Overview: Traffic Modeling and Queueing Analysis

Mark Daniel Ward
joint with William S. Cleveland and Jin Xia
Department of Statistics
Purdue University

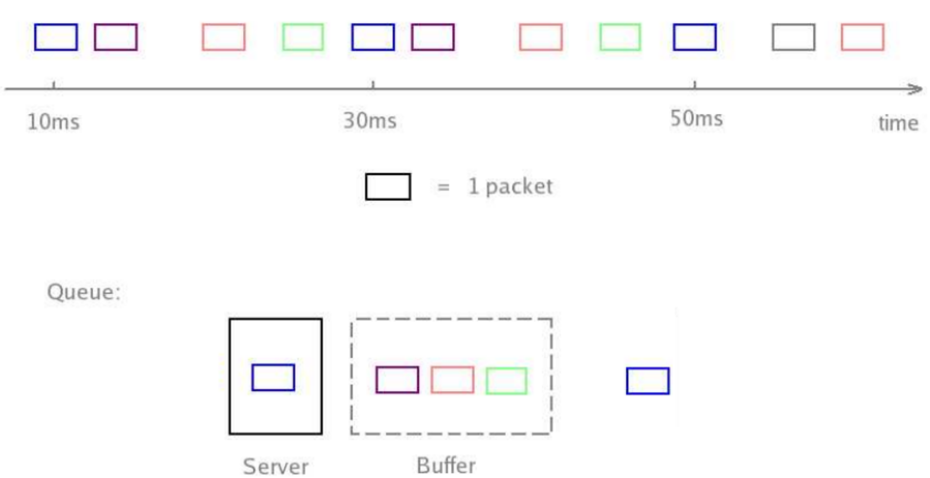
MAASCOM
Control of Multi-Scale Networks

Data Collection

- Global Crossing VoIP (completed)
  - partnership with Global Crossing (GBLX) and Cariden, Inc.
  - 100 mbps Ethernet link on the GBLX IP network
    - voice gateway
    - edge router
  - two days of traces
  - times stamps plus UDP, RTP, and IP headers
  - both directions of 328,631 calls
  - 1/2 terabyte of data
- Properties of calls
  - calls have 1600 bits per packet
  - each call has 1 packet every 20 ms
  - call arrival process is Poisson
  - call durations are Pareto
  - silence suppression and packet-level traffic further complicate the model

VoIP Semi-Empirical Model

- First-In-First-Out, 1-Server Queue, Infinite Buffer
- Superimposed Packet Process
  - aggregated calls in a semi-empirical model
  - call arrivals follow a randomly generated, homogeneous Poisson process
  - empirical packet traces chosen from database



Simulation Design

- Run simulations at various arrival rates and link speeds (service times)
  - traffic rates 1 mbps, 3.16 mbps, 10 mbps, 31.6 mbps
  - utilization rates $0.1, \ldots, 0.9$ (traffic rate / link speed)
  - replicate simulation on three servers
- Packet Delay
  - Calculate delay for each packet in the queue
  - Treat delay as “data” that we seek to characterize by models
- Theory vs Practice
  - use approximations of traffic models to derive queueing properties
  - formulate mathematical model
  - check if derived results are good approximations to the delay models
- Implementation
  - original simulation in S-PLUS, 1 packet trace per file
  - faster simulation using database, UNIX utilities, Perl, and S-PLUS for statistical analysis
  - future plans: distributed data, parallel processing, Hadoop, etc.

Local Poisson Approximation

- Extremely Short Time Intervals (< 20ms)
  - packet arrivals follow a Poisson process
  - provides motivation for using M/D/1 queueing model
- M/D/1 model
  - inter-arrival times between packets are exponential
    - i.e., a Poisson number of arrivals in a fixed interval
  - service time is deterministic
  - 1 server for the queue
- Density of M/D/1 model
  - traffic rate T, utilization rate U
  - number of jobs remaining unprocessed in the queue is Y
  - limiting distribution is

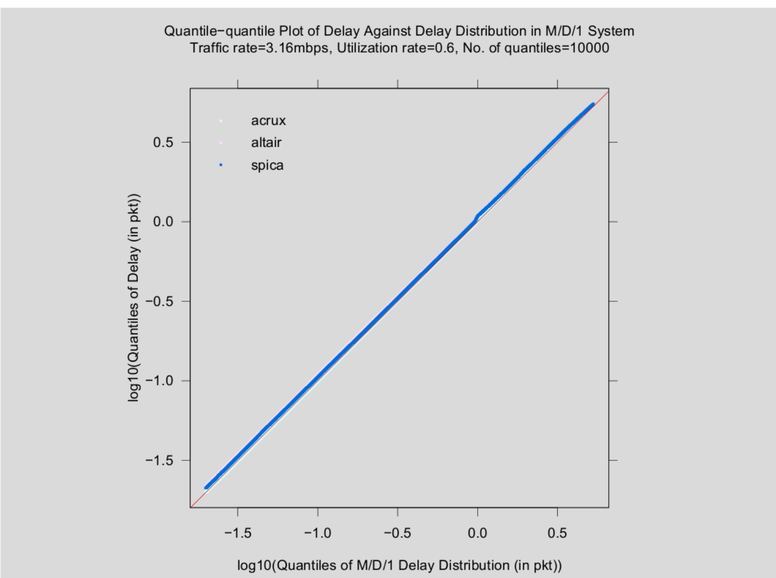
$$F(y) = P(Y \leq y) = (1-U)\sum_{n=0}^{\lfloor y \rfloor} \frac{[U(n-y)]^n}{n!} e^{-U(n-y)}$$

Local Poisson Approximation

- (Relatively) Short Time Intervals
  - packet arrivals approximately follow a Poisson process
  - complicated by regularity of packet arrivals (every 20 ms) with each call
  - higher traffic rates diffuse the effect of the regularity of packet arrivals within a call
- (Relatively) Long Time Intervals
  - long-range dependence (LRD) appears to make a significant impact on queueing
  - at higher traffic rates, higher utilization rates are required before seeing LRD effect

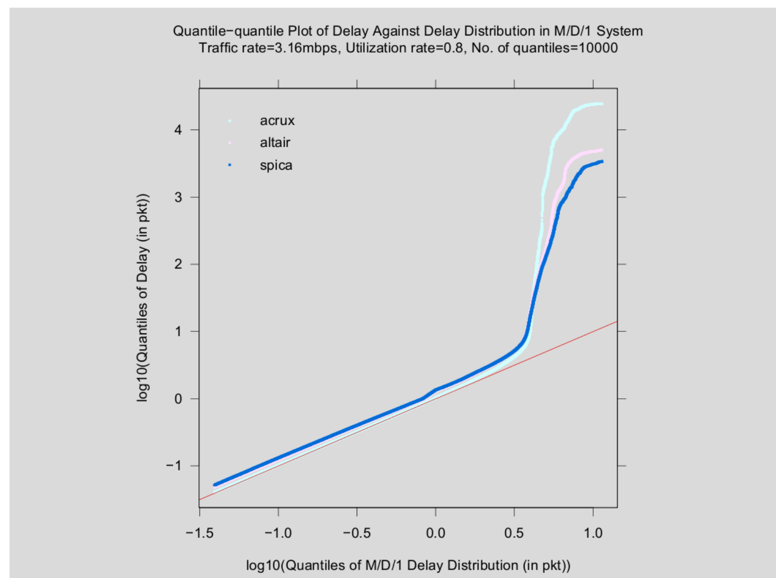
Quantile-Quantile plot of Delay vs M/D/1 Delay (1st Example

- Q-Q Plot of Queue Length vs M/D/1 Delay, Rate 3.16 mbps, Utilization 0.6



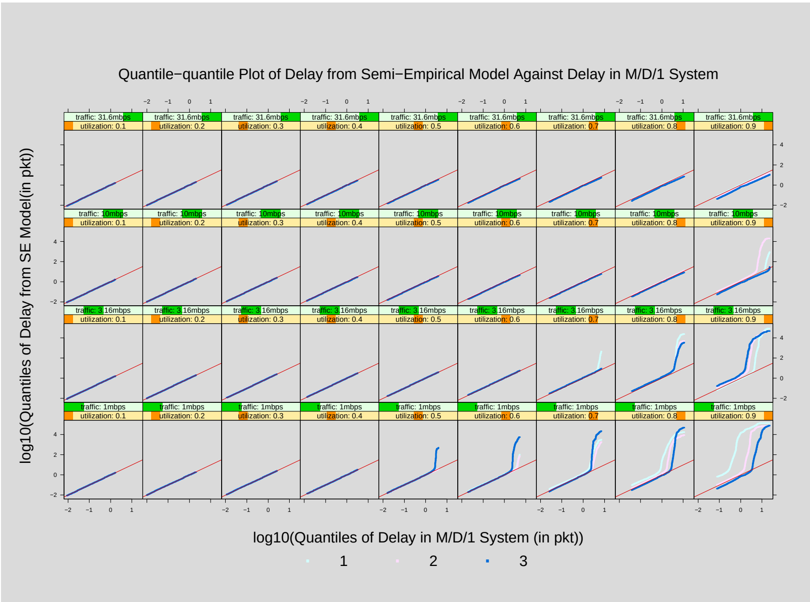
Quantile-Quantile plot of Delay vs M/D/1 Delay (2nd Exampl

- Q-Q Plot of Queue Length vs M/D/1 Delay, Rate 3.16 mbps, Utilization 0.8



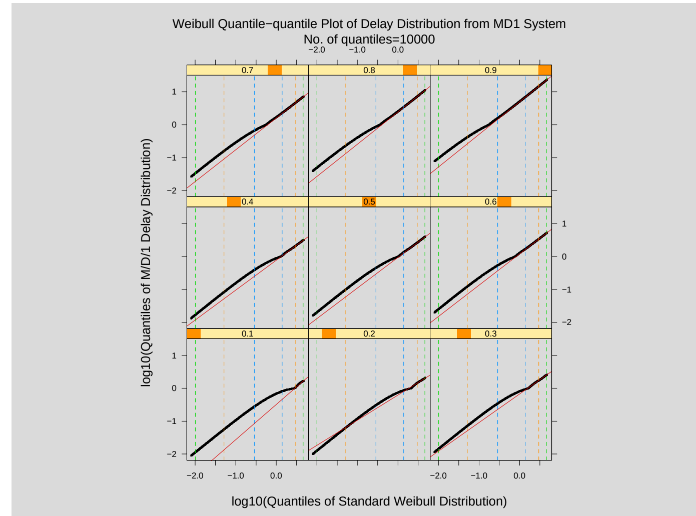
Trellis Display of Delay vs M/D/1 Delay

- Q-Q Plots for 4 traffic rates, 9 utilizations, and 3 servers



Q-Q plot of M/D/1 Delay vs Weibull

- Compare M/D/1 Delay Distribution vs Weibull Distribution
  - trellis for 9 utilizations
  - M/D/1 Delay Distribution is Weibull in the tail



Provisioning

- Quality of Service (QoS)
  - determine appropriate traffic rate
    - performance is measured by packet delay and jitter
    - provision traffic to keep QoS variables below standards with high probability
    - maintain high performance
  - seek to characterize and clarify the nature of multiplexing gains
- Global Crossing Company's VoIP Delay and Jitter Requirements

end-to-end delay (domestic)	$\leq$	150 ms
end-to-end delay (international)	$\leq$	184 ms
maximum absolute end-to-end jitter	$\leq$	9.875 ms
average end-to-end jitter (continental)	$\leq$	2 ms
average end-to-end jitter (trans-continental)	$\leq$	5 ms

Mark Daniel Ward's Research Interests

- Analysis of Algorithms and Data Structures
  - probabilistic methods
  - combinatorial methods
  - analytic (complex) methods
- Asymptotic Analysis
  - sequences and trees
    - suffix trees
- Game Theory
  - combinatorial game theory
  - von Neumann / Nash game theory
- Information Theory
  - data compression
- Integrating Algorithmic and Computational Thinking into the Statistics Curriculum