- **Internet** as a network of networks
  - The Internet is essentially a network of many networks, each of which are owned and operated by different organizations, each of which can be called an *Internet Service Provider* (**ISP**).
  - The Internet is loosely hierarchical. This can be represented by edge networks or **access networks**, and **core networks**. ISPs at the bottom of the hierarchy are access ISPs such as residential ISPs, university ISPs, and enterprise ISPs. ISPs at the top of the hierarchy are called tier-1 ISPs and typically include long-haul intra- and intercontinental fiber links. Tier-n ISPs provide service (for a price) to tier-(n+1) ISPs. Each ISP is *independently* managed, making it difficult to deploy common services. However, ISPs employ a common protocol suite called the Internet Protocol, which is better known as IP, to connect them (discussed in Chapter 4).

- **Protocol**
  - A protocol essentially defines the rules how two or more entities on different end systems (hosts) can communicate with each other. Since the only way that they could communicate is through the exchange of messages, a protocol defines the format and order of messages exchanged between two or more communication entities, as well as the actions taken on the transmission and/or receipt of a message or other event. (You will see many protocols !)

- **Client/server** versus **Peer-to-Peer model**
  - In a client and server mode, a client always requests certain service(s) from a server, and the server is expected to be **always** available. In a client-server architecture, two clients never communicate with each other directly, while server is always online with a well-known address (IP address and port number to be discussed), so a client can always contact a server by this address
  - In a Peer-to-Peer (P2P) model, a peer can request service (content) from another peer, while also providing service (content) for other peer, but there is no requirement for any peer to be available all the time. Therefore, one of the fundamental problems (as you can imagine) in a P2P system is to deal with peer dynamics, i.e., peers can come and leave at any time. The rationale behind P2P model is that there exist a reasonably large number of peers that are willing to provide certain content to other peers in the Internet.

- **Circuit Switching**
  - Traditional digital telephone network, which is completely different from the Internet.
  - Resources along a path are **reserved** and **dedicated** for a connection even

if there is no traffic along the path. There is always a connection (i.e., a circuit) set-up process (phase) prior to any communications.

- Quality-of-Service or QoS is guaranteed once the circuit (connection) is established; this implies that if the system can not satisfy the requirement of a new connection, this connection cannot be established (rejected). This function is part of a control mechanism called call *admission control*. There is no such a function in today's Internet, in which there is no such QoS guarantee.
- The reserved connection bandwidth is "wasted" whenever the end systems are not sending data.
- Example is the digital telephone network. FDM and TDM are commonly used to divide a large link capacity into small pieces in circuit-switched networks. They can also be used in combination, for example, in GSM cellular networks, they use a combination of FDM and TDM, i.e., firstly it divides the bandwidth using FDM into small frequency bands, then further uses TDM to divide each small frequency band into smaller time units.

- **Packet Switching**
  - Resources are statistically shared by all traffic, with no reservation, referred as **statistical multiplexing**. You will see the notions of multiplexing and demultiplexing again in all the subsequent chapters. This enables the most effective use of the link bandwidth, which is the major proponent for packet switching. The example on Slide 1-42 illustrates that packet switching can accommodate more users if the users are bursty, i.e., they inject different amount of traffic into the Internet at different times.
  - The above example also shows that it will be difficult to guarantee the QoS of each user with packet switching, since there is always a small probability that traffic can overload the system (in this particular example, the chance is very small). In reality, there could be well more than 35 users sending traffic to this link, since there is no limit in the Internet. The link (or the network) can be congested at certain times. This is one of the prices that packet switching has to pay for its gain of the better link utilization
  - Another interesting point in this example (Slide 1-42) is that circuit switching can provide QoS guarantee with **over-provisioning**. This means that each of the 10 users is essentially given the maximum bandwidth that each user ever needs (100 kbps). In most cases, this dedicated bandwidth allocation to each user will not be utilized all the time, or is always under-utilized. This is the key price circuit switching pays in order for the service guarantee

- **Delay in the Internet**
  - **Processing delay** examines the packet so to figure out where to forward the packet to the next hop, and also possibly for error checking (you will see the details in Chapter 4 when we discuss routing). This is typically in the

order of microseconds in high-speed routers

- **Queuing delay**, the time waiting to be transmitted in routers/switches, this can be zero, or on the order of microseconds to milliseconds depending on the traffic along the path (route) from the source to the destination.
- **Transmission delay**, assuming the packet length is L (in bits), and link bandwidth (rate) is R (bits per second), the transmission time is L/R, usually microseconds or millisecond. High-speed link results in small transmission time since R is large. It is the amount of time it takes to push the packet onto the link or out of routers/switches
- **Propagation delay**, the distance between two routers (d) divided by the propagation speed (speed of light, s), d/s. In a wide-area network, this is in the order of milliseconds. The propagation delay over a link is the time it takes one bit to travel from one end of the link to the other.
- There is no time called reception time! When the last bit reaches the destination node, all other bits (in front of it) have been received by the destination. So typically, ignoring processing and queuing time, the transmission time plus and propagation delay is the time required for a packet to leave the source and reach the directly connected destination (received or stored completely by the destination).

- **Layered architecture**
  - Layering approach provides the modularity and functional decomposition, which is effective when dealing with large and complex software system. One negative aspect is that this could be inflexible and might cause redundancy of certain functions in different layers (seen in later chapters).
  - Internet protocol layers are divided into five protocol stacks:
    - **Application layer** is the application specific protocol;
    - **Transport layer** adds certain control between two communication processes in two different end systems such as reliability control, congestion control, multiplexing and de-multiplexing. Since this deals with two processes at the end systems, this is often referred as an *end-to-end protocol*. In another word, no routers (not end systems) have any transport layer functions;
    - **Network layer protocol** is primarily responsible for routing a packet from a source to a destination, or to next router/switch that is one hop closer to the destination;
    - **Link layer protocol** deals with the packet transmission between two nodes directly connected, including functions such as channel access control if there are more than one nodes that can transmit over the same channel when it is shared (shared medium access control), packet framing and perhaps reliability again;
    - **Physical layer** deals with the bit transmission in physical medium, apparently different mediums such as fiber optics, copper, wireless have different characteristics.