IV. Intro to MAC Sub-Layer and Queueing Theory	
Instructor: Patrick Tague	**Date**: 16 January 2008
EE 565: Computer-Communication Networks I	Winter Quarter 2008

1 Introduction to the MAC Sub-Layer

Recall from the first lecture that networks can be partitioned into two classes depending on whether nodes communicate using unicast (or point-to-point) transmission or broadcast (also, multicast) transmission. In broadcast communication, there must be a mechanism to handle the key issues:

1. Who gets to use the channel?
2. For how long?
3. When?
4. Why?

This function of controlling access to the physical medium is delegated to the **Medium Access Control (MAC) Sub-Layer**, which technically falls somewhere between the physical and link layers.

The central issues informally listed above can be addressed by the **Channel Allocation Problem**, namely *how to allocate a single channel among competing users*? This is a long-standing problem, so let's first look at some of the traditional approaches which partition resources equally among users.

Traditional approaches to the channel allocation problem involve partitioning users physically or logically in such a way as to dedicate a unique resource to each of the N users. Common approaches include Frequency, Time, and Code Division Multiplexing (FDM, TDM, CDM, respectively) which utilize orthogonal (or independent) sets of resources for different users. Protocols based on these approaches typically involve the allocation of a fixed resource to each user. However, if the number of users is variable, there may be unallocated resources or users that do not receive service. Further, since traffic tends to be bursty (i.e. users do not have a constant supply of data frames to send), allocated resources often go unused. FDM, TDM, and CDM can be further generalized into the framework of **Orthogonal Frequency Division Multiplexing (OFDM)** (check out EE 567, taught by Prof. Hui Liu).

As the entire class of OFDM-based protocols rely on physical or logical partitioning of users based on deterministic properties, they are not entirely applicable to network environments, especially in the bursty-traffic regime and in networks with a high degree of variability (e.g. mobile ad-hoc networks) or uncertainty (e.g. wireless sensor networks) in the structure of the network. For the duration of this topic, we will focus our attention on unstructured broadcast or *random access* channels instead of OFDM-based approaches.

In order to completely understand the behavior of MAC protocols for a random access channel, we first have to investigate the underlying models for user traffic generation, random access of the shared medium, and behavior of the service being provided. The models we'll adopt here are addressed in detail in Tanenbaum's *Computer Networks* textbook, compressed into the following assumptions.

1. **Station Model**: The N users or stations are independent and traffic arrives with constant average rate λ, i.e. $\Pr[\text{frame generated in } \Delta t \text{ seconds}] = \lambda \Delta t$.
2. **Single Channel**: All N users transmit and receive each message on a single broadcast channel.
3. **Collisions**: If multiple users transmit frames in overlapping time intervals, all overlapping frames are lost in the collision, and all nodes involved can detect the collision. We assume collisions are the only source of frame loss and error.
4. **Time**: Access to the shared medium is either (a) asynchronous/continuous or (b) synchronous/discrete/slotted.
5. **Carrier Sensing**: Users may (a) be able to detect the presence of carrier signals on the medium or (b) not able to carrier sense.

Given these models, we need a mathematical basis to compute such quantities as:

- the probability that a transmission or medium access is successful, as a function of the traffic generation rate, transmission rate, service, rate, number of users, etc.),
- the expected queueing delay incurred by message frames, and
- the expected throughput, collision rate, retransmission rate, etc.,

motivating the need for a study of **Queueing Theory**.

2 Introduction to Queueing Theory

Simply stated, queueing theory is the *mathematics of waiting in line*. This topic is far more interesting than it first seems, and we'll go into a sufficient level of detail for the purposes of this course. We'll begin our treatment of queueing theory by introducing a large collection of concepts from probability theory, including many random variables and random processes of interest. Following common notation, we'll denote discrete-time random processes in terms of the random variables X_n in the random process $\mathbf{X} = \{X_1, X_2, \ldots\}$ and continuous-time processes in terms of the random variables $X(t)$ as in the random process $\mathbf{X} = \{X(t_1), X(t_2), \ldots\}$.

2.1 Markov Random Processes

A **Markov process** is defined as a random process $\mathbf{X}$ with the property that the probability distribution of the process reaching the *state* x_{n+1} is dependent only on the current state x_n of the process and not upon any of the states $x_0, \ldots, x_{n-1}$ or the times that state transitions occurred in the past. A Markov process for which the states x_i are elements of a discrete (or countable) state space is referred to as a *Markov chain.* Markov chains can be classified based on whether state transitions occur at integer time steps $0, 1, \ldots, n, \ldots$ (discrete-time Markov chain) or at continuous times $t_0 < t_1 < \cdots < t_n < \cdots$ (continuous-time Markov chain). In the case of a continuous-time Markov chain, state transitions may take place at any time instant, leading us to investigate the random variable describing the amount of time spent in each state. Because the Markov property requires independence of transitions from past transition times, this transition time must be a memoryless random variable. Hence, the transition time must be exponentially or geometrically distributed, depending respectively on whether the variable is continuous or discrete. In general, the Markov property can be stated mathematically as

$$\begin{aligned}
&\Pr[X(t_{n+1}) = x_{n+1} \mid X(t_n) = x_n, X(t_{n-1}) = x_{n-1}, \ldots, X(t_0) = x_0] \\
=&\Pr[X(t_{n+1}) = x_{n+1} \mid X(t_n) = x_n].
\end{aligned}$$

2.2 Birth-and-Death Chains

A special case of a Markov chain referred to as a **birth-and-death chain** (or birth-death process) is given by the restriction of state transitions to itself and “neighboring” states only. Without loss of generality, we can take the state space to be the set of (non-negative) integers and define the neighboring states of a state i to be $i-1$ and $i+1$ (unless $i = 0$, in which case only state 1 is a neighbor). Specifically, if $X_n = i$, then the probability that X_{n+1} is outside the set $\{i-1, i, i+1\}$ is zero.

3 Discrete-Time Markov Chains

We’ll next investigate discrete-time Markov chains in further detail. Following the definition of a general Markov process above, we say the discrete-time sequence $X_1, X_2, \ldots$ forms a discrete-time Markov chain (DTMC) if for all non-negative integers n and all possible values

i_n of X_n in the state space, we have

$$\Pr[X_n = j \mid X_{n-1} = i_{n-1}, \ldots, X_0 = i_0] = \Pr[X_n = j \mid X_{n-1} = i_{n-1}].$$

We assume there is a probability distribution $\Pr[X_0 = j]$ to characterize the initial state of the DTMC. If it turns out that the state transition probabilities above are independent of the time index n, the DTMC is **homogeneous**, and we denote the state transition probabilities by

$$p_{ij} = \Pr[X_n = j \mid X_{n-1} = i]$$

and let P denote the matrix of transition probabilities. For homogeneous DTMCs, we further define the m-step transition probabilities as

$$p_{ij}^{(m)} = \Pr[X_{n+m} = j \mid X_n = i].$$

Using the Markov property, we can show that the m-step transition matrix $P^{(m)}$ is equal to the m^{th} power P^m of the transition matrix P. Further, we say that a DTMC is **irreducible** if for every i, j there is some value of m such that $p_{ij}^{(m)} > 0$, meaning that every state j is reachable from every other state i with non-zero probability. For an irreducible DTMC, let γ_j denote the greatest common divisor (GCD) of all values m such that $p_{jj}^{(m)} > 0$, referred to as the *periodicity* of state j. A state j is periodic if $\gamma_j > 1$, and the chain is periodic (resp. aperiodic) if any (resp. none) of its states is periodic. Finally, we define the probability $\pi_j^{(n)}$ to be the probability that we find the chain in state j at time n and state the first result of interest.

In an irreducible and aperiodic homogeneous DTMC, the limiting probabilities

$$\pi_j = \lim_{n\to\infty} \pi_j^{(n)}$$

always exist and are independent of the initial distribution $\pi_j^{(0)}$. Furthermore, the row-vector π with j^{th} entry equal to π_j satisfies $\pi P = \pi$. In certain cases (strongly recurrent DTMCs, when the expected time to return to each state j is finite), the vector π will satisfy the equality $\pi\mathbf{1} = 1$, where $\mathbf{1}$ is the all-one column vector of appropriate length, implying that π is a probability distribution. In this case, π is referred to as the **stationary distribution** or the **steady-state distribution**.

4 Continuous-Time Markov Chains

We now turn our attention to continuous-time Markov chains. The initial definition of the Markov property characterizes a continuous-time Markov chain (CTMC) and can be alternately defined as satisfying the condition

$$\Pr[X(t) = j \mid X(\tau) \text{ for } \tau_1 \le \tau \le \tau_2 < t] = \Pr[X(t) = j \mid X(\tau_2)].$$

The treatment of CTMCs follows quite naturally from that of DTMCs with the exception of the state transition times, which, as stated earlier, must satisfy a memoryless probability distribution. Analogous to that of a DTMC, we define the state transition probability $p_{ij}(s,t)$ as

$$p_{ij}(s,t) = \Pr[X(t) = j \mid X(s) = i].$$

Letting $H(s,t)$ denote the transition matrix, the m-step transitions can be characterized using the fact that $H(s,t) = H(s,u)H(u,t)$ for $s \leq u \leq t$ (and letting $H(t,t) = I$, the identity matrix). In this case, we see that the transition matrix H carries information for any time interval $s - t$, so we investigate the limit over an infinitesimal time step Δt. We thus define the **transition rate matrix** $Q(t)$ as

$$Q(t) = \lim_{\Delta t \to 0} \frac{H(t, t+\Delta t) - I}{\Delta t}.$$

For $i \neq j$, the element $q_{ij}(t)$ of $Q(t)$ characterizes the rate of transitions from state i to state j, while for the diagonal entries, $-q_{ii}(t)$ characterizes the rate of transitions away from state i. We see immediately that the sum of each row of $Q(t)$ is zero.

Similar to the case of DTMCs, a homogeneous CTMC exhibits a degree of time-independence, in that the transition probability $p_{ij}(s, s+t)$ is independent of s (hereafter denoted $p_{ij}(t)$), $q_{ij}(t)$ is independent of t (hereafter denoted q_{ij}), $H(s, s+t)$ is independent of s (hereafter denoted $H(t)$), and $Q(t)$ is independent of t (hereafter denoted Q). From the above definitions, we can show that $H(t) = e^{Q(t)}$. For the state probabilities $\pi_j(t)$ (and the corresponding vector $\pi(t)$), we can show that $\frac{d\pi(t)}{dt} = \pi(t)Q$. Furthermore, in the limit $t \to \infty$, $\pi_j(t)$ approaches the steady state probability π_j (and $\pi(t) \to \pi$) and we can show that $\pi Q = 0$.