The Network Layer

Michael Brodskiy
Professor: E. Bernal Mor

November 8, 2023

- Network Layer Overview
  - Transport segment from sending to receiving host
    - Sender: encapsulates segments into packets, passes to link layer
    - Receiver: extracts segments from packets and delivers segments to transport layer protocol
- Network Layer Functions
  - Forwarding: move packets from router's input link to appropriate router's output link
  - Routing: determine route taken by packets from source to destination
    - Routing Algorithms
  - Analogy: Taking a Trip
    - Forwarding: process of getting through single intersection
    - Routing: process of planning trip from source to destination
- Data Plane
  - Local, per-router function
  - Determines hoe packet arriving on router input port is forwarded to router output port
- Control Plane
  - Network-wide logic
  - Determines how packet is routed among routers along end-end path from source host to destination host
  - Two control-plane approaches

    * Traditional routing algorithms: implemented in routers
    * Software-Defined Networking (SDN): implemented in (remote) servers

- Traditional Control Plane Algorithms
  - Individual routing algorithm components in each and every router interact in the control plane
- SDN Control Plane
  - Remote controller interacts with local Control Agents (CAs) to compute, install forwarding tables in routers
- Network Layer Service Model
  - A network layer service model defines the characteristics of end-to-end transport of packets between sending and receiving hosts
  - Examples of possible services (this is only a partial list, there are countless variants):
    * Guaranteed delivery
    * Guaranteed delivery with bounded delay
    * In-order packet delivery
    * Guaranteed minimum transmission rate
    * Security
  - Services provided by the network layer: two main options
    1. Connection-oriented service
       * A path from source all the way to destination must be established before any data packets can be sent
         · This connection is called a Virtual Circuit (VC)
         · The network is called a virtual-circuit network
         · Each VC requires router table space and reservation of resources
       * Designed to provide some quality of service (QoS) (*i.e.* maximum delay guarantees, minimum losses, minimum throughput guarantees, etc.)
       * Example: Asynchronous Transfer Mode (ATM) $\rightarrow$ popular in the 90s early 200, being replaced by all-IP architecyres
    2. Connectionless service
       * Best-effort service
       * Packets are injected into the network individually and routed independently of each other
       * No advance setup is needed
       * No error or flow service functionalities provided

      - The transport layer might do something end-to-end
      - The link layer might do something at the link level
    - For example, IP (internet protocol)

- Reflections on Best-Effort Service
  - Simplicity of mechanism has allowed Internet to be widely deployed and adopted
  - Sufficient provisioning of capacity allows performance of real-time applications (*e.g.* interactive voice, video) to be "good enough" for "most of the time"
  - Replicated, application-layer distributed services (data centers, content distribution networks) connecting close to clients' networks, allow services to be provided from multiple locations
  - Congestion control at the transport layer of "elastic" services helps

- Input Ports
  - Decentralized Switching:
    - Using header field values, lookup output port using forwarding table in input port memory ("match plus action")
      - Destination-based forwarding: forward based only on destination IP address (traditional)
      - Generalized forwarding: forward based on any set of header field values
      - Input port queueing: if packets arrive faster than forwarding rate into switch fabric

- Input Port Queueing
  - If switch fabric slower than input ports combined → queueing may occur at input queues
    - Queueing delay and loss due to input buffer overflow
  - Head-of-the-Line (HOL) blocking: queued packet at front of queue prevents others in queue from moving forward

- Output Ports
  - Buffering required when packets arrive from fabric faster than link transmission rate
  - Drop policy: which packets to drop if no free buffers?
  - Scheduling discipline chooses among queued packets for next transmission
    - FCFS (First Come, First Served), priority, ...

- The Internet Protocol

  - The glue that holds the whole Internet together (data plane)
    - Designed with internetworking in mind
  - Provides a best-effort (no guaranteee) way to transport IP packets (aka datagrams) from source to destination
    - Without regard to whether these machines are on the same network or whether there are other networks between them
  - There are two versions of IP in use today
    - IPv4 (IP version 4)
      - The first "major version" of IP and currently the dominant protocol of the Internet
    - IPv6

- IP Fragmentation
  - Network links have MTU (maximum transmission unit)
    - MTU: largest possible payload in link-level frame → maximum IP packet size
    - Different link types, different MTUs
  - Problem: IP packet larger than MTU of output link
    - Solution: Fragmentation?
      - Typically, IPv6 does not allow fragmentation
      - Typically, TCP does not allow fragmentation

- IP Alternative to Fragmentation
  - If fragmentation is no allowed → "path MTU discovery"
  - Path MTU Discovery
    - Each IPv4 packet is sent with its header bits set to indicate that fragmentation is not allowed to be performed (flag DF=1)
    - Added start-up delat
    - The transport layer can learn about the MTU to adapt the Maximum Segment Size (MSS)

- IP Addressing: Introduction
  - IPv4 Address: 32-bit identifier associated with each host or router interface
  - Interface: connectio between host/router and physical link
    - Router's typically have multiple interfaces
    - Host typically has one or two interfaces (e.g, wired Ethernet, wireless 802.11)

- Subnets

- Device interfaces that can physically reach each other without passing through an intervening router
- IP Addresses have structure:
  - Network portion (aka subnet portion): high order bits
    - Devices in same subnet have common network portion
  - Host portion: remaining low order bits

- IP Addresssing in Subnets: CIDR
  - CIDR: Classless Inter Domain Routing (pronouned "cider")
    - Network portion (aka prefix) of address of arbitrary length
    - Address format (by convention): A.B.C.D.X, where X os the number of bits in the network portion of the address
  - Network address (subnet address): network portion and 0s in the host portion/X
  - Subnet mask: binary mask of 1s in teh subnet portion and 0s in the host portion $\rightarrow$ X
    - The subnet mask can be ANDed with an IP address to obtain the network address
  - Recipe for identifying subnets
    - Detach each interface from its host or router, creating "islands" of isolated networks
    - Each isolated network is a subnet

- Longest Prefix Matching
  - When looking for forwarding table entry for a given destination address, use longest address prefix that matches destination address.

- Forwarding in Access Networks
  - Forwarding tables in routers of an access network have an entry for their subnets
  - When a datagram reaches a router in an access network, it looks at the destination address of the datagram, and checks which subnet inside the network it belongs to. How?
    - AND the destination address with the mask for each subnet entry in the table
    - Check to see if the result is the prefix in the entry

- Forwarding in the Network Core
  - Routers in ISPs and backbones in the middle of the internet must know which way to go to get to every network and no simple default will work

        - This can make for a very large table
            - Routers must perform a lookup in this table for every datagram they forward

- Hierarchical Addressing: Route Aggregation
    - Hierarchical addressing allows efficient advertisement for routing information

- How Are IP Addresses Assigned?
    - Hard-coded by system administrator → fixed IP address
    - DHCP: Dynamic Host Configuration Protocol
        - Can renew its lease on address in use
        - Allows reuse of addresses (only hold address while connected/on)
        - Support for mobile users who join/leave network

- DHCP: More than IP Addresses
    - DHCP can return more than just allocated IP addresses on a subnet:
        - Address of first-hop router for client
        - Name and IP address of local DNS server
        - Subnet mask (indicating network versus host portion of address)

- NAT: Network Address Translation
    - All devices in local network have 32-bit IP addresses in a "private" IP address space that can only be used in local networks
    - "Private" IP address space corresponds to prefixes: 10.0.0.0/8, 172.16.0.0/12 and 192.168.0.0/16
        - Defined by IANA (Internet Assigned Numbers Authority) → department of ICANN
    - Advantages:
        - Private IP addresses can be reused in different private networks
        - Just one IP address needed from provider ISP for all devices
    - Implementation: NAT router (or NAT box)
    - Outgoing datagrams: replace (source IP address, port #) of every outgoing datagram to (NAT IP address, new port #)
        - Remote clients/servers will respond using (NAT IP address, new port #) as destination address
    - Store in NAT translation table every (source IP address, port #) to (NAT IP address, new port) translation pair

- Incoming datagrams: replace (NAT IP address, new port #) in destination fields of every incoming datagram with corresponding (source IP address, port #) stored in NAT table
- NAT has been controversial:
  * Routers "should" only process up to layer 3
  * Address "shortage" should be solved by IPv6
  * Violates end-to-end argument (port number manipulation by network-layer device)
  * NAT traversal: what if client wants to connect to server behind NAT?
- But NAT is here to stay:
  * Extensively used in home and institutional networks, 4G/5G cellular networks

- IPv6: Motivation
  - Initial motivation: 32-bit IPv4 address space would be completely allocated
  - Additional motivation:
    * Speed processing/forwarding: 40-byte fixed length header
      · Extension headers: optional headers can be added after the fixed IPv6 header
    * Enable different network-layer treatment of "flows"

- Transition from IPv4 to IPv6
  - Not all routers can be upgraded simultaneously
    * No "flag delays"
    * How will network operate with mixed IPv4 and IPv6 routers?
  - Tunneling: Packet within a packet
    * IPv6 datagram carried as payload in IPv4 datagram among IPv4 routers ("packet within a packet")
    * TUnneling used extensively in other contexts (4G/5G)

- Flow Table Abstraction
  - Flow: defined by header fields
  - Generalized Forwarding
    * Match: pattern values in packet header fields
    * Actions: for matched packet: drop, forward, modify matched packet or send matched packet to controller
    * Priority: disambiguate overlapping packets
    * Coutners: # bytes and packets

- OpenFlow Abstraction
  - Match and Action: abstraction unifies different kinds of devices
  - Router:
    - Match: longest destination IP prefix
    - Action: forward out a link
  - Firewall:
    - Match: IP addresses and TCP/UDP port numbers
    - Action: permit of deny
  - Swtich
    - Match: destination MAC adress (link layer address)
    - Action: Forward or flood
  - NAT
    - Match: IP address and port
    - Action: rewrite address and port
- Middlebox
  - Any intermediary box performing functions apart from normal, standard functions of an IP router
  - Initially: proprietary (closed) hardware solutions
  - Move towards "whitebox" hardware implementing open API
    - Move away from proprietary hardware solutions
    - Programmable local actions via match and action
    - Move towards innovation/differentiation in software
  - SDN: (logically) centralized control and configuration management often in private/public cloud
    - Network Function Virtualization (NFV): programmable services over white box networking, computation, and storage
      - Allows for programmable network devices