Basics of traffic modeling I

Lecturer: Dmitri A. Moltchanov
E-mail: dmitri.moltchanov@tut.fi

OUTLINE:

• Traffic description;

• Point processes;

• Poisson process;

• Service processes;

• Internet traffic properties;

• Step-by-step traffic modeling;

• Exponential-based time interval distributions;

• Example of modeling: two-moments fitting.
1. Traffic definition: packet (IP) level

In packet-switching networks we talk about transmission needs. Let:

- any packet can be of \( s \) units in length (units, e.g. bits or bytes);
- any link is characterized by a capacity \( \phi \) (units per second).

Then the service time for a customer (so-called transmission time) is:

\[
\frac{s}{\phi}, \text{ seconds.} \tag{1}
\]

Utilization \( \rho \) of the link is:

\[
\rho = \frac{\lambda s}{\phi}, \quad 0 < \rho < 1. \tag{2}
\]

- \( \lambda \) is arrival rate of packets per time unit (1/s.).
2. Traffic definition: call (session) level

In circuit-switching networks we talk about occupation times. Let:

- a transmission line is (fully) occupied for $x$ seconds;
- out of some fixed reference interval of $\Delta$ second.

Then, the measurement unit is:

$$\frac{x}{\Delta}, \text{Erlang},$$

which is dimensionless (this is good as it is easy to deal with);

- one Erlang: a line if fully busy for a reference unit.

Applications:

- heavily used in circuit-switching networks;
- can be used for virtual switching networks (e.g. MPLS);
- can be used to describe session level load in packet-switching networks.
3. Point processes

Two major properties are:

- how do customers arrive to the system:
- how is the service process described:

**Point processes:** math abstraction to describe arrivals to the system!
3.1. Examples of point processes

Example: a telephone exchange

- we consider call arriving process to that exchange.

![Diagram of call arriving process at a telephone exchange]

Figure 2: Call arriving process at a telephone exchange.

Note: due to such visual representation such process is called point process.
Example: a hypothetical router

- we consider packet arriving process to the output port.

![Diagram of a hypothetical router](image)

Figure 3: Packet arriving process at the output buffer.
4. Description of point processes

Let us assume:

- we are given point process $N(t), t \geq 0$;
- **fundamental property**: multiple arrivals do not occur!

Does this represents reality?

- does not hold for discrete systems;
- everything is slotted, clocks are synchronized;
- may produce fair approximation;
- often called **regularity** or **batchless** property.
Denote the time distance between two successive arrivals:

\[ X_i = T_i - T_{i-1}, \quad i = 1, 2, \ldots \]  

(4)

Note the following:

- \( X_i, \ i = 1, 2, \ldots \) are called interarrival times;

- distribution of \( X \) is called interarrival time distribution:
  - note that \( X_i, \ i = 1, 2, \ldots \) may have different distributions;
  - we are interested in the case when they have the same distribution.
4.1. Interval and number representations

Using RVs $N(t)$ and $X_i$ arrival process can be characterized:

- **number representation $N(t)$**
  - time interval is constant $[t_1, t_2)$;
  - we observe number of arrivals in $[t_1, t_2)$.

- **interval representation**
  - number of arriving customers is kept constant $n = 1, 2, \ldots$;
  - we observe time until $n = 1, 2, \ldots$ arrivals occur $X_i^{(n)}$;
  - **note**: $n = 1$ giving rise to $X_i, i = 1, 2, \ldots$ is used in practice.

![Network analysis and dimensioning I](image-url)
4.2. Basic properties of interval and number representations

Properties of number representation \{N(t), t > 0\}:

- \(N(t)\) is always a discrete RV;
- when \(t\) increases \(N(t)\) never decreases:
  \[N(t_2) \geq N(t_1), \quad t_2 \geq t_1.\] (5)

Properties of interval representation \(\{X_i^{(n)}, i = 1, 2, \ldots\}\) for fixed \(n\):

- \(X_i\) could be continuous or discrete RV (discrete or continuous time);
- \(X_i > 0\) for all \(i = 1, 2, \ldots\):

Important notes:

- in practice interval representation \(\{X_i^{(1)}, i = 1, 2, \ldots\}\) is always used;
- \(\{X_i^{(1)}, i = 1, 2, \ldots\}\) is denoted simply by \(\{X_i, i = 1, 2, \ldots\}\)
Graphical relation between $N(t)$ and $X_i$, $i = 1, 2, \ldots$ (discrete-time case).

Figure 4: Graphical relation between $N(t)$ and $X_i$, $i = 1, 2, \ldots$. 

Lecture: Basics of traffic modeling I
4.3. Relation to traffic measurements

Note the following:

- it is up to you which representation to work with;
- the choice of representation is often dictated by measurements!

Basic measurement methods:

- 1st method:
  - recording number of arrivals at constant time intervals;
  - **shortcoming**: you cannot get interval representation;
  - **advantage**: easy to implement and requires less resources.

- 2nd method:
  - recording time when arrival occurs;
  - **advantage**: you can get both representation;
  - **shortcoming**: require more resources than the first approach.
Get number representations:

- start observing arrivals, set timer to $t = 0$ and start it;
- stop observing when $t = T$, write $N(t) = n$, reset timer and start observing again.

Get interval representations:

- start observing, set timer to $t = 0$ and start it;
- observe the arrival, write $X_i = t$, reset timer and start observing again.
4.4. Properties of the number representation

The following properties hold:

- the total and mean number of arrivals at time interval \([t_1, t_2]\) is:
  \[
  H(t_1, t_2) = N(t_2) - N(t_1), \quad E[H(t_1, t_2)] = E[N(t_2) - N(t_1)].
  \]  

- time average of arriving calls at time \(t\) is:
  \[
  \lambda(t) = \lim_{\Delta t \to 0} \frac{N(t + \Delta t) - N(t)}{\Delta t}.
  \]  
  - we will always assume that \(\lambda(t)\) is finite;
  - \(\lambda(t)\) is instantaneous intensity with which arrivals occur at time \(t\)

- to describe variation one can use index of dispersion for counts (IDC):
  \[
  C(t) = \frac{D[N(t)]}{E[N(t)]^2}.
  \]  
  - \(E[N(t)]\): mean of the \(N(t)\);
  - \(D[N(t)]\): variance of the \(N(t)\).
4.5. Properties of the interval representation

The following properties hold:

- $X_i$ is fully characterized by its CDF:
  \[ F_{X_i}(t) = Pr\{X_i \leq t\}. \]  
  \[ (9) \]

- mean is the arrival average and given by:
  \[ E[X_i] = \int_0^\infty t \, dF_{X_i}(t) = \int_0^\infty tf_{X_i}(t) \, dt. \]  
  \[ (10) \]

- to describe variation one may use index of dispersion for intervals (IDI):
  \[ IDI = \frac{D[X_i]}{(E[X_i])^2}, \]  
  \[ (11) \]
  - $D[X_i]$ and $E[X_i]$ are variance and mean, respectively;
  - note: estimating IDI some type of smoothing is always introduced!

Important: point process for which sequential interarrival times are iid is called renewal one.
5. Poisson process

A point process with the following properties:

- regular: no multiple arrivals in $dt$;
- memoryless: probability of arrival in $dt$ is independent of what happened up to $dt$;
- why all these? intervals between arrivals follow exponential distribution!

Why the title ”Poisson?”

- interval rep.: $X_i$ are exponential with parameter $\lambda$;

$$
F_X(t) = 1 - e^{-\lambda t}, \quad f_X(t) = \lambda e^{-\lambda t}.
$$

- number rep.: number of arrivals in a fixed interval is Poisson with parameter $\lambda$

$$
P\{N = k\} = \frac{\lambda^k}{k!} e^{-\lambda} \quad (13)
$$

Lecture: Basics of traffic modeling I
6. Service times

Service times:

• can be considered as a process:
  – 1st service time, 2nd,....

• are given by interval representation:
  – time to serve a customer.

• deterministic or probabilistic:
  – constant service time;
  – service time given by RV.

• service times are usually iid:
  – service time of the customer does not depend on the service time of the previous customer.

We assume: service times are realization of a certain RV.
6.1. Some examples

Consider possible servers associated with queuing systems:

- rate of the outgoing link from the switch service packets;
- exchange’s processor service incoming calls;
- man at the cash desk.

Example 1: supermarket, service time of a customer:

- two customers bought 5 and 10 items respectively;
- adding price to receipt requires approximately the same time for any item;
- service times:
  - the service time of an item is constant $\Delta$;
  - the service time of a customers is variable.

Question: how to model this situation?
What we have:

- arrival process: point process;
- service process: $S_i = n\Delta$, $n$ is RV describing how many items customers buy.

Figure 6: Service time of customers.
Example 2: outgoing link of IP router:

- service time of every bit is constant;
- service time of packets are variable due to variable size of IP packets.

Example 3: telephone exchange:

- time that call is on is, of course, variable;
- service time of each call is RV.

What is important:

- you should be creative defining service and arrival process;
- you should first decide what are the parameters of interest;
- based on these parameters you should define service and arrival processes!
7. Facts about the Internet traffic

The most important fact: Internet traffic constantly changes in time (every \( \tilde{5} \) years).

Observations, trends and facts on the Internet traffic up to 2005:

- TCP accounts for most of the packet traffic in the Internet;
- traffic flows are bidirectional, but often asymmetric;
- most TCP sessions are short-lived;
- the packet arrival process in the Internet is not Poisson;
- the session arrival process may be approximated by Poisson distribution;
- packet sizes are bimodally distributed;
- unknown stochastic properties of packet arrivals;
- Internet traffic continues to changes.
8. Step-by-step traffic modeling procedure

**Step-by-step procedure:**

- determine the point of interest;
- determine the level (layer) of interest;
- measure traffic at the point of interest;
- decide what statistics should be captured;
- estimate statistics of traffic observations;
- choose a candidate model;
- fit parameters of the model;
- test accuracy of the model.
9. Exponential-based time interval distributions

Why exponential-based distributions:

- analytically tractable:
  - reason: retains (to the some extent) memoryless properties;
  - can be used in analytical studies.

- include a wide class of distributions:
  - example: phase-type distribution allows to model an arbitrary continuous CDF.

- easy to generate interarrival times in simulation studies:
  - you just have to be able to generate exponential RVs.

Why not:

- all are renewal models: does not allow to take into account autocorrelation;

- practically, limited to simple cases of exponential combinations:
  - analytical analysis of PH/PH/-/-/- is complicated.
9.1. Exponential interval distribution

The CDF and pdf of $X_i$ are as follows:

$$F_X(t) = 1 - e^{-\lambda t}, \quad f_X(t) = \lambda e^{-\lambda t}, \quad t > 0, \lambda > 0. \quad (14)$$

Some important notes:

- we may use it to model arrival and service times;
- sometimes called a negative exponential distribution;
- completely characterized by a single parameter $\lambda$;
- the most important property is memoryless behavior: $F_X(t + x|\lambda) = F_X(t)$!

The mean, variance and squared coefficient of variation:

$$E[X] = \frac{1}{\lambda}, \quad D[X] = \frac{1}{\lambda^2}, \quad C^2[X] = \left(\frac{\sigma[X]}{E[X]}\right)^2 = \frac{D[X]}{(E[X])^2} = 1. \quad (15)$$

- $C^2[X]$ is a convenient metric when comparing randomness of RVs!
9.2. Why combinations of exponential distributions?

**Example:** repair with two consecutive repair points with the same rate $\lambda$:

![Phase Diagram of Two Consecutive Repair Points](image)

Figure 7: Phase diagram of two consecutive repair points with the same rate $\lambda$.

**Mean, variance and SCV (note the independence of phases!):**

$$
E[X] = \mu = \frac{1}{\lambda} + \frac{1}{\lambda} = \frac{2}{\lambda}, \quad \sigma^2[X] = \sigma^2 = \frac{1}{\lambda^2} + \frac{1}{\lambda^2} = \frac{2}{\lambda^2}, \quad C^2[X] = C^2 = \frac{\sigma^2}{\mu^2} = \frac{1}{2} \neq 1. \quad (16)
$$

**Note the following:**

- we cannot model it using exponential distribution: $\mu \neq \sigma \ (C^2 \neq 1)$ as it should be;
- we can model it as is: two phases with exponential distribution.
9.3. Erlang and generalized Erlang distribution

Generalized Erlang RV of order $k$, $k = 2, 3, \ldots$: sum of $k$ exponentials.

Note the following:

- intensity of distributions may be different;
  - Erlang RV: $\lambda_1 = \lambda_2 = \cdots = \lambda$;
  - generalized Erlang RV: at least one rate is different.

Erlang RV of order $k$, $k = 2, 3, \ldots$: sum of iid exponentials with the same intensity.

![Figure 8: Phase diagram of Erlang distribution.](image)
CDF of Erlang RV $X$ of order $k$ if found using convolution:

$$F_X(x) = 1 - \sum_{j=0}^{k-1} \frac{(\lambda x)^j}{j!} e^{-\lambda x}, \quad \lambda > 0, \quad t \geq 0,$$

(17)

- $k = 1, 2, \ldots$ is the number of exponentials.

Taking derivative, we can find a pdf of Erland distribution:

$$f_X(x) = \frac{(\lambda x)^{k-1}}{(k-1)!} e^{-\lambda x}.$$  

(18)

The mean, variance and squared coefficient of variation are:

$$E[X] = \frac{k}{\lambda}, \quad D[X] = \frac{k}{\lambda^2}, \quad C^2 = \frac{1}{k} \leq 1.$$

(19)

**Important note!** $C^2 \leq 1$, for exponential it was exactly 1!
Figure 9: Example of pdfs of Erlang distribution with mean 1 and different $C^2$. 

Lecture: Basics of traffic modeling I
Generalized Erlang RV (phases have different rates):

- pdf is convolution of individual pdfs and LT is given by:
  \[ F(s) = \frac{\lambda_1}{s + \lambda_1} \times \cdots \times \frac{\lambda_k}{s + \lambda_k}. \]  
  \[ (20) \]

- mean and variance given by:
  \[ E[X] = \sum_{i=1}^{k} \frac{1}{\lambda_i}, \quad \sigma^2[X] = \sum_{i=1}^{k} \frac{1}{\lambda_i^2}. \]  
  \[ (21) \]

- note when \( \lambda_1 = \lambda_2 = \cdots = \lambda_k \) they reduce to:
  \[ E[X] = \frac{k}{\lambda}, \quad \sigma^2[X] = \frac{k}{\lambda^2}. \]  
  \[ (22) \]

- coefficient of variation:
  \[ C^2 = \frac{\left( \sum_i \frac{1}{\lambda_i^2} \right)}{\left( \sum_i \frac{1}{\lambda_i} \right)^2} \leq 1. \]  
  \[ (23) \]
  - varies between \( k^{-(1/2)} \) and 1 but always less than 1!
9.4. Hyperexponential distribution

Hyperexponential RV of order $k$, $k = 2, 3, \ldots$: weighted sum of $k$ exponentials.

Note the following:

- no sense to set $\lambda_1 = \lambda_2 = \cdots = \lambda$ as we are going to get exponential with rate $\lambda$!
Factors $p_i, i = 1, 2, \ldots, k$ such that they sum up to one:

$$\sum_{i=1}^{k} p_i = 1. \quad (24)$$

CDF and pdf of hyperexponential distribution are:

$$F_X(x) = 1 - \sum_{i=1}^{k} p_i e^{-\lambda_i x}, \quad f_X(x) = \sum_{i=1}^{k} p_i \lambda_i e^{-\lambda_i x} \quad \lambda_i > 0, \quad x \geq 0. \quad (25)$$

Moments are given by:

$$\mu_j = \int_0^{\infty} x^j f_X(x) dx = j! \sum_{i=1}^{k} \frac{p_i}{\lambda_i^j}. \quad (26)$$

First and second moments are:

$$E[X] = \sum_{i=1}^{k} \frac{p_i}{\lambda_i}, \quad E[X^2] = \sum_{i=1}^{k} \frac{2}{\lambda_i^2} p_i, \quad C^2 \geq 1. \quad (27)$$

**Important note:** $C^2 \geq 1$ (Erlang: $C^2 \leq 1$, exponential: $C^2 = 1$).
Figure 11: Example of pdf of hyperexponential distribution with mean 1 and different $C^2$. 

Lecture: Basics of traffic modeling I
9.5. Cox distribution

More general compared to Erlang and hyperexponential.

Figure 12: Two representation of a random variable having Cox distribution.
9.6. Phase type distribution

Arbitrary combination of exponentials.

Phase distributed RV:
- time to absorb in the Markov chain;
- state 0, 1, ..., r are transient, state 0 is absorbing.

Note: renewal process with phase distributed times is denoted by PH-RP.
10. Simple modeling: two-moments fitting

We capture first two moments:

- mean;
- variance via SCV, $C^2 = \mu^2/\sigma^2$.

Choosing a model:

- $C^2 \approx 1$: exponential distribution;
- $C^2 < 1$: erlang distribution;
- $C^2 > 1$: hyperexponential distribution;

Important notes:

- can be used for both arrival and service processes;
- may sometimes result in very ‘coarse’ modeling;
- correlation is not taken into account at all.
10.1. Two moments fitting: $C^2 < 1$

You may use mixture of Erlang distributions:

- mix of Erlang distribution $E_{k-1,k}$:
  - $k - 1$: exponentials with the same mean with probability $p$;
  - $k$: exponentials with the same mean with probability $(1 - p)$;
  - pdf is given by:

$$f(x) = p\lambda \frac{(\lambda x)^{k-2}}{(k-2)!}e^{-\lambda x} + (1 - p)\lambda \frac{(\lambda x)^{k-1}}{(k-1)!}e^{-\lambda x}, \quad x \geq 0.$$ \hspace{1cm} (28)

- when $p$ goes from 0 to 1, $C^2$ goes from $1/(k - 1)$ to $1/k$!

The procedure:

- determine $k$ from: $1/k \leq C^2 \leq 1/(k - 1)$;
- get $p$ and $\lambda$ from the following:

$$p = \frac{1}{1 + C^2(kC^2 - [k(1 + C^2) - k^2C^2]^{1/2})}, \quad \lambda = \frac{k - p}{E[X]}.$$ \hspace{1cm} (29)

Lecture: Basics of traffic modeling I
10.2. Two moments fitting: \(1/\sqrt{2} < C^2 < 1\)

You may use generalized Erlang distribution:

- special case: sum of two exponentials with different rates \(\lambda_1\) and \(\lambda_2\);
- this may produce:

\[
1/\sqrt{2} < C^2 < 1, \tag{30}
\]

– recall, \(C\) varies between \(m^{-1/2}\) and 1 for generalized Erlang distribution.

If we are given mean \(\bar{E}[X]\) and SCV \(\overline{C^2}\):

- find \(\lambda_1\) and \(\lambda_2\) as follows:

\[
\lambda_i^{-1} = \frac{\bar{E}[X]}{2}(1 \pm \sqrt{2\overline{C^2} - 1}). \tag{31}
\]
10.3. Two moments fitting: $0 < \bar{C}^2 < 1$

You *may* use shifted exponential distribution:

- pdf is given by:

$$f_X(x) = \lambda e^{-\lambda(x-d)}, \quad x \geq d. \quad (32)$$

- $\lambda$ is the rate of exponential, $d$ is constant.

- mean and variance are given by:

$$E[X] = \frac{1}{\lambda} + d, \quad \sigma^2[X] = (E[X])^2C^2 = \frac{1}{\lambda^2}. \quad (33)$$

Parameters are given by:

$$\lambda = \frac{1}{\sigma[X]} m \quad d = \mu - \frac{1}{\lambda}. \quad (34)$$

Note the following:

- this distribution is shifted by $d$ on $X$-axis;

- such form may not be proper at all for many applications!
10.4. Two moments fitting: $\bar{C}^2 > 1$

You may use hyperexponential distribution of order 2: $H_2(p_1, p_2, \lambda_1, \lambda_2)$:

- we use hyperexponential distribution with balanced means:
  \[ \frac{p_1}{\lambda_1} = \frac{p_2}{\lambda_2}. \]  
  \[ (35) \]

- probabilities $p_1$ and $p_2$ can be found as:
  \[ p_1 = \frac{1}{2} \left( 1 + \sqrt{\frac{C^2 - 1}{C^2 + 1}} \right), \quad p_2 = 1 - p_1. \]  
  \[ (36) \]

- rates $\lambda_1$ and $\lambda_2$ are given by
  \[ \lambda_1 = \frac{2p_1}{E[X]}, \quad \lambda_2 = \frac{2p_2}{E[X]}. \]  
  \[ (37) \]

Note: when $\bar{C}^2 \approx 1$, use exponential distribution.

Lecture: Basics of traffic modeling I
10.5. Two moments fitting: $\overline{C^2} > 0.5$

When $C^2 > 0.5$ you may use Cox-2 distribution:

- parameters are given by:

$$\lambda_1 = 2\overline{E[X]}, \quad p_1 = \frac{1}{2\overline{C^2}}, \quad \lambda_2 = \lambda_1p_1.$$  \hspace{1cm} (38)

![Figure 14: Illustration of the Cox-2 distribution.](image-url)