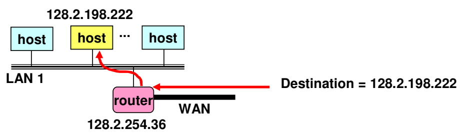
CS 640: Introduction to Computer Networks

Aditya Akella

Lecture 9 -
ARP, IP Packets and Routers

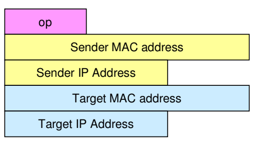
Finding a Local Machine



- Routing Gets Packet to Correct Local Network
  - Based on IP address
  - Router sees that destination address is of local machine
- Still Need to Get Packet to Host
  - Using link-layer protocol
  - Need to know hardware address
- Same Issue for Any Local Communication
  - Find local machine, given its IP address

2

Address Resolution Protocol (ARP)

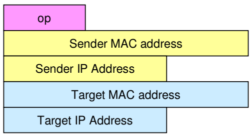


- op: Operation
  - 1: request
  - 2: reply
- Sender
  - Host sending ARP message
- Target
  - Intended receiver of message

  - Diagrammed for Ethernet (6-byte MAC addresses)
- Low-Level Protocol
  - Operates only within local network
  - Determines mapping from IP address to hardware (MAC) address
  - Mapping determined dynamically
    - No need to statically configure tables
    - Only requirement is that each host know its own IP address

3

ARP Request

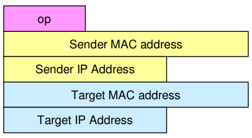


- op: Operation
  - 1: request
- Sender
  - Host that wants to determine MAC address of another machine
- Target
  - Other machine

- Requestor
  - Fills in own IP and MAC address as "sender"
    - Why include its MAC address?
- Mapping
  - Fills desired host IP address in target IP address
- Sending
  - Send to MAC address `ff:ff:ff:ff:ff:ff`
    - Ethernet broadcast

4

ARP Reply



- op: Operation
  - 2: reply
- Sender
  - Host with desired IP address
- Target
  - Original requestor

- Responder becomes "sender"
  - Fill in own IP and MAC address
  - Set requestor as target
  - Send to requestor's MAC address

5

IP Delivery Model

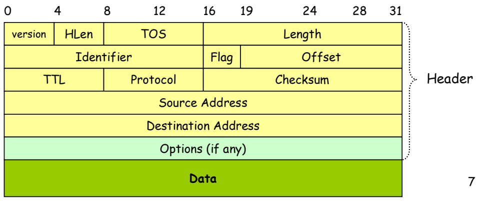
- *Best effort service*
  - Network will do its best to get packet to destination
- Does NOT guarantee:
  - Any maximum latency or even ultimate success
  - Sender will be informed if packet doesn't make it
  - Packets will arrive in same order sent
  - Just one copy of packet will arrive
- Implications
  - Scales very well → simple, dumb network; "plug-n-play"
  - Higher level protocols must make up for shortcomings
    - Reliably delivering ordered sequence of bytes → TCP
  - Some services not feasible
    - Latency or bandwidth guarantees
    - Need special support

6

IP Packets

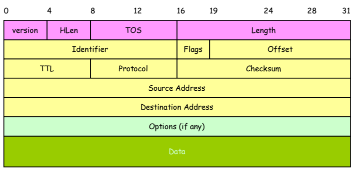
- Low-level communication model provided by Internet
  - Unit: "Datagram"
- Datagram
  - Each packet self-contained
    - All information needed to get to destination
  - Analogous to letter or telegram

IPv4 Packet Format



7

IPv4 Header Fields

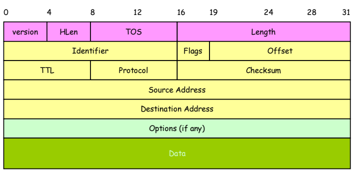


- Version: IP Version
  - 4 for IPv4
  - 6 for IPv6
- HLen: Header Length
  - 32-bit words (typically 5)
- TOS: Type of Service
  - Priority information
- Length: Packet Length
  - Bytes (including header)
- Header format can change with versions
  - First byte identifies version
  - IPv6 header are very different – will see later
- Length field limits packets to 65,535 bytes
  - In practice, break into much smaller packets for network performance considerations

8

IPv4 Header Fields

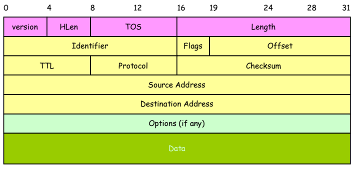
- Identifier, flags, fragment offset → used primarily for fragmentation
- Time to live
  - Must be decremented at each router
  - Packets with TTL=0 are thrown away
  - Ensure packets exit the network



- Protocol
  - Demultiplexing to higher layer protocols
  - TCP = 6, ICMP = 1, UDP = 17…
- Header checksum
  - Ensures some degree of header integrity
  - Relatively weak – only 16 bits
- Options
  - E.g. Source routing, record route, etc.
  - Performance issues at routers
    - Poorly supported or not at all

9

IPv4 Header Fields

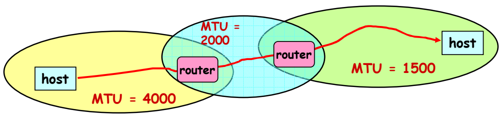


- Source Address
  - 32-bit IP address of sender
- Destination Address
  - 32-bit IP address of destination

- Like the addresses on an envelope

10

IP Fragmentation



- Every Network has Own Maximum Transmission Unit (MTU)
  - Largest IP datagram it can carry within its own packet frame
    - E.g., Ethernet is 1500 bytes
  - Don't know MTUs of all intermediate networks in advance
- IP Solution
  - When hit network with small MTU, fragment packets
    - Might get further fragmentation as proceed farther

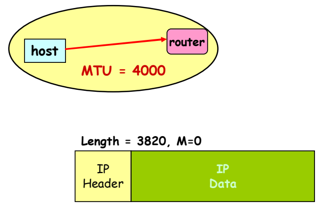
11

Fragmentation Related Fields

- Length
  - Length of IP fragment
- Identification
  - To match up with other fragments
- Fragment offset
  - Where this fragment lies in entire IP datagram
- Flags
  - "More fragments" flag
  - "Don't fragment" flag

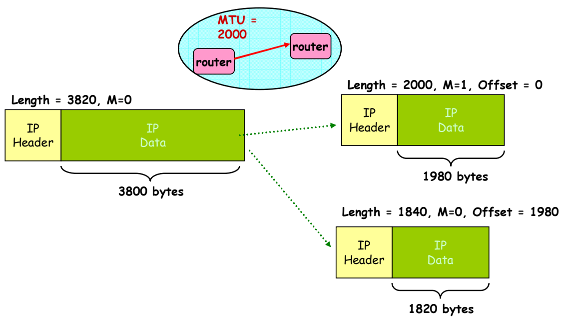
12

IP Fragmentation Example #1



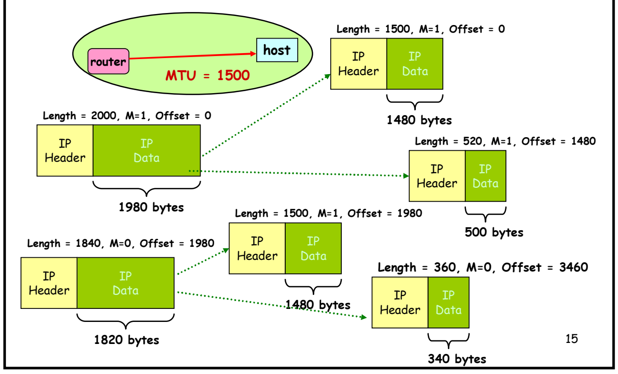
13

IP Fragmentation Example #2



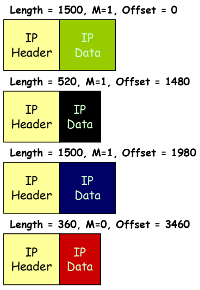
14

IP Fragmentation Example #3



15

IP Reassembly



- Fragments might arrive out-of-order
  - Don't know how much memory required until receive final fragment
- Some fragments may never arrive
  - After a while, give up entire process

Reassembly

- Where to do reassembly?
  - End nodes or at routers?
- End nodes -- better
  - Avoids unnecessary work where large packets are fragmented multiple times
  - If any fragment missing, delete entire packet
- Intermediate nodes -- Dangerous
  - How much buffer space required at routers?
  - What if routes in network change?
    - Multiple paths through network
    - All fragments only required to go through to destination

Fragmentation and Reassembly

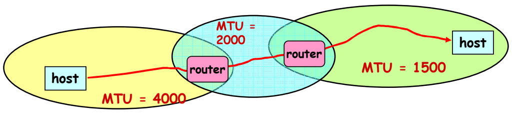
- Demonstrates many Internet concepts
  - Decentralized
    - Every network can choose MTU
  - Connectionless
    - Each fragment contains full routing information
    - Fragments can proceed independently and along different routes
  - Complex endpoints and simple routers
    - Reassembly at endpoints
- Uses resources poorly
  - Forwarding, replication, encapsulations costs
  - Worst case: packet just bigger than MTU
  - Poor end-to-end performance
    - Loss of a fragment
- How to avoid fragmentation?
  - **Path MTU discovery protocol** → determines minimum MTU along route
  - Uses ICMP error messages

Internet Control Message Protocol (ICMP)

- Short messages used to send error & other control information
- Examples
  - Echo request / response
    - Can use to check whether remote host reachable
  - Destination unreachable
    - Indicates how far packet got & why couldn't go further
  - Flow control (source quench)
    - Slow down packet delivery rate
  - Timeout
    - Packet exceeded maximum hop limit
  - Router solicitation / advertisement
    - Helps newly connected host discover local router
  - Redirect
    - Suggest alternate routing path for future messages

19

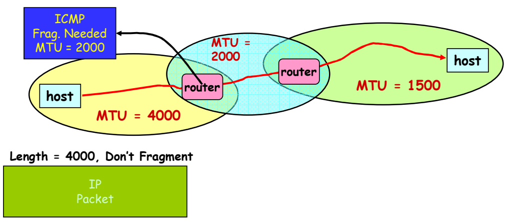
IP MTU Discovery with ICMP



- Operation
  - Send max-sized packet with "do not fragment" flag set
  - If encounters problem, ICMP message will be returned
    - "Destination unreachable: Fragmentation needed"
    - Usually indicates MTU encountered

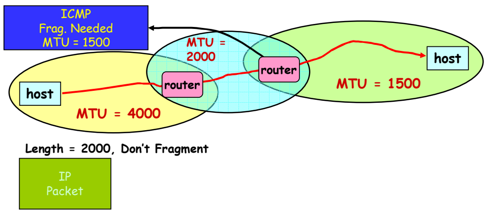
20

IP MTU Discovery with ICMP



21

IP MTU Discovery with ICMP

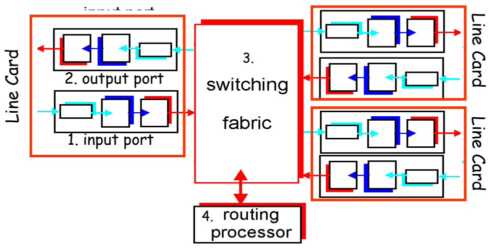


22

Router Architecture Overview

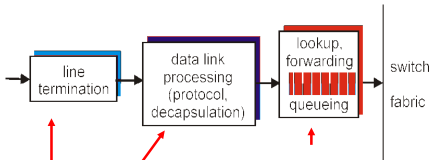
Two key router functions:

- Run routing algorithms/protocol (RIP, OSPF, BGP)
- *Switching* datagrams from incoming to outgoing link



23

Line Card: Input Port



Physical layer:
bit-level reception

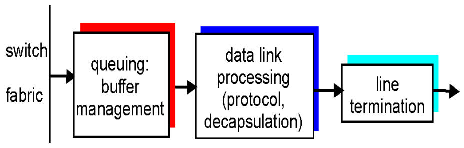
Data link layer:
e.g., Ethernet

Decentralized switching:

- Process common case ("fast-path") packets
  - Decrement TTL, update checksum, forward packet
- Given datagram dest., lookup output port using routing table in input port memory
- Queue needed if datagrams arrive faster than forwarding rate into switch fabric

24

Line Card: Output Port



- Queuing required when datagrams arrive from fabric faster than the line transmission rate

25

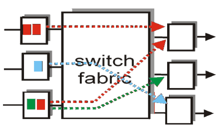
Buffering

- 3 types of buffering
  - Input buffering
    - Fabric slower than input ports combined → queuing may occur at input queues
      - Can avoid any input queuing by making switch speed = N x link speed
      - But need output buffering
  - Output buffering
    - Buffering when arrival rate via switch exceeds output line speed
  - Internal buffering
    - Can have buffering inside switch fabric to deal with limitations of fabric
- What happens when these buffers fill up?
  - Packets are THROWN AWAY!! This is where (most) packet loss comes from

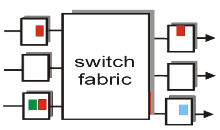
26

Input Port Queuing

- Which inputs are processed each slot – schedule?
- Head-of-the-Line (HOL) blocking: datagram at front of queue prevents others in queue from moving forward



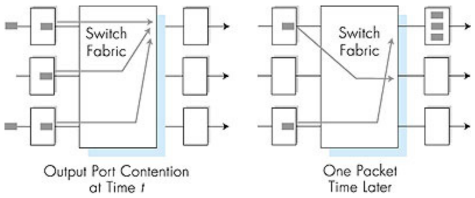
output port contention at time t - only one red packet can be transferred



green packet experiences HOL blocking

27

Output Port Queuing



- Scheduling discipline chooses among queued datagrams for transmission
  - Can be simple (e.g., first-come first-serve) or more clever (e.g., weighted round robin)

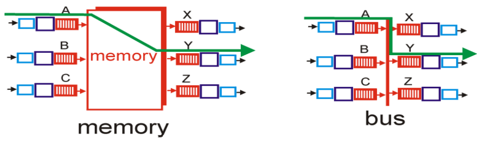
28

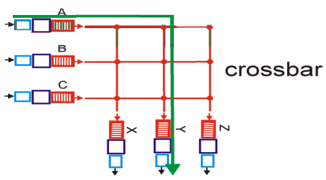
Network Processor

- Runs routing protocol and downloads forwarding table to forwarding engines
- Performs "slow" path processing
  - ICMP error messages
  - IP option processing
  - Fragmentation
  - Packets destined to router

29

Three Types of Switching Fabrics

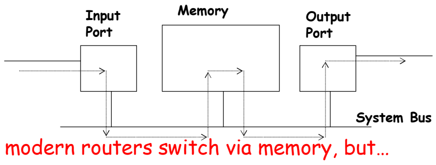




30

Switching Via a Memory

First generation routers → looked like PCs
- Packet copied by system's (single) CPU
- Speed limited by memory bandwidth (2 bus crossings per datagram)

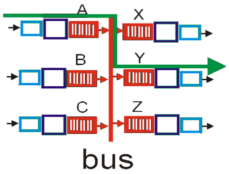


Most modern routers switch via memory, but…
- Input port processor performs lookup, copy into memory
- Cisco Catalyst 8500

31

Switching Via a Bus

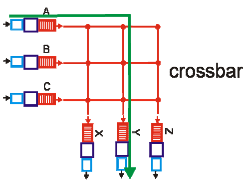
- Datagram from input port memory to output port memory via a shared bus
- Bus contention: switching speed limited by bus bandwidth
- 1 Gbps bus, Cisco 1900: sufficient speed for access and enterprise routers (not regional or backbone)



32

Switching Via an Interconnection Network

- Overcome bus and memory bandwidth limitations
- Crossbar provides full NxN interconnect
  - Expensive
  - Uses 2N buses
- Cisco 12000: switches Gbps through the interconnection network



33