Packet Switching

Nature seems . . . to reach many of her ends by long circuitous routes.

—Rudolph Lotze

The directly connected networks described in the previous chapter suffer from two limitations. First, there is a limit to how many hosts can be attached. For example, only two hosts can be attached to a point-to-point link, and the Ethernet specification allows no more than 1,024 hosts. Second, there is a limit to how large of a geographic area a single network can serve. For example, an Ethernet can span only 2,500 m, wireless networks are limited by the ranges of their radios, and even though point-to-point links can be quite long, they do not really serve the area between the two ends. Since our goal is to build networks that can be global in scale, the next problem is therefore to enable communication between hosts that are not directly connected.

P R O B L E M

Not All Networks Are Directly Connected

This problem is similar to one addressed in the telephone network: Your phone is not directly connected to every person you might want to call, but instead is connected to an exchange that contains a *switch*. It is the switches that create the impression that you have a connection to the person at the other end of the call. Similarly, computer networks use *packet switches* to enable packets to travel from one host to another, even when no direct connection exists between those hosts. This chapter introduces the major concepts of packet switching, which lies at the heart of computer networking.

A packet switch is a device with several inputs and outputs leading to and from the hosts that the switch interconnects. The core job of a switch is to take packets that arrive on an input and *forward* (or *switch*) them to the right output so that they will reach their appropriate destination. There are a variety of ways that the switch can determine the "right" output for a packet, which can be broadly categorized as connectionless and connection-oriented approaches.

3

A key problem that a switch must deal with is the finite bandwidth of its outputs. If packets destined for a certain output arrive at a switch and their arrival rate exceeds the capacity of that output, then we have a problem of *contention*. The switch queues (buffers) packets until the contention subsides, but if it lasts too long, the switch will run out of buffer space and be forced to discard packets. When packets are discarded too frequently, the switch is said to be *congested*. The ability of a switch to handle contention is a key aspect of its performance.

This chapter introduces the issues of forwarding and contention in packet switches. We begin by considering the various approaches to switching, including the connectionless and connection-oriented models. We then examine two particular technologies in detail. The first is *LAN switching*, which has evolved from Ethernet *bridging* to become one of the dominant technologies in today's LAN environments. The second noteworthy switching technology is *asynchronous transfer mode (ATM)*, which was initially developed to meet the needs of telecommunications service providers in wide area networks. Finally, we consider some of the aspects of switch design that must be taken into account when building large-scale networks.