Random Access and Wireless LANs

- OFDMA systems can be viewed as a centralized resource allocation mechanism for packet multiplexing in wireless networks.
- Random access and wireless LANs, on the other hand, can be viewed as a distributed resource allocation mechanism for packet multiplexing in wireless networks.
- We consider networks in which all the nodes use the same part of the spectrum.
- Our interest is in the use of random access based medium access control (MAC) protocols for distributed access control.
Around a receiver, we define an interference region (also called the carrier sense region) and a subset of it called the decode region.

Consider a transmitter and its receiver in the network. If the transmitter is in the interference region of the receiver, the received power at the receiver is significant and causes a collision at the receiver if another transmission is being received at the same time. If the transmitter is also in the decode region, then the SNR at the receiver is greater than the prescribed threshold, say $\theta$, and the receiver can decode the transmission.

The received power from transmitters outside the interference region is assumed insignificant.
Figure 7.1
MAC protocols

- In single hop networks, every node is within the decode region of every other node; that is, all nodes are one hop away from each other. Such networks are also called broadcast networks because a transmission from a node can be decoded by every other node in the network. They are also called colocated networks.

- The channel is also called the medium and a medium access control (MAC) protocol regulates use of medium by prescribing the rules to initiate a transmission and continue with it.

- In random access networks, collisions may occur and the MAC protocol has to resolve collisions; it has to arbitrate among the nodes contending to use the medium.
The arbitration is a distributed algorithm that typically prescribes forced silences on the nodes.

Thus, some amount of transmission time is lost to collisions and arbitrations. The fraction of time so lost is a measure of the efficiency of the protocol.

Simple protocols, even if of low efficiency, are useful if per node throughput that the protocol obtains is significant compared to the throughput required by the nodes in the network.
Protocols without Carrier Sensing: Aloha and Slotted Aloha

- The Aloha, also called the pure Aloha, is the earliest random access protocol.
- The idea is simple: If a node has a packet to transmit, it just transmits.
- If there was another transmission at the same time, there would be a collision and neither packet can be decoded correctly by the corresponding receivers.
- If the packet is not correctly received, the packet will have to be transmitted using a suitable retransmission algorithm.
- Next figure shows a space-time diagram of a transmission and reception in a network with large propagation delays and the futility of deferring to a carrier.
Figure 7.2
Simplifying Assumptions

- We will assume fixed length packets and time will be measured in terms of packet transmission time; the packet transmission times are of unit duration. The nodes are located along a straight line of length $\eta$. The maximum propagation delay in the network is also $\eta$.

- The packet transmission attempts in the network are assumed to form a Poisson process of rate $G$ attempts per second. The location of the transmitting node is chosen uniformly in $[0, \eta]$ and independently of the other transmissions. Thus, each packet transmission attempt is characterized by an order pair $(t, y)$, where $t \geq 0$ is the time at which the transmission started, and $0 \leq y \leq \eta$ is the location of the transmitting node.

- A sample realization of this space-time attempt process in the region $([0, \infty]) \times [0, \eta]$ is shown in the next figure.
Figure 7.3
Simplifying Assumptions (ctd.)

- Note that the number of attempts in A and B are independent. Further, the number of attempts in A has a Poisson distribution with mean \( (G/\eta) \times (\text{Area of A}) \). Thus, the space-time attempt process in the region \([0, \infty] \times [0, \eta]\) is a two-dimensional Poisson point process of rate \( G/\eta \) attempts per meter-second.

- This model means that there is an infinite number of nodes in the network and at any instant, each node has at most one packet to transmit. This is a good model for a network with a large number of nodes and a low packet arrival rate per node.

- Consider a node at a location T transmitting a packet to a node at location R. For this transmission, we can define a collision window in time at each location in \([0, \eta]\). If a transmission is begun at that location in the collision window, then it will arrive at R when it is receiving the packet T, thus causing a collision.

- The union of the collision window at all points in the network gives us a collision cone. Any transmission attempt in this collision cone will cause a collision at the intended receiver, R in the example, as in the next figure.
Figure 7.4
Transmissions initiated in the interval \((b, c)\) at \(\eta\) and in the interval \((f, e)\) at \(0\) will cause a collision at \(a\). The space-time area covered by the collision cone is clearly \(2\eta\).

Since \(G/\eta\) is the space-time arrival rate, multiplying it by \(2\eta\) gives the mean of the Poisson distribution of the number of arrivals in the collision cone. Thus, the probability that a reception is successful, \(P_s\), is given by

\[
P_s = \Pr(\text{No transmission attempt in collision cone}) = e^{-\left(\frac{G \times 2\eta}{\eta}\right)} = e^{-2G}
\]

Defining the throughput, \(S\), as the mean number of successful attempts in unit time, we get \(S = G \cdot P_s = S \exp(-2G)\). The maximum value of \(S\) is achieved for \(G = 0.5\) and \(S_{\text{max}} = 1/2e \approx 0.18\).
Slotted Aloha

There is a simple way to make pure Aloha more efficient. Instead of allowing a node to begin transmission at any time, let time be slotted and the nodes be allowed to begin transmission only at the beginning of a slot. The slot length is made equal to the sum of the packet transmission time (unity) and the maximum propagation delay in the network. This is the slotted Aloha (S-Aloha) MAC protocol and it has found wide application in the practice.

Assume as before that transmission attempts arrive according to a Poisson process of rate $G$ and the source node is uniformly distributed in $[0, \eta]$. The Poisson rate of packet arrivals that can cause a collision is thus the expected number of Poisson arrivals in a slot, $G(1+\eta)$.

The probability that a transmission is successfully received is $P_s = \exp(-G(1+\eta))$. Then the throughput is $S = G P_s = G \exp(-G(1+\eta))$. The maximum achievable throughput, $S_{\text{max}}$, is $S_{\text{max}} = 1/(e(1+\eta))$. 
Figure 7.5
Example Applications of Slotted Aloha

- In GSM cellular networks, a control channel on the reverse link from the mobile node to the base station, called the Random Access Channel (RACH), is used by the GSM mobile stations to send messages to the network. The types of messages include those to initiate new calls, register locations of the mobile stations, and reply to paging queries. The messages are small and generated at very low rate compared to the capacity of the RACH channel. The number of mobile nodes in a cell is not fixed and also quite large and signaling bandwidth cannot be allocated statically to these nodes. Hence, slotted Aloha is used on this channel. After transmitting on the RACH using slotted Aloha protocol, the mobile station waits for a fixed duration to know if the transmission was successful. If an acknowledgement is not received before this duration, a retransmission is attempted.
Another application of slotted Aloha is in very small aperture terminal (VSAT) networks. A VSAT network is a satellite network in which there are several geographically widespread, small terminals. These terminals are attached to individual computers or to the local area networks of small organizations through the digital interface unit (DIU). The terminals share a satellite link to a large hub. The nodes can communicate only with the hub and all inter-node communications are over two hops via the hub. Hence, the inbound channel from the terminals to the hub needs to be shared. The terminals request for reservation on the slots in the inbound channel to the hub. When a remote node wants to transmit, it first request for a reservation on the slots in the inbound channel to the hub. This reservation request is made using slotted Aloha protocol on the uplink from the remote station to the hub. This reservation scheme can be very efficient if the bandwidth allocated for the reservation requests is small and the amount of reserved bandwidth is large.
Figure 7.6
In our previous discussion, we had implicitly assumed that packets involved in collision are lost. We now analyze the S-Aloha protocol under the more realistic assumption that a packet that has suffered a collision stays in the network and makes retransmission attempts until it is successful. The number of such packets is called the backlog.

Let $A_k$ be the number of new packet arrivals into the network and $D_k$ the number of successful transmission (departures from the network) in slot $k$. We assume that all fresh arrivals during a slot will attempt a transmission at beginning of the next slot. Let $B_k$ denote the backlog at the beginning of slot $k$. It is easy to see that $B_k$ evolves as

$$B_{k+1} = B_k + A_k - D_k$$
We assume that new packet form a Poisson process of rate $\lambda$ independent of everything else in the network. We also assume that the backlogged nodes attempt retransmission independently in each slot with probability $r$. Under these assumptions, $\{B_k\}$ is a discrete time Markov chain. Our interest is in the stability of this Markov chain.

For stability analysis of $\{B_k\}$, consider $d(n)$ defined by

$$d(n) := E(B_{k+1} - B_k \mid B_k = n) = E((A_k - D_k) \mid B_k = n)$$

$d(n)$ is the expected change in the backlog in one slot when the backlog is $n$ and is called the drift from state $n$. 
Instability of Aloha (ctd.)

- Then we can write

\[
\begin{align*}
Pr(A_k - D_k = +m | B_k = n) &= (\lambda^m/m!)e^{-\lambda}, \text{ for } m \geq 2 \\
Pr(A_k - D_k = +1 | B_k = n) &= \lambda e^{-\lambda} (1-(1-r)^n) \\
Pr(A_k - D_k = -1 | B_k = n) &= e^{-\lambda} nr(1-r)^{n-1}
\end{align*}
\]

- Using this and simplifying, we get

\[
d(n) = (-e^{-\lambda}nr(1-r)^{n-1}) + (\lambda e^{-\lambda} (1-(1-r^n))) + (\sum_{m=2}^{\infty} m \frac{\lambda^m}{m!} e^{-\lambda})
\]

\[
= \sum_{m=0}^{\infty} m \frac{\lambda^m}{m!} e^{-\lambda} - \lambda e^{-\lambda} (1-r)^n - e^{-\lambda} nr(1-r)^{n-1}
\]

\[
= \lambda - e^{-\lambda} (1-r)^n (\lambda + \frac{nr}{1-r})
\]

- This proves that \( \{B_k\} \) is not positive recurrent for any \( \lambda, r > 0 \).
An obvious issue now is to design mechanisms to make the network stable for some $\lambda > 0$ so that the network can support new packet arrival at that rate. This is done by making the retransmission probabilities adaptive. We assume that all nodes know the size of the backlog at the beginning of every slot and also the stationary packet arrival rate. Then, the probability of a successful transmission when the backlog is $n$, $P_s(n)$, is given by

$$P_s(n) = \lambda e^{-\lambda}(1-r)^n + e^{-\lambda} n r (1-r)^{n-1}$$

The retransmission probability that maximizes $P_s(n)$, say $r(n)$, is obtained as $r(n) = (1-\lambda)/(n-\lambda)$. The corresponding drift will be $d(n) = \lambda - e^{-\lambda} ((n-1)/(n-\lambda))^{n-1}$. We note that $d(n) \to \lambda - e^{-1}$ as $n \to \infty$.

We conclude that for $\lambda < 1/e$, the Markov chain $\{B_k\}$ will be positive recurrent and the S-Aloha network with this type of adaptive retransmission will be stable.
It is not practical for the nodes to know $B_k$, and a node should learn the network state from the events that it can observe. Let $Z_k$ be the event in slot $k$. $Z_k$ takes the following values:

$$Z_k = \begin{cases} 
0 & \text{idle slot; no transmission is attempted} \\
1 & \text{successful transmission; exactly one transmission is attempted} \\
e & \text{error (collision); more than one transmission is attempted}
\end{cases}$$

To learn the state, a learning variable, $S_k$ is used. $S_k$ is updated based on $Z_k$. A node transmits with probability $1/S_k$ in slot $k$ if it has either a fresh packet or a backlogged packet. Typically, two kinds of updates are used:

$$S_{k+1} = \max \{ 1, S_k + aI_{\{Z_k=0\}} + bI_{\{Z_k=1\}} + cI_{\{Z_k=e\}} \}$$

Here $a=-1$, $b=0$, and $c=1$ is a common choice.
Another possibility is as follows:

\[ S_{k+1} = \max\{1, a(Z_k) \cdot S_k\} \]

It has been shown that there exists \( a(Z_k) \) that achieve the maximum possible throughput of \( 1/e \).

To make the protocol more robust, a node will make the \( m \)-th transmission attempt after a backoff period of \( x_m \) unites of time. Here \( x_m \) is uniformly distributed random integer in the interval \([0, B_{m-1}]\). A typical update equation has the form

\[
B_m = \begin{cases} 
\min(a \times B_{m-1}, B_{\text{max}}) & \text{if } (m-1)\text{-th transmission collides} \\
\max(B_{m-1} - b, B_{\text{min}}) & \text{if } (m-1)\text{-th transmission is successful}
\end{cases}
\]

where \( a, b, B_{\text{min}}, \) and \( B_{\text{max}} \) are predefined.
Carrier Sensing Protocols

- In networks in which the propagation delay is small compared to the packet transmission time, it is possible to infer channel state (busy or idle) through carrier sensing and thereby obtain a random access protocol that is more efficient than the pure random access strategy of Aloha.

- In such networks, if a node senses the channel to be busy and yet transmits, it can cause a collision at the receiver of the ongoing transmission. Further, it is likely that the ongoing transmission is being heard at the receiver of the new transmission and a collision will occur there as well. Thus both transmissions are lost.

- Hence, a node should listen to the channel before beginning to transmit and defer to an ongoing transmission. This is the principle of the carrier sense multiple access (CSMA) protocol. In this protocol, once a node begins transmitting, it transmit the complete packet.
Collision Window for the CSMA

- The **collision window** for the CSMA is as follows. It is the time since the beginning of a transmission during which another node (not having heard the ongoing transmission) can begin its own transmission, and hence collide with the first transmission.

- The **maximum duration** of a collision will be \((t_{\text{trans}} + 2t_{\text{propgn}})\), where \(t_{\text{trans}}\) is the packet transmission time and \(t_{\text{propgn}}\) is the maximum propagation delay.

- A further **improvement** over CSMA is possible by the node **continuing** to monitor the channel after beginning transmission. If it **senses** a collision, then the node can immediately **stop** transmission and **minimize** the loss of channel capacity. This is called CSMA with **collision detect** or **CSMA/CD**. The **maximum collision duration** in the network is reduced to \(3t_{\text{propgn}}\), which is also the maximum collision duration seen at a node.
(a) A collision and a successful transmission in a CSMA network.

(b) A collision in a CSMA/CD network.

Figure 7.7
Consider the next figure. A and B are interference regions of transmitter a and b, respectively, and X is the intersection of A and B.

Consider an ongoing transmission from a to c. Since b is outside the interference region of a, it cannot sense the carrier from this transmission and can decide to transmit. If b transmits at the same time as a, there will be a collision at the receivers in the region X including c. However, a will not know of the collision at c and will continue to transmit. In this scenario, we say that b is hidden from a with reference to a transmission to c.

Now consider the transmitter d whose interference region is shown as D. d has to send a packet to e when a is transmitting to c. Node d is in the interference region of a, and can therefore sense the carrier from a. The two transmissions, d-e and a-c, can coexist because c is outside the interference region of d, and e is outside the interference region of a. Yet, on sensing the carrier from a, d will be forced to defer transmission. Here we say that node d is exposed to a transmission from a.
Collision due to a hidden node

Deferral by an exposed node

- $b$ senses idle channel and starts transmitting
- $d$ senses busy channel and defers transmission

Figure 7.8
Collision avoidance (CA) mechanisms **prevent** collisions due to transmissions by hidden node. These mechanisms assume that the interference regions, and also the decode regions, for transmission and reception are **identical**.

A simple CA mechanism is to have a narrowband auxiliary signaling channel in addition to the data channel. A node actively receiving data on the data channel transmits a busy tone on the signaling channel to enable the hidden nodes to defer to receiving nodes in their interference regions.

The effect of the busy tone is achieved by preceding the actual data transfer by a **handshake** between the transmitter and the receiver. This handshake is used to convey an imminent reception to the hidden node.

Before transmitting a data packet, a source node transmits a (short) **request to send (RTS)** packet to the destination. If the destination receives the RTS correctly, it means that it is not receiving any other packet and it acknowledges the RTS with a **clear to send (CTS)** packet. CTS informs the neighborhood of a receiver about an impending packet reception. The source then begins packet transmission. If CTS is **not** received within a specified timeout period, the source assumes that the RTS had a collision at the receiver and **retransmission** is attempted.
The RTS serves to inform nodes in the decode region of the transmitter about the imminent transmission of a packet, and CTS serves the same purpose for nodes in the decode region of the receiver.

If the transmission duration information is included in the RTS and CTS packets, then the nodes in the decode region of both the transmitter and the receiver can maintain a network allocation vector (NAV) that indicates the remaining time in the current transmission and schedule their transmissions to avoid collision.

Thus this handshake is a collision avoidance scheme and the protocol is the carrier sense, multiple access with collision avoidance (CSMA/CA).
The CSMA/CA protocol has been adopted for wireless LANs in the IEEE 802.11 series of standards.

Many improvements to this basic protocol have been suggested with the most popular being called MACAW. An additional feature in MACAW is the use of an acknowledgement from the receiver after the successful reception of a packet.

MACAW also specifies the transmission of a short Data Sending (DS) packet preceding the actual data transfer.

The DS packets provides information to the exposed nodes about the beginning and end of transmission times.

It has since been decided that DS is not necessary and the IEEE 802.11 standard does not include this message in its handshake protocol.
The basic ideas of the CSMA/CA protocol of MACA and MACAW have been formalized in IEEE 802.11 (Wi-Fi) wireless LAN standards. According to this standards, the network configuration could be either of the following two modes.

- **Independent or ad hoc network** mode. In this mode, the nodes form an independent multi-hop wireless network and they communicate directly with one another. A routing protocol and a corresponding routing algorithm will need to be used so that the packets find paths to their destination.

- **Infrastructure** mode. Here data communication is always between a mobile station (MS) and an access point (AP). The AP is connected to the wired network and provides a service similar to the base station of a cellular network. In this mode, the MSs need to associate with an AP using an association protocol. An AP and the MSs associated with it form a basic service set (BSS), and a set of BSSs is called an extended service set (ESS). The association, and dissociation, allows the MSs to be mobile within the ESS.
Figure 7.10
Figure 7.11
The First version of the IEEE 802.11 Standard for the Physical Layer

- The initial 802.11 standard had three PHY standards:
  - Infrared
  - 1 and 2 Mbps over frequency hopping spread spectrum (FHSS) in the 2.4 GHz band, and
  - 1 and 2 Mbps over direct sequence spread spectrum (DSSS) in the 2.4 GHz band.
- The transmitter and the receiver can choose the data rate to suite the channel conditions.
2\textsuperscript{nd} version of the IEEE 802.11 Standard for the Physical Layer

- \textit{802.11a} in the 5.0 GHz band using OFDM. Each BSS uses a bandwidth of 20 MHz, which is further divided into 52 OFDM carriers of which 48 are for data. Depending on the channel conditions, the data rates could be any of 6, 9, 12, 18, 24, 36, 48, or 54 Mbps.

- \textit{802.11b} in the 2.4 GHz band using DSSS. Depending on the channel conditions, the data rates could be any of 1, 2, 5.5, or 11 Mbps.

- \textit{802.11g} is an extension of the 802.11b and uses DSSS, OFDM, or both to support data rates in the range of 5.5 to 54 Mbps.

<table>
<thead>
<tr>
<th>Standard</th>
<th>Band</th>
<th>Data rates</th>
<th>Num of channels</th>
</tr>
</thead>
<tbody>
<tr>
<td>IEEE 802.11 (dated)</td>
<td>2.4 GHz</td>
<td>2 and 1 Mbps</td>
<td>1</td>
</tr>
<tr>
<td>IEEE 802.11b</td>
<td>2.4 GHz</td>
<td>11, 5.5, 2 and 1 Mbps</td>
<td>14</td>
</tr>
<tr>
<td>IEEE 802.11a</td>
<td>5.0 GHz</td>
<td>54, 48, 36, 24, 18, 12, 9 and 6 Mbps</td>
<td>12</td>
</tr>
<tr>
<td>IEEE 802.11g</td>
<td>2.4 GHz</td>
<td>1–54 Mbps</td>
<td>14</td>
</tr>
<tr>
<td>IEEE 802.11n</td>
<td>work in progress, data rate up to 540 Mbps</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\textit{Table 7.1 Summary of the 802.11 PHY Standards.}
Two basic protocols are defined: a polling-based protocol called the point coordination function (PCF) and a random access protocol called the distributed coordination function (DCF). PCF and DCF can coexist in the same BSS.

The BSS has a point coordinator that is typically the AP of the BSS. Time is divided into superframes and each superframe has two parts:
- The contention free period (CFP) and
- The contention period (CP).

PCF is used in CFP and DCF in CP. The CFP and CP alternate.

The AP will begin trying to initiate a new CFP, after the target beacon transmission time (TBTT) has elapsed from the time that the previous one was initiated. The TBTT specifies the period of a superframe.
Point Coordination Function (PCF) Protocol

- PCF is initiated by the AP by transmitting a bacon frame. The eligible nodes in the BSS then are polled and the data that need to be transmitted are transmitted along with the polling message.
- If the polled node has packets to transmit, it will also transmit them in response to the polling packet.
- The PCF ends when all nodes are polled by the AP. The end of the PCF mode of medium access is signaled using the End frame. This is also marks the end of the contention-free period.
- This is followed by a contention period using the DCF-based MAC, which continues until the end of the superframe period.
Figure 7.12
Distributed Coordination Function (DCF) Protocol

- The DCF is derived from the CSMA/CA MACAW protocol that we have described earlier. In addition to RTS and CTS based handshake mechanism before the transmission of the data packet, the standard specifies the following:
  - **Minimum silence** periods between transmissions; **different** kinds of packets have to compulsorily wait for different length of time after the medium is sensed idle to begin transmissions. The **minimum** idle sensing time prioritizes different transmissions—shorter minimum waiting time implies higher priority.
  - **Backoff** mechanism to resolve collisions. Like in any backoff mechanism, backoff durations are measured in **multiples** of a basic slot time, the length of which depends on the version.
A single hop network is considered. At the end of the data transmission, there is a short inter-frame space (SIFS), which allows the receiving node to turn around its radio and send back a MAC level ACK packet. When this ACK transmission ends, the channel is sensed to be idle by all the nodes, and each one of them starts a DCF inter-frame space (DIFS) timer. The DIFS duration is more than SIFS.

When the DIFS timers expire, each node enters a backoff phase. Even though the channel is idle, random backoff is used to try to order the transmissions so that a collision does not occur. The node that just completed its data transmission samples a new random backoff value. The backoff durations are multiples of the basic slot time. When a new backoff is sampled, this multiple is sampled uniformly from integers \( \{0, 1, \ldots, CW_{\text{min}} - 1\} \).

A collision occurs if two nodes finish their backoff within one slot of each other. It is assumed that the maximum propagation delay in the network is such that all nodes are able to sense a transmission within one slot time. In this case, both RTS packets collide. A CTS time-out then follows, after which the colliding nodes sample a backoff from a doubled collision window; that is from \( \{0, 1, \ldots, 2 \cdot CW_{\text{min}} - 1\} \).
After the collision event, the nodes that were not involved in the collision continue their backoffs with their residual backoff timers. An extended inter-frame space (EIFS) is taken before the backoff durations are started. Repeated collisions lead to a doubling of the collision window until it reaches $CW_{\text{max}}$, after which the collision window remains fixed.

The interframe spacing, the minimum time after a transmission for which a node has to wait before transmitting, is a prioritization mechanism. For example, since the SIFS < DFIS, the CTS, data, and ACK packets have priority over initiation of new transmissions.
Figure 7.13

node defers; backoff counter frozen
backoff period
### Parameters for Different IEEE802.11 Versions

<table>
<thead>
<tr>
<th>Version</th>
<th>Slot</th>
<th>SIFS</th>
<th>PIFS</th>
<th>DIFS</th>
<th>$CW_{\text{min}}$</th>
<th>$CW_{\text{max}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>IEEE 802.11a</td>
<td>9 μsec</td>
<td>16 μsec</td>
<td>25 μsec</td>
<td>34 μsec</td>
<td>15</td>
<td>1023</td>
</tr>
<tr>
<td>IEEE 802.11b</td>
<td>20 μsec</td>
<td>10 μsec</td>
<td>30 μsec</td>
<td>50 μsec</td>
<td>31</td>
<td>1023</td>
</tr>
<tr>
<td>IEEE 802.11g</td>
<td>9 μsec</td>
<td>10 μsec</td>
<td>19 μsec</td>
<td>28 μsec</td>
<td>16</td>
<td>1024</td>
</tr>
</tbody>
</table>

Table 7.2 Interframe spacings and transmission times and the values of the $CW_{\text{min}}$ and $CW_{\text{max}}$ for the different 802.11a and 802.11b. The handshake packets RTS, CTS and ACK packets are 20, 14, and 14 octets, respectively, and transmitted at the lowest transmission rate. The payload in a packet could be up to 2312 bytes. The MAC header and trailers constitute 34 octets. In addition, there will be PHY headers of 192 bits.
ETSI has defined a standard that is similar to the IEEE 802.11 and is called the high performance radio LAN (HIPERLAN). HIPERLAN is essentially like the IEEE 802.11 at the physical layer but has significant differences with it in the channel access method, which is called the channel access control (CAC). Packets in HIPERLAN are assigned a priority.
Figure 7.14
Saturation Throughput of a Colocated IEEE 802.11-DCF Network

- **Saturation throughput analysis** is an important development in understanding the performance of the CSMA/CA protocol in 802.11. Here, we assume that nodes *always* have packets to transmit; at a node, a successfully transmitted packet is replaced immediately by another packet that needs to be transmitted. This is also called the *infinite backlog* model. The throughput of the network under this saturation assumption is called the *saturation throughput*.

- We consider a *colocated* network of *n saturated* nodes. This model is applicable in the infrastructure mode when all nodes are associated with the same access point, or in the ad hoc mode when the geographical spread of the network is such that each node is within the decode range of all the other nodes. We further assume *homogeneous* nodes; where the *parameters* of the backoff process and the state machine that implements it are *identical* at all nodes.
Saturation Throughput (ctd.)

- The nodes freeze their backoff counters when they sense activity in the medium and resume the count after the mandated silence period. Thus the backoff process is active only when the medium is idle and it is instructive to view the backoff process by considering only the idle times on the channel.

- The busy period could correspond to either a successful transmission or to a collision in which two or more nodes transmit an RTS.

- We assume that in each slot each node attempts a transmission with probability $\beta$ independent of the attempts of other nodes.
Figure 7.15

random backoff at node

successive backoffs

successful attempt

failed attempt

Attempt process in the network
Figure 7.16
Saturation Throughput (ctd.)

- Let $\delta$ denote the slot duration specified by the protocol.
- Let $I$ denote the mean number of slots in the idle period.
- At the end of the idle period, a collision will occur if a second backoff completes within $\delta$ of the first one.
- Let $T_c$ denote the average duration of the collision period. If there is no collision, the transmission will be successful. Let $T_s$ denote the length of the average successful period.
Saturation Throughput (ctd.)

- We will also assume that the collision durations and the packet lengths are independent. We can then conclude that the instants at which idle periods begin are renewal points. The mean time between the successive renewal points is the sum of the mean idle period and the mean busy period. The mean busy period is given by $\gamma T_c + (1-\gamma)T_s$. Thus, the mean time is

$$\text{(I + ((1-\gamma)T_s + \gamma T_c)}$$

- From the renewal reward theorem, the normalized network throughput, $S(\gamma,\beta)$, defined as the fraction of time that the network is involved in transmitting successful packets, is

$$S(\gamma,\beta) = (((1-\gamma)T_s)/(I + ((1-\gamma)T_s + \gamma T_c))$$
Saturation Throughput: Adaptive Backoff

- Here we assume that in each slot each node attempts a transmission with probability $\beta$ independent of the attempts of other nodes.
- Thus, the number of slots in an idle period would follow a geometric distribution and have a mean $\frac{1}{1-(1-\beta)^n}$.
- Then the probability that a transmission attempt suffers a collision is given by $\gamma = 1 - (1 - \beta)^{n-1}$.
- The idle period is followed by a successful transmission if exactly one node attempts a transmission in the slot, conditioned on one or more nodes attempting a transmission. The probability for this event, $P_s(n)$, given by

$$P_s(n) = \frac{n\beta (1 - \beta)^{n-1}}{1 - (1 - \beta)^n}$$
The normalized throughput is

\[ S(\gamma, \beta) = \frac{n\beta (1 - \beta)^{n-1}}{1 - (1 - \beta)^n} T_s \]

Let (K+1) be the maximum number of collisions that a packet can experience before it is discarded. Let \(b_k\) be the mean backoff duration of a node after the k-th collision, \(k=0, 1, 2, \ldots, K\).

Let \(A_j\) be the number of transmission attempts and \(B_j\) be the total backoff duration (in countdown slots) for the j-th packet from a node. \(A_j\) has a truncated (at (K+1)) geometric distribution; that is, the probability that the j-th packet makes k transmission attempts is \((\gamma^{k-1} (1-\gamma))\) for \(k=1,\ldots,K\) and \(\gamma^K\) for \(k=K+1\).
Figure 7.17
Saturation Throughput: Adaptive Backoff (ctd.)

Therefore, we see that

\[ E(A_j) = 1 + \gamma + \gamma^2 + \ldots + \gamma^K \]
\[ E(B_j) = b_0 + \gamma b_1 + \gamma^2 b_2 + \ldots + \gamma^K b_K \]

We can use the renewal reward theorem to say that \( A_j \) is the reward in the renewal period \( B_j \) and obtain

\[ \beta = \frac{E(A_j)}{E(B_j)} \]

Then, we get

\[ G(\gamma) := \beta = \frac{1 + \gamma + \gamma^2 + \ldots + \gamma^K}{b_0 + \gamma b_1 + \gamma^2 b_2 + \ldots + \gamma^K b_K} \]
\[ \Gamma(\beta) := \gamma = 1 - (1 - \beta)^{n-1} \]
Fixed Point Solution

- Solving for $\gamma$ from the two equations in the previous slide is equivalent to asking for a solution to the following fixed point equation:

$$\gamma = \Gamma(G(\gamma))$$

- Since $\Gamma(G(.))$ is a continuous function that maps the interval $[0, 1]$ into itself, Brouwer’s fixed point theorem guarantees that there is a solution to this fixed point equation.

- To validate our analytic method, an exact simulation model that captures all the details of the protocol specification and obtain the performance parameters of interest is constructed. It is heartening to note the close match between the analytical result and that from simulation over a large range of $n$. 
Figure 7.18
Figure 7.19

Saturation throughput (Mbps) vs. Number of nodes (n) for different C values:
- C = 11 Mbps
- C = 5.5 Mbps
- C = 2.0 Mbps

Analysis and Simulation curves are shown.
QoS at the MAC layer can be provided by either of two mechanisms:

- **Per-flow time reservation** with admission control.
  - In this method, MAC-level flows are defined and each flow is guaranteed a certain fraction of time during which the node can transmit. The actual rate of transmission will depend on the characteristics of the medium between the transmitter and the receiver.

- **Service differentiation** by dividing traffic (or node) into different classes and guaranteeing a service quality to each aggregate.
  - In this technique, traffic of the same class compete with one another and receive best-effort-within-class service, and the different classes receive different grades of service in the aggregate. Absolute guarantees of QoS parameters like delay and loss are not provided. Thus, this is also called “better than best effort” service and is suitable for elastic traffic.
A simple method to provide service differentiation would be to assign absolute **priorities** to the classes and provide a **non-preemptive priority** service. However, this can cause **starvation** of the lower priorities and usually is avoided in LAN environment.

A second method to provide service differentiation is to **reserve** capacity for the different classes. Packets of the same class compete among themselves for channel access. In IEEE 802.11 WLAN standard, this can be accomplished in two ways, (1) a **polling** mechanism and (2) an enhanced version of the **DCF**.

Recall that **polled access** is provided during **CFP**. This can be used to serve **different** classes of traffic in **different** ratios by **varying** the polling rate and the duration for which a node is allowed to transmit each time it is polled. This is the method used in the **HCF (hybrid coordination function) controlled channel access (HCCA)**.
The second method, the enhanced DCF (EDCF), is an extension to the IEEE 802.11 DCF. It classifies and prioritizes medium access among the traffic classes, which are called access classes (AC). Recall that the duration of inter-frame spacing can be used to provide priority to different types of packets. Different DIFS durations can be used to give different ACs priorities in transmitting RTS. Further, different backoff windows and backoff window growths are defined for the different classes. This provides service differentiation by changing the probability of obtaining channel access.
Parameters for Access Classes (ACs)

- **Arbitration inter-frame space (AIFS)**
  - This specifies the minimum number of slots for which the AC should sense the channel to be free before attempting a transmission. Higher priority nodes will start backoff countdown earlier than lower priority nodes and hence, will have a higher access probability.

- **Different minimum and maximum contention windows**
  - Different values for $CW_{\text{min}}$ and $CW_{\text{max}}$ are specified for each class. Clearly, having a lower $CW_{\text{min}}$ will increase the probability of success at the first and subsequent early attempts. Similarly, a lower $CW_{\text{max}}$ will increase the success probability if the packet experiences many collisions.

- **The transmission opportunity (TxOP) limit**
  - This specifies the maximum time for which a node can transmit after acquiring the channel. Allowing high priority nodes a larger TxOP implies that their contention cost per bit (or packet) can be reduced.
In the analysis of the EDCF, we assume that the effect of the different MAC parameters described earlier essentially translates to an attempt rate, $\beta_i$ for class i and hence a different conditional collision probability $\gamma_i$. Thus, we have

$$E(A_j) = 1 + \gamma_i + \gamma_i^2 + \ldots + \gamma_i^{K_i}$$

$$E(B_j) = b_{i,0} + \gamma_i b_{i,1} + \gamma_i^2 b_{i,2} + \ldots + \gamma_i^k b_{i,k} + \ldots + \gamma_i^{K_i} b_{i,K_i}$$

$$G_i(\gamma_i) := \beta_i = \frac{1 + \gamma_i + \gamma_i^2 + \ldots + \gamma_i^{K_i}}{b_{i,0} + \gamma_i b_{i,1} + \gamma_i^2 b_{i,2} + \ldots + \gamma_i^k b_{i,k} + \ldots + \gamma_i^{K_i} b_{i,K_i}}$$

$$\gamma_i = \Gamma(\beta_1, \ldots, \beta_n) = 1 - \prod_{j=1, j \neq i}^n (1 - \beta_j)$$

$$= \Gamma(G_i(\gamma_1), \ldots, G_n(\gamma_n))$$

Here, $b_{i,j}$ is the average backoff duration for AC i for the (j+1)-th attempt and $(K_i + 1)$ is the maximum number of attempts that a packet of AC_i will make. The last equation above can be written as

$$\gamma = \Gamma(G(\gamma))$$

This is the vector fixed point equation and it can be shown that the fixed point exists.
Applications on a networked node are typically one of the following types. A **TCP-based session** involving **elastic data** transfer using a closed loop control, for example, we browsing using HTTP over TCP. Another class of application could be a **streaming session**, for example, a packet voice application, also called **Voice over IP (VoIP)**.

We consider a **single hop WLAN** in which a number of STAs associate with an AP, which in turn provides **Internet access** to the STAs via an **uplink**. This is by far the **most widely** used configuration for 802.11 based access.
Data over WLAN

A typical data transfer environment will involve a laptop equipped with an 802.11 interface that is downloading files, like the inbox in mail application, or HTTP documents in a web browsing session. The laptop, or the STA, will be associated with an AP. For both these kinds of downloads, the TCP transport protocol will be used. We now make the following simplifying assumptions:

- The STA will send a TCP-ACK packet for every TCP-data packet the it receives.
- A TCP-ACK packet is generated at an STA as soon as a TCP-data packet is received. All TCP-ACK packets are immediately queued at the MAC for transmission.
- There is no packet loss due to either channel errors or due to buffer overflows.
- The TCP protocol at the sender does not time out due to late ACKs and the WLAN hop on the path from the sender to the STA is the bottleneck and the AP always has a TCP-data packet to transmit.
Figure 7.21
Let $S(t)$ denote the number of nodes competing to transmit on the channel at time $t$. Let $t_k$ be the instant at which the $k$-th successful transmission in the network was completed. We can define the evolution equation of $S(t_k)$ as follows.

$$S(t_k) = S(t_{k-1}) + 1, \text{ if AP is successful;} \quad S(t_k) = S(t_{k-1}) - 1, \text{ if STA is successful.}$$

Let $T_k = t_{k+1} - t_k$. We will also assume that when $n$ nodes are active, the attempt probability is $\beta_n$ and the collision probability is $\gamma_n$. We will assume that $\beta_n$ and $\gamma_n$ are the same as the attempt rate and collision probability, respectively, in an $n$-node saturated network.

Hence, we can embed a Markov chain at $\{t_k\}$ and $\{S(t_k), T_k\}$ will be a Markov renewal process.
Figure 7.22
Let \( \pi_n \) be the stationary probability of there being \( n \) contending nodes, that is, of the event \( \{S(t_k)=n\} \). We can now solve the Markov chain \( \{S(t_k)\} \). We do so by writing the balance equation \( \pi_n / n = n \pi_{n+1} / (n+1) \). From this, we can obtain the recursion \( \pi_{n+1} = (n+1/n^2) \pi_n \). Using the normalizing condition, we find

\[
\pi_n = \frac{n}{(n-1)! (2e)}
\]

Now consider the \( k \)-th renewal period \( (t_{k+1}, t_k) \). Let \( n \) nodes including the AP, be contending for the channel in this period. Let \( R \) be the number of collision in this period. Then \( T_k \) would be the sum of \( (R+1) \) idle periods each of mean duration \( 1/(1-(1-\beta_n)^n) \), \( R \) collision periods each of duration \( T_c \) and one successful transmission period. \( R \) takes values 0,1,.... Further, \( R \) is geometrically distributed with parameter \( \gamma \).
Now, let $H(t)$ be the number of times that AP transmits in the interval $(0, t)$. Our interest is in the rate at which TCP packets can be downloaded, i.e.,

$$
\lim_{t \to \infty} \frac{H(t)}{t} = \frac{1}{E(T_k)} \sum_{i=1}^{\infty} \pi_n \frac{1}{n+1}
$$

From the discussion in the previous slide, we obtain

$$
E(T_k) = \sum_{n=1}^{\infty} \pi_n \left( \frac{1}{1 - \gamma_n} \frac{1}{(1 - \beta_n)^n} \right) + \left[ \frac{\gamma_n}{1 - \gamma_n} T_c \right] + \left[ \frac{1}{n} T_{DATA} + \frac{n-1}{n} T_{ACK} \right]
$$

Here the term in the first square brackets corresponds to the mean idle period, the term in the second square brackets corresponds to the mean collision duration, and the term in the third square brackets corresponds to the mean duration of a successful transmission in $T_k$. Note that we have assumed that $\beta_n$ and $\gamma_n$ have the same values as in an $n$-node saturated network.
Figure 7.23

The graph shows the Aggregate AP throughput ($\Theta_{AP-ftp}$) in Mbps as a function of the Number of FTP connections, $N$. The analysis for $N=1$ and $N=\infty$ is compared with simulation results.

- Analysis; $N=1$: For $N=1$, the aggregate throughput remains constant at approximately 2 Mbps.
- Analysis; $N=\infty$: For $N=\infty$, the aggregate throughput increases linearly with the number of connections.
- Simulation: The simulation results show a similar trend but with slight variations compared to the analysis.

Key points:
- At $N=1$, the throughput is close to 2 Mbps.
- As $N$ increases, the throughput for both analysis and simulation increases.
- The simulation data points align closely with the linear trend predicted by the analysis for $N=\infty$.

Legend:
- × Analysis; $N=1$
- — Analysis; $N=\infty$
- ○ Simulation
Voice over WLAN
Figure 7.24

Voice Packet Format

| IP Hdr_{20} | UDP Hdr_{8} | RTP Hdr_{12} | Voice Payload |

Uplink to Wired Network
Association in IEEE 802.11 WLANs
Figure 7.25