Application Models and Performance Issues

- We will discuss the major applications that wireless networks are used for and the performance issues related to these applications.
A Big Picture

Let us first have a big picture of telecommunication networks as they exist today, showing all the elements, including the phone network, the Internet, and various wireless access networks.

- **Public Switched Telephone Networks (PSTN);** has carried telephone calls for nearly a century.
- **Cellular Networks** have provided mobile access since the early 1980s, and have already evolved through three generations of commercial deployment.
- **Wireless Local Area Networks (WLANs),** providing access to Internet in campuses and enterprises through devices such as laptops and personal digital assistants (PDA).
- **Wireless Metropolitan Area Networks (WMANs).**
- **Ad Hoc Multi-hop Wireless Mesh Networks.**
Figure 3.1
Types of Traffic and QoS Requirements

- We can distinguish three types of traffic as follows:
  - Elastic Traffic; e.g., WWW browsing, FTP file transfer, and electronic mail.
  - Real-Time Stream Traffic; e.g., packet voice telephony.
  - Store-and-Forward Stream Traffic; e.g., streaming movies or music over the Internet.
Elastic Traffic

- Elastic traffic usually do not have an intrinsic temporal behavior, and can be transported at arbitrary transfer rates. The following are some QoS requirements of this type of traffic:
  - Transfer delay and delay variability can be tolerated. An elastic transfer can be performed over a wide range of transfer rates, and the rate can even vary over the duration of the transfer.
  - The application cannot tolerate data loss. This does not mean, however, that the network cannot lose data. Packet can be lost in the network provided that lost packets are recovered by an automatic retransmission procedure. Thus, effectively the application would see a lossless transport service. Since elastic sources do not require delay guarantees, the delay involved in recovering lost packets can be tolerated.
The following are the typical QoS requirements of real-time stream sources:

- **Delay (average and variation) need to be controlled.** For example, for **wide area packet telephony**, the delay may need to be **less than 200 ms** with a probability of **0.99**. Packets that do **not conform** to the delay bound are considered **lost**.

- **There is tolerance to data loss.** Owing to the high levels of redundancy in speech and images, a certain amount of data loss is imperceptible. For example, for **packet voice in which each packet carries 20 ms of speech**, and the receiver does lost-packet interpolation, **5 to 10%** of the packets can be lost without significant degradation of speech quality.
Figure 3.2

- First packet in a burst
- Peak rate = $R$ bits/sec
- Source output
- Network output
- Input to playout device
- Playout delay
Store-and-Forward Stream Traffic

- This is a kind of traffic that is generated by applications such as streaming audio and video. Such applications basically involve a one-way transfer of an audio or video file stored on disk of a media server.

- Thus, the problem of transporting streaming audio or video becomes just another case of transferring elastic data, with appropriate receiver adaptation. Following are the QoS requirements of such traffics:
  - The average transfer rate provided in the network should match (in fact, should be greater than) the average rate at which the stored media has been encoded.
  - The transfer rate variability should not be too large.
Real-Time Stream Session: Delay Guarantees (in voice telephony)

- **CBR Speech**: In order to carry a CBR voice source, it is necessary for the network to use a service rate greater than or equal to the voice bit rate. Further, if the source is allocated exactly the constant bit rate then there will not be queuing. Hence, for CBR sources it is sufficient to allocate the CBR rate.

- Consider a voice call between a PSTN phone A and the cell phone C. If the gateway GW converts PCM speech arriving over the PSTN to CBR speech at rate R, then the cellular network can just allocate resources so that the voice call is provided a service rate of R. This is typically what is done in an FDM-TDMA cellular system (such as GSM), or in a CDMA cellular system.
Real-Time Stream Session: Delay Guarantees (in voice telephony)

- **VBR Speech:**
  - In speech generated by interactive telephony, there are low energy periods that correspond to silences while the speaker listens, or gaps between words, sentences, and utterances. The coder output corresponding to these inactive periods can be discarded or encoded at a lower rate. This yields a variable bit rate (VBR) coded speech. The VBR speech can be handled as variable rate byte stream, or can be packetized for transport over a packet network.

- Suppose, the VBR speech source is allocated the service rate $C$ in cellular network. Let us denote by $R$ the peak rate of the VBR source, and by $r$ the average rate. It is clear that we need to have $r < C < R$.

- To avoid any rate mismatch, one approach is to use peak rate allocation. This is the approach usually used in FDM-TDMA (e.g., GSM) and CDMA cellular systems.
buffer fills when source rate > service rate

\[
\tilde{r} < C < R
\]
Real-Time Stream Session: Delay Guarantees (in voice telephony)

- **Speech Playout:**
  - Consider a VBR coded packetized speech telephony between devices B and D, or B and C. Suppose that the cellular network handles only CBR speech; then the problem of converting the speech packets, arriving asynchronously over the Internet, to CBR speech over the cellular network will be a task of the gateway, GW, between the Internet and the cellular network.
  - An obvious method to be used is the policy of deferred playout. A playout delay is applied to each packet to allow trailing packets to catch up.
  - Suppose that we are able to determine a value T such that the packet delay rarely exceeds T. Packets that are delayed more than T are lost and may be interpolated.
Figure 3.4

this target end-to-end delay $T$
will avoid any packet loss
Figure 3.5
Elastic Transfers: Feedback Control

- Elastic traffic is generated by applications whose objective is to move chunks of data between disks of two computers connected to the network. Elastic data can be speeded or slowed down depending on the number of flows contending for the capacity of the network. Other examples are applications using FTP and HTTP protocols.

- In all these cases the basic problem is to transfer each file in its entirety from the source machine to the destination machine. There is no intrinsic rate at which the files must be transferred.
Figure 3.6
Figure 3.7

bandwidth of this pipe is shared between a varying number of elastic flows

feedback control necessary to slow down or speed up the traffic sources

users downloading data from the web servers
TCP Performance over Wireless Links

- **Independent Packet Losses**: Consider a simple scenario in which a mobile host is doing TCP controlled file transfer from a file server on a wired LAN. The LAN wireless router network would be located at the base station. The propagation delay between the base station and the mobile host is negligible. The BER on the wireless link is such that the packet are lost with probability $p$. The packets are lost independently; correlated losses owing to channel fading are not modeled here.
Figure 3.8
Figure 3.9

Packet throughput normalized to speed of lossy link

- NewReno: $K=3$
- Tahoe: $K=3$
- Reno: $K=3$
- OldTahoe

Packet Error Probability

$10^{-3}$ $10^{-2}$ $10^{-1}$ $10^{0}$
Correlated Packet Losses

- Let us consider the performance of TCP controlled file transfer over a fading channel. Note that the fading is correlated in time. Thus, for a given average BER there are periods when the BER is greater than average, and periods during which the BER is less than average.

- A simple approach is to model the channel as being in one of two states: a Good state (during which the packet transmission is successful), and a Bad state (during which a packet transmission is unsuccessful). A further simplification is to model the state process as a two-state Markov process, on state space \{good, Bad\}. The durations in each state are taken to be multiples of the packet transmission time. The transition probabilities of the Markov chain are obtained by specifying the amount of fading that leads to a bad transmission.
Figure 3.10

Diagram showing transitions between 'Good' and 'Bad' states with probabilities $(1-g)$ and $(1-b)$. Arrows indicate the direction of transitions.
Figure 3.12

Packet throughput, normalized to speed of wireless link

Mean Signal to Noise Ratio (dB)

TCP Tahoe; no fading
TCP Tahoe; fade = 1 pkts
TCP Tahoe; fade = 2 pkts
* limit, speed → 0

Figure 3.12