KTH ROYAL INSTITUTE
OF TECHNOLOGY

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

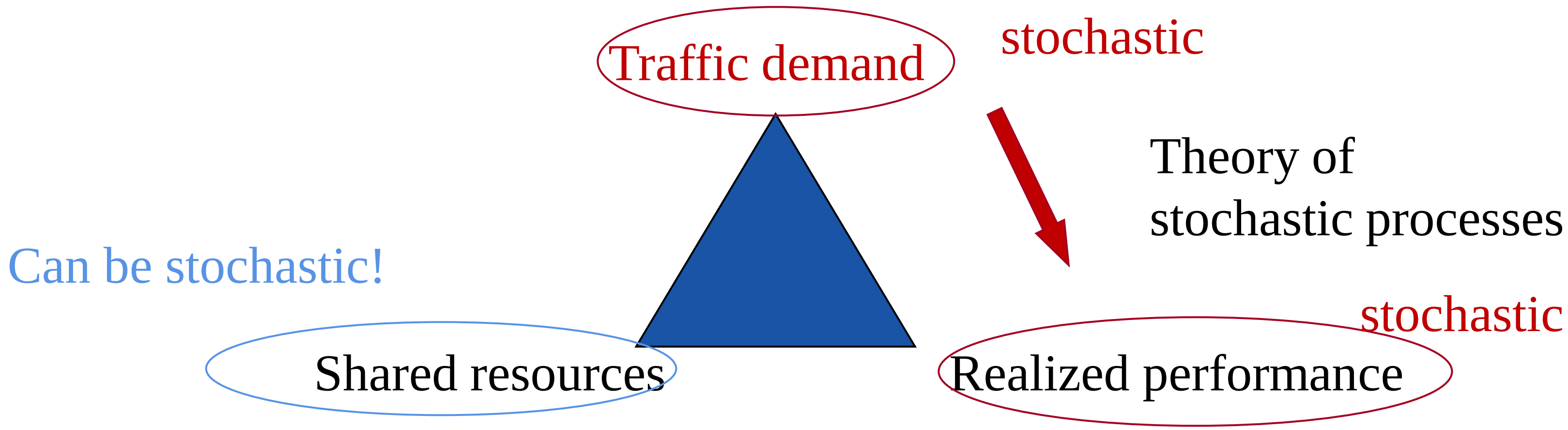
- Small project – during or after the course (for +3ECTS)
  - select a lecture topic
  - prepare a small simulator to support a mathematical model or problem definition, the simulator could be used for demonstration
  - see examples under Course material (will be updated soon)

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

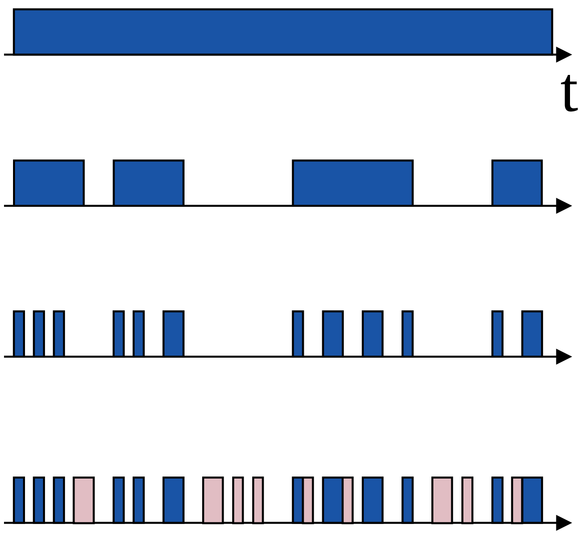
- Teletraffic theory:
  - to model dynamic resource sharing systems
  - to explain the traffic-performance relation



- Traffic: 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



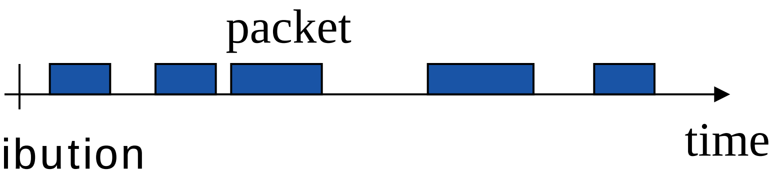
- Flow (one instance of communication, TCP or UDP session) — Skype call
- Burst (Active/passive periods) — Talk/listen
- Sequence of packets — IP packets
- Multiplexed packets — IP packets at a router

Traffic modeling

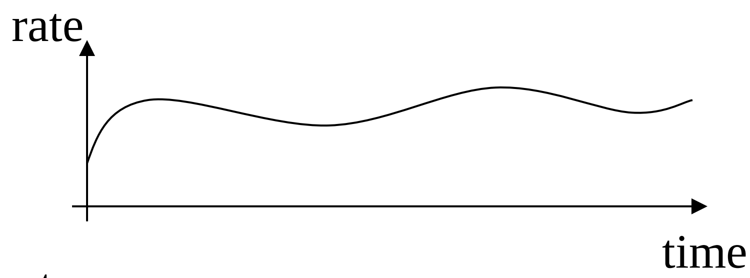
- Packet level – characteristics of the sequence of packets
  - packet arrival process
    - according to some stochastic/deterministic arrival process (e.g. Poisson arrival at a router…)
    - saturated source model: there is always packet to send at the source
  - packet size distribution
- Flow level (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

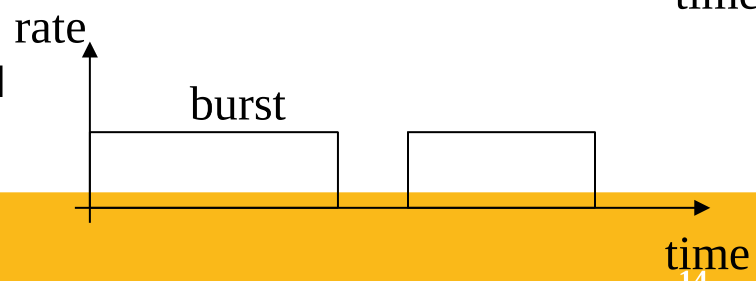
Models that describe the distribution of the sequence of packets for a flow level model



- 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



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?

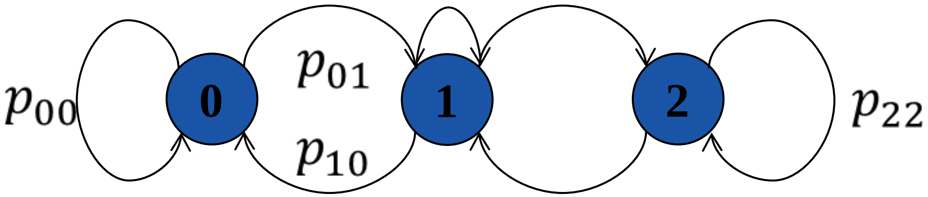
1. What is the probability that a packet collides and therefore needs to be retransmitted when using CSMA/CA protocol?
2. Several Skype calls are using the same communication link. What is the utilization of the link *(utilization={average rate of traffic} / {link transmission rate})*
3. Several flows 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

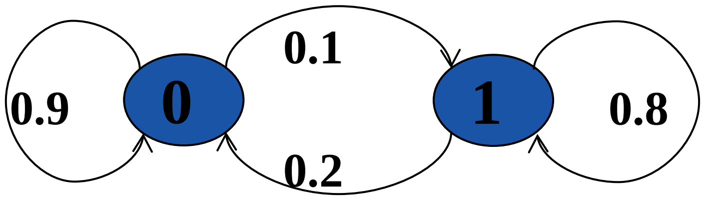


- Discrete time Markov chains
  - State transition probability matrix $\boldsymbol{P} = \{p_{ij}\}$
  - $\underline{p}_{i+1} = \underline{p}_i \boldsymbol{P}$
  - If steady state exists, the stationary state probability is given by $\underline{p} = \underline{p}\boldsymbol{P}$
  - Holding time of a state is geometric with parameter $1 - p_{ii}$ (memoryless)

Recall – 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}

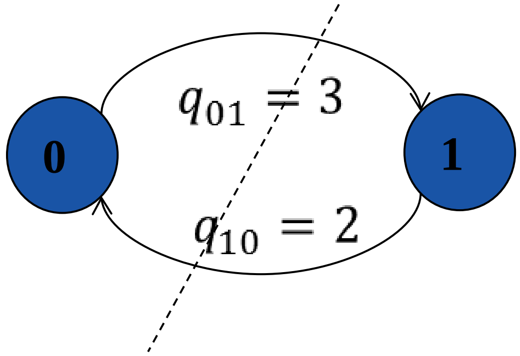
  → Packets lost in a raw $\sim Geo(1 - p_{11})$, in average 1/ (1-p_{11})



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



Markovian traffic modeling

- Traditional telephone networks (from Erlang)
  - Poisson call arrival
  - exponential call duration
  - constant rate

  } $\Rightarrow$ nice Markovian models (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
- $Cov(k)=E[(X_i-E[X])(X_{i+k}-E[X])]=E[X_iX_{i+k}]- E[X]^2=0$

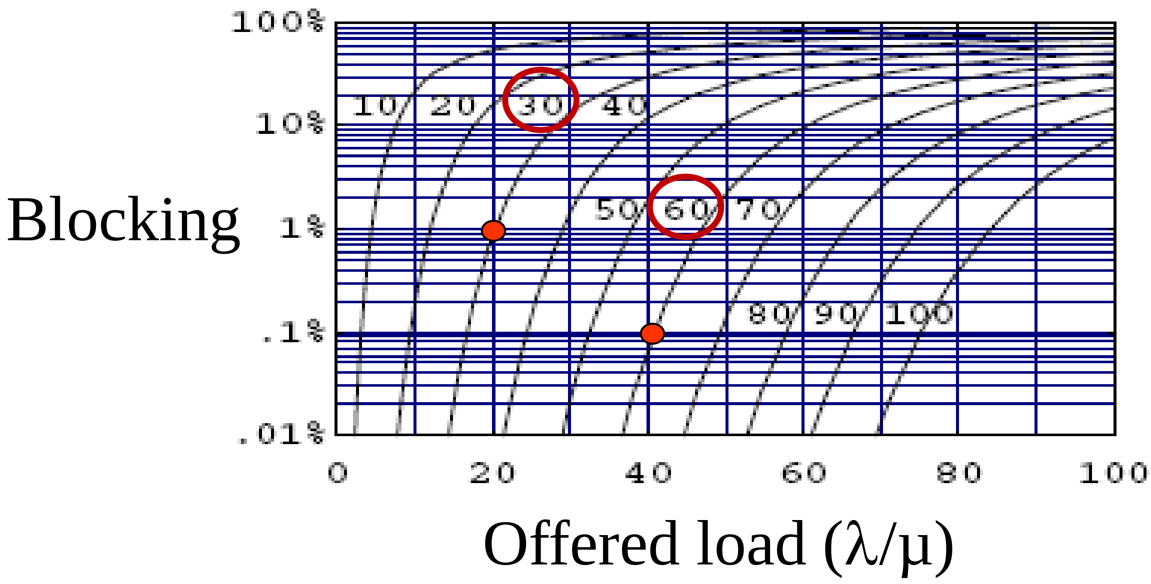
Markovian traffic models

- 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 → higher offered load
- Blocking given by the Erlang-B curves



B(load,servers)

B(20,30)≈1%
B(40,60)≈0.1%

Markov modulated models

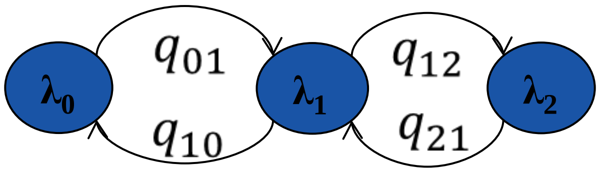
- However, we know that packet arrivals are not Poissonian
  - the arrival rate changes with time (traffic control, coding)
  - immediate result, auto-covariance should not be zero: $Cov(k)=E[(X_i-E[X])(X_{i+k}-E[X])]=E[X_iX_{i+k}]-E[X]^2 \neq 0$

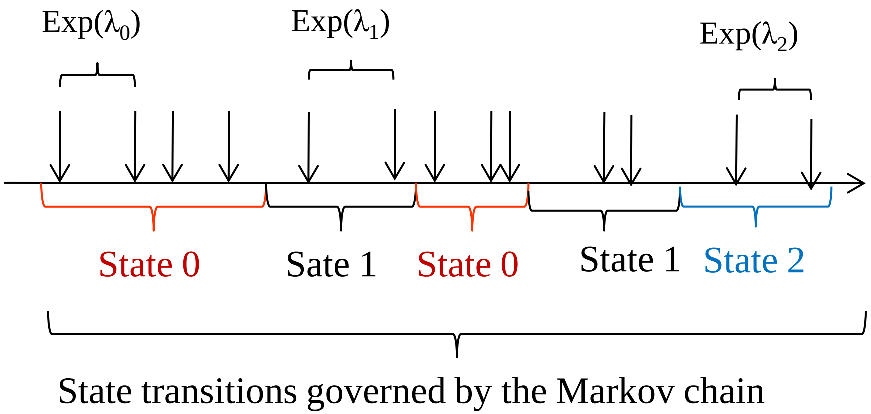
- First step towards modeling traffic sources:
- Markov-modulated traffic models
  - to capture “burstiness” (changing arrival rate)
  - while keeping the simplicity of modeling

Markov modulated models

Packet scale models

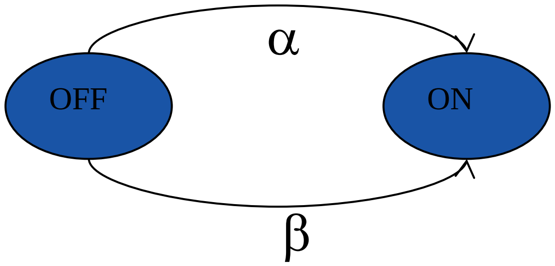
- Markov-modulated Poisson Process (MMPP)
  - A Markov chain is given that describes the state of the source
  - The packet generation process is Poisson in each state, but with different intensity (state i $\rightarrow \lambda_i$)
  - Burstiness is captured by the state transitions in the Markov chain



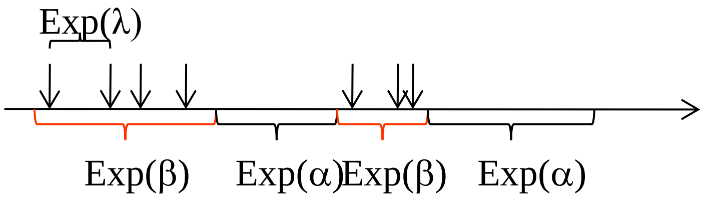


Markov modulated models

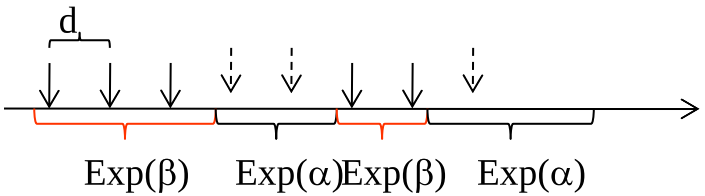
Packet scale models with two states



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

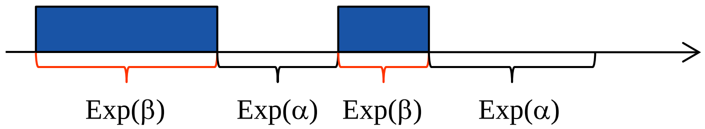


- ON-OFF model
  - two states, no arrivals in state 0 and fixed (d) packet interarrival times in state 1 (deterministic arrival process)



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



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

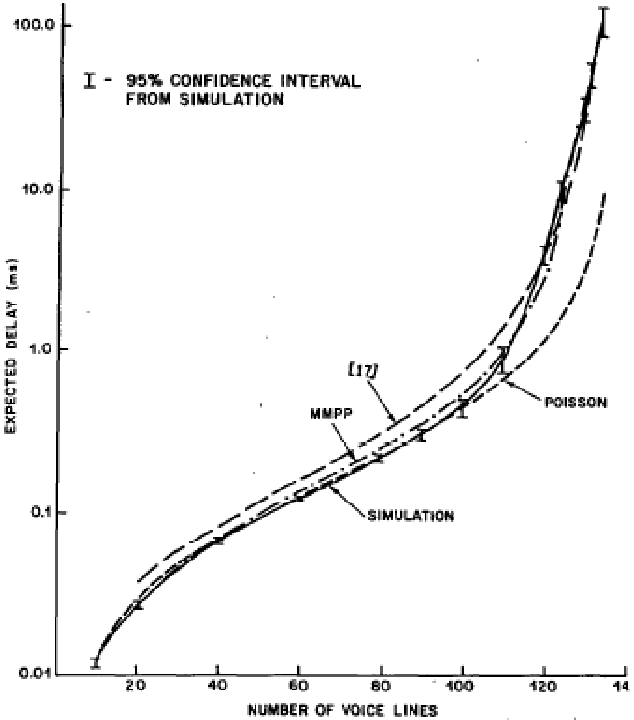
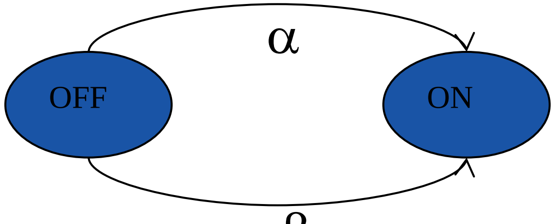


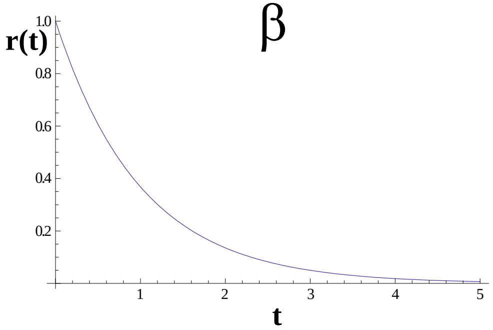
Fig. 2. Expected delay for a packetized voice multiplexer.

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 → *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.