EP2210 – Performance evaluation of communication networks

Viktoria Fodor
Laboratory for Communication Networks
School of Electrical Engineering, KTH

Course homepage
https://www.kth.se/social/course/EP2210/

Fall 2015 (HT 2015)

EP2210 – Performance evaluation of communication networks

Course objectives:

- Advanced networking course
- Discuss mathematical modeling in some main areas of networking
  - Learn techniques to address performance related questions
  - Discuss some of the significant results – and read the original papers
  - Improve our "paper reading" (and writing) skills

Topics

1. Traffic modeling

2. Multiple access protocols

3. Congestion control

Can I use simple "random" packet arrival to evaluate my protocol?

The random access control I have implemented has zero throughput... what is going on?

What throughput should I expect for my TCP session?

Topics

3. Scheduling

4. Fairness

5. Multimedia communication

How difficult scheduling should I implement to satisfy flows with different needs?

Is my network fair? What is fairness, by the way? Equality?

Should I add redundancy, or should I retransmit? Or maybe I should not even try…

Course setup

- Scheduled activities:
  - 12 lectures of 2 hours
  - make-up test opportunity, right after the last lecture
  - project presentations
- 2 lectures per subject
  - first lecture – introduction and simple models
  - second lecture – advanced models,discussion of papers, phd student presentations
- Midterm tests (5 tests altogether, lectures 3, 5, 7, 9, 12)
- Home assignments (3 home assignments altogether, submitted at lectures 6, 9, 12)
- Project

Requirements

- Read all the papers
  - covering the lecture and for home reading
- Home assignments
  - questions to answer
  - numerical examples (e.g., matlab)
  - independent solutions, submit one paper copy at the lecture or at STEX
  - tell me in advance if you can not submit on time (minus points)
- Tests
  - ca. 20 minutes
  - questions on the lecture material and about the papers (open book/computer)
  - make-up test after the course (missed or weak results)

Requirements

- Project
    - in groups of ca. 2 students
    - subject selected from subject list or on your own (discuss with the instructor)
    - comparative review of 3-5 papers in the area
    - written report of 4-5 pages
    - presentation of the project
    - schedule: defined later
    - groups: defined later by the instructor based on the study results in the course….
    - good reports from earlier years are available on the web

Grading

- Tests: 50%
- Home assignments: 30%
- Project 20% (same for all project members)
  - detailed on the web-page under Projects

- Grading guidelines (approx):
  - 90%:-A, 80%-B, 70%-C, 60%-D, 50%-E, 45%-Fx

Requirements – graduate students

- Paper presentation (for 9ECTS)
  - select a lecture topic as soon as possible
  - ca. 20 minutes presentation on one of the lectures (second lecture of a topic)
  - short meeting right after todays lecture abut the details

- Small project – during or after the course (for +3ECTS)
  - select a lecture topic
  - prepare a small simulator to support a mathematical model or problem definition, the simulator could be used for demonstration
  - see examples under Course material (will be updated soon)

Traffic theory - Traffic models

- Topics:
  - Traffic modeling – traffic objects
  - Markov processes recall
  - Traffic models: markovian and non-markovian models

- Lecture material:
  - A. Adas, "Traffic models in broadband networks," IEEE Communications Magazine, July 1997.
  - J. Roberts, "Traffic theory and the Internet," IEEE Communications Magazine, January 2001.
  - V. Frost, B. Melamed, "Traffic modeling for telecommunications networks", IEEE Communications Magazine, March 1994.
  - I. Kaj, „Stochastic modeling", 5.2.2-5.3.1.

Teletraffic theory

- Teletraffic theory:
  - to model dynamic resource sharing systems
  - to explain the traffic-performance relation

Traffic demand

stochastic

Theory of stochastic processes

Can be stochastic!

stochastic

Shared resources

Realized performance
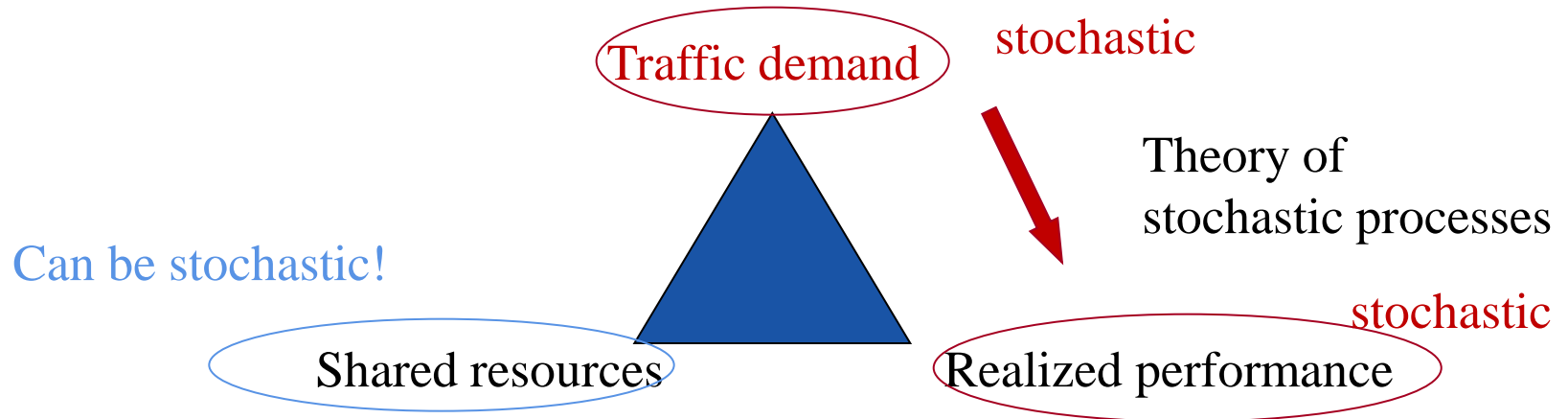
- Traffic: arrival intensity, holding time, packet length (distribution or moments)
- Resources: link bandwidth, router buffer, server capacity
- Performance: utilization, loss, delay, delay variation, perceptual quality

Traffic modeling

- To describe the network traffic demand
- Statistical characterization
- Traffic objects
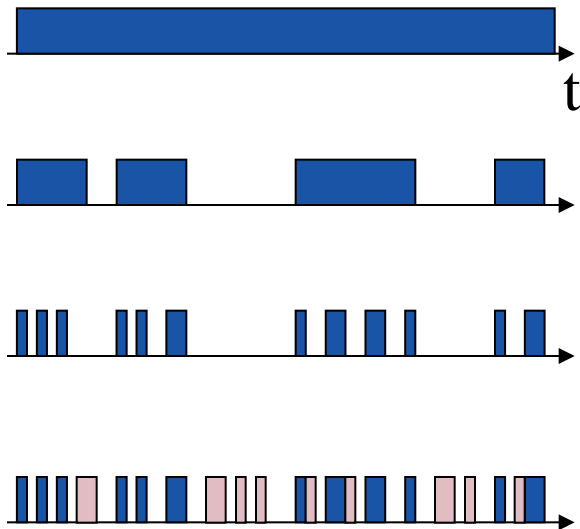
- Flow (one instance of communication, TCP or UDP session) — Skype call
- Burst (Active/passive periods) — Talk/listen
- Sequence of packets — IP packets
- Multiplexed packets — IP packets at a router

Traffic modeling

- Packet level – characteristics of the sequence of packets
  - packet arrival process
    - according to some stochastic/deterministic arrival process (e.g. Poisson arrival at a router…)
    - saturated source model: there is always packet to send at the source
  - packet size distribution
- Flow level (burst level is similar too, but rarely used):
  - flow arrival process
    - e.g., flows from all the laptops in a WLAN are generated according to a Poisson process
  - flow duration distribution
  - flow characteristics – how traffic is generated within a flow

Flow characteristics

Models that describe the distribution of the sequence of packets for a flow level model

- **packet scale model**
  - arrival process and packet size distribution
  - queuing theory
  - used typically in this course
  - may lead to very complex models on flow level
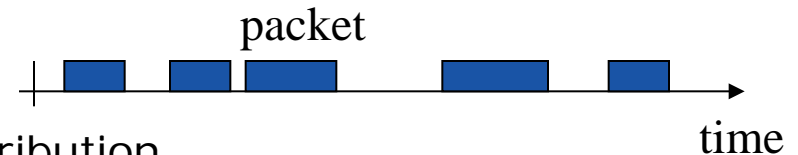
- **fluid models**
  - transmission as a continuous stream
  - parameter: flow rate r(t)
  - system of differential equations
  - often more tractable on the flow level

packet

time

rate

time

rate

burst

time
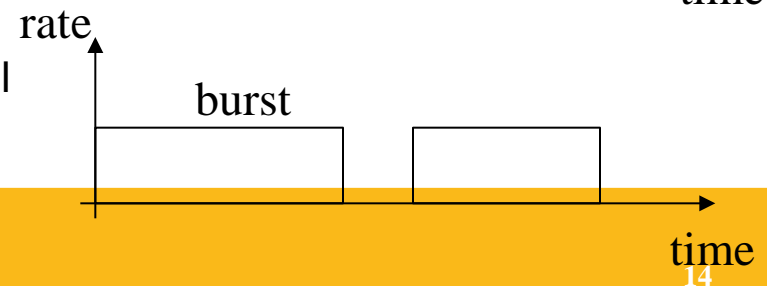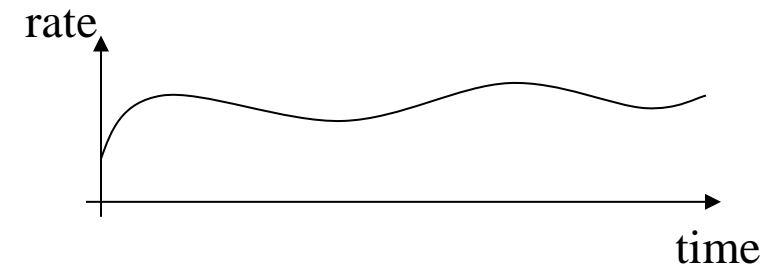
Flow types - Terminology

- Flow - one instance of an application
  - Reasonable to classify according to application types
- Elastic flow
  - The application requires the transmission of a given amount of information, some delay is acceptable – that is, transmission is elastic in time
  - E.g., file transfer over TCP
  - Flow characteristics is determined by the transport protocol (e.g., TCP) and the background traffic
- Streaming flow
  - The application has strict delay limitations, late information is dropped
  - E.g., VoIP over UDP
  - Flow characteristics is determined by source characteristics (e.g., coding)

Traffic modeling

- Should we use packet or flow level models in the following problems?
  - buffer dimensioning – sequence of packets
  - error control – loss of individual packets
    - PACKET LEVEL MODELS
  - video rate control
  - routing
    - FLOW LEVEL MODELS

Group work

Should we use packet or flow level models in the following problems? In the case of flow level models, what kind of flow characterization is necessary?

1. What is the probability that a packet collides and therefore needs to be retransmitted when using CSMA/CA protocol?

2. Several Skype calls are using the same communication link. What is the utilization of the link *(utilization={average rate of traffic} / {link transmission rate})*

3. Several flows are multiplexed at a router with limited buffer. What is the probability that consecutive packets of a flow are dropped due to buffer overflow?

Mathematical modeling

- Recall: Markov chains
- Markovian traffic models
- Home reading: non Markovian models

Recall – Markov chains

- Basic tools of queuing theory
- Stochastic process
  - Discrete state space
  - Discrete or continuous time (change of state)
  - Markovian property: the future of the process does not depend on the past, only on the present
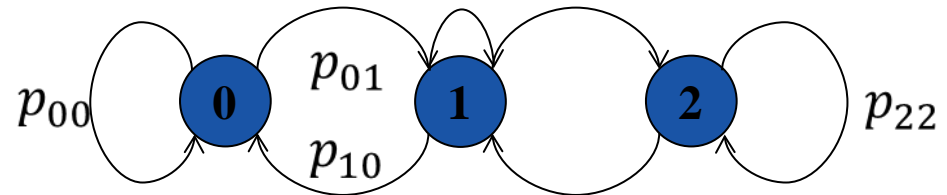
- Discrete time Markov chains
  - State transition probability matrix $\boldsymbol{P} = \{p_{ij}\}$
  - $\underline{p}_{i+1} = \underline{p}_i \boldsymbol{P}$
  - If steady state exists, the stationary state probability is given by $\underline{p} = \underline{p}\boldsymbol{P}$
  - Holding time of a state is geometric with parameter $1 - p_{ii}$ (memoryless)

Recall – Continuous time Markov chains

- **Continuous time Markov chains**
  - State transition is possible at any time
  - State transition intensity matrix $\mathbf{Q} = \{q_{ij}\}$, $q_{ii} = -\sum q_{ij}$
  - $\underline{p'(t)} = \underline{p(t)}\mathbf{Q}$
  - If steady state exists, the stationary state probability is given by $\underline{0} = \underline{p}\mathbf{Q}$
  - Holding time of a state is Exponential with parameter $-q_{ii}$, with mean $1/(-q_{ii})$
  - The exponential distribution is memoryless

- E.g., good (0) or bad (1) state of a wireless channel
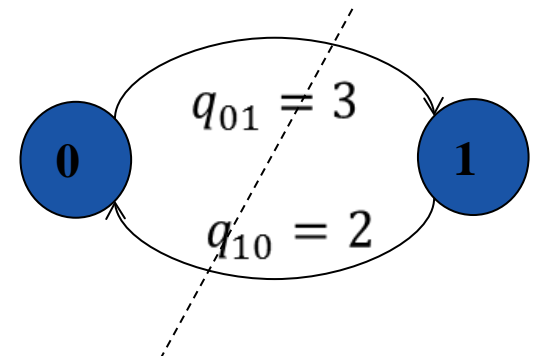
Recall –
Discrete time Markov chains

- E.g., to model the packet loss process at a receiver
  - States: packet received or lost (0,1)
  - Captures the burstiness of the loss process (see Gilbert model later in the course)
    - If a packet is lost (state 1), the next one is lost with probability p_{11}
    - If a packet is received (state 0), the next one is received with probability p_{00}
  - $\rightarrow$ Packets lost in a raw $\sim Geo(1 - p_{11})$, in average 1/ (1-p_{11})
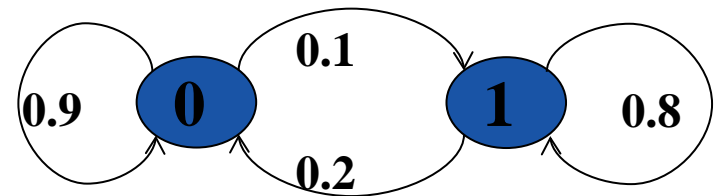
Markovian traffic modeling

- Traditional telephone networks (from Erlang)
  - Poisson call arrival
  - exponential call duration $\Rightarrow$ nice Markovian models
  - constant rate (M/M/*/*)

- Similar models are possible for data networks
  - Poisson flow/packet arrival process
  - Exponential flow size (e.g., file length), packet size

Markovian traffic models

- Poisson process: $P\{N(t)=n\}=e^{-\lambda t}(\lambda t)^n/n!$
- Exponential distribution: $P(X \leq t)=1-e^{-\lambda t}$, $f(t)= \lambda e^{-\lambda t}$

- Recall – some basic results

- Exponentially distributed interarrival and service times
- Possion arrival: exponential interarrival time
- Exponential distribution is memoryless – simple modeling
- Tail function $P(t>T)=e^{-\lambda T}$ – exponential decay in t
  - e.g., the probability that a packet size is larger than T decreases exponentially in T.
- Consecutive values (interarrival time, service time) are independent, therefore auto-covariance is zero
- $Cov(k)=E[(X_i-E[X])(X_{i+k}-E[X])]=E[X_iX_{i+k}]- E[X]^2=0$

Markovian traffic models

- Exponential interarrival and service times in queues (M/M/*/*)
- Buffering is efficient, does not cause large delays

- E.g, distribution of the number of users in an M/M/1 queue: $p(n)=(1- \rho)\rho^n$, $\rho=\lambda x$
- $P(n \geq N)=\rho^N$ – the probability that the queue length is at least N decays exponentially fast (exponential decay)

Markovian traffic models

- Multiplexing is efficient, decreases the blocking probability
- E.g, M/M/m/m
  - Multiplexing: higher aggregate arrival intensity $\rightarrow$ higher offered load
- Blocking given by the Erlang-B curves

Blocking

Offered load (λ/μ)
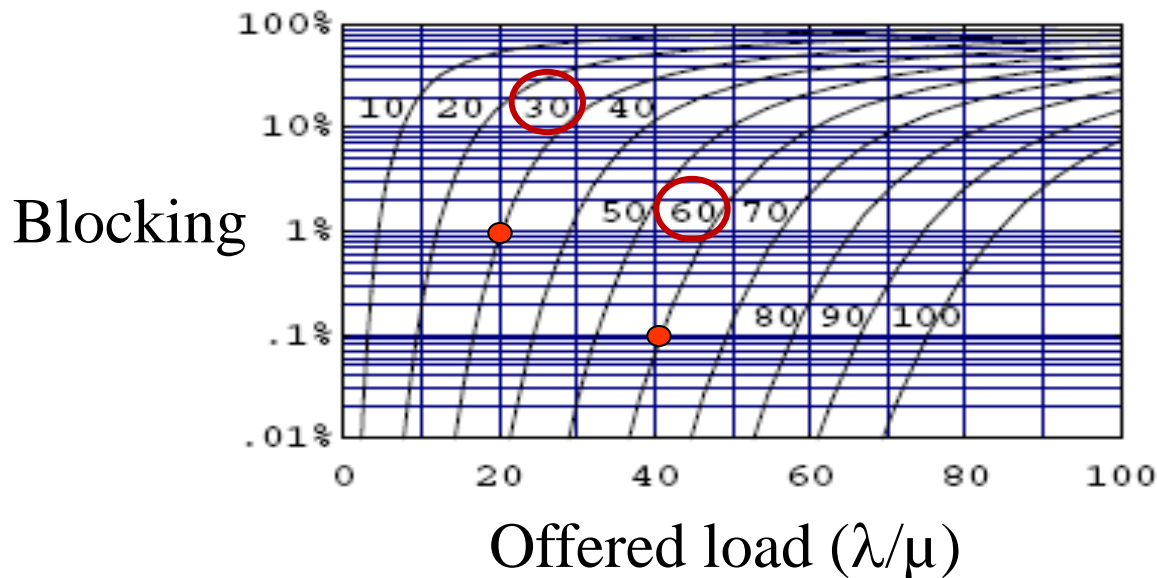
B(load,servers)

B(20,30)≈1%
B(40,60)≈0.1%

Markov modulated models

- However, we know that packet arrivals are not Poissonian
  - the arrival rate changes with time (traffic control, coding)
  - immediate result, auto-covariance should not be zero:
    $Cov(k)=E[(X_i-E[X])(X_{i+k}-E[X])]=E[X_iX_{i+k}]- E[X]^2 \neq 0$

- First step towards modeling traffic sources:
- Markov-modulated traffic models
  - to capture "burstiness" (changing arrival rate)
 - while keeping the simplicity of modeling

Markov modulated models

Packet scale models

- Markov-modulated Poisson Process (MMPP)
  - A Markov chain is given that describes the state of the source
  - The packet generation process is Poisson in each state, but with different intensity (state i $\rightarrow$ λ_i)
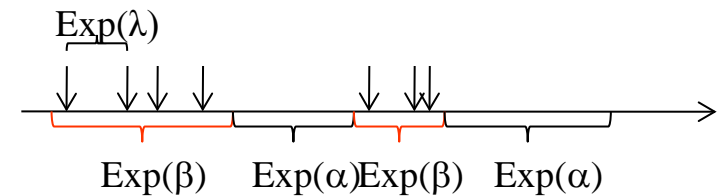 - Burstiness is captured by the state transitions in the Markov chain

State transitions governed by the Markov chain
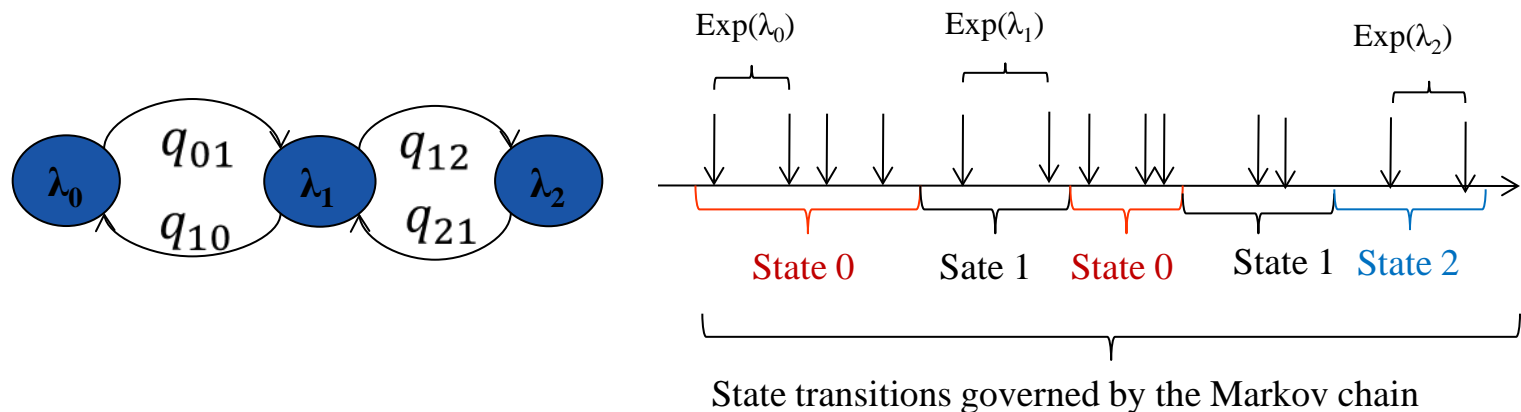
Markov modulated models

Packet scale models with two states
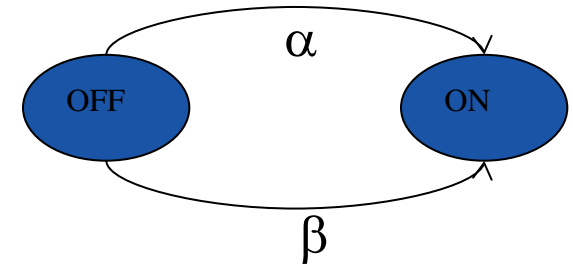
- Interrupted Poisson Process (IPP)
  - Most simple MMPP
 - two states $\lambda_0=0$, $\lambda_1=\lambda$

- ON-OFF model
  - two states, no arrivals in state 0 and fixed (d) packet interarrival times in state 1 (deterministic arrival process)
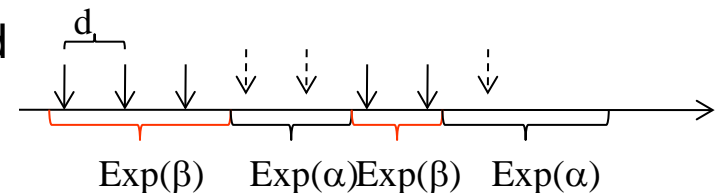
Markov modulated models

- ## Fluid models
  - When individual units (e.g., packets) have little impact
- ## Markov modulated fluid model
  - Traffic as a continuous stream with a parameterized flow rate (state i $\rightarrow$ r_i)
  - Flow rate changes described by a Markov-chain

$$\text{Exp}(\beta) \qquad \text{Exp}(\alpha) \qquad \text{Exp}(\beta) \qquad \text{Exp}(\alpha)$$
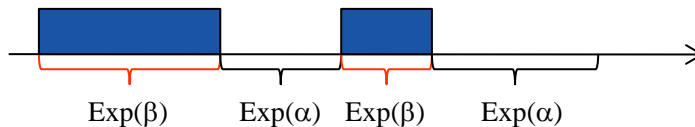
- ## Semi Markov models and embedded Markov chains
  - If the state holding times are not Exponential
  - The sequence of states visited can be described with a discrete time Markov chain -> embedded Markov chain
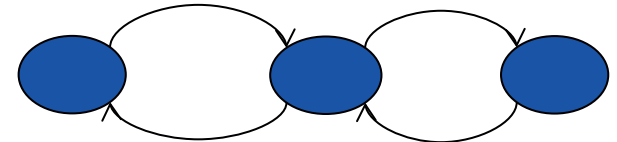
Markovian traffic models
Modeling voice traffic

- Compare the average delay at a multiplexer, if
  - Real voice source packets are multiplexed in a simulator
  - Poisson arrival is assumed with the same average rate
  - 2 state MMPP model is used
  - Some advanced technique is used

- Results:
  - Poisson arrival approximation underestimates delays (queue lengths)
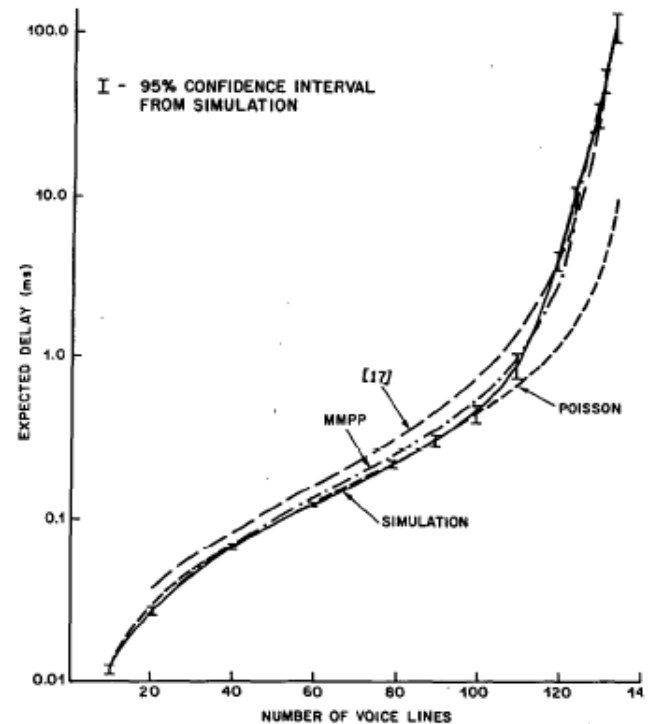 - MMPP seems to fit well at high load regime as well

Fig. 2. Expected delay for a packetized voice multiplexer.

Markovian traffic models

- Auto-covariance and auto correlation function decays exponentially
  - Auto-covariance: $Cov(t)=E[(X_i-E[X])(X_{i+t}-E[X])]$
  - Auto-correlation: $r(t)=Cov(t)/V[X]$

- Simplest example: on-off fluid model
  - The auto-correlation of the state of the system (on or off)
  - $r(t)=e^{-(\alpha+\beta)t}$

- What does it mean: the system has some memory about the past, but only for a short time – (we introduce the concept of short range dependence later)
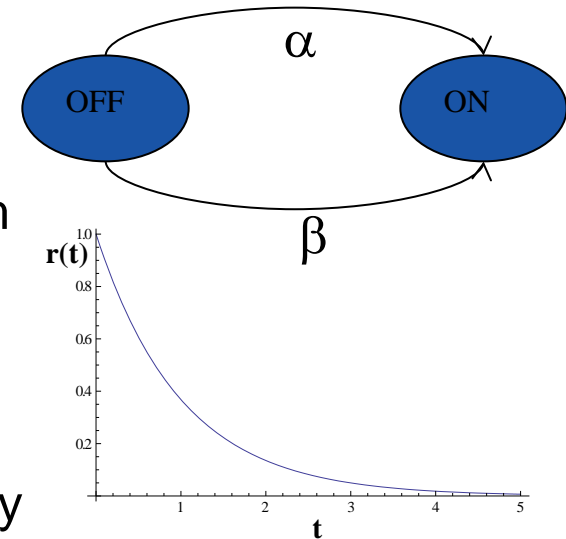
Traffic modeling

- Are Markovian traffic models enough to model network traffic sources?
- Or do we need other models?

Modeling Internet traffic

Read: J. Roberts, "Traffic theory and the Internet,"

"As a first approximation, it is not unreasonable to assume that individual flows also occur as a Poisson process. To ignore the correlation between flow arrivals within the same session is not necessarily significant when the number of sessions is large. It is also true that results derived under the simple Poisson assumption are also often true under more general assumptions.

The size of elastic flows (i.e., the size of the documents transferred) is extremely variable and has a so-called heavy-tailed distribution: most documents are small (a few kilobytes) but the number which are very long tend to contribute the majority of traffic. The precise nature of the size distribution is important in certain circumstances, such as describing the resulting self-similar packet arrival process, and can have a significant impact on performance in some multiplexing schemes.

The duration of streaming flows also generally has a heavy-tailed distribution. Furthermore, the packet arrival process within a variable rate streaming flow is often self-similar."

Modeling Internet traffic

- Elastic flows - controlled by congestion control
  - e.g., file transfer
  - arrival of flows: independent activity of a large number of users $\rightarrow$ *Poisson*
  - size: *heavy tail*
  - traffic characteristics: extreme variability introduced by TCP and heavy tailed flows
  - *self-similar* packet arrival process
- Streaming flows - determined by the source coding
  - arrival of flows: *Poisson*
  - duration: extreme variability, *heavy tail*
  - traffic characteristics (rate): often *self-similar* due to coding
- Conclusion:
  - Simple Markovian or Markov Modulated source models may not work

Home reading

Home reading for Wednesday next week: A. Adas, "Traffic Models in Broadband Networks", IEEE Communications Magazine, July 1997

- Markov and Embedded Markov models in detail
  - including the MMPP example for video coding
- Regression models *are not part* of the course material, but are interesting reading
- Long-range dependent traffic models, *not including* fractional ARIMA and fractional Brownian Motion
- See "Reading Assignment" on the course web

End of first lecture.

Traffic modeling - recall

- Teletraffic theory – performance triangle
- Traffic modeling
  - Traffic objects: flow (elastic, streaming), burst, packet
  - Modeling levels:
    – Packet level: arrival process, packet length distribution
    – Flow level: arrival process, flow length distribution, flow characteristics (packet scale or fluid)
- Classical traffic models
  - Poisson arrival, exponential service time (packet, flow length) – M/M/*/*
    – Simple model
    – Exponential decay (interarrival time, service time, queue length), no time correlation of input parameters
    – Efficient multiplexing, efficient buffering
- Markov modulated traffic models
  - Poisson arrival, exponential service time (packet, flow length) – M/M/*/*
    – Still tractable model
    – Captures burstiness, auto-covariance is not zero, but decays Exponentially
    – Seem to model well the effect of burstiness on buffering

Discrete time Markov chains

- E.g., to model the packet loss process at a receiver
  - States: packet received or lost (0,1)
  - Captures the burstiness of the loss process (see Gilbert model later in the course)
    - If a packet is lost (state 1), the next one is lost with probability p_{11}
    - If a packet is received (state 0), the next one is received with probability p_{00}
  - $\rightarrow$ Packets lost in a raw $\sim Geo(1 - p_{11})$, in average 1/ (1-p_{11})
  - $\rightarrow$ Steady state probability of receiving or loosing a packet:

$$\{p_0, p_1\} = \{p_0, p_1\} \begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix}$$

$$\{p_0, p_1\} = \{p_0, p_1\} \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}$$
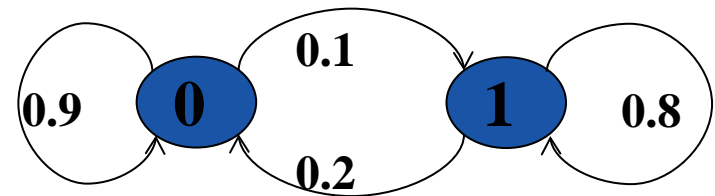
  - What is the probability that a packet gets lost (p_1)?
  - What is the average number of packets lost in a raw?

Continuous time Markov chains

- Continuous time Markov chains
  - State transition is possible at any time
  - State transition intensity matrix $\mathbf{Q} = \{q_{ij}\}$, $q_{ii} = -\sum q_{ij}$
  - $\underline{p'(t)} = \underline{p(t)}\mathbf{Q}$
  - If steady state exists, the stationary state probability is given by $\underline{0} = \underline{p}\mathbf{Q}$
  - Holding time of a state is Exponential with parameter $-q_{ii}$, with mean $1/(-q_{ii})$
  - The exponential distribution is memoryless

- E.g., good (0) or bad (1) state of a wireless channel
- Steady state probabilities:

$$\{0,0\} = \{p_0, p_1\} \begin{bmatrix} -3 & 3 \\ 2 & -2 \end{bmatrix}$$

- What is the probability that the system is in state 1?
- What is the average holding time of bad state?

$q_{01} = 3$

0   **1**

$q_{10} = 2$
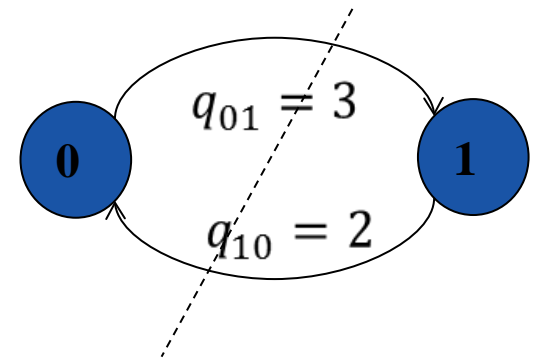
A traffic source is modeled as follows:
- 2 state MMPP
- Transition intensity $q_{12}=0.5$, $q_{21}=0.5$ (transitions per sec)
- Transmission rates: $\lambda_1=100$ packets/s and $\lambda_2=400$ packets/s
- Packet size: 500Bytes

1. Draw the Markov-chain, give all the parameters, give the Q matrix

2. What is the mean time in states 1 and 2 respectively?
3. What is the probability that the source is in state 1 (state 2) at an arbitrary point of time?

4. What is the average packet interarrival time in state 1?
5. What are the transmission rates in the two states in bit per sec?
6. What is the average transmission rate?

7. If 5 such sources are multiplexed, what is the probability that the instantaneous rate is 8Mbps or larger?
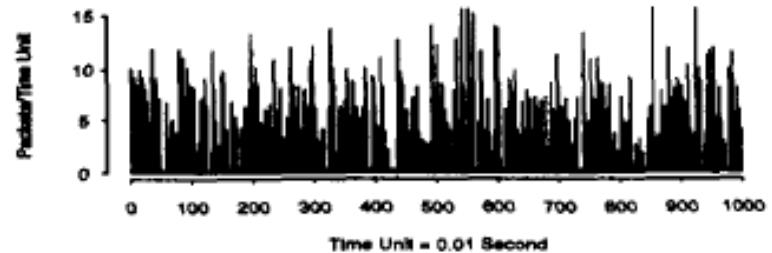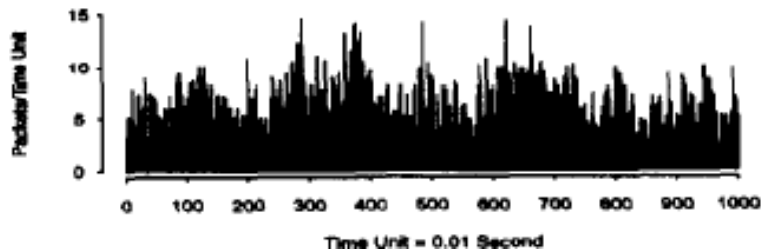
Heavy-tail distributions, self-similarity, and long-range dependence

- What are the limitations of Markovian models?
- Example 1
  - Telephone call holding time measurements (holding time, s)
    - Exponential assumption: $P(s>x)=e^{-\mu x}$
    - Statistics (for large s): $P(s>x) \sim x^{-\alpha}$, $\alpha>0$
  - Decay is slower than exponential: heavy-tail distribution
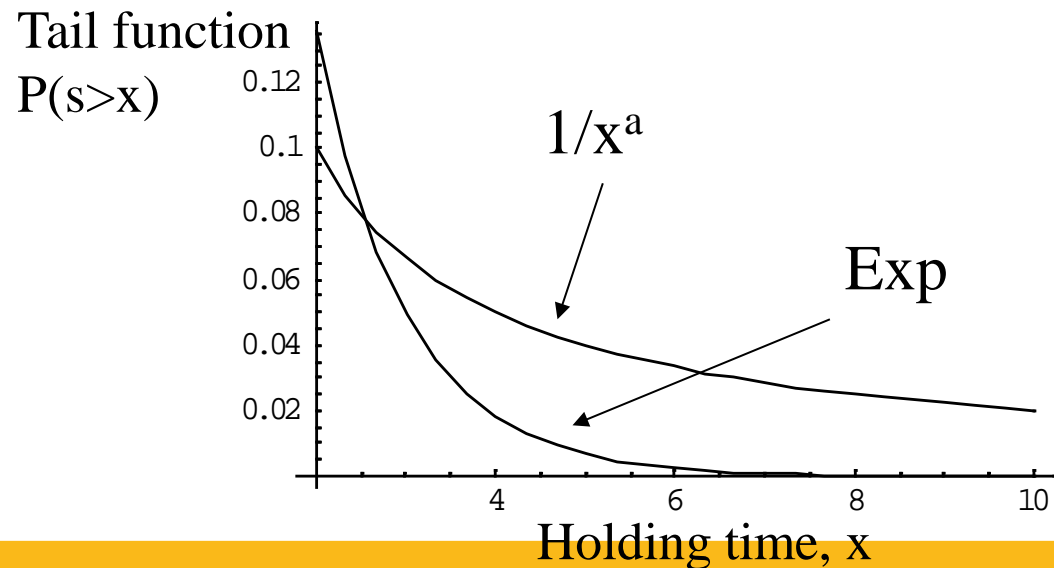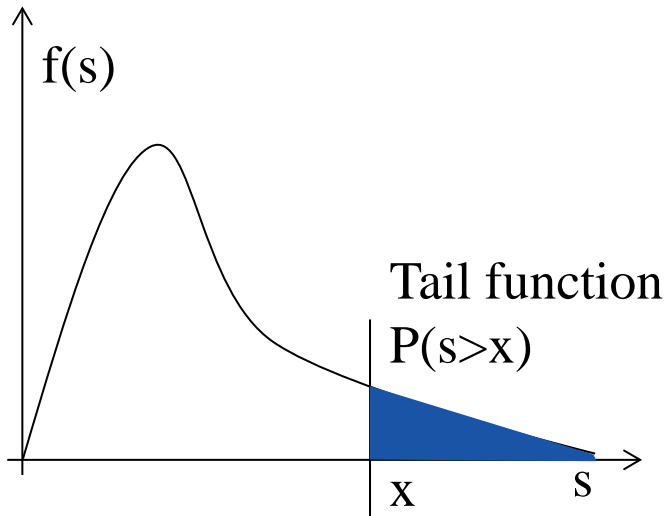
Heavy-tail distributions, self-similarity, and long-range dependence

- Example 2
  - Packet arrivals in 40 hours Ethernet traffic (Bellcore '89)
  - Number of packet arrivals in increasing time intervals
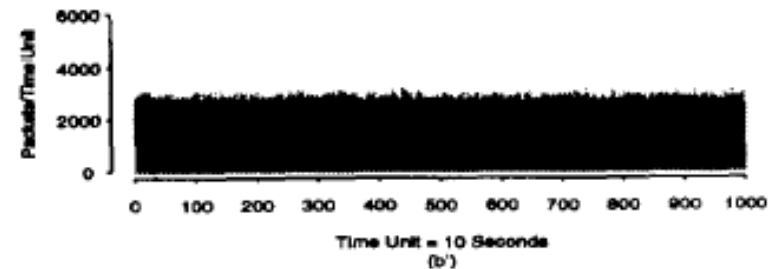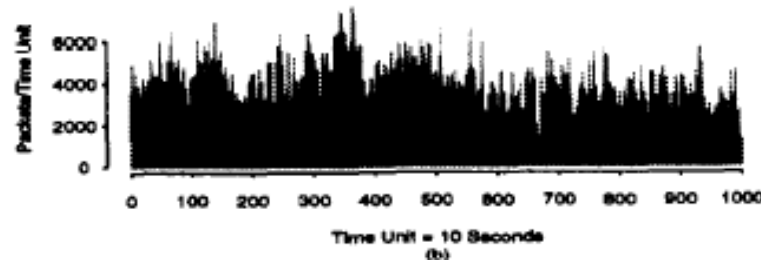

Ethernet measurement   Poisson

- Self-similar nature of packet arrival process

Heavy-tail distributions, self-similarity, and long-range dependence

- Example 3 – the effect of (long range dependent) self similarity

- LRD-SS source characteristics changes the network performance significantly.
  - E.g., mean queue-length at routers/multiplexers
  - Blocking and loss probabilities
- Therefore
  - we have to take it into account at the performance evaluation
 - have to understand how it emerges and whether it is possible to avoid it
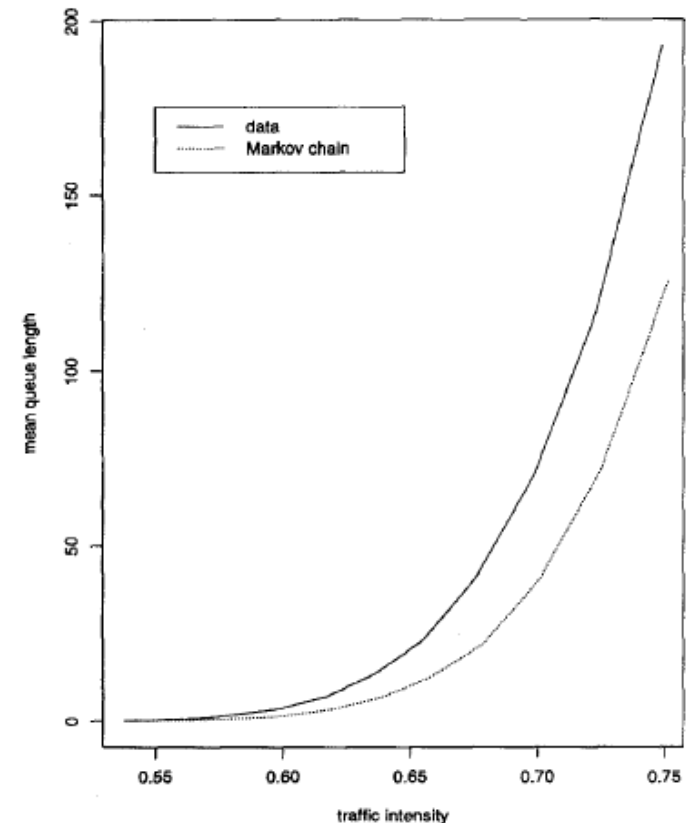
Heavy-tail distributions, self-similarity, and long-range dependence

- We have to address the followings
  - what is heavy-tail distribution
  - what is self-similarity (and related: what is long range dependence)
  - how are these related to each other
  - when is it possible to apply Markovian models

- First a bit of group work

Heavy-tail distributions

- Exponential distribution: $P(X>x)=e^{-\mu x}$
- Heavy-tail distribution:
  - $P(X>x) \sim x^{-a}$, $x \rightarrow \infty$, $a>0$
  - the asymptotic shape is hyperbolic
- Pareto distribution: often used heavy tail distribution (e.g., for file size length):
  - $f(x)= ab^a/(x^{a+1})$,
    - $a>0$ (shape),
    - b is the minimum possible value (base)
  - $P(X>x)=1-F(x)=(b/x)^a$
  - $E[X]=ab/(a-1)$ for $a>1$ otherwise the mean is not finite

P(s>x), b=1, a=1.5, 2

Pareto

Exp with the same mean
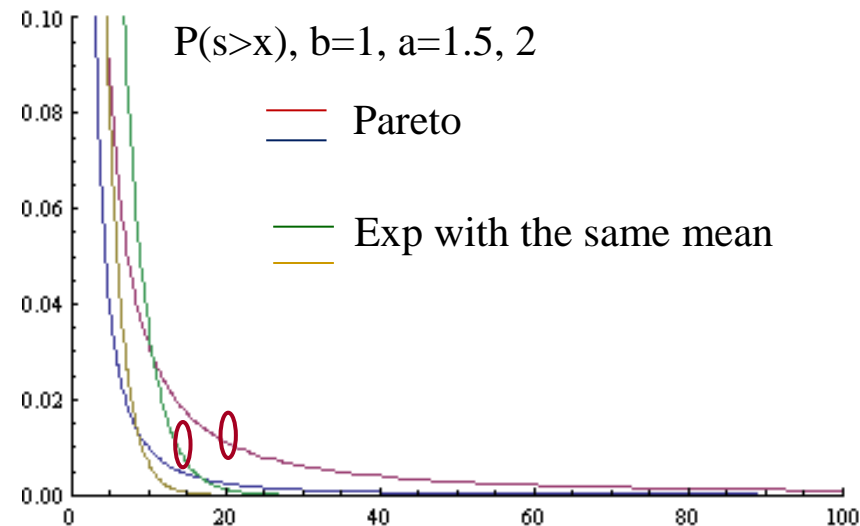
Heavy-tail distributions – Waiting for the bus revisited

- Distribution of remaining service time (remaining time to wait for the bus…)

$$R_t(x) = P(X > x + t | x > t) = \frac{\bar{F}(x + t)}{\bar{F}(t)}$$

- Exponential distribution: $P(X > x) = e^{-\mu x}$, $R_t(x) = e^{-\mu x}$ (the memoryless property)

- Pareto distribution:

$$f(x) = \frac{ab^a}{x^{a+1}}, P(X > x) = \left(\frac{b}{x}\right)^a$$

$$R_t(x) = \frac{\left(\frac{b}{x+t}\right)^a}{\left(\frac{b}{t}\right)^a} = \left(1 + \frac{x}{t}\right)^{-a}$$

- That is, the remaining service time increases with t, the time already spent on service!

Heavy-tail distributions

- Why is heavy-tail a problem?
- Can be proved:
  - Superposition of ON-OFF processes where the distribution of the ON periods is heavy tailed (e.g., Pareto) gives long-range dependent self-similar process
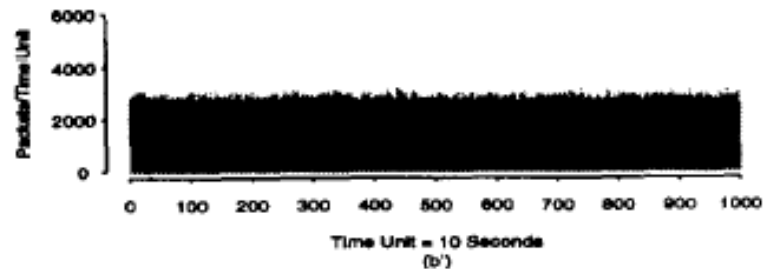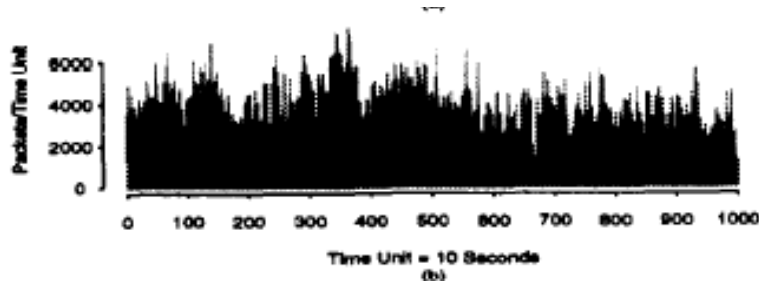
Long-range dependence

- Consider X_i stochastic process, i=1,2,3 (discrete time)
  - Discrete process, samples from a continuous time process or integral over the interval
  - E[X], V[X] finite
  - Auto-covariance: $Cov(k)=E[(X_i-E[X])(X_{i+k}-E[X])]$
  - Auto-correlation: $r(k)=Cov(k)/V[X]$
- Short-range dependent:
  - $\sum_{k=1}^{\infty} r(k) < \infty$: the consecutive samples are correlated, but the correlation decreases fast with k
- Long-range dependent:
  - $\sum_{k=1}^{\infty} r(k) = \infty$: the consecutive samples are correlated, the correlation is preserved for long period.
- Note: long-range dependence is an asymptotic definition for large lags (k).

Long-range dependence

- Short-range dependent:
  - $\sum_{k=1}^{\infty} r(k) < \infty$: the consecutive samples are correlated, but the correlation disappears fast
- Long-range dependent:
  - $\sum_{k=1}^{\infty} r(k) = \infty$: the consecutive samples are correlated, the correlation can be preserved for long period.

- MMPP is short range dependent. E.g., on-off fluid:

$$r(k) = e^{-(\alpha+\beta)k} \quad \text{(geometric serie)}$$

$$\sum_{k=1}^{\infty} r(k) = \sum_{k=1}^{\infty} e^{-(\alpha+\beta)k} < \infty \quad \text{since } e^{-(\alpha+\beta)} < 1$$

Exp(β)   Exp(α)   Exp(β)   Exp(α)
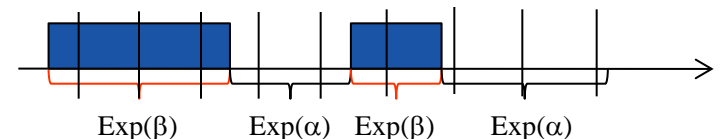
- So, what is the relationship between long range dependence and self-similarity?

Self-similarity

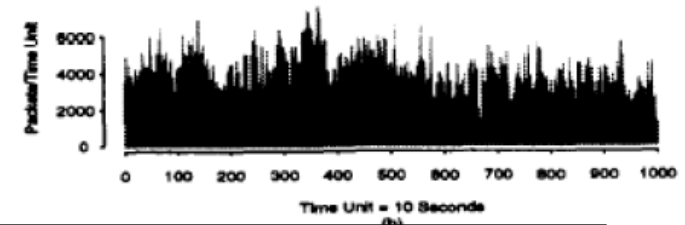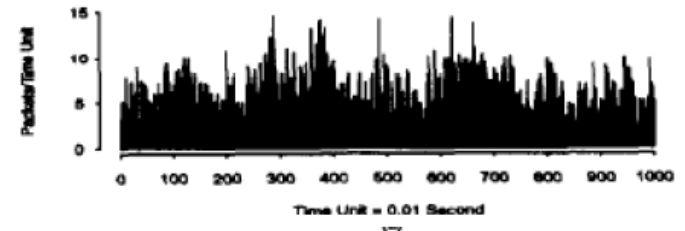- Consider X_i stochastic process, i=1,2,3 (discrete time)
- E[X], V[X] finite

$X_1, X_2, \ldots$

$$X_1^{(2)} = \frac{X_1 + X_2}{2}, X_2^{(2)} = \frac{X_3 + X_4}{2}$$

$$X_1^{(m)} = \frac{X_1 + \cdots + X_m}{m}, X_2^{(m)} = \frac{X_{m+1} + \cdots + X_{2m}}{m}$$

$$r^{(m)}(k) = \frac{Cov^m(k)}{V[X^m]}$$

time unit: $0.01s \rightarrow X_1, X_2 \ldots$

time unit: $10s \rightarrow X_1^{(1000)}, X_2^{(1000)} \ldots$

- (Second order) Self-similar: auto-correlation $r^m(k) = r(k)$, for all m and k
- Asymptotically self-similar: if above true for large m and k

50

Self-similarity and long-range dependence (LRD)

- Second-order self-similar: $r^m(k) = r(k)$, for all m and k
- r(k) has specific form (can be proved):

$$r(k) \sim H(2H-1)\frac{1}{k^{2(1-H)}}, \qquad 0 < H < 1,\ H \neq 0.5$$

- H: Hurst parameter, the parameter of a self-similar process
- Self-similarity and LRD

$$\sum_{k=1}^{\infty} r(k) = H(2H-1)\sum_{k=1}^{\infty}\frac{1}{k^{2(1-H)}} \quad \text{is a hyper-harmonic serie}$$

$$\sum_{k=1}^{\infty} r(k) = \infty \ \text{ if } 2(1-H) < 1 \rightarrow H > 0.5$$

→ A self-similar process is LRD if 0.5<H<1. This is the interval when SS makes trouble.

- Often the terms self-similarity and long-range dependence are used for the same thing.

Heavy tail, self-similarity (SS) and long-range dependence (LRD)

LRD

SS
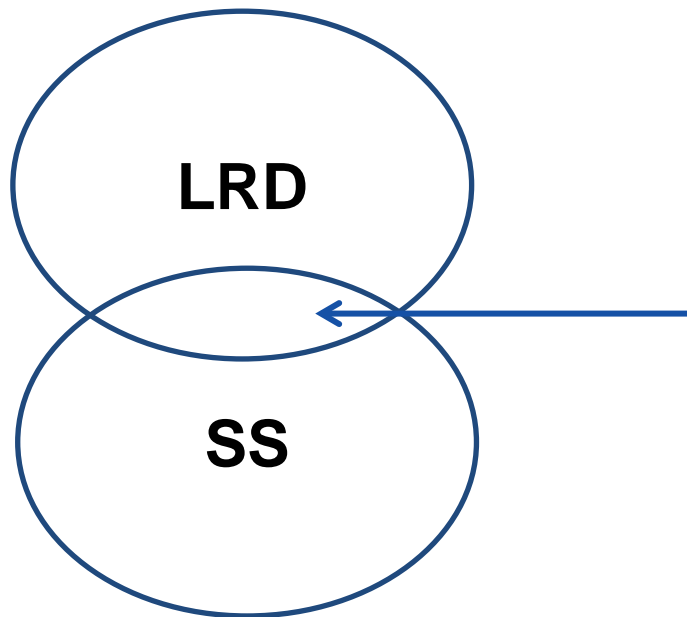
- LRD-SS is the "problematic area
- Multiplexed heavy tail on-off sources give LRD-SS process

Markovian vs. SS/LRD models

How should we choose traffic model for performance evaluation?

Examples

Fukuda, K., Takayasu, M., Takayasu, H., **A cause of self-similarity in TCP traffic**, *International Journal of Communication Systems*, v 18, n 6, Aug. 2005, p 603-17

Abstract: We analyze the statistical properties of aggregated traffic flows generated by TCP, in order to clarify a possible cause of self-similarity in Internet traffic. Using ns-2 simulation… The main contribution of our work is to show that TCP itself contains a mechanism for generating self-similarity…

Examples

Tsybakov, B. , Georganas, N.D., **Overflow and losses in a network queue with a self-similar input,** *Queueing Systems, Theory and Applications,* 2000.

Abstract: …considers a discrete time queuing system that models a communication network multiplexer which is fed by a self-similar packet traffic. The model has … an input traffic which is an aggregation of independent source-active periods having Pareto-distributed lengths and arriving as Poisson batches. The new asymptotic upper and lower bounds to the buffer-overflow and packet-loss probabilities are obtained.

Examples

Kasahara, S., **Internet traffic modeling: Markovian approach to self-similar traffic and prediction of loss probability for finite queues**, *IEICE Transactions on Communications,* 2001.

Abstract: It has been reported that IP packet traffic exhibits self-similar nature and causes the degradation of network performance… However recent studies have revealed that the loss probability of finite queuing systems can be well approximated by the Markovian input models. This paper studies the time-scale impact on the loss probability of MMPP/D/1/K system where the MMPP is generated so as to match the variance of the self-similar process over specified time-scales.

Markovian vs. SS/LRD models

- How should we choose traffic model for performance evaluation?
  - SS/LRD
    - complex models, possible to use for simulation but mathematical models are not that tractable
    - LRD captures asymptotic behavior but not short time characteristics
  - Markovian models
    - can capture correlations on arbitrary – finite – time scale
    - easier to use in mathematical models
- We have to choose models according to the dominant time scale we consider.

Summary (1/2)

- Network traffic modeling
  - Flows, bursts and packets
  - Elastic and streaming flows
  - Packet scale and fluid models for flow characterization

- Markovian traffic models
  - Markov modulated traffic models
  - The rate is modulated by a Markov chain to capture burstiness
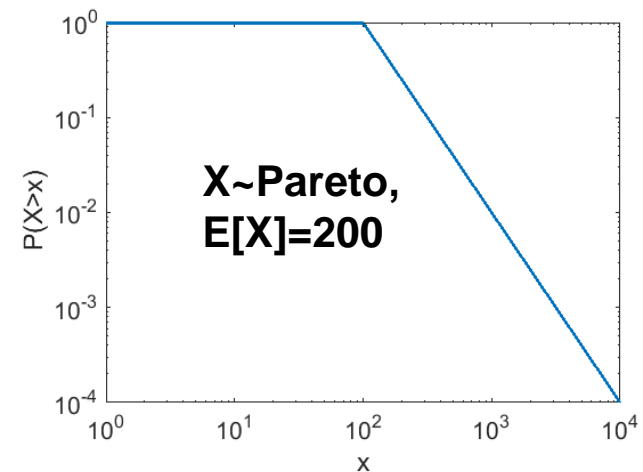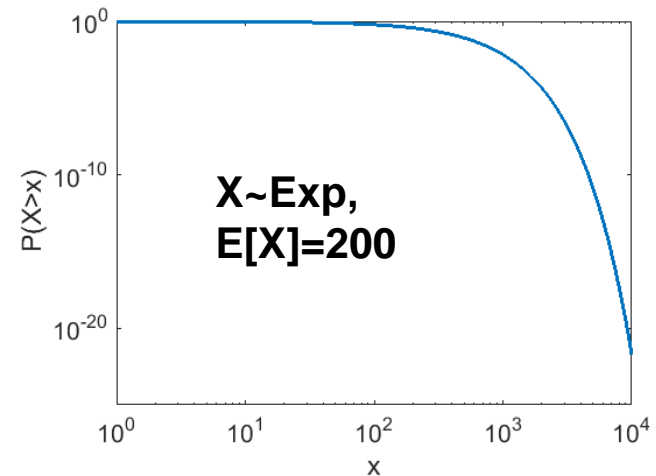 - Can describe short term correlation

Summary (2/2)

- Long-range dependence, self-similarity and heavy-tail
    - Asymptotic characteristics
    - Heavy-tail: the tail function of the distribution has only hyperbolic decrease: $P(s>x) \sim x^{-a}$, $x \to \infty$, multiplexing heavy-tail flows leads to self-similarity
    - Long-range dependence: correlation is preserved over long timescales: $\sum_{k=1}^{\infty} r(k) = \infty$
    - Self-similarity: the correlation is preserved irrespective of time aggregation: $r^m(k) = r(k)$
    - Self similarity is characterized by H, the Hurst parameter, and the SS process is LRD if $0.5 < H < 1$
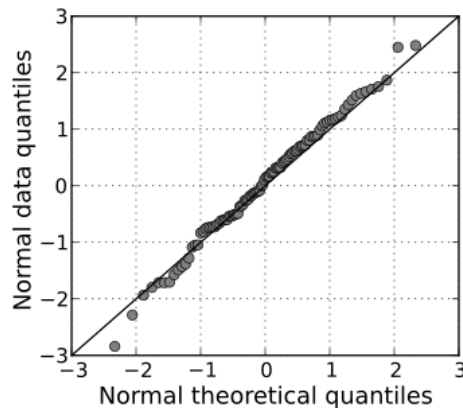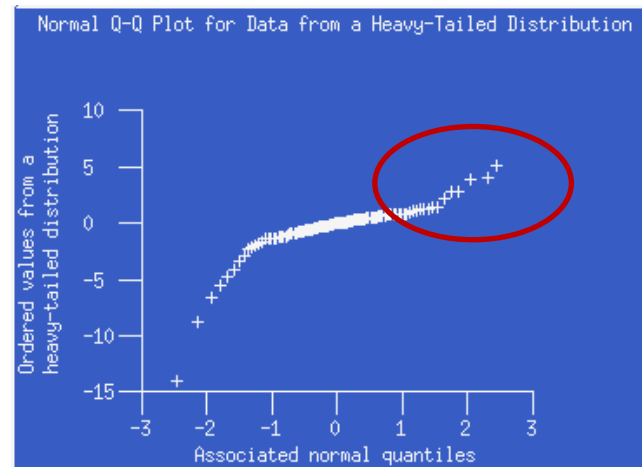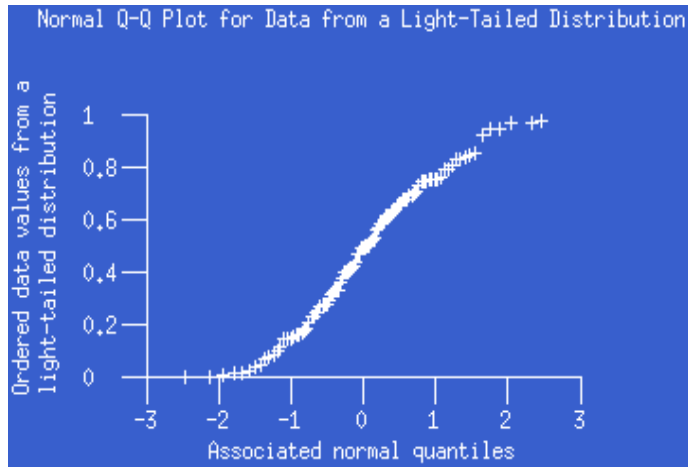   - LRD-SS flows lead to inefficient multiplexing and long queues

Heavy tail plots
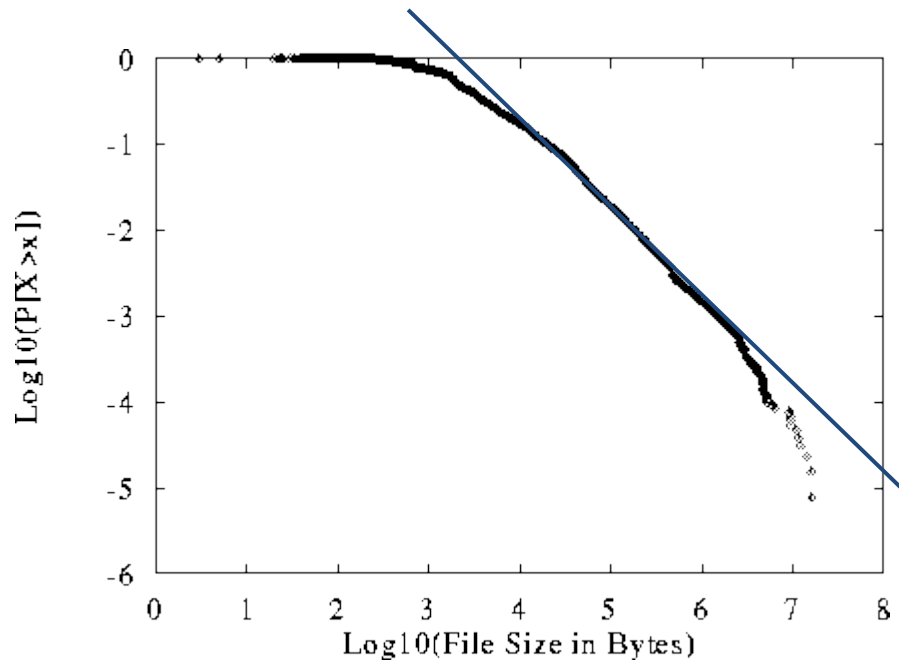
The figure shows a nearly linear plot on the log-log scale, which reflects a decay $\sim x^{-a}$, that is, heavy tail.

X~Exp, E[X]=200

X~Pareto, E[X]=200

Heavy tail plots

Normal Q-Q Plot for Data from a Light-Tailed Distribution

Normal Q-Q Plot for Data from a Heavy-Tailed Distribution

"Normal normal qq" by Skbkekas - Own work. Licensed under CC BY-SA 3.0 via Commons -
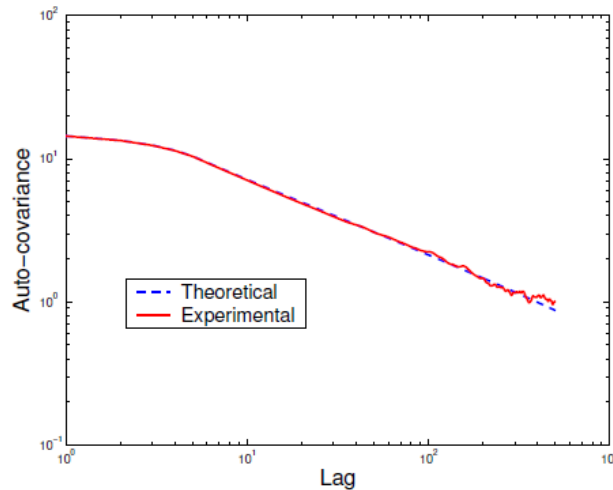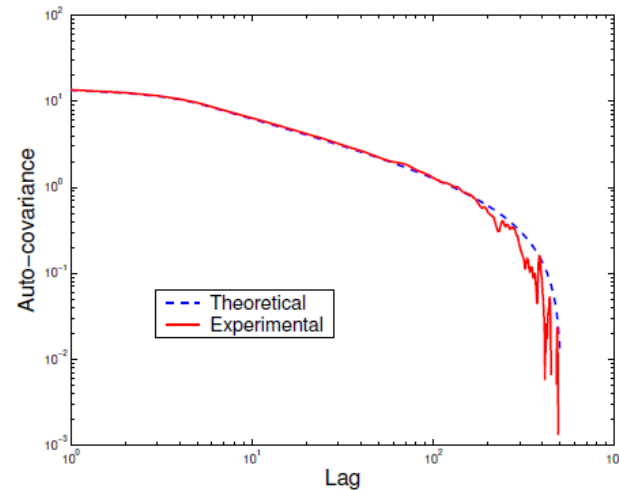
Q-Q plot: compares ordered sequence of samples from two distributions.

The up-going curve on the Q-Q plot reflects that the tail is heavier than the one of the normal distribution.

Long-range dependence

(a) Pareto Service Time

(b) Pareto+Exponential Service Time

The figure shows the auto-covariance ($r(k)$) as a function of the lag (k). Linear line in the log-log scale shows slow decay in $r(k)$, which may result in long range dependence. In contrast, $r(k)$ diminishes fast in the second figure, showing short range dependence.
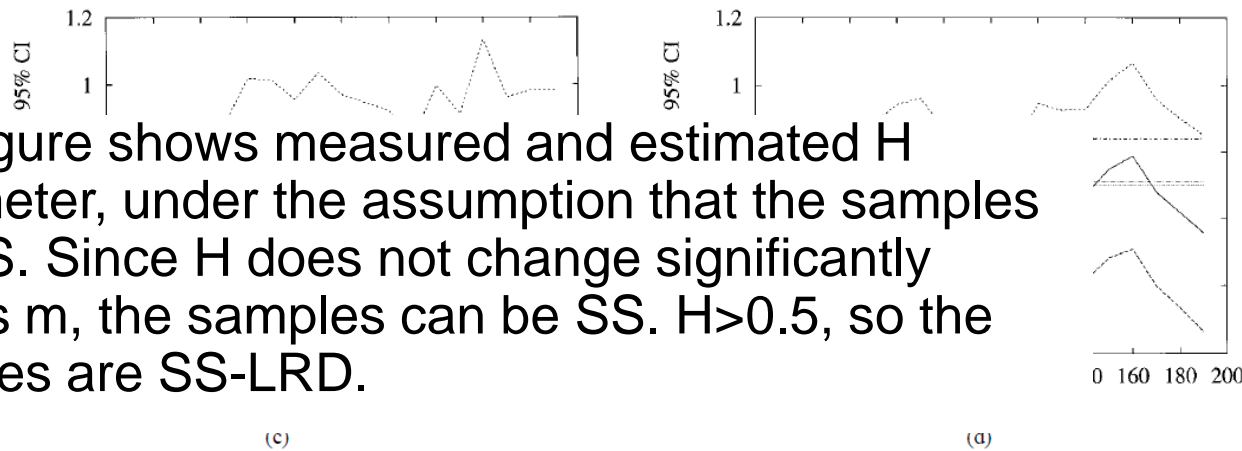
Self similarity

High load

Different estimators
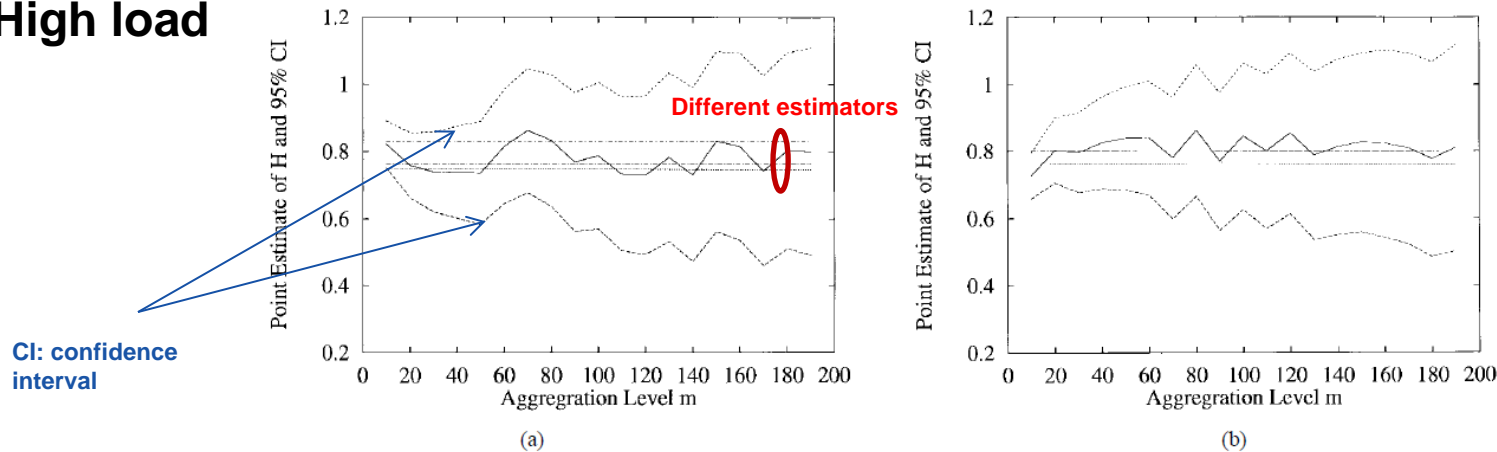
CI: confidence interval

(a)

(b)

The figure shows measured and estimated H parameter, under the assumption that the samples are SS. Since H does not change significantly across m, the samples can be SS. H>0.5, so the samples are SS-LRD.

Low load

(c)

(d)