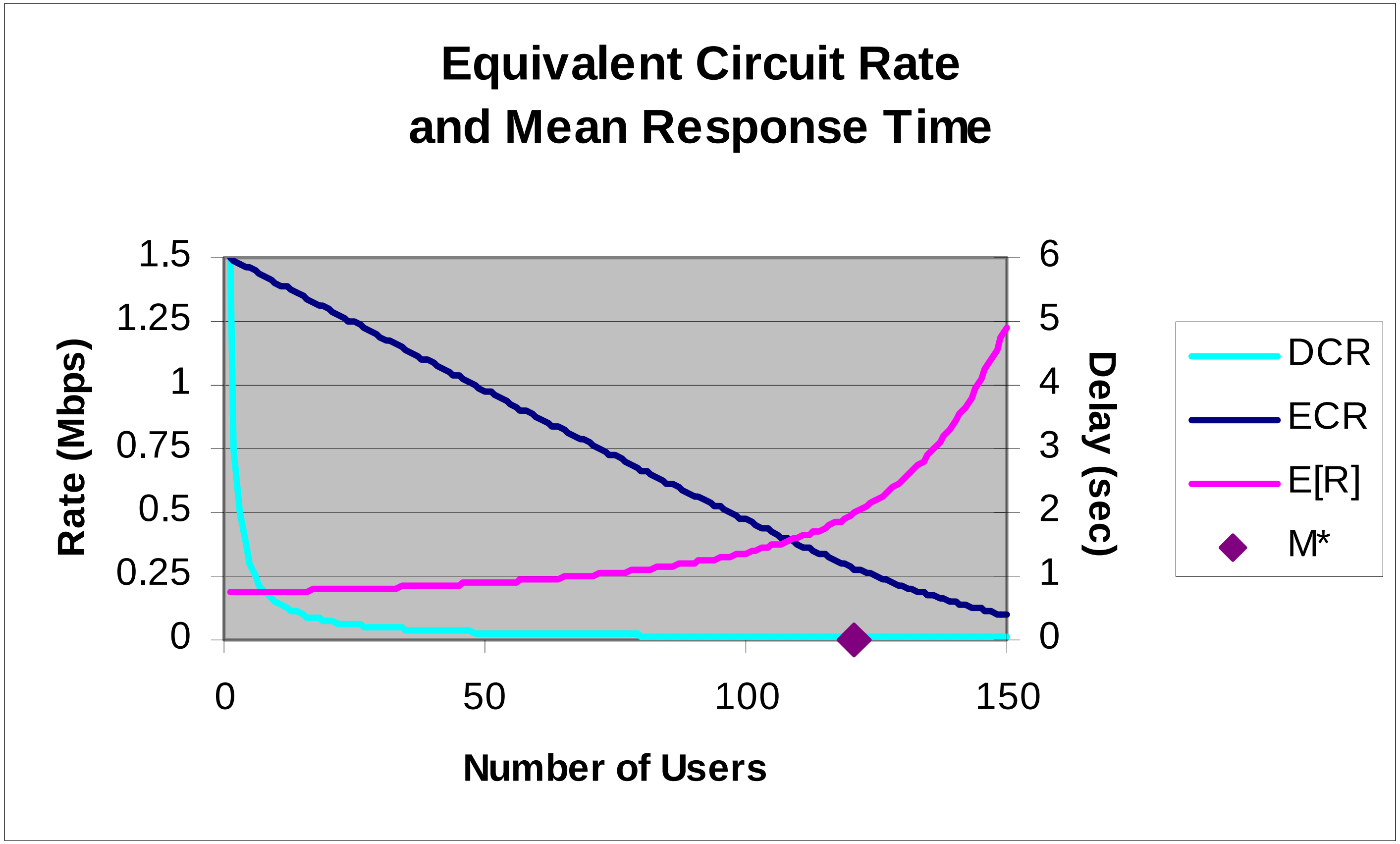
Talk Outline

- ECR (Equivalent Circuit Rate) measure
- Analytical model for interactive traffic
- Shared-TDMA downlink simulation
- ECR used for MBWA

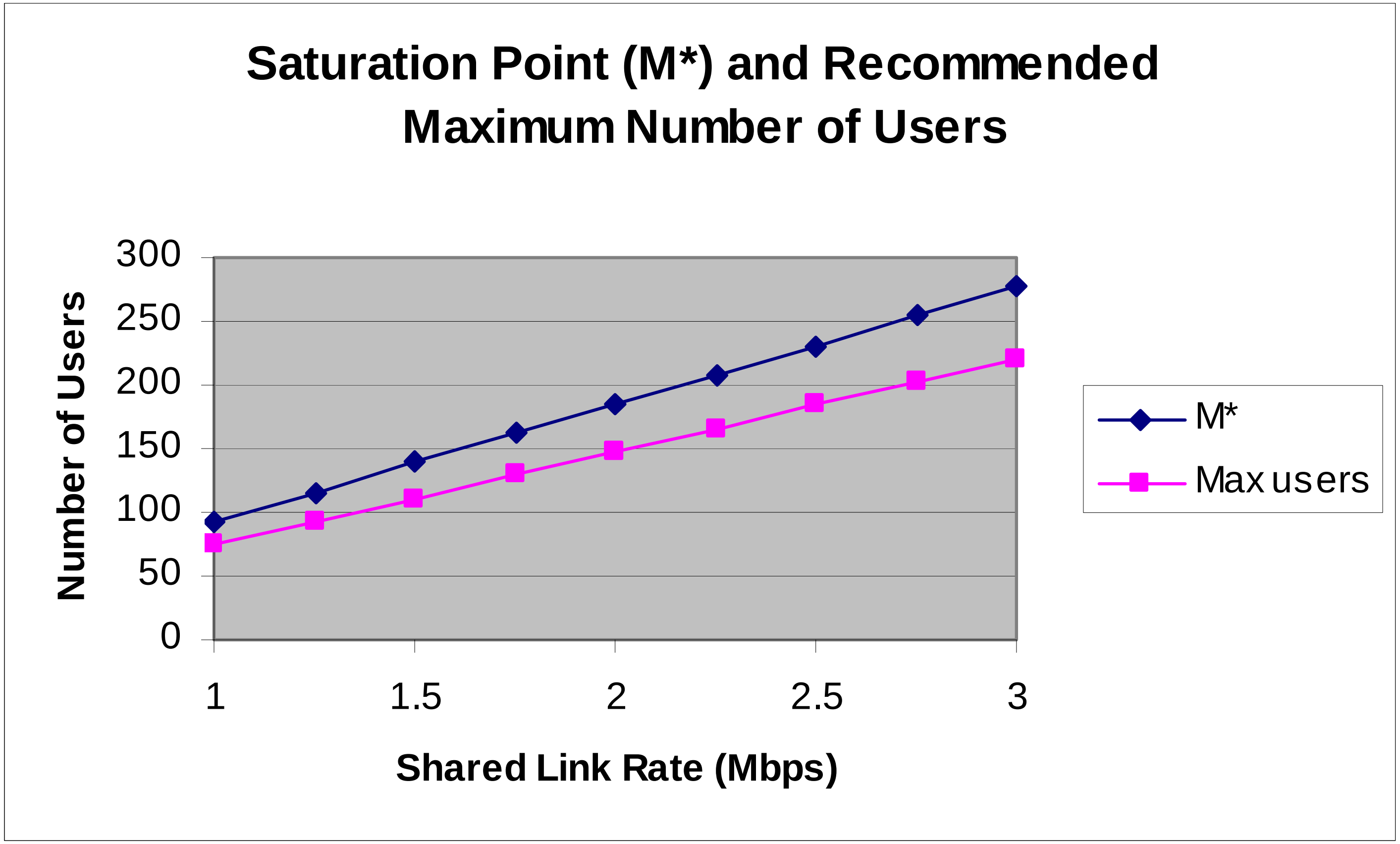
Using ECR for MBWA Dimensioning: E.g., Number of Active Users

- Based on modeling Web traffic (dominant)
- Web traffic model details…
  - Mean “Web-request” size: 53 Kbytes
  - Mean “Viewing” time: 40 sec
- Additional system model details…
  - Shared Link Rate: 1.5 Mbps
  - E[D]: 450 ms (accounts for RTT and TCP slow start)
- Corresponding User Workload: ~10.8 kbps

Saturation Point suggests number of active users that should be concurrently supported



Effect of changing shared link rate



Improving Fidelity of Model for MBWA Dimensioning

- Selecting the Shared Link Rate
  - Function of user distribution in cell/sector?
  - Function of interference?
  - System might not offer full aggregate rate to one user
- Modeling mix of individual user rate limits
  - Due to channel conditions
  - Due to administrative policy, e.g., policing
- Modeling a mix of users with different workloads
  - Need traffic models for other interactive applications (e.g., instant messaging?)
  - Need to defined mix of user traffic models
- Inclusion of non-interactive traffic models
  - Aggregate rate of CBR traffic flows (e.g., VoIP) can be subtracted from Shared Link Rate prior to modeling/analysis

ECR in MBWA Simulations

- Separate computation of ECR reference (e.g.
  - Pre-compute equivalence metric (e.g. mean delay) v. circuit rate for a single user with dedicated circuit connection for same conditions of {traffic model, protocol model, backbone network etc.} - this is independent of MBWA system
  - Map MBWA user performance to ECR using lookup
- Runtime computation of Effective Serving Rate ~ ECR
  - ECR = ESR ~ (Σ user bytes transferred)/(Σ “serving time”)
  - “serving time” = time when user queue is not empty
- ECR may be useful as comparison tool
- Easy to track ECR - useful as input for research, e.g. analytical model can be extended

Value of using ECR for MBWA

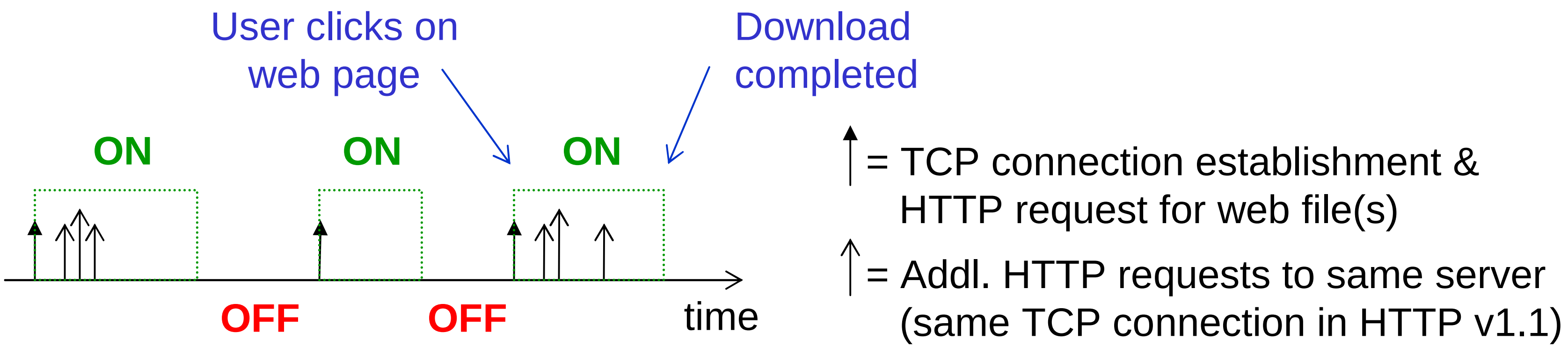
- Technical
  - Forces one to think about, and quantify various rates
    - aggregate rate, shared rate, user rate
  - ECR v. utilization relationship may reveal "capacity" behavior under load
  - Insights into system behavior
- Planning, Operations, and Marketing
  - Simple model, simple rules of thumb
  - End users & planners can easily relate to ECR
  - ECR highlights multiplexing advantage of shared systems for bursty traffic

Conclusion

- Useful and robust new measure of user-perceived performance: ECR (Equivalent Circuit Rate)
- Simple analytical model for interactive data in a shared packet network
- ECR & network sizing depend on simple parameters: (shared channel rate, average user rate)
- Results validated by TCP-based simulations of typical Web workloads in shared TDMA wireless system

Backup Slides

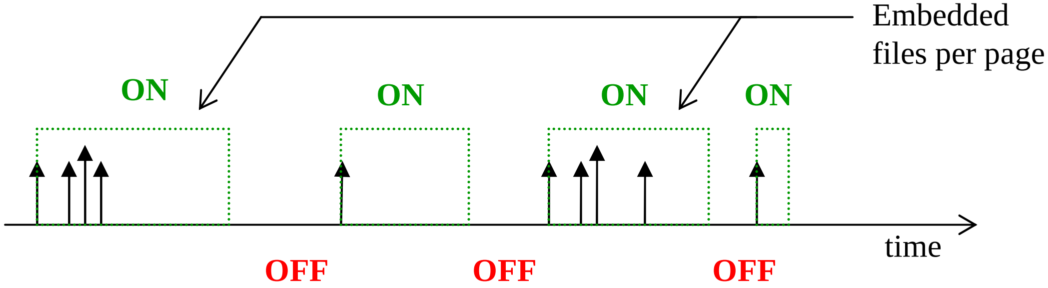
Workload Model for Web/Interactive Data



- Focus on Web-based applications since they dominate Internet usage
- Users alternate between ON (busy) state and OFF (idle) state
- ON state duration (Web page delay) depends on:
  - Page size statistics - no/size of files in page - typically heavy-tailed
  - Server and network resources
  - HTTP and TCP protocol dynamics
- OFF state duration (Think time)
  - Determined by human behavior - typically heavy-tailed
  - Independent of ON state => feedback, upper limit on workload

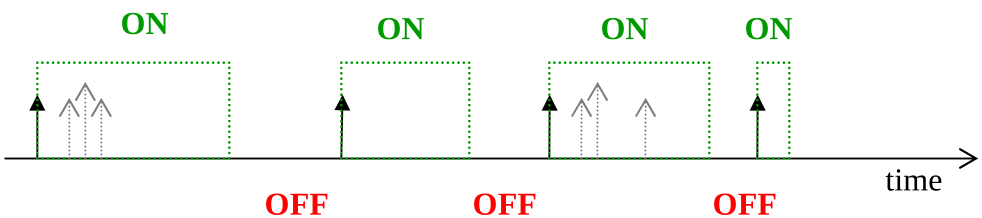
Web Workload Model

HTTP v1.0



= HTTP request and TCP connection establishment

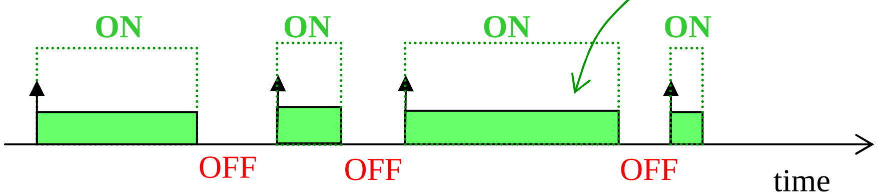
HTTP v1.1



= HTTP request and TCP connection establishment followed by period during which the web **page** is received by the user

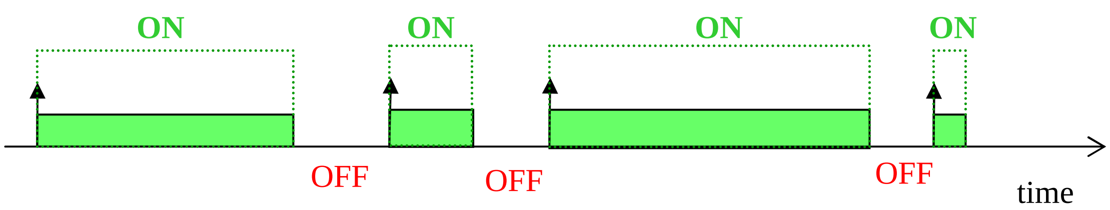
Web Browsing Workload Model

Web page statistics*
- Lognormal distribution
- 10 KB median, 20 KB mean
- 100 B min, 100 KB max



OFF (think) time statistics*
- Pareto distribution
- 4 sec median, 12 sec mean
- 2 sec min, 10 mins. max

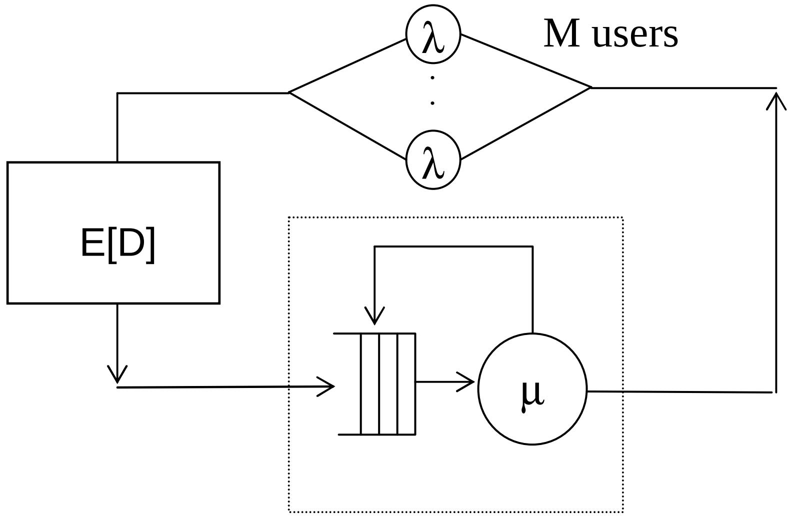
Lower channel bit rate OR more users ...
- Longer ON times (Web page delay)
- Same Web pages, Same OFF (Think) times



- 9 to 13 kbps avg user rate
- Laptop scenario
- Lower rates expected with PDAs and smartphones

* Web traffic statistics based on Boston University SURGE model (1998)

We can assume exponential distributions ..

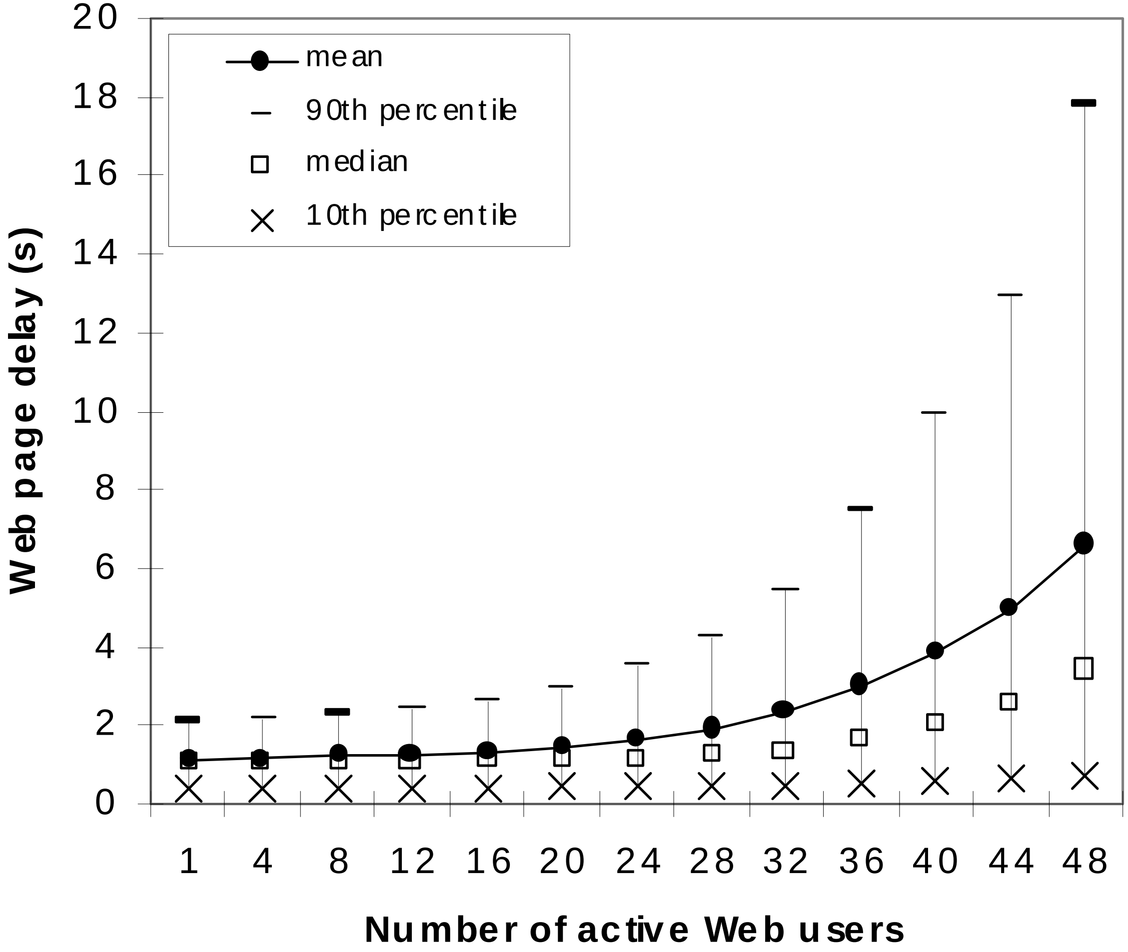


- Formulas can be easily derived by assuming **exponential distributions** for **service time and think time**.
- For processor-sharing (round-robin) service policy, the formulas involving mean statistics are also valid:
  - for any **rational service time** and **think time distributions**
  - for any **general service time distribution** and exponential think time distribution

Table I. Simulation parameters for network and traffic model.

Access network bit rate	176 to 440 kbps
Round-trip backbone propagation delay	2 to 400 ms
Queueing policy	Round robin
TCP window size	64 KB
TCP MSS	512 bytes
Web page size distribution	Log-normal (μ = 9.5, σ = 1.8) in bytes $f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-(\ln x - \mu)^2 / 2\sigma^2}$ 100 B min, 100 kB max
Median/Mean Web page size	10 kB / 20 kB
OFF time distribution	Pareto (k = 2 s, α = 2) $f(x) = \alpha k^{\alpha} / x^{\alpha+1} \ , x \geq k$ 2 s min, 10 min max
Median/ Mean OFF time	4 s / 12 s
Workload for single user	9 to 13 kbps

Web page delay behavior



- 440 kbps rate
- No errors
- 200 ms latency
- 9 to 13 kbps user rate