CHAPTER 2

Common Network Technology

2.1 Introduction

The two most successful computer networks at present time are Telephone Networks and Internet. The telephone network inter connects more than a billion telephones worldwide and the network is controlled by a set of special purpose computers inter connected by an Internet like computer network [15]. Internet is the other major networking technology. It started as a research project in late 1960’s but has since grown exponentially to interconnect tens of millions of computers around the world. At present, the Internet carries computer data traffic almost exclusively with voice and video traffic. It is interesting to contrast the styles of the Internet and the telephone network. The internet philosophy is to put intelligence at the end points, and to assume that the network may drop, mangle and reorder data. The network is expected to maintain any information about the performance requirements of an end point. Here all packets are treated alike. Thus an application cannot obtain a guarantee of performance, such as assurance that the network will transfer information from end to end within a time bound. Thus experience shows that the Internet does a good job at carrying data, but a poor one at carrying any thing that requires an end to end quality of service.

In contrast, the telephone network assumes that the end points are dumb, and that all intelligence ought to be placed in the network. Moreover network is explicitly designed for a single application, digitized voice. Voice calls are guaranteed to receive sufficient bandwidth, a low delay and zero loss, but it achieves this at greater expense. Specifically, even when a caller is not speaking, the network continues to reserve resources for the call, wasting network resources, which could otherwise be devoted to other callers. This is the fundamental reason why the telephone network is more
expensive to operate than the Internet. But telephone network offers an end to end
guarantee of quality of service.

The Holy Grail of computer networking is to design a network that has the
flexibility and low cost of Internet, yet offers the end to end quality of service as
provided by the telephone network [1].

One approach to such a design is to add a quality of service mechanism to Internet. The other approach is to design a new network from scratch that has the desired
properties. This is the stated goal of Asynchronous Transfer Mode (ATM) networks. Although it is too early to judge the extent to which ATM networks have achieved their
goals, they are sure to play an important role in the future.

2.1.1 Telephone Networks

Though the global telephone network is the world’s largest computer network, it is
not recognized as such because it is specialized to carry voice. Besides, voice, the
network also carries video, facsimile, and telemetry data. It is enormous both in its reach
and in the service it provides. More than a billion telephones interconnect one fourth of
world’s population.

2.1.1.1 Concepts

The telephone network offers a single basic service to its users; two way switched
voice service with small end to end delays and a guarantee that a call, once accepted, will
run to completion. It achieves this quality of service by setting up a circuit between the
two end points. The network guarantees enough resources to each circuit to ensure that
the service quality is met [44]. The four essential components of telephone system are
i) End system
ii) Transmission
iii) Switching
iv) Signaling
i) End System

The end system attached to a telephone network is usually a telephone instrument that consists of five parts.

- A voice to electrical signal transducer
- An electrical signal to voice transducer
- A dialer, which sends a series of tones or pulses to a central office
- A ringer
- A switch hook

ii) Transmission

The telephone network carries signals generated by a telephone instrument over transmission links. Many link technologies are in common use today. These include unshielded twisted pair (UTP), coaxial cable, microwave, satellite links and fiber optics which can be used depending on the bandwidth and cost requirements.

iii) Switching

Unlike television network, which provides one-to-many communication, the telephone network provides one-to-one communication. A telephone switch has two parts, the switching hardware, which carries voice, and the switch controller, which handles requests to set up and tear circuits.

iv) Signaling

The last element of the telephone network is signaling. As we know a switching system consist of a switch, which carries data, and a switch controller, which is responsible for establishing a path from a calling telephone to a destination. When a subscriber dials a number, the pulses or tones representing the destination's number are sent to the exchange at the local central office. Here, the switch controller interprets the signals to decide the destination of the call. If the call is local, then the controller sets up a path in the local exchange from input to the correct output. It then sends a signal on the
wire connecting the central office to the telephone instrument to cause a bell to ring. When the handset is picked up, voice from either end is typically digitized and sent over the circuit. The controller also sets up billing records with the calling and called telephone numbers, the time of call, and the duration. When either party hangs up, the switch controller clears the circuit from the switch and marks the corresponding lines “Idle”.

Generally, switch controller used to exchange special tones to set up and tear down calls. For example a “disconnect” tone would tear down a call. Other tones were used to disable billing or test switch features. Unfortunately, these tones attracted the attention of ‘phone phreaks’, people who discovered that special tone sequences permitted free long distance calls.

Telephone companies counter attacked this by changing the mechanism for exchanging messages between switch controllers. Instead of using the same channel for both voice and signaling tones, a separate Common Channel Interoffice Signaling (CCIS), network interconnects switch controllers. This network is not directly accessible to subscribers. Because control information is separate from the voice channel, CCIS is called out of band signaling. CCIS is more flexible than in-band signaling because peer switch controller can exchange arbitrarily complicated messages instead of just tone sequences.

2 1.1.2 Challenges
i) Multimedia

Multimedia communication is the simultaneous transmission of voice, video and data traffic. The existing telephone network is inadequate for multimedia traffic in three ways:

- Video requires between 0.35 and 3.5 Mbps of digital information per system. But the best of the current modems can achieve no more than 120 Kbps.
• Both compressed video traffic and data traffic are inherently bursty, that is, the traffic rate over short periods is higher than the rate over long periods. For typical data traffic, the peak to average rate can be as large as 1000.
• Network trunk capacities are decided based on the statistical behavior of large number of voice calls. With the arrival of multimedia communication, it is not clear whether voice call statistics are appropriate any longer.

ii) Backward Compatibility

Since telephone companies in developed nations have a huge investment in existing infrastructure, changes in the network have to be backward compatible. New equipment must interoperate with existing equipment and service provided to customer must continue to be supported.

iii) Regulation

Most countries recognize that telephones provide an essential economic infrastructure and so telephone companies around the world are subject to strict regulation. In many countries a government monopoly provides telephone service, and telephone service is often called a “natural monopoly”. This tradition of regulation usually tends to stifle innovation, particularly when the telephone company is a monopoly. For example, until recently, answering machines were illegal in Germany. Thus, German consumers were denied access to a convenient service because of out of date regulations. Technological innovations require rapid response, and the challenge is to do this within the regulatory environment.

iv) Inefficiencies within the system

One legacy of long history is a proliferation of systems and formats that are similar, but incompatible. For example the AT & T billing system processes more than two thousand types of billing records to generate a bill. Some formats date back several decades and must still be supported in the 21st century unless the billing system is
completely overhauled - a daunting and expensive exercise. The challenge facing the system is to recognize and eliminate inefficiencies due to engineering decisions necessary in the past, but no longer relevant.

2.1.2 The Internet

The Internet connects tens of millions of computers around the world, allowing them to exchange messages and share resources. Users of Internet can exchange Electronic mail, read and post electronic bulletin boards, access files anywhere in the network, and publish information for other users.

The Internet which started as a research project connecting four computers in 1969 had grown by late 1996 to a network of more than twenty million computers linking an estimated sixty million people.

The Internet is a loose collection of networks organized into a multilevel hierarchy using a wide variety of interconnection technologies. At the lowest level, ten to hundred computers may be connected to each other, and to a router, by a local area network. A large business or university campus may have many routers. These routers are usually linked by a campus wide network to a campus router that handles all traffic entering and leaving the campus. Campus routers are connected by leased lines to routers, belonging to an Internet Service Provider (ISP). These, in turn connect to routers in a high-speed wide area network called a backbone. A country usually has a handful of backbones--linking all its ISP's. Finally, national backbones interconnect in a mesh using international trunk lines.

A single authority usually administers campuses, so that all computers within a campus trust each other. Networks where all the computers trust each other are said to form an “intranet”. At higher levels, however, the network is heterogeneous and individual administrative domains rarely trust each other.
2.1.2.1 Concepts

The Internet is bound together by addressing, routing and the Internet protocol (IP). Addressing provides a uniform way of identifying a destination in the network. Routing allows information to traverse the network from end to end. Finally, IP allows data to be interpreted consistently as they travel across the network. A common definition of Internet is “The set of computers that are reachable with IP”. The easiest way to understand these concepts is to look at three steps involved in making a host “Internet-Capable”.

✔ First, we need an Internet address for addressing data. This can be obtained by requesting the network manager of our own organization or from an Internet Service Provider (ISP). If our intention is to provide Internet service to others, we can request a chunk of Internet addresses from an Internet addressing and numbering authority.

✔ Next, data originating from us has to somehow travel to the destination. If there is only one connection to the Internet, all the data will be carried on that unique link. However, if one decides to get two Internet links, then one will need to run a routing protocol to decide which link has a better path to each destination. This routing protocol must communicate with peer protocols running at neighboring computers and announce our presence to the rest of the network.

✔ Finally, when we send data, we must have software on the computer that formats it using the IP format, so that routers along the path know what they are getting.

Once these three steps are completed, then one can send data into the network, which will eventually arrive at its destination.

One reason the Internet has spread so quickly is that the three basics, addressing, routing and IP are designed to scale. For example, the Internet does not have a centralized addressing authority. Any ISP has the authority to give away part of address space.
allocated to it without further references to a higher authority. Similarly routing changes
dynamically if hosts are added or removed or if transmission lines go up and down.

2.1.2.1 Basic Internet Technology

Two key ideas in Internet technology are packets and store and forward transmission.

i) Packets

The Internet carries all information using packets. A packet has two parts. One is
the information content called the payload, and the other is the information about the
payload, called header or metadata. The header consists of fields such as the source
and destination address, data length, sequence number and data type, which helps packet
reach the destination safely.

ii) Store and forward

Header allows a form of data forwarding called store and forward, in which a
packet is stored at a series of routers, and eventually delivered to the destination. The
advantages with this method are:

- We can store all the packets as backlog before accessing an expensive transmission
  link, and then can be forwarded. By this, packet network allows the cost of that link
to be shared among many users.
- If a link goes down, incoming packets can be stored and delivered when the link
  comes back.
- It is less expensive to operate than telephone network. But users of network suffer
  three problems:

  1. It is hard for users to control how long their packets will be delayed in the
     network
  2. Since packets must be stored, switches need memory for buffers, which can be
     expensive.
3. If many users decide to send packets to same destination simultaneously, and there is no enough buffer space to hold all incoming data, then packets will be lost.

2.1.2.1.2 Internet Link Technology

We can divide Internet link technologies in to roughly three categories: those used within a campus, those used in the wide area, and those used to access the Internet from home.

Within a campus, Internet packets are carried over LANs. Each building in the campus provides desktop Internet access using either 10 Base T Ethernet or 100 Mbps fast Ethernet. Each building has a router. Groups of about 10 routers are connected with FDDI ring, which in turn interconnects one router in each such ring.

Wide area links (anything longer than 5 kms) makes use of telephone network. A common approach is to lease a point to point T1 or T3 line to connect two wide area routers, such as a campus router and a router belonging to an ISP. As SONET enters the telephone infrastructure, leased OC3 links are gradually becoming available. But these leased lines are very expensive.

A second solution to wide area connection is ‘frame relay’. A frame relay network is a connection oriented packet switched network that supports variable packet sizes and currently, only permanent virtual circuits. Since all connections are permanent, there is no overhead in signaling or admission control, making the network easier to build and maintain. The main service available from a frame relay network is a virtual link (corresponding to permanent virtual circuit i.e. PVC) that guarantees a minimum bandwidth called ‘Committed Information Rate (CIR)’. Since a frame relay PVC can have an arbitrary bandwidth, and users are charged only for bandwidth they use, a frame relay link proves to be much cheaper than a leased line. Moreover frame relay allows the single leased line between the frame relay access point on a campus and the frame relay
network to be multiplexed among several PVCs, so that a campus needs only one leased line for several virtual links. Because of these benefits, frame relay service is growing enormously in popularity.

Home access to Internet is traditionally through modems. These are restricted to 50 Kbps because of the way voice is sampled during analog to digital conversion. Somewhat higher bandwidth can be obtained using ISDN lines with higher cost. These provide up to 128 Kbps of bandwidth. However several new technologies for high speed Internet are on the anvil. Several local telephone companies offer 'Asymmetric Digital Subscriber Link (ADSL)' modems, which offer 6 Mbps to the home and 1.5 Mbps from the home, but these can only be used for homes within about 18,000 feet of the central office. In areas where the cable operator has installed two way cables, homes can access the Internet at a speed of about 10 Mbps, but it is too expensive.

A promising solution is to use the telephone network to provide the return path, with the cable TV plant providing only one way access. This gives users, a peak access rate of about 1.5 Mbps. Technologies in all three areas are evolving rapidly, and a picture may well change in the next couple of years.

2.1.2.2 Challenges

1) IP Address Space Shortage

One of the most technological challenges facing Internet is the shortage of IP address. As we know, IP address is hierarchically structured, with a network number and an interface number. Once a central authority gives an organization a network number, the organization can create address within that network whenever it wants. Unfortunately if the organization chooses not to use the entire available address space, this space is not available to other organizations.
Efforts are under way to increase the IP address space to 128 bits. This scheme has to be backward compatible with 32 bit address already in the Internet, and eventually the new scheme has to be adopted all over the network.

2) Problems of Decentralized Control

With decentralization, reliable service can be guaranteed which was essential to the Internet’s scalability.

But with decentralization there is no control over adding routers to the network, which can cause disaster in near future. Another problem is security. Packets sent on the Internet are visible to any one who wishes to examine them. Third problem is that there is no uniform way to do accounting on the Internet and therefore mechanism for billing and cost sharing are non existent. This has led to flat rate pricing based on the bandwidth of the access link. Finally, decentralized control means that routing can sometimes be non-optimal. Since no single authority is responsible for globally optimal decisions, a series of locally optimal decisions can result in global chaos.

3) Multimedia

Multimedia applications that need real time performance guarantees that bonded delay and a minimum throughput are not well supported in current Internet.

These parameters are often called ‘quality of service’. The major challenges facing the Internet today is integrating quality of service requirements into the existing architecture.

2.2 Common Protocols

Now we examine some of the important protocols in telephone network and Internet.
2.2.1 Telephone Network Protocols

In this section we survey some of the common protocols in telephone networks. We start with description of telephone network protocol stack and then we deal with traditional digital transmission hierarchy and its replacement, which is called SONET in the United States and SDH by the international telecommunication union.

Protocol Stack

Table 2.2.1.1 shows a typical telephone network protocol stack. Here the control plane is used for call establishment and network management, and the data plane moves 8 bits samples generated at a constant bit rate. The data plane has only the physical, data link and application layers. The application layer consists of telephone and telephone like application layer such as fax and modems. The control plane is essentially the signaling system 7 stack.

<table>
<thead>
<tr>
<th>Data Plane</th>
<th>Control Plane</th>
</tr>
</thead>
<tbody>
<tr>
<td>Application</td>
<td>Voice/fax</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Session</td>
<td></td>
</tr>
<tr>
<td>Transport</td>
<td></td>
</tr>
<tr>
<td>Network</td>
<td></td>
</tr>
<tr>
<td>Data link</td>
<td>SONET/P.H</td>
</tr>
<tr>
<td>Physical</td>
<td>Many</td>
</tr>
</tbody>
</table>

The traditional digital transmission hierarchy

Telephone companies find it economical to multiplex long distance calls on to high-speed trunks, because this reduces the number of physical long distance trunks needed in the network. If each multiplexed trunks were able to carry a different number
of calls, each trunk would need specialized equipment running at its own transmission rate, which is not cost effective. International standards bodies have therefore declared a few standard multiplexed rates that form the digital transmission hierarchy.

The base level in the hierarchy is the equivalent of a signal digitized voice call, that requires a bandwidth of 64 Kbps. Higher levels of hierarchy carry more calls over higher speed trunks. The transmission hierarchy differs from region to region.

We say that a system is plesiochronous (which means almost synchronous) if each component generates data at nominally the same bit rate, but these bit rates may vary within a tightly controlled tolerance. Traditionally, telephone networks are plesiochronous, because they only require each component to have a good quality clock, and components do not need to be synchronized with each other explicitly. However, a plesiochronous network suffers some serious problems. Among which the three significant one’s are:

First, each part of world has its own digital transmission hierarchy. Thus, connecting calls internationally requires expensive interface equipment. Second, the process of justification spreads data from a tributary all over the output frame. Thus, extracting data from single tributary is hard. Finally for the same reason, it is hard to build switches that switch bundles of voice calls instead of individual calls.

**SONET/SDH**

The problems with plesiochrony described earlier disappear if all the tributaries entering a multiplexer are precisely synchronized. Because synchronous tributaries need no justification, they occupy a fixed portion of multiplexed frame, and thus can easily be extracted. In theory, in a synchronous network, even a single voice call can be extracted from multiplexed trunk containing many calls. This is the basis of **Synchronous Optical Network (SONET)** standard.
Synchronous Hierarchy

SONET defines a multiplexing hierarchy in the same way as does the plesiochronous digital hierarchy. A link at a higher speed is formed by byte-interleaving data from lower speed links. The higher speed trunks are exact multiples of the speed of lower speed trunk. This is because the higher speed trunks do not contain additional justification overhead.

SONET Frame Structure

SONET defines a multiple frame type, where each type corresponds to a particular link speed. Because SONET is synchronous, higher level frames are formed simply by byte-interleaving the lowest level (OC-1) frame. Therefore we need to study the structure of only the OC-1 frame.

---

*Figure: 2.2.1.1: SONET Frame Structure*
An OC-1 frame consists of 810 bytes, arranged in 9 rows and 90 columns as shown in figure 2.2.1.1. Each switch or multiplexer sends each frame in exactly 125 microseconds, so that each byte in the payload of a frame corresponds to one 64 Kbps call. The first three columns are overhead bytes, and the remaining 87 contain byte-interleaved data from multiple tributaries. Data from each tributary is placed in a payload container, also called the synchronous payload envelope. Each container has its own descriptive header. The overhead portion of a SONET frame contains many Operations, Administrative and Maintenance (OA & M) fields, parity bytes for error detection, and a pointer to the start of payload container in the frame.

**Signaling System 7**

The plesiochronous digital hierarchy and SONET standards describe the data transport protocols in the telephone network. Network control including call establishment, routing and enhanced services is accomplished using the SS7 protocol stack, which is shown in table 2.2.1.2
Table 2.2.1.2: SS7 Protocol Stack

<table>
<thead>
<tr>
<th>OSI Layer Name</th>
<th>SS7 Layer Name</th>
<th>Functionality</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Application</td>
<td>Application service Element (ASE)</td>
<td>Application level Functionality, such as Interpreting signaling Messages</td>
<td>FTP</td>
</tr>
<tr>
<td></td>
<td>Transaction capabilities Application part-1 (TCAP)</td>
<td>Allows a system to invoke procedure calls on remote Machines</td>
<td>RPC</td>
</tr>
<tr>
<td>Transport</td>
<td>Signaling connection control part-2 (SCCP)</td>
<td>Connection, sequence numbering, segmentation and reassembly, window flow control</td>
<td>TCP</td>
</tr>
<tr>
<td>Network</td>
<td>Message transfer part-3 (MTP3)</td>
<td>Routing</td>
<td>IP</td>
</tr>
<tr>
<td>Data link</td>
<td>Message Transfer part-2 (MTP-2)</td>
<td>Framing, link-level error detection, and retransmission</td>
<td>Ethernet</td>
</tr>
<tr>
<td>Physical</td>
<td>Message Transfer part-1 (MTP-1)</td>
<td>Physical transfer, usually on a TI circuit</td>
<td>Ethernet</td>
</tr>
</tbody>
</table>

Note that the digital transmission hierarchy and SS7 represent a separation of data and control. Protocol layers in SS7 are called parts. The SS7 transport subsystem consists of the message transfer part (MTP) and the signaling connection control part (SCCP).

The overall responsibility of MTP is to reliably transfer signaling messages from a source to a destination over the signaling network, compensating for link and router failures. MTP is partitioned into MTP level-1, MTP level-2 and MTP level-3, corresponding to the physical, data link, and network layers respectively.
*MTP level-1 is a physical link, typically a standard T1 voice circuit devoted to signaling.*

*MTP level-2 provides message framing on a link, CRC-16 error detection, per link go-back-n error correction and static window flow control. If too many errors are detected on a link, or the receiver withholds acknowledgements for too long, the link is declared down, and SS7 routing at MTP level-3 is informed about this change.*

*MTP level-3 is responsible for datagram routing, SS7 identifies endpoints with 24 bit address called point codes. When a signaling message is received at STP, MTP-3 at the STP uses the destination point code and routing table to forward the message on the correct outgoing link. MTP Level-3 also updates routing tables when a link goes down or comes up. SCCP, which corresponds to the transport layer, sits above MTP-3. Note that MTP-3 provides only datagram, that is, connectionless, service.*

*SCCP optionally provides connection, layered over MTP's datagrams. It also provides sequence numbering, segmentation and reassembly, and window flow control.*

One of the main applications on SS7 is the Telephone User Part (TCP), which is responsible for setting up voice calls. The TCP interprets dialed digits to route calls, reserves resources for calls, maintains accounting information, and provides services such as three-way calling and 800-number service. In modern telephone networks, the TCP has been replaced by the Integrated Service Digital Network-User Part (ISDN-UP), which subsumes TCP functionality and additionally provides 64 Kbps data channels. ISDN-UP uses both SCCP (when the higher-level functionality is needed) and MTP-3 (when performance is critical).

Another application layer that uses SCCP is the Transaction capabilities application part (TCAP), which provides remote procedure calls (allowing an application to invoke a procedure on a remote switch controller). TCAP thus makes it easier to develop distributed application, which are called Application Service Elements or ASEs in SS7 jargon. An example of an ASE is the operations Maintenance and
Administration Part-ASE (OMAP-ASE) which provides network management. OMAP-ASE allows network managers to ping a remote destination, trace routes to see if they are up or retrieve status information from remote switch controller. SS7 forms the basis not only for signaling but also for creating enhanced services.

### 2.2.2 Internet Protocols

In this section, we study some representative protocols in the Internet. IP is the workhorse of the Internet, and every Internet protocols must be layered at some point, over IP. ICMP, which is a control plane protocol, that works hand-in-hand switch with IP. TCP, which provides end-to-end reliable service, works with HTTP, which is the basis for the World Wide Web.

#### Internet Protocol Stack

Though the Internet protocol has only a few protocols defined at the network and transport layers, it has many more at the application layer. Table 2.2.2.1 shows a typical Internet Protocol Stack, where we have chosen the Hypertext Transfer Protocol (HTTP) as an example application-level protocol. The main protocols in the data plane are Internet Protocol (IP), Transmission Control Protocol (TCP), and User Datagram Protocol (UDP) which provide network and transport-layer services.

Table 2.2.2.1: Internet Protocol Stack

<table>
<thead>
<tr>
<th>Layer</th>
<th>Data Plane</th>
<th>Control Plane</th>
</tr>
</thead>
<tbody>
<tr>
<td>Application</td>
<td>Web browser</td>
<td>RSVP/OPF</td>
</tr>
<tr>
<td></td>
<td>HTTP</td>
<td></td>
</tr>
<tr>
<td>Session</td>
<td>Socket/streams</td>
<td></td>
</tr>
<tr>
<td>Transport</td>
<td>TCP/UDP</td>
<td></td>
</tr>
<tr>
<td>Network</td>
<td>IP</td>
<td>IP/ICMP</td>
</tr>
<tr>
<td>Data link</td>
<td>Many</td>
<td>Many</td>
</tr>
<tr>
<td>Physical</td>
<td>Many</td>
<td>Many</td>
</tr>
</tbody>
</table>
The Internet Protocol provides two network-layered services in the Internet. First, it provides every endpoint with an IP address. Second, packets injected into the Internet from an arbitrary endpoint are forwarded to the destination whose IP address is in the packet’s header. When forwarding packets, IP hides the details of link-level technologies from the endpoints and provides the abstraction of an unreliable, best effort, end-to-end link.

* By reliable, we mean that IP does not guarantee that a packet will reach its destination.

* By best-effort, we mean that IP does not guarantee packet a quality of service. However, if a packet is too large to be carried by a particular data link-layer protocol, IP will fragment the packet and reassemble it at its destination.

* Finally, by end-to-end, we mean that IP carries a packet from the source to the destination by consulting routing tables established by an independent routing protocol.

Because the Internet does not specify a physical or data link-layered standard, we can layer IP over practically any physical and data link-layer technology, such as Ethernet (CSMA/CD), FDDI (token ring), wide area telephone trunks (SONET or PDH), wireless links (CSMA/CD), satellite links (ALOHA), and other networking technologies such as X.25 and ISDN.

This allows the Internet to interconnect networks built with widely varying technologies, and allows network managers to upgrade the underlying technology without affecting higher-level protocols and applications. This is an important reason for the Internet's popularity.
We can understand IP's functionality by studying the IP header as shown in Table 2.2.2.2

<table>
<thead>
<tr>
<th>4-bit Version</th>
<th>4-bit header Length</th>
<th>8-bit type of service (TOS)</th>
<th>16-Bytes Total length</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>16-bit Identification</td>
<td>3-bit flags</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8-bit time to live (TTL)</td>
<td>13-bit fragment offset</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8-bit protocol</td>
<td>16-bit header checksum</td>
</tr>
<tr>
<td></td>
<td></td>
<td>32-bit source IP address</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>32-bit destination IP address</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Options (if any)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Data</td>
<td></td>
</tr>
</tbody>
</table>

The first field is the version number, which allows the network to upgrade to newer versions of IP.

The next field is the length of the header in 4-bytes words. Since the field is 4 bits long, the longest header length is \(2^4 - 1 = 15\) words or 60 bytes long. The header length field allows a router or endpoint to distinguish between header and data bytes.

The third field, type of service, or TOS, allows endpoints to choose a quality of service for their packets.

The fourth field, length is the length of the datagram in bytes. It is 16 bytes long so the longest IP packet is \((64\text{ Kbytes} - 1) = (535)\text{ bytes}\) long.
Fragmentation

The fifth field, ID, uniquely identifies an IP packet among those that have a given source and destination address. Some link technologies limit the packet size. For example, Ethernet does not allow packets larger than 1500 bytes. If a router receives an IP packet that is too large for the next hop, IP allows the router to fragment the packet into smaller packets. The destination then uses the ID field to reconstruct the original packet.

The next 3-bit field, flags, also supports fragmentation. Only 2 of the 3 bits are significant. One of them is the “More Fragments” flag. If this is 1, the destination knows that this packet is not the last fragment in the packet. It therefore places the fragment in a queue, waiting for a fragment with the “More Fragments” flag set to 0.

The other flag is the “Don’t Fragment” flag. If this is set, a router that gets a packet that is too large for the next hop must discard the packet. It also sends a control message using the Internet Control Message Protocol (ICMP) to the source, telling it that the packet was too large. A source that receives this message should reduce its packet size and transmit the packet, if necessary.

The fragment-offset field is also required for supporting fragmentation and reassembly. This field tells the destination which part of the original packet, described by a byte offset and length is contained in the current fragment.

TTL

The time-to-live (TTL) field serves many purposes. Current implementations decrement the TTL field each time a packet is forwarded, instead of once every second. As before, when the TTL field reaches zero, the packet is discarded. This is not quite right, because a router may take substantially less (or more) than one second to forward a packet. Thus, a source cannot strictly bound a packet’s maximum lifetime. Nevertheless,
even this use of the TTL field protects against looping packets in case of transient routing loops, which can form with both distance vector and link state routing.

We can exploit the TTL field to determine the route from a source to a destination. This technique depends on the fact that, when a router decrements a packet’s TTL to zero, it is supposed to send an ICMP “timer-expired” message to the source. A source that wants to find its path to a destination, first sets TTL to 1 and sends it to the destination. This evokes an error from the first router along the path. The source then sets the field to 2 and tries again, thus telling it the second router in the path. Continuing in this fashion, the source can find the entire path. This technique is used in the trace route. The TTL field is also used to clean up fragments at the destination.

**Remaining Fields**

The protocol field tells IP to which upper layer protocol it should pass the packet. This is typically TCP or UDP.

The header-checksum field protects against corruption in the packet header. IP does not checksum data. The checksum is a 16 bits one’s complement of the header, which is easy to implement in software and reasonably effective in practice.

The next two fields, source and destination IP address, are used to route packets to the destination and for the destination to reply to the source. These are required in any datagram network.

**Options**

The last set of fields, the option fields, allows further expansion of the IP header for special purposes. Five sets of options are commonly used.

- **Security and handling options** are mainly used in the military portion of the Internet.
- When the record route option is set, each router along the path stores its 4-byte IP address in the header’s option field. Since the header length field is 4 bits long and counts the number of 4-bytes words, the longest IP header is 15 words or 60 bytes.

- When the time stamp option is set, each router along the path adds its own IP address and a time stamp to the packet header.

- The loose source routing option allows a source to specify a set of routers that must be traversed in reaching the destination.

- Finally, the strict source routing option allows a source to choose the entire path to reach the destination.

**ICMP**

The Internet control message protocol carries error messages from the network to an IP source.

The ICMP header, shown in figure 2.2.2.1, contains only type, code and checksum fields. The type and code field encodes the error message. The following are some important error messages:

<table>
<thead>
<tr>
<th>8-bit type</th>
<th>8-bit code</th>
<th>16-bit checksum</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Content depend on type and code)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Figure 2.2.1: ICMP header*

- **Destination unreachable**: Sent by a router to a source when it cannot find a route.

- **Source quench**: Informs a source that it is sending data too fast, and should reduce its rate.
• **Redirect**: Informs a source to use another router for reaching a particular destination.

• **Router advertisement**: Advertises the presence of a router on a broadcast LAN.

• **Time exceeded**: A router has decremented a packet’s TTL to zero. Used by trace route.

• **Fragmentation needed but “Don’t-Fragment” bit set**: Used by path-MTU discovery.

Notice that ICMP is closely tied to IP. Thus, one can think of ICMP as the extension of IP for carrying error messages.

**TCP**

The Transmission Control Protocol (TCP) provides a multiplexed, duplex, connection oriented, reliable, flow-controlled, byte-stream service.

* By multiplexed, we mean that many applications can share access to a single TCP layers coordinate actions to correctly route packets to the desired destination.

* By duplex, we mean that both ends of a TCP connection can simultaneously read and write packets.

* By connection-oriented, we mean that TCP does connection establishment before data transfer in each direction.

* By reliable, we mean that TCP uses timeouts and retransmission to ensure that a destination receives a transmitted packet.

* By flow-controlled, we mean that a TCP source uses feedback flow control to adjust its transmission rate to the rate currently supportable in the network.

* Finally, by byte-stream, we mean that TCP provides the abstraction of carrying bytes in sequence from source to destination, ignoring message boundaries. A source may send two messages of length 50 and 100 bytes, which it can read in as many messages as it wants.
These aspects of TCP can be seen from the TCP packets header, shown in Figure 2.2.2.2

<table>
<thead>
<tr>
<th>0</th>
<th>15</th>
<th>16</th>
<th>31</th>
</tr>
</thead>
<tbody>
<tr>
<td>16-bit Source port Number</td>
<td>16-bit destination port Number</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32-bit sequence Number</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32-bit acknowledge Number</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-bit header</td>
<td>reserved (6-bit)</td>
<td>U</td>
<td>A</td>
</tr>
<tr>
<td>Length</td>
<td>G</td>
<td>K</td>
<td>H</td>
</tr>
<tr>
<td>16-bit TCP check sum</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16-bit urgent</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Options (if any)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data (if any)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 2.2.2.2: TCP packet header

**Port Numbers**

The first two fields in the TCP header are the source and destination port numbers. A single interface, identified by a single IP address, may carry data from many different applications (think of a server that replies to simultaneous requests from many clients). To differentiate between these applications sharing the same IP address, each is assigned a different port number by the operating system. The destination port number therefore identifies the application program that should receive the incoming TCP packet. The source port number allows the application to reply to the sender. The four quantities, the source IP address, source port number, destination IP address, and destination port number, uniquely identify a TCP connection.
Some port numbers are ‘well known’ in the sense that end-systems always assign them to certain common applications. For example, the Telnet remote login application uses port 23, the simple Mail Transfer Protocol uses port 25, and the Hypertext Transfer Protocol uses port 80.

**Sequence and Acknowledgement Number**

The 32-bit TCP sequence number is the byte offset of the start of the packet from the initial sequence number used in the stream. The acknowledgement number uses the same format as the sequence number and is a cumulative acknowledgement used for providing reliable transmission.

**Header Length**

We need the header length field because the TCP header is variable length and, like the IP header, may contain several option fields.

**Flags**

The TCP header has six flag bits. The first bit, URG, indicates that the urgent pointer is valid. The ACK bit indicates that the acknowledgement field contains valid information, that is, the packet contains a piggybacked acknowledgement. The PSH (push) bit, request the receiver to hand over the packet to the application program as soon as possible. The RST bit resets the receiver during a three-way handshake. The SYN and FIN bits are also used during a three-way handshake.

**Window size**

The window size field allows a receiver to control how much data the sender is allowed to send before it must wait for an acknowledgement. It is usually set to the size of the buffer at the receiver that holds out-of-order packets. The last byte the sender may send is the acknowledgement sequence number added to the window size. Since a 16-bit
field describes this size, the largest possible flow control window size is 64 Kbytes – 1, which may be too small for a long delay, high-bandwidth connection. In this case, the window size can be scaled. When a sender wants to use window scaling, during the three-way handshake, an option in its SYN packet proposes that, window sizes be interpreted as the most significant bits of a large window size. If the receiver agrees, then both sides can use a much larger window size. For example, if the endpoints agree to treat the window as the MSB of a 20 bit window (by shifting the size 4 positions to the left before interpreting it), the effective largest window size increases to 1 MB.

The window size depends on the amount of buffer space left in the receiver. If the receive buffer is full, then the receiver sets the window size to zero, preventing the sender from sending any more data. The sender periodically sends a 1-byte packet, checking to see whether the window size has increased. This is called sender persistence, and the timer used to send a probe packet is called the persist timer.

Checksum

The TCP checksum is computed not only over the TCP header, but also over the data in the packet. Thus, the TCP checksum protects against corruption in the data packet.

Urgent pointer

This points to the last byte in the packet that contains ‘urgent’ data. Bytes from the start of the packet to the byte pointed to by the urgent pointer are considered urgent. The receiving operating system should pass this data to the application as soon as it can.

Options

The most common option is the choice of the maximum segment size (MSS). In TCP jargon, each packet is called a segment, and the MSS is the largest segment a sender may send. If a receiver wants to control the MSS of the sender, it advertises it in the option field during the three-way handshake.
HTTP

We focus on the Hypertext Transfer Protocol (HTTP) as an example of a common application-layer protocol on the Internet. Its use in the World Wide Web is nearly synonymous with the Internet.

HTTP is request-response protocol. Servers are passive entities that reply to client requests. Much of the complexity in the World Wide Web is in the client's browser software. A browser needs to decide when and where to send a request, and how to interpret the response (for example, uncompressing and displaying graphics, or playing a sound file). Since HTTP is layered over TCP, it uses the error-control, flow-control, and in sequence ordering features of TCP. HTTP merely provides a way to represent queries and replies.

An important concept in HTTP is the notion of a Universal Resource Locator (URL). A URL is "a way to encapsulate a name in any registered namespace", and label it with the namespace, producing a member of the universal set of URLs. An example URL is http://www.aw.com/ where http is the name space, and www.aw.com is a path in this name space that, by convention, refers to the directory associated with the HTTP protocol on the machine named www.aw.com. Similarly, in the URL mail to: skeshav@cs.cornell.edu, mail to is the name space, and skeshav@cs.cornell.edu is a name in this space that corresponds to an address to which email can be sent. Loosely speaking, a URL is a way to refer to an arbitrary resource on a server, such as a file or a server, so that it can be accessed by a client. URLs are always of the form <name space> • <path>, where both the namespace and the path use 7-bit ASCII characters. This allows the URL space to be extensible, complete, and printable.

An HTTP request is a GET, HEAD, or POST request. The GET request is of the form GET request URL, followed by zero or one modifiers on subsequent lines.
Modifiers include restrictions such as ‘If-Modified-Since’ and ‘Authorization’. The former allows a client to retrieve a resource only if it has been modified since a given date. The latter allows the server to authenticate the client. The HEAD request is identical to a GET, except that server returns only a short header instead of actual resources. A client uses it to test the validity of a link. Finally, a POST request of the form POST URL <blank line> body allows a client to give the server some information to post to some location in the server, which might correspond to a form or a bulletin board.

An HTTP response is of the form status line <line break> header<line break><blank line> body allows a client to give the server some information to post to some location in the server, which might correspond to a form or a bulletin board.

An HTTP response is of the form status line <line break> headers <line break><blank line><body>. The status describes the server’s version of HTTP and contains some standard response course. The headers modify the response, for example, with an expire field which indicates how long that information is valid. The actual text of the response is in the body field.

After studying the two technologies (the telephone and Internet technologies), let us study different teletraffic models in chapter 3.