EP2120 Internetworking

IK2218 Protocols and Principles of the Internet

IP, ARP, more IP and ICMP

Lecture 5

György Dán
KTH/EE/LCN

Literature:

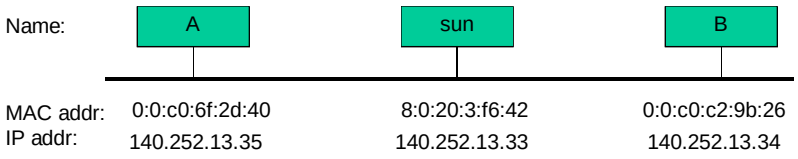
Forouzan, TCP/IP Protocol Suite

(3ed Ch 7-9, 27)(4ed Ch 7-9, 27, 28)

Address Resolution

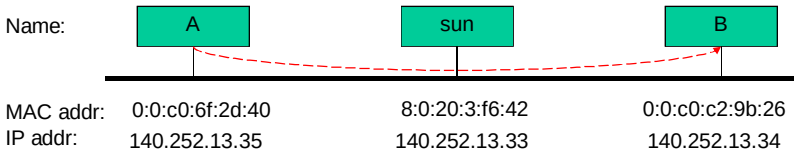
(Address Resolution Protocol – ARP)
(Neighbor Discovery Protocol – ICMPv6)

Logical and Physical Addresses



- A host's network interface card (NIC) has:
  - *MAC address* - hardcoded (physical)
    - e.g., 48-bit Ethernet address
  - *IP address* – configured (logical)
  - *Name* – configured

Communicating with a next-hop



- Problem: send IP datagram from *A* to *B*
  - direct delivery
- Getting the IP address of B
  - Static configuration
  - DNS: Name → Address (Later lectures)
- Getting the MAC address of B
  - Static configuration
  - Dynamic
    - Address Resolution – ARP, Neighbor Discovery - NDP

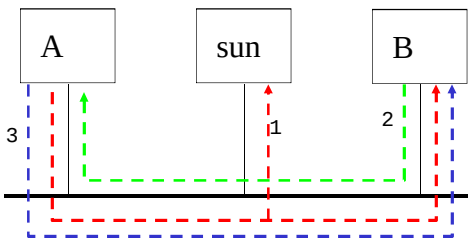
Address Resolution

- Problem:
  - Want to send a packet to an interface on a directly attached network
  - Know the IP-address of the destination but not the MAC address.
- Idea:
  - Send a request
    - "On which MAC address can IP-address *B* be reached?"
  - Wait for reply
    - Host/router with the address *B* replies with its MAC address
- When does the idea work well?

ARP Example (IPv4)

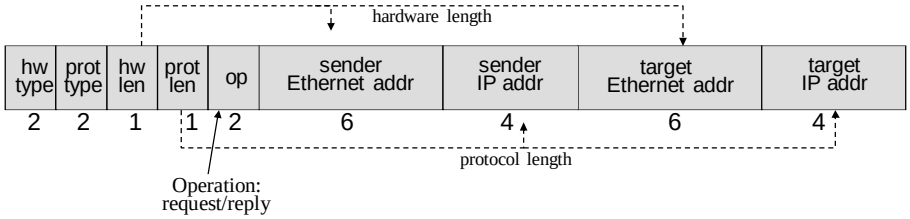
A wants to send an IP datagram to *B* (140.252.13.34)

1. Send an *ARP request* on broadcast to all stations:
   - "Who has 140.252.13.34?"
2. *B* identifies it as its own address and sends an *ARP reply* on unicast back to *A*
   - "I have 140.252.13.34 and my mac address is 0:0:c0:c2:9b:26"
3. *A* sends the datagram to *B* using the resolved MAC address
4. Note that *sun* and *B* can update their ARP caches with *A*!



ARP Packet

- Two length fields
  - Hardware (Ethernet address length: 6)
  - Protocol (IP address length: 4)
- Sender hardware (e.g., *Ethernet)* and protocol (e.g., *IP*) address
- Target hardware (e.g., *Ethernet)* and protocol (e.g., *IP*) address
- ARP packet encapsulated into a link layer frame (e.g., Ethernet)



HW type: Ethernet=1, Prot Type: 0800 = IPv4

ARP Optimizations

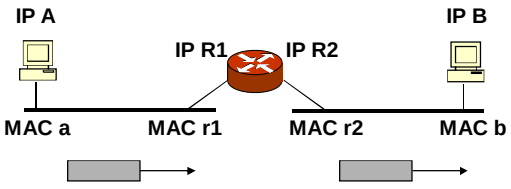
- ARP takes time ⇒ packets can queue up
- Idea: save resolved addresses ⇒ ARP cache
  - Exploit correlations in use of addresses ⇒ Less ARP traffic
  - Entries in the ARP cache should time out (when?)
- Idea: learn from broadcast ARP traffic ⇒ ARP snooping
  - Sender's Internet-to-Physical address binding is in every ARP broadcast
  - Receivers can update their caches before processing the ARP request

ARP Timeouts

- If there is no reply to an ARP request
  - The machine is down or not responding
  - Request was lost, therefore retry (but not too often)
  - Eventually give up (When?)
- ARP cache timeouts
  - completed entry in 20 minutes (BSD Unix)
    - What if the host disappears without notice?
  - incomplete entry in 3 minutes (BSD Unix)

Indirect/Direct Delivery and ARP

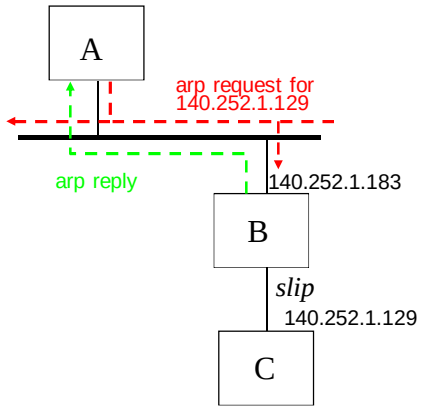
- Ethernet links connect *A* and *B* to *R*
- *A* wants to send an IP packet to *B*



IP Header	**Src: A, Dst: B**	**Src: A, Dst: B**
Ethernet Header	**Src: a, Dst: r1**	**Src: r2, Dst: b**
	Indirect delivery	***Direct delivery***

Proxy ARP (RFC 826)

- Proxy ARP
  - Respond to ARP requests on someone else's behalf
- Allows sub-networks to be hidden
- Example
  - *C* is hidden behind *B*: *B* responds on behalf of *C*



Gratuitous ARP

- Host sends an ARP request of its own address
- Why?
  - Inform other hosts of the IP address at boot time (possibly a new address) – they can update their cache entries immediately
  - Notify other hosts after reconfiguring the IP address
  - Detect IP address conflicts ⇒ "duplicate IP address sent from Ethernet address a:b:c:d:e:f"
- Note
  - hosts hear the broadcast, so they can cache this information - this is one of the ways the proxy ARP server could know the mapping
  - faking that you are another machine can be used to provide failover for servers

IP Network layer functions (again)

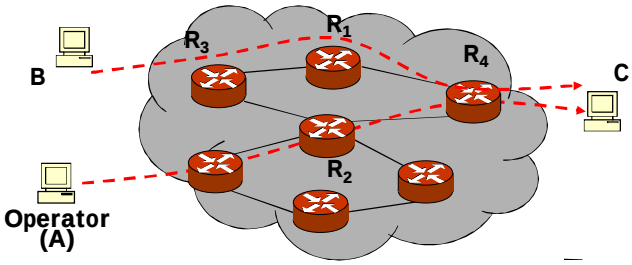
- Logical addressing
- Routing
- Forwarding
- Fragmentation
- Multiplexing/demultiplexing
- Error detection + avoidance
- QoS (??????)

Network maintenance

Diagnosis

Error handling

Motivation - A simple example



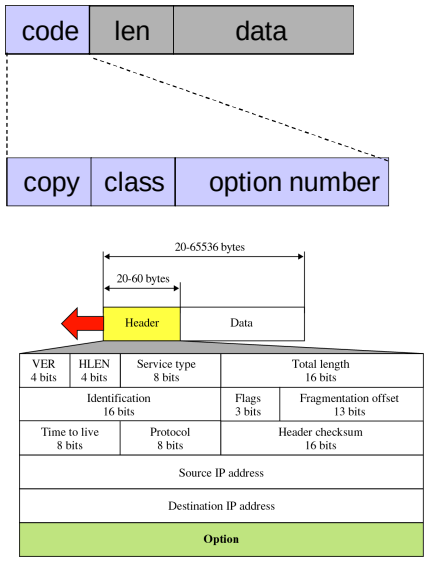
- How are the forwarding tables configured?
- Why cannot host B communicate with host C?
- Why is TCP slow from host B to host C?

IP diagnosis and control

- Purpose
  - Control, testing and debugging of the network functionality
- Implementation
  - Variable size (and optional…)
  - Come after the fixed header
  - Contiguous (no separators)
  - But very different in IPv4 and IPv6…
    - IPv4: Max 40 bytes (Max header length is 60 bytes in IPv4)
    - IPv6: No limitations
- Processing
  - All IP implementations must recognize options/extension headers
    - In practice some implementations do not…
  - Can ignore unknown options

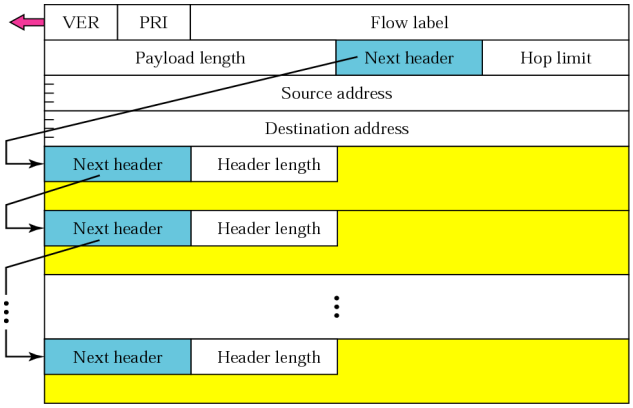
IPv4 Options Encoding

- Two kinds
  - Single byte (only code)
  - Multi-byte
- Option Code: 1 byte
  - Copy (to fragments) (1 bit)
  - Class (2 bits)
    - 0 (00): Datagram or network control
    - 2 (10): Debugging and measurement
  - Number (5 bits)
- Option Length (len): 1 byte defines total length of option (including code and len fields)
- Data: option specific



IPv6 Extension headers Encoding

- Chain of extension headers after IPv6 header
  - No practical limit on number of extension headers
  - Order specified, occurs at most once (except Destination options)
- Extension header length multiple of 8 octets

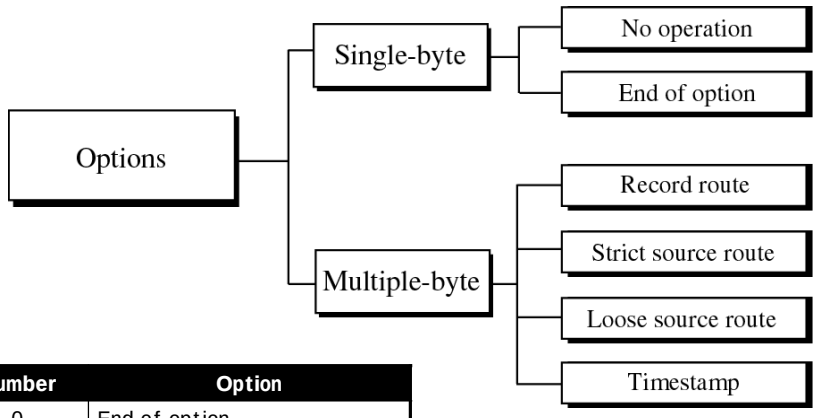


IPv4 options *vs.* IPv6 extension headers

IP diagnosis and control

Purpose	IPv4 Option	IPv6 Extension header
Source routing	Loose/Strict Source Route and Record	Routing
Route recording	Record Route	-
Fragmentation	-	Fragment
Jumbograms		Jumbo payload option (in Hop-by-hop options)
Delay measurement	Timestamp	-
Special attention	Router alert *(proposed)*	Router alert option (in Hop–by-hop options)
Security	AH, ESP* (as payload)	AH, ESP

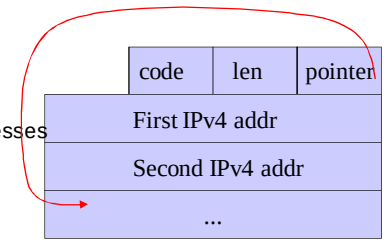
Categories of IPv4 Options

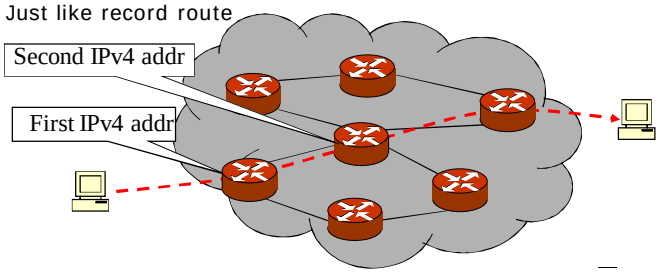


Number	Option
0	End of option
1	No option
3	Loose source route
4	Timestamp
7	Record route
9	Strict source route

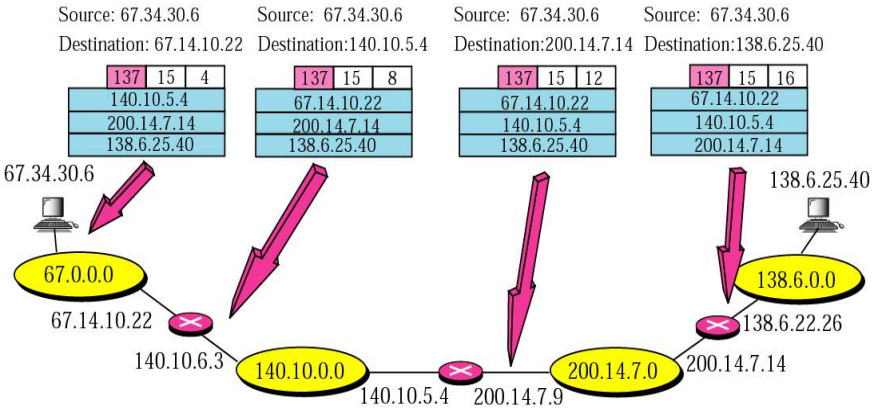
IPv4 Options: Source Route

- Purpose
  - Strict Source routing (SSRR)
    - The path is *exactly* as specified
  - Loose Source Routing (LSRR)
    - The path *includes* the specified addresses
- Operation
  - Sender prefills routers to be visited
  - The routers record their addresses
    - Just like record route





IPv4 Options: Source Route Example

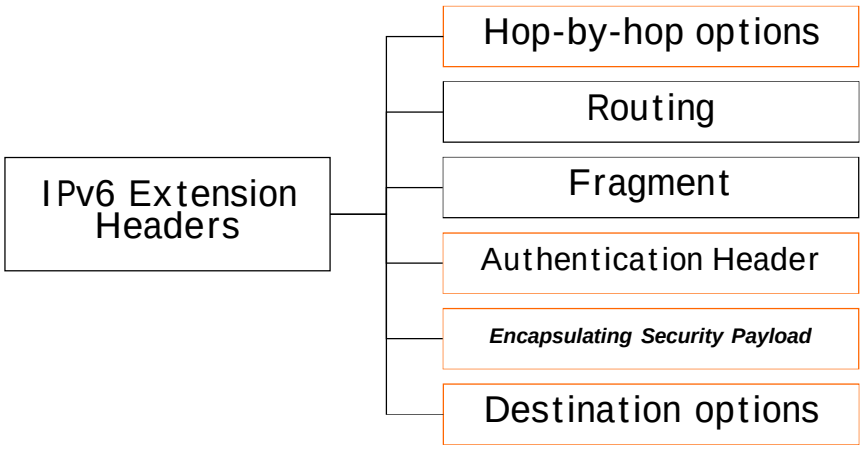


The fig. is erroneous: it records the IP address of the outgoing interfaces.
Why?

Source route: when it comes handy

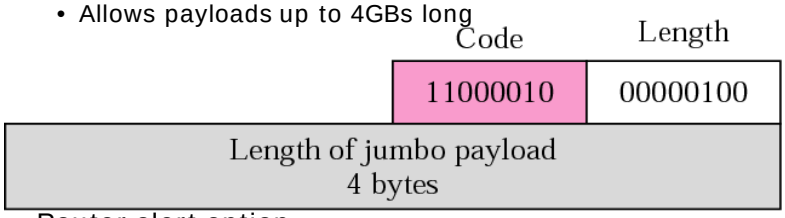
- Troubleshooting
  - Figure out from point "A" why machines “B" and “C" cannot communicate
- Mapping the network
  - Used with traceroute in order to find all the routes between two points on the network
- Performance
  - Force an alternate link to avoid congesting the correct routes w/o changing the forwarding tables (management)
  - Create independent paths for MDC or FEC
- Hacking
  - Can send packets to a host via a trusted third party
    - Normally disabled in routers…

IPv6 Extension Header Types



IPv6 Hop-by-hop options header

- Information to be passed to all routers on the path
  - Must immediately follow the IPv6 header
- Can carry a number of options of different types
  - Pad1 and PadN (padding for alignment)
  - Jumbo Payload option [rfc2675]
    - Allows payloads up to 4GBs long

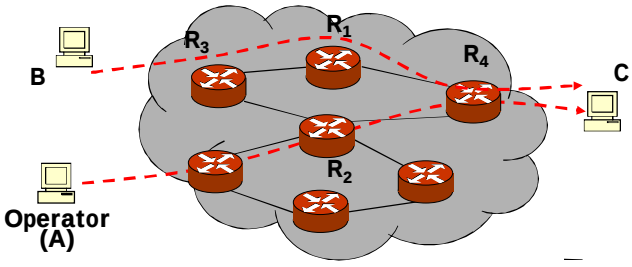


  - Router alert option
    - Datagram requires special attention from router

ICMP

Internet Control Message Protocol v4
RFC 792
Internet Control Message Protocol v6
RFC 2463

More motivation - A simple example



- Is router R1 up and running?
- What is the path MTU between A and C?
- Why do my datagrams get lost?
- Plugged in the cable, but cannot send packets?

Internet Control Message Protocol ICMP

- Signalling protocol for IP
  - Report IP problems back to sender
  - Control and management
- Considered a part of IP, but uses IP for delivery

Query/Informational messages

- Information reporting
- Management

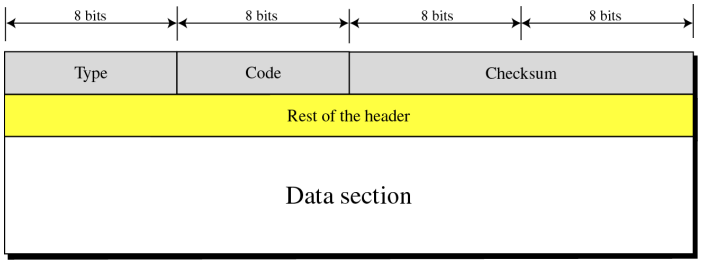
Error reporting messages

- Report errors in
  - Forwarding table configuration
  - IP datagram format
  - Host/service status

ICMPv4 and ICMPv6 Messages

Group	Message	ICMPv4 Type	ICMPv6 Type
Error reporting	Destination unreachable	3	1
	Source quench	4	-
	Time exceeded	11	3
	Parameter problem	12	4
	Redirect	5	-
	Packet too big	-	2
Query (v4)/ Informational (v6)	Echo request/reply	8/0	128/129
	Timestamp request/reply	13/14	-
	Address mask request/reply	17/18	-
	Router solicitation/ advertisement	10/9	133/134
Neighbor Discovery Protocol (NDP,v6)	Neighbor solicitation/ advertisment	-	135/136
	Redirect	-	137
Multicast Listener Discovery Protocol (MLDP)	Multicast Listener Query/Report/Done/Report (v2)	-	130/131/ 132/143

General Format of ICMP Messages



- Type: Specifies type of ICMP message
  - e.g., destination unreachable
- Code: Specifies reason for the particular message type
  - E.g., host/net/port unreachable
- Checksum: covers header and data

ICMP Error Reporting

- Motivation and purpose
  - IP is an unreliable protocol, and errors do occur…
  - Reports errors but does not correct errors
    - One of the main responsibilities of ICMP
    - Error correction left to higher level protocols
- Operation
  - Error message *always* sent back to the *original source*
    - use the source address of the IP datagram to send the error message back to the (probable) originator
  - Error message carries part of the original datagram header and payload
  - Error message *not* sent under certain conditions

Why?

ICMP Error reporting restrictions

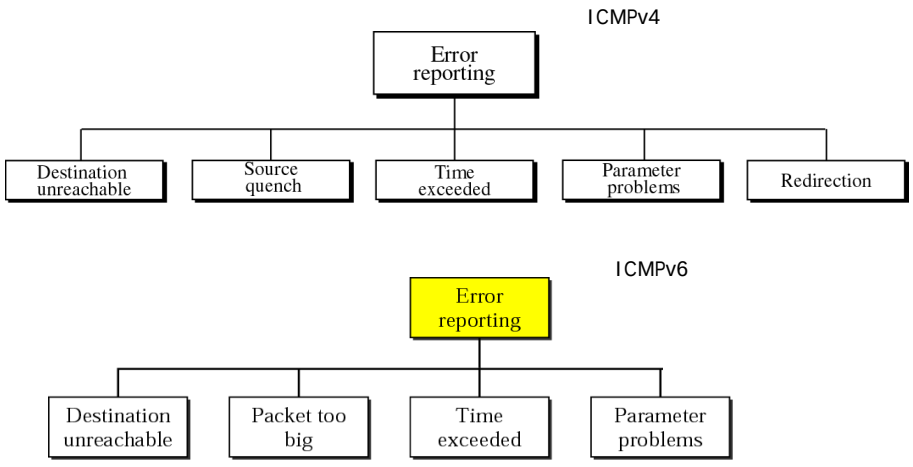
ICMP Error not returned for

- A datagram carrying another ICMP Error
- A datagram destined to IP broadcast or multicast address
- A datagram sent as link-layer broadcast (e.g., Ethernet)
- An IP fragment other than the first
- A datagram whose source address does not define a single host
  (e.g., 0.0.0.0)

Reason

- Avoid loops
- Avoid packet explosions (broadcast storms)

ICMP Error Reporting Messages



ICMP Time Exceeded

Sent in 2 cases

- Router
  - When TTL is zero after decrementation, discards the datagram and sends this back to the source (Code 0)
- Destination host
  - When all fragments of a datagram do not arrive at the destination host within a certain time limit (Code 1)
    - Timer is started upon reception of the first fragment

Do you know a tool that uses ICMP time exceeded to provide network diagnosis?

Tool Using ICMP: Traceroute

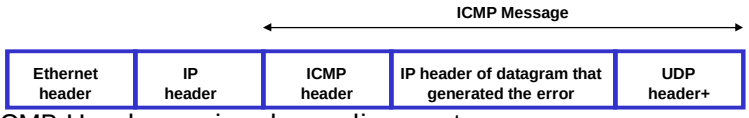
- Traceroute traces a path to a destination by exploring every IP hop on the way
  - Note: only the *receiving* interfaces are traced, not the sending interfaces.
- Traceroute algorithm uses two steps:
  1. Set small TTL fields and receive ICMP time exceeded incrementally
  2. When final host reached, use unlikely UDP port and get ICMP port unreachable back
- Alternative:
  - IP datagram with *record route* option,but
    - RR option not always implemented (e.g., in IPv6)
    - Limited number of hops can be traced due to maximum size of IPv4 options
    - RR records IP addresses of outgoing interfaces

ICMPv6 Packet Too Big

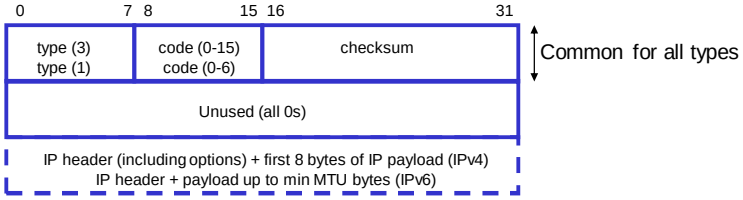
- Router receives datagram larger than MTU
  - Discards datagram
  - Sends message to source incl. MTU
- Can be used for path MTU discovery
  - Like ICMPv4 Destination Unreachable (Fragmentation needed)

ICMP Error reporting header

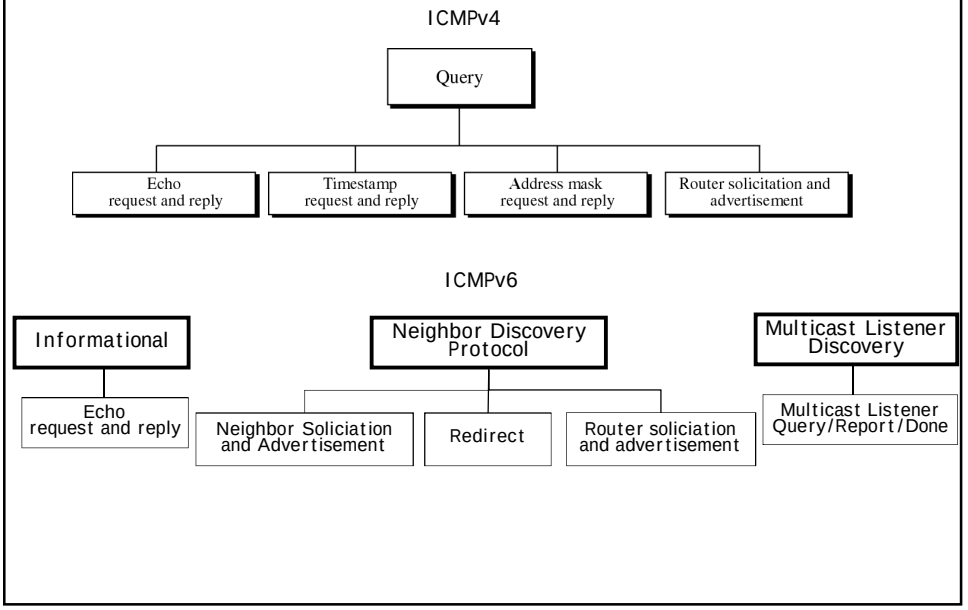
- Carries part of the original IP datagram
  - ICMPv4: IP header (+ options) and 8 bytes of payload (≤576 bytes total datagram size)
  - ICMPv6: IP header + payload (≤min IPv6 MTU)
    - Example: ICMP Destination Unreachable (UDP packet)



- ICMP Header varies depending on type
  - Example: ICMP Destination Unreachable (type 3/type 1)

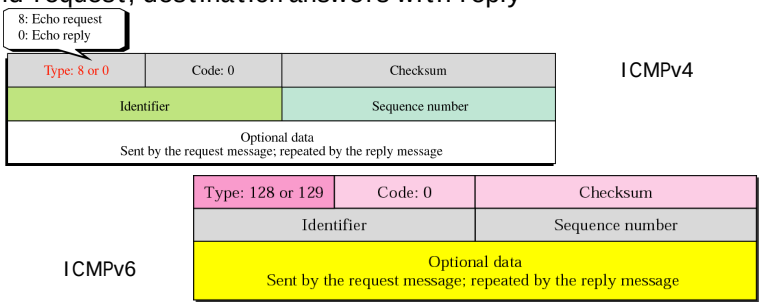


ICMP Query/Informational Messages



ICMP Echo Request and Reply

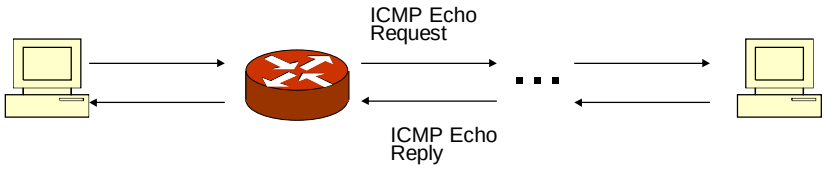
- Purpose: check host reachibility
- Operation
  - Send request, destination answers with reply



- Can you think of a widely used program that uses ICMP Echo request/reply?

The 'Ping' Utility

- Purpose
  - test host reachability
- Operation
  - Sends an ICMP echo request to a node
  - Server replies with ICMP echo reply
    - Almost all IP implementations support Ping server
  - With IPv4 record route (RR) option, the route of the ping datagram can be traced (up to IHL limit)
    - try "ping –r 9 'hostname'"



IPv6 Neighbor Discovery Protocol

- Purpose
  - Assist forwarding of datagrams
    - Address resolution (ARP in IPv4)
    - Forwarding table updates (Redirect in ICMPv4)
    - Router address
- Operation
  - Neighbor solicitation = query
    - Includes the solicitor's physical address in the solicitation for easier processing at the receiver
  - Neighbor advertisement = reply
  - Redirect = forwarding table update necessary
  - Router solicitation/advertisement

ICMP Summary

Error reporting	ICMPv4	ICMPv6
Destination unreachable	Yes	Yes
Source quench	Yes	No
Packet too big	No	Yes
Time exceeded	Yes	Yes
Parameter problem	Yes	Yes
Redirect	Yes	Yes

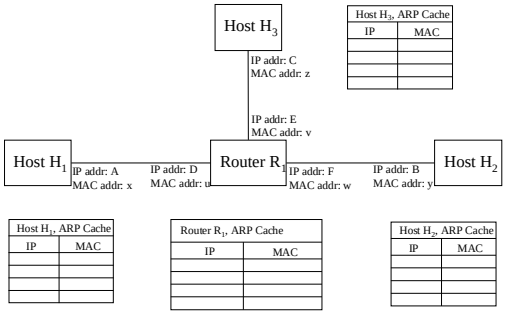
Query/Informational/NDP/MLDP	ICMPv4	ICMPv6
Echo request and reply	Yes	Yes
Timestamp request and reply	Yes	No
Address mask request and reply	Yes	No
Router solicitation and advertisement	Yes	Yes
Neighbor solicitation and advertisement	ARP	Yes
Group membership management	IGMP	Yes

IP and ICMP Summary

- IP is engineered to provide
  - e2e reachability
  - best-effort datagram service
- Each field in the IP header is related to some functionality
  - Logical addressing, Fragmentation, QoS, Bit erro
    Multiplexing/Demultiplexing
- Error reporting provided by ICMP

- IPv4 is a very successful protocol
  - With some design flaws and unused features
  - IPv6 cleaned up the network (IP) layer

ARP - Example



Host H_3, ARP Cache	
IP	MAC

Host H_1, ARP Cache	
IP	MAC

Router R_1, ARP Cache	
IP	MAC

Host H_2, ARP Cache	
IP	MAC

- Three hosts H1, H2 and H3
- Routed network running IPv4
- The IP and MAC addresses of the hosts and the router's interfaces are given in the figure.
- The ARP caches of the hosts and the router are shown. Assume that the ARP caches are initially empty and that no packets have been sent yet.

- Host H1 wants to send an IPv4 unicast datagram to host H3.
  - Fill in the state of the four ARP caches as they will appear after the IPv4 unicast datagram has been delivered to host H3, that is, after dynamic ARP resolution has been made.

UDP, TCP

Summary

- Address resolution
  - ARP
  - ~~(RARP)~~
- IP options
  - Network testing, debugging
  - Network control
- ICMP
  - Error reporting
  - Query
  - NDP, MLDP

ARP, IP, ICMP