Traffic Engineering

Part 2: Queuing Theory in Telecommunications

By: Farhad Faghani (Ph.D.)

Queueing Theory

- is used to model a wide range of problems for teletraffic analysis.
- is used every time a network resource is shared by competing بـ requestsپ
- When service requests arrive temporarily according to a higher rate than the time needed to fulfill each of them, a waiting list is needed at queue, provided that it has enough rooms to store all requests.

Typical problems

- OSI Layer 1:
  - **Blocking phenomena** of a traffic flow (i.e., a call) due to unavailable resources in at least one link in the path from source to destination.
- OSI Layer 2:
  - Queuing is generated by different packets **sharing the transmission resources** of a link connecting two adjacent nodes
- OSI Layer 3:
  - Queuing is experienced **by routing requests** at the layer 3 signaling processor.

Typical examples

- Performance analysis for the transmission on links and corresponding buffer dimensioning;
- Network planning
  - planning of the capacity needed to interconnect the different nodes of a telecommunication network
- Performance evaluation of access protocols where different بـ users پcontend for the same resources.

Queue characterization

- **arrival process** of service requests,
- **a waiting list** of the requests to be processed,
- **a discipline** according to which the requests in the queue are selected to be served
- **a service process**

Queues

- Queues are special cases of **stochastic processes** that are represented by a **state X(t)**, denoting the **number of service requests** or **“entities”** or **“customers”** queued at time t.
- $f_{X(t)}(\tau)$
  - denotes the pdf of process X at time □

stochastic process

- can be characterized as follows:
  - **state space:**
    - that is the set of all the possible values, which can be taken by X(t).
    - Such space can be continuous or discrete
      - if the state space is discrete, the stochastic process is called **chain**
  - **Time variable:**
    - variable t can belong to a continuous set or a discrete one.
  - **Correlation characteristics**
    - among X(t) random variables at different instants t.

Definition

- to account for the process correlation,
  - we describe X(t) in terms of its joint probability distribution function, sampling the process at different instants

$$\mathbf{t} = \{t_1, t_2, \ldots, t_n\} \text{ for any } n:$$

$$\mathrm{PDF}_X(\mathbf{x}, \mathbf{t}) = \mathrm{Prob}\{X(t_1) \leq x_1, X(t_2) \leq x_2, \ldots, X(t_n) \leq x_n\}$$

$$\text{vector } \mathbf{x} = \{x_1, x_2, \ldots, x_n\}$$

- Expected value:

$$E[X(t)] = \int_{-\infty}^{+\infty} \tau f_{X(t)}(\tau)\mathrm{d}\tau$$

- Autocorrelation:

$$R(t_1,\ t_2) = E[X(t_2)X(t_1)]$$

strict-sense stationary

- X(t) is strict-sense stationary (SSS)
  - if the following equality holds for any n and t
  - i.e., distribution $PDF_x(x,t)$ is invariant to time shifts

$$\mathrm{PDF}_X(\mathbf{x}, \mathbf{t} + \tau) = \mathrm{PDF}_X(\mathbf{x}, \mathbf{t})$$

wide-sense stationary

- X(t) is wide-sense stationary (WSS)
  - if its **expected value**

$$E[X(t)] = \mu$$

  - and its **autocorrelation** are independent of t:

$$R(t, t + \tau) = E[X(t)X(t + \tau)] = R(\tau)$$

- SSS implies the wide-sense stationarity

independence

- A process is independent if we have for any n and t:

$$\mathrm{PDF}_X(\mathbf{x}, \mathbf{t}) = \mathrm{Prob}\{X(t_1) \leq x_1\}\mathrm{Prob}\{X(t_2) \leq x_2\} \cdots \mathrm{Prob}\{X(t_n) \leq x_n\}$$

Markov chain

- is characterized by the fact that its state value at instant $t_{n+1}, X(t_{n+1})$, depends only on its state value at the previous instant $t_n, X(t_n)$.

$$\text{Prob}\left\{X(t_{n+1}) = x_{n+1} \,\middle|\, X(t_n) = x_n, X(t_{n-1}) = x_{n-1}, \ldots, X(t_1) = x_1\right\}$$

$$= \text{Prob}\left\{X(t_{n+1}) = x_{n+1} \,\middle|\, X(t_n) = x_n\right\}$$

- This **memoryless** characteristic implies:
  - state sojourn times are exponentially distributed for a continuous-time chain or geometrically distributed for a discrete-time chain

subclasses of Markov chains

- Renewal processes
- Birth-death Markov chains
- Semi-Markov chains

Renewal Processes:

- These are ﺑ pointﭘprocesses like the arrival of points on the time axis. Intervals.
- A special case of renewal process is the Poisson arrival process, where interarrival times are exponentially distributed with a constant rate;

Birth-death Markov chains

- The transitions from the generic state X = i are only towards state X = i-1 or towards state X = i + 1.
- These chains will be used to model Markovian queues (M/M/. . .)

Semi-Markov chains

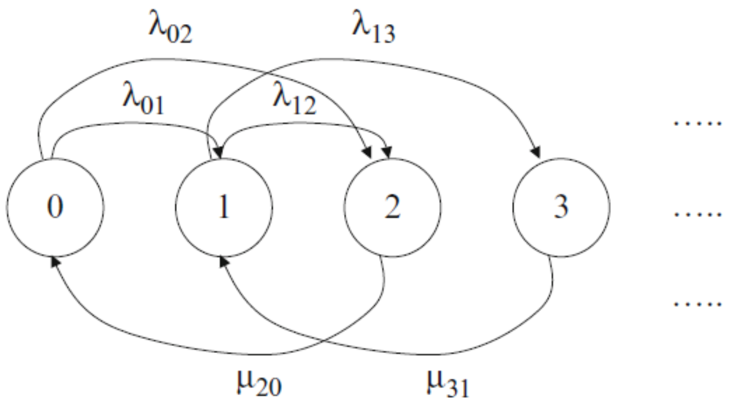
- we obtain an imbedded Markov chain, which can be considered as a discrete-time Markov chain.
- Also in this case we have a state probability distribution.
- will be used to model M/G/1 queues

Markov chain

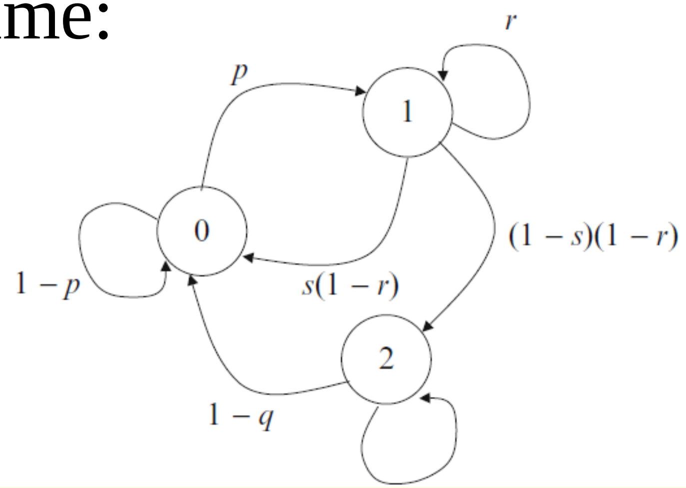
- continuous-time
  - transitions may occur at any time and are characterized by exponentially distributed intervals
    - with mean rates shown above the arcs of the transitions
- discrete-time
  - transitions occur at given instants
  - probabilities are used to characterize the transitions that correspond to geometrically distributed intervals.
  - states may have transitions into themselves
  - the sum of all the transitional probabilities leaving a state must be equal to 1

Markov chain

- continuous-time:



- discrete-time:

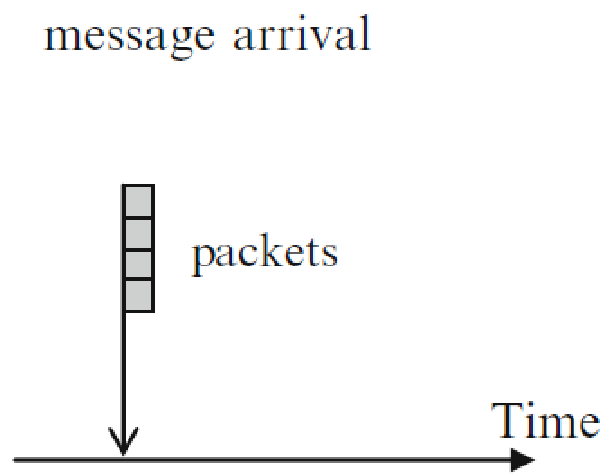


Compound Arrival Processes

- consider the case where each arrival carries multiple **“service requests”** or **“objects”**:
  - for instance, the arrival of a message that carries multiple packets simultaneously
    - this could be case of an IP packet fragmented into many layer2 packets arriving at a MAC layer queue.
  - Different names:
    - bulk arrival process,
    - batched arrival process, and
    - compound arrival process.
- difference in the compound arrival processes between continuous-time cases and discrete-time ones

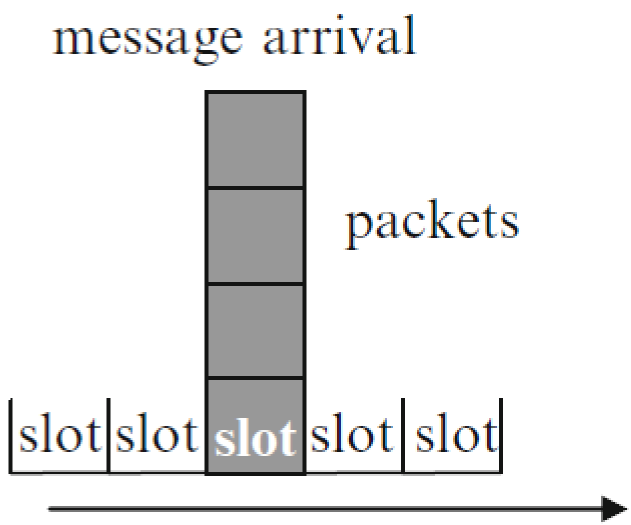
continuous-time cases

- all the ب objectsپof a group arrive simultaneously at a queuing system.
- consider that all of these ب objectsپare generated by the operating system at a speed extremely faster than the service rate of the queue.



discrete-time cases

- (compound) arrivals are synchronized with time slots:
  - a message arrival needs a slot to become available to the queue.
  - This is consistent with the **store-and-forward** model and refers to the case where messages arrive at a network element propagating along a communication link:
    - a message must first be stored in the queue of this node element and then it is ready for transmission.


message arrival
packets
slot
slot
slot
slot
slot