Traffic Modeling (1)

Introduction

We need to either design a perfect network by employing appropriate data traffic models or rely on traffic management techniques, in order to provide QoS in data networks

Traditional traffic models for voice-centric networks

1) *Poisson – packet and connection arrival*
2) *Exponential - packet interarrival*

To evaluate the network performance : queuing performance, congestion control, designing process

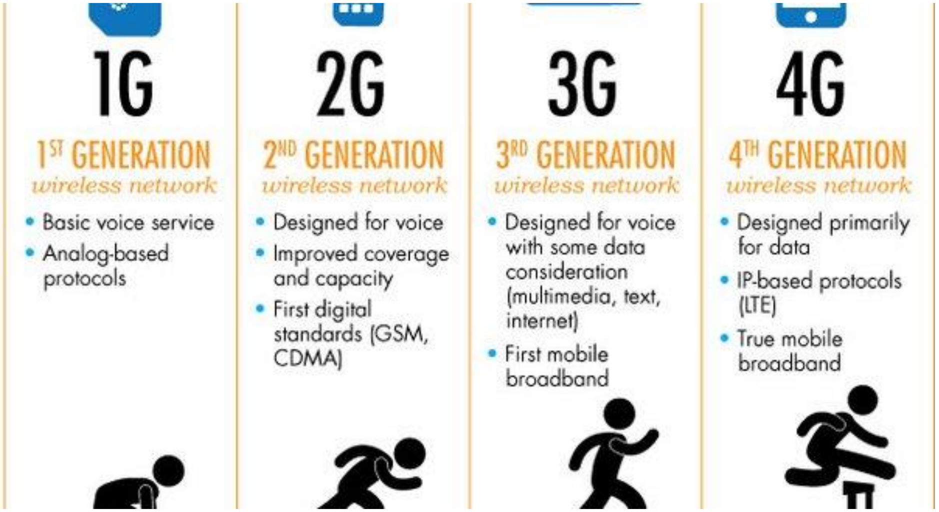
Erlang formulas have provided universal solutions to network problems for both wireline and wireless circuit-switched networks

Objective

- Traffic modeling is the key for determining the performance of the system.
- The more accurate is the traffic model the better is the system quantified in terms of its performance.
- Traffic modeling in the evaluation methodology document should focus on capturing the accents of the application which posts special demand on the system performance.

Emerging trend of the mobile traffic

- Global System for Mobile communications is the most widely used 2G
- Main focus of the 3G mobile will be anywhere, anytime communications for both voice and other types of data transmission



Emerging trend of the mobile traffic

- Internet and multimedia traffic can be characterized by frequent transitions between active and inactive states (ON/OFF patterns)
- ON – file downloading time
- OFF – user reading time
- For the emerging future network traffic, current circuit-switched technique and the Erlang formulas are no longer appropriate to use

What is the traffic modeling?

- Meaning of modeling is anything that represents something else, usually on a smaller scale

- Traffic modeling is the problem of representing our understanding of dynamic demands by stochastic processes.

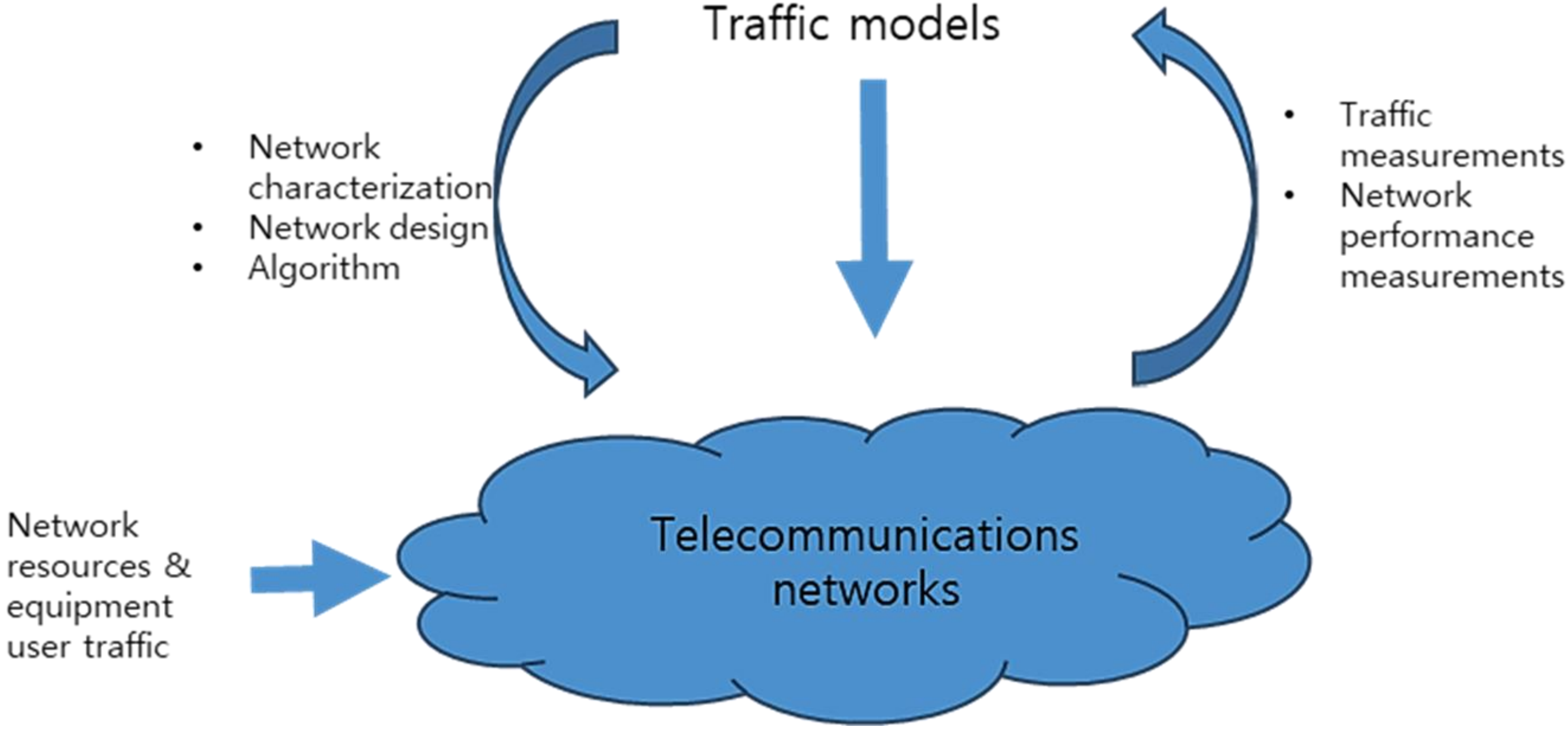
- Accurate traffic models are necessary for service providers to properly maintain quality of service. Many traffic models have been developed based on traffic measurement data.

Importance of traffic modeling

- The aim of traffic modeling is to find stochastic processes to represent the behavior of traffic.
- Support efficient network-dimensioning procedures and traffic management functions
- Assist in characterizing and modeling traffic behavior that is used for accessing QoS
- Help estimate the resource utilization in a network environment

Importance of traffic modeling

- Analysis of the traffic provides information like the average load, the bandwidth requirements for different applications, and numerous other details.
- Traffic models enables network designers to make assumptions about the networks being designed based on past experience and also enable prediction of performance for future requirements.



The role of traffic modeling in telecommunication networks

Traffic modeling criteria

- **General** enough to provide a good approximation to the field data
- **Simple** enough to obtain analytically tractable results for performance evaluation, in terms of
  - Mathematical analysis
  - Programming
  - Computing (fast simulation and numerical analysis)

Traffic modeling criteria

- Three traffic characteristics in traffic modeling

  i) Queuing performance - buffer size and parameters

  ii) Marginal distribution - statistical multiplexing and source traffic control

  iii) Autocorrelation – prediction of queuing behavior

Basis of probability distribution

- Definition of probability distribution
- The conditions of probability distribution
- Probability Density Function (PDF) and Cumulative Distribution Function (CDF)
- Poisson, Exponential, Geometric distribution in statistics
- Markov chain

Definition of probability distribution

- Probability distribution yields the possible outcomes for any random event. It is also defined based on the underlying sample space as a set of possible outcomes of any random experiment. .

- Random experiments are defined as the result of an experiment, whose outcome cannot be predicted.

- The probability distribution gives the possibility of each outcome of a random experiment. It provides the probabilities of different possible occurrences.

The conditions of the Probability distribution

- The probabilities for random events must lie between 0 to 1

$$0 \leq P(A_i) \leq 1$$

$$P(A_i)\text{: Probability of event } A_i,$$
$$i = 1, 2, \dots n \; in \; sample \; space \; S$$

- The sum of all the probabilities of outcomes should be equal to 1.

$$\sum_{i=1}^{n} P(A_i) = 1 \text{ or } \int_{x \in S} p(x)dx = 1$$

PDF and CDF

- The probability density function (PDF) is the probability that a random variable, say X, will take a value exactly equal to x.

$$f(x) = P(X = x)$$

- Probability density function formula

$$P(a \leq X \leq b) = \int_a^b f(x), \qquad x \in \mathbb{R}, a \leq x \leq b$$

where x is a contiunous random variable.

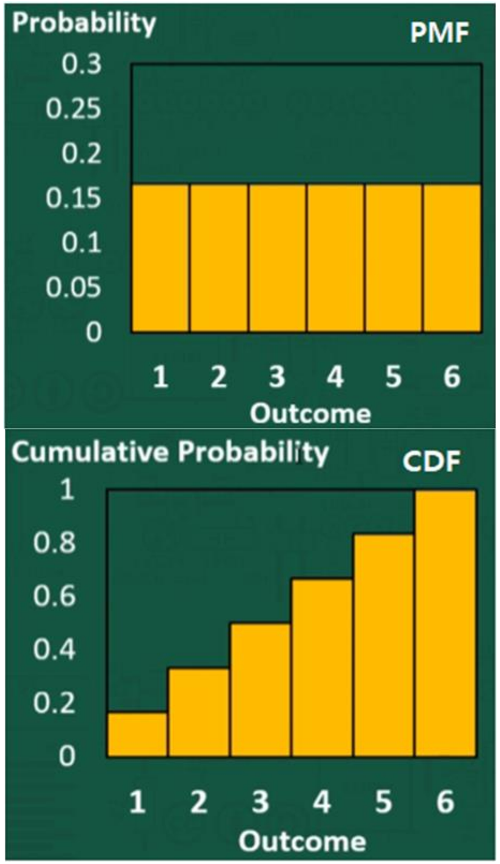
PDF and CDF

- The Cumulative Distribution Function (CDF) is the probability that a random variable, say X, will take a value less than or equal to x.

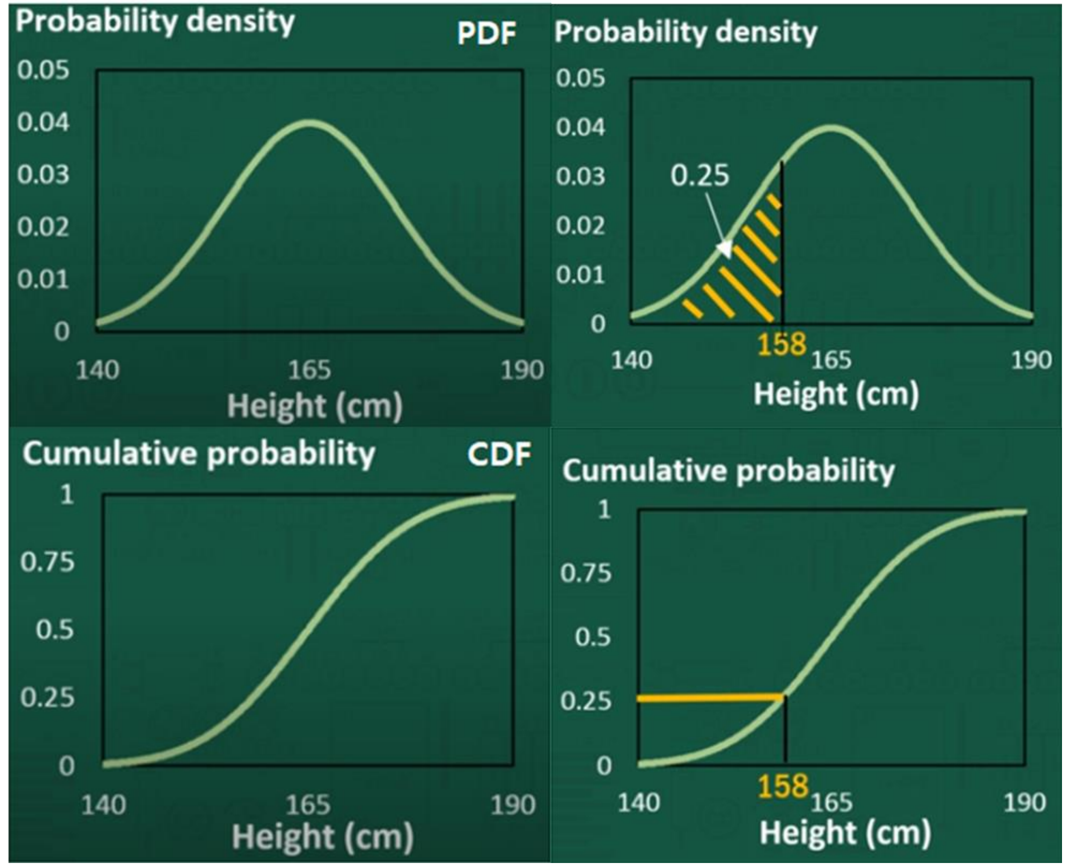
- The cumulative distribution function of a real-valued random variable X is the function given by

$$F(x) = P(X \leq x)$$

- The cumulative distribution function CDF will be more accurate because it has points for each unique data value.



(a) PMF(Probability Mass Function) and CDF for discrete random variables



(b) PDF and CDF for continuous random variables

What Is a Poisson Distribution?

- The Poisson Distribution is a discrete distribution. Its variable can only take the values 0, 1, 2, 3, etc., with no fractions or decimals.

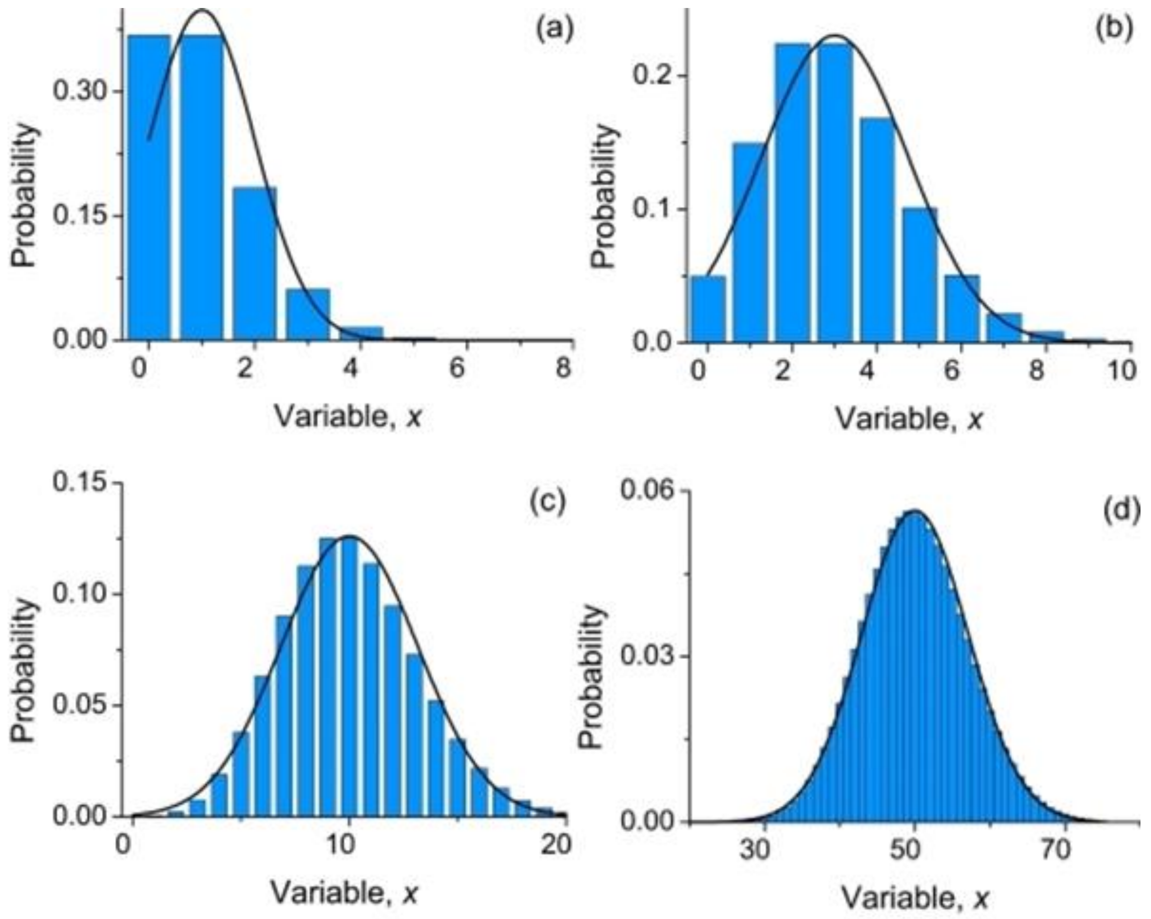
- In statistics, a Poisson distribution is a probability distribution that is used to show how many times an event is likely to occur over a specified period.

$$\boldsymbol{P(x) = \frac{(\lambda)^{x} e^{-\lambda}}{x!}}$$

x: The number of occurrences ($x = 0, 1, 2, \ldots$)

e is Euler's constant $e = 2.71828$ ...

The positive real number λ is equal to the expected value of X, $\lambda = E[X]$



Poisson distribution is usually close enough to Gaussian

The evolution of the Poisson distribution as the mean increases.
(a) When the mean of Poisson distribution is 1
(b) When the mean of Poisson distribution is 3
(c) When the mean of Poisson distribution is 10
(d) When the mean of Poisson distribution is 50

What is a exponential distribution?

- Relationship between exponential distribution and Poisson distribution can be explained by defining random variable of exponential distribution as 'time gap' between Poisson distribution with λ.

- In probability theory and statistics, the exponential distribution is the probability distribution of the time between events in a Poisson point process.

What is a exponential distribution?

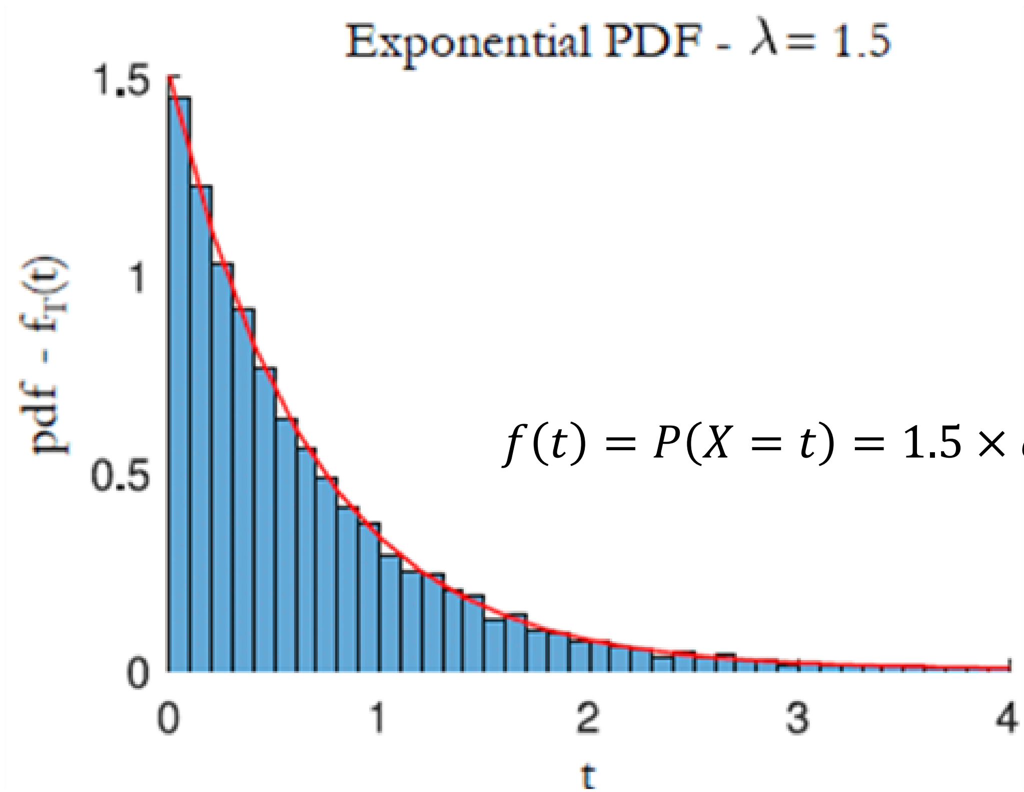
- When X follows the exponential distribution, $X \sim EXP(\lambda)$

- Random variable X expresses the time until the next event occurs.

- So exponential distribution is continuous distribution because it deals with time which is continuous.

$$f(t) = \lambda \times e^{-\lambda t}\ (\lambda > 0)\ ,\quad P(0 \le t \le 1) = \int_0^1 f(t)dt$$

- The same parameter 'lambda λ' which denotes the average rate of events during the specific period.



$$f(t) = P(X = t) = 1.5 \times e^{-1.5t}$$

Example of Exponential PDF where λ is 1.5

What is a Geometric distribution?

- Geometric distribution can be defined as a discrete probability distribution that represents the probability of getting the first success after having a consecutive number of failures.

- Geometric distribution is a type of probability distribution that is based on three important assumptions.

What is a Geometric distribution?

i) The trials being conducted are independent.

ii) There can only be two outcomes of each trial: success or failure.

iii) The success probability, denoted by p, is the same for each trial.

$$P(X = x) = p(1 - p)^{x-1}$$

$x = 1, 2, \dots$: a discrete random variable, $0 \leq p \leq 1$

- The probability of failure is given by q.
  Here, $q = 1 - p$.

What is a Markov Chain?

- A Markov chain or Markov process is a stochastic model describing a sequence of possible events in which the probability of each event depends only on the state attained in the previous event.

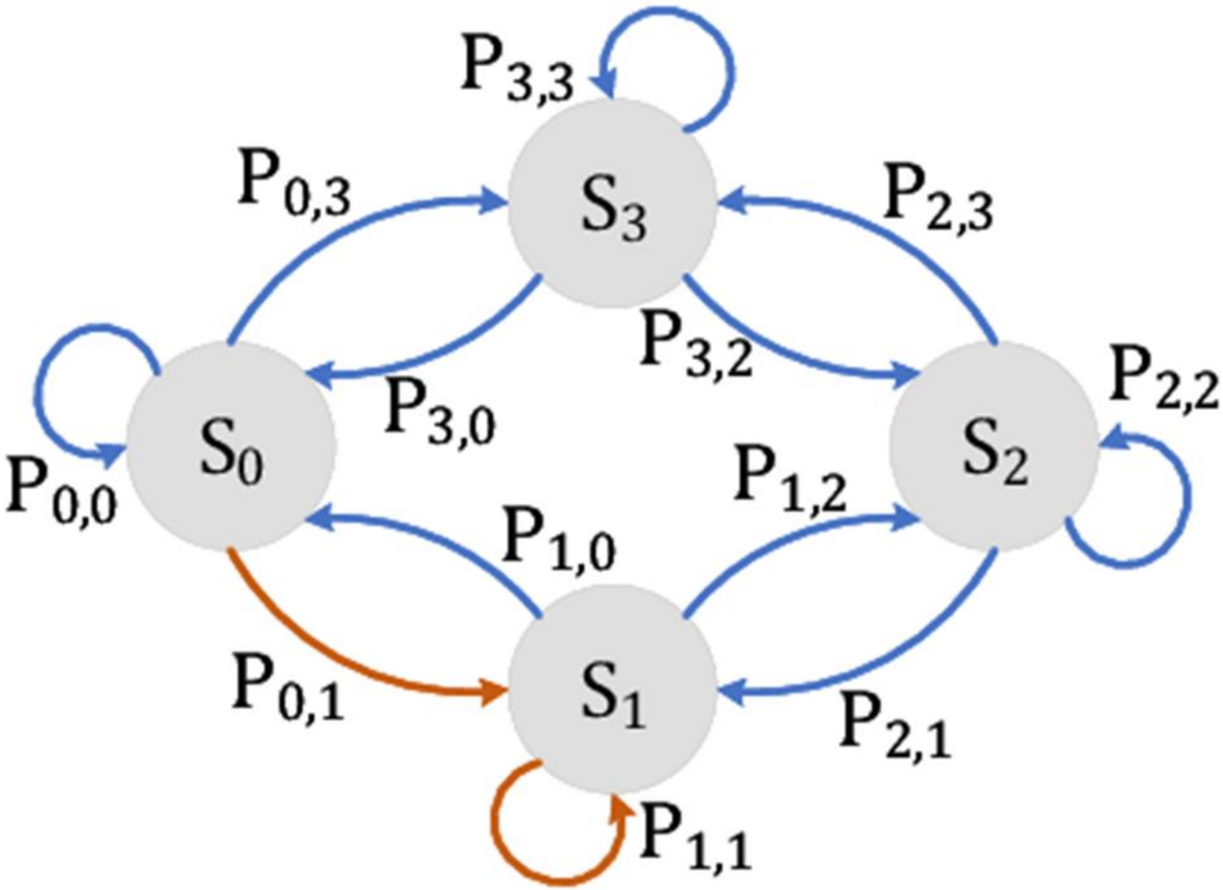
- The conditional probability distribution of the state in the future, given the state of the process currently and in the past, depends only on its current state and not on its state in the past – the future and past states are independent

$$P(X_{n+1} = x | X_n = x_n, \ldots, X_1 = x_1, X_0 = x_0) \\ = P(X_{n+1} = x | X_n = x_n)$$

$x_1, x_2, \ldots, x_n$: **Random variables**

What is a Markov Chain?

- The changes of state of the system are called transitions.
- The probabilities associated with various state changes are called transition probabilities.
- Since the system changes randomly, it is generally impossible to predict with certainty the state of a Markov chain at a given point in the future.
- However, the statistical properties of the system's future can be predicted.



M

$$\begin{bmatrix} P_{0,0} & P_{0,1} & P_{0,2} & P_{0,3} \\ P_{1,0} & P_{1,1} & P_{1,2} & P_{1,3} \\ P_{2,0} & P_{2,1} & P_{2,2} & P_{2,3} \\ P_{3,0} & P_{3,1} & P_{3,2} & P_{3,3} \end{bmatrix}$$

M

$$\forall x \sum_{i=0}^{N} P_{x,i} = 1$$

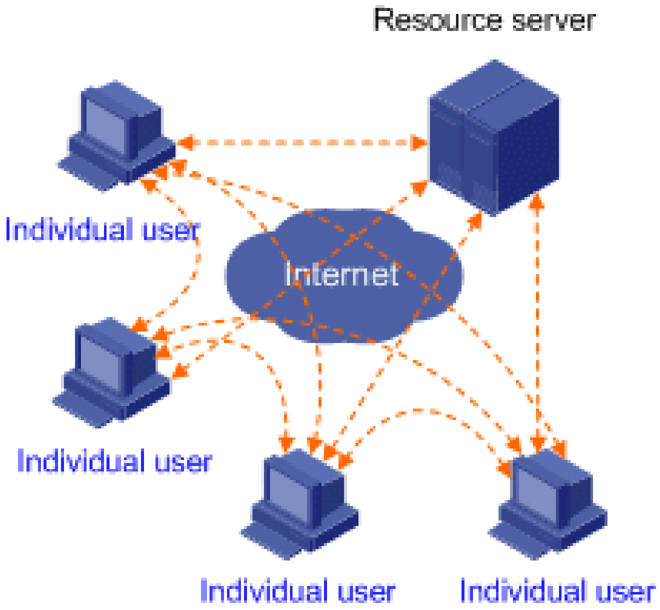
$P_{x,i}: transition\ probability$

Transition: The change from state x to i

Example of Markov chain for workload prediction

Traditional traffic models

- Because of the long history of traditional telephony networks, there are plenty of traffic models available for voice-centric network traffic.



History of Poisson model

- Poisson models have been in use in the literature since the advent of computer networks, and before that in the telecommunications arena.

- Memoryless models are very attractive from an analytical point of view, and a Poisson model can be fit to most network traffic traces reasonably well for short periods with proper selection of parameters .

Poisson model

- The Poisson model is suitable for traffic applications that physically comprise a large number of independent traffic streams.

$$P(x) = \frac{(\lambda t)^x e^{-\lambda t}}{x!}$$

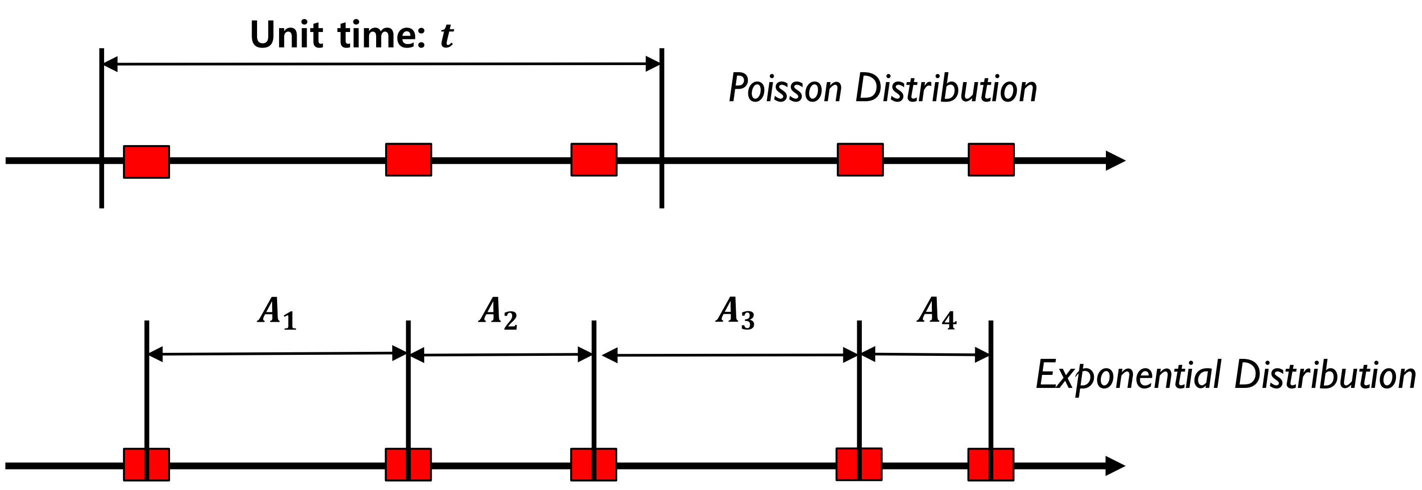
λ:Average number of events that occur

x:The number of individual traffic streams $(x = 0, 1, 2, \ldots, n)$

t: Unit time

Poisson model(continued)

- The time interval between packet arrival $\{A_i\}$ are exponentially distributed. $(i = 1, 2, \ldots, n)$
  - $\rightarrow H(x) = P(x)$: *Poisson Distribution*
  - $\rightarrow G(A_i) = P(A_i \leq t)$=1-$e^{-\lambda t}$: *Exponential Distribution*
  - $\rightarrow G(x), H(i)$ are independent



Limitation of Poisson model

- The model fails to capture the autocorrelation of traffic as it vanishes identically for all nonzero lags
- It is expected that burst data traffic will dominate the future network traffic and it is essential to capture the autocorrelated nature of the traffic for predicting the performance.
- In high-speed data networks, the Poisson process is no longer appropriate

History of Markov Process

- Simple Poisson models exhibit too little burstiness for realistic traffic. Particularly in the telecommunications realm, packet arrivals come in bursts (when people are talking) and pause between bursts (when people are silent).

- In 1986, voice and data were beginning to share telecommunications links, so research into traffic models began to consider the superposition of traffic from multiple types of sources.

History of Markov Process

- The real process underlying a voice network was considered too complex to model effectively.
- Researchers sought a simplified model that still exhibited realistic characteristics.
- What was needed was a traffic model where the arrival rate varied. However, in keeping with the voice model, it made sense to vary the arrival rate in a quantized manner.

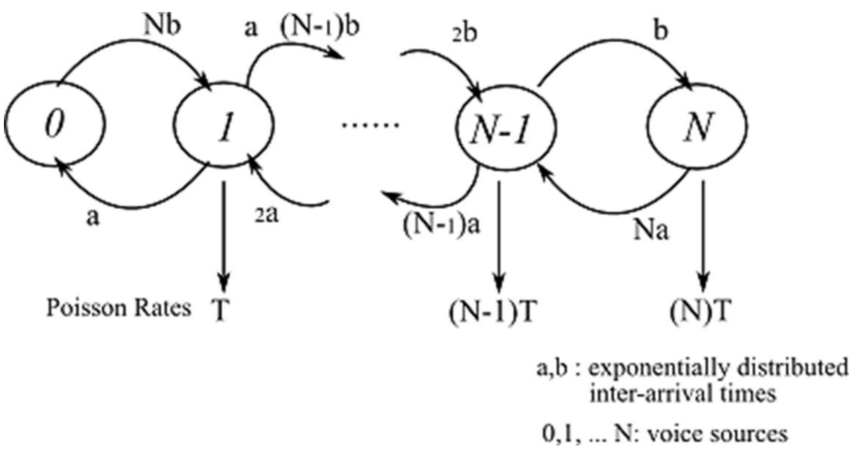
Markov model

- The Markov model introduces some dependency into the random sequence $\{X_1, X_2, \dots\}$; therefore, it capture the traffic 'burstiness'.

$$P(X_{n+1} = x | X_n = x_n, \dots, X_1 = x_1, X_0 = x_0) = P(X_{n+1} = x | X_n = x_n)$$

$x_1, x_2, \dots$: Random variables

- Any traffic modeling requires a multistate Markov and each state adds several free parameter, but it's time consuming to estimate these parameters



Poisson and Markov models

- Inadequate physical explanation for the observed self-similar nature of measured traffic from today's packet networks.

- Lack of studies on its impact on the network, and protocol design and performance analysis.

- Since the traditional traffic models are inadequate to capture today's network characteristics, the packet-switched data traffic models have been developed on the basis of measurements from actual data networks. However, their availability in wireless network modeling is still to be proven.

Three important traffic characteristics

Traffic characteristics	Description
Queuing performance	Buffer size and parameters
Marginal distribution	Statistical multiplexing and source traffic control
Autocorrelation	Prediction of queuing behavior

Three important traffic characteristics in traffic modeling

Three important traffic characteristics

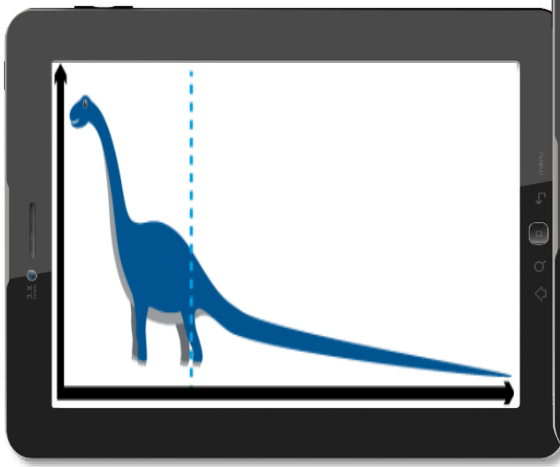
- A good traffic model should be able to capture three of the most important traffic characteristics of the measured data.
- The suitability of a traffic model is primarily determined by its ability to predict the queuing performance.
- More refined models predict a better marginal distribution and autocorrelation of the modeled traffic but usually at the cost of increase in model complexity.

Limitation and the need for new traffic models

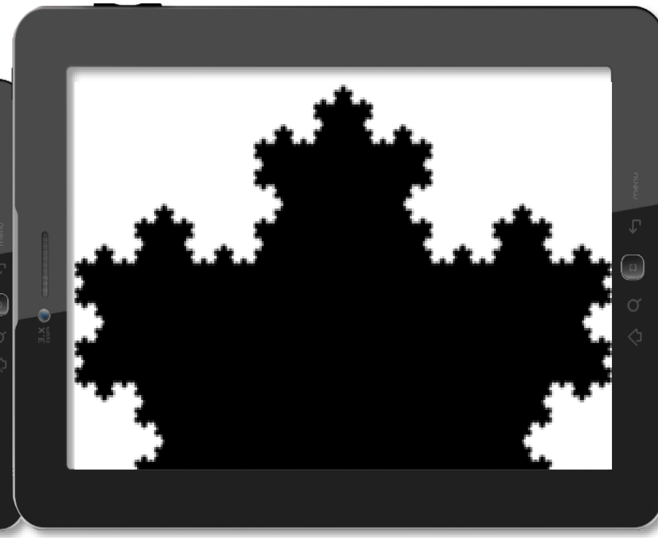
- The aggregate traffic behaves like white noise and fails to capture any of the three most important traffic characteristics above slide
- Recent traffic reveal the prevalence of a LRD on packet-switched networks
- Essential to understand the characteristics of data traffic in order to utilize the network resources and to optimize the network performance
- There's no such model that fits an Erlang-like formula

	Traditional traffic	Emerging network traffic
ON/OFF traffic distribution	*Exponential* or *geometric* distribution(i.e. finite variance distribution)	Heavy-tailed *distribution*(i.e. infinite variance distribution)
Burstiness	Multiplexing traffic streams tend to produce '*smoothed out*' aggregate traffic with reduced burstiness	Aggregate self-similar traffic streams can actually *intensify* burstiness
Aggregate traffic	Gaussian	LRD
Queuing performance	Queue length decreases *exponentially* with increase in buffer size	Buffer gain is *linear* so that queue length decreases linearly
Admission control	Extensive studies are done	Subject of future studies
Congestion control	Extensive studies are done	Subject of future studies

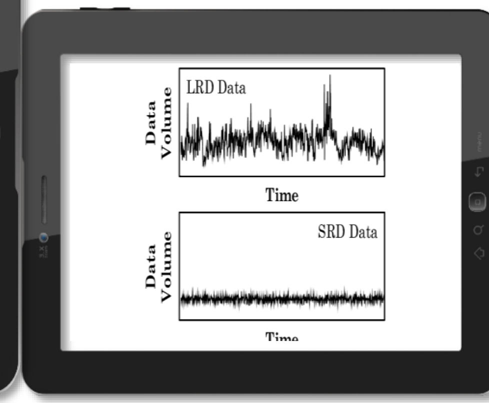
Comparison between traditional and emerging network traffic



Characteristics of
the emerging traffics



Self-similar and LRD
traffic models



Short-range and Long-range
dependence models

Next