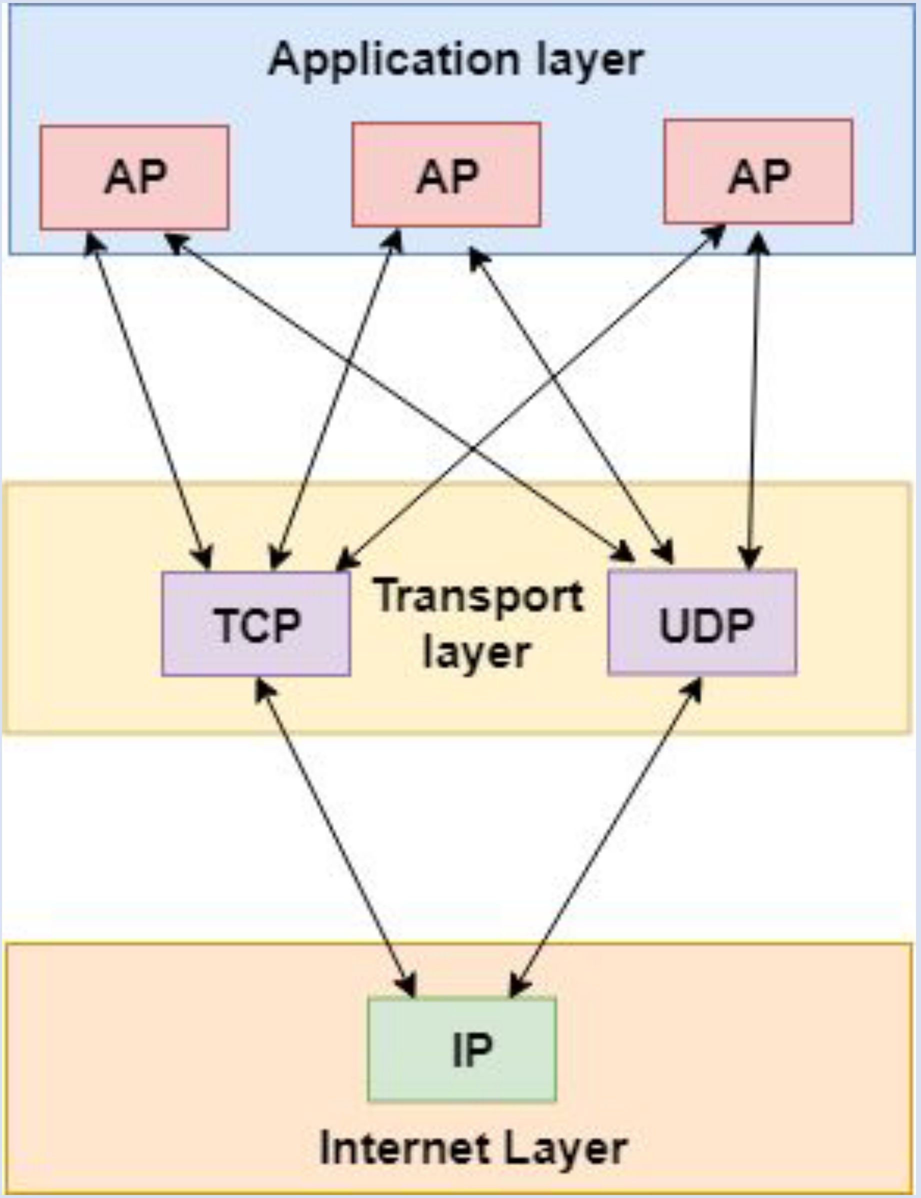
Transport Layer

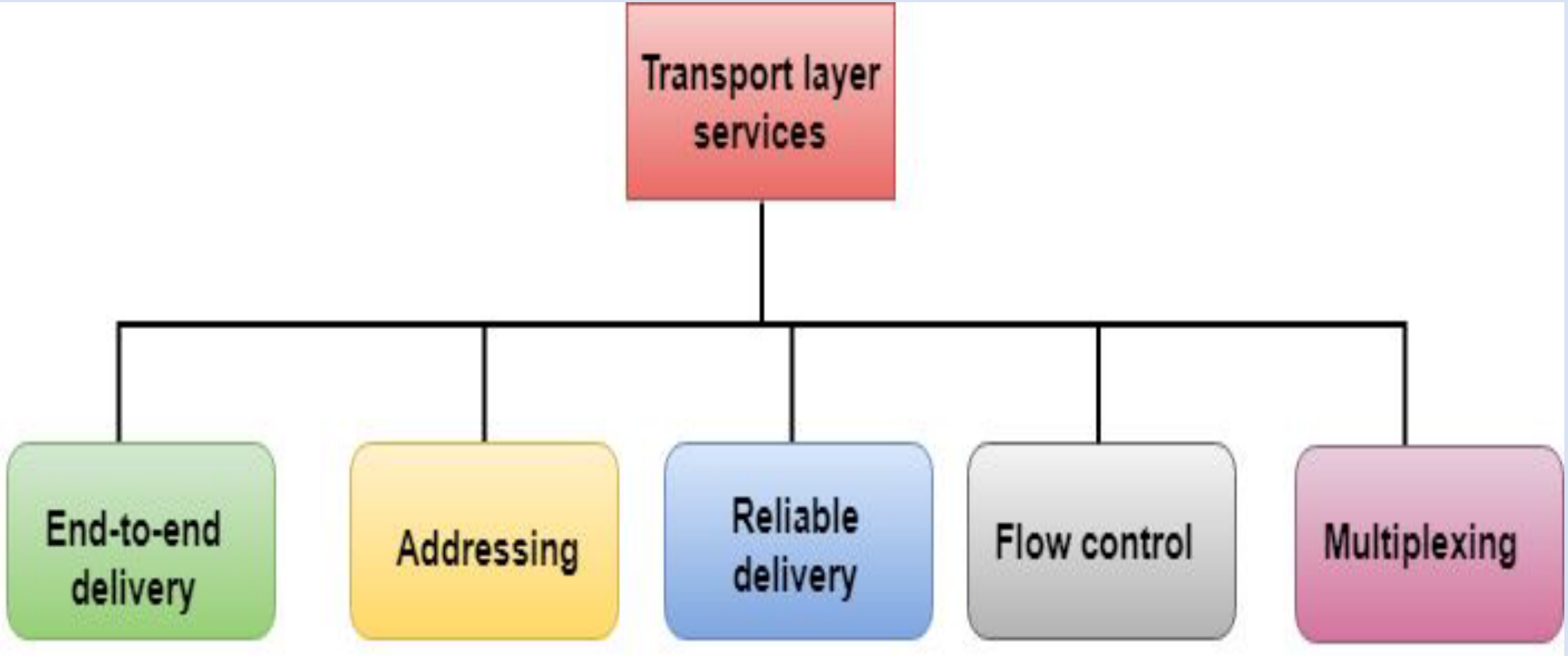
- Transport Layer Services,
- UDP: Datagram services
- Applications
- Berkley Sockets
- Addressing
- Connection establishment
- Connection release
- Flow control and buffering
- Multiplexing
- TCP: Services, Features,
- Segment,
- TCP Timer management,
- TCP Congestion Control,
- Real Time Transport protocol(RTP),
- Stream Control Transmission Protocol (SCTP),
- Quality of Service (QoS),
- Differentiated services, TCP and UDP for Wireless.

- The transport layer is a 4th layer from the top.
- The main role of the transport layer is to provide the communication services directly to the application processes running on different hosts.
- The transport layer provides a logical communication between application processes running on different hosts. Although the application processes on different hosts are not physically connected, application processes use the logical communication provided by the transport layer to send the messages to each other.
- The transport layer protocols are implemented in the end systems but not in the network routers.
- A computer network provides more than one protocol to the network applications. For example, TCP and UDP are two transport layer protocols that provide a different set of services to the network layer.
- All transport layer protocols provide multiplexing/demultiplexing service. It also provides other services such as reliable data transfer, bandwidth guarantees, and delay guarantees.
- Each of the applications in the application layer has the ability to send a message by using TCP or UDP. The application communicates by using either of these two protocols. Both TCP and UDP will then communicate with the internet protocol in the internet layer. The applications can read and write to the transport layer. Therefore, we can say that communication is a two-way process.


Application layer
AP
AP
AP
TCP
Transport layer
UDP
IP
Internet Layer

Services provided by the Transport Layer

- The services provided by the transport layer are similar to those of the data link layer.
- The data link layer provides the services within a single network while the transport layer provides the services across an internetwork made up of many networks.
- The data link layer controls the physical layer while the transport layer controls all the lower layers.



Transport layer services

End-to-end delivery

Addressing

Reliable delivery

Flow control

Multiplexing

End-to-end delivery:

- The transport layer transmits the entire message to the destination. Therefore, it ensures the end-to-end delivery of an entire message from a source to the destination.
- Reliable delivery:
- The transport layer provides reliability services by retransmitting the lost and damaged packets.
- **The reliable delivery has four aspects:**
- Error control
- Sequence control
- Loss control
- Duplication control

Error Control

- The primary role of reliability is **Error Control**. In reality, no transmission will be 100 percent error-free delivery. Therefore, transport layer protocols are designed to provide error-free transmission.
- The data link layer also provides the error handling mechanism, but it ensures only node-to-node error-free delivery. However, node-to-node reliability does not ensure the end-to-end reliability.
- The data link layer checks for the error between each network. If an error is introduced inside one of the routers, then this error will not be caught by the data link layer. It only detects those errors that have been introduced between the beginning and end of the link. Therefore, the transport layer performs the checking for the errors end-to-end to ensure that the packet has arrived correctly.

Sequence Control

- The second aspect of the reliability is sequence control which is implemented at the transport layer.
- On the sending end, the transport layer is responsible for ensuring that the packets received from the upper layers can be used by the lower layers.
- On the receiving end, it ensures that the various segments of a transmission can be correctly reassembled.

Loss Control

- Loss Control is a third aspect of reliability. The transport layer ensures that all the fragments of a transmission arrive at the destination, not some of them. On the sending end, all the fragments of transmission are given sequence numbers by a transport layer. These sequence numbers allow the receivers transport layer to identify the missing segment.

Duplication Control

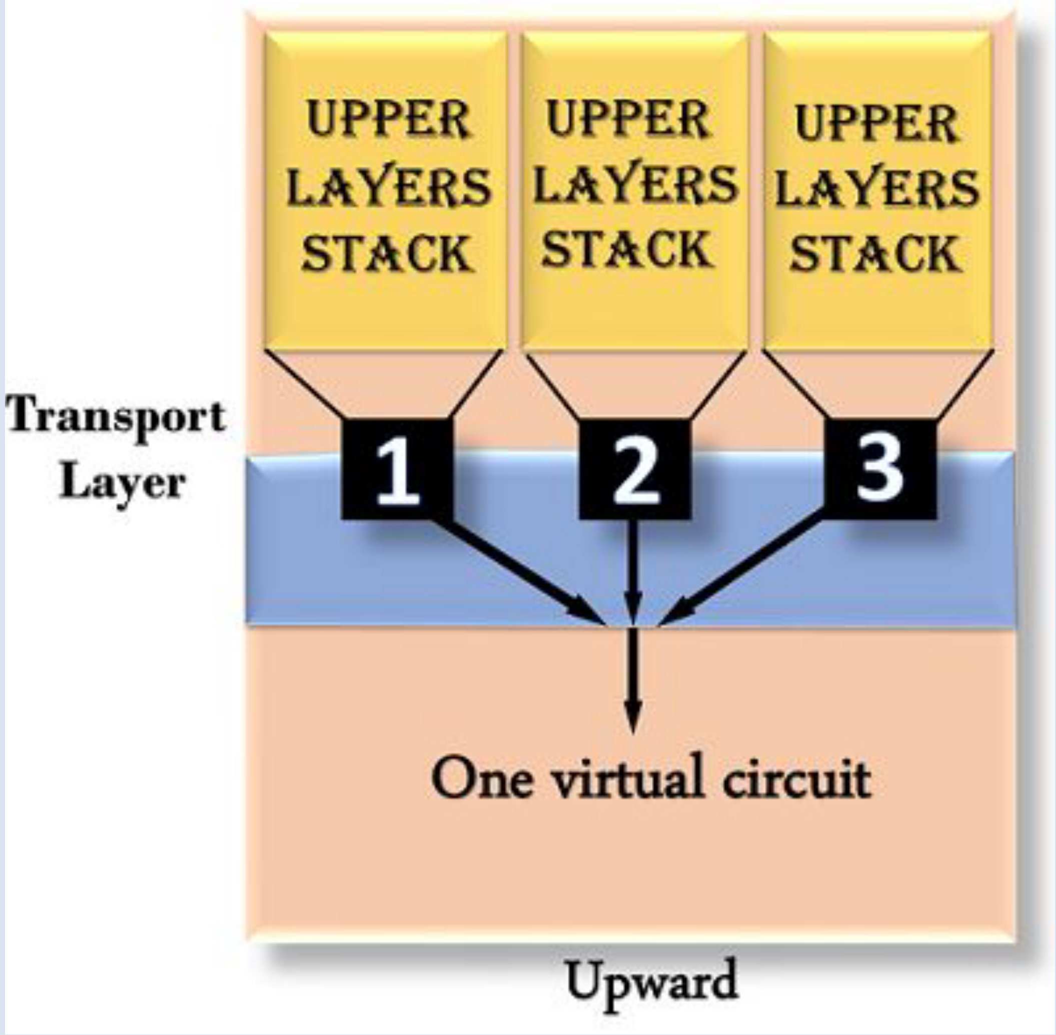
- Duplication Control is the fourth aspect of reliability. The transport layer guarantees that no duplicate data arrive at the destination. Sequence numbers are used to identify the lost packets; similarly, it allows the receiver to identify and discard duplicate segments.

Flow Control

- Flow control is used to prevent the sender from overwhelming the receiver.
- If the receiver is overloaded with too much data, then the receiver discards the packets and asking for the retransmission of packets.
- This increases network congestion and thus, reducing the system performance.
- The transport layer is responsible for flow control.
- It uses the sliding window protocol that makes the data transmission more efficient as well as it controls the flow of data so that the receiver does not become overwhelmed.
- Sliding window protocol is byte oriented rather than frame oriented.

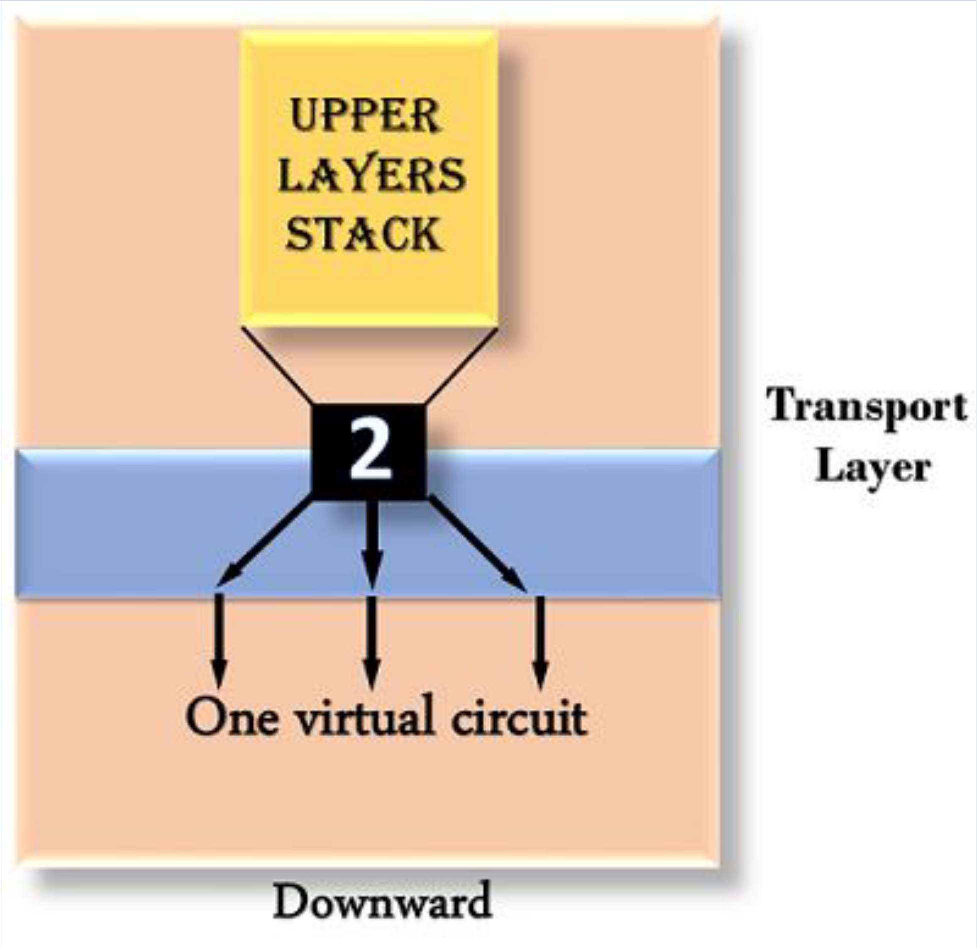
Multiplexing

- The transport layer uses the multiplexing to improve transmission efficiency.
- **Multiplexing can occur in two ways:**
- **Upward multiplexing:** Upward multiplexing means multiple transport layer connections use the same network connection. To make more cost-effective, the transport layer sends several transmissions bound for the same destination along the same path; this is achieved through upward multiplexing.



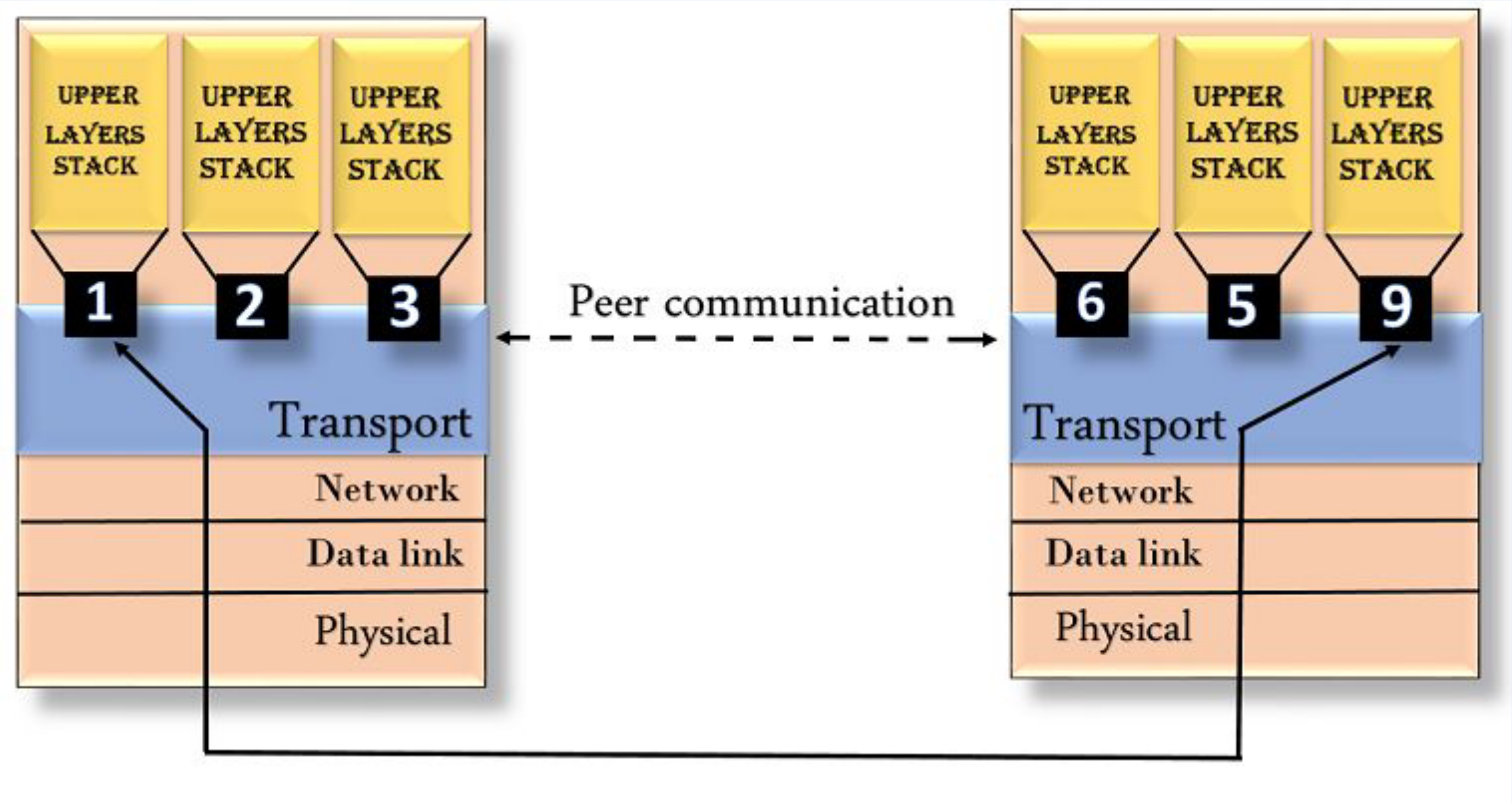
Downward multiplexing:

- Downward multiplexing means one transport layer connection uses the multiple network connections. Downward multiplexing allows the transport layer to split a connection among several paths to improve the throughput. This type of multiplexing is used when networks have a low or slow capacity.



Addressing

- According to the layered model, the transport layer interacts with the functions of the session layer.
- Many protocols combine session, presentation, and application layer protocols into a single layer known as the application layer.
- In these cases, delivery to the session layer means the delivery to the application layer.
- Data generated by an application on one machine must be transmitted to the correct application on another machine. In this case, addressing is provided by the transport layer.
- The transport layer provides the user address which is specified as a station or port.
- The port variable represents a particular TS user of a specified station known as a Transport Service access point (TSAP).
- Each station has only one transport entity.
- The transport layer protocols need to know which upper-layer protocols are communicating.


UPPER LAYERS STACK
UPPER LAYERS STACK
UPPER LAYERS STACK
1
2
3
Transport
Network
Data link
Physical
Peer communication
UPPER LAYERS STACK
UPPER LAYERS STACK
UPPER LAYERS STACK
6
5
9
Transport
Network
Data link
Physical

Transport Layer protocols

- The transport layer is represented by two protocols: TCP and UDP.
- The IP protocol in the network layer delivers a datagram from a source host to the destination host.
- Nowadays, the operating system supports multiuser and multiprocessing environments, an executing program is called a process. When a host sends a message to other host means that source process is sending a process to a destination process. The transport layer protocols define some connections to individual ports known as protocol ports.
- An IP protocol is a host-to-host protocol used to deliver a packet from source host to the destination host while transport layer protocols are port-to-port protocols that work on the top of the IP protocols to deliver the packet from the originating port to the IP services, and from IP services to the destination port.
- Each port is defined by a positive integer address, and it is of 16 bits.

UDP: **User Datagram Protocol**.

- UDP stands for UDP is a simple protocol and it provides non sequenced transport functionality.
- UDP is a connectionless protocol.
- This type of protocol is used when reliability and security are less important than speed and size.
- UDP is an end-to-end transport level protocol that adds transport-level addresses, checksum error control, and length information to the data from the upper layer.
- The packet produced by the UDP protocol is known as a user datagram.

User Datagram Format

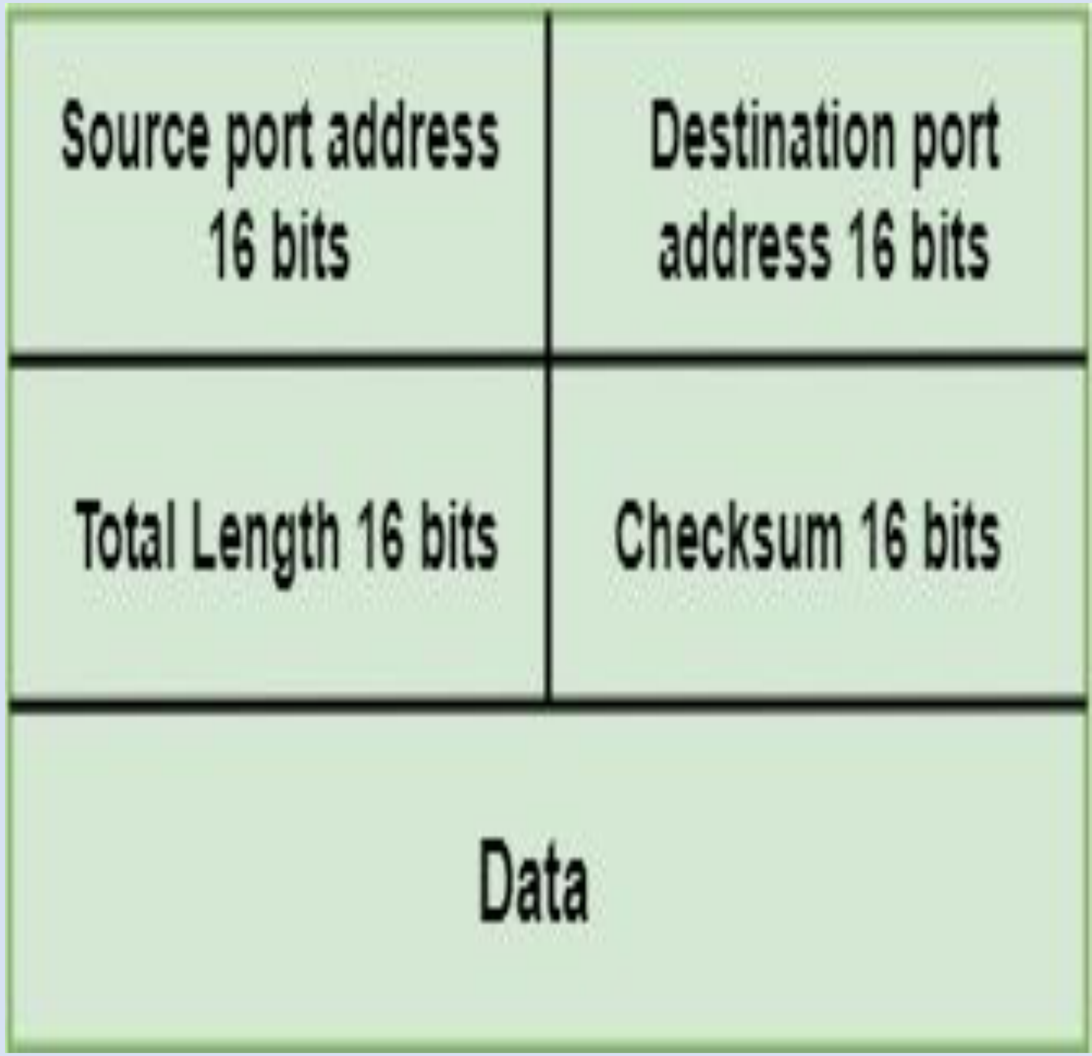
Where,

Source port address: It defines the address of the application process that has delivered a message. The source port address is of 16 bits address.

Destination port address: It defines the address of the application process that will receive the message. The destination port address is of a 16-bit address.

Total length: It defines the total length of the user datagram in bytes. It is a 16-bit field.

Checksum: The checksum is a 16-bit field which is used in error detection.



Source port address 16 bits	Destination port address 16 bits
Total Length 16 bits	Checksum 16 bits
Data	

Disadvantages of UDP protocol

- UDP provides basic functions needed for the end-to-end delivery of a transmission.
- It does not provide any sequencing or reordering functions and does not specify the damaged packet when reporting an error.
- UDP can discover that an error has occurred, but it does not specify which packet has been lost as it does not contain an ID or sequencing number of a particular data segment.

UDP application

- Domain Name Services
- Simple Network Management Protocol
- Trivial File Transfer Protocol
- Routing Information Protocol
- Kerberos

TCP

- TCP stands for Transmission Control Protocol.
- It provides full transport layer services to applications.
- It is a connection-oriented protocol means the connection established between both the ends of the transmission. For creating the connection, TCP generates a virtual circuit between sender and receiver for the duration of a transmission.

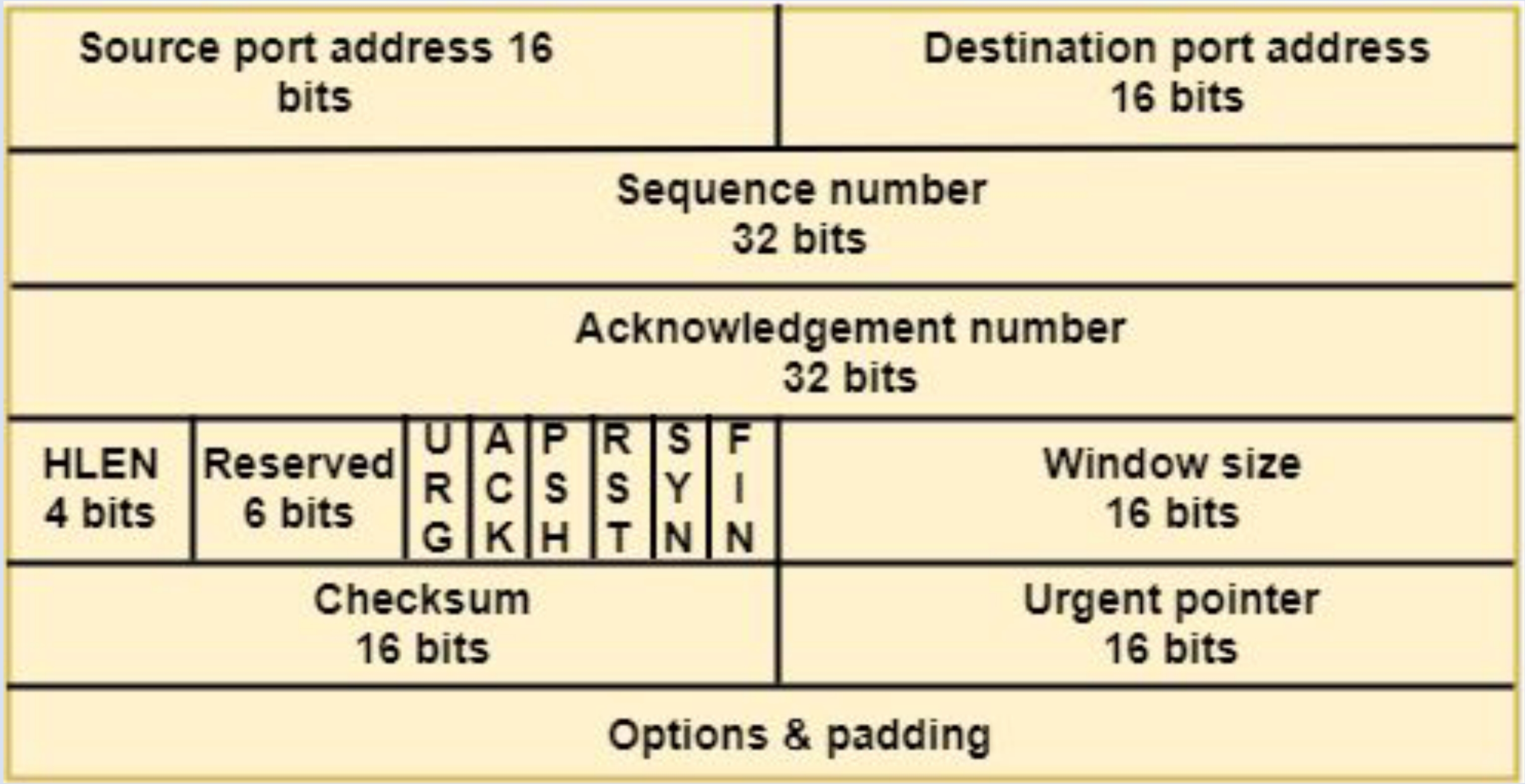
Features Of TCP protocol

- **Stream data transfer:** TCP protocol transfers the data in the form of contiguous stream of bytes. TCP group the bytes in the form of TCP segments and then passed it to the IP layer for transmission to the destination. TCP itself segments the data and forward to the IP.
- **Reliability:** TCP assigns a sequence number to each byte transmitted and expects a positive acknowledgement from the receiving TCP. If ACK is not received within a timeout interval, then the data is retransmitted to the destination. The receiving TCP uses the sequence number to reassemble the segments if they arrive out of order or to eliminate the duplicate segments.

- **Flow Control:** When receiving TCP sends an acknowledgement back to the sender indicating the number the bytes it can receive without overflowing its internal buffer. The number of bytes is sent in ACK in the form of the highest sequence number that it can receive without any problem. This mechanism is also referred to as a window mechanism.
- **Multiplexing:** Multiplexing is a process of accepting the data from different applications and forwarding to the different applications on different computers. At the receiving end, the data is forwarded to the correct application. This process is known as demultiplexing. TCP transmits the packet to the correct application by using the logical channels known as ports.
- **Logical Connections:** The combination of sockets, sequence numbers, and window sizes, is called a logical connection. Each connection is identified by the pair of sockets used by sending and receiving processes.

- **Full Duplex:** TCP provides Full Duplex service, i.e., the data flow in both the directions at the same time. To achieve Full Duplex service, each TCP should have sending and receiving buffers so that the segments can flow in both the directions. TCP is a connection-oriented protocol. Suppose the process A wants to send and receive the data from process B. The following steps occur:
    - Establish a connection between two TCPs.
    - Data is exchanged in both the directions.
    - The Connection is terminated.

TCP Segment Format

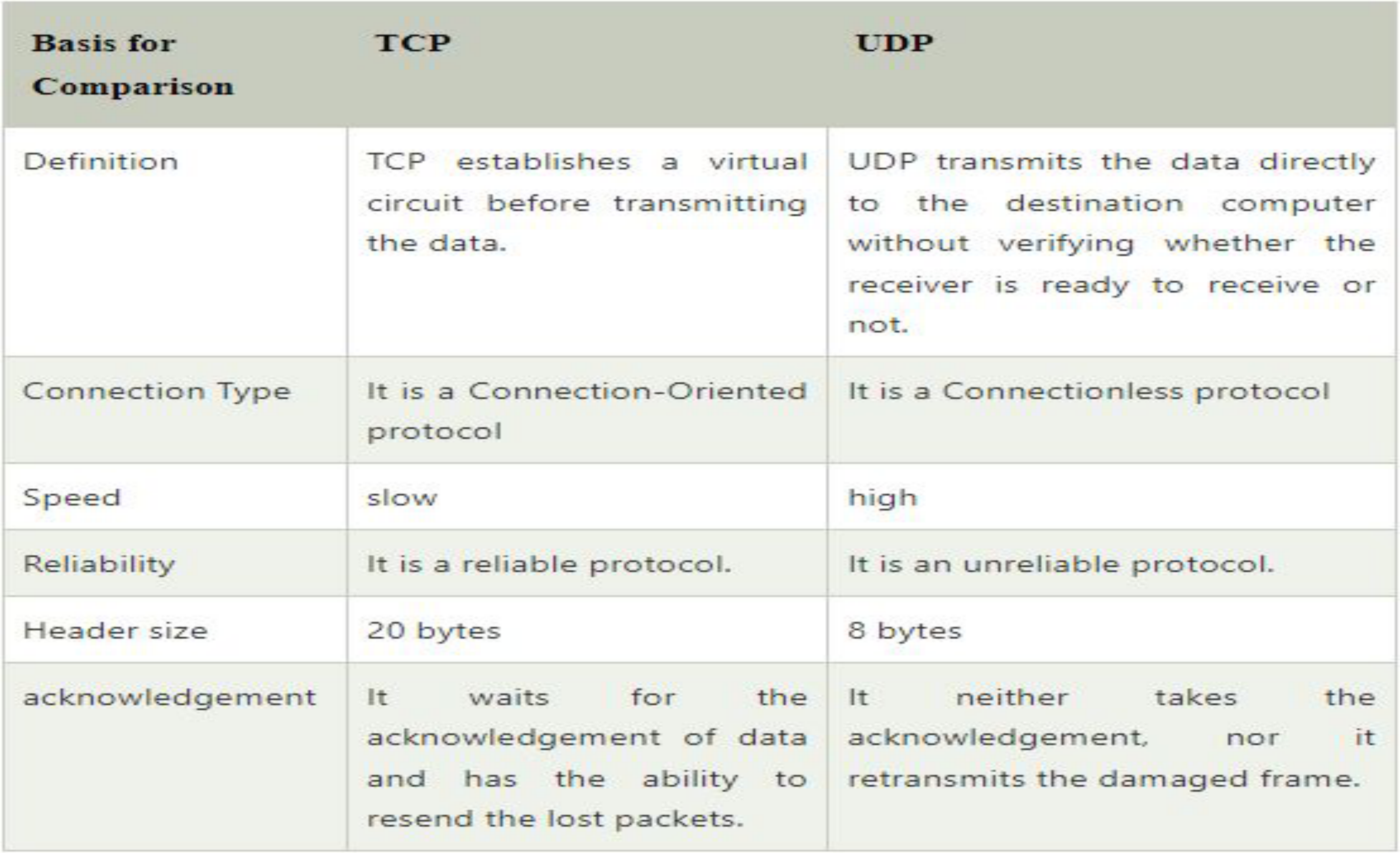


Source port address 16 bits								Destination port address 16 bits
Sequence number 32 bits								
Acknowledgement number 32 bits								
HLEN 4 bits	Reserved 6 bits	URG	ACK	PSH	RST	SYN	FIN	Window size 16 bits
Checksum 16 bits								Urgent pointer 16 bits
Options & padding								

- **Source port address:** It is used to define the address of the application program in a source computer. It is a 16-bit field.
- **Destination port address:** It is used to define the address of the application program in a destination computer. It is a 16-bit field.
- **Sequence number:** A stream of data is divided into two or more TCP segments. The 32-bit sequence number field represents the position of the data in an original data stream.
- **Acknowledgement number:** A 32-field acknowledgement number acknowledge the data from other communicating devices. If ACK field is set to 1, then it specifies the sequence number that the receiver is expecting to receive.
- **Header Length (HLEN):** It specifies the size of the TCP header in 32-bit words. The minimum size of the header is 5 words, and the maximum size of the header is 15 words. Therefore, the maximum size of the TCP header is 60 bytes, and the minimum size of the TCP header is 20 bytes.
- **Reserved:** It is a six-bit field which is reserved for future use.
- **Control bits:** Each bit of a control field functions individually and independently. A control bit defines the use of a segment or serves as a validity check for other fields.

- There are total six types of flags in control field:
- **URG:** The URG field indicates that the data in a segment is urgent.
- **ACK:** When ACK field is set, then it validates the acknowledgement number.
- **PSH:** The PSH field is used to inform the sender that higher throughput is needed so if possible, data must be pushed with higher throughput.
- **RST:** The reset bit is used to reset the TCP connection when there is any confusion occurs in the sequence numbers.
- **SYN:** The SYN field is used to synchronize the sequence numbers in three types of segments: connection request, connection confirmation (with the ACK bit set), and confirmation acknowledgement.
- **FIN:** The FIN field is used to inform the receiving TCP module that the sender has finished sending data. It is used in connection termination in three types of segments: termination request, termination confirmation, and acknowledgement of termination confirmation.
    - **Window Size:** The window is a 16-bit field that defines the size of the window.
    - **Checksum:** The checksum is a 16-bit field used in error detection.
    - **Urgent pointer:** If URG flag is set to 1, then this 16-bit field is an offset from the sequence number indicating that it is a last urgent data byte.
    - **Options and padding:** It defines the optional fields that convey the additional information to the receiver.

Differences b/w TCP & UDP



Basis for Comparison	TCP	UDP
Definition	TCP establishes a virtual circuit before transmitting the data.	UDP transmits the data directly to the destination computer without verifying whether the receiver is ready to receive or not.
Connection Type	It is a Connection-Oriented protocol	It is a Connectionless protocol
Speed	slow	high
Reliability	It is a reliable protocol.	It is an unreliable protocol.
Header size	20 bytes	8 bytes
acknowledgement	It waits for the acknowledgement of data and has the ability to resend the lost packets.	It neither takes the acknowledgement, nor it retransmits the damaged frame.

Real-time Transport Protocol

- The **Real-time Transport Protocol** (**RTP**) is a network protocol for delivering audio and video over IP networks. RTP is used in communication and entertainment systems that involve streaming media, such as telephony, video teleconference applications including WebRTC, television services and web-based push-to-talk features.
- RTP typically runs over User Datagram Protocol (UDP). RTP is used in conjunction with the RTP Control Protocol (RTCP). While RTP carries the media streams (e.g., audio and video), RTCP is used to monitor transmission statistics and quality of service (QoS) and aids synchronization of multiple streams. RTP is one of the technical foundations of Voice over IP and in this context is often used in conjunction with a signaling protocol such as the Session Initiation Protocol (SIP) which establishes connections across the network.

Stream Control Transmission Protocol (SCTP)

- SCTP stands for Stream Control Transmission Protocol. It is a new reliable, message oriented transport layer protocol. SCTP, however, is mostly designed for Internet applications that have recently been introduced.
- SCTP combines the best features of UDP and TCP. SCTP is a reliable message-oriented protocol. It preserves the message boundaries, and at the same time, detects lost data, duplicate data, and out-of-order data. It also has congestion control and flows control mechanisms.

Features of SCTP

Transmission Sequence Number

- The unit of data in TCP is a byte. Data transfer in TCP is controlled by numbering bytes by using a sequence number. On the other hand, the unit of data in SCTP is a DATA chunk that may or may not have a one-to-one relationship with the message coming from the process because of fragmentation.
- Stream Identifier
- In TCP, there is only one stream in each connection. In SCTP, there may be several streams in each association. Each stream in SCTP needs to be identified by using a stream identifier (SI). Each data chunk must carry the SI in its header so that when it arrives at the destination, it can be properly placed in its stream. The 51 is a 16-bit number starting from O.
- Stream Sequence Number
- When a data chunk arrives at the destination SCTP, it is delivered to the appropriate stream and in the proper order. This means that, in addition to an SI, SCTP defines each data chunk in each stream with a stream sequence number (SSN).
- Packets

- In TCP, a segment carries data and control information. Data is carried as a collection of bytes; control information is defined by six control flags in the header. The design of SCTP is totally different: data is carried as data chunks; control information is carried as control chunks.
- Flow Control
- Like TCP, SCTP implements flow control to avoid overwhelming the receiver.
- Error Control
- Like TCP, SCTP implements error control to provide reliability. TSN numbers and acknowledgement numbers are used for error control.
- Congestion Control
- Like TCP, SCTP implements congestion control to determine how many data chunks can be injected into the network.

Quality of Service (QoS)

- Quality of service (QoS) is **the use of mechanisms or technologies that work on a network to control traffic and ensure the performance of critical applications with limited network capacity**. It enables organizations to adjust their overall network traffic by prioritizing specific high-performance applications.
- **Need for QoS –**
- Video and audio conferencing require bounded delay and loss rate.
- Video and audio streaming requires bounded packet loss rate, it may not be so sensitive to delay.
- Time-critical applications (real-time control) in which bounded delay is considered to be an important factor.
- Valuable applications should be provided better services than less valuable applications.

- Differentiated services (DiffServ) is one technique for implementing Quality of Service (QoS) policies. Using DiffServ in your network allows you to directly configure the relevant parameters on the switches and routers rather than using a resource reservation protocol.
- **QoS Specification –**
  QoS requirements can be specified as:
- Delay
- Delay Variation(Jitter)
- Throughput
- Error Rate

- There are two types of QoS Solutions:
- **Stateless Solutions –**
  Routers maintain no fine-grained state about traffic, one positive factor of it is that it is scalable and robust. But it has weak services as there is no guarantee about the kind of delay or performance in a particular application which we have to encounter.
- **Stateful Solutions –**
  Routers maintain a per-flow state as flow is very important in providing the Quality-of-Service i.e. providing powerful services such as guaranteed services and high resource utilization, providing protection, and is much less scalable and robust.