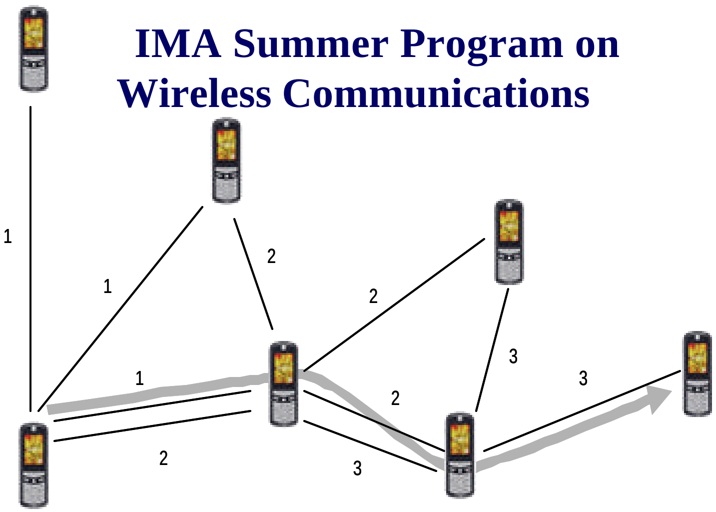
VoIP over a wired link



Phil Fleming
Network Advanced Technology Group
Motorola, Inc.

Acknowledgements

- Rajesh Pazhyannur and Ivan Vukovic, Motorola
- Ilkka Norros, VTT, Finland

Outline

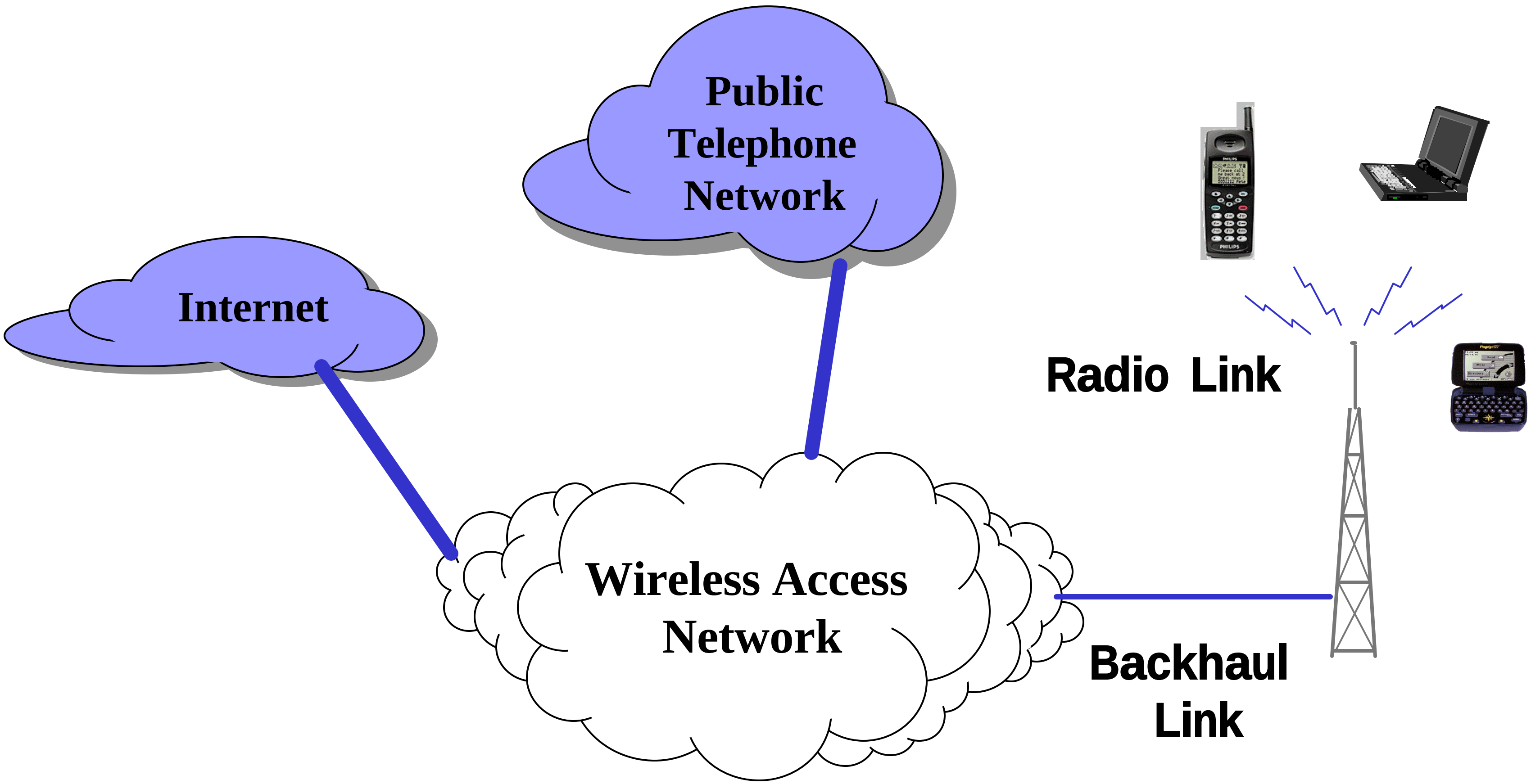
Part 1: VoIP over a wired link

- **Traffic models of VoIP**
- **The hockey-stick nature of the queueing delay distribution**
- **Large deviations analysis (ala Norros, et. al.)**
- **The two-law approximation**
- **VoIP over wireless teaser**

Part 2: (Friday) VoIP over wireless

- **3GPP2 EV-DO**
- **3GPP HSDPA/HSUPA**
- **802.16d/e**

Packet Radio Access Network



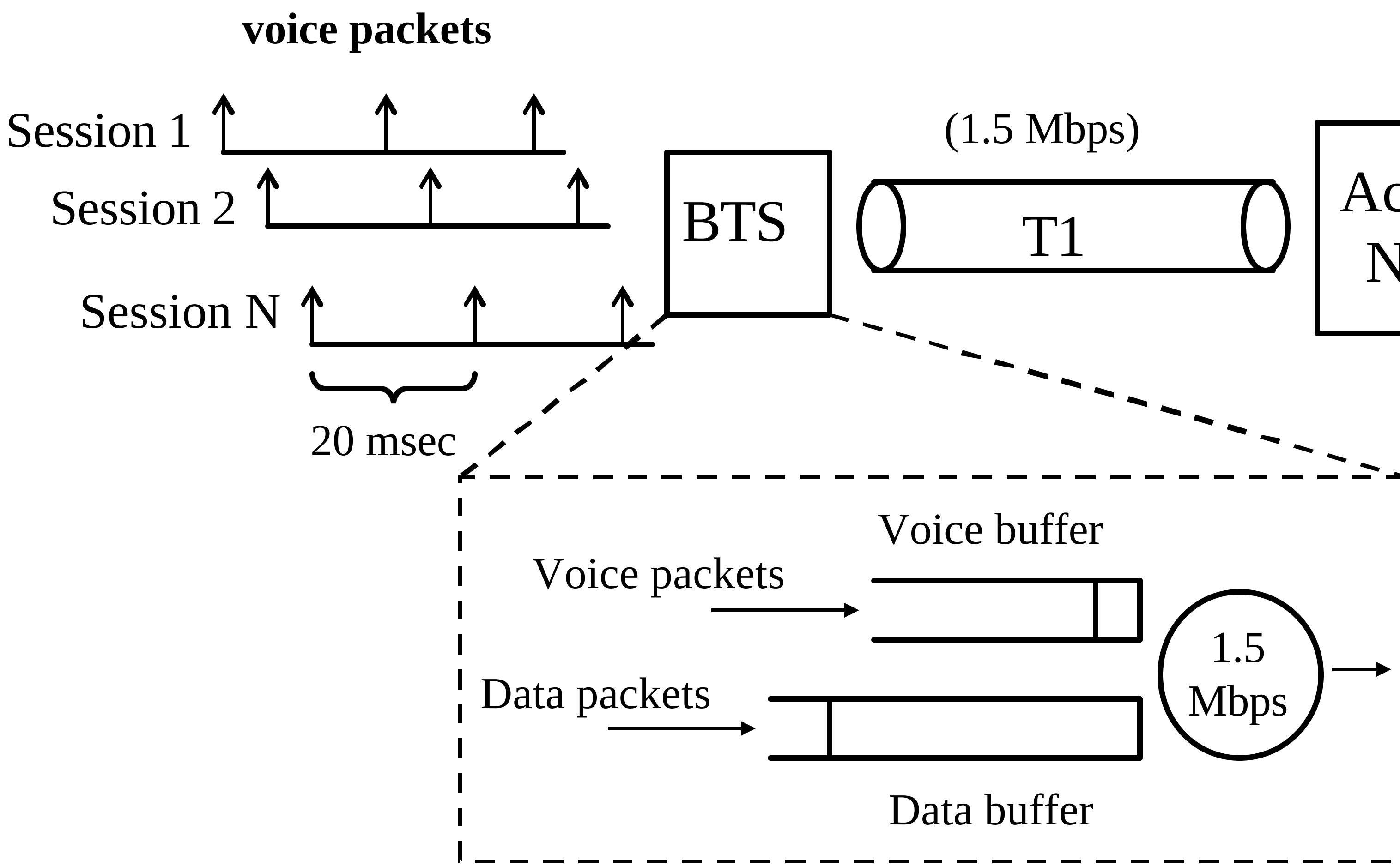
The Backhaul Link

- T1 or E1 in almost all current cellular systems (1.5 or 2.0 Mbps, resp.)
- Fiber or xDSL will be rare over the next five years
- Up to 20 kilometers
- Hundreds of backhaul links per Access Network in 2G systems and possibly thousands in 3G systems.
- Accounts for up to 60% of recurring operating cost

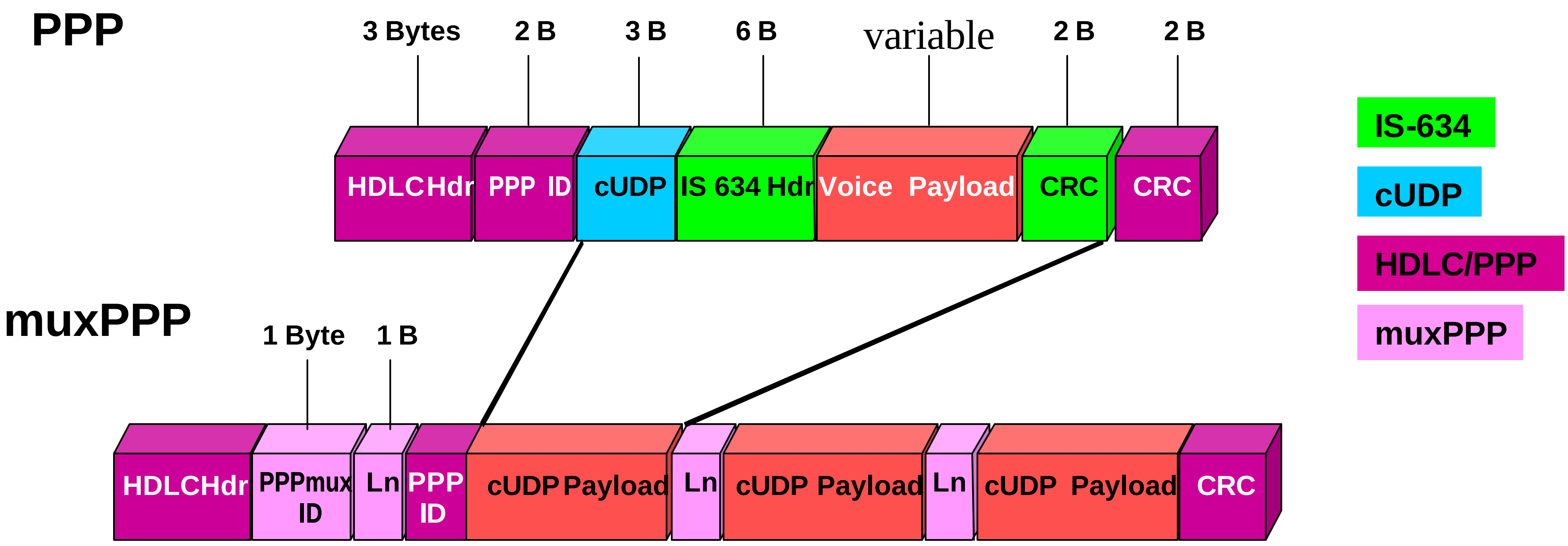
Packetized Voice in Cellular Telephony

- The *vocoder* outputs 50 voice frames per second
- Voice frames are carried by IP packets
- An 8 Kbps vocoder (e.g. EVRC) produces a non-empty voice payload every frame.
- Payloads (voice + control bits) range from 10 to 40 bytes every 20 msec.
- Network protocol headers add more bytes.
- Vocoder output is modeled by a two-state Markov Chain.

Queueing Model



The Protocol Stack

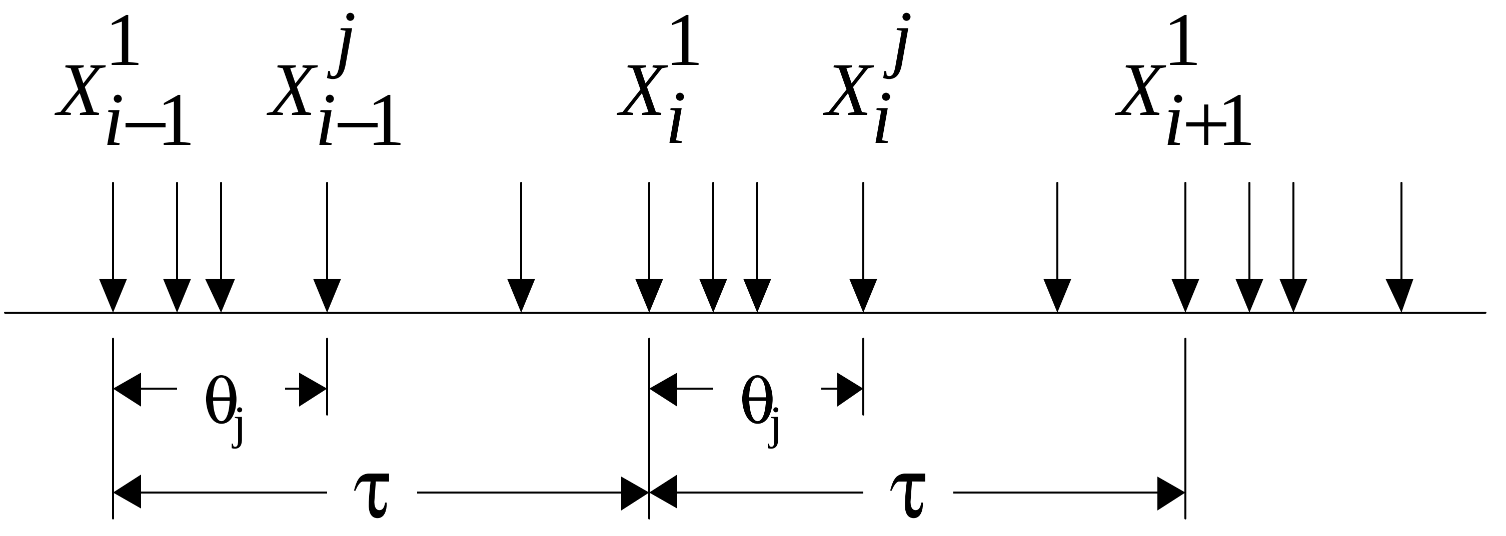


Aggregated Voice Traffic

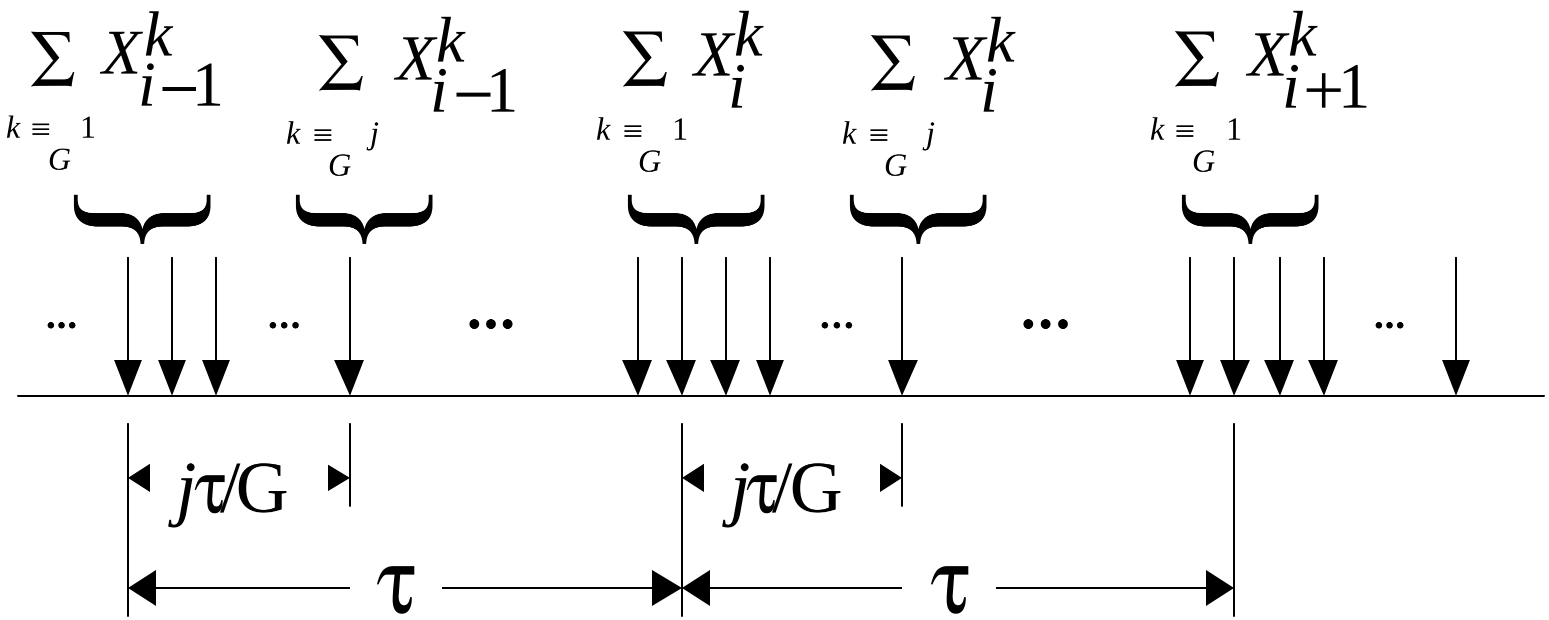
- Consider N voice sources each with period τ.
- The ith source has offset value q_i between 0 and τ .
- The number of bits from the j^{th} source in the i^{th} frame is the random variable $X_{i,j}$.
- The total number of bits arriving at the buffer in the interval $[0, t]$ is then

$$A(t) = \sum_{q_j + i t \leq t} X_{i,j}$$

General arrival process for aggregated voice traffic with random offsets



Arrival Process of Aggregated Voice Traffic: Offset Groups



Voice Sources

We assume the stochastic processes

$$\{X_{i,j}\ , i=0,1,2,\ldots\}$$

are I.I.D. two-state markov chains.

Let l, s^2 and r_k denote the mean, variance and lag k correlation of the process $X_{i,j}$.

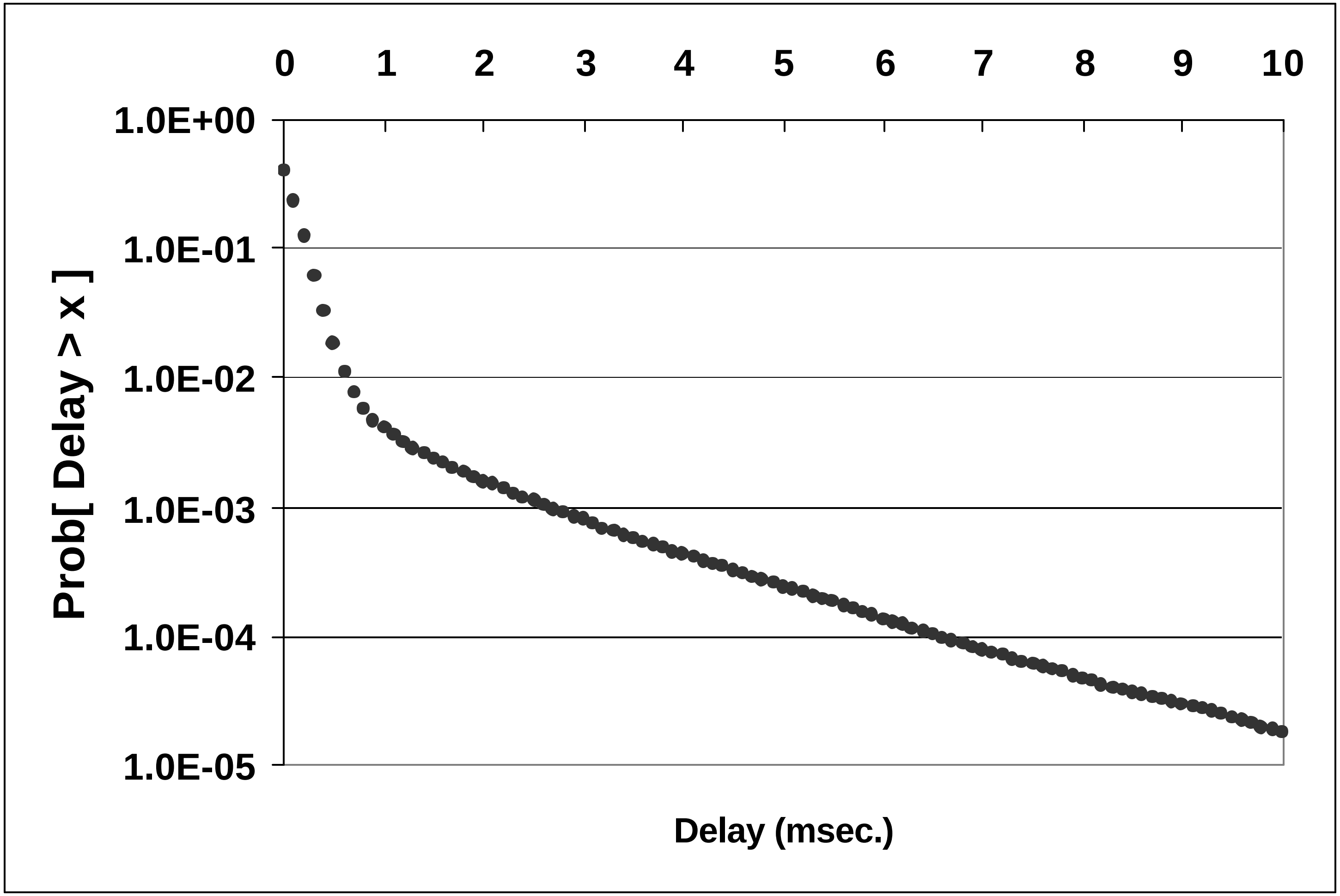
$$E[X_{i,j}] = l, \quad Var[X_{i,j}] = s^2$$

$$r_k = Corr(X_{0,j}, X_{k,j}) \equiv \frac{Cov(X_{0,j}, X_{k,j})}{Var(X_{0,j})}$$

Since $\{X_{i,j}\}$ is a two-state Markov Chain, $r_k = r^k$

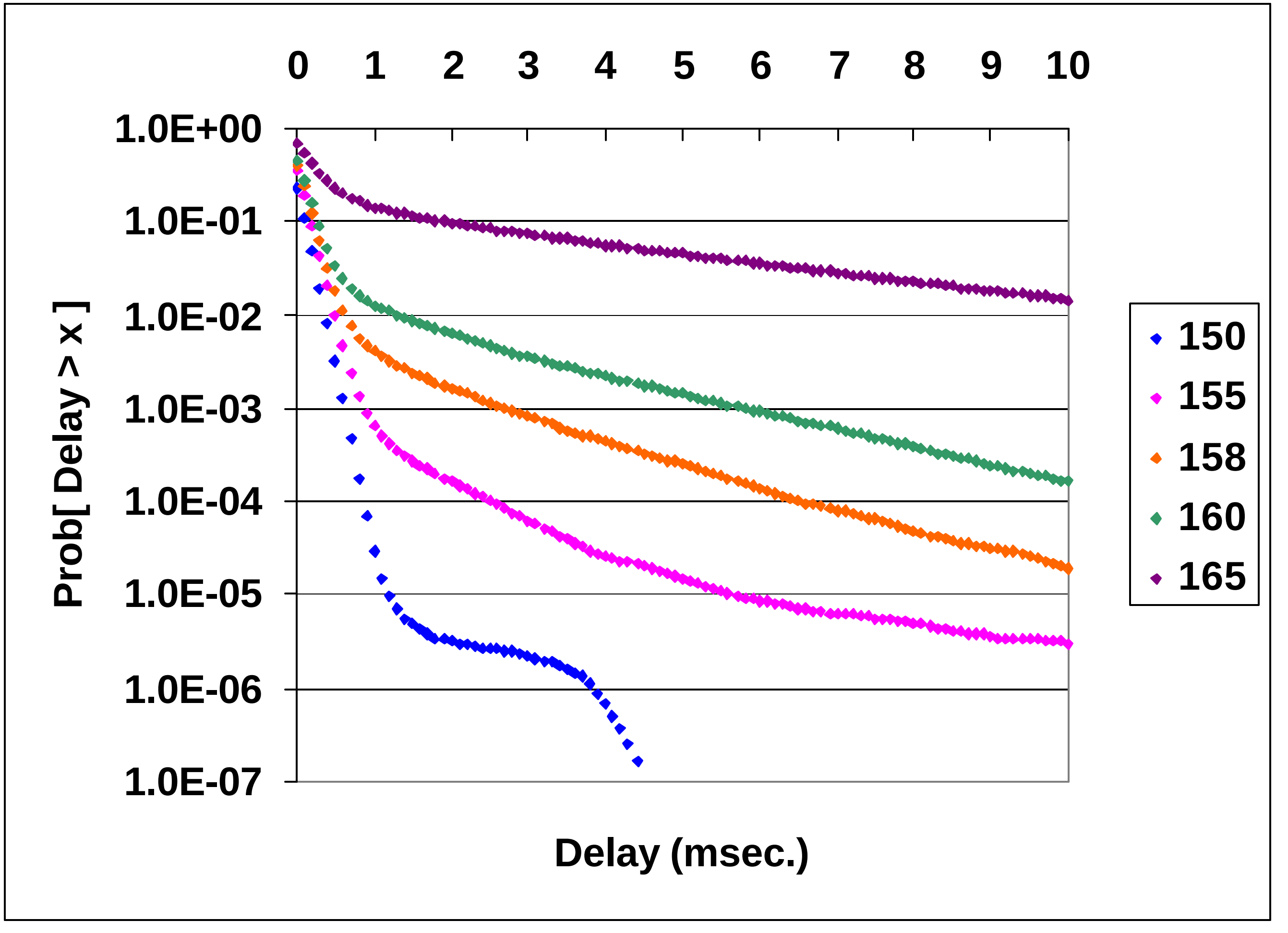
Simulation Results

(157 users over a T1 at 1.536 MbpS)



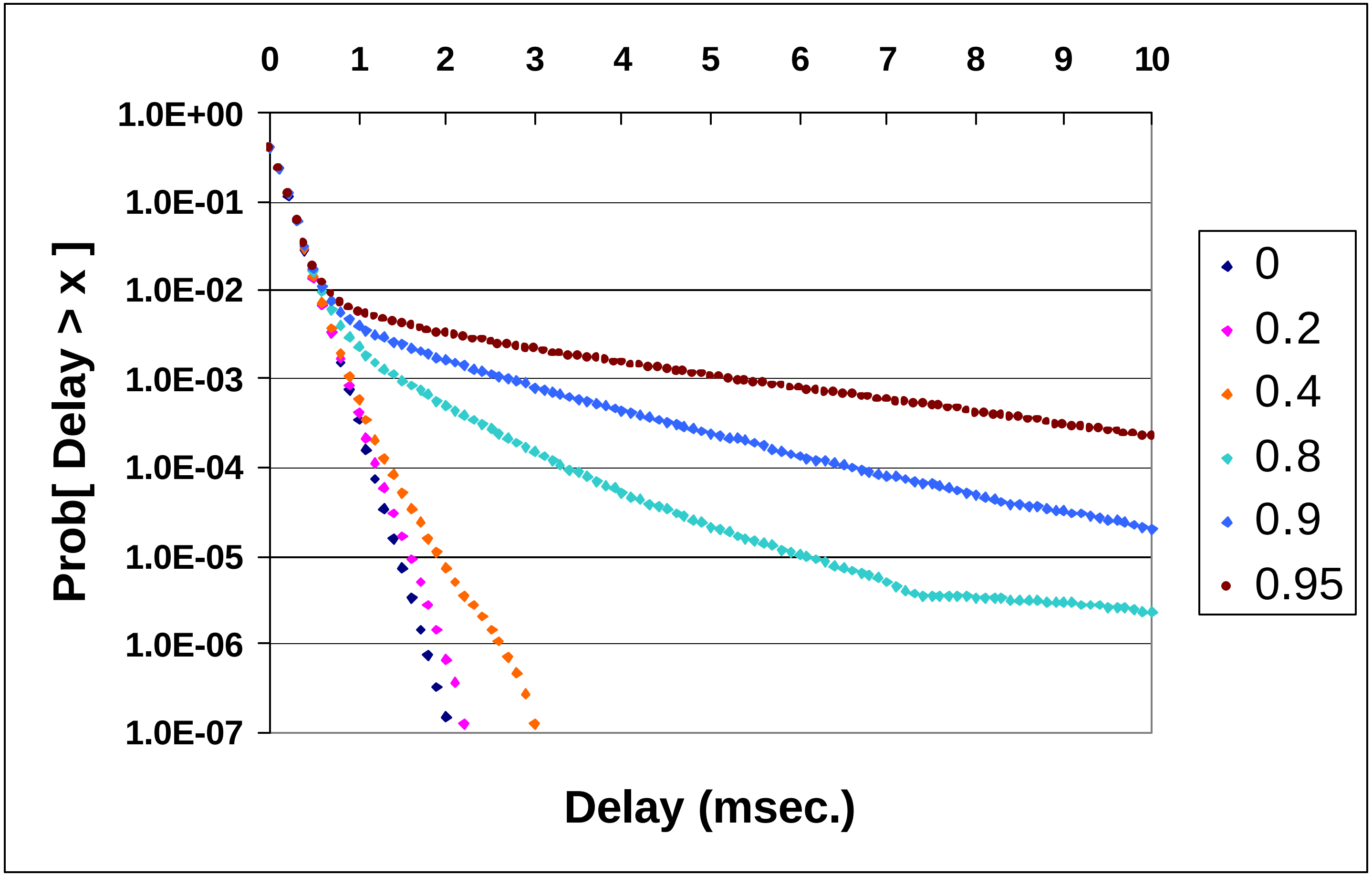
Effect of Loading

number of sessions



Effect of Correlation

one-step correlation shown



Gaussian Traffic Model

$\{ Z_t, t > 0 \}$ is a *centered Gaussian process with stationary increments* iff it is centered and for any $t_1,\ldots,t_k$ and s the random vector

$$(Z_{t_1} - Z_s, \ldots, Z_{t_k} - Z_s)$$

has a multivariate Gaussian distribution that is independent of s.

Gaussian Traffic Model

Arrival traffic (number of bits) in $(0,t]$ is modeled by

$$A(t) = mt + Z_t$$

where

- m is the mean rate
- Z_t is a centered Gaussian process with stationary increments

Denote

$$v(t) \equiv Var(A_t) = E(Z_t^{\ 2})$$

Queue Length

The *queue length process* with server rate c is

$$V_t = \sup_{s \le t}(A_t - A_s - c(t-s))$$
$$= \sup_{s \le t}(Z_t - Z_s - (c-m)(t-s))$$

The tail distribution of the queue length process is

$$P[V_t > x] = P[\sup_{s \le t}(Z_t - Z_s - (c-m)(t-s)) > x]$$

Queue Length Approximation

$$P[V_t > x] = P[\sup_{s \le t}(Z_t - Z_s - (c-m)(t-s)) > x]$$

$$\ge \sup_{s \le t} P[Z_t - Z_s - (c-m)(t-s)) > x]$$

$$= \sup_{u \ge 0} P[Z_u - (c-m)u > x]$$

$$= P(Z_{u*} - (c-m)u* > x]$$

$$= \overline{\Phi}\left(\frac{x+(c-m)u*}{\sqrt{v(u*)}} \right) \approx \exp\left(-\frac{(x+(c-m)u*)^2}{2v(u*)} \right)$$

Queue Length Approximation

The quantity u^* can be found as the value of t that minimizes

$$\frac{(x+(c-m)t)^2}{v(t)}$$

Ref: Ilkka Norros, *Most probable path techniques for Gaussian queueing systems*, Network 2002, Pisa, Italy

Scaling Factor

- A *scaling factor* is required to approximate $P[\ Q > 0]$.
- For the case of k offset groups, we use

$$\mathrm{P}[\,\mathrm{N}(Nl\,/\,k,\ s^{2}/\,k) > m\,/\,k\,]$$

- For random offsets use

$$0.5 = \lim_{n\to\infty} \mathrm{P}[\,\mathrm{N}(Nl\,/\,n,\ s^{2}/\,n) > m\,/\,n\,]$$

The Variance of *A(t)*

Define a simplified arrival process, $A^0(t)$, for a single normalized voice source as

$$A_t^0 = \sum_{i=1}^{\lfloor t \rfloor} X_i + I_{\{u < t - \lfloor t \rfloor\}} X_{\lfloor t \rfloor + 1}$$

where $E(X_i)=0$, $Var(X_i)=1$ and $Cov(X_i, X_j)=\rho_{(i-j)}$

For N “real” voice sources, we have

$$A_t =_D N l t + N s^2 A_t^0$$

The Variance of *A(t)*

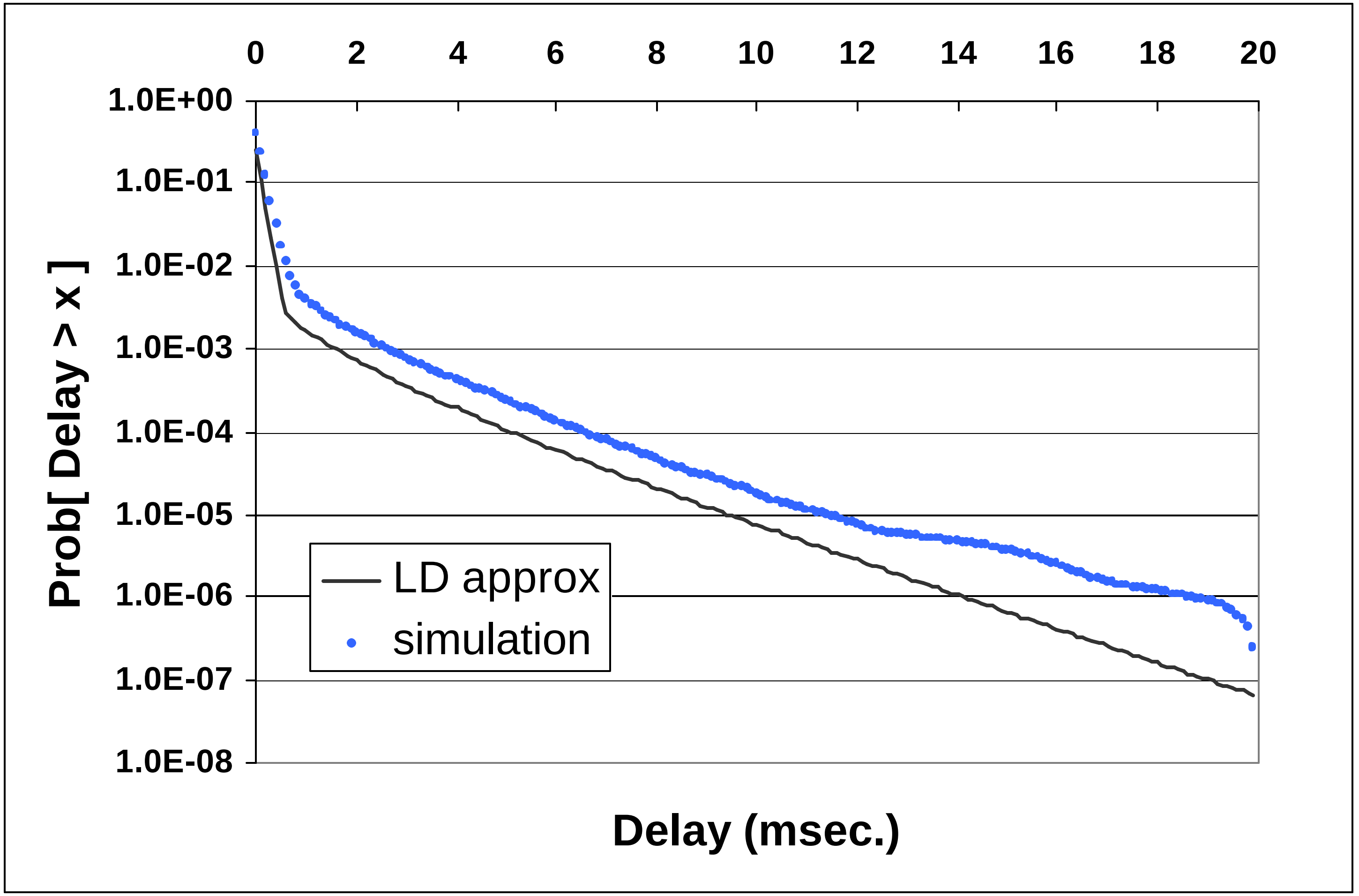
$$E(A_t^0)^2 = E\left(\left(\sum_{i=1}^{\lfloor t \rfloor} X_i\right)^2 + 2I_{\{u<t-\lfloor t \rfloor\}} X_{\lfloor t \rfloor+1} \sum_{i=1}^{\lfloor t \rfloor} X_i + I_{\{u<t-\lfloor t \rfloor\}} X^2_{\lfloor t \rfloor+1}\right)$$

$$= tE(X_1^{\ 2}) + 2\sum_{i=1}^{\lfloor t \rfloor}(t-i)X_1 X_{i+1} - t^2 (EX_1)^2$$

$$= t + 2\sum_{i=1}^{\lfloor t \rfloor}(t-i)\, r_i$$

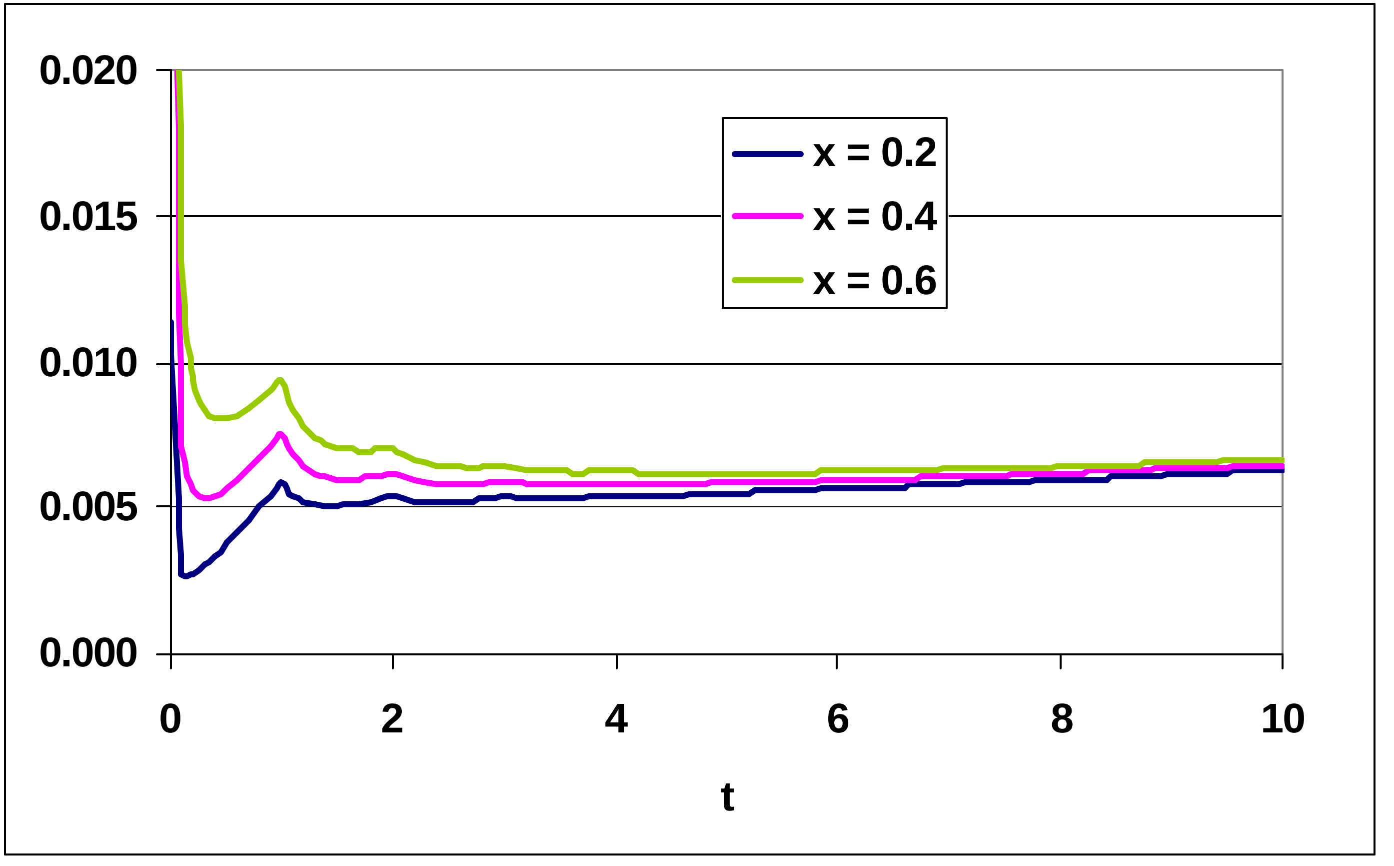
The Large Deviations Approximation

(157 users over 1.536 Mbps)



Why the hockeystick?

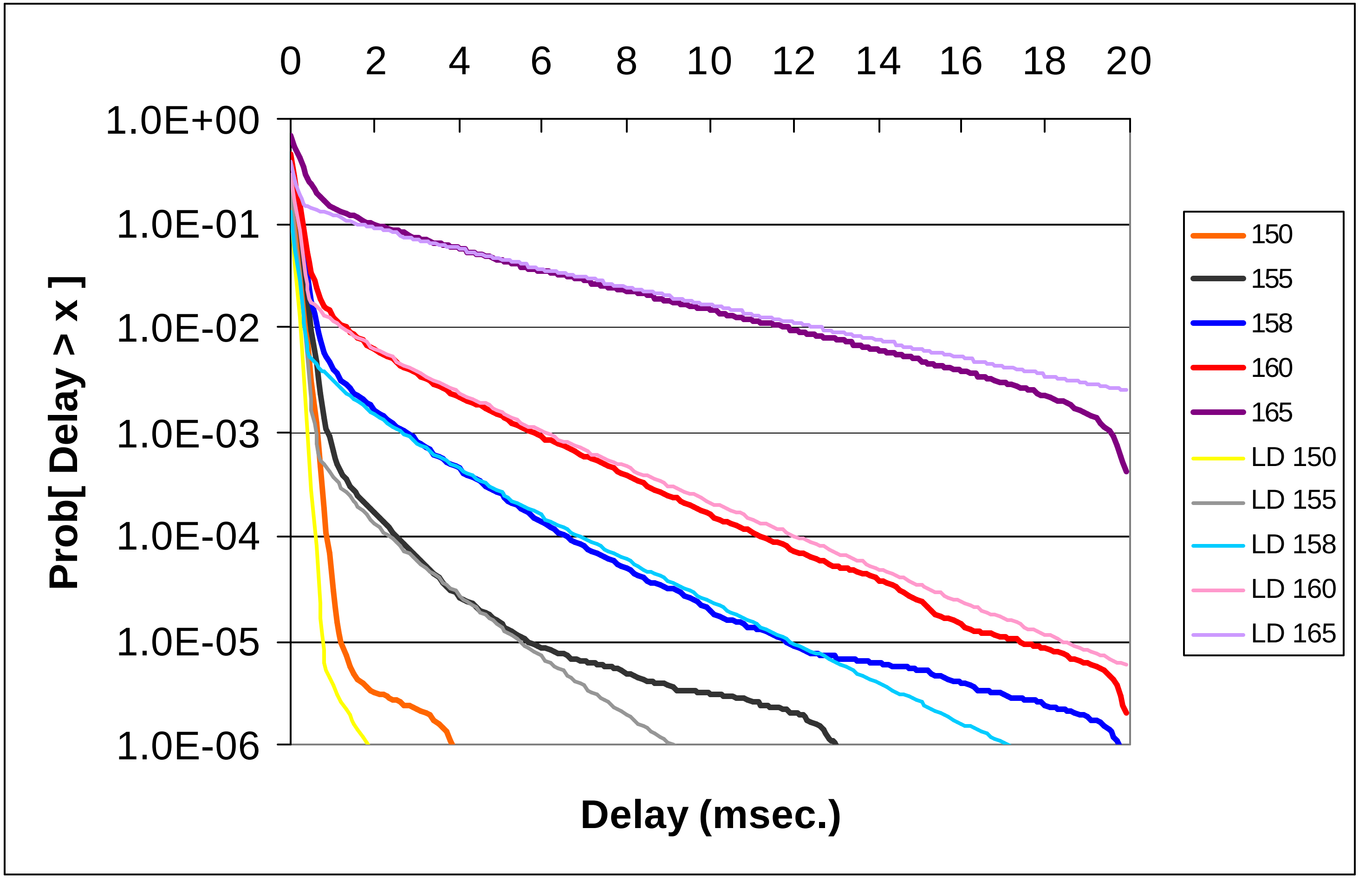
$$\frac{(x+(c-m)t)^2}{v(t)}$$



Intuition

- The delay of a packet that arrives during a busy period that has lasted less than 20 msec will not depend on the correlation structure of the arrival process.
- “Short” busy periods are common and lead to short delays
- The delay of a packet that arrives during a busy period that has lasted longer than 20 msec will depend on the correlation structure.
- “Long” busy periods are somewhat rare and lead the longer delays .

Validation


0
2
4
6
8
10
12
14
16
18
20
1.0E+00
1.0E-01
1.0E-02
1.0E-03
1.0E-04
1.0E-05
1.0E-06
Prob[Delay > x]
Delay (msec.)
150
155
158
160
165
LD 150
LD 155
LD 158
LD 160
LD 165

Heavy Traffic Analysis

- The system evolution is given by

$$Q(t) = Q(0) + A(t) - c\int_0^t 1_{\{Q(s)>0\}}\, ds$$

- We consider a scaled version of Q

$$Y_n(t) = \frac{Q(nt)}{\sqrt{n}}$$

- Assume $Y_n(0)=Q(0)=A(0)=0$

Heavy Traffic Analysis

- Consider a sequence of systems indexed by n in which the mean arrival rate in the nth system is $N\lambda_n$
- The scaled queue length can then be written

$$Y_n(t) = \frac{\tilde{A}(nt)}{\sqrt{n}} + \sqrt{n}(N l_n - m)t + \Lambda_n(t) + O(n^{-1/2})$$

Heavy Traffic Analysis

$$Y_n(t) = \frac{\tilde{A}(nt)}{\sqrt{n}} + \sqrt{n}(Nl_n - m)t + ?_n(t) + O(n^{-1/2})$$

where $\dfrac{\tilde{A}(nt)}{\sqrt{n}} \Rightarrow Ns^2(1 + 2\sum_{k=1}^{\infty} r_k)B(t) \equiv N\hat{s}^2 B(t)$

If we assume $\sqrt{n}(Nl_n - m) \rightarrow (Nl - m)$ then

$$Y_n(t) \Rightarrow Y(t) = N\hat{s}^2 B(t) + (Nl - m)t + ?(t)$$

where $?(t) = \max(0, \sup_{s \le t}(-(N\hat{s}^2 B(t) + (Nl - m)s))$

So *Y(t)* is Reflected Brownian Motion

Stationary Density

From this analysis one concludes that *Y(t)* is a Reflected Brownian Motion with Drift and has a stationary density, *f*, given by

$$f(y) = \frac{2(m - Nl)}{N\hat{s}^{2}} \exp\left(- \frac{2y(m - Nl)}{N\hat{s}^{2}} \right)$$

The exponent is the same as in the Gaussian LD approximation for sufficiently large y and for small y.

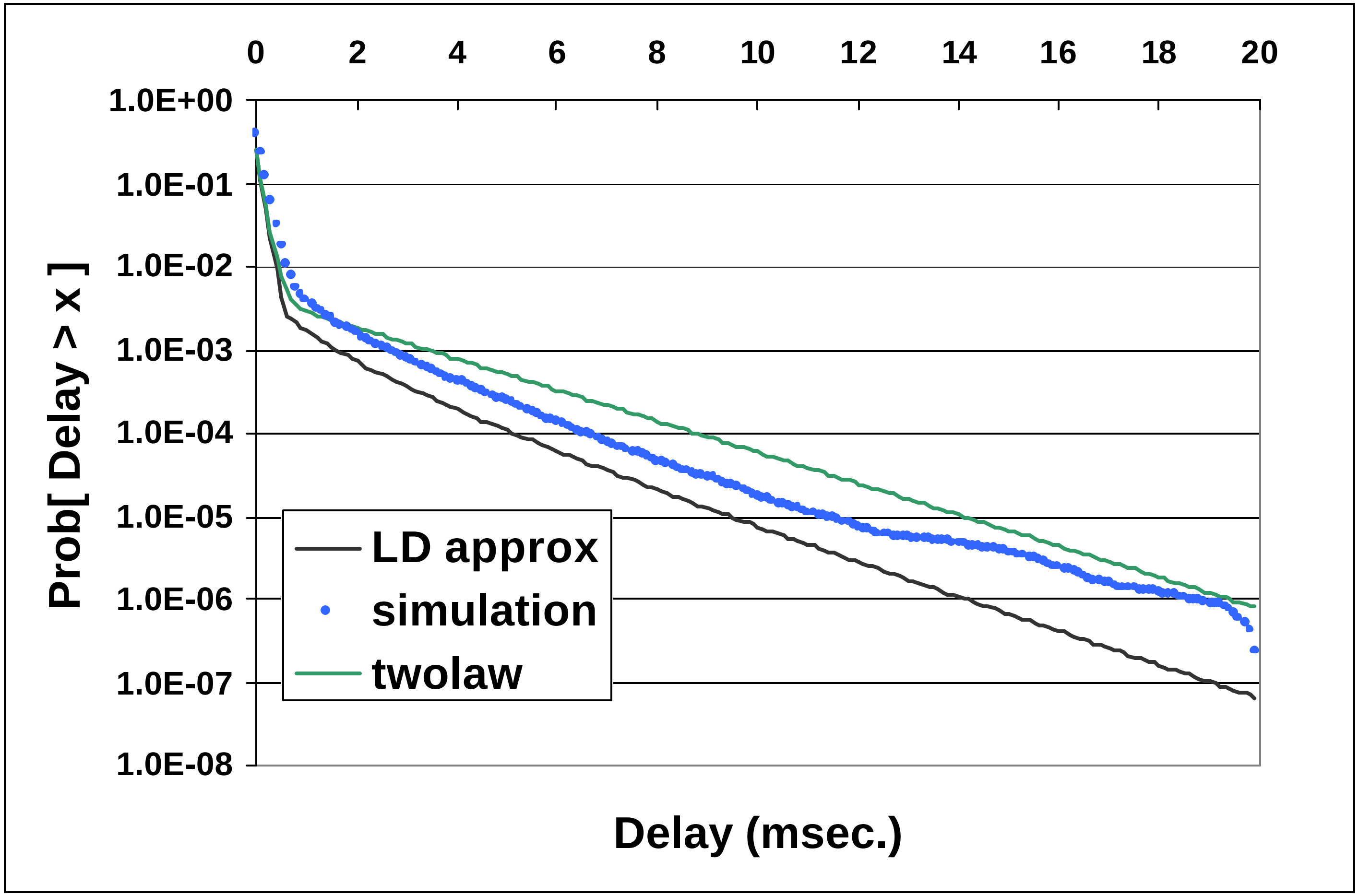
Waiting Time Approximation

The heavy traffic analysis leads us to an approximation which is the sum of two scaled exponential distributions.

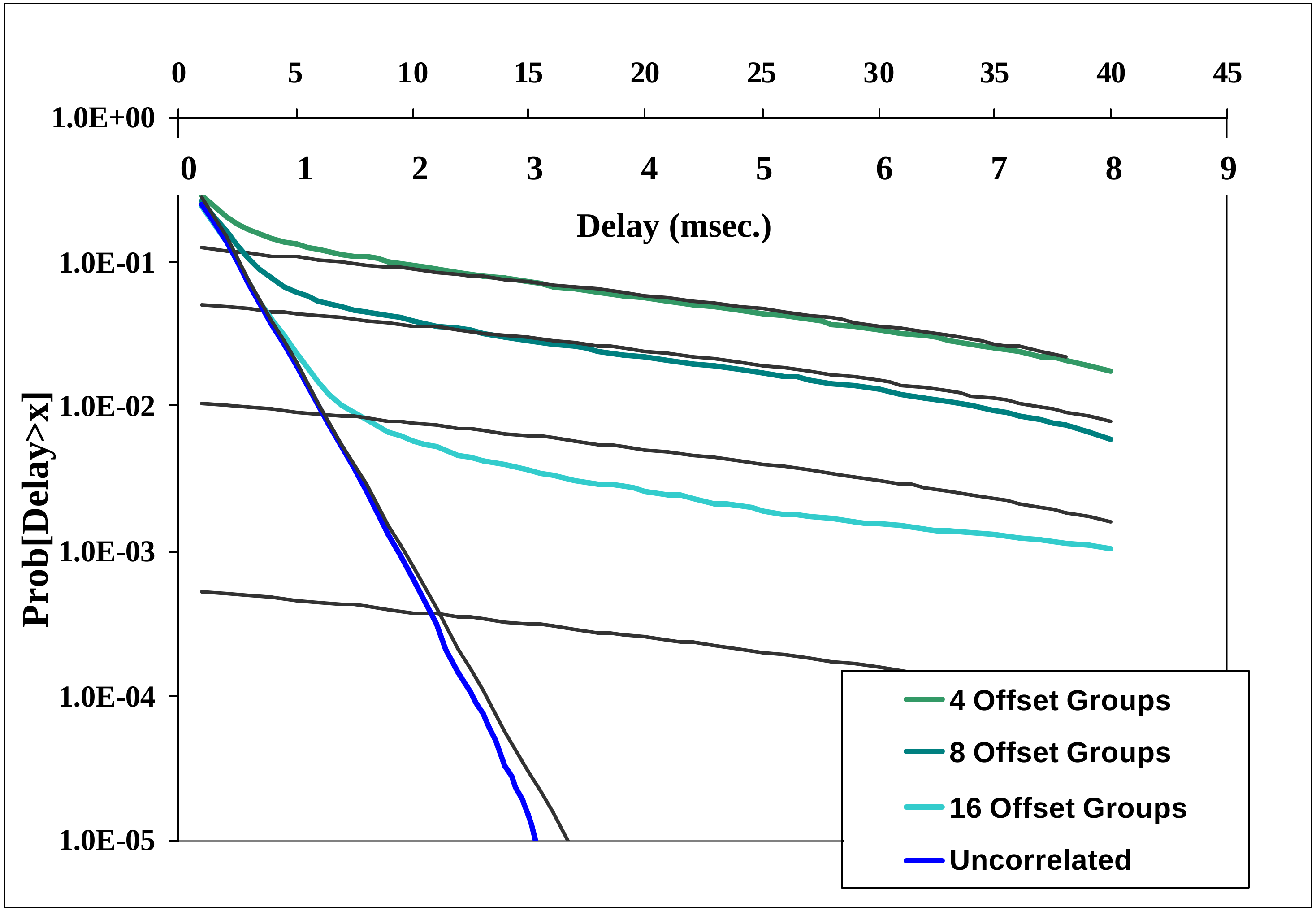
$$P[W_k > t] \approx \mathrm{P}[\mathbf{N}(Nl, N\hat{s}^2) > m]\exp\left(-\frac{2tm(m - Nl)}{N\hat{s}^2}\right)$$
$$+\mathrm{P}[\mathbf{N}(Nl/k, Ns^2/k) > m/k]\exp\left(-\frac{2tm(m - Nl)}{Ns^2}\right)$$

Validation

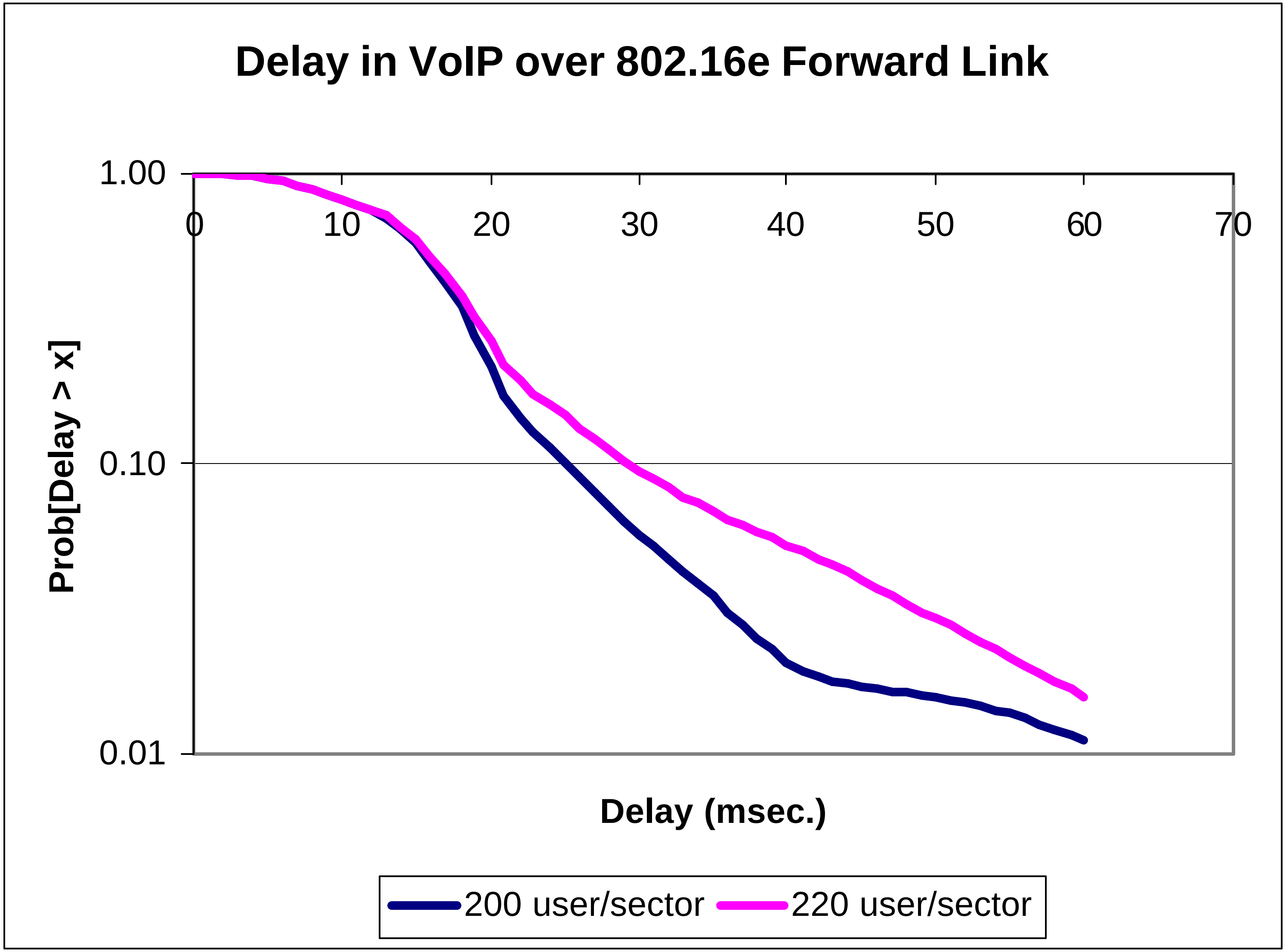
(157 users over 1.536 Mbps)



Validation



0
5
10
15
20
25
30
35
40
45
1.0E+00
0
1
2
3
4
5
6
7
8
9
Delay (msec.)
1.0E-01
1.0E-02
1.0E-03
1.0E-04
1.0E-05
Prob[Delay>x]
4 Offset Groups
8 Offset Groups
16 Offset Groups
Uncorrelated

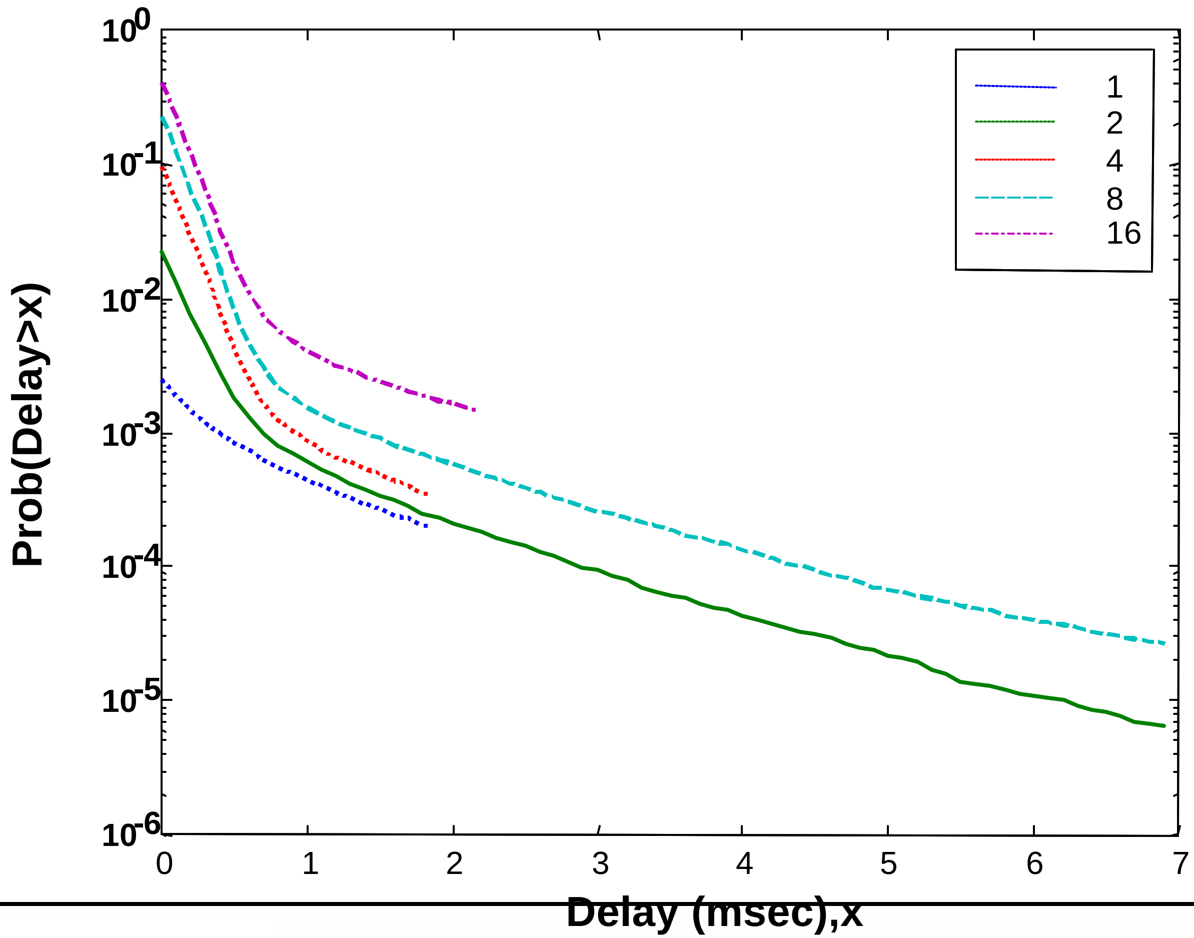


Delay in VoIP over 802.16e Forward Link
1.00
0.10
0.01
0
10
20
30
40
50
60
70
Prob[Delay > x]
Delay (msec.)
200 user/sector
220 user/sector

Ongoing work

- Models of framed voice+data (e.g., I. Norris' approach)
- Network sizing tools for system engineers
- Scheduling and admission control implications

Impact of the Number of Offset Groups on Waiting Times



Packetized Voice in Cellular Telephony (numerical example: EVRC)

Let X_i = number bits in the i^{th} voice frame (before protocol overhead)

$$E[X_i] = 162 \text{ bits}$$

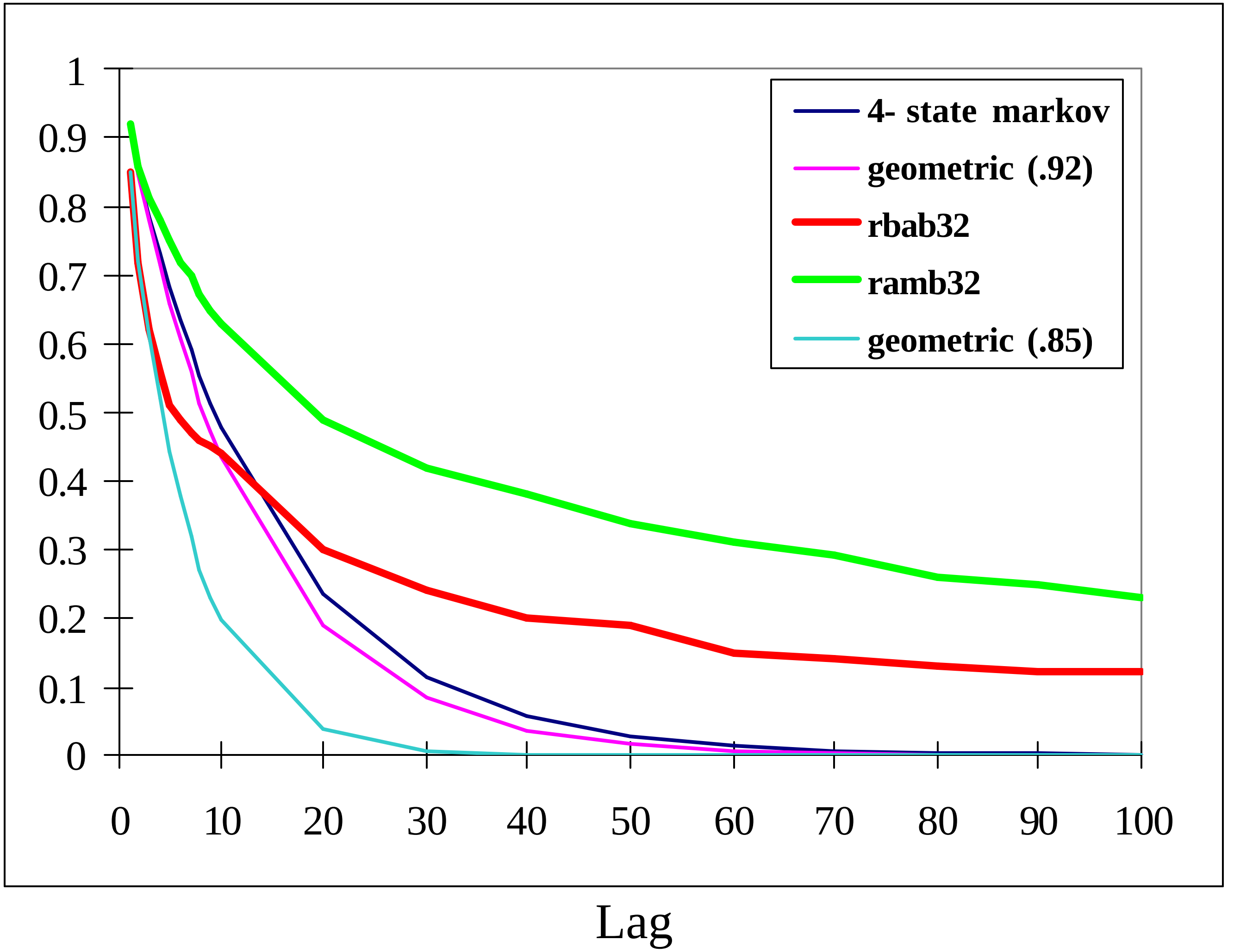
$$\sqrt{Var[X_i]} = 74 \text{ bits}$$

$$Corr[X_i, X_{i+j}] = r^j, \; j \geq 0$$

Typically, $0.90 \leq r \leq 0.99$

X_i is a two-state markov chain

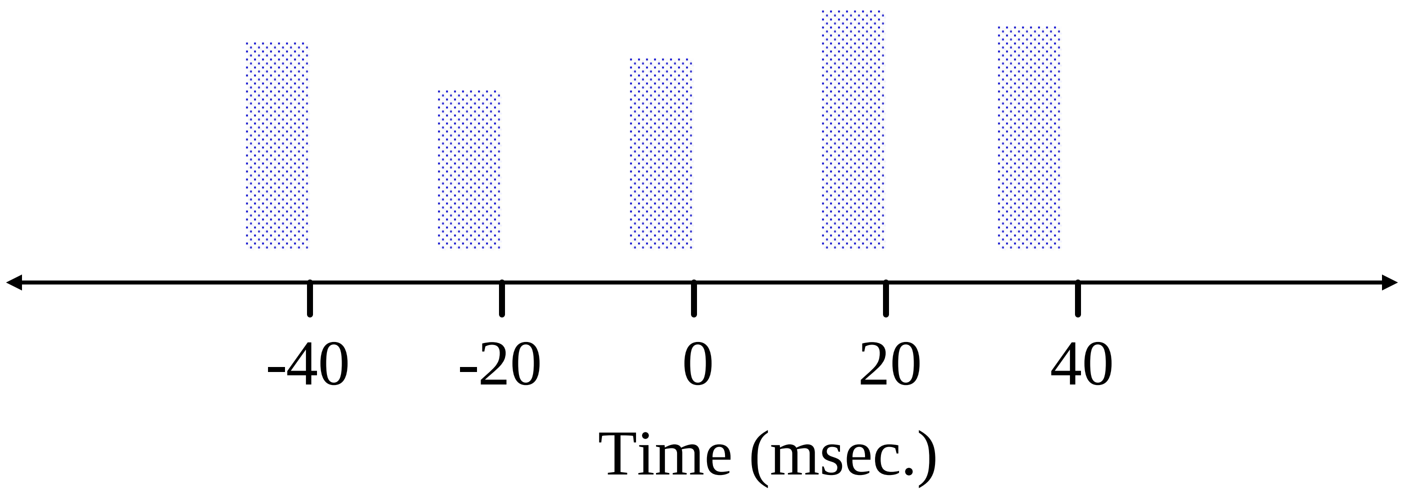
Correlation Structure of Vocoder Output



Voice Arrival Process

We approximate the aggregate voice process $\{Z_i, i=1,\ldots\}$ as a (truncated) Gaussian process (a “discrete Ornstein-Uhlenbeck” process)

$$\tilde{Z}_{i+1} = \mathbf{N}(m + r(\tilde{Z}_i - m), s^2(1 - r^2))$$



Traffic Shaping

- Voice traffic is shaped in 3G systems by dividing voice flows into *offset groups*
- Note:
  - **traffic within an offset group is correlated**
  - **traffic between offset groups is independent**