Advanced Computer Networking

Summary

Author: Thomas Pettinger

2017–03–03

Advanced Computer Networking
Technische Universität München

Table of Contents

1 Introduction **1**

1.1 Protocols 1

1.2 Node Forwarding Performance 1

2 Link Layer **3**

2.1 Ethernet 3

2.2 Limitations of Layer 2 4

2.3 MAC addresses 4

2.4 Layer 2 Switching 4

3 Network Layer **6**

3.1 Internet Protocol 6

3.2 ICMP 7

3.3 Active Network Measurements 7

3.4 Address Resolution Protocol (ARP) 7

1 Introduction

Terminology:

Protocols control sending and receiving of messages

Internet loosely hierarchical global network

Internet Standards • RFC: Request for comment

- IETF: Internet Engineering Task Force
- IANA: Internet Assigned Numbers Authority

1.1 Protocols

Protocols take care of addressing, fragmentation & re-sequencing, error control, congestion control, compression, privacy and more.
The internet has an layered architecture of protocols. On the sender side, protocols take the PDU (Protocol Data Unit) from layer N+1, add their header and trailer and pass the SDU (Service Data Unit) to layer N-1. On the receiver side, the corresponding protocol takes the PDU from layer N-1, strips header and trailer again and passes the SDU to layer N+1.

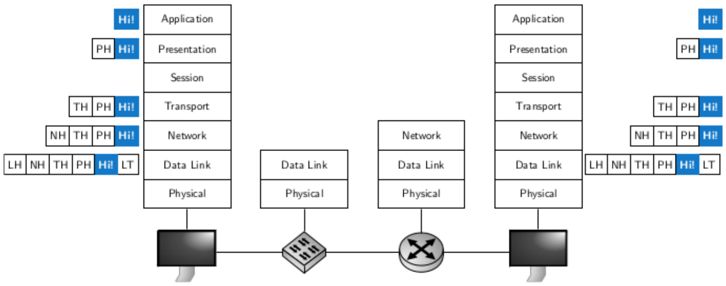

Figure 1: Internet Layers

Protocol layering is necessary because one does not want to implement everything to the physical layer when writing a networking application. On the other hand, layering also introduces some problems like protocol layers are sometimes reusing techniques of other layers like ARQ (Automatic Repeat Query) and layers might need informations of other layers.

1.2 Node Forwarding Performance

During transmission, packets might get delayed or even lost for several reasons. First, the packets need some time to get written to router buffers, secondly the packet arrival rate might exceed the output link capacity and lastly the packets need to wait again for being sent from the packet queue in routers.
The sources for these delays are listed below.

1. Processing delay: interrupt handling when receiving new packets and processing for further transmission
2. Queuing delay: waiting time in output queue
3. Transmission delay: time to send bits into link: $= \frac{\text{packet length L (bit)}}{\text{link bandwidth (bps)}}$
4. Propagation delay $= \frac{\text{length of physical link d}}{\text{propagation speed}\approx 2\cdot 10^8 m/s}$

The total amount of delay is then $d_{nodal} = d_{proc} + d_{queue} + d_{trans} + d_{prop}$
To reduce total packet delays for a connection consisting of several links one can use circuit switching, where packets do not have to be received entirely to be sent to the next link. Another alternative is to split packets into (very) small sub-parts (= segmenting) and using pipelining (parallel computing of packets).

2 Link Layer

Terminology:

- Hosts and routers are nodes
- Communication channels between adjacent nodes are links
- A layer 2 packet is a frame and encapsulates a layer 3 packet called datagram

The data-link layer has the responsibility of transferring a datagram from one node to an adjacent node over a link.

Services

- Framing, link access, MAC addressing
- Reliable delivery between adjacent nodes (mostly in wireless transmission)
- Flow control: Pacing between sending and receiving nodes
- Error detection
- Error correction
- Half- and full-duplex (half = both ends can transmit, but not simultaneously)

Multiple Access Protocols

When sharing a single channel, a distributed algorithm manages how nodes share it. This management is done via the same channel as the actual communication and does not require a separate one coordination.

Medium Access Control (MAC) Protocols Taxonomy

Channel Partitioning divides channel into smaller pieces (time, frequency, . . .)

Random Access does not divide channels, but try to recover from collisions

Taking turns Nodes take turns, requesting turns by polling or token passing

2.1 Ethernet

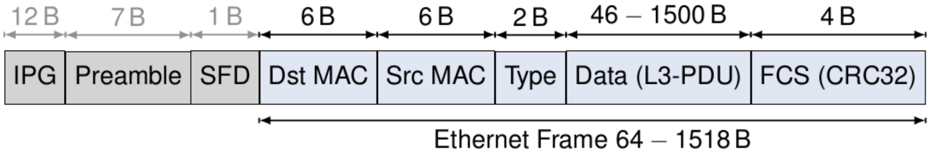

Figure 2: Ethernet Frame

IPG = Inter packet gap, minimum idle period Preamble = 7 byte (10101010. . .) SFD = Start-of-frame delimiter (10101011) Type = Ethernet II: Protocol type of payload, Ethernet I: length of payload in bytes PAD = Padding if data length smaller than 46 byte FCS = Frame check sequence (CRC-32)
There are several Ethernet standards, but they all share a common MAC protocol and frame format. They provide different bandwidth (from 10M to 200/400G (planned for 2017)) and have different physical layer media like twisted pairs (xBase-T), optical fibres or even chip to chip interfaces on NIC.

2.1.1 Carrier Sense Multiple Access - Collision Detection (CSMA/CD)

CSMA/CD is used for detecting and reacting to collisions. Its steps are

1. NIC receives datagram and creates frame
2. If NIC sees channel idle, it starts transmission, if channel busy, wait until idle
3. If NIC does not detect another transmission during its own transmission, it is done
4. If NIC does detect another transmission, jam signal is sent and transmission is aborted
5. NIC enters exponential backoff: after m-th collision, NIC chooses k at random from 0, 1, $\ldots 2^m - 1$ and waits $k \cdot 512bit$ times and returns to step 2. Bit time is $0.1\mu s$ for 10MbE

2.2 Limitations of Layer 2

- Flat addresses
- No hop count (dangerous when having loops)
- Missing protocols like ICMP
- Missing features: fragmentation, error messages, congestion feedback

2.3 MAC addresses

MAC addresses are 6 Byte long unique identifiers for NICs. Manufacturers can buy portions of the total MAC address space from the IEEE Registration Authority, which assures uniqueness. The first 3 bytes of the address in transmission order represent the Organization Unique Identifier (OUI). If the 2nd least significant byte is 0, the MAC is OUI enforced, otherwise its locally administered. MACs are transmitted in canonical form which stands for sending the least significant bit of each byte first (in memory, token ring and FDDI it is the other way around).

2.4 Layer 2 Switching

Hubs

Hubs are repeaters which means they send every bit arriving out to all other links. Because of this, frames from all connected nodes can collide with each other. Furthermore there is no frame buffering or CSMA/CD.

Switch

Switches are a lot smarter when compared to routers. They store and forward Ethernet frames only to the node that the destination MAC address belongs to. Furthermore they use CSMA/CD to access links. Hosts do not need to be aware of the presence of switches and they do not need to be configures and learn themselves. Learning is done when receiving packets: The switch then knows the location of the sender MAC address and stores it in a switch table. An entry expires after a specified amount of time. If a packet arrives, the switch table is checked if the destination is known. If yes, the packet is only sent to that node, otherwise it is sent to all.

If more switches are involved, the **spanning tree protocol** is used. It calculates a loop-free subnet of the given physical network and determines routing. The calculation steps are as followed:

1. Select root bridge, i.e. bridge with lowest bridge_ID (concatenation of 16bit bridge_priority and MAC address)

2. determine least cost paths to root
    - Every bridge determines cost of each path to root
    - Every bridge picks least cost path
    - port connecting to that path is root port
    - Bridges on network segment determine bridge port with least-cost-path to root, i.e. designated port
3. disable all other ports

Bridge Protocol Data Units (BPDUs) are used to transmit configuration information about bridge_IDs and root path costs, to notify about topology changes (TCN = Topology Change Notification) and for TCN acknowledgements.

3 Network Layer

The network layer serves the following functions:

- IP protocol for addressing, datagram format and packet handling conventions
- Routing protocols for path selection
- ICMP protocol for error reporting and router signaling

3.1 Internet Protocol

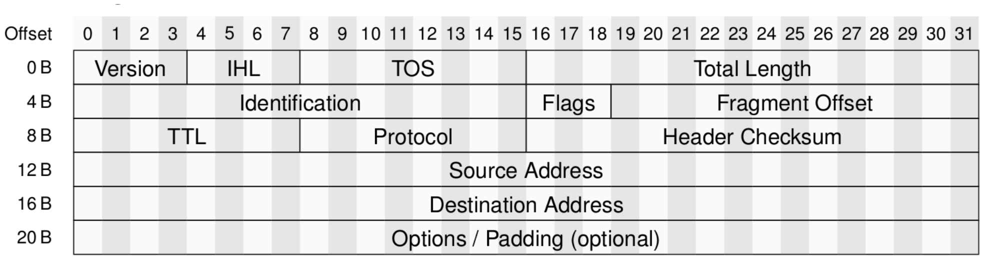

Figure 3: IPv4 Datagram

IPv4 Addressing

IPv4 addresses are 32-bit identifiers for every host and router interfaces where interfaces represent the connection between host/router and physical link.
Subnets are device interfaces with the same subnet part of the IP address which can physically reach each other without intervening router.
Splitting the IP address into network and host part is done in the following way (for the address 192.168.128.1/17):

$$192_{10}.\ 168_{10}.\ 128_{10}.\ 1_{10}$$

$$1100000_2.10101000_2.1_2 \quad 0000000_2.00000001_2$$

network part host part

From 1982 to 1993, IP addresses were classfully divided as shown in Figure 4. In 1993, Classless Inter-Domain

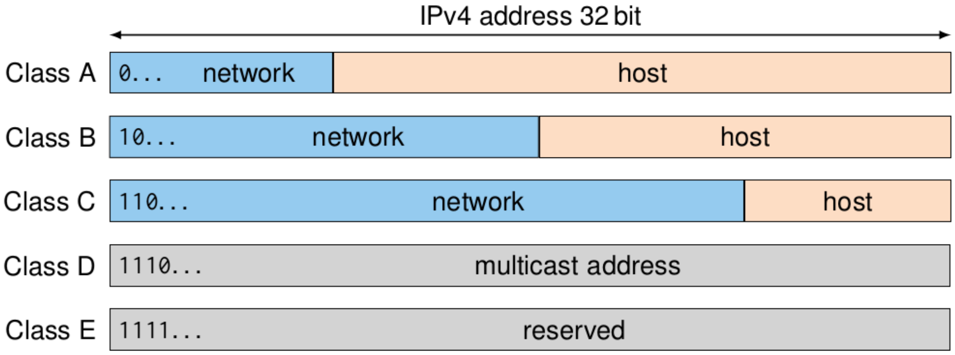

Figure 4: Classful IPs

Routing (CIDR) was introduced which allowed arbitrary subnet length. To route packets, prefix matching is used which checks which entry in the routing table fits best for the incoming packet's network prefix.

3.2 ICMP

The Internet Control Message Protocol (ICMP) are located above IP but can be considered as part of the IP layer. It is used for communicating error messages and other attention requiring conditions for IP and TCP or UDP. Two classes of ICMP messages are possible:

1. Query messages: only kind that generates other ICMP messages
2. Error messages: contain IP header and first 8 bytes (today as much as possible up to 572 bytes) of datagram that caused the ICMP message which allows the receiver to put it into context

The structure of an ICMP message is shown in Figure 5.

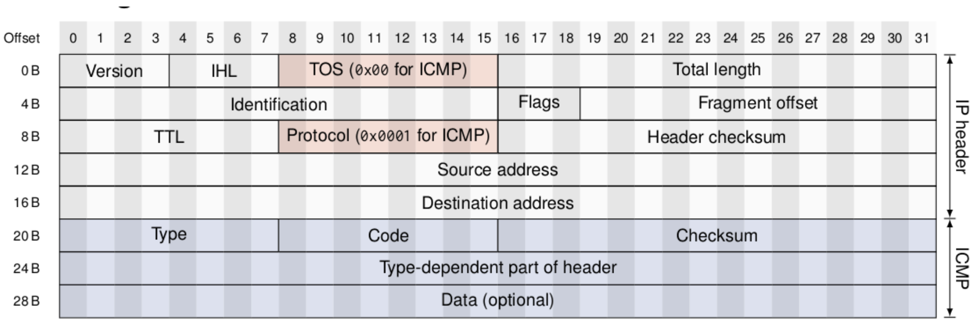

Figure 5: ICMP Message

3.3 Active Network Measurements

Network is actively measured by several parties like network providers (to manage traffic or reduce cost), service providers (to adjust service, get information about clients, ...), clients (to check services, get best one) or researchers (for performance evaluation of algorithms). Furthermore malicious traffic can be detected.

Measurements are done with probe packets and looking at the packet loss, one-way delay, RTTs or packet inter-arrival times.

(Paris-) Traceroute

Traceroute uses different TTLs in the IP header to get the route from the source to the destination. In case of load balancing though, traceroute might fail due to the appearance of ghost paths when successive packets are routed on different routes.
Load balancing routers usually use the IP-5-Tuple to determine routes, so to fixthis Paris traceroute uses different fields than normal traceroute (e.g. destination port for tcp) to do measurements.

3.4 Address Resolution Protocol (ARP)

The ARP is used to map IP addresses to MAC addresses. For that, an ARP broadcast is sent by the sender of an IP packet to get the MAC address of the next hop. The node with the specified IP address responds and the sender caches the mapping and is able to send the resulting Ethernet frame. In case we have a network with routers, the router then again does an ARP request for the IP address specified in the received

IP header and so the procedure begins again at that point. Cached information times out when not refreshed in a certain threshold.

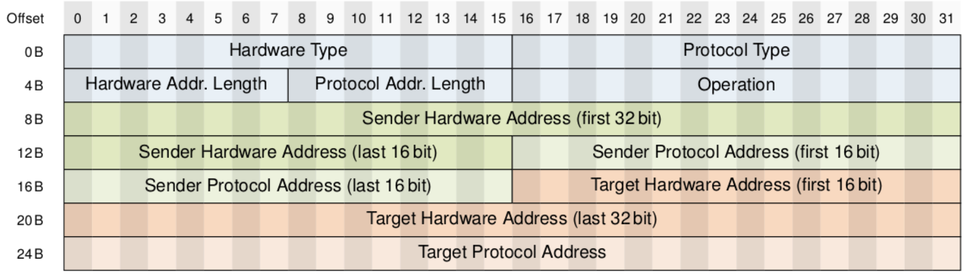
Offset
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
0 B
Hardware Type
Protocol Type
4 B
Hardware Addr. Length
Protocol Addr. Length
Operation
8 B
Sender Hardware Address (first 32 bit)
12 B
Sender Hardware Address (last 16 bit)
Sender Protocol Address (first 16 bit)
16 B
Sender Protocol Address (last 16 bit)
Target Hardware Address (first 16 bit)
20 B
Target Hardware Address (last 32 bit)
24 B
Target Protocol Address

Figure 6: ARP Packet

Reverse ARP also exists, but is rarely used.
Proxy ARP also responds for ARP request of one of its networks with ARP responses for hosts of another network. This enables transparent subnet gatewaying (two LANs with in same subnet), Host joining LAN via VPN and host separated via firewalls.

Since ARP is stateless and not authenticated, ARP responses can easily be forged to poison the cache of hots which can be used to redirect traffic.