A Computer Network is a collection of devices or nodes (Personal Computer (PC), server, router, printer etc.) connected together by means of a wired or wireless medium. Resource sharing and communication are two principal reasons for building and using computer networks. A simplified description of the uses of computer networks might be as follows (Norman 2005):

- **Sharing of Hardware/Resources:** Several PCs might be networked together in a wired or wireless local area network (LAN) to share a printer/scanner.
- **Sharing of Information:** The sharing involves both LANs and wide area networks (WANs). Distributed databases, e-mail, the World Wide Web and so on are examples of this.

There are many aspects of building a successful computer network. Hadi (2011) expressed that, a successful network design takes account of several important factors, such as cost, security, integrity, scalability, reliability and robustness. Computer networks are continuing to play more vital roles in our everyday lives. These networks hold sensitive data, mission critical applications, and time-sensitive functionalities such as health management, defense systems, and credit information. Their reliability and robustness, therefore, is of grave importance. Most real networks are not designed in a structured manner or with a keen eye for reliability and robustness (Albert et al. 2000). The typical expansions to these networks are not carried out with proper consideration for fault tolerance fundamentals. In fact, most networks grow in a preferential manner, where the more active and connected nodes on the network continue to be the preferred candidates for the new nodes to connect to (rich get richer). Today networks touch upon virtually every aspect of computing. Therefore, better and more reliable network configurations are needed to meet the fault tolerance and efficiency requirements of computer networks. A network that is designed with special attention to these efficiency factors will generally provide better throughput, is less costly, and is easier to manage. The Internet is the largest and most popular
packet switched network. Packet switched networks transfer information between specified destinations as per the need of user or an application. Routing of traffic to desired destination is most important function performed by packet switched networks. The topology design and routing plays a major role in efficient network design.

The brief organization of chapter is as follows: The classification of computer networks on the basis of topology, geographical area covered, switching methods and transmission technology is carried out in section 1.1. Section 1.2, 1.3 and 1.4 gives a brief introduction of protocols involved, services provided and performance issues in computer networks respectively. Concepts of routing are explained in section 1.5. Section 1.6 introduces about various issues in dynamic routing protocols. A description of various areas that need optimization is carried out in section 1.7. Section 1.8 introduces about various optimization techniques for single and multi-objective problems. Section 1.9 deals with evolutionary computation techniques as an optimization tool for NP-hard problems. In evolutionary algorithms, genetic algorithms are considered as most adequate algorithms for optimization problem. Summary is given in section 1.10.

1.1 Classification of Computer Networks

Classification can be carried out on the basis of geographical area covered, methods of connecting the nodes in network, transmission technology and many other factors. A brief classification is given below:

1.1.1 Topology Based

Topology defines a way of interconnecting the nodes in a network. A medium is required to connect the nodes with each other, which can be wired or wireless. The topology design is a major issue in computer networks. The use of different type of communication medium and way of interconnecting the nodes play a major role in reliability and cost of setting up a network. Topology design is always a typical task that needs continuous optimization for minimizing the setup cost and maximizing the reliability and throughput using different types of communication mediums. Following are the wired and wireless mediums that are used to connect nodes in a network (Karthik 2011):
- **Coaxial cable**: It is widely used for cable television systems, office buildings, and other work-sites for local area networks. Coaxial cable is more resistant to interference and attenuation than twisted-pair cabling. Generally these are used up to a distance of 500 meters with speed of 10 Mbps.

- **Twisted pair wire**: It is the most widely used medium for telecommunication. Transmission speed ranges from 1 Mbps to 10,000 Mbps.

- **Optical fiber cable**: It transmits light which can travel over extended distances. Fiber-optic cables are not affected by electromagnetic radiation. Transmission speed may reach trillions of bits per second.

- **Terrestrial microwave**: Terrestrial microwaves use Earth-based transmitter and receiver. The equipment looks similar to satellite dishes. Terrestrial microwaves use low-gigahertz range, which limits all communications to line-of-sight. Path between relay stations spaced approx. 48 km apart.

- **Communications satellites** – The satellites use microwave radio as their telecommunications medium which are not deflected by the Earth's atmosphere. The satellites are stationed in space, typically 35,400 km (for geosynchronous satellites) above the equator. These Earth-orbiting systems are capable of receiving and relaying voice, data, and TV signals.

- **Cellular and PCS systems** – It uses several radio communications technologies. The systems are divided into different geographic areas. Each area has a low-power transmitter or radio relay antenna device to relay calls from one area to the next area.

- **Wireless LANs** – Wireless local area network uses a high-frequency radio technology similar to digital cellular and a low-frequency radio technology. Wireless LANs use spread spectrum technology to enable communication between multiple devices in a limited area.

- **Infrared communication**: It can transmit signals between devices within small distances of typically no more than 10 meters. In most cases, line-of-sight propagation is used, which limits the physical positioning of communication.

There are several common physical and logical structures, used in network design (Anon 2011). These are given below:
• **Bus** – In this all the nodes of the network are connected to a common transmission medium which has exactly two endpoints— all data that is transmitted between nodes in the network is transmitted over this common transmission medium and is able to be received by all nodes in the network virtually simultaneously. The two endpoints of the common transmission medium are normally terminated with a device called a terminator.

• **Star** – In this topology each of the nodes of the network is connected to a central node with a point-to-point link in a "hub". All data that is transmitted between nodes in the network is transmitted to this central node, which is usually some type of device that then retransmits the data to some or all of the other nodes in the network, although the central node may also be a simple common connection point without any active device to repeat the signals.

• **Ring** – In this each of the nodes of the network is connected to two other nodes in the network and with the first and last nodes being connected to each other, forming a ring – all data that is transmitted between nodes in the network travels from one node to the next node in a circular manner and data generally flows in single direction only.

• **Mesh** – The mesh topology, each of the nodes of the network is connected to each of the other nodes in the network with a point-to-point link – this makes it possible for data to be simultaneously transmitted from any single node to all of the other nodes. The physical fully connected mesh topology is generally too costly and complex for practical networks, although the topology is used when there are only a small number of nodes to be interconnected.

• **Tree (hierarchical)** – In this topology, a central "root" node (the top level of the hierarchy) is connected to one or more other nodes that are one level lower in the hierarchy (i.e., the second level) with a point-to-point link between each of the second level nodes and the top level central "root" node, while each of the second level nodes that are connected to the top level central "root" node will also have one or more other nodes that are one level lower in the hierarchy (i.e., the third level) connected to it, also with a point-to-point link, the top level central "root" node being the only node that has no other node above it in the hierarchy – the hierarchy of the tree is symmetrical, each node in the network having a specific fixed number of nodes ‘f’ connected to it at the
next lower level in the hierarchy. The number ‘f’ being referred to as the "branching factor" of the hierarchical tree.

- **Hybrid** - The hybrid topology is a type of network topology that is composed of one or more interconnections of two or more networks that are based upon different physical topologies mentioned above.

### 1.1.2 Location and Distance Based

Classification may be carried out on the basis of the geographical area covered by the network. Following are the popular network structures.

- **Personal Area Networks (PAN):** For setting such type of network the interconnection of devices should be within the range of an individual person, typically a range of 10 meters. For example, a wireless network connecting a Laptop and a Mobile or two or more mobiles through Bluetooth.

- **Local Area Networks (LAN):** These are Privately-owned networks typically connecting computers within a small geographic area, like a home, office, college laboratory, or a campus (William 2000). They are widely used to share resources like printers, scanners and exchange of information. Traditional LANs run at speeds of 10 Mbps to 100 Mbps, have low delay (microseconds or nanoseconds), and make very few errors. Newer LANs operate at a speed of 10 Gbps to 100 Gbps.

- **Metropolitan Area Networks (MAN):** Such types of networks are spread around a town or city. It is smaller than Wide Area Network (WAN), but larger than a LAN (William 2000). It covers a larger geographical area, ranging from several blocks of buildings to entire cities. MANs can also operate on communications channels of moderate-to-high data rates. A MAN might be owned and operated by a single organization, but it usually will be used by many individuals and organizations. They will often provide means for internetworking of LANs.

- **Wide Area Networks (WAN):** These connect networks over a large geographical area, such as between different buildings, towns or even countries (William 2000). The inter-processor distance can be from 100 kilometers to 10000 kilometers and the processors are located in a country or a continent.
*Internetworks:* The inter-processor distance is 10,000 kilometers or more and a popular example is the INTERNET. An Internet has intermediate systems that send data to next intermediate node/system or to the destination. An interconnected intermediate system saves cable and bandwidth. A network can use different types of channels (twisted pair, coaxial cables, fiber optics, radio waves etc.) to connect the nodes in intermediate network. Intermediate systems are known under various terms depending on the context (Kurose and Ross 2000): routers (TCP/IP, AppleTalk), switches (X.25, Frame Relay, ATM, telephone) etc.

### 1.1.3 Transmission Technology Based

Transmission technology specifies the way of exchanging the information among nodes of network. On the basis of transmission technology, networks can be classified as “Broadcast” and “Point-to-Point” (Tanenbaum 1996). As a general rule, smaller geographically localized networks tend to use broadcasting, whereas larger networks usually are point-to-point.

- **Broadcast Networks:** These have a single communication channel that is shared by all the machines on the network. Short messages, called packets in certain contexts, sent by any machine are received by all the other. An address field with the packet specifies for whom it is intended. Upon receiving a packet, intended for itself, it processes the packet; if the packet is intended for some other machine, it is just ignored. Example is an airport announcement asking all “flight number 521” passengers to report to gate 16. Broadcast systems generally also allow the possibility of addressing a packet to all destinations by using a special code in the address field. When a packet with this code is transmitted, it is received and processed by every machine on the network. This mode of operation is called broadcasting. Some broadcast system also support transmission to a subset of the machines, something known as multicasting.

- **Point-to-Point Networks:** In this type of network there are many connections between individual pairs of machines. To go from the source to the destination, a packet on this type of network may have to first visit one or more intermediate machines. Often multiple routes, of different lengths are possible, so routing algorithms play an important role in point-to-point networks for finding optimal path.
1.1.4 Forwarding/Routing Based

It specifies the methods that can be adopted for sending data from sender to receiver. On the basis of routing technique two network types are in use – Circuit switched networks and Packet switched networks.

- **Circuit Switching**: This technology is used by telephone networks (Jean-Yves et al. 2011). When computers are connected by a circuit switched network, they establish a direct data link over the circuit. Modern circuit switches are based on byte multiplexing and are thus similar to packet switches, with the main difference that they perform non-statistical multiplexing.

- **Packet Switching**: With packet switching technology, data packets from many applications can be transferred together on the same link (Kurose and Ross 2000). They are differentiated by information in headers. Packet switches/routers have queues to hold and process the incoming packets. Packet switching is the basis for all data networks today, including the Internet, and other public data networks like Frame Relay and ATM (Jean-Yves et al. 2011).

1.2 Protocols

A protocol is set of rules to govern the communication between communicating entities. It defines the format of data units, rules of operations like data delivered, retransmission required, transmission speed, packet size etc. The Protocols may be connection oriented or connectionless. In Connection oriented transmission some synchronization of all involved parties is required before communication takes place (Kurose and Ross 2000). The telephone system is connection oriented and mail or postal system is connectionless examples. Examples of protocols are Hyper Text Transfer Protocol (HTTP), File Transfer Protocol (FTP), Transmission Control Protocol (TCP), User Datagram Protocol (UDP), Internet Protocol (IP) etc. Open System Interconnection (OSI) defines the modal for a network system. International Standards Organization (ISO) developed OSI layered model for supporting communications architecture. According to OSI model, networking functions are structured as a layered model:
Layer n communicates with other layer n entities using layer n process data units (PDUs).
Layer n uses the service of layer n-1 and offers a service to layer n+1.
Every layer has a collection of protocols to use the services offered by lower layer and to provide services to upper layer and users.
Entities at the same layer are said peer entities
Layering of protocol entities is reflected by the term of a protocol stack.

The TCP/IP Architecture (William 1997), or the Internet Architecture is described by a collection of Internet standards, published in documents called RFCs (Requests for Comments). There exist other architectures (Kurose and Ross 2000), each of them having a different set of layers and names for layers.

- **Proprietary Architectures:** SNA (IBM), Decnet (Digital), AppleTalk (Apple), XNS (Xerox), UUCP (UNIX internal protocols), etc.
- **ITU Architecture:** Defines public networks for telephony, telex, fax, data networks (X.25, Frame Relay, mail and directory services) and ATM.
- **IEEE LAN Architecture:** It defines layers 1 and 2 for local area networks.

Having several architectures is a nuisance; everything would be simpler if there would be only one. Today, the TCP/IP architecture has become dominant.

### 1.3 Services Provided by Computer Networks

Computer networks allow machines and human beings to communicate, using a number of services (Jean-Yves et al. 2011). To name a few, File transfer, e-mail, remote login, remote processing, news, resource sharing (printers and servers), name service, Voice over Internet Protocol (VoIP) video-on-demand, digital library, e-commerce, distributed database etc. are the services provided by computer networks using distributed applications like web server or name server. According to Martin (Martin et al. 1999) the rapid rise in popularity of the Internet is mainly due to the World Wide Web (WWW) and e-mail systems that enable free exchanges of information. Today the growth of Internet has reached to a level where every business, university, college etc. has to get connected to it in order to stay relevant and compete. The
traditional text-based transactions have been replaced by web-based applications with multimedia contents.

1.4 Network Performance

There are two types of performance measures that should be considered for the network (Martin et al. 1999). One is the throughput requirement and the other is the response time. Both of these factors need to be considered when designing the network. The scalability of the network with respect to the performance requirements must also be considered.

1.4.1 Throughput

It defines how much data can be moved by time unit or shortest possible time. It is equal to the bit rate if there is no protocol. However, in most practical cases, the throughput is less than the bit rate for two reasons:

i. *Protocol Overhead:* Protocols like UDP use some bytes to transmit protocol information. This reduces the throughput. If you send one-byte messages with UDP, then for every byte, you create an Ethernet packet of size 53 bytes, thus the maximum throughput you could ever get at the UDP service interface if you use a 64 kb/s channel would be 1.2 kb/s.

ii. *Protocol Waiting Times:* Some protocols may force you to wait for some event.

1.4.2 Response Time

It defines how long a user must wait before a result is returned from the system. Following are the terms that effects the response time of a system:

- *Propagation Time:* It is the time taken by the front of a signal to reach destination.
- *Transmission Time:* The bit rate and the number of bits to transmit determine the transmission time.
- *Bandwidth:* The bandwidth is the width of the frequency range that can be used for transmission over the channel. It determines how much data; link can transfer per unit of
time. It is generally represented by a data rate, such as 622 megabits per second (Mbps). The bandwidth limits the maximal bit rate that can be obtained using a given channel.

- **Bit Rate**: The bit rate of a channel is the number of bits transmitted per second.

### 1.5 Routing in Computer Networks

Network routing refers to the ability of an electronic communication network to send a unit of information from one node (Source) to another node (Destination) by determining a path through the network, efficiently and quickly (Medhi and Ramasamy 2007). Routing is performed for many kinds of networks, including the telephone network, electronic data networks (such as the Internet), and transportation (transport) networks. In packet switching networks, routing directs forwarding, the transit of logically addressed packets from their source toward their ultimate destination through intermediate nodes; typically hardware devices called routers, bridges, gateways or switches (Moiz 2011). A numbers of factors needed to be governed, for determining of an efficient path.

#### 1.5.1 Network Traffic

Traffic in a network refers to packets generated by different applications, such as web or email. When many users trying to download from the same website; then, packets that are generated can possibly be queued at routers or even dropped.

#### 1.5.2 Network Addressing

A network address is a unique identifier for a node or computer on a network. A node can determine the addresses of other computers on the network and use these addresses to send messages to each other. There are basically two types of addressing schemes used in computer networks, named as physical addressing or Media Access Control (MAC) and logical addressing or Internet Protocol (IP) addressing.

*Media Access Control*: This addressing method is also known as physical addressing and generally used in LAN. MAC addresses are provided by manufacturers of network adapters, burn into their products to uniquely identify them.
**IP Addressing:** If one has to send data to any host in the Internet, there is a need to uniquely identify the hosts in the Internet. Thus, there is need of global addressing scheme in which no two hosts have the same address. The addressing in the Internet is referred to as IP addressing (Medhi and Ramasamy 2007). There are two IP packet formats: IPv4 and IPv6. In IPv4, addresses consist of four bytes (32 bits) and IPv6 address consists of sixteen bytes (128 bits) that uniquely identify all computers on the public Internet. An IP address defines two parts: In Internet terminology, they are known as the netid and the hostid, to identify a network and a host address, respectively. A netid identifies a contiguous block of addresses. A host is a general term used for indicating different entities in network; the most common ones are a desktop, laptop, web-server, an email server, or any computer we use for accessing the Internet. Generally a host is the start point or end point of communication in the Internet. The Internet’s conceptual framework, known as TCP/IP (Transmission Control Protocol/Internet Protocol), relies on a delivery model in which TCP is in charge of the reliable delivery of information, while IP is in charge of routing, using the IP addressing mechanism. A classful and classless addressing scheme can be used for finding the host on Internet.

a) **Classful Addressing:** To support different sizes for the (netid, hostid) part, a good rule on how to partition the total IP address space of $2^{32}$ addresses was needed, i.e., how many network addresses will be allowed and how many hosts each of them will support. Thus, the IP address space was originally divided into three different classes, Class A, Class B, and Class C for networks and hosts. Each class was distinguished by the first few initial bits of a 32-bit address. For readability, IP addresses are expressed as four decimal numbers, with a dot between them. This format is called the *dotted decimal notation*. The notation divides the 32-bit IP address into 4 groups of 8 bits and specifies the value of each group independently as a decimal number separated by dots. Because of 8-bit breakpoints, there can be at most 256 ($= 2^8$) decimal values in each part. Since 0 is an assignable value, no decimal values can be more than 255. Thus, an example of an IP address is 192.168.0.1 consisting of the four decimal values, separated by a dot or period.

Class A address has the first bit set to 0 and is followed by 7 bits for the network part, resulting in a maximum of 128 ($= 2^7$) networks; this is then followed by a 24-bit host
part. Thus, Class A supports a maximum of \(2^{24} - 2\) hosts per network. This calculation subtracts 2 because 0s and 1s in the host part of a Class A address may not be assigned to individual hosts; rather, all 0s that follow a netid such as 10.0.0.0 identify the network, while all 1s that follow a netid such as 10.255.255.255 are used as the broadcast address for this network. Each Class B network address has the first two bits set to “10,” followed by a 14-bit network part, which is then followed by a 16-bit host part. A maximum of \(2^{14}\) networks can be defined with up to \(2^{16} - 2\) hosts per network. Finally, a Class C network address has the first three bits set as “110” and followed by a 21-bit network part, with the last 8 bits to identify the host part. Class C provides support for a maximum of \(2^{21} (= 2,097,152)\) networks with up to 254 \((2^8 - 2)\) hosts. In each class, a set of network addresses is reserved for a variety of purposes. Three address classes discussed so far are used for unicasting in the Internet, that is, for a host-to-host communication. There is another class of IP addresses, known as Class D addressing, that is used for multicastring in the Internet; in this case, the first four bits of the 32-bit address are set to “1110” to indicate that it is a multicast address. A host can use a multicast address as the destination address for a packet generated to indicate that the packet is meant for any hosts on the Internet.

The original rationale behind classes of different sizes was to provide the flexibility to support different sized networks, with each network containing a different number of hosts. Thus, the total address length can still be kept fixed at 32 bits, an advantage from the point of view of efficient address processing at a router or a host. As the popularity of the Internet grew, several disadvantages of the addressing scheme came to light. The major concerns were the rate at which the IP address blocks that identify netids were being exhausted, especially when it was necessary to start assigning Class C level netids. Recall from our earlier discussion that IP netids are non-geographic; thus, all valid netids are required to be listed at the core routers of the Internet along with the outgoing link, so that packets can be forwarded properly. Imagine all Class C level netids being assigned, then there are over 2 million entries that would need to be listed at a core router; no current routers can handle this number of entries without
severely slowing packet processing. This issue, first recognized in the early 1990s, led to the development of the concept of *classless* addressing. In order to understand this concept, Understanding of subnetting/netmask is required.

To understand subnetting/netmask consider the IP address 192.168.40.3 that is part of Class C network 192.168.40.0. A subnet or sub-network is defined through a network mask boundary using the specified number of significant bits as 1s (Medhi and Ramasamy 2007). Since Class C defines networks with a 24-bit boundary, therefore the most significant 24 bits are 1s, and the lower 8 bits are 0s. This translates to the dotted decimal notation 255.255.255.0, which is also compactly written as “/24” to indicate how many most significant bits are 1s. A bit-wise logical “AND” operation can be performed between the host address and the netmask to obtain the Class C network address as shown below:

\[
\begin{align*}
11000000 & \quad 10101000 \quad 00101000 \quad 00000011 \rightarrow 192.168.40.3 \\
\text{AND} & \quad 11111111 \quad 11111111 \quad 11111111 \quad 00000000 \rightarrow \text{netmask (/24)} \\
11000000 & \quad 10101000 \quad 00101000 \quad 00000000 \rightarrow 192.168.40.0
\end{align*}
\]

Here both the host address and the netmask have 1s in the first two positions from the left; thus, the “AND” operation results in 1s for these two positions. For the third position from left, the host has 0 while the netmask has 1; thus, the result of the “AND” operation is zero; and so on. Note that for network addresses such as Class C address, the netmask is implicit and it is on a /24 subnet boundary. Now consider that we want to change the netmask explicitly to /21 to identify a network larger than a 24-bit subnet boundary. If a bit-wise operation is performed here:

\[
\begin{align*}
11000000 & \quad 10101000 \quad 00101000 \quad 00000011 \rightarrow 192.168.40.3 \\
\text{AND} & \quad 11111111 \quad 11111111 \quad 11110000 \quad 00000000 \rightarrow \text{netmask (/21)} \\
11000000 & \quad 10101000 \quad 00101000 \quad 00000000 \rightarrow 192.168.40.0
\end{align*}
\]
The network address is again 192.168.40.0. However, in the later case, the network boundary is 21 bits. In the second example “192.168.40.0/21”, the first part is the netid and the second part is the mask boundary indicator. In this notation, the original Class C address can be written as 192.168.40.0/24 and thus, there is no ambiguity with 192.168.40.0/21.

b) Classless Interdomain Routing: Classless Interdomain Routing (CIDR) uses an explicit netmask with an IPv4 address block to identify a network, such as 192.168.40.0/21. An advantage of explicit masking is that an address block can be assigned at any bit boundaries, be it /15 or /20; most important, the assignment of Class C level addresses for networks that can show up in the global routing table can be avoided or minimized. For example, a contiguous address block can be assigned at the /21 boundary which can be thought of as an aggregation of subnets at the /24 boundary. Because of this, the term super-netting or variable-length subnet masking (VLSM) is also used in reference to the explicit announcement of the netmask. Through such a process, and because of address block assignment at boundaries such as /21, the routing table growth at core routers can be delayed. In the above example, only the netid 192.168.40.0/21 needs to be listed in the routing table entry, instead of listing eight entries from 192.168.40.0/24 to 192.168.47.0/24. Thus, it shows how the routing table growth can be curtailed.

1.5.3 Purpose of Routers in Routing

A router is an intermediate node in the network which runs the routing protocol for efficient forwarding of packets to the destination. The major functions carried out by router are as:

i. A router (Intermediate node) functions are to read the destination address marked in an incoming IP packet, to consult its internal information to identify an outgoing link to which the packet is to be forwarded, and then to forward the packet. A router also maintains a finite amount of space, known as a buffer, to temporarily store backlogged packets. From an efficient delivery point of view, it is desirable not to have any packet loss (or at least, minimize it) during transit. This is because the reliable delivery notion
works on the principle of retransmission and acknowledgment and any drop would mean an increase in delay due to the need for retransmission. In addition, during transit, it is also possible that the content enclosed in a data packet is possibly corrupted due to, for example, an electrical signaling problem on a communication link. This then results in garbling of a packet. From an end-to-end communication point of view, a garbled packet is the same as a lost packet. Thus, for efficient delivery of packets, there are several key factors to consider (Medhi and Ramasamy 2007): (1) routers with a reasonable amount of buffer space, (2) links with adequate bandwidth, (3) actual transmission with minimal error (to minimize packets being garbled), and (4) the routers’ efficiency in switching a packet to the appropriate outgoing link.

ii. For efficient processing of packets a routers maintains a table in the router, known as the routing table, that contains the identifiers for the next router, known as the next hop, for a given destination netid. The router updates the table periodically. In order to update such a table, the router would need to store all netids it has learned about so far; second, if a link downstream is down or congested or a netid is not reachable for some reason, it needs to know so that an alternate path can be determined as soon as possible. This means that a mechanism is required for communicating congestion or a failure of a link or non-reachability of a netid. This mechanism is known as the routing protocol mechanism. The information learned through a routing protocol is used for generating the routing table ahead of time. If new information is learned about the status of links or nodes, or the reachability of a netid through a routing protocol, a routing algorithm is then invoked at a router to determine the best possible next hop for each destination netid in order to update the routing table. For efficient packet processing, another table, known as the forwarding table, is derived from the routing table that identifies the outgoing link interfaces. The forwarding table is also known as the Forwarding Information Base (FIB). It should be noted that a routing algorithm may need to take into account one or more factors about a link, such as the delay incurred to traverse the link, or its available bandwidth, or path reliability in order to determine the best possible path among a number of possible paths. If a link along a path does not have adequate bandwidth, congestion or delay might occur. To minimize delay, an important function, called traffic engineering, is performed. Traffic engineering is concerned with ways to improve the
operational performance of a network and identifies procedures or controls to be put in place ahead of time to obtain good network performance.

1.5.4 Architectures

There is another important term associated with networking in general and network routing in particular, labeled as architecture. From a router’s perspective, architecting a network refers to how it is organized internally for a variety of functions, from routing protocol handling to packet processing. From a network perspective, this means how the network topology architecture should be organized, where routers are to be located and bandwidth of links determined for efficient traffic engineering, and so on. Architectures cover many different aspects of networking environments, so routing must account for each of the following architectural components. Some aspects of the architectures listed below are critical to routing issues:

i. **Service Architecture:** A service model gives the basic framework for the type of services a network offers.

ii. **Protocol Stack Architecture:** It defines how service delivery may require different functions to be divided along well-defined boundaries so that responsibilities can be decoupled. It does not describe how actual resources might be used or needed.

iii. **Router Architecture:** A router is a specialized computer that is equipped with hardware/software for packet processing. It is also equipped for processing of routing protocols and can handle configuration requirements. A router is architected differently depending on its role in a network, such as a core router or an edge router, although all routers have a common set of requirements.

iv. **Network Topology Architecture:** For efficient operation as well as to provide acceptable service to its users, a network is required to be organized based on a network topology architecture that is scalable and allows growth. In order to address efficient services, there is also a direct connection among the topology architecture, traffic engineering, and routing.

v. **Network Management Architecture:** A network needs to provide several additional functions in addition to carrying the user traffic from point A to point B; for clarity, the user data traffic forwarding is considered as the “data plane”. From an operational point
of view, a “management plane” handles the configuration responsibility of a network, and a “control plane” addresses routing information exchanges.

1.5.4.1 Service Architecture

An important aspect of a networking architecture is its service architecture. Every networking environment has service architecture, much like the postal delivery system. Service models associated with IP networks are discussed here.

1.5.4.1.1 Best Effort Service Architecture

The basic information unit of an IP network is a packet or a datagram which is forwarded from one router to another towards the destination using a switching concept, referred as “packet switching”. This means that a router makes decisions by identifying an outgoing link on a packet-by-packet basis instantaneously after the packet arrives. At the conceptual level, it is assumed that no two packets are related, even though they might arrive one after another and possibly for the same web-page downloaded. Also, recall that at the IP level, the packet forwarding function is provided without worrying about reliable delivery, because TCP handles reliability issues; in this sense, IP makes its best effort to deliver packets. Because of this, the IP service paradigm is referred to as the “best-effort service”.

1.5.4.1.2 Integrated Service Architecture

The best-effort service model was initially developed for the reliable delivery of data services, since it was envisioned that services would be data-oriented services that can tolerate delay, but not loss of packets. This model worked because the data rate provided during a session can be adaptive. The concept for integrated services (“int-serv”) architecture was developed in the early 1990s to allow functionalities for services that are real-time, interactive, and that can tolerate some loss, but require a bound on the delay. Furthermore, each session or connection requires a well-defined bandwidth guarantee and a dedicated path. For example, interactive voice and multimedia applications fall into this category. Note that the basic best-effort IP framework works on the notion of statelessness; that two consecutive packets that belong to the same connection are to be treated independently by a router. Yet, for services in the integrated services
architecture that require a connection or a session for a certain duration of time, it became necessary to provide a mechanism to indicate the longevity of the session, and the ability for routers to know that resources are to be reserved for the entire duration. Since the basic IP architecture works on the notion of statelessness, and it was infeasible to completely change the basic IP service architecture, a soft-state concept was introduced to handle integrated-services. To do that, a session setup and maintenance protocol was also developed known as the resource reservation protocol (RSVP) that can be used by each service. The basic idea was that once a session is established, RSVP messages are periodically generated to indicate that the session is alive. The idea of integrated services was a novel concept that relies on the soft-state approach. A basic problem is the scalability of handling the number of RSVP messages generated for all sessions that might be simultaneously active at a router or a link.

1.5.4.1.3 Differentiated Service Architecture

The differentiated services ("diff-serv") architecture was developed to provide prioritized service mechanisms without requiring connection-level information to be maintained at routers. Specifically, this approach gives priority to services by marking IP packets with diff-serv code points located in the IP header. Routers along the way then check the diff-serv code point and prioritize packet processing and forwarding for different classes of services. Second, this model does not require the soft-state concept and thus avoids the connection-level scalability issue faced with RSVP. Diff-serv code points are identified through a 6-bit field in the IPv4 packet header; in the IPv6 packet header, the traffic class field is used for the same purpose.

1.5.4.1.4 Supplementing a Service Architecture

Earlier in this section, the best-effort service model was introduced. In a realistic sense, and to provide acceptable quality of service performance, the basic concept can be supplemented with additional mechanisms to provide acceptable service architecture, while functionally it may still remain as the best-effort service architecture. For example, although the basic conceptual framework does not require it, a router can be designed to do efficient packet processing for packets that belong to the same web-page requested by a user since they are going to the same destination. That is, a sequence of packets that belongs to the same pair of origination and
destination IP addresses, to the same pair of source and destination port numbers, and to the same transport protocol (either TCP or UDP) can be thought of as a single entity and is identified as a micro-flow. Thus, packets belonging to a particular micro-flow can be treated in the same manner by a router once a decision on forwarding is determined based on the first packet for this micro-flow. Another way to fine-tune the best-effort service architecture is through traffic engineering. That is, a network must have enough bandwidth so that delay or backlog can be minimal; routers must have adequate buffer space, and so on, so that traffic moves efficiently through the network. In fact, both packets process at a router and traffic engineering work in tandem for providing efficient services. Similarly, for both integrated-services and differentiated-service architecture, packet handling can be optimized at a router. Furthermore, traffic engineering can be geared for integrated services or differentiated services architectures.

1.5.4.2 Protocol Stack Architecture

Another important facet of a networking environment is the protocol stack architecture. First OSI (Open Systems Interconnections) reference model is discussed followed by the IP protocol stack architecture and its relation to the OSI reference model.

1.5.4.2.1 OSI Reference Model

The OSI reference model was developed in the 1980s to present a general reference model for how computer network architecture should be functionally divided. As part of OSI, many protocols have also been developed. Here, the basic reference model is presented.

The OSI reference model uses a seven layered hierarchy to separate functions, where the layering is strictly enforced. That is to say that an \( N \)-layer uses services provided by layer \( N - 1 \); it cannot receive services directly from layer \( N - 2 \). In the OSI model, seven-layer architecture is defined. The seven layers are also referenced by layer numbering counting from bottom up. From a functional point of view, layer 1 provides physical layer functions. Layer 2 provides the data link function between two directly connected entities. Layer 3 is the network layer, where addressing and routing occurs. Layer 4 is the transport layer that can provide either reliable or unreliable transport services, with or without defining a connection (“connection-oriented” or
“connection-less”). Layer 5 is the session layer, addressing communication that may transcend multiple connections. Layer 6 is the presentation layer that addresses structured information and data representation. Layer 7 is where the application layer protocols are defined. Every computer networking environment do not strictly adhere to the OSI reference model, it does provide an easy and simple way to check and compare what a particular networking environment might have to consider.

1.5.4.2.2 TCP/IP Protocol Stack Architecture

The IP architectural model can be classified into the following layers: the network interface, the IP layer, the transport layer, and the application layer. It can be easily observed that it does not exactly map into the seven-layer OSI reference model. Actual applications are considered on the top of the application layer, although the IP model does not strictly follow layering boundaries as in the OSI reference model. For example, it allows an application to be built without using a transport layer; ping is such an example. It is discussed earlier that IP includes both the destination and the source address – this is accomplished through a header part in the IP packet that also contains additional information. The IP model does not explicitly declare how the layer below the IP layer needs to be; this part is simply referred to as the network interface that can support IP. Figure 1.1 shows the major protocols of TCP/IP architecture.

![Figure 1.1: Protocol Layering in IP Architecture](image)

1.5.4.3 Router Architecture

A router provides several important functions in order to ensure proper packet forwarding, and to do so in an efficient manner. A router is a specialized computer that handles three primary
functions:

i. **Packet Forwarding:** On receiving an incoming packet, a router checks whether the packet is error free. After inspecting the header of a packet for destination address, it performs a table lookup function to determine how to find the appropriate outgoing link.

ii. **Routing Protocol Message Processing:** A router also needs to handle routing protocol packets and determine if any changes are needed in the routing table by invoking a routing algorithm, when and if needed.

iii. **Specialized Services:** In addition, a router is required to handle specialized services that can aid in monitoring and managing a network.

1.5.4.4 **Network Topology Architecture**

The network topology architecture encompasses how a network is to be architected in an operational environment while accounting for future growth. What does topology mean? It refers to the form a network will adopt, such as a star, ring, or a fully mesh topology, or a combination of them. The topological architecture then covers architecting a network topology that factors in economic issues, different technological capabilities, and limitations of devices to carry a certain volume of expected traffic and types of traffic, for an operational environment. Certainly, network topology architecture also needs to take into account routing capability, including any limitation or flexibility provided by a routing protocol. It is up to a network provider, also referred to as a network operator or a service provider, to determine the best topological architecture for the network. It is important to note that the operational experience of an existing network can contribute to the identification of additional features required from a routing protocol, or the development of a new routing protocol, or the development of a new routing algorithm or modification of an existing algorithm.

1.5.4.5 **Network Management Architecture**

Routing information exchange uses the same framework as the user data traffic in the Internet. For an operational network, it is important to have a network management architecture where various functions can be divided into “planes.” Here it is divided into three different planes: the
management plane, the control plane, and the data plane.

i. The management plane, addresses router configuration and collection of various statistics, such as packet throughput, on a link. Router configuration refers to configuration of a router in a network by assigning an IP address, identifying links to its adjacent routers, invoking one or more routing protocols for operational usage, and so on. Statistics collection may be done, for example, through a protocol known as Simple Network Management Protocol (SNMP). The management plane of a router is closely associated with network operations.

ii. The control plane exchanges control information between routers for management of a variety of functions, such as setting up a virtual link. The control plane is also involved in identifying the path to be taken between the endpoints of this virtual link, which relies on the routing information exchange. Another clarification is important to point out. Since these functions are different, the routing-related functions are in the control plane.

iii. The data transfers, such as the web or email, are in the data plane. All these planes use IP for communication, so at the IP layer; there is no distinction between these functional planes.

For efficient traffic engineering of a network, certain information is also required from different routers. Such information exchanges can be conducted either through the control plane or through the management plane. In certain networking environments, some functions can overlap across different planes.

1.5.5 Routing Classification

Routing can be divided into two broad categories, Static and Dynamic routing. Static routing is useful for small networks only. As the network grows, overheads increases, which are not possible to handle manually and dynamic routing comes into picture. In today’s networks dynamic routing is implemented.

1.5.5.1 Static Routing

In static routing information about all networks must be entered manually by the administrator
on all routers. Its Default Administrative Distance is 1. As routing data is entered by the administrator there is no requirement to perform any metric calculations to determine paths. Routes are fixed at boot time and useful only for simple cases. Following are the advantages and disadvantages:

**Advantages:**

- Doesn't exchange any routing tables hence it doesn’t consume any bandwidth.
- Again the bandwidths which may be consumed in dynamic updates are saved.
- Low processing power may suffice; hence we can use cheaper routers.
- Higher Security as there is no information being regarding routing on the connecting links.

**Disadvantages:**

- Remote network information is written manually.
- Reconfiguration is problematic and tedious if there is a change in the topology.
- Network administrator tells the router to send traffic with a destination IP address, to a router with an IP address of z.z.z.z. It is possible for small networks with very few routes, but it can become very cumbersome as the network grows. To keep a large network fully connected via static routes, you need to create a route on every router for every other router. This is unmanageable and unworkable for big networks.

1.5.5.2 **Dynamic Routing**

Dynamic routing is necessary in large internets. The router dynamically learns routes after an administrator configures a routing protocol that helps determine the routes. On configuration of router for dynamic routing, routing process automatically updates routes whenever new topology information is received. The router learns and maintains routes to the remote destinations by exchanging routing updates with other routers. Routing table is initialized at boot time. A routing protocol defines the rules that are used by a router when it communicates with the neighboring routers. Dynamic routing relies on a routing protocol to disseminate knowledge. Following are the information that routing protocols describe:
How updates are conveyed
What knowledge is conveyed?
When to convey knowledge
How to locate recipients of the updates

The advantages and disadvantages of dynamic routing are as given below:

**Advantages:**

- The main advantages of dynamic routing over static routing are scalability and adaptability. A dynamically routed network can grow more quickly and larger without human involvement, and is able to adapt to changes in the network topology brought about by the growth or by the failure of one or more network components.
- The problems of static routing are solved.

**Disadvantages:**

- Dynamic routing requires more resources (memory, bandwidth and processing power) and complex routing protocol for finding optimal paths and managing other issues like packet handling etc.

1.6 Dynamic Routing Protocols

A dynamic routing protocol is a set of processes, algorithms, and messages that are used to exchange routing information and populate the routing table with best paths chosen by the routing protocol. The purpose of a routing protocol includes (Rick and Allan 2007):

- Discovering remote networks
- Maintaining up-to-date routing information
- Choosing the best path to destination networks
- Having the ability to find a new best path if the current path is no longer available

There are basically three components that define a dynamic routing protocol:
i. **Data structures:** Some routing protocols use tables or databases for their operations. The data structure is kept in RAM for processing.

ii. **Algorithm:** It is a finite list of steps used to complete a task. Routing protocols use algorithms for processing routing information and for best-path determination.

iii. **Routing protocol messages:** Routing protocols use various types of messages to discover neighboring routers, exchange routing information, and do other tasks to learn and maintain accurate information about the network.

To cope up the needs and maintain good performance the routing protocols need continuous optimization in terms of reducing the size of routing table and better routing algorithm.

### 1.6.1 Evolution of Dynamic Routing Protocols

Dynamic routing protocols have been used in networks since the early 1980s. An overview of routing protocols evolution and IP routing strategies is given below in figure 1.2 and figure 1.3 respectively.

![Protocols Evolution](image)

**Figure 1.2:** Protocols Evolution

![Unicast Routing Protocols](image)

**Figure 1.3:** Unicast Routing Protocols
The first version of Routing Information Protocol (RIP) was released in 1982, but some of the basic algorithms within the protocol were used on the ARPANET as early as 1969. As networks have evolved and become more complex, new routing protocols have emerged (Rick and Allan 2007, Dirk 2000).

1.6.2 Characteristics based Classification of Dynamic Routing Protocols

The classification is carried out on basis of the domain where protocol can be implemented, the algorithm it uses and the addressing scheme it uses to locate a host or entity in network. Table 1.1 shows the classification of routing protocols.

- Interior Gateway Protocols (IGP) or Exterior Gateway Protocols (EGP)
- Distance vector or Link-state
- Classful or Classless

**Table 1.1: Classification of Routing Protocols**

<table>
<thead>
<tr>
<th></th>
<th>Distance Vector</th>
<th>Link State</th>
<th>Path Vector</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Classful</strong></td>
<td>RIP, IGRP</td>
<td>---</td>
<td>EGP</td>
</tr>
<tr>
<td><strong>Classless</strong></td>
<td>RIP v2, EIGRP</td>
<td>OSPFv2, IS-IS</td>
<td>BGPv4</td>
</tr>
<tr>
<td><strong>IPv6</strong></td>
<td>RIPng, EIGRP for IPv6</td>
<td>OSPFv3, IS-IS for IPv6</td>
<td>BGPv4 for IPv6</td>
</tr>
</tbody>
</table>

1.6.2.1 Interior Gateway Protocols and Exterior Gateway Protocols

An autonomous system (AS) is a collection of routers under a common administration (Rick and Allan 2007). Typical examples are a company’s internal network and an ISP’s network. Because the Internet is based on the autonomous system concept, two types of routing protocols are required: interior and exterior routing protocols.

1.6.2.1.1 Interior Gateway Protocols (IGP)

These are used for intra-autonomous system routing, that is, routing inside an autonomous system. IGPs are used for routing within a routing domain, those networks within the control of a
single organization. An autonomous system is commonly composed of many individual networks belonging to companies, schools, and other institutions. An IGP is used to route within the autonomous system and also used to route within the individual networks themselves. IGPs for IP include RIP, IGRP, EIGRP, OSPF, and IS-IS. Routing protocols (the algorithm used by that routing protocol) use a metric to determine the best path to a network. The metric used by RIP is hop count, which is the number of routers that a packet must traverse in reaching another network. OSPF uses bandwidth metric to determine the shortest path.

a) *RIP V1:* It’s an open standard protocol. It can be used by routers of all vendors. The administrative distance (AD) is 120. It’s a classful routing protocol which means it follows an older system of addressing. It works by DVA. It can route a packet up to maximum 15 hops. It makes use of Split Horizon and Poison Reverse.

b) *RIP V2:* RIP V2 is an update of RIP V1. It can route up to 255 hops, by default 100. It supports authentication and auto-summarization. Authentication is used for secure routing and auto-summarization reduces traffic. It’s a classless routing protocol. It works on an advanced version of DVA. It exchanges its routing table using multicasting.

c) *Interior Gateway Protocol (IGRP):* It’s propriety of Cisco. It can be used only by Cisco routers. It’s a classful routing protocol. It works on DVA. It can route up to 255 hops (by default 100). Its AD is 100. IGRP uses autonomous systems.

d) *Enhanced IGRP (EIGRP):* It’s a classless routing protocol. The AD is 90. It’s a hybrid routing protocol. It uses characteristics of both DVA and Link State Algorithm. Metric calculation for EIGRP includes Bandwidth, Delay, Load, Reliability and Max. Transferable Unit (MTU) and is done by using the DUAL (Diffusing update dual algorithm) algorithm. It has Very fast convergence time. It makes use of Reliable Transport Protocol to stop loops. It multicasts updates to all neighbors in every 5 seconds. It supports auto-summarization.

e) *Open Shortest Path First (OSPF):* It’s an open standard routing protocol. It’s an internet standard routing protocol (Moiz 2011). It’s a classless routing protocol. The AD is 110. It works on Link State Algorithm. Multicast is made using 224.0.0.5 for all OSPF routers and 224.0.0.6 for OSPF Designated Router. It makes a Routing Table, Database Table and Neighbor ship table. It makes use of process numbers in its configuration. The
process ID is a unique arbitrary no. that is selected to identify the routing process. It has local significance. It is not important for 2 neighbors to have same process no., However Area No. must match. Area no. range is from 0 to 4.2 billion. Area 0 is known as backbone area and it is used for communication between multiple areas. There are around 40 to 80 routers in an area. The associated terms are:

A. **Benefits of Area Numbers**

- Smaller routing Table
- Lesser Processing Power is required
- Lesser DRAM is required
- Fewer problems in troubleshooting

B. **Designated Router (DR):** It is used to reduce adjacencies or neighbor ship. There is only one DR and BDR in an area. Only DR is assigned a neighbor ship table.

1.6.2.1.2 **Exterior Gateway Protocols (EGP)**

These are used for inter-autonomous system routing, that is, routing between autonomous systems. These are designed for use between different autonomous systems that are under the control of different administrations. Border gateway protocol (BGP) is the only currently viable EGP and is the routing protocol used by the Internet. BGP is a path vector protocol that can use many different attributes to measure routes. At the ISP level, there are often more important issues than just choosing the fastest path. BGP is typically used between ISPs and sometimes between a company and an ISP.

1.6.2.2 **Distance Vector and Link-State Routing Protocols**

Interior gateway protocols can be classified as two types:

 a) Distance vector routing protocols
 b) Link-state routing protocols
a) Distance Vector Routing Protocol: Distance vector means that routes are advertised as vectors of distance and direction. Distance is defined in terms of a metric such as hop count, and direction is simply the next hop router or exit interface. Distance vector protocols typically use the Bellman-Ford algorithm for the best-path route determination. Some distance vector protocols periodically send complete routing tables to all connected neighbors. In large networks, these routing updates can become enormous, causing significant traffic on the links. Although the Bellman-Ford algorithm eventually accumulates enough knowledge to maintain a database of reachable networks, the algorithm does not allow a router to know the exact topology of an internetwork. The router only knows the routing information received from its neighbors. Distance vector protocols use routers as signposts along the path to the final destination. The only information a router knows about a remote network is the distance or metric to reach that network and which path or interface to use to get there. Distance vector routing protocols do not have an actual map of the network topology. To perform Metric calculation Router will first see the administrative distance of routing protocol and route the packet towards the destination with lower AD. Administrative distance is an integer from 0 to 255. Then the hop count shall be checked and the packet shall be routed to destination with lower hop count If the hop count is same then load balancing will occur. A routing protocol with a lower AD is more trustworthy than one with higher AD. Distance vector protocols work best in situations where

- The network is simple and flat and does not require a hierarchical design.
- The administrators do not have enough knowledge to configure and troubleshoot link-state protocols.
- Specific types of networks, such as hub-and-spoke networks, are being implemented.
- Worst-case convergence times in a network are not a concern.

b) Link-State Protocol

In contrast to distance vector routing protocol operation, a router configured with a link-state routing protocol can create a “complete view,” or topology, of the network by gathering information from all the other routers. Think of using a link-state routing
protocol as having a complete map of the network topology. The signposts along the way from source to destination are not necessary, because all link-state routers are using an identical “map” of the network. A link-state router uses the link-state information to create a topology map and to select the best path to all destination networks in the topology. With some distance vector routing protocols, routers send periodic updates of their routing information to their neighbors. Link-state routing protocols do not use periodic updates. After the network has converged, a link-state update is only sent when there is a change in the topology. It exchanges its complete table with neighbor router once a time, after that it will send only hello packets. Whenever change occurs in a topology then routing will send partial or triggered update. It creates 3 tables. First neighbor ship table is formed. All paths for communication with each network are kept in a topology table. The best paths to reach the destination network are stored in the routing table. Link-state protocols work best in situations where

- The network design is hierarchical, usually occurring in large networks.
- The administrators have a good knowledge of the implemented link-state routing protocol.
- Fast convergence of the network is crucial.

### 1.6.2.3 Classful and Classless Routing Protocols

All routing protocols can also be classified as either

a) Classful routing protocols or
b) Classless routing protocols

a) **Classful Routing Protocols:** Classful routing protocols do not send subnet mask information in routing updates. The first routing protocols, such as RIP, were classful. This was at a time when network addresses were allocated based on classes: Class A, B, or C. A routing protocol did not need to include the subnet mask in the routing update because the network mask could be determined based on the first octet of the network address. Classful routing protocols can still be used in some of today’s networks, but because they do not include the subnet mask, they cannot be used in all situations.
Classful routing protocols cannot be used when a network is subnetted using more than one subnet mask. In other words, classful routing protocols do not support variable-length subnet masks (VLSM). There are other limitations to classful routing protocols, including their inability to support discontiguous networks. Classful routing protocols include RIPv1 and IGRP.

b) **Classless Routing Protocols:** Classless routing protocols include the subnet mask with the network address in routing updates. Today’s networks are no longer allocated based on classes, and the subnet mask cannot be determined by the value of the first octet. Classless routing protocols are required in most networks today because of their support for VLSM, discontiguous networks, and other features. Classless routing protocols are RIPv2, EIGRP, OSPF, IS-IS, and BGP.

All Intra-domain and Inter-domain protocols (Cisco 2011) are summarized as shown in table 1.2 and table 1.3:

**Table 1.2: Characteristics of IP Routing Protocols**

<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>Proprietary</th>
<th>Function</th>
<th>Updates</th>
<th>Metric</th>
<th>VLSM</th>
<th>Summarization</th>
</tr>
</thead>
<tbody>
<tr>
<td>RIP</td>
<td>DV</td>
<td>NO</td>
<td>Interior</td>
<td>30 sec</td>
<td>Hops</td>
<td>No</td>
<td>Auto</td>
</tr>
<tr>
<td>RIPV2</td>
<td>DV</td>
<td>NO</td>
<td>Interior</td>
<td>30 sec</td>
<td>Hops</td>
<td>YES</td>
<td>Auto</td>
</tr>
<tr>
<td>IGRP</td>
<td>DV</td>
<td>YES</td>
<td>Interior</td>
<td>90 sec</td>
<td>Compound</td>
<td>NO</td>
<td>Auto</td>
</tr>
<tr>
<td>EIGRP</td>
<td>ADV</td>
<td>YES</td>
<td>Interior</td>
<td>Trig</td>
<td>Compound</td>
<td>YES</td>
<td>Both</td>
</tr>
<tr>
<td>OSPF</td>
<td>LS</td>
<td>NO</td>
<td>Interior</td>
<td>Trig</td>
<td>Cost</td>
<td>YES</td>
<td>Manual</td>
</tr>
<tr>
<td>IS-IS</td>
<td>LS</td>
<td>NO</td>
<td>Interior/Exterior</td>
<td>Trig</td>
<td>Cost</td>
<td>YES</td>
<td>Auto</td>
</tr>
<tr>
<td>BGP</td>
<td>Path vector</td>
<td>NO</td>
<td>Exterior</td>
<td>Incr</td>
<td>N/A</td>
<td>Yes</td>
<td>Auto</td>
</tr>
</tbody>
</table>
### Table 1.3: Comparison of Routing Protocols

<table>
<thead>
<tr>
<th></th>
<th>Link State</th>
<th>Traditional Distance Vector</th>
<th>Advanced Distance Vector</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scalability</td>
<td>Good</td>
<td>Low</td>
<td>Excellent</td>
</tr>
<tr>
<td>Bandwidth</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>CPU</td>
<td>High</td>
<td>Low</td>
<td>Moderate</td>
</tr>
<tr>
<td>Memory</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Convergence</td>
<td>Fast</td>
<td>Slow</td>
<td>Fast</td>
</tr>
<tr>
<td>Configuration</td>
<td>Moderate</td>
<td>Easy</td>
<td>Easy</td>
</tr>
</tbody>
</table>

#### 1.6.3 Operation of Dynamic Routing Protocol

The general strategy of all routing algorithms is same: to learn about remote networks and to quickly adapt whenever there is a change in the topology. To accomplish this, routing protocol depends upon its operational characteristics and the algorithm it uses. The operations of a dynamic routing protocol vary depending on the type of routing protocol and the specific operations of that routing protocol. In general, the operations of a dynamic routing protocol can be described as follows (Rick and Allan 2007):

- The router sends and receives routing messages on its interfaces.
- The router shares routing messages and routing information with other routers that are using the same routing protocol.
- Routers exchange routing information to learn about remote networks.
- When a router detects a topology change, the routing protocol can advertise this change to other routers.

#### 1.6.4 Routing Metrics in Dynamic Routing Protocols

Routing protocols use metrics to determine which route is the best path. There are cases when a routing protocol learns of more than one route to the same destination. To select the best path, the routing protocol must be able to evaluate and differentiate among the available paths. For this purpose, a metric is used. A metric is a value used by routing protocols to assign costs to reach
remote networks. The metric is used to determine which path is most preferable when there are multiple paths to the same remote network. Each routing protocol calculates best route in a different way. For example, RIP uses hop count, EIGRP uses a combination of bandwidth and delay, and the Cisco implementation of OSPF uses bandwidth. Hop count is the easiest metric to envision. The hop count refers to the number of routers a packet must cross to reach the destination network.

1.6.4.1 Metric Parameters

Metrics used in IP routing protocols include the following:

- **Hop count**: A simple metric that counts the number of routers a packet must traverse.
- **Bandwidth**: Influences path selection by preferring the path with the highest bandwidth.
- **Load**: Considers the traffic utilization of a certain link.
- **Delay**: Considers the time a packet takes to traverse a path.
- **Reliability**: Assesses the probability of a link failure, calculated from the interface error count or previous link failures.
- **Cost**: generally a value is determined by network administrator to indicate preference for a route. Cost can represent a metric, a combination of metrics, or a policy.

1.6.4.2 Metric Field in the Routing Table

The routing table displays the metric for each dynamic and static route. Static routes always have a metric of 0. The metric for each intra-domain dynamic routing protocol are as:

- **RIP**: Hop count: Best path is chosen by the route with the lowest hop count.
- **IGRP and EIGRP**: Bandwidth, delay, reliability, and load: Best path is chosen by the route with the smallest composite metric value calculated from these multiple parameters. By default, only bandwidth and delay are used.
- **IS-IS and OSPF**: Cost: Best path is chosen by the route with the lowest cost. The Cisco implementation of OSPF uses bandwidth to determine the cost.
The metric associated with a certain route can be best viewed using the show ip route command. All the routing protocols discussed are capable of automatically load balancing traffic for up to four equal-cost routes by default. EIGRP is also capable of load balancing across unequal-cost paths.

1.6.5 Shortest Path Selection Algorithms

In literature the number of shortest path algorithms which have been developed and published. The new variants are still appearing. The best known algorithms such as Dijkstra (Dijkstra 1995) and the Bellman-Ford algorithm (Ford and Fulkerson 1962) are implemented in routing and run in low order polynomial time. The standard Dijkstra algorithm, for example, runs in time $O(n^2) + O(m)$ where $n$ is the number of nodes in the network and $m$ is the number of links in the network. Although the algorithms are low-order polynomial several difficulties arise when applying them to a running network (Willmott and Flatings 1999).

Dijkstra shortest path Algorithm

The algorithm does not work if some weights are negative. The nodes implementing this algorithm have complete information regarding network topology and link costs. This is accomplished via “link state broadcast”. All nodes have same info and computes least cost paths from one node (“source”) to all other nodes for preparing forwarding table for that node. After $k$ iterations, know least cost path to $k$ destinations.

Notation:

$c(x, y)$: link cost from node $x$ to $y$; $\infty$ if not direct neighbors

$D(v)$: current value of cost of path from source to destination $v$

$p(v)$: predecessor node along path from source to $v$

$N'$: set of nodes whose least cost path definitively known
Algorithm:

1. Initialization:
2. \( N' = \{u \} \)
3. For all nodes \( v \)
4. If \( v \) adjacent to \( u \)
5. Then \( D(v) = c(u, v) \)
6. Else \( D(v) = \infty \)
7. Loop
8. Find \( w \) not in \( N' \) such that \( D(w) \) is a minimum
9. Add \( w \) to \( N' \)
10. Update \( D(v) \) for all \( v \) adjacent to \( w \) and not in \( N' \):
11. \( D(v) = \min(D(v), D(w) + c(w, v)) \) /*new cost to \( v \) is either old cost to \( v \) or known Shortest path cost to \( w \) plus cost from \( w \) to \( v \)*/
12. until all nodes in \( N' \)

1.6.6 Convergence of Dynamic Routing Protocols

An important characteristic of a routing protocol is how quickly it converges when there is a change in the topology. Convergence is when the routing tables of all routers are at a state of consistency. The network has converged when all routers have complete and accurate information about the network. Convergence time is the time it takes routers to share information, calculate best paths, and update their routing tables. A network is not completely operable until the network has converged; therefore, most networks require short convergence times. Convergence is both collaborative and independent. The routers share information with each other but must independently calculate the impacts of the topology change on their own routes. Because they develop an agreement with the new topology independently, they are said to converge on this consensus.

Convergence properties include the speed of propagation of routing information and the calculation of optimal paths. Routing protocols can be rated based on the speed to convergence;
the faster the convergence, the better the routing protocol. Generally, RIP and IGRP are slow to converge, whereas EIGRP, OSPF, and IS-IS are faster to converge.

1.7 Need of Optimization in Computer Networks

The Major components of a network are, an end system (A PC or Laptop), communication medium (wired or wireless) Intermediate devices like switches, routers and gateways, server systems (web server, mail server etc.), a network operating system (Windows 2008, Linux or Sun Solaris etc.), a network topology and a communication model/architecture (Apple, TCP/IP etc.).

a) **End System:** It is a personal computer or a laptop with a specific configuration of processor, memory (Primary as well as secondary) and an operating system. For enhancing the performance it is required to upgrade hardware.

b) **Communication Medium:** It provides the connection between two nodes in a network. To enhance the speed of communication between two nodes a channel with higher bandwidth can be employed. Further to enhance the reliability of communication another path can be added by addition of new node and a backup link with additional cost and management efforts.

c) **Server System:** It is a machine with high power processor (Many times parallel processor architecture), memory, server software and a powerful network based operating system. For enhancing the performance it is required to upgrade one or combination of above mentioned components.

d) **Switches/Routers/Gateways:** These are specialized devices that perform the forwarding of information chunks known as packets of an application. These devices perform the process of store and forward for processing of packets. For this function these devices maintains the buffers. To enhance the performance either the device should be upgraded or buffer size should be increased to avoid the packet loss due to bursty traffic flows.

e) **Network Topology:** It defines the connectivity of nodes in a network. There are various types of topologies like star, bus, tree, mesh, partial mesh, hybrid etc for configuring the physical structure of a network. For enhancing reliability and performance a combination of above technologies can be configured as per the need. The optimal topology design is
a NP-hard problem which needs continuous optimization.

f) **Communication Architecture:** It is a set of protocols forming a protocol stack that governs the type of services that can be provided through network. Currently TCP/IP protocol is most popular architecture employed in Internet. The network layer in the TCP/IP has the responsibility of moving the packets from one node to another node in a network for different applications. For finding the path from one node to another node for various types of services with efficient utilization of network resources by avoiding congestion is the responsibility of network layer. For this network layer employs the routing protocols. Network protocols use the routing algorithms for finding the best routes. The routing algorithms uses various metrics like hop count, delay, load, reliability to find the efficient path between two nodes (Kumar and Kumar 2008). The design and implementation of adequate strategy is a complex issue in today’s networks due to exponentially increment of traffic in network. The good routing strategy makes better utilization of network resources as selection of optimal path is routing algorithm dependent. Therefore routing issues also need continuous optimization.

As the amount of data that businesses need to process increases, best use of network utilization becomes critical. Every business needs to make sure that every aspect of its physical network is performing as efficiently as possible. The answer to these data problems is not always providing more bandwidth. In view of the increasing scale, complexity, and diversity of networks, optimization for certain objectives such as maximum efficiency, maximum reliability and minimum energy has become an essential critical issue in real-world applications. The design of network topology and routing algorithm are the major issues in a network that affects the performance of a network. These two areas need the optimization for providing reliable and efficient services with overall low cost (Kumar and Kumar 2010b). Broad scope of network routing is to address routing algorithms, routing protocols, and architectures, with architectures encompassing several different aspects for efficient routing.

### 1.8 Optimization Techniques

Optimization is a mathematical discipline that concerns the finding of minima and maxima of functions, under the given constraints, by maximizing desired factors and minimizing undesired
ones (Andres 2011). An optimization algorithm is one that accepts a solution set (search space) to the problem (single objective or multi-objective) as input, processes the search space by applying its different operators considering the specified constraints and return an optimal solution with minimum cost. The cost is problem dependent.

Broadly, optimization algorithms can be grouped in two basic classes: deterministic and probabilistic algorithms. Deterministic algorithms are most often used if a clear relation between the characteristics of the possible solutions and their utility for a given problem exists. The search space can efficiently be explored using some method like divide and conquer scheme. If the relation between a solution candidate and its “fitness” are not so obvious or too complicated, or the dimensionality of the search space is very high, it becomes harder to solve a problem deterministically. Trying it would possible result in exhaustive enumeration of the search space, which is not feasible even for relatively small problems.

Deterministic algorithms usually employ heuristics in order to define the processing order of the solution candidates. Probabilistic methods may only consider those elements of the search space in further computations that have been selected by the heuristic. Heuristics are a part of global optimization algorithms that helps in deciding which possible solutions from a solution is to be tested next or how the next individual can be produced. Heuristics are problem class dependent. An example for such a strategy is informed search.

Probabilistic algorithms are class of Monte Carlo based approaches. They trade in guaranteed correctness of the solution for a shorter runtime. This does not mean that the results obtained using them are incorrect—they may just not be the global optima. On the other hand, a solution a little bit inferior to the best possible one is better than one which needs thousands of years to be found.

1.8.1 Mathematical Representation of an Optimization Problem

Mathematically, a problem can be represented as:

Optimize \( y = f(x_1, x_2, \ldots, x_n) \)-----------------------------(1)
Subject to $g_j=(x_1, x_2, ..., x_n) \ {\leq, =, \geq}, \ b_j \ j=1, 2, ..., m$----(2)

Equation (1) describes the objective function and Equation (2) constitutes the set of constraints imposed on the desired solution. The decision variables are used to describe the objective function $y= f(x_1, x_2, ...., x_n)$. The $x_1, x_2, …., x_n$ (x_i variables) represent the set of decision variables (Google 2011). The term optimization means either maximize or minimize the value of the objective function depending on the problem (New Age Publishers 2011). The Equation (2) indicates that each constraint can take the form of an equality (=) or an inequality ($\leq$ or $\geq$) relationship. The local and global optimum concept is shown in figure 1.4 (New Age Publishers 2011).

![Figure 1.4: Local and Global Optimum for Single Variable Problem](image)

### 1.8.1.1 Components of Optimization Problem

An optimization problem has to be minimized or maximized. For minimization or maximization some attributes of the problem requires optimization. Therefore problem can be sub-divided in different components:

- **Objective function**: It is quantitative performance function of a system that we want to maximize or minimize (Andres 2011).
- **Variables**: Variables are building blocks whose decision influence the objective function
- **Constraints**: These are the conditions or set of relations that variables are forced to satisfy the unconstrained optimization problem is simplest situation to be considered. In such problems no constraints are imposed on the decision variables, and differential calculus can be used to analyze them. Another relatively simple form of the general optimization
problem is the case in which all the constraints of the problem can be expressed as equality (=) relationships. The technique of Lagrangian multipliers can be used to find the optimal solution for many of these problems.

There are basically four factors that can make optimization problems fairly complex and difficult to solve:

i. Existence of multiple decision variables in a problem
ii. Decision variables complex relationships and the associated outcome
iii. Possible existence of one or more complex constraints on the decision variables
iv. Presence of uncertainty

1.8.2 Classification of Search and Optimization Methods

There are various methods for providing optimal solution to different type of problems. On basis of algorithmic approach these are classified as under in figure 1.5:

![Classification of Search Optimization Techniques](image)

**Figure 1.5:** Classification of Search Optimization Techniques
1.8.2.1 Mathematical Optimization Techniques

Mathematical optimization refers to the selection of a best element from some set of available alternatives using some information about the main characteristics of the underlying process (Mohammad 2011). In the simplest case, this means solving problems in which one seeks to minimize or maximize a real function by systematically choosing the values of real or integer variables from within an allowed set. This formulation, using a real-valued objective function, is probably the simplest example. More generally, it means finding "best available" values of some objective function given a pre-specified region. These methods of optimization are useful in finding the optimum solution of continuous and differentiable functions (Kumar 2011). These methods are analytical and make use of the techniques of differential calculus in locating the optimum points. For example: Given $y=f(x)$, take the derivative of $f$ w.r.t. $x$, and solve for $x$. It works perfectly but for simple analytical functions. Three main types of problems can be handled by the classical optimization techniques:

- Single variable functions
- Multivariable functions with no constraints,
- Multivariable functions with both equality and inequality constraints. In problems with equality constraints the Lagrange multiplier method can be used. If the problem has inequality constraints, the Kuhn-Tucker conditions can be used to identify the optimum solution.

There are a range of algorithms to deal with optimization problems including linear programming, integer programming, nonlinear programming, and etc.

a) **Linear Programming**: It studies the case in which the objective function $f$ is linear and the set “A” (design variable space) is specified using only linear equalities and inequalities (Andres 2011). These constitute the most important class for which efficient solution techniques have been developed. In a linear-programming problem, both the objective and the constraint relationships are expressed as linear functions of decision variables. A linear relationship of the variables $x_1, x_2, \ldots, x_n$ is a function of the form:

$$a_1x_1 + a_2x_2 + \ldots + a_n x_n ;$$

where all the $x$ variables have exponents of 1.
b) **Integer Programming**: Here some (or all) of the decision variables are required to take on integer values.

c) **Quadratic Programming**: It allows the objective function to have quadratic terms, while the set A must be specified with linear equalities and inequalities. A quadratic function contains either squared terms ($x_i^2$) or cross-product terms ($x_i x_j$).

d) **Nonlinear Programming**: It studies the general case in which the objective function or the constraints or both contain nonlinear parts.

e) **Stochastic Programming**: It studies the case in which some of the constraints depend on random variables.

### 1.8.2.2 Advanced Optimization Techniques

The meta-heuristic optimization techniques have attracted increasing attentions in recent years for solving complex optimization problems. They are more robust than traditional methods based on formal logics or mathematical programming for many real world applications (Mohammad 2011). These techniques have obviously progressed and new ideas in the area of optimization algorithms have been developed during recent decades.

The biologically-inspired computation rapidly advanced over the past decade is a covering term for different computational methods that are based on principles or models of biological systems. This class of methods such as evolutionary algorithms (EAs), ant colony optimization (ACO), particle swarm optimization (PSO), Genetic Algorithm (GA) and artificial neural networks complements traditional techniques in the sense that these techniques can be applied to large-scale applications where little is known about the underlying problem and heuristic approaches encounter difficulties such as the risk of trapping in local minima. Therefore, bio-inspired methods are becoming increasingly important in face of the complexity of today's demanding applications, and accordingly they have successfully been used in various fields.

- **Evolutionary (Evolve in Genetics)**: Genetic algorithm, Memetic algorithm etc.
- **Nature Inspired (Evolve in Social behavior)**: Particle swarm optimization, Ant colony optimization etc.
- **Traditional Heuristic approaches**: Simulated Annealing, Tabu Search and Hill climbing
a) **Hill Climbing**: It is a graph search algorithm where the current path is extended with a successor node which is closer to the solution than the end of the current path. In simple hill climbing, the first closer node is chosen whereas in steepest ascent hill climbing all successors are compared and the closest to the solution is chosen. Both forms fail if there is no closer node. This may happen if there are local maxima in the search space which are not solutions. Hill climbing is used widely in artificial intelligence fields, for reaching a goal state from a starting node.

Given \( y=f(x) \), pick a point \( x_0 \); compute the gradient \( \nabla f(x_0) \) and step along the gradient to obtain \( x_1 = x_0 + \alpha \nabla f(x_0) \). Repeat until extremum is obtained. This technique requires existence of derivatives and easily gets stuck on local extreme.

b) **Simulated Annealing (SA)**: The name and inspiration come from annealing process in metallurgy, a technique involving heating and controlled cooling of a material to increase the size of its crystals and reduce their defects. The heat causes the atoms to become unstuck from their initial positions (a local minimum of the internal energy) and wander randomly through states of higher energy; The slow cooling provides atoms more chances of finding configurations with lower internal energy than the initial one. In the simulated annealing method, each point of the search space is compared to a state of some physical system, and the function to be minimized is interpreted as the internal energy of the system in that state. Therefore the goal is to bring the system, from an arbitrary initial state, to a state with the minimum possible energy.

For better results SA can also be used within a standard GA algorithm by starting with a relatively high rate of mutation and decreasing it over time along a given schedule (Peter 2005). Simulated annealing is quite similar to local search. Technically, simulated annealing can be seen as an evolutionary algorithm with population size 1, but it is generally seen as another class of algorithms.

c) **Tabu Search (TS)**: Tabu search is another general combinatorial optimization technique (Peter 2005). The idea is basically the same as with local search, but when a local optimum is found, the algorithm does not stop, Instead the scheduling might continue, by disallowing points previously visited (hence ‘tabu’). Tabu search (TS) is similar to
simulated annealing in that both traverse the solution space by testing mutations of an individual solution. While simulated annealing generates only one mutated solution, tabu search generates many mutated solutions and moves to the solution with the lowest energy of those generated. In order to prevent cycling and encourage greater movement through the solution space, a tabu list is maintained of partial or complete solutions. It is forbidden to move to a solution that contains elements of the tabu list, which is updated as the solution traverses the solution space.

d) *Ant colony optimization (ACO):* It uses many ants (or agents) to traverse the solution space and find locally productive areas. In the real world, ants (initially) wander randomly, and upon finding food return to their colony while laying down pheromone trails. If other ants find such a path, they are likely not to keep traveling at random, but instead follow the trail laid by earlier ants, returning and reinforcing it if they eventually find food. Over time, however, the pheromone trail starts to evaporate, thus reducing its attractive strength. The more time it takes for an ant to travel down the path and back again, the more time the pheromones have to evaporate. A short path, by comparison, gets marched over faster, and thus the pheromone density remains high. Pheromone evaporation has also the advantage of avoiding the convergence to a locally optimal solution. If there were no evaporation at all, the paths chosen by the first ants would tend to be excessively attractive to the following ones. In that case, the exploration of the solution space would be constrained. Thus, when one ant finds a good (short) path from the colony to a food source, other ants are more likely to follow that path, and such positive feedback eventually leaves all the ants following a single path.

The idea of the ant colony algorithm is to mimic this behavior with "simulated ants" walking around the search space representing the problem to be solved. Ant colony optimization algorithms have been used to produce near-optimal solutions to the traveling salesman problem. While usually inferior to genetic algorithms and other forms of local search, it is able to produce results in problems where no global or up-to-date perspective can be obtained, and thus the other methods cannot be applied. They have an advantage over simulated annealing and genetic algorithm approaches when the graph may change dynamically. The ant colony algorithm can be run continuously and can adapt to changes
in real time.

e) **Particle swarm optimization (PSO):** It is a computational method for multi-parameter optimization which also uses population-based approach. A population (swarm) of candidate solutions (particles) moves in the search space, and the movement of the particles is influenced both by their own best known position and swarm's global best known position. Like genetic algorithms, the PSO method depends on information sharing among population members. In some problems the PSO is often more computationally efficient than the GAs, especially in unconstrained problems with continuous variables (Hassan et al. 2005).

f) **Genetic Algorithm (GA):** Genetic Algorithm (GA) uses an initial population called as chromosomes of all feasible solutions of the problem. On basis of fitness function it evaluates each chromosome to select the best pair for crossover and mutation operation by performing a number of iterations to achieve the optimal solution.

- Does not require derivatives, just an evaluation function (a fitness function)
- Samples the space widely, like an enumerative or random algorithm, but more efficiently
- Can search multiple peaks in parallel, so is less hampered by local extrema than gradient-based methods
- Crossover allows the combination of useful building blocks, or schemata (mutation avoids evolutionary dead-ends)
- Robust

g) **Memetic Algorithms:** The Incorporating of other heuristics or algorithms with evolutionary computing leads to memetic algorithms (or hybrid-genetic algorithms). For example, during evolution, local search might be used first to find better solutions and minimize search complexity. The idea is that only by incorporating problem-specific information, better algorithms can be created.
1.9 Evolutionary Computation (EC)

Metaheuristic methods broadly fall within stochastic optimization methods. Evolutionary computation is a sub-field of the meta-heuristic methods that follow some natural phenomena (Genetic inheritance and Darwinian strive for survival) in search process. Evolutionary computation (Wikipedia 2011a) uses iterative progress for growth in a population using parallel processing to reach an optimum solution. Literature reflects that artificial intelligence and soft computing has contributed in a lot of activities for improvement. One of such tools of soft computing is evolutionary algorithms (Zelinka 2011). As comparative to other global optimization techniques evolutionary approach provides some key features like population-based collective learning process, self-adaptation, parallelism and robustness. Evolutionary algorithms domain includes genetic algorithms (Holland 1975), evolution strategies (Rechenberg 1973, Schwefel 1977), evolutionary programming (Fogel et al. 1966), and genetic programming (Koza 1992), (Michalewicz and Michalewicz 1997). All these techniques share a common conceptual base of simulating the evolution of individual structures via processes of selection, crossover and mutation. Compared to other global optimization techniques, evolutionary algorithms (EA) are easy to implement and very often they provide adequate solutions.

1.9.1 Basics of Evolutionary Computing

Evolution says that you should always have some solutions to a problem, called population, and then try to make that population better and better. At some point you will find a solution that is good enough for you, or you have already done lots of attempts. At that stage selection of best individual solution, will end the solution. Therefore at each point in time, during problem solving, a solution is available. This is called the anytime behavior of an evolutionary algorithm. Generally the starting population with the starting solutions is generated at random. An evolutionary algorithm usually soon gives pretty good solutions.

1.9.2 When to use Evolutionary Computing

This technique is usually used for hard optimization problems (Peter 2005). It is usually well suited under the following conditions:
• **Large solution space**: Evolutionary Computing is able to tackle large problems with huge search space. If the solution space is not too large, exhausted search is probably to be preferred.

• **Best solution not required**: Evolutionary computing usually finds a good solution, but it is not guaranteed that this solution is the best solution possible. It also does not provide information on how far away the found solution is from the global optimum.

• **No very fast results needed**: Evolutionary computing algorithms may take up some time due to number of iterations performed.

• **No exact heuristics available**: When problem specific knowledge is not available, evolutionary computing is well suited for little understood problems. If good, fast and exact heuristics are available, that lead to an optimum value, they should be used instead.

• **Constraints or multiple objectives are present**: Evolutionary Computing is well able to handle difficult problems with constraints and multiple objectives.

• **Robust solutions are needed**: Evolutionary Computing can be used in combination with noise and dynamic environments. The algorithm usually adapts well to a new environment.

### 1.9.3 Evolutionary Algorithms

An evolutionary algorithm is a sub-field of evolutionary computing (Wikipedia 2011). The well known techniques under the domain are:

i. **Evolution strategies**: Evolution strategies (ES) evolve individuals by means of mutation and intermediate or discrete recombination (Eiben et al. 1994). ES algorithms are designed particularly to solve problems in the real-value domain. They use self-adaptation to adjust control parameters of the search. De-randomization of self-adaptation has led to the contemporary Covariance Matrix Adaptation Evolution Strategy (CMA-ES).

ii. **Evolutionary programming**: Evolutionary programming (EP) involves populations of solutions with primarily mutation and selection and arbitrary representations. They use self-adaptation to adjust parameters, and can include other variation operations such as combining information from multiple parents.
iii. *Genetic programming:* Genetic programming (GP) is a related technique popularized by John Koza in which computer programs, rather than function parameters, are optimized. Genetic programming often uses tree based internal data structures to represent the computer programs for adaptation instead of the list structures typical of genetic algorithms.

iv. *Genetic Algorithm:* On basis of fitness function it evaluates each chromosome (from population) to select the best pair for crossover and mutation operation by performing a number of iterations to achieve the optimal solution. GA methodology is most popular technique to solve the complex optimization problem.

Genetic algorithms are most popular optimization technique for multi-objective optimization problems among evolutionary algorithms.

### 1.9.3.1 Genetic Algorithms

GA was developed by John Holland at the University of Michigan in the early 1970’s. Genetic algorithms are stochastic search methods, theoretically and empirically proven to provide robust search in complex spaces (Goldberg 1989). These operate on a population of potential solutions, applying the principle of survival of the fittest to generate improved estimations to a solution. At each generation, a new set of approximations is created by the process of selecting individuals according to their level of fitness and breeding them together using genetic operators inspired by natural genetics. This process leads to the evolution of better populations than the previous populations. Genetic algorithms are typically implemented using computer simulations in which an optimization problem is specified. The GA consists of an iterative process (Goldberg 1989), (Wikipedia 2011) that evolves a working set of individuals called a population toward an objective function, or fitness function.

a) *Fitness function:* Fitness function is the measure of the quality of an individual. The fitness function should be designed to provide assessment of the performance of an individual in the current population. The selection step is preceded by the fitness assignment which is based on the objective value. This fitness is used for the actual selection process.
b) Selection: There are many types of selection methods used in genetic algorithms, including:

- Rank-based fitness assignment
- Roulette wheel selection
- Local selection
- Tournament selection
- Steady-state selection

A decision about the method of selection to be applied is one of the most important decisions to be made.

c) Crossover: This is a version of artificial mating. Individuals with high fitness should have high probability of mating. Crossover represents a way of moving through the space of possible solutions based on the information gained from the existing solutions. Examples of crossover are one point, two point and multipoint crossovers. New crossover methods can be developed specific to the problem.

d) Mutation: Mutation is another recombination technique. It is used to make sure all the elements in a population are not homogeneous and diversity is maintained. Mutation can help genetic algorithms escape local optima in the search for the global optima. Mutation rate is usually ranges not higher than 30% (Lima et al. 2005).

e) GA State of the field: GAs have been successfully applied to many significant problems in machine learning and data mining, most notably classification, pattern detectors (Rizki et al. 2002), predictors (Chan et al. 2003), and payoff-driven reinforcement learning (Goldberg 1989). The general strengths of genetic algorithms lie in their ability to explore the search space efficiently through parallel evaluation of fitness (Cantu-Paz 2000) and mixing of partial solutions through crossover (Goldberg 2002); maintain a search frontier to seek global optima (Goldberg 1989); and solve multi-criterion optimization problems.

Some limitations of GAs are that in certain situations, they are overkill compared to more Straight-forward optimization methods such as hill-climbing. In global optimization scenarios, GAs often manifests their strengths: efficient, parallelizable search; to evolve
solutions with multiple objective criteria (Llora and Goldberg 2003). GAs has been successfully used to solve variety of parameter optimization problems in computerscience and engineering (Sheta and Jong 1996, Sheta 2006, Sheta and Turabieh 2006).

1.9.4 Applications of Evolutionary Computation

Evolutionary Computation has many uses in the world of telecommunications. It is used in locating antenna’s, designing networks, routing data in networks, and all kinds of other combinatorial optimization problems (Peter 2005). Telecommunications is one of the most active application areas of EC.

Even though evolutionary computation has been widely accepted for solving several important practical applications in engineering, business, commerce, etc., yet in practice sometimes they deliver only marginal performance. For several problems a simple Evolutionary algorithm might be good enough to find the desired solution. As reported in the literature, there are several types of problems where a direct evolutionary algorithm could fail to obtain a convenient (optimal) solution ((Morgan et al. 1997), (Chang and Lo 2000), (Somasundaram et al. 2005), (Tseng and Liang 2005)). This clearly paves way to the need for modifying the basic evolutionary algorithm or hybridizes with other techniques. Some of the possible reasons for hybridization are as follows (Sinha and Goldberg 2003):

- To improve the performance of the evolutionary algorithm (For example: speed of convergence)
- To improve the quality of the solutions obtained by the evolutionary algorithm
- To incorporate the evolutionary algorithm as part of a larger system

In computer routing for even single demand, finding a route using two additive or multiplicative metrics causes complexity which makes the problem as NP-hard (Spall 2003). NP problems can be solved better with probabilistic or evolutionary approaches (Sivanandam and Deepa 2008). GA is a most popular global optimization technique in literature that has been used for solving various NP problems (VijayLakshmi and Radhakrishnan 2007). Therefore researches explored the use of Genetic Algorithms in computer network routing.
A Computer Network is a collection of nodes connected together by means of a communication medium. Resource sharing and communication are two principal reasons for building and using computer networks. The Classification can be carried out on the basis of geographical area covered, methods of connecting the nodes in network, transmission technology and many other factors. A successful network design takes account of several important factors, such as cost, security, integrity, scalability, reliability and robustness. The reliability and robustness are of grave importance.

To manage the data on network a protocol is required. It governs the communication between communicating entities. It defines the format of data units, rules of operations like data delivered, retransmission required, transmission speed, packet size etc. TCP, IP, HTTP, OSPF etc. are examples of protocols. Today, the TCP/IP architecture is most dominant.

Routing is divided into two broad categories, Static and Dynamic routing. Static routing is useful for small networks only. As the network grows, overheads increases, which are not possible to handle manually and dynamic routing comes into picture. In today’s networks dynamic routing is implemented. Routing protocols use metrics to determine which route is the best path. To select the best path, the routing protocol must be able to evaluate and differentiate among the available paths. For this purpose, a metric is used. A metric is a value used by routing protocols to assign costs to reach remote networks. The metric is used to determine which path is most preferable when there are multiple paths to the same remote network. Each routing protocol calculates best route in a different way. For example, RIP uses hop count, EIGRP uses a combination of bandwidth and delay, and the Cisco implementation of OSPF uses bandwidth. The best known shortest path algorithms are Dijkstra and the Bellman-Ford algorithm. The most popular OSPF uses Dijkstra algorithm.

The design of network topology and routing algorithm are the major issues in a network that affects the performance of a network. These two areas need the optimization for providing reliable and efficient services with overall low cost. Optimization is a mathematical discipline that concerns the finding of minima and maxima of functions, under the given constraints, by
maximizing desired factors and minimizing undesired ones. An optimization algorithm is one that accepts a solution set (search space) to the problem as input, processes the search space by applying its different operators considering the specified constraints and returns an optimal solution with minimum cost. The cost is problem dependent.

An optimization problem has to be minimized or maximized. For minimization or maximization some attributes of the problem requires optimization. Therefore problem can be sub-divided in different components like objective function, variables and constraints.

Evolutionary Computation has many uses in the world of telecommunications. It is used in locating antenna’s, designing networks, routing data in networks, and all kinds of other combinatorial optimization problems. The bio-inspired methods are becoming increasingly important in face of the complexity of today's demanding applications, and accordingly they have successfully been used in various fields. This class of methods such as evolutionary algorithms, ant colony optimization, particle swarm optimization, and artificial neural networks complements traditional techniques in the sense that these techniques can be applied to large-scale applications where little is known about the underlying problem and heuristic approaches encounter difficulties such as the risk of trapping in local minima.

As comparative to other global optimization techniques evolutionary approach provides some key features like population-based collective learning process, self-adaptation, parallelism and robustness. GA is a most popular global optimization technique in literature that has been used for solving various NP problems.