when a segment arrives, the receiver knows whether the segment experienced congestion at any point. Unfortunately, the sender, not the receiver, needs to learn about congestion. Therefore, the receiver sets the *ECE* code bit in the next ACK to inform the sender that congestion occurred. The sender then responds by reducing its congestion window and setting the *CWR* code bit.

In addition to the pair of code bits in the TCP header, ECN uses two bits in the IP header to allow routers to record congestion. Bits in the IP header are taken from unused bits in the TYPE OF SERVICE field. A router can choose to set either bit to specify that congestion occurred (two bits are used to make the mechanism more robust).

11.22 Congestion, Tail Drop, And TCP

We said that communication protocols are divided into layers to make it possible for designers to focus on a single problem at a time. The separation of functionality into layers is both necessary and useful — it means that one layer can be changed without affecting other layers, but it means that layers operate in isolation. For example, because it operates end-to-end, TCP remains unchanged when the path between the endpoints changes (e.g., routes change or additional networks routers are added). However, the isolation of layers restricts inter-layer communication. In particular, although TCP on the original source interacts with TCP on the ultimate destination, it cannot interact with lower-layer elements along the path†. Thus, neither the sending nor receiving TCP receives reports about conditions in the network, nor does either end inform lower layers along the path before transferring data.

Researchers have observed that the lack of communication between layers means that the choice of policy or implementation at one layer can have a dramatic effect on the performance of higher layers. In the case of TCP, policies that routers use to handle datagrams can have a significant effect on both the performance of a single TCP connection and the aggregate throughput of all connections. For example, if a router delays some datagrams more than others‡, TCP will back off its retransmission timer. If the delay exceeds the retransmission timeout, TCP will assume congestion has occurred. Thus, although each layer is defined independently, researchers try to devise mechanisms and implementations that work well with protocols in other layers.

The most important interaction between IP implementation policies and TCP occurs when a router becomes overrun and drops datagrams. Because a router places each incoming datagram in a queue in memory until it can be processed, the policy focuses on queue management. When datagrams arrive faster than they can be forwarded, the queue grows; when datagrams arrive slower than they can be forwarded, the queue shrinks. However, because memory is finite, the queue cannot grow without bound. Early routers used a *tail-drop* policy to manage queue overflow:

† The Explicit Congestion Notification scheme mentioned above has not yet been adopted.

‡Variance in delay is referred to as *jitter*.