ECE 463
Introduction to Computer Networks

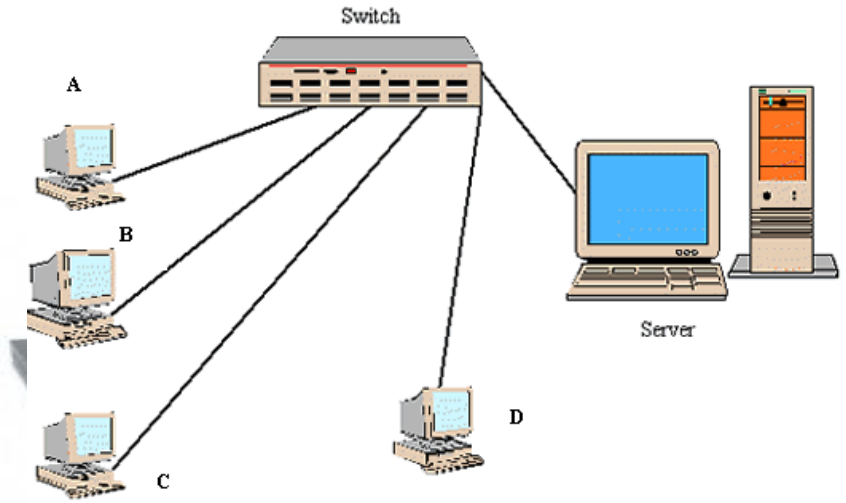
Lecture: Fragmentation

Sanjay Rao

ARP Demo

- /sbin/arp -n (shows arp table)
- netstat -rn (shows IP routing table)
- /usr/sbin/tcpdump -n arp host [hostip]
  (tcpdump of ARP traffic)

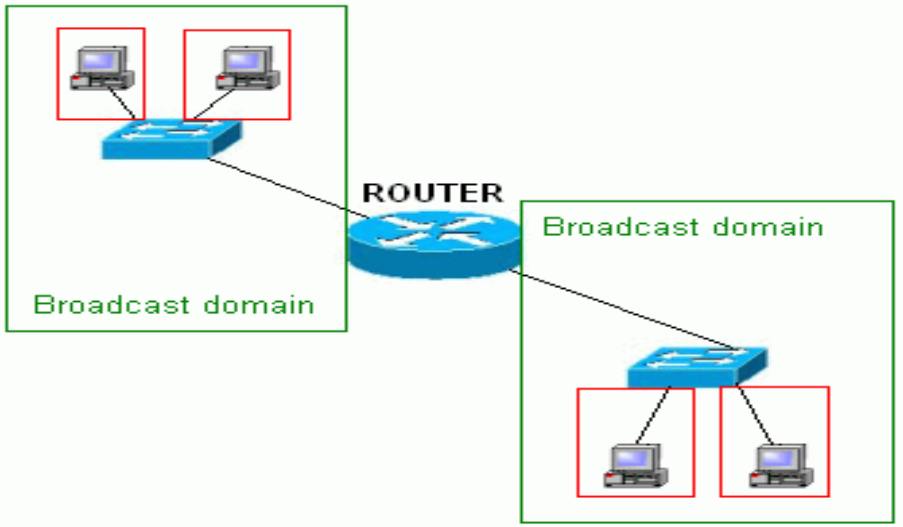
What is a Switch?



- Switch: Networking device that operate on Ethernet frames – Layer 2 devices
- Forward pkts to the destination's MAC address.
- Plug-and-Play devices - Self-configuring without hardware or software changes.
- Will propagate broadcast packets (ARP)
- Will not show up in output of "traceroute"

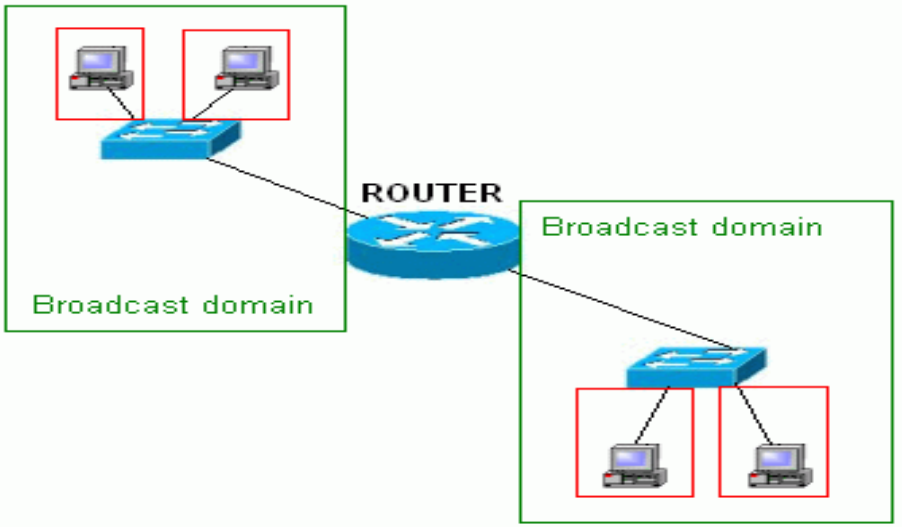
- Example:
  - Host A can have arbitrary MAC address. No "rule" dictating what it should be.
  - Switch will learn address of A over time.
  - A need not be configured with information about the switch.

What is a Router?



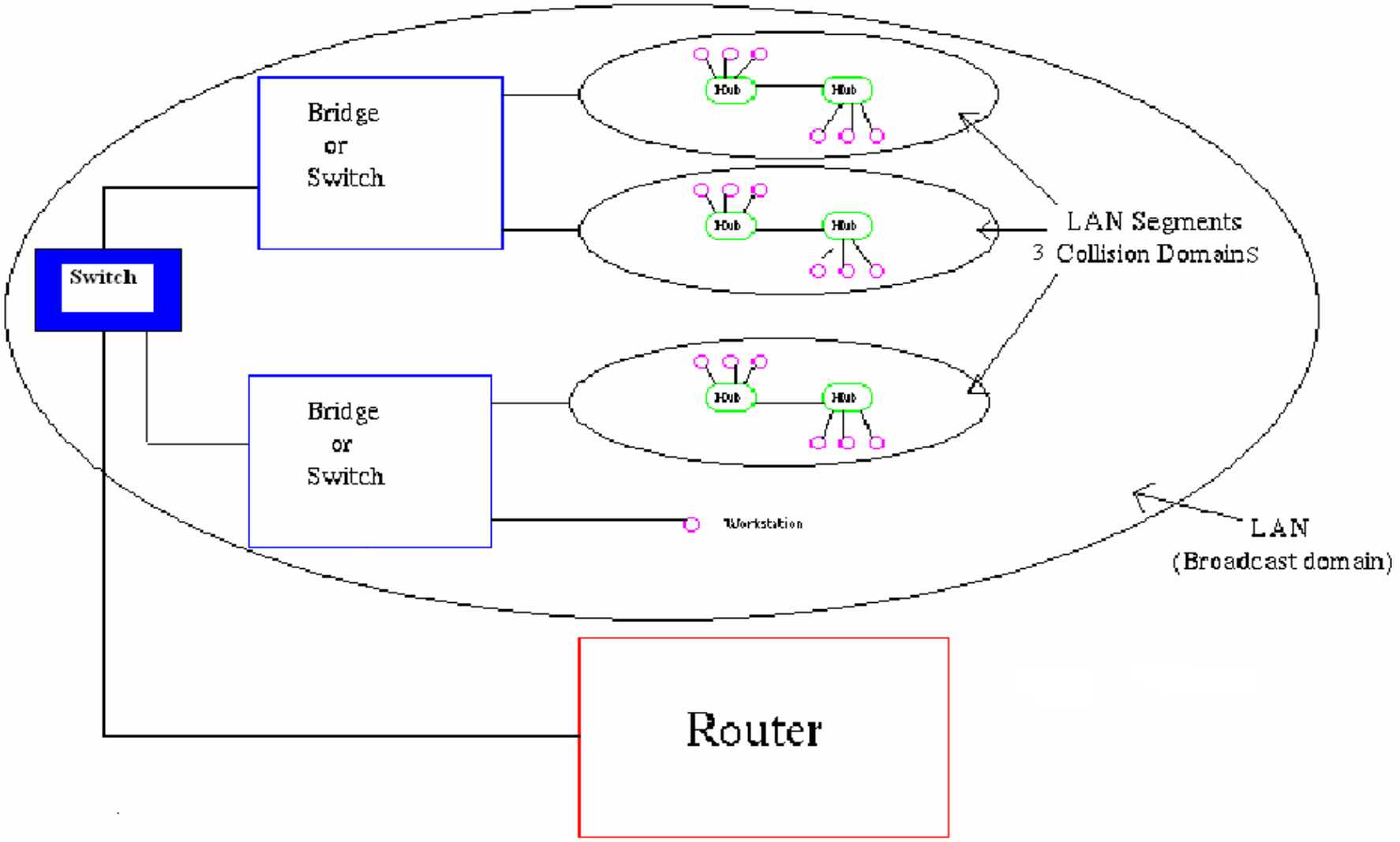
- Router: networking device - forwards data packets across an inter-network.
- Routers operate on Layer 3 of OSI model and use IP address information of the destination to forward the packet.
- Will not propagate broadcast packets (ARP)
- Are not Plug-and-play devices, hosts connected to the routers need their IP addresses to be configured.
  - Router must be configured to indicate packets of certain subnet must be directed on particular interface.
  - IP address of host must be carefully configured to match subnet it is on.
  - Host also configured with router information (typically)

Broadcast Domain



- Broadcast Domain: 2 hosts in same broadcast domain if a broadcast packet (e.g. ARP packet) sent by one of the hosts will also reach the other host.
- Different than “Collision Domain”.
- Collision Domain: logical network segment where data packets can "collide" with one another for being sent on a shared medium at the same time.

Physical View of a LAN

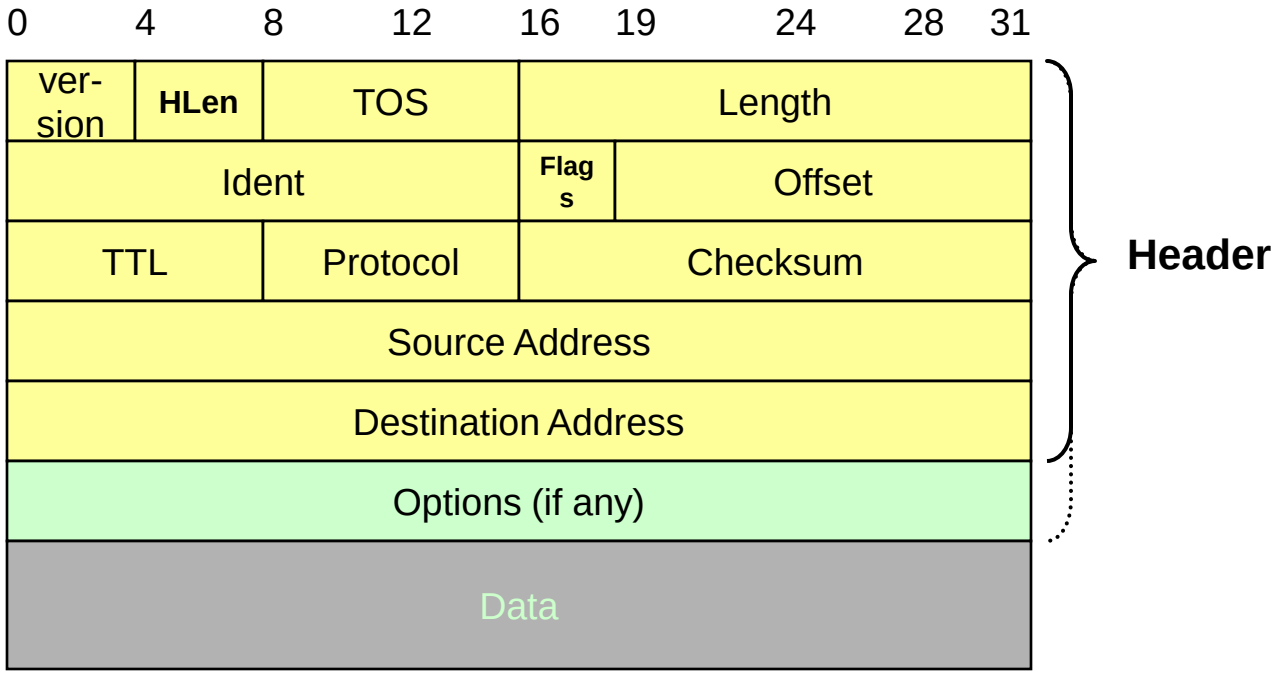


IP Service Model

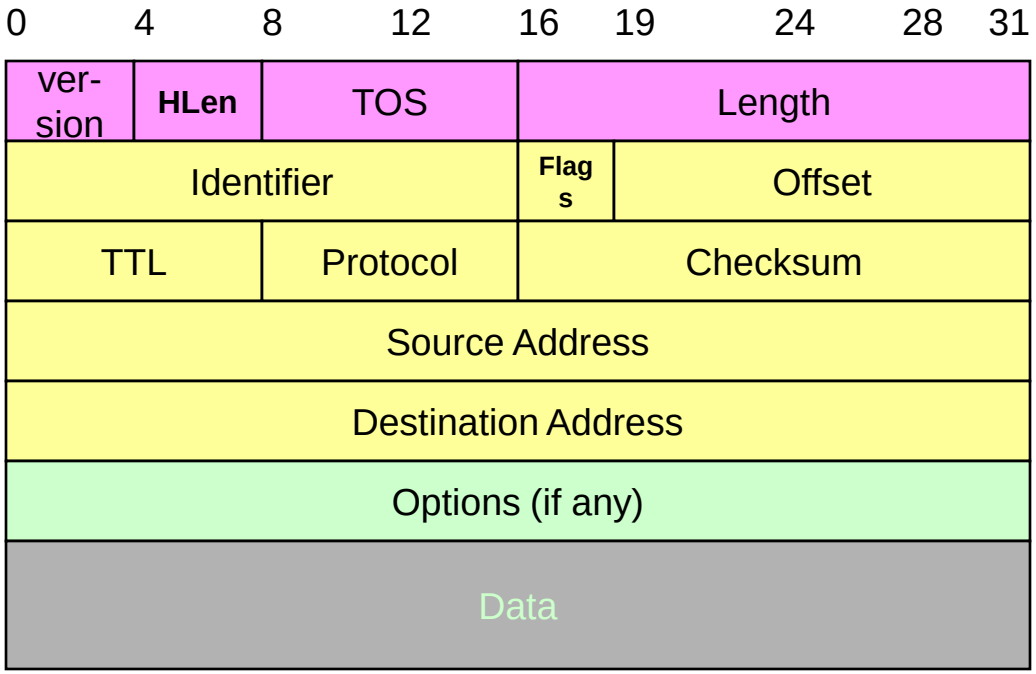
- Low-level communication model provided by Internet

- Datagram
  - Each packet self-contained
    - All information needed to get to destination
    - No advance setup or connection maintenance
  - Analogous to letter or telegram

IPv4 Packet Format



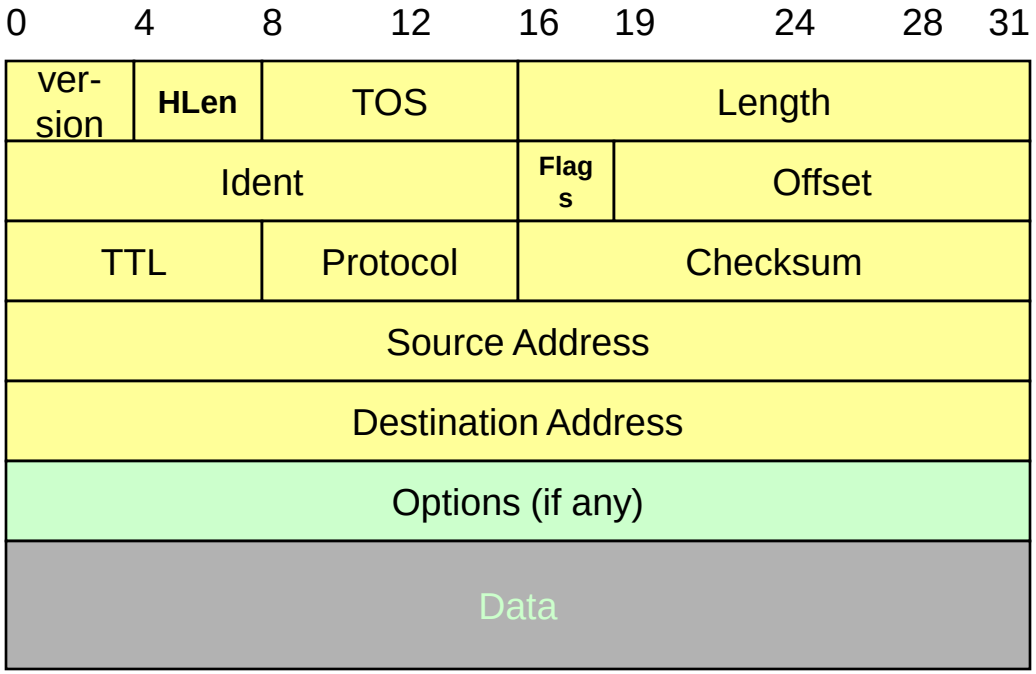
IPv4 Header Fields: Word1



- Version: IP Version
  - 4 for IPv4
- 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
- Length field limits packets to 65535 bytes
  - In practice, break into much smaller packets for network performance considerations

Other Fields

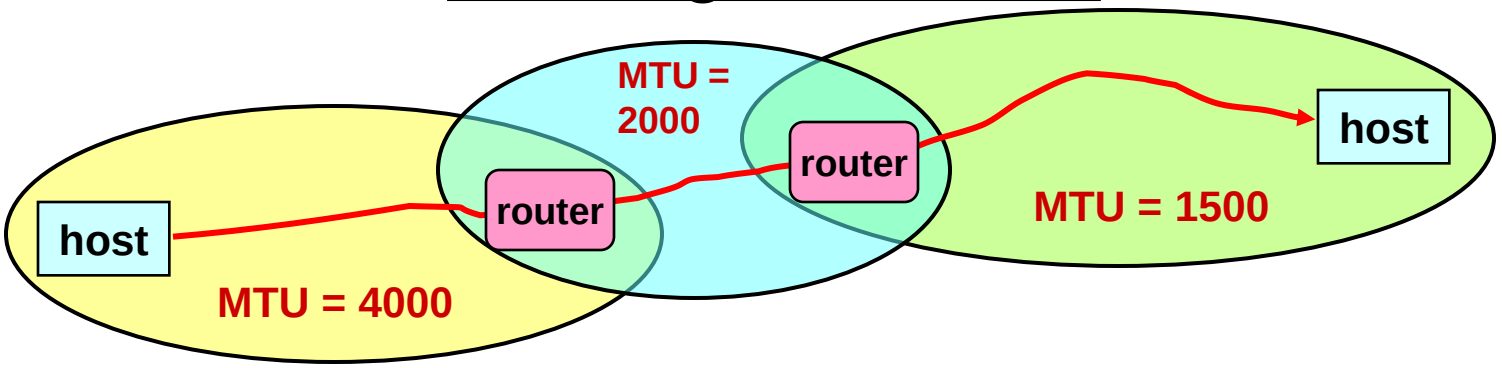


TTL: "Time to Live": decremented by 1 hop at each router

Protocol: For Demultiplexing. E.g. UDP/TCP

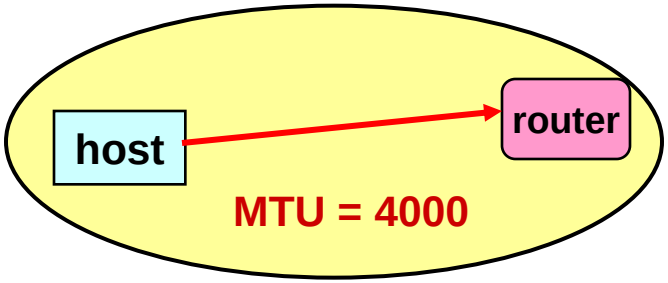
Checksum: On header

IP Fragmentation



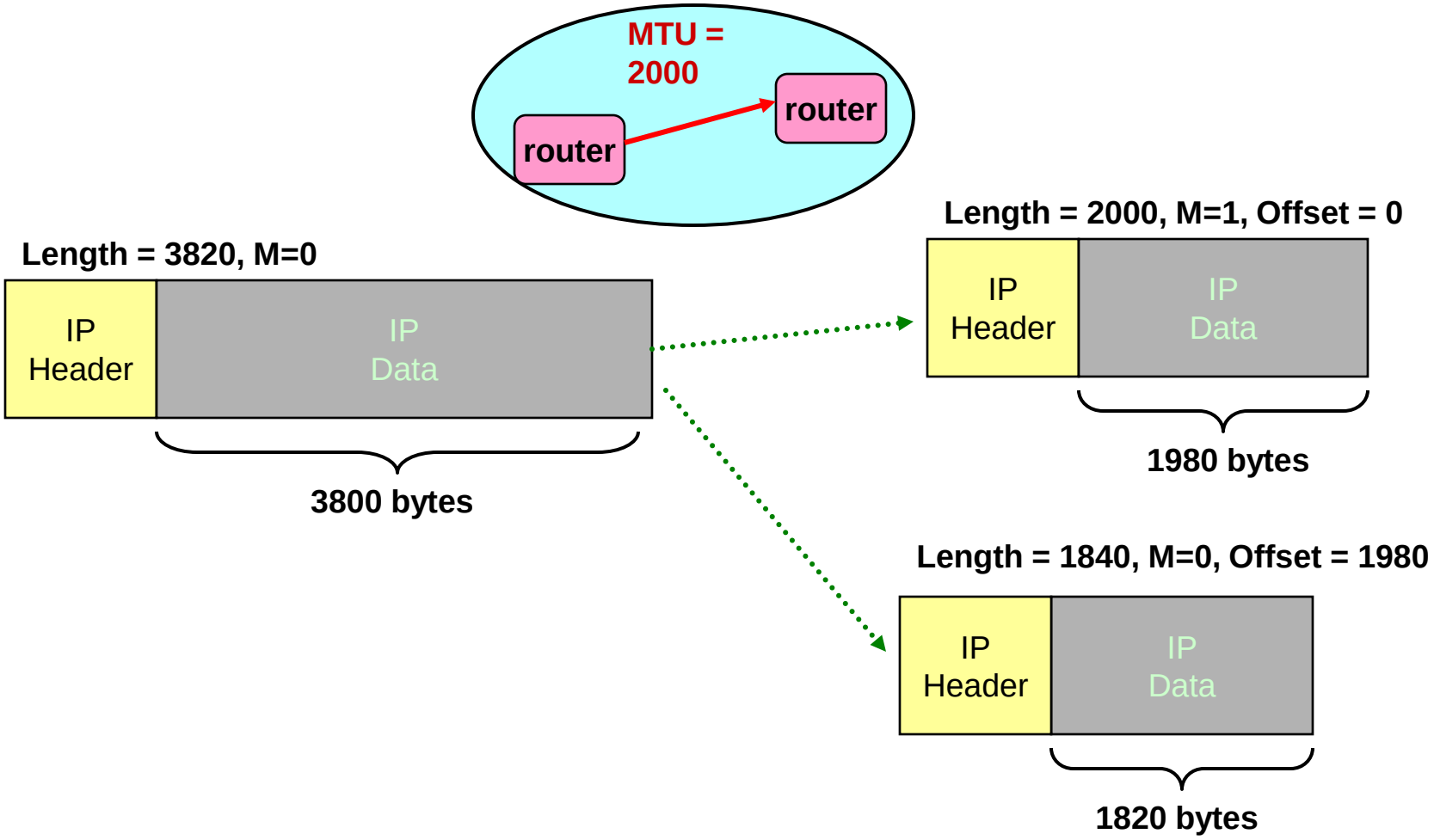
- Every Network has Own Maximum Transmission Unit (MTU)
  - Largest IP datagram it can carry within its own packet frame
    - E.g., Ethernet is 1500 bytes
  - Don't know MTUs of all intermediate networks in advance
- IP Solution
  - When hit network with small MTU, fragment packets
    - Might get further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

IP Fragmentation Example #1

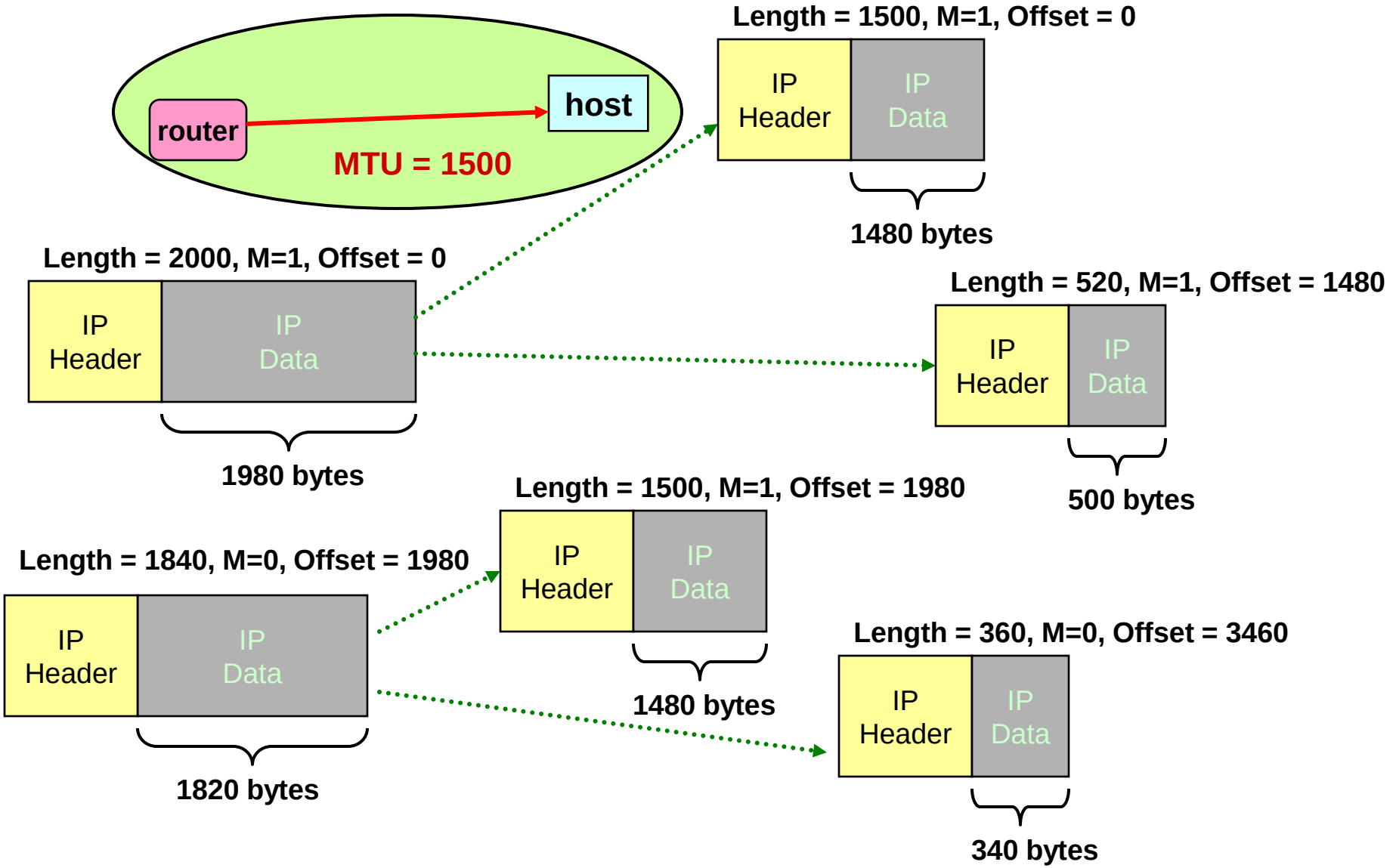


Length = 3820, M=0

IP Fragmentation Example #2

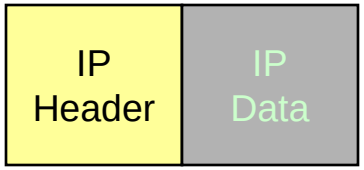


IP Fragmentation Example #3



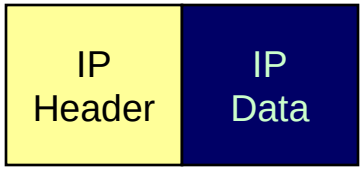
IP Reassembly

Length = 1500, M=1, Offset = 0

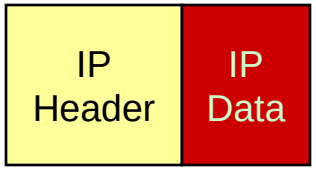


Length = 520, M=1, Offset = 1480

Length = 1500, M=1, Offset = 1980



Length = 360, M=0, Offset = 3460



  - Performed at final destination
- Challenges
  - Fragments might arrive out-of-order
    - Don’t know how much memory required until receive final fragment
  - Some fragments may be duplicated
    - Keep only one copy
  - Some fragments may never arrive
    - After a while, give up entire process
  - Significant memory management issues
    - See code in book

Frag. & Reassembly Concepts

  - Demonstrates Many Internet Concepts
- Decentralized
  - Every network can choose MTU
- Connectionless Datagram Protocol
  - Each (fragment of) packet contains full routing information
  - Fragments can proceed independently and along different routes
- Fail by Dropping Packet
  - Destination can give up on reassembly
  - No need to signal sender that failure occurred
- Keep Most Work at Endpoints
  - Reassembly

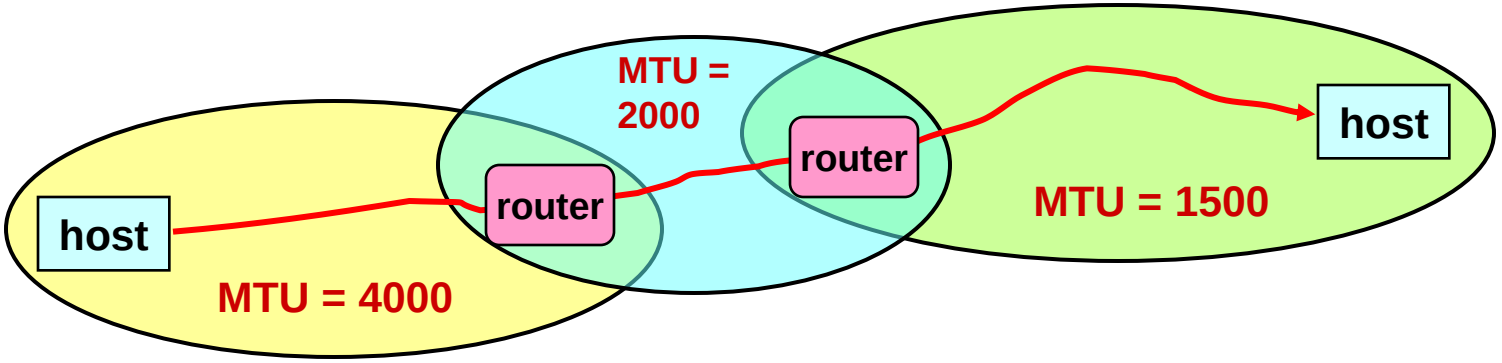
Frag. & Reassembly Reality

- Reassembly Fairly Expensive
  - Copying, memory allocation
  - Want to avoid
- MTU Discovery Protocol
  - Protocol to determine MTU along route
    - Send packets with “don’t fragment” flag set
    - Keep decreasing message lengths until packets get through
  - Assumes every packet will follow same route
    - Routes tend to change slowly over time
- Common Theme in System Design
  - Assure correctness by implementing complete protocol
  - Optimize common cases to avoid full complexity

Internet Control Message Protocol (ICMP)

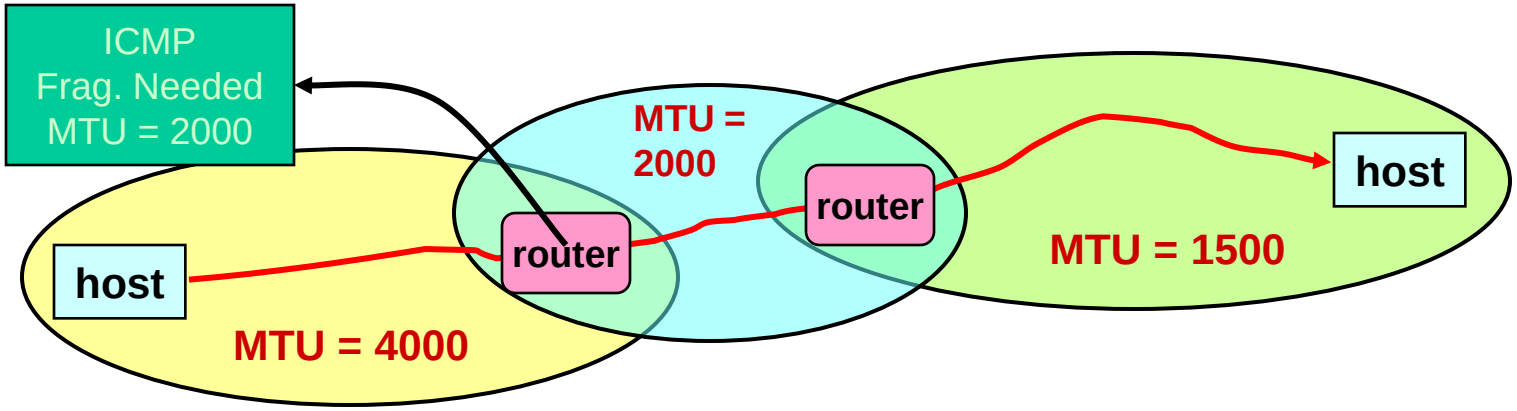
- Short messages used to send error & other control information
- Examples
  - Ping request / response
    - Can use to check whether remote host reachable
  - Destination unreachable
    - Indicates how packet got & why couldn't go further
  - Flow control
    - Slow down packet delivery rate
  - Redirect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

IP MTU Discovery with ICMP



- Exploit Properties
  - Typically send series of packets from one host to another
  - Typically, all will follow same route
    - Routes remain stable for minutes at a time
  - Makes sense to determine path MTU before sending real packets
- Operation
  - Send max.-sized packet with “do not fragment” flag set
  - If encounters problem, ICMP message will be returned
    - “Destination unreachable: Fragmentation needed”
    - Indicates MTU encountered

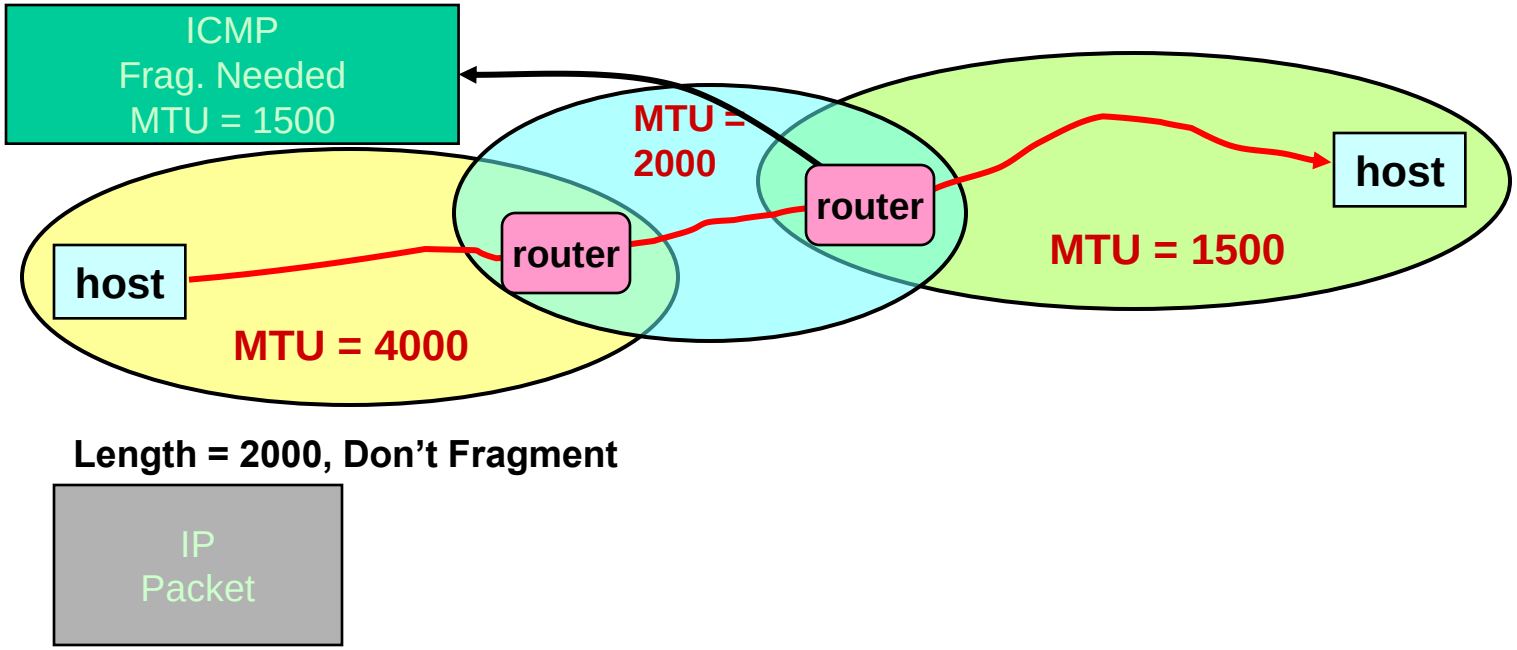
IP MTU Discovery with ICMP



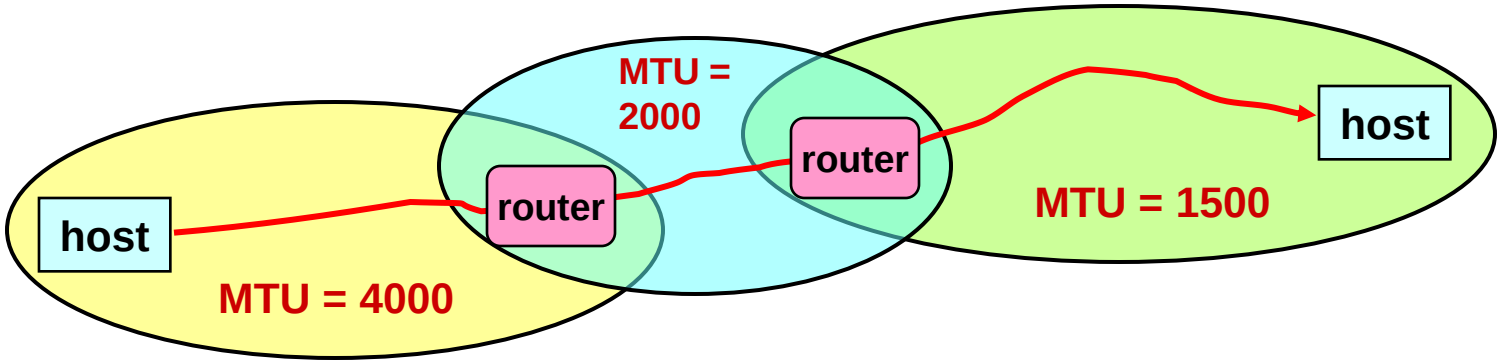
Length = 4000, Don't Fragment

IP
Packet

IP MTU Discovery with ICMP



IP MTU Discovery with ICMP



Length = 1500, Don't Fragment

IP
Packet

- When successful, no reply at IP level
  - "No news is good news"
- Higher level protocol might have some form of acknowledgement

Summary: Important Concepts

- Ideas in the Internet
  - Base-level protocol (IP) provides minimal service level
    - Allows highly decentralized implementation
    - Each step involves determining next hop
    - Most of the work at the endpoints
  - Use ICMP for low-level control functions
  - Fragmentation/Reassembly

Longest Prefix Matching Example

Routing Table

Network	Next Hop	3rd Octet
128.96.170.0/23	Interface 0	1010 1010
128.96.168.0/23	Interface 1	1010 1000
128.96.166.0/23	R2	1010 0110
128.96.164.0/22	R3	1010 0100
(default)	R4	

Packet to destination 128.96.167.151: 1010 0111

- Matches two entries.
- Forwarded to R2 (Longer Prefix Match)