Traffic Modeling

Traffic Modeling

- Models that describe and generate telecommunication traffic are important for several reasons :
- **Traffic description:** Network users might be required to give a traffic description to the service provider. Based on that, the service provider decides whether the new connection can be admitted with a guaranteed quality of service and without violating the quality of service for established connections.
- **System simulation:** Future networks and new equipment could be designed and the expected network performance checked.

Traffic Modeling

- Simple traffic models are sometimes called **point processes** since they are basically **counting processes** that count the number of packets that arrive in a time interval. These point processes sometime give the random sequence representing the time separations between packets.
- The other extreme for traffic modeling is to use fluid flow models. Fluid flow modeling groups the traffic into flows that are characterized by average and burst data rates.

Poisson Traffic Properties

1. $\Lambda(t)$ has independent increment.
2. $Pr[\Lambda(t) = j] = exp(-\lambda t)\ (\lambda t)^j /j!,\quad j = 0, 1, 2...$
3. $\Lambda(t)$ has stationary increments: distribution of $\Lambda(t + \Delta t) - \Lambda(t)$ is independent of t.
4. For a short interval, the probability of one arrival is proportional to the interval.
5. For a short interval, the probability of more that one arrivals is negligible.

Poisson Traffic Properties (cont'd)

6. Poisson arrival means the interarrival times are i.i.d. exponential distributed R.V.s.
7. The converse of the above is true.
8. Merging of Poisson processes is still a Poisson process.
9. Splitting of a Poisson process is still Poisson.
10. PASTA (Poisson Arrival Sees Time Average): prob. of an arrival sees k customers in system equals prob. of there are k customers in system.

Interarrival Time for Poisson Traffic

- One of the main characteristics of traffic is the lull period in which no packets arrive.
- We can think of the inter-arrival time between two successive packets as a random variable (R.V.) *T*.
- This R.V. is continuous for Poisson traffic.

Interarrival Time for Poisson Traffic

To study the random variable T we need to study the probability $p(0)$ that no packets arrive in the period t.

Let us start by assuming Poisson traffic with probability $p(k)$ that k packets arrive in a time period t which is given by

$$p(k) = \frac{(\lambda_a\, t)^k}{k!} e^{-\lambda_a\, t}$$

where λ_a (packets/s) is the average packet arrival rate.

Interarrival Time for Poisson Traffic

For the interarrival time, we ask a different sort of question: what is the probability that the time separation between adjacent packets is t?

To derive an expression for the PDF distribution for the interarrival time, we need to find the probability that no packets arrive in period t.

Interarrival Time for Poisson Traffic

Probability that no packets arrive in a time period t is obtained by substituting $k = 0$ in the above equation

$$p(0) = e^{-\lambda_a\, t}$$

Interarrival Time for Poisson Traffic

$p(0)$ is equivalent to the event $A : T > t$ and we can write

$$\begin{aligned} p(A : T > t) &= p(0) \\ &= e^{-\lambda_a\, t} \end{aligned}$$

The event A is basically the event that no packets arrived for a time period t. What happens after this time period is not specified. A packet might arrive or no packets arrive.

Interarrival Time for Poisson Traffic

In order to find the PDF associated with the interarrival time we need to define event B which is complementary to A as follows:

$$B : T \leq t$$

Interarrival Time for Poisson Traffic

The probability associated with event B is

$$\begin{aligned} p(B : T \leq t) &= 1 - p(A) \\ &= 1 - e^{-\lambda_a\, t} \end{aligned}$$

The CDF for the random variable T is given by

$$F_T(t) = p(T \leq t) = 1 - e^{-\lambda_a\, t}$$

Interarrival Time for Poisson Traffic

The PDF for this random variable is obtained by differentiating the above equation

$$f_T(t) = \lambda_a e^{-\lambda_a t}$$

Thus the PDF for the interarrival time of Poisson traffic follows the **exponential distribution**

Example

Find the average value for the exponentially distributed interarrival time.

The average time between arriving packets T_a is given by

$$\begin{aligned} T_a &= \int_{t=0}^{\infty} t\, \lambda_a e^{-\lambda_a t}\, dt \\ &= \frac{1}{\lambda_a} \qquad \text{s} \end{aligned}$$

We see that as the rate of packet arrival decreases ($\lambda_a \ll 1$), the average time between packets increases as expected.

Example

Consider an ATM channel where a source transmits data with an average data rate of 500 kbps. Derive the corresponding Poisson distribution and find the probability that 10 cells arrive in a period of 1 ms. Write down the pdf for the interarrival time.

Convert the arrival rate from bps to cells/s.

$$\lambda_a = \frac{500 \times 10^3}{8 \times 53} = 1.179\,2 \times 10^3 \qquad \text{cells/s}$$

Probability of 10 cells arriving in time t is

$$p(10) = \frac{(\lambda_a\, t)^{10}}{(10)!} e^{-\lambda_a\, t} = 4.407 \times 10^{-7}$$

For interarrival time we have

$$f_T(t) = 1.179\,2 \times 10^3\, e^{-1.1792\, t \times 10^3}$$

Realistic Models for Poisson Traffic

The Poisson distribution and the associated interarrival time do not offer much freedom in describing realistic traffic sources since they contain one parameter only: λ (packets/s) that reflected the average data arrival rate.

The minimum value for the interarrival time is zero.

This implies that the time interval between two packet headers could be zero.

An interarrival time value of zero implies two things: that our packets have zero length and the data rate could be infinity.

Both of these conclusions are not realistic.

Realistic Models for Poisson Traffic

A realistic bursty source is typically described using some or all of these parameter:

λ_a	the average data rate
σ	the maximum data rate expected from the source
A	the average packet length

A source is said to be conforming if its data rate is below a certain specified value. Typically when the source is nonconforming, its data rate is σ.

Realistic Models for Poisson Traffic

PDF of interarrival time is the exponential distribution

$$f_T\,(t) = \lambda_a e^{-\lambda_a\; t}$$

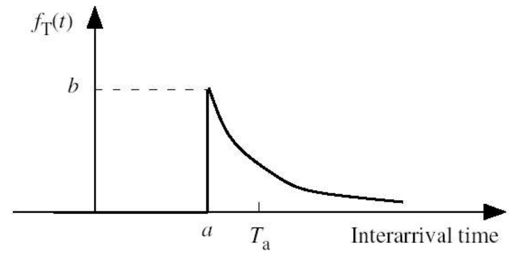
The **biased exponential distribution** as follows.

$$f_T(t) = \begin{cases} 0 & t < a \\ b\ \exp -b\,(t-a) & t \geq a \end{cases}$$

where $a \geq 0$ is the **position parameter** (units s) and $b > 0$ is the **shape parameter** (units s^{-1}).

Realistic Models for Poisson Traffic

$$f_T(t) = \begin{cases} 0 & t < a \\ b\ \exp -b\,(t-a) & t \geq a \end{cases}$$



A biased exponential distribution with two design parameters: position parameter a and shape parameter b.

Example

Find the average value for the exponentially distributed interarrival time with PDF given by the biased exponential distribution.

The average time separation between arriving packets T_a is given by

$$T_a = \int_{t=a}^{\infty} t\, b e^{-b(t-a)}\; dt = a + \frac{1}{b} \qquad \text{s}$$

When b is large ($b \gg 1$), the exponential function will approach a delta function and the interarrival time will have its **minimum value** $T_a \approx a$.

$$\sigma^2 = \frac{1}{b^2}$$

Large values for b will result in traffic with low burstiness approaching CBR. Lower values for b result in more bursty traffic.

Extracting Poisson Traffic Parameters

We show how to find the values of the position parameter a and shape parameter b for a source whose average rate λ_a and burst rate σ are known.

Once we know a and b we will be able to

1. Generate accurate traffic for numerical simulations.
2. Develop accurate queuing parameters for analytical simulations.

Extracting Poisson Traffic Parameters

The position parameter a is equivalent to the minimum time between two adjacent packets.

In a time period t, the **maximum number** of packets that could be produced by the source is given by

$$N_m = \sigma\; t$$

Extracting Poisson Traffic Parameters

The minimum time between two adjacent packets was defined as a and is given by

$$a = \frac{t}{N_m} = \frac{1}{\sigma} \qquad \text{s}$$

Extracting Poisson Traffic Parameters

In a time period t, the **average number** of packets that could be produced by the source is given by

$$N_a = \lambda_a \, t$$

Extracting Poisson Traffic Parameters

The average time between two adjacent packets is given by

$$T_a = \frac{t}{N_a} = \frac{1}{\lambda_a} \qquad \text{s}$$

But the average interarrival time as

$$T_a = a + \frac{1}{b} \qquad \text{s}$$

Extracting Poisson Traffic Parameters

From the above two equations we are able to obtain a value for the shape parameter b

$$\frac{1}{\lambda_a} = a + \frac{1}{b}$$

Therefore we have

$$b = \frac{\sigma \, \lambda_a}{\sigma - \lambda_a} \qquad \text{s}^{-1}$$

Example

A data source follows the Poisson distribution and has an average data rate $\lambda_a = 10^3$ packets/s and maximum burst rate of $\sigma = 3 \times 10^3$ packets/s. Estimate the exponential distribution parameters that best describe that source.

The position parameter is given by

$$a = \frac{1}{3 \times 10^3} = 3.33 \times 10^{-4} \qquad \text{s}$$

The shape parameter b is given by

$$b = 1500 \qquad \text{s}^{-1}$$

The PDF for the interarrival time is given by

$f_T(t)$=1500 exp (-1500(t-0.00033)) for t>0.00033 s

Self-similar Traffic

We are familiar with the concept of periodic waveforms. A periodic signal repeats itself with **additive** translations of time. For example the sine wave $\sin \omega t$ will have the same value if we add an integer multiple of the period $T = 2\pi/\omega$ since

$$\sin \omega t = \sin \omega (t + i\, T)$$

Self-similar Traffic

Self-similar signal repeats itself with **multiplicative** changes in the time scale.

Thus a self-similar waveform will have the same shape if we scale the time axis up or down. In other words, imagine we observe a certain waveform on a scope when the scope is set at 1 ms/division. We increase the resolution and set the scale to 1 μs/division. If the incoming signal is self-similar, the scope would display the same waveform we saw earlier at a coarser scale.

Self-similar Traffic

Self-similar traffic describes traffic on Ethernet LANs and variable-bit-rate video services.

These results were based on analysis of millions of observed packets over an Ethernet LAN and an analysis of millions of observed frame data generated by variable bit rate (VBR) video.

Self-similar Traffic

The effect of self-similarity is to introduce long range (large lag) autocorrelation into the traffic stream which is observed in practice.

This phenomenon leads to periods of high traffic volumes even when the average traffic intensity is low.

A switch or router accepting self-similar traffic will find that its buffers will be overwhelmed at certain times even if the expected traffic rate is low.

Thus switches with buffer sizes selected based on simulations using Poisson traffic will encounter unexpected buffer overflow and packet loss.

Pareto Traffic Distribution

The Pareto distribution is used here to describe realistic traffic sources that have bursty behavior.

The Pareto distribution is describe by the PDF

$$f(x) = \frac{b\, a^b}{x^{b+1}}$$

where a is the position parameter, b is the shape parameter, and the random variable X has values limited in the range $a \le x < \infty$.

Pareto Traffic Distribution

The mean and variance for X are

$$\mu = \frac{b\, a}{b-1}$$

$$\text{Var} = \frac{b\, a^2}{(b-1)^2\,(b-2)}$$

Pareto Traffic Distribution

A realistic bursty source is typically described using some or all of these parameter:

- λ_a the average data rate
- σ the maximum data rate expected from the source
- A the average packet length

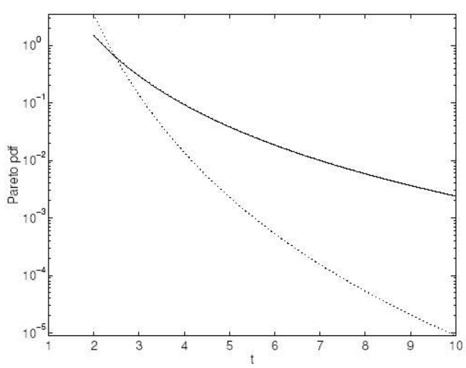
Interarrival Time of Pareto Traffic

The interarrival time following the Pareto distribution has a PDF that is given by

$$f_T(t) = \frac{b\,a^b}{t^{b+1}} \qquad \text{with } a \le t < \infty$$

where a (units seconds) is the position parameter and $b \geq 1$ is the shape parameter.

Interarrival Time of Pareto Traffic



The PDF distribution for the case when $a = 2$ and $b = 3$ (solid line) and $b = 7$ (dashed line).

Extracting Pareto Interarrival Time Statistics

In a time period t, the maximum number of packets that could be produced by the source is given by

$$N_m = \sigma\, t$$

We use this estimate to calculate the minimum time between two adjacent packets as follows.

$$a = \frac{t}{N_m} = \frac{1}{\sigma} \qquad \text{s}$$

The position parameter depends only on the average packet size and burst rate.

Extracting Pareto Interarrival Time Statistics

In the time period t, the average number of packets that could be produced by the source is given by

$$N_a = \lambda_a\, t$$

Extracting Pareto Interarrival Time Statistics

The average time between two adjacent packets is given by

$$T_a = \frac{t}{N_a} = \frac{1}{\lambda_a} \qquad \text{s}$$

But from the Pareto PDF distribution, the average interarrival time is given by

$$T_a = \int_{t=a}^{\infty} t\, \frac{b\,a^b}{t^{b+1}}\, dt = \frac{b\,a}{b-1} \qquad \text{s}$$

From the above two equations, we are able to obtain a value for the shape parameter b.

Extracting Pareto Interarrival Time Statistics

$$\frac{1}{\lambda_a} = \frac{b\,a}{b-1}$$

Therefore we have

$$b = \frac{\sigma}{\sigma - \lambda_a}$$

The shape parameter depends only on the average rate λ_a and burst rate σ.

Extracting Pareto Interarrival Time Statistics

Furthermore, the shape parameter lies between the following extreme values

$$\begin{array}{lcll} b & = & 1 & \text{when } \sigma \gg \lambda_a \\ b & \to & \infty & \text{when } \lambda_a \to \sigma \end{array}$$

The first expression applies to a fairly bursty source and the second expression applies to a constant bit rate source where the average data rate equals the burst rate.

Extracting Pareto Interarrival Time Statistics

Thus the range of the shape parameter b can be expressed as

$$1 \leq b < \infty$$

Lower values of b imply bursty sources and higher values of b imply sources with little variations in the interarrival times since we would have a constant rate source.

Example

A bursty source produces data at an average rate of 5 Mbps and its maximum burst rate is 20 Mbps. Estimate the Pareto parameters that best describe that source assuming that the average packet size is 400 bits.

The position parameter is

$$a = \frac{1}{\sigma} = 20 \qquad \mu\text{s}$$

The average data rate is used to determine the shape parameter b

$$b = \frac{\sigma}{\sigma - \lambda_a} = 1.333$$

Fluid Flow Pareto Traffic Description

PDF for the instantaneous data rate in the form

$$f_R(\lambda) = \begin{cases} 0 & \text{when} \quad \lambda < \lambda_{min} \\ b\,a^b/\lambda^{b+1} & \text{when} \quad \lambda \geq \lambda_{min} \end{cases}$$

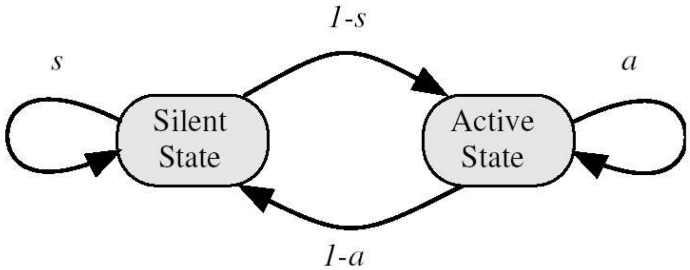
Fluid Flow Pareto Traffic Description

The values of the two parameters a and b can be found for a source with traffic descriptors $(\lambda_{min}, \lambda_a, \sigma)$ as follows.

$$\begin{array}{lcl} a & = & \lambda_{min} \\ b & = & \dfrac{\lambda_a}{\lambda_a - \lambda_{min}} \end{array}$$

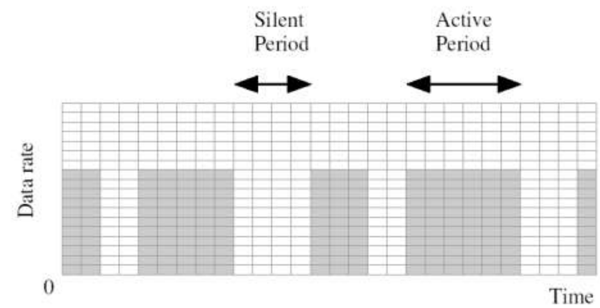
On-off Model

The source switches between an active state, producing packets, and a silent state where no packets are produced.



An on-off source model.

On-off Model



Traffic pattern for an on-off source model.

On-off Model

The probability that the length of the active period is n time steps is given by the geometric distribution

$$A(n) = a^n(1-a) \qquad n \geq 1$$

The average duration of the active period is given by

T_a= 1+ a/(1-a) = 1/(1-a) time steps

On-off Model

the probability that the length of the silent period is n time steps is given by the geometric distribution

$$S(n) = s^n(1-s) \qquad \text{n>=1}$$

The average duration of the silent period is given by

T_s= 1+ s/(1-s) = 1/(1-s) time steps

On-off Model

Assume that λ is the data rate when the source is in the active state. In that case, the average data rate is obtained as

$$\begin{aligned} \lambda_a &= \frac{\lambda \times T_a}{T_a + T_s} \\ &= \frac{\lambda}{1 + T_s/T_a} \leq \lambda \end{aligned}$$

Example

Assume a 64 kbit/s voice source which is modeled as an on-off source with an average duration of the active period of $T_a = 0.45$ s and average duration of the silent period is $T_s = 1.5$ s. Estimate the source parameters and the average data rate.

Choose time step duration 0.01 second.
T_a = 45 time steps,
T_s = 150 time steps.

Probability the source remains in active state is
$a = (T_a - 1)/T_a = 44/45$

Probability the source remains in silent state is
$s = (T_s - 1)/T_s = 149/150$

The average data rate is
λ_a = 14.77 Kbps

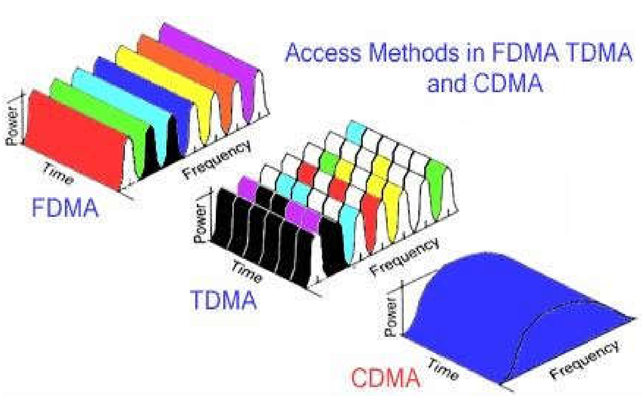
Example

- Following the previous example, if we have five such voice sources sharing a link, what is the probability that more than three sources are in active period simultaneously?

- Solution:
  $S_a = T_a/(T_a+T_s) = 45/195$
  Pr{k active among N users}
  $=P(k)= [N!/(k!(N-k)!)]\ S_a^{\ k}\ S_s^{\ N-k}$
  Pr{more than 3 active among N users}
  $= P(4) + P(5) = 0.0842$

Case Study: Voice activity factor for CDMA systems

Access methods

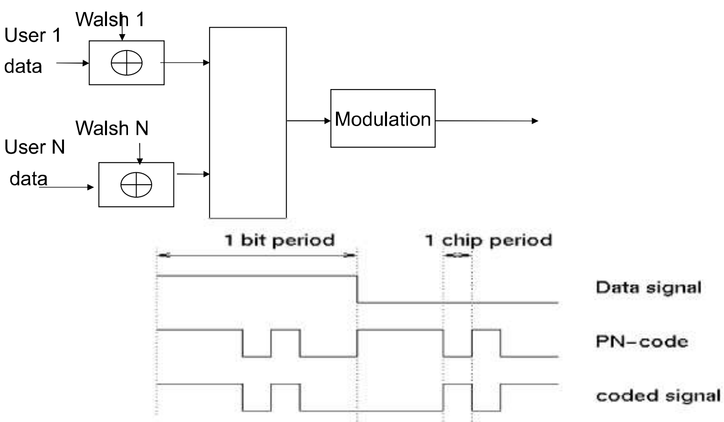


http://3.bp.blogspot.com/_UatLVeBY0O0/S9kyHfby8dI/AAAAAAAAAFo/rL9p_ciEbZc/s1600/access-methods-in-fdma-tdma-and-cdma1.jpg

CDMA technologies

- Frequency-hopping spread spectrum: invented by actress Hedy Lamarr and composer George Antheil, for anti-jamming
- Direct-sequence, time-hopping, chirp spread spectrums, and their combinations

CDMA system, downlink



– Walsh code: orthogonal sequence

CDMA system, how it works

- Example: user 1 uses a Walsh code of "1 0 1 0" (i.e., [1 -1 1 -1]), and user 2 uses a Walsh code of "1 1 0 0" (i.e., [1 1 -1 -1]).
- When user 1 transmits a bit "1" and user 2 transmits a bit "0", after Walsh code, the sequences are [1 -1 1 -1] and [-1 -1 1 1]. After summation, the sequence is [0 -2 2 0]
- At the receiver side,

  user 1: $[0\ -2\ 2\ 0]\ [1\ -1\ 1\ -1]^T = 4$ (>0) -> bit "1"

  user 2: $[0\ -2\ 2\ 0]\ [1\ 1\ -1\ -1]^T = -4$ (<0) -> bit "0"

CDMA Capacity

In CDMA the total noise density is

$$N_o^{'} = N_o + I_o = N_o + U\frac{E_b R_b}{W_s}$$

$$\frac{E_b}{N_o + I_o} = \frac{E_b}{N_o^{'}} = \frac{E_b}{N_o + U\frac{E_b R_b}{W_s}}$$

where U is the number of users

CDMA Capacity (cont'd)

In most practical cases the thermal noise N_o is negligible in comparison to mutual interference (also called self-noise). Then

$$\frac{E_b}{N_o^{'}} = \frac{E_b}{N_o + U\frac{E_b R_b}{W_s}} \approx \frac{1}{U\frac{R_b}{W_s}} \rightarrow \mathrm{U} \approx \frac{G_p}{\frac{E_b}{N_o^{'}}}$$

$\frac{W_s}{R_b} = \frac{R_c}{R_b} = G_p$ is known as the processing gain

Notice that here the minimum Nyquist bandwidth is assumed for W_s (while Rappaport assumes twice the minimum bandwidth).

Consider Voice Activity Factor

- If the carrier is turned off when the speaker is silent, the interference reduces from $U\ E_b R_b/W_s$ to

  $V\ U\ E_b R_b/W_s$, where the voice activity factor V is typically 0.35 to 0.4.
- Thus, approximately, the number of users can be increased to 1/V of the one without considering the voice activity factor.

Example

- *A CDMA system uses direct-sequence BPSK modulation with a data rate R= 5 kbps. Assume 30 equal power data users. Ignore the thermal noise.*
  - *a) Determine the minimum chip rate (number of pulses per second) to obtain a BER of 10^{-5}.*
  - *b) Repeat the calculations if 15 users are data the other 15 users are voice users with voice activity factor of 0.4.*

Solution

- *a) BER of 10^{-5} requires E_b/N_0'=10*

$$U = \frac{G_p}{\frac{E_b}{N_o'}} \rightarrow 30 = \frac{G_p}{10} \rightarrow G_p = 300$$

This means a chip rate of
$R_c = 5 \times 300 = 1.5\text{Mcps}$

Solution (cont'd)

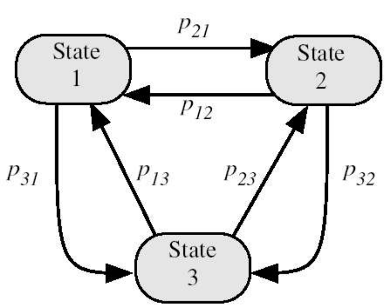
(b) 15 voice activated users are equivalent to 0.4x15=6 data users, then the total is equivalent to 21 data users.

$$U = \frac{G_p}{\frac{E_b}{N_o'}} \rightarrow 21 = \frac{G_p}{10} \rightarrow G_p = 210$$

This means a chip rate of
$R_c = 5 \times 210 = 1.05\text{Mcps}$

Markov Modulated Poisson Process (MMPP)

This is a generalization of the on-off source model.



A three-state MMPP source model.

Video traffic modeling

- Raw, uncompressed video traffic has constant bit rate (CBR). E.g.,
  - A video stream has a rate of 30 frame/sec;
  - using North American standard, each frame has 250,000 pixels;
  - If 8-bit gray scale is used (256 levels of amplitude), the video has 2Mbits/frame, or 60Mbps.
  - For three-color, the video has 180Mbps.

Video traffic modeling

- With compression (MPEG), the data rate will be reduced to several Mbps **on average**.
- But the traffic has variable bit rate (VBR).

Video traffic modeling

- In the literature, we can use mini-source model for VBR video traffic.
- A video traffic can be viewed as a multiplexing of *N* On/Off sources.
- The value of N is related to the burstiness of video.
- The On/Off model parameters should be chosen to match the video statistics.

Packet Length Statistics

- Unlike ATM, many protocols produce packets that have variable lengths.
- Examples of protocols that have variable length packets are IP, Frame Relay, and Ethernet.
- Knowledge of the packet size is essential if one wants to estimate the buffer space required to store data having a certain arrival distribution statistics.

Packet Length Statistics

Exponential distribution could be used to provide a simple model for packet length statistics.

The probability of receiving a packet of length x is given by

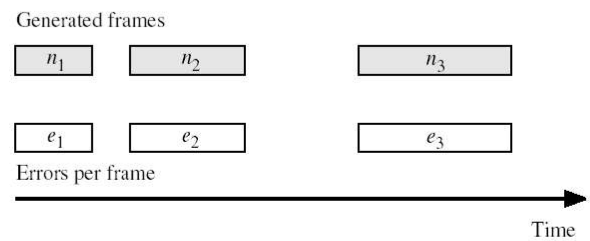
$$p(x) = \frac{1}{\mu} e^{-x/\mu}$$

where μ is the average packet length.

Packet Transmission Error Description

- The previous sections dealt with issues related to network traffic such as data rate variation, packet length variation, and packet destination.
- When the packets are in transit, they are corrupted due to channel impairment or they could be totally lost due to congestion or address errors.

Packet Transmission Error Description



Time series sequence of generated data and channel errors. A received frame is in error if it is generated at the same time that an error is generated.

Example

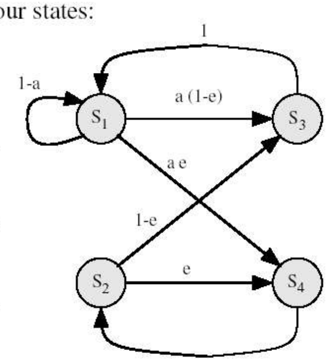
Assume an on-off data source that generates equal length frames with probability a per time step.

Assume for simplicity that each frame contains only one packet. The channel introduces errors in the transmitted frames such that the probability of a packet is received in error is e.

Perform a Markov chain analysis of the system and derive its performance parameters.

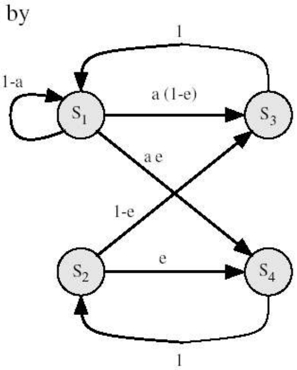
The Markov chain model we use has four states:

State	Significance
1	Source is idle
2	Source is retransmitting a frame that was in error
3	Frame is transmitted with no errors
4	Frame is transmitted with an error



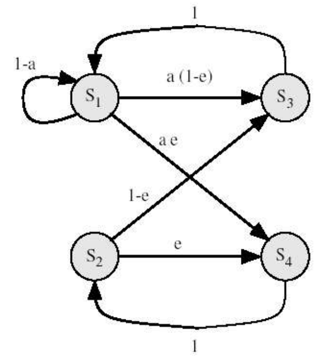
Transition matrix for the system is given by

$$\mathbf{P} = \begin{bmatrix} 1-a & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ a(1-e) & 1-e & 0 & 0 \\ a\,e & e & 0 & 0 \end{bmatrix}$$



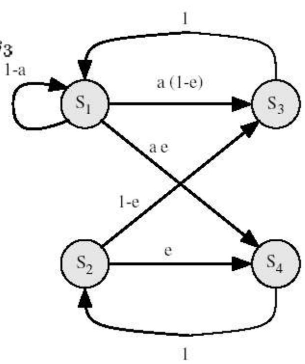
The system throughput is given by

$Th = s_3$



The average number of lost packets per time step is given by

$$N_a(lost) = N_a(in) - N_a(out)$$
$$= a - Th = a - s_3$$



The probability that the packet will be transmitted is

$$p_a = \frac{Th}{a} = \frac{s_3}{a}$$

The packet loss probability is

$$L = \frac{N_a(lost)}{a} = 1 - \frac{s_3}{a}$$

The average number of retransmissions is given by

$$W = \frac{1-p_a}{p_a} = \frac{a}{s_3} - 1$$

$e1 = 0.1$
(solid line)
$e2 = 0.6$
(dotted line)