Internet Architecture and Protocols Review

CS2520/TelCom2321
Wide Area Network
Spring Term, 2019

Prof. Taieb Znati
Department Computer Science
Telecommunication Program

Outline

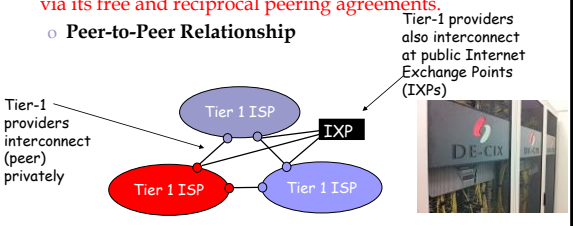
- Internet Views and Structure
  - Internet Service Provision and Peering
- Internet Protocol Architecture
- Internet Application Design Issues
  - End-to-End Design Issues
- Internet Service Model
  - Datagram Services
- Internet Network Protocol

Internet Structure

Internet – Network of Netoworks

Internet Structure

- Roughly hierarchical strutcture
- **At center: "Tier-1" ISPs** - Have access to the entire ***Internet Region,*** a portion of the Internet network typically bounded by a **country's geographical boundaries**, solely via its free and reciprocal peering agreements.
  - **Peer-to-Peer Relationship**



- IXP consists of one or more network switches, to which each of the participating ISPs connect.
- Mostly, Ethernet and FDDI switches

2019	Name	City	Createtd	Members	Max Thruput (Gb/s)	Avg Thruput (Gb/s)
DE-CIX	Deutscher Commercial Internet Exchange	Frankfurt, Hamburg, Munich, Düsseldorf, New York City, Dubai (as UAE-IX), Palermo, Marseille, Istanbul, Dallas, Madrid, Lisbon, Mumbai (as Mumbai IX)	1995	863	6,738	4,300
IX.br	Brazil Internet Exchange	Aracaju, Belém, Belo Horizonte, Brasília, Campina Grande, Campinas, Caxias do Sul, Cuiabá, Curitiba, Florianópolis, Fortaleza, Foz do Iguacu, Goiânia, João Pessoa, Lajeado, Londrina, Maceio, Manaus, Maringá, Natal, Porto Alegre, Recife, Rio de Janeiro, Salvador, Santo Maria, São Carlos, São José dos Campos, São José do Rio Preto, São Luis, São Paulo, Teresina, Vitória	2004	3,252	6,020	3,650
AMS-IX	Amsterdam Internet Exchange	Amsterdam, Haarlem, Schiphol-Rijk, Willemstad,Hong Kong, New York City, Chicago, San Francisco Bay Area,Mumbai	1997	818	5,858	3,976
LINX	London Internet Exchange	London, Manchester, Edinburgh, Cardiff, Northern Virginia	1994	825	4,340	2,850
MSK-IX	MSK-IX	Moscow, SaintPetersburg, Novosibirsk, Rostov-on-Don, Stavropol, Samara, Kazan, Ekaterinburg, Vladivostok, Riga	1995	504	3,249	1,530
DATA-IX	DATA-IX	Moscow, Saint-Petersburg, Novosibirsk, Samara, Ufa, Perm, Ekaterinburg, Chelyabinsk, Krasnoyarsk, Khabarovsk, Omsk	2009	406	3,126	N/A

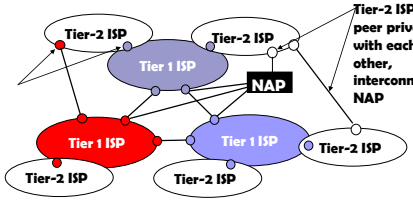
IXPs around the World

Source: TeleGeography World IX Map, http://www.internetexchangemap.com/

Internet Structure – Tier-2

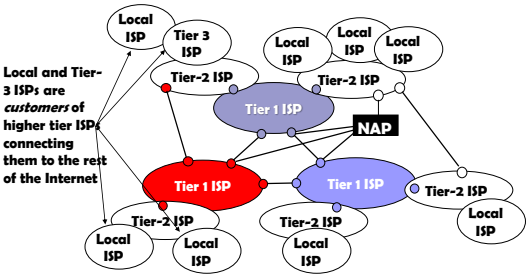
- **"Tier-2" ISPs: Smaller (often regional) ISPs**
  - Connect to one or more tier-1 ISPs, possibly other tier-2 ISPs

Tier-2 ISP pays tier-1 ISP for connectivity to rest of Internet
- **Tier-2 ISP is *customer* of Tier-1 provider**



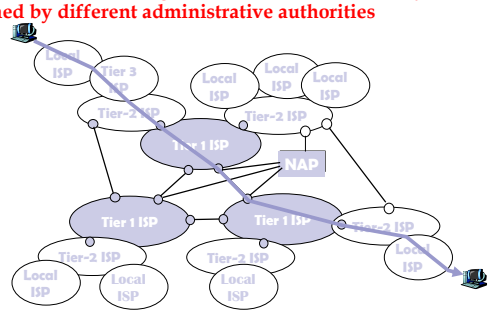
Internet Structure – Tier-3

- "Tier-3" ISPs and Local ISPs
  - Last Hop ("access") Network (closest to end systems)



Internet Structure – End-to-End Path

- A packet passes through several networks, possibly owned by different administrative authorities



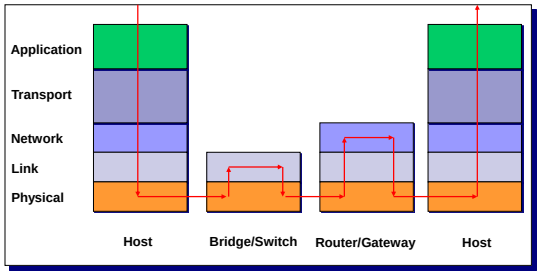
Internet Architecture and Basic Protocols

Internet Protocol Stack

- Application: Supporting network applications
  - FTP, SMTP, STTP
- Transport: Host-host data transfer
  - TCP, UDP
- Network: routing of datagrams from source to destination
  - IP, Routing Protocols – Interior Routing Protocols and Exterior Routing Protcols
- Link: Data transfer between neighboring network elements
  - PPP, Ethernet
- Physical: Bits "on the wire"

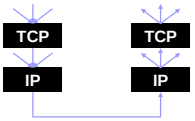
IP Layering

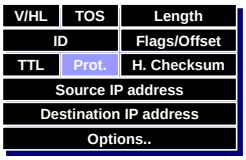
- A relatively simple layered architecture



Multiplexing and Demultiplexing

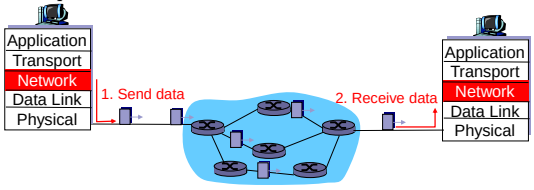
- There may be multiple implementations of each layer.
  - How does the receiver know what version of a layer to use?
- Each header includes a demultiplexing field that is used to identify the next layer.
  - Filled in by the sender
  - Used by the receiver
- Multiplexing occurs at multiple layers. E.g., IP, TCP, …





Datagram networks -- The Internet Model

- No call setup at network layer
- Routers: no state about end-to-end connections
  - No network-level concept of "connection"
- Packets typically routed using destination host ID
  - Packets between same source-dest pair may take different paths



Network Edge: Connection-Oriented Service

Transmission Control Protocol [RFC 793] – Internet's connection-oriented service

- TCP service – *reliable, in-order* byte-stream data transfer
  - Packet Loss
    - Acknowledgements and retransmissions
  - Flow Control
    - Sender won't overwhelm receiver
  - Congestion Control
    - Senders "slow down sending rate" when network is congested

Network Edge: Connectionless Service

User Datagram Protocol [RFC 768]: Internet's connectionless service

- UDP Service – unreliable data transfer
  - No flow control
  - No congestion control
- Applications using TCP
  - HTTP (WWW), FTP (file transfer), Telnet (remote login), SMTP (email)
- Applications using UDP
  - Streaming media, teleconferencing, Internet telephony

Internet Architecture and Protocols – Overview

Outline

- Internet Network Service Model and Protocol
  - Best-Effort Network Model
  - Internet Protocol
  - Internet Services

Part I – Internet Protocol (IP)

INTERNET SERVICE MODEL

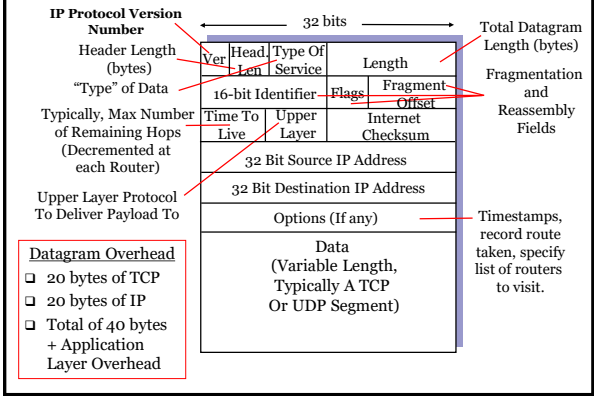
Internet "Best-Effort" Service

- Internet "best-effort" semantics
  - Unreliable, connectionless packet delivery system
  - The service makes the earnesst attempt to deliver packets
    - Delivery is not guaranteed
    - Packets may be lost, duplicated, delayed or delivered out-of-sequence
    - Packets are treated independently
- This service is defined by the Internet Protocol

Internet Protocol

- The Internet Protocol defines:
  - The basic unit of data transfer, Internet datagram
  - A routing algorithm
  - A set of rules that characterize the "best effort" delivery system

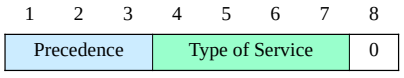
IP datagram format



Datagram Overhead

- 20 bytes of TCP
- 20 bytes of IP
- Total of 40 bytes + Application Layer Overhead

Datagram Service Type



Precedence Field

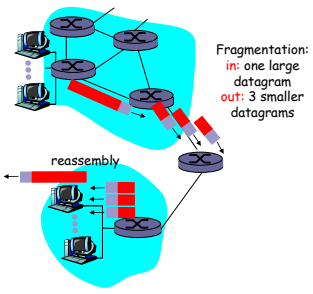
- Precedence is a measure of the nature and priority of the datagram
  - 0 : Routine
  - 1 : Priority
  - 2 : Immediate
  - 3 : Flash
  - 4 : Flash override
  - 5 : Critical
  - 6 : Internet Control
  - 7 : Network Control

Type of Service

- Specifies the type of service values
  - 1000 : Minimize delay
  - 0100 : Maximize throughput
  - 0010 : Maximize reliability
  - 0001 : Minimize monetary cost
  - 0000 : Normal service

IP Fragmentation

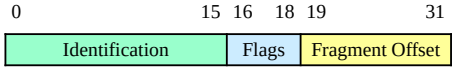
- Network links have MTU (max.transfer size) - largest possible link-level frame.
  - different link types, different MTUs
- Large IP datagram divided ("fragmented") within net
  - One datagram becomes several datagrams
  - "Reassembled" only at final destination
  - IP header bits used to identify, order related fragments



Fragmentation

- Datagrams have maximum length of 65,535 bytes
- Datagrams are encapsulated into physical frames
  - Ideal case, IP datagram fits into a physical frame
  - Often, fragmentation is required
  - IP expects routers to handle datagrams of up to 576 bytes
- Reassemly of fragments takes place at the ultimate destination
  - May lead to inefficiency
    - When timers expire, the receiving machine drops the fragments

Fragmentation Control

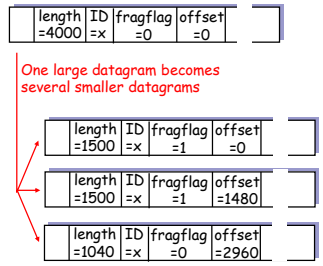


- Total length refers to the size of the fragment, so it cannot be used to collect all fragments
  - Identification: unique integer for every fragmented packet – all fragments share the same identification
  - Flags: No fragment, More fragments
    - No fragment forbids fragmenting of a packet
    - More fragments indicates it is not the last packet fragment
  - Fragment offset – offset in multiples of 8 bytes of given fragment in the packet
    - Size of all fragments except for the last one must be divisible by 8

IP Fragmentation and Reassembly

Example

- 4000 byte datagram
- MTU = 1500 bytes



Time to Live (TTL)

- The field specifies how long, in seconds, a datagram is allowed to live in the Internet
  - Maximum time is estimated when a datagram is injected
- In theory:
  - Each router decrements by one the TTL of each datagram it processes
  - Each router decrements the TTL by the number of seconds the datagram waited inside the router
  - When TTL reaches 0, the datagram is dropped
- In practice:
  - TTL field is reduced by one on every hop
    - TTL field is renamed "hop limit" in IPv6 to reflect this practice
- Guarantees that datagram cannot circulate indefinitely
- Can be used to alleviate congestion

Protocol Number

- Indicates the higher level protocol to which IP should deliver the data
  - 0 : Reserved
  - 1 : ICMP
  - 2 : IGMP
  - 3 : GGP
  - 4 : IP encapsulation
  - 5 : stream
  - 6 : TCP
  - 17 : UDP
  - 89 : OSPF

Header Checksum

- It is a checksum for the header only and does not include the data
  - Checksum is calculated as the 16-bit one's complement sum of all 16-bit words in the header
    - For the purpose of this calculation, the checksum is assumed to be 0

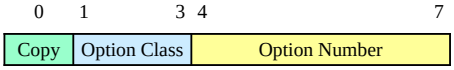
Options

- Options are variable length
  - An IP implementation is not required to be capable of generating options in the datagram it creates
    - All implementations, however, are required to be able to process a datagram containing options
- There may be zero or more options in a datagram

Option Format

- Two option formats are possible
  - A type byte alone
  - A type byte, a length byte and one or more data bytes
    - The type byte has the same structure in both cases
- The format of each option is dependent on the value of the option number found in the first byte of option

Datagram Options –Type Byte Format



- Copy : when set that option should be copied into all fragments
- Option Class:
  - 0 : Datagram and Network Control
  - 2 : Debugging and Measurement
  - 1,3 : Unused

Option Number

- 0 : End of option list
- 1 : No operation
  - This option may be used to align fields in the datagram
- 2 : Security
- 3 : Loose source routing
- 4 : Internet time stamp
- 7 : Record route
- 8 : Stream
- 9 : Strict source routing

Record Route Option

- Allows a source to create an empty list of IP addresses
- Each router that handles the datagram adds its IP address to the list
- Used for evaluation of routing paths

Source Route Option

- Provides a way for the sender to dictate a routing path
- Two forms of source routing
  - Strict source routing
    - The address specify the exact path the datagram should follow
    - If the route cannot be satisfied, the gateway drops the datagram
  - Loose source routing
    - Specifies the path that must be followed by the datagram, but allows for multiple hops between succesive addresses in the list
- Useful for testing paths and reachability

Time Stamp Option

- Allow a source to create an empty list of IP addresses and time stamps (current time and date)
- The options are determined by a flag
  - 0 : record time stamps only, omit IP addresses
  - 1 : precede each time stamp with an IP address
  - 3 : IP address is specified by the sender
    - A router records a time stamp only if its address matches the next IP address