Network layer

1: Introduction to TCP/IP, IP design
2: IP addressing, Address resolution
3: IP Routing

Overview Slide

Key Issues:
- Addressing
- Routing
- Group Communication
- QoS
- Security

INTERNET

Autonomous System (AS)
Host
External Router

Internet Protocol

- Network layer protocol for routing data across Internet.
- Objective: Fulfil DoD requirements:
  - Accommodate use of hosts/routers built by different vendors.
  - Encompass a growing variety of network types.
  - Enable network to grow without interrupting service.
  - Support higher layer sessions/message oriented services.
- IP network layer architecture designed to meet these needs.
- IP network layer: Part of TCP/IP protocol suite.
**TCP/IP: History of Internet**

- An industry standard suite of protocols designed for WANs.
- **1960s**: Mainframes were standalone, computers from different manufacturers could not communicate.
- Advanced Research Project Agency (ARPA) in the Department of Defense (DoD) initiated a research project for connecting different computers together.
  - For Researchers.
- Result: ARPANET comes into existence in **1969**: A small network of connected computers.
  - Consists of 4 nodes and a specialized computer called Interface Message Processor (IMP).

**History of Internet**

- Birth of Internet: Vinton Cerf and Robert Kahn develop the Gateway.
- Some important milestones:
  - **1972**: Ad Hoc Telnet Specification (RFC 317)
  - **1973**: Kahn and Cerf’s landmark paper on TCP for end-to-end packet delivery.
  - ARPA hands ARPANET to Defence Communications Agency (DCA).
  - **1981**: TCP split into two protocols TCP and IP.
  - IP standard published in RFC 791.

**1981**: UNIX modified by Berkeley to include TCP/IP
- Unix: an open implementation.
- TCP/IP official protocol for ARPANET.
- ARPANET split into 2 networks
  - MILNET and ARPANET.
- **1984**: DNS introduced.
- **1990**: ARPANET replaced by NSFNET.
- NSFNET evolved into ANSNET (**1991**) and finally into Internet.

**TCP/IP**

- Standard, routable enterprise networking protocol.
- Technology for connecting dissimilar systems.
- Robust, scalable, cross-platform client-server framework.
- Method of gaining access to Internet.
Internet Standards Process

- Development associated with administration of internet.
- No organization owns Internet.
- Internet Society (ISOC).
- Internet Architecture Board (IAB).
  - Internet Engineering Task Force (IETF).
  - Internet Assigned Number Authority (IANA).
  - Internet Research Task Force (IRTF).

TCP/IP Protocol Stack

- Four layer conceptual model.
- Network Interface layer: (OSI physical, data link layer functionality).
  - Protocols defined by underlying networks.
- Internet layer: (OSI network layer)
  - Addressing, Routing
  - Houses other support protocols.

TCP/IP Protocol Stack

- Transport layer (OSI Transport-Session layer functionality)
  - Connection Oriented, reliable (TCP)
  - Connectionless, Unreliable (UDP)
- Application Layer (OSI Application-Presentation layer Functionality)
TCP/IP Versions

- Six versions of TCP/IP since its inception.
- Three later versions are:
    - Address length 32 bits.
    - Address space divided into classes.
    - Not sufficient to handle projected number of users.
    - Classes limit the available addresses.
  - Version 5: Based on OSI model
    - Never went beyond the proposal stage.
    - Also known as IPng (ng: New generation).
    - Uses 128 bit addresses.
    - Simplified/flexible packet format.
    - Support for security features.
    - Handles congestion and route discovery better than version 4.

Internet Architecture

- Internet consists of groups of networks and routers.
- Each group: Single administrative authority.
- Group: Autonomous system.
- Autonomous systems: Multiple independent networks.
  - Internal routes governed by local authority.
  - Routes consistent and viable.
- Reachability to Internet via advertisements.

Global consistency of routing information: Routing Arbiter (RA).
RA: Replicated authenticated database of reachability information.
Major ISPs interconnect at Network Access Point (NAP). (Core AS)
Each ISP represents one or more ASs.
NAP: Boundary between several ASs.
NAP includes Route Server (RS).

Internet Architecture
Internet Architecture

- RS: Supplies each ISP with reachability information.
- Each ISP designates one router to interact with Route server.

Internet Protocol

- Maps onto OSI Network layer protocol.
- Offers
  - Connectionless
  - Unreliable
  - Datagram delivery service.
- Two tools used by IP
  - Addressing (Subnet mask)
  - Routing (IP Routing table).

Primary IP Function

- Accept data from upper transport layer (TCP/UDP)
- Create a datagram
- Route datagram through internetwork.
- Deliver datagram to recipient transport layer and finally the intended application.
- Need to examine datagram before we discuss the above bullet points.

IP Datagram

- Packets in IP layer called datagrams.
- Maximum length of the datagram is 65,536 bytes.
- Packets:
  - Variable length
  - Header and data.

Reference: C4 (Comer) chapter 7
IP Datagram

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Version</td>
<td>IP version</td>
</tr>
<tr>
<td>IHL</td>
<td>.Length</td>
</tr>
<tr>
<td>DS service type</td>
<td></td>
</tr>
<tr>
<td>Total Length</td>
<td></td>
</tr>
<tr>
<td>Identification</td>
<td>Identification number</td>
</tr>
<tr>
<td>Flags</td>
<td>Flags</td>
</tr>
<tr>
<td>Fragment offset (13)</td>
<td>Fragment offset</td>
</tr>
<tr>
<td>Time to Live (TTL)</td>
<td>TTL</td>
</tr>
<tr>
<td>Protocol</td>
<td>Protocol</td>
</tr>
<tr>
<td>Header Checksum</td>
<td>Header Checksum</td>
</tr>
<tr>
<td>Source Address</td>
<td>Source Address</td>
</tr>
<tr>
<td>Destination Address</td>
<td>Destination Address</td>
</tr>
<tr>
<td>Options</td>
<td>Options</td>
</tr>
<tr>
<td>Data</td>
<td>Data</td>
</tr>
</tbody>
</table>

Figure 3: IP Options

![Option Format Diagram]

**IP Datagram Options Fields**

- Options Sub-Field: Used for network testing and debugging.
- Contains 4 fields:
  - A code that identifies option
  - Option length
  - Pointer: As an offset
  - Specific data

**IP Datagram Options Fields**

- **Code**: Contains three sub-fields:
  - Copy
  - Class
  - Number
- Copy: Controls presence of options in fragments.
- Class: Defines general purpose of option.
- Number: Defines the type of options.
**IP Datagram (Options Field)**

- Options defined by **Number** sub-field:
  - Record Route: used to record Internet routers that handle the datagram.
  - Strict Source Route: Source predetermines a route a datagram travels through Internet. Why??
  - Loose Source Route: Similar to strict route option but datagram can visit other routers as well.
  - Timestamp: Records time of datagram processing by a router.
    - O-Flow
    - Flags

**Timestamp Option**

<table>
<thead>
<tr>
<th>Code</th>
<th>Length</th>
<th>Pointer</th>
<th>O-Flow</th>
<th>Flags</th>
</tr>
</thead>
</table>

**IP Datagram Options Fields**

- **No Operation**: Padding between options
- **End of Option List**: Padding after last option

- **Length**: Defines total length of option.
- **Pointer**: An offset integer field containing the byte number of the first empty entry.
- **Value**: Contains the data that specific options require.
Encapsulation

- Technique used by layered protocols.
- Higher level protocol packet placed in data portion of lower level protocol packet.
- E.g. IP datagram within a Data Link frame.
  - This makes sense
  - IP resides above data link frame.
  - Sometimes IP packets are encapsulated within IP packets.
  - Can you think of a situation when this would happen and why?

Ideal Case: One entire datagram fits into the link layer frame.

This may not always be the case.

How to solve this Problem??

IP Fragmentation

- Datagram can travel through different networks.
  - Different networks have different physical layer specifications.
    - Token ring, Ethernet
    - Format, size of frame are governed by specifications.
    - Maximum Transfer Unit (MTU)
    - Largest size data unit that a network can handle.
    - Limitation imposed by DL layer.
    - Value of MTU varies.
  - Solution: Path MTU Discovery/IP Fragmentation

Path MTU Discovery

- Source specifies a “don’t fragment” option.
  - Initially a large sized packet is sent.
    - Maximum size of local MTU.
    - Maximum segment size negotiated by TCP Handshake.
  - If a network is encountered with a small MTU, ICMP error message generated to wards the source.
    - MTU size specified.
    - Source learns of the correct size.
IP Fragmentation

- Fragmentation: Division of packet into smaller units to accommodate a protocol’s MTU.
- Each fragment has its own header.
- Fragment can be further fragmented.
- Datagram fragmented at source or any other router in the path.
- Reassembly done only at destination.
  - Why??

Fragmentation Example

Reassembly at Destination

- Advantage
  - No need to assemble datagrams immediately after it passes a network with a small MTU.
- Disadvantages
  - Smaller fragments traverse networks with large MTU capability.
  - If fragments are lost, then datagram cannot be reassembled.
Fragmenting a Fragment

- Last fragment has MF=0; set MF=1 in all but last new fragments.
- First and middle fragments have MF=1; set MF=1 in all new fragments.
- Only 1 fragment has MF set to 0 regardless of the number of times fragmentation is performed.
- Offset always calculated relative to the original datagram.
- Receiver does not see multiple layers of fragment.

IP Design

- Simplified bare bones design of IP: Involves eight components
  - Header adding module
  - Processing Module
  - Routing Module
  - Fragmentation Module
  - Reassembly module
  - Routing table
  - MTU table
  - Reassembly table

IP Components

- Header Adding Module
  - Encapsulate data in IP datagrams.
  - Calculate Checksum, insert in checksum field.
  - Send data to corresponding input queue.
  - Return
IP Components

- Processing Module
  - Remove one datagram from input queue.
  - If destination local:
    - Send datagram to reassembly module.
    - Return.
  - If machine (A router)
    - Decrement TTL.
    - If TTL = 0
      - Discard datagram.
      - Send ICMP Error message.
      - Return
    - Send datagram to routing module.
    - Return

- Routing Module (Discussed in detail in routing subsection)
  - Receive IP packet from processing module.
  - If destination local, deliver packet using link layer.
  - If destination remote
    - Find IP address of next station towards destination.
    - Determine interface to deliver the packet.
    - Packet sent to fragmentation Module.

IP Components

- Fragmentation Module
  - Receive packet from Routing module.
  - Extract size of datagram.
  - If size >MTU (Of corresponding network)
    - If D bit set
      - Discard datagram, send ICMP error message.
      - Return.
    - Else
      - Calculate maximum size, divide datagram into fragments.
      - Add header, required options to each fragment.
      - Send datagram
      - Return.

- Reassembly Module.
  - Receive IP datagram from processing module.
  - Find datagram to which a fragment belongs.
  - Reassemble fragments of datagram (when all fragments arrive)
  - Time out expires, fragment missing: Discard datagram.
  - Return.
IP Components

- Queues:
  - Input
  - Output.
- Routing Table:
  - To determine next hop.
- MTU:
  - To find Max. Transfer Unit of interface.

Summary: IP Working

- IP on the router
  - Packet received at router.
  - IP decrements TTL by 1. If TTL = 0, discard packet.
  - IP may fragment packet.
  - If fragmented, add new header.
    - Fragment Id, Identification Flag, Fragment Offset.
  - IP calculates new checksum.
  - IP obtains destination hardware address of next router.
  - IP forwards packets.
  - Process repeated until packet reaches final destination.

Section 2

IP Addressing/Address Resolution

IP Addressing

- Internet: a combination of different physical networks.
- Packets travel different physical networks to destination.
- A global identification system necessary for hosts (and routers).
- Solution: Unique identifier is IP address in Internet.
- IP address:
  - Unique
  - Universal.
- Location not names.
Address Format

- 32 bits long
  - Four 8 bit fields, called octets.
- Each octet represented as decimal number in the range 0-255.
  - Octets separated by periods.
  - Dotted decimal notation.
- Address divided into two parts:
  - Netid: identifies systems located on same physical segment.
  - Hostid: Identifies individual entity (host, server, router) within the segment.

IP Address Format

<table>
<thead>
<tr>
<th>Network ID</th>
<th>Host ID</th>
</tr>
</thead>
</table>

Address Classes

- **Class**: Defines
  - bits for **netid** and **hostid**.
  - number of networks/number of hosts per network.
- Five classes defined
  - A, B, C, D, and E.
- We shall examine Class A, B and C.
- Class D is a Multicast Address.
- Class E is for Experimental use (Not defined).
Address Class Summary

<table>
<thead>
<tr>
<th>Class</th>
<th>Number of networks</th>
<th>Number of Hosts per Network</th>
<th>Range of Network Ids (First Octet)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class A</td>
<td>126</td>
<td>16,777,214</td>
<td>1 - 126</td>
</tr>
<tr>
<td>Class B</td>
<td>16,384</td>
<td>65,534</td>
<td>128 - 191</td>
</tr>
<tr>
<td>Class C</td>
<td>2,097,152</td>
<td>254</td>
<td>192 - 223</td>
</tr>
</tbody>
</table>

Multihomed Devices

- Internet address defines node’s connection to its network.
- Multihomed: A device connected to more than one network.
  - Different address for each network.
  - Address can be of different classes.
  - Can you think of such a multihomed device??

Special Addresses

- Some parts of address space (A,B,C) are used for special addresses.
- Network address: A,B,C, address with hostid set to zero.
  - Defines the network itself.
- Direct Broadcast Address
  - True for A, B, C.
  - All hostid’s are set to 1.
  - Packets sent to all hosts in a specific network.

Special Addresses

- Limited Broadcast Address
  - True for A, B, C.
  - All 1s in the netid and host id.
  - Broadcast within the current network.
- This host on this network
  - True for A only.
  - All zeroes address.
- Specific host on this network
  - True for A only.
  - Netid of all zeroes.
Special Addresses

- Loopback Address.
  - (This is a class A address)
  - First byte set to 127.
  - To test IP software in a machine.
  - Packet returns to protocol software: Does not leave machine.

Unicast, Multicast and Broadcast

- Unicast: One to one communication.
  - Addresses belong to class A, B, or C.
- Multicast: One to many communication.
- Broadcast: One to all communication.
  - Internet limits broadcast to local level.
  - No global level broadcast. Why??
  - Example: Limited broadcast, direct broadcast.

Private Networks

- For organisations not requiring internet access.
- Organisation uses TCP/IP.
- Addressing scheme options:
  - Apply for unique address and NOT connect to Internet.
  - Use any class A, B, C without registering with authority.
  - Use reserved block of private addresses.

Applying for IP addresses

- Organization applies for netid. (A, B, C)
- Problem: Class A addresses depleted, very few B addresses left.
- Proof for procuring A.
- Organization assigns hostids from the given netid.
- Application for netid: Network Information Centers.
Assigning NetIDs and HostIDs

- Network Id (Netid): identifies host located on same physical segment.
- Unique id for each router interface
- Figure 5:
  - Networks 1 and 3 represent two routed networks.
  - Network 2: Wide area connection between two routers.
  - Network 2 requires network id so that interfaces between two routers can be assigned unique host ids.

Assigning Host IDs

- Host id: identifies a host in network.
  - Unique to network id.
  - Assigned to: hosts, interfaces to routers.
- Hostid of router interface: Default gateway for other hosts.
- Suggestions
  - No rules for assigning valid host ids.
  - Guidelines:
    - Assign host ids in groups based on host or server type.
    - Designate routers by their IP address.
Table 2: Valid Host IDs

<table>
<thead>
<tr>
<th>Address Class</th>
<th>Beginning Range</th>
<th>Ending range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class A</td>
<td>w.0.0.1</td>
<td>w.255.255.254</td>
</tr>
<tr>
<td>Class B</td>
<td>w.x.0.1</td>
<td>w.x.255.254</td>
</tr>
<tr>
<td>Class C</td>
<td>w.x.y.1</td>
<td>w.x.y.254</td>
</tr>
</tbody>
</table>

Subnet Mask

- What is a subnet mask?
  - Distinguishes network id from host id.
  - Specifies if destination host is local or remote.
- Default subnet mask
  - Used on networks not divided into subnets.
- What is a subnet??
  - All hosts on TCP/IP network need a subnet mask.
  - Its use depends on address class.
  - All bits corresponding to network id are set to 1.
  - All bits corresponding to host id are set to 0.

Subnetting

- Subnet
  - Division of a single class A, B, or C network into smaller pieces.
  - Each piece: A physical network in TCP/IP environment.
  - Uses IP address derived from single network ID.
  - Result: Single network (Single Netid) divided into smaller subnets.
    - Each subnet has different network ID.
  - Determining local vs remote hosts
    - ANDing: Internal process to determine whether a packet is local or remote.
    - Host's IP address and remote address ANDed with default subnet mask.
    - If both match, they belong to same network.
Need for Subnetting

- IP addresses rapidly getting saturated.
- Internet Routing tables beginning to grow.
- Local administrators must request another network number from Internet before a new network could be installed.

Subnetting:
- Organization gets one Net ID from InterNIC.
- One Net ID divided into multiple sub IDs.
- Each sub ID can represent a single physical network.

Subnetting

Benefits
- Mix different technologies.
- Overcome limitations of current technologies.
- Reduce network congestion.
- Subnetworks appear as a single network to the rest of the Internet.

Note: Subnetting defined in RFC 950.

Fig 7: Extended Network Prefix

| Network-Prefix | Subnet Number | Host-Number |

Subnet Design Considerations

Key Questions
- How many total subnets does the organization need today?
- How many total subnets will the organization need in the future?
- How many hosts are there on the organization’s largest subnet today?
- How many hosts will there be in the organization’s largest subnet in the future?
**Special Subnet Addresses**

- Initially, it was prohibited to use all-0s and all-1s subnet.
  - Reason: To eliminate situations that could potentially confuse a router.
- Today a router has functionality to support this feature.
- In our examples we will label these as special addresses.

**Subnetting Procedure**

- Defining subnet mask:
  - Determine number of physical segments required.
  - Convert this number **minus 1** to binary.
  - Count number of bits required to represent number of physical segments in binary.
  - Convert required number of bits to decimal format in high order.

**Subnetting example**

- Class B network:
  - Subnet mask is 255.255.0.0
  - E.g. 131.73.0.0/16
  - 6 subnets
    - 5 in Binary 101: 3 bits for subnet
    - Subnet mask bits 11100000
    - Decimal 128+64+32 = 224
    - Subnet mask is 255.255.224.0

**Subnetting Procedure**

- Defining subnet IDs.
  - List number of bits in high order used for subnet id.
  - Convert the bit with lowest value to decimal format.
    - Increment value.
    - Starting with zero, increment value for each bit combination until next value is 256.
  - Tip: If you know number of bits you need, you can raise 2 to the power of bit, then subtract 2 to determine possible bit combinations.
### Subnet IDs Example

- Network 131.73.0.0/16
- 8 subnets: 131.73.0.0/19
  - mask 255.255.224.0
- First network ID (binary) 00100000
  - Decimal 32
- Network IDs:
  - 131.73.0.0/19 131.73.128.0/19
  - 131.73.32.0/19 131.73.160.0/19
  - 131.73.64.0/19 131.73.192.0/19
  - 131.73.96.0/19 131.73.224.0/19
  - (131.73.256.0 check)

### Subnetting Procedure

- To determine number of hosts per subnet.
  - $N = \text{How many bits are available for the host ID?}$
  - Raise 2 to the power of $N$
  - Subtract 2 (all 0 host ID disallowed, all 1 disallowed).
- Example
  - Class B, 8 subnets, subnet mask 255.255.224.0
  - Binary 11111111.11111111.11100000.00000000
  - 13 bits for hostID $\rightarrow 2^{13} - 2 = 8192 - 2 = 8190$ hosts/subnet.

### Variable Length Subnetting

- Divide a subnet further into smaller subnets.
  - Class B 131.73.0.0/16 8 subnets 131.73.x.0/19
  - 16 subnets of 131.73.160.0 as 131.73.y.0/23
- Subnetted network uses more than one subnet mask.
- Mask can be of variable length: Variable Length Subnet Mask (VLSM)
- VLSM Benefits:
  - Efficient use of Organization’s assigned IP address space.
  - Route aggregation.

### Subnetting Strategy for VLSM

- 140.25.0.0/16
  - Level 1: 10 subnets 140.25.x.0/20
  - 0 1 2 3 4 5 6 7 8 9
  - Level 2: 140.25.64.0/20
  - 5 subnets 140.25.y.0/23
  - 0 1 2 3
  - Level 2: 140.25.128.0/20
  - 4 subnets 140.25.y.0/22
  - 0 1 2 3
  - Level 3: 140.25.132.0/22
  - 6 subnets 140.25.z.a/25
  - 0 1 2 3 4 5
VLSM Design Considerations

- How many total subnets does this level need today?
- How many total subnets will this level need in the future?
- How many hosts are there on this level’s largest subnet today?
- How many hosts will there be on this level’s largest subnet in the future?

CIDR

- Classless Inter-Domain Routing
- Assigns blocks of addresses to an organisation
  - E.g. 128.211.168.0/21 – part of old class B
  - E.g. 193.11.12.0/22 – block of 4 old class C
- Allows more sensible block sizes than classes
- Helps solve shortage of address space
  - More details later...

Address Resolution

- At network level: Host id is logical address: IP
  - Jurisdiction is universal.
- At physical level, hosts, routers recognized by physical address.
  - Jurisdiction is local.
  - MAC addresses.
- Two unique identifiers (logical and physical) essential. Why??
- Delivery of packet needs two levels of addressing.
- Mapping essential.
Address Resolution Protocol (ARP)

- Map IP (Logical) address to a hardware (Physical) address.
  - Called Address resolution
- ARP uses local broadcast to obtain a hardware address.
- Address mappings are stored in cache for future reference.
- Two cases of resolution:
  - Local
  - Remote

ARP: Resolving local IP address

- ARP request initiated
- Check IP address: Is it local/remote?
- If local:
  - Check ARP cache.
  - If no cached mapping found
    - Generate ARP request
    - Broadcast request to all hosts.
  - Concerned target host
    - Recognises IP address in request.
    - Generates ARP reply
  - Sender updates its cache.

ARP: Remote Addresses

- IP address is remote:
  - Determine IP address of default gateway
  - Find hardware address of gateway as normal (cache/ARP)
  - Send data packet to gateway
- Gateway:
  - Follows same address resolution procedure:
    - If remote, determine default gateway
    - If local, use cache/ARP to find hardware address

ARP Cache

- Maintains static/dynamic entries.
- Dynamic entries added/deleted automatically.
- Hardware broadcast address maintained as static entry.
- Each entry has potential lifetime: 10 minutes.
- Each entry, when added is timestamped.
- Default timeout: 2 minutes for unused entries, 10 minutes for used entries.
ARP Command

- Cache examined using `arp` command.
  - `-a`: displays all entries in ARP cache.
  - `-d`: deletes entry from arp cache.
  - `-s`: adds entry into arp cache
- Sender’s ARP data included in ARP request
- What does Reverse ARP do? (C4 chap 6)
- Gratuitous ARP: An ARP message broadcasting your IP address and your corresponding physical address.
  - Why is this used?
- What is Proxy ARP? (C4 p150)