

Network Layer

Arnav Gupta

October 24, 2024

Contents

1	Overview	1
1.1	Best Effort Service	2
2	Data Plane: The Internet Protocol	3
2.1	Fragmentation and Assembly	4
2.2	IPv4 Addressing	4
2.3	IPv6 Addressing	5
3	What's Inside a Router	5
3.1	Switching Fabrics	6
3.1.1	Switching via Memory	6
3.1.2	Switching via Bus	6
3.1.3	Interconnection Network	6
3.2	Port Queuing	7
3.2.1	Input Port Queuing	7
3.2.2	Output Port Queuing	7
4	Control Plane: ICMP, Routing	7

1 Overview

Network layer protocols transport segments from sending to receiving hosts. The sender encapsulates segments into datagrams and passes them to the link layer. The receiver delivers segments to the transport layer protocol.

Every Internet device uses network layer protocols.

Routers examine header fields in all IP datagrams passing through them

and move datagrams from input interfaces to output interfaces to transfer datagrams along end-to-end paths.

Key network-layer functions:

- **forwarding:** move packets from a router’s input link to appropriate router output links (done for all datagrams very fast)
- **routing:** network-wide process that determines route taken by packets from source to destination to fill forwarding tables (done in the background, takes longer), uses routing algorithms

The **data plane** is a local, per-router function and determines how a datagram arriving on router input interface is forwarded to router output interface.

The **control plane** uses network-wide logic to determine how a datagram is routed among routers along end-to-end paths from source host to destination host.

Control plane approaches:

- **traditional routing algorithms:** implemented in routers (both planes implemented monolithically within a router)
 - individual routing algorithm components in every router interact in the control plane
- **software-defined networking:** explicitly separate the two planes by implementing the control plane as a service in remote servers
 - remote controller computes and installs forwarding tables in routers

1.1 Best Effort Service

The Internet runs on a “best effort” service model, so there are no guarantees on:

- successful datagram delivery to the destination
- timing or order of delivery
- bandwidth available for end-to-end flow

Best effort is simple and has allowed Internet to be widely adopted. Successful provisioning of bandwidth allows performance of real-time applications to be good enough for most of the time.

Replicated, application-layer distributed services connect close to client networks which allows services to be provided from multiple locations.

2 Data Plane: The Internet Protocol

Data plane is:

- connectionless (datagram-based)
- best-effort delivery
 - packets can be lost, delivered out of order, or delayed
- a common packet format for IPv4 and new packet format for IPv6
- global addressing for identifying all hosts (ARP)
- sister protocol that performs error reporting and enables signaling between routers: ICMP (v4, v6)

IPv4 datagram has

- IP version number
- header length (bytes)
- type of service
- total datagram length (bytes)
- 16-bit identifier
- flag
- fragmentation/reassembly and offset info
- time to live: remaining max hops
- upper layer protocol (TCP or UDP)
- header checksum
- 32-bit source IP address
- 32-bit destination IP address
- options (timestamp, record route taken, etc)
- payload data

When no options, overhead is 20 bytes. Upper layer protocol can be:

- ICMP: 1
- TCP: 6
- UDP: 17
- IPv4: 4
- IPv6: 41

If header checksum detects an error, datagram is dropped. Must be recomputed at every hop because of TTL and options.

2.1 Fragmentation and Assembly

Link layer protocols have maximum transfer unit size, which is the largest possible data size in a frame.

Large IP datagrams can be divided (fragmented) within a network, and then reassembled at the final destination. IP header bits are used to identify and order related fragments.

A receiver cannot hold fragments forever and fragments can arrive out of order, so loss of fragments can mean loss of entire datagram. The receiver starts a timer when the first fragment of a datagram arrives. If the timer expires before all the fragments are received, those already received are discarded.

Fragmentation complicated routers and end-systems, which is used by attackers.

2.2 IPv4 Addressing

IP address: 32-bit identifier associated with each host or router interface (about 4 billion total)

Interface: connection between host/router and physical link

- routers typically have multiple interfaces
- hosts typically have 1 or 2 interfaces

2.3 IPv6 Addressing

Possibly not enough 32-bit IPv4 addresses. Also, IPv4 is slow (variable length header). IPv6 allows different network-layer treatment of flows and better mobility management.

IPv6 datagram has:

- IP version
- priority among datagrams in flow
- flow label: identify datagrams in same flow
- payload length
- next header
- hop limit
- 128-bit source address
- 128-bit destination address
- data payload

Compared to IPv4, has no checksum, fragmentation/reassembly, and options.

Not all routers can be upgraded simultaneously so must operate with mixed IPv4 and IPv6.

Tunneling: IPv6 datagram is carried as payload in IPv4 datagram among IPv4 routers (packet within a packet)

44.5% of clients access services via IPv6, so takes time to deploy.

3 What's Inside a Router

High-level view of generic router architecture has **routing processor** (control plane) and **high-speed switching fabric** (data plane).

Input ports have a physical layer, link layer, and decentralized switching (using header field values, lookup output port using forwarding table in input port memory).

Destination-based forwarding: forward based only on destination IP address

Generalized forwarding: forward based on any set of header field values

3.1 Switching Fabrics

Transfer packets from input link to appropriate output link.

Switching rate: rate at which packets can be transferred from inputs to outputs, measured as multiple of input/output line rate

For N inputs, switching rate of N times the line rate is desirable.

Major types of switching fabrics are:

- memory
- bus
- interconnection network

3.1.1 Switching via Memory

Used traditionally, with switching under direct control of CPU.

The packet is copied into system memory and speed is limited by memory bandwidth (2 bus crossings per datagram).

3.1.2 Switching via Bus

Datagram from input port memory to output port memory via shared bus.

Bus contention: switching speed is limited by bus bandwidth

3.1.3 Interconnection Network

Initially developed to connect processors in multiprocessor.

Multistage switch: $n \times n$ switch from multiple stages of smaller switches

With parallelism:

- fragment datagram into fixed length cells on entry
- switch cells through the fabric and reassemble datagram at exit

Can scale by using multiple switching planes in parallel.

Cisco CRS router:

- basic unit has 8 switching planes
- each plane has a 3 stage interconnection network
- up to 100s of Tbps switching capacity

3.2 Port Queuing

3.2.1 Input Port Queuing

If switch fabric slower than input ports combined, queuing may occur at input queues. This can lead to queuing delay and loss due to input buffer overflow.

Head of the Line (HOL) blocking: queued datagram at front of queue prevents others in queue from moving forward

3.2.2 Output Port Queuing

Buffering occurs when arrival rate via switch exceeds output line speed.

Queuing delay (and loss) due to output port buffer congestion (overflow).

Buffering is required when datagrams arrive from fabric faster than the link transmission rate. Must have **drop policy** to decide which datagrams to drop if no free buffers.

Scheduling discipline chooses among queued datagrams for transmission. With priority scheduling, this decides who gets best performance.

4 Control Plane: ICMP, Routing