Networking Basics 3: TCP/IP Protocols

1. TCP/IP Introduction
2. Routing and Ports
3. **TCP/IP Protocols**
   - **IP**
   - **UDP**
   - **TCP**
   - **SCTP**
   - **ICMP**
4. Ethernet
5. DNS

The Internet Protocol (IP)

IP creates a *virtual network* that hides the characteristics of the underlying physical network(s).

It is an **unreliable, connectionless packet delivery protocol**:

- information is sent in **packets** called **datagrams**
- does not guarantee connection to target machine
- packets may not arrive or may be duplicated
- packets may arrive out of order
- no **flow control** (e.g., if network congestion)
- transport-level protocol must handle these issues

IP (contd.)

Each IPv4 packet consists of two basic parts:

- **header**
  - 20 bytes long
  - source and destination IP addresses
  - protocol number/type: e.g., TCP, UDP, ICMP
  - length, checksum, **time-to-live** (TTL)
  - **fragmentation** flags and fragment offset
- **data section/field (body)**
  - *encapsulates* (contains) a packet for a transport-level protocol like TCP
  - total packet size limited to 65k bytes

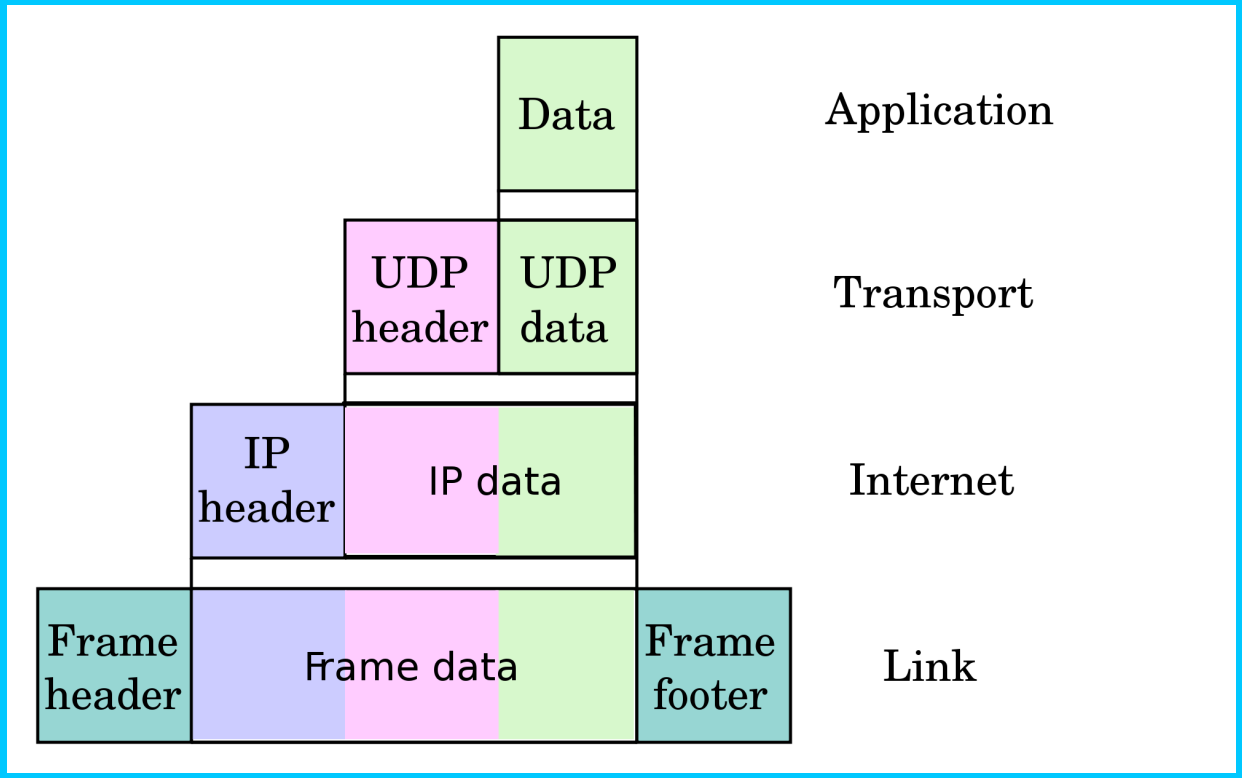
IP (contd.)

Each IPv6 packet consists of two basic parts:

- **header**
  - 40 bytes long
  - source and destination IP addresses
  - protocol number/type: e.g., TCP, UDP, ICMP
  - payload length, checksum, **hop limit**
  - **traffic class** (priority number)
- **payload**
  - *encapsulates* (contains) a packet for a transport-level protocol like TCP
  - total packet size up to 4GB ("**jumbograms**")

Encapsulation

Higher-layer packets become the data/payload component of lower-layer packets: (figure from Wikipedia)



Transport-Level Protocols

The two most popular transport-level protocols that are layered on top of IP are:

- **Transmission Control Protocol** (**TCP**)
- **User Datagram Protocol** (**UDP**)

Other transport-level protocols:

- **Stream Control Transmission Protocol** (**SCTP**)
- **Datagram Congestion Control Protocol** (**DCCP**)

(SCTP and DCCP are supported in Linux, but not widely used.)

UDP: User Datagram Protocol

UDP is an **unreliable, connectionless, message-based** protocol:

- basically an application interface to IP
- adds concept of ports (processes) to IP datagrams
- can be viewed as a a **multiplexer/demultiplexer** for IP datagrams/packets
- i.e., receives/sends packets from/to the appropriate ports
- issues of dropped, duplicated, out of order packets must be handled by the application if important
- **message-based** (each `sendto` results in a separate datagram and each `recv` reads one datagram's data)

UDP (contd.)

UDP **datagram** ("packet") format is simple:

- source and destination ports
- length (of datagram)
- checksum
- data

UDP less costly, useful for small messages.

Standard servers that use UDP:
DNS, RPC, SNMP, NTP, etc.

Lost messages typically "handled" simply in that requests will be repeated if no response is received.

TCP: Transmission Control Protocol

TCP is a **reliable, connection-oriented, stream-based** protocol:

- acts as if there is a *dedicated two-way connection* between the two hosts
- verifies machines connected/reachable
- automatically takes care of dropped, duplicated, and out of order packets (so applications need not handle)
- supports flow control and congestion control
- **stream-based** (creates apparent byte stream, so appears to be like reading/writing from/to a file)

TCP (contd.)

TCP **segment** ("packet") format more complex:

- source and destination ports
- **sequence number**
- **acknowledgment number**
- **flags**: SYN, ACK, FIN, RST, URG, PSH
- checksum
- **window size** (number of bytes sender can receive)
- options: MSS, window size scaling, SACK, etc.
- data/payload

TCP (contd.)

The sequence and acknowledgement numbers allow:

- duplicated packets to be detected/ignored by receiver
- dropped packets to be detected and retransmitted
- out-of-order packets to be buffered and reordered

TCP is a more costly protocol than UDP, in order to guarantee connection, data reception, order, etc.

TCP is the transport protocol most widely used in servers: FTP, SSH, HTTP, POP, IMAP, SMTP, etc.

TCP (contd.)

Flags denote purpose of packet:

- **SYN** – synchronize sequence numbers (start connection)
- **ACK** – acknowledgement (part of ongoing connection)
- **FIN** – finish/close
- **RST** – reset connection (i.e., error)
- **URG** – urgent pointer is valid
- **PSH** – push (unsent) data to application

Only certain flag combinations are valid, and illegal combinations are often used to probe machines for servers and OS version.

TCP (contd.)

Three-way handshake establishes TCP connection:

1. Process 1 sends: SYN w/seq=n
2. Process 2 sends: SYN+ACK w/seq=m, ack=n+1
3. Process 1 sends: ACK w/ack=m+1

Four-way handshake terminates connection:

1. Process 1 sends: FIN
2. Process 2 sends: ACK
3. Process 2 sends: FIN
4. Process 2 sends: ACK

TCP (contd.)

TCP ports can be in many **states**:

- **LISTEN** – server end only, waiting for connection
- **ESTABLISHED** – after connection established
- **CLOSE_WAIT** – received FIN, sent ACK, waiting for local process to terminate to send FIN
- **TIME_WAIT** – sent final closing ACK, but waiting to be certain no duplicates will be received, by default wait 2MSL (twice the **Maximum Segment Lifetime**—i.e., the maximum roundtrip transmission time)
- **CLOSED**

SCTP

SCTP (**Stream Control Transmission Protocol**) is a relatively new IP transport protocol (2000).

It combines features of TCP and UDP, but also has several unique features:

- **message-based** like UDP
- **reliable** like TCP
- supports either **connection-oriented** or **connectionless** mode
- able to read *out of order data* (avoids blocking issues)
- **multi-homing**: multiple IP addresses for endpoints
- **multi-streaming**: multiple streams within connection
- four-way init handshake with cookies prevents attacks

Internet Control Message Prot. (ICMP)

ICMP packets are used to report routing errors and send basic routing information.

ICMP acts like a transport-level protocol (layered on IP) but is an integral part of IP.

Key ICMP message types:

- **Echo Request** – `ping` request
- **Echo** – `ping` response
- **Desination Unreachable** – router cannot route packet
- **Time Exceeded** – TTL expired, used by `traceroute`
- **Redirect** – change routing (very dangerous since can be abused, so routers often set to ignore)