- **Forwarding and routing**
  - **Forwarding**: moving a packet from an input port to an output port. This is usually done by using the destination IP address or an identifier in the packet header as an index to look up the forwarding table. The design of high spend routers requires this lookup process to be done very fast (often in hardware) considering that a router potentially has to forward million of packets per second.
  - **Routing**: this is far more complex; it is a series of protocols running in routers that collectively calculate and determine the route, i.e., fill the entry in the routing (forwarding) table. In another words, routing determines the end-to-end routes between sources and destinations.

- **Virtual Circuit**
  - Connection set up process like circuit switching, but still one type of packet switching network since the bandwidth along the path is statistically sharing, unlike in the circuit switching network where the bandwidth and other resources are dedicated to one circuit.
  - Each switch or router has one routing table at each input port, unlike the Internet router that has only one routing table that all input ports use. What this implies is that in the Internet, when a router receives a packet, it does not matter which input port this packet arrives, what it matters is only the destination IP address (index to the routing table).
  - The VC number only has local significance, that is to say that each port selects a VC number locally available, since selecting a globally unique number is not feasible (requiring to know all VCs used currently in a large network), so a complication in the virtual circuit network is to change a VC each time a packet is routed to another switch/router. This is called **label swapping**.

- **Fragmentation and Reassembly**
  - It is inevitable for fragmentation some time as each physical network s has a restriction on the maximum transmit unit (MTU).
  - There are three things to notice in fragmentation and reassembly, 1) each packet has a unique ID, in addition to the source and destination address, so fragments of the same IP packet copy that; 2) fragment flag is set to 1, meaning this is not the last fragment (meaning that more fragment(s) after this), 0 means either there is no fragment after this or this is the last fragment; 3) offset is the chunk of 8 bytes data from the beginning of the packet (used to assemble the fragments together at the destination).

- **IP address**
  - IP address is hierarchical in that it is divided into a subnet address (network address and host ID with the network). This is important for scalability, so in each router it has **only one** entry for each network, not for each host. So

we often say a router routes packet (datagram) by network. The negative side is that this network address once assigned to an organization, cannot be used anywhere in the world even if the organization cannot use up all of the addresses within the network.

- IP addresses are divided into classes, A, B, C, D and E.

Class	First 4 bits	# bits (network)	# bits (host)	Network range
A	0xxx	7	24	1.0.0.0 – 127.255.255.255
B	10xx	14	16	128.0.0.0-191.255.255.255
C	110X	21	8	192.0.0.0-223.255.255.255
C	1110	28	multicast	224.0.0.0.-239.255.255.255
E	1111		reserved	240.0.0.0- 247.255.255.255

- The class A has up to 126 networks each can have 16 million hosts, class B has 16,382 networks with up to 64K hosts each, and class C has up to 2 million networks with up to 254 hosts each. This class-based address clearly has limitations, as class C address is too small for most of the organizations, and we have too few class A and class B addresses.
- First byte address 127 is for loopback, all 1 (4 bytes) address is for broadcast on the local network, and all 1 address (in the host ID) is for broadcast in a distant network (identified by the network ID).
- **CIDR**: *Classless InterDomain Routing* address allows the subnet portion of the address to be arbitrary length, in the format of **a.b.c.d/x,** instead of at the one byte boundary. In practice, this usually combines multiple class C address (2, 4 or 8 ..) into one big address, and specifying the number of bits in the address part, for example, 200.23.16.0/23 contains two class C subnets.
- **DHCP**: *Dynamic Host Configuration Protocol* allows a host to dynamically obtain IP address from network server when it first joins the network.
- **NAT**: *Network Address Translation* uses one IP address for all hosts within a network. This separates the publically known IP address from the local host addresses. Address translation is needed with the assistance of port number for packet multiplexing (outgoing packets) and demultiplexing (incoming packets). This offers flexibility and security.
- IPv6 use 128 bit address instead of 32 bit IPv4 address. IPv6 packets can be carried as "payload" in IPv4 datagram between IPv4 routers; this is called **tunneling**.

- **Distance Vector Routing and RIP**
  - RIP is a distance vector (DV) protocol used in the Internet. It uses Bellman-Ford algorithm to calculate the path between two end points.
  - **Hop count** is the cost in RIP, in which initially each router maintains a list in the routing table that the routers it can directly reach
  - Periodically (typically 30 seconds) routing table is shared between

neighboring routers via a routing protocol.

- If two identical paths to the same network exist, only the one with the smallest hop-count is kept. When the new table has been cleaned up, it may be used to replace the existing routing table used for packet forwarding
- The new routing table is then communicated to all neighbors of this router. The routing information will spread and eventually all routers know the routing path to each network, which router it shall use to reach this network, and to which router it shall route next.
- *Distance-vector routing protocols* are simple and efficient in small networks, and require little, if any management. However, they do not scale well, and have poor convergence properties, which has led to the development of more complex but more scalable protocols for use in large networks.

- **Link State Routing and OSPF**
  - The link-state protocol is performed by each router in the network. The basic concept of link-state routing is that every router receives a map of the connectivity of the network, in the form of a graph showing which nodes are connected to which other nodes. Each node then independently calculates the best **next hop** from it for every possible destination in the network. (It does this using only its local copy of the map, and without communicating in any other way with any other node.) The collection of best next hops forms the routing table for the node.
  - This contrasts with **distance-vector routing protocols**, which work by having each node share its routing table with its neighbors. In a link-state protocol, the only information passed between the nodes is information used to construct the connectivity maps.
  - There are two steps in the link-state protocol, 1) distributing link states (network topology) that is done by flooding the link-state advertisement; 2) each node iteratively computes the local map using the Dijkstra's shortest path routing algorithm
  - The primary advantage of link-state routing is that it reacts more quickly, and in a bounded amount of time, to the connectivity changes. The primary disadvantage of link-state routing is that it requires more storage and more computing than a distance-vector routing protocol.
  - The key difference between a link-state algorithm and a distance-vector algorithm is that a link-state algorithm computes the least-cost path between source and destination using complete, global knowledge about the network. In a distance-vector protocol, the calculation of the least-cost path is carried out in an iterative, distributed manner; each node only knows the neighbor to which it should forward a packet on the way to the destination along the least-cost path. This results in different performance in terms of convergence, routing loops and etc. The trade-off is that a link-state algorithm generates more link-state broadcast messages.

- **Border Gateway Protocol (BGP)**
  - The **Border Gateway Protocol (BGP)** is the core routing protocol of the Internet. It works by maintaining a table of IP networks or 'prefixes' which designate network **reachability** between autonomous systems (AS). It is described as a path vector protocol. BGP does not use technical metrics, but makes routing decisions based on network policies or rules.
  - BGP neighbors, or peers, are established by manual configuration between routers to create a TCP session on port 179. Among all routing protocols, BGP is unique in using TCP as its transport protocol. When BGP is running inside an autonomous system (AS), it is referred to as Internal BGP. When BGP runs between AS, it is called External BGP. If the role of a BGP router is to route iBGP traffic, it is called a transit router. Routers that sit on the boundary of an AS and that use EBGP to exchange information with the ISP are called border or edge routers.
  - There are two attributes in BGP, **AS-PATH** contains the Ass through which prefix advertisement has passed, which can be used to detect and prevent looping advertisements and also choosing among multiple paths to the same prefix. **NEXT-HOP** indicates the specific internal-AS router to next-hop AS.