**Basic Sockets**

**2.1 Socket Addresses**

The InetAddress abstraction represents a network destination, encapsulating both names and numerical address information. The class has two subclasses, Inet4Address and Inet6Address, representing the two versions in use. Instances of InetAddress are immutable: once created, each one always refers to the same address.

To get the addresses of the local host, the program takes advantage of the Network Interface abstraction. Recall that IP addresses are actually assigned to the connection between a host and a network (and not to the host itself). The NetworkInterface class provides access to information about all of a host’s interfaces. This is extremely useful, for example when a program needs to inform another program of its address.

**2.2 TCP Sockets**

Java provides two classes for TCP: Socket and ServerSocket. An instance of Socket represents one end of a TCP connection. A TCP connection is an abstract two-way channel whose ends are each identified by an IP address and port number. Before being used for communication, a TCP connection must go through a setup phase, which starts with the client’s TCP sending a connection request to the server’s TCP. An instance of ServerSocket listens for TCP connection requests and creates a new Socket instance to handle each incoming connection. Thus, servers handle both ServerSocket and Socket instances, while clients use only Socket

**2.2.1 TCP Client**

The client initiates communication with a server that is passively waiting to be contacted. The typical TCP client goes through three steps:

1. Construct an instance of Socket: The constructor establishes a TCP connection to the specified remote host and port.

2. Communicate using the socket’s I/O streams: A connected instance of Socket contains an InputStream and OutputStream that can be used just like any other Java I/O stream

3. Close the connection using the close() method of Socket.

**2.2.2 TCP Server**

The server’s job is to set up a communication endpoint and passively wait for connections from clients. The typical TCP server goes

through two steps:

1. Construct a ServerSocket instance, specifying the local port. This socket listens for incoming connections to the specified port.

2. Repeatedly:

a. Call the accept() method of ServerSocket to get the next incoming client connection. Upon establishment of a new client connection, an instance of Socket for the new connection is created and returned by accept().

b. Communicate with the client using the returned Socket’s InputStream and OutputStream.

c. When finished, close the new client socket connection using the close() method of Socket.

**2.2.3 Input and Output Streams**

The basic I/O paradigm for TCP sockets in Java is the stream abstraction. A stream is simply an ordered sequence of bytes. Java input streams support reading bytes, and output streams support writing bytes. In our TCP client and server, each Socket instance holds an InputStream and an OutputStream instance. When we write to the output stream of a Socket, the bytes can (eventually) be read from the input stream of the Socket at the other end of the connection.

OutputStream is the abstract superclass of all output streams in Java. Using an OutputStream, we can write bytes to, flush, and close the output stream.

**2.3 UDP Sockets**

UDP provides an end-to-end service different from that of TCP. In fact, UDP performs only two functions: 1) it adds another layer of addressing (ports) to that of IP, and 2) it detects some forms of data corruption that may occur in transit and discards any corrupted messages. Because of this simplicity, UDP sockets have some different characteristics from the TCP sockets.

Another difference between UDP sockets and TCP sockets is the way that they deal with message boundaries: UDP sockets preserve them. A final difference is that the end-to-end transport service UDP provides is best-effort: there is no guarantee that a message sent via a UDP socket will arrive at its destination, and messages can be delivered in a different order than they were sent (just like letters sent through the mail). A program using UDP sockets must therefore be prepared to deal with loss and reordering.

Given this additional burden, why would an application use UDP instead of TCP? One reason is efficiency: if the application exchanges only a small amount of data—say, a single request message from client to server and a single response message in the other direction—TCP’s connection establishment phase at least doubles the number of messages (and the number of round-trip delays) required for the communication. Another reason is flexibility: when something other than a reliable byte-stream service is required, UDP provides a minimal-overhead platform on which to implement whatever is needed.

Java programmers use UDP sockets via the classes DatagramPacket and DatagramSocket. Both clients and servers use DatagramSockets to send and receive DatagramPackets.

**2.3.1 DatagramPacket**

Instead of sending and receiving streams of bytes as with TCP, UDP endpoints exchange self-contained messages, called datagrams, which are represented in Java as instances of DatagramPacket.

In addition to the data, each instance of DatagramPacket also contains address and port information, the semantics of which depend on whether the datagram is being sent or received.

**2.3.2 UDP Client**

A UDP client begins by sending a datagram to a server that is passively waiting to be contacted. The typical UDP client goes through three steps:

1. Construct an instance of DatagramSocket, optionally specifying the local address and port.

2. Communicate by sending and receiving instances of DatagramPacket using the send() and receive() methods of DatagramSocket.

3. When finished, deallocate the socket using the close() method of DatagramSocket.

**2.3.3 UDP Server**

Like a TCP server, a UDP server’s job is to set up a communication endpoint and passively wait for clients to initiate communication; however, since UDP is connectionless, UDP communication is initiated by a datagram from the client, without going through a connection setup as in

TCP. The typical UDP server goes through three steps:

1. Construct an instance of DatagramSocket, specifying the local port and, optionally, the local address. The server is now ready to receive datagrams from any client.

2. Receive an instance of DatagramPacket using the receive() method of DatagramSocket. When receive() returns, the datagram contains the client’s address so we know where to send the reply.

3. Communicate by sending and receiving DatagramPackets using the send() and receive() methods of DatagramSocket.

**2.3.4 Sending and Receiving with UDP Sockets**

A subtle but important difference is that UDP preserves message boundaries. Each call to receive() on a DatagramSocket returns data from at most one call to send(). Moreover, different calls to receive() will never return data from the same call to send().