Internetworking and IP Datagrams
Ch. 20–23.14

Networks For All Occasions

There are many types of networking hardware.

A computer is usually attached to only one.

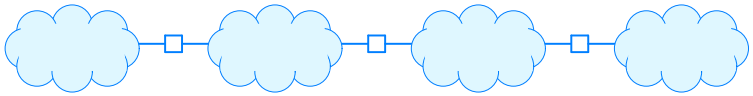
Different nets use different protocols.
Different formats, sizes, addressing schemes.

Exchange between computers on different nets is difficult.

Routers

Networks are connected by routers.

Routers behave like bridges, but the connected networks may be of different types.



Heterogeneous Networking

Even with routers, cross-communication is difficult.

How do you address the recipient?
All senders would need to know all possibilities

How do you know which router(s) to use?

What's the maximum packet size?
Probably the smallest of whatever it's passing through.

So the maximum packet size depends on the route.

Do all the nets support the same high-level protocols?

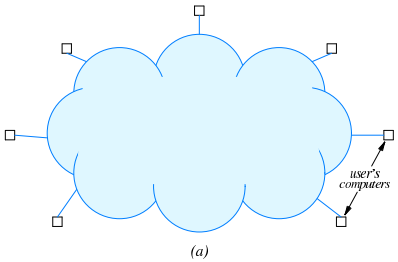
E Pluribus Unum

Internet software provides a unified appearance.

An internet is essentially a virtual network built atop a collection of real networks.

An internet specifies packets and protocols independent of the particular hardware.

A Virtual Network



The abstraction (public view).

The implementation (private view).

Transmission Control Protocol / Internet Protocol (TCP/IP)

An internet protocol.

TCP/IP was the first such protocol.

TCP/IP has never been replaced.
It has evolved.

First funded by ARPA, then NFS.

TCP/IP Stack (Again)

Application	← *LAYER 5*
Transport	← *LAYER 4*
Internet	← *LAYER 3*
Network Interface	← *LAYER 2*
Physical	← *LAYER 1*

The Internet layer provides the abstract network.

Layers above use the abstraction.
Same on any computer.

Layers below depend on the type of physical network.

TCP and IP

The IP is the Internet layer in the stack.

IP provides the abstract network.

This group of slides concerns IP.

TCP is one of several related transport protocols described in the next set of slides.

Hosts and Routers

A host is any attached computer that runs applications.

Routers are attached, but do not run applications.

Plain routers need only IP protocol layers 1, 2, and 3.

Firewalls need 4.

A Virtual Net Requires Virtual Addresses

Each TCP/IP host is assigned an IP address.

These bear no relationship to any hardware device.

These addresses are 32-bit numbers.

Dotted Decimal Notation

Break the 32 bits into 4 bytes.

String the four decimal values together, separated with .

32-bit Binary Number	Equivalent Dotted Decimal
10000001 00110100 00000110 00000000	129.52.6.0
11000000 00000101 00110000 00000011	192.5.48.3
00001010 00000010 00000000 00100101	10.2.0.37
10000000 00001010 00000010 00000011	128.10.2.3
10000000 10000000 11111111 00000000	128.128.255.0

Networks

Addresses are hierarchical, having a network part and a host part.

Routers use the network part to address other routers.

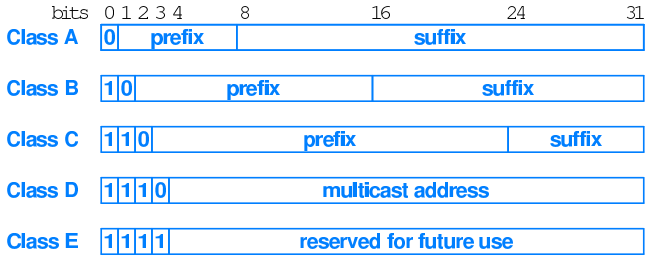
The 32-bit address is divided into network and host part.

All network addresses differ.
Hosts within a network differ.

Different addresses are divided in different places.
Sizes of the network and host parts vary, but always total 32.

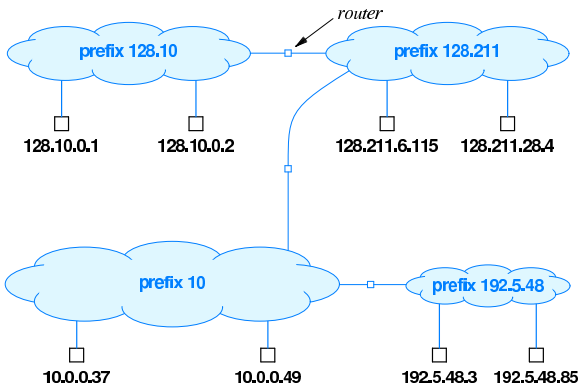
Original Division Scheme

Classful addressing



Class	Range of Values
A	0 through 127
B	128 through 191
C	192 through 223
D	224 through 239
E	240 through 255

Routing Under Classful Addressing



The first few bits tells how to divide each address.

How Many Hosts Would You Like With That Network, Sir?

Address Class	Bits In Prefix	Maximum Number of Networks	Bits In Suffix	Maximum Number Of Hosts Per Network
A	7	128	24	16777216
B	14	16384	16	65536
C	21	2097152	8	256

Classes are not of equal size.

Not Very Efficient

Classful addressing wastes addresses.

Original researchers never expected there would ever be a shortage.

Solution: Allow addresses to be divided in arbitrary places.
More flexible.
Allows networks to be re-divided.

Network Masks

A network is specified as two 32-bit numbers:

- The network number.
- A mask which tells which of the bits in the network number matter
- The bit positions which contain 1's are part of the network number.
- The bit positions which contain 0's are part of the host number.

For Example

Network: 204.198.64.0, Mask: 255.255.192.0

1 1 0 0 1 1 0 0 1 1 0 0 0 1 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0

The network number is:
1 1 0 0 1 1 0 0 1 1 0 0 0 1 1 0 0 1

Is An Address In A Net?

Perform bit-wise AND between the mask and the address to test.

Compare the result to the net number.

Test: 204.198.127.58

1 1 0 0 1 1 0 0 1 1 0 0 0 1 1 0 0 1 1 1 1 1 1 1 0 0 1 1 1 0 1 0
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0
1 1 0 0 1 1 0 0 1 1 0 0 0 1 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0

Which is 204.198.64.0: Yep

And Another One!

Test: 204.199.72.30

1 1 0 0 1 1 0 0 1 1 0 0 0 1 1 1 0 1 0 0 0 1 0 0 0 0 0 1 1 1 0

1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0

1 1 0 0 1 1 0 0 1 1 0 0 0 1 1 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0

Which is 204.199.64.0, not 204.198.64.0: Nope

CIDR

Usually, masks are leading ones followed by trailing zeros.

Save time: Just give the number of 1's.

For

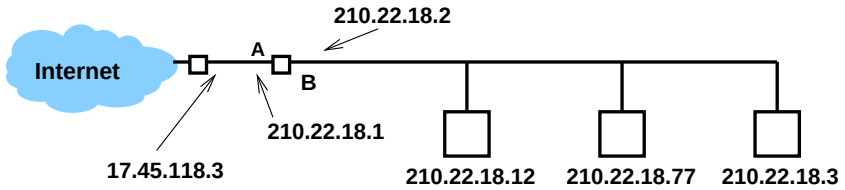
net 216.100.0.0 mask 255.255.0.0

say

216.100.0.0 / 16

Classless Inter-Domain Routing

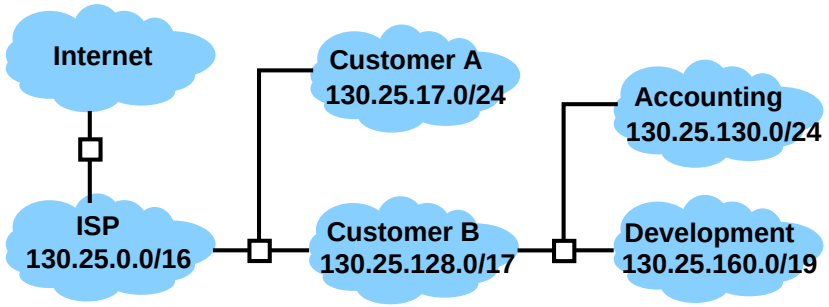
Routing at the Edge



The trivial router table.

210.22.18.0	255.255.255.0	Direct Through B
default		Send to 17.45.118.3 Through A

Masks for Subdivision



Masks allow subdivision of address blocks.

Some IP Addresses Are Special

Prefix	Suffix	Type Of Address	Purpose
all-0s	all-0s	this computer	used during bootstrap
network	all-0s	network	identifies a network
network	all-1s	directed broadcast	broadcast on specified net
all-1s	all-1s	limited broadcast	broadcast on local net
127	any	loopback	testing

How Many IP Numbers Do You Need, Anyway?

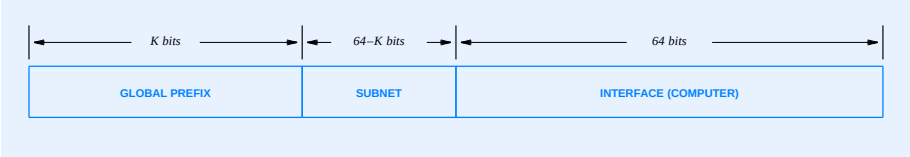
Some computers have more than one IP number.

Routers
High-availability
Performance

An IP address does not identify a host, but a network interface.

IPv6 Addresses

IP v. 6 addresses are 128 bits.



The Global Prefix names the owning organization.
Some values have special meaning.

The Sub-net names a part of that organization.

Interface chooses a specific interface (computer).

IPv6 Address Classes

Unicast
Delivered to a single destination.

Anycast
Delivered to any one of a group of interfaces.
No coding distinction from unicast.

Multicast
Delivered to each of a group of interfaces.
Addresses begin with a byte of ones: ff00::/8

Writing IPv6 Addresses

Write in hex, groups of 16 bits, separated by colon:
69DC:8864:FFFF:FFFF:0:1280:80C0A:FFFF

Often contain many zeros, and leading zeros may be omitted:
FF0C:0:0:0:0:0:0:B1

The longest run of zeros may be replaced by two colons:
FF0C::B1

Only one :: may be used.

IPv6 Link-Local Address

Address starting with FE80::/10 (in practice, FE80::/64) are link-local.

The remaining 64 bits are constructed from the 48-bit MAC address.
Insert FFFE in the middle
Invert the #7 bit (xor first byte with 0x02)

Link-local can only be used on the same network segment.

Virtual Packets

An internet must provide a standard service.

Packet formats vary with various hardware.

The internet provides a virtual packet format.
Implemented on all hardware.

All transmissions on the virtual net use the virtual packet.

IP Datagram

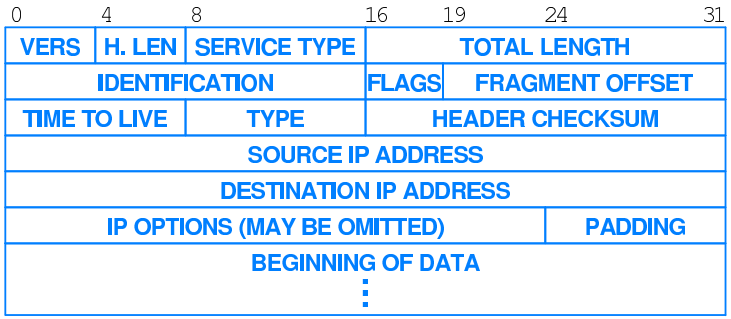
Virtual Packet

Datagram size is determined by the application.

1 to 64k payload bytes.

More flexible than most hardware.

Header Format



Header Fields

VERS	IP Version, which is 4
H. LEN	Header length, words
SERVICE TYPE	A priority indicator, often ignored.
TOTAL LENGTH	Size of whole datagram in bytes.

IDENTIFICATION, FLAGS, and FRAGMENT OFFSET
Next topic.

A packet is dropped after being routed TIME TO LIVE times.
Prevents a routing loop from keeping packets on the net forever.

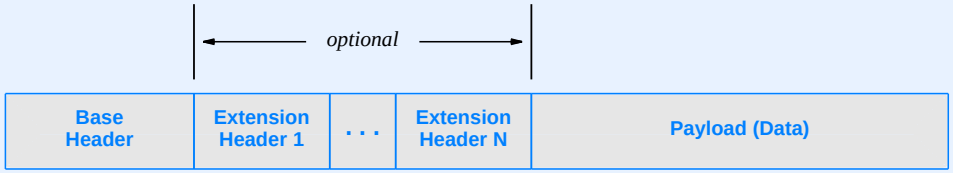
TYPE says what higher-level protocol is using this datagram.
For instance, TCP is code 6.

TYPE tells the receiver how to interpret the payload.

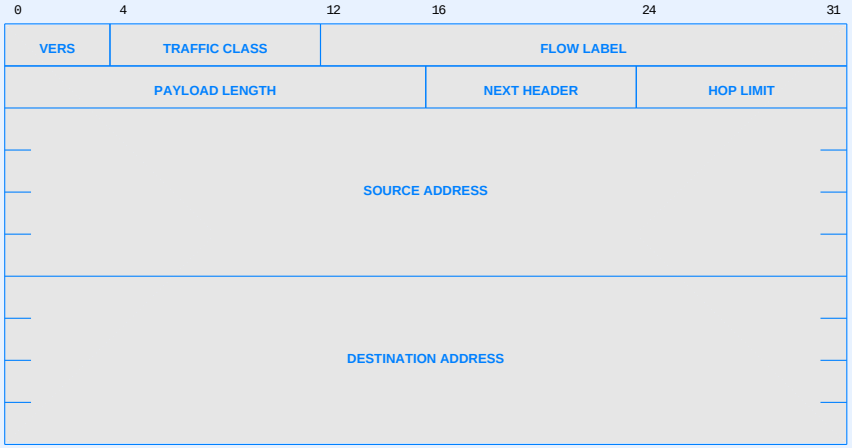
IPv6 Format

The IPv4 header is intended to be complete, with options optional.

The IPv6 is designed as a base header, followed by blocks as needed. It is expected that some will be needed.



IPv6 Base Header



IPv6 Base Header

VERS	IP Version, which is 6.
TRAFFIC CLASS	Same as IPv4 Service Type.
FLOW LABEL	Identifies packets in the same flow, such a media stream Routers may try to treat the stream consistently.
PAYLOAD LENGTH	Just that, in bytes.
NEXT HEADER	Type of what follows, another header or the payload type.
HOP LIMIT	Same as IPv4 time-to-live.
SOURCE ADDRESS	Where from.
DESTINATION ADDRESS	Where to.

Extension Headers

There are many, and more may be added.

Some are fixed size, others contain a length field.

RFC 2460 indicates the following are required:
Hop-by-Hop Options Routing Fragment
Destination Options Authentication Encapsulating Security Payload

Only the first of these needs to be read before the final destination.

Routing Tables

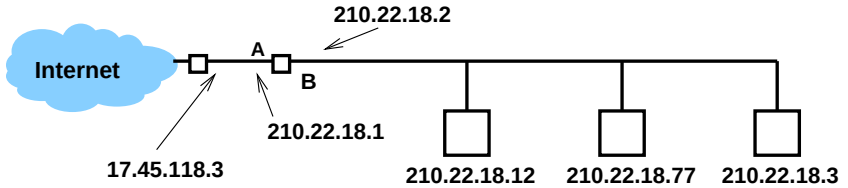
IP hosts use *routing tables* to decide where to send packets.

Each entry has a network number, a mask, and what to do with packets whose destinations match the entry.

There is usually a default entry.

The action is to deliver the packet directly to the recipient
or
Deliver to the next router on the packet's path.

Trivial Routing

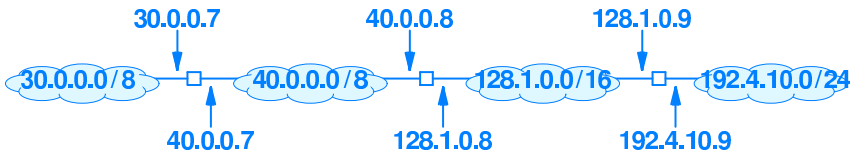


The trivial router table.

210.22.18.0	255.255.255.0	Direct Through B
default		Send to 17.45.118.3 Through A

The average desktop has a routing table like this.

Routing



(a)

Destination	Mask	Next Hop
30.0.0.0	255.0.0.0	40.0.0.7
40.0.0.0	255.0.0.0	deliver direct
128.1.0.0	255.255.0.0	deliver direct
192.4.10.0	255.255.255.0	128.1.0.9

(b)

Table for the middle router.

Matching The Addresses

Route table entries can overlap.
112.130.0.0 / 255.255.0.0
112.130.210.0 / 255.255.254.0

A destination may match multiple entries.

Simple routers match in order.

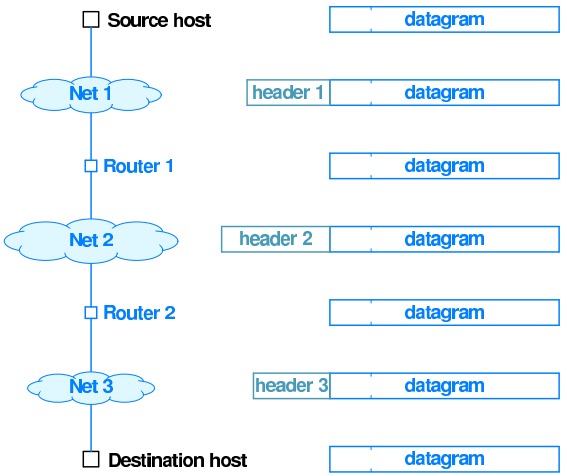
Complicated routers use the match with the longest prefix
Largest network part

Encapsulation

IP packets are sent as the data portion of
actual hardware packets.

Recall that one of the type fields in the Ethernet header is IP.
That means the segment is carrying an IP datagram.

Transmission



Re-encapsulation

Packets will be re-encapsulated for each transmission.
Each time the packet is routed.

Each router removes the datagram from the frame that brought it.

The datagram is then sent out again in a new frame.

Best-Effort Delivery

The network delivers packets to their destinations with *best effort.*

This permits:
Duplication Delay Out-of-Order Delivery
Corruption Loss

These are properties of most network hardware.

As we'll see later, TCP provides reliable communications.

Limits

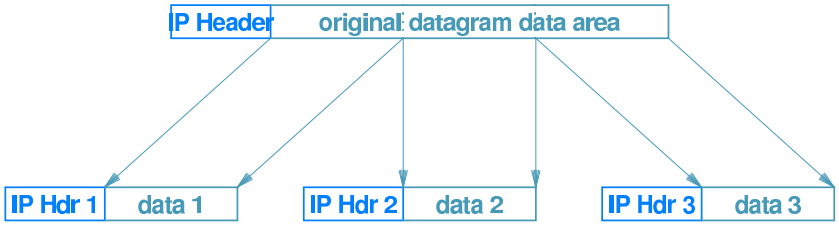
Each hardware limits its packet size.
Maximum Transmission Unit: MTU

IP packets which are too large must be broken up.

For instance, Ethernet frames contain 46-1500 bytes.

An IP datagram can hold up to 64K bytes.
Doesn't fit too well in a single Ethernet frame.

Fragmentation



Fragment headers are copies of the original with a few changes.

Who's Your Datagram?

Fragments have the fragment flag set.
All fragments but the last, actually.

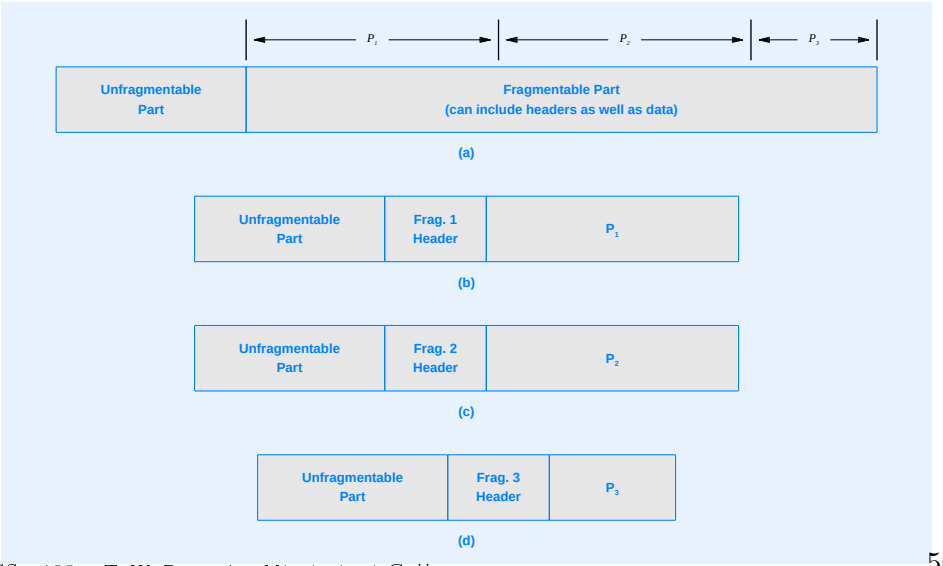
The fragment offset field tells where in the original datagram this fragment goes.
Units of eight bytes from the start of the original.

Each datagram is given a unique identification number when sent.

Its fragments retain their original identification.

So all the fragments of a datagram have the same identifier.

IPv6 Fragmentation



IPv6 Fragmentation

The **Unfragmentable Part** includes the base header, any header which must be read by routers.

The Unfragmentable Part is duplicated to each fragment.

The presence of a fragment header indicates a fragment. This header contains the same information as the IPv4 fragment fields.

Reassembly

Final destination reassembles.

Fragment offset tells where to put each piece.

When the first fragment arrives, a timer is started.

If all the fragments arrive within the time limit, the datagram is reconstructed.

If not, all fragments (and the datagram) are discarded.

More Fragments

In IPv4, a datagram is fragmented by any router when it is too large to send on the outbound network.

The arriving datagram may already be a fragment, so fragments may be further fragmented.

There is no difference between a fragment of an original and the fragment of a fragment.

In IPv6, the sender must create fragments small enough to make the whole trip. Routers do not fragment.

No Fragments, Please

IPv4 has a do-not-fragment flag.

If set, and fragmentation is required, the packet is dropped.

An error message will generally be sent.
ICMP messages — later topic.

This can be used to find the path MTU.
The smallest MTU on the path.

Fragments To Be Avoided

Current practice is for sender to limit its packet size so fragmentation is not needed:
Send only packets below the MTU.

Algorithms can discover the path MTU.
Observe the fate of various-sized no-fragment packets.
Binary search the size.

Too small a value is inefficient.

Values in the range 1000-2000 seem usual.
Ethernet is 1500

MTU in IPv6

Version 6 makes finding the MTU more standard.

Since the sender must make small enough fragments to travel the whole path, it must know the path MTU.

Finding The Receiver

Internet messages are sent to an IP address.

IP addresses are virtual.
Hardware won't help much.

IP addresses must be mapped to hardware addresses.
Address Resolution.

Address Resolution Protocol

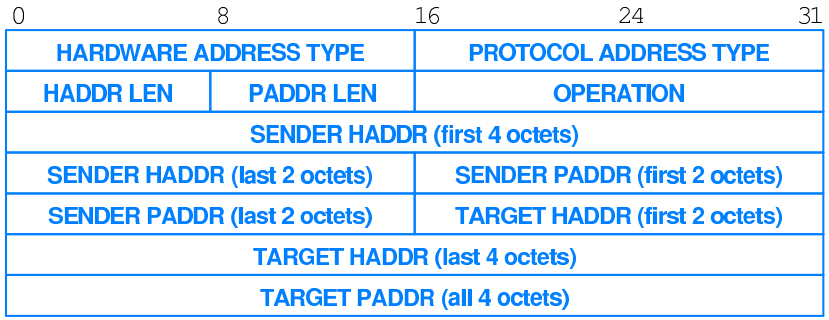
Typically used on a LAN.

Host broadcasts a query: Who has IP number x?

The request message contains the hardware address of the requester.

If some host has IP number x, it responds.

Message Format



Send as a packet on the LAN

ARP Message Format

Not limited to IP over Ethernet.

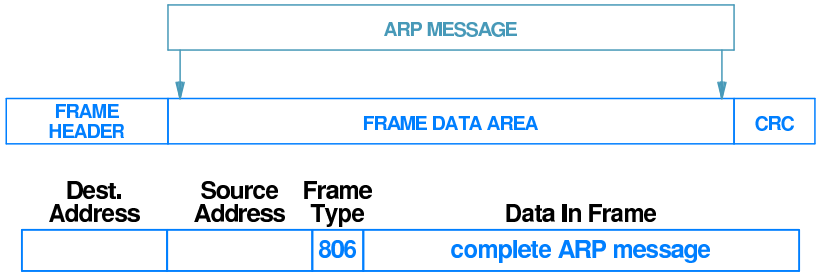
Codes and sizes for both hardware and protocol address type.

Hardware and protocol address, both of sender and recipient.
Four total

Both requests and responses use the same format.
Determined by operation code.

Unknown parts of a request are usually just filled with zeros.

ARP Over Ethernet



ARP segments are not IP packets.

Caching

Hosts generally cache IP addresses which they have requested.

Hosts usually cache the IP of the sender when responding to a request.

IPv6 Address Translation

IPv6 doesn't use ARP, even though ARP is flexible enough.

Uses IPv6 Neighbor discovery.

Similar, but uses a multicast address on which all IPv6 hosts must listen.
IPv6 doesn't have a broadcast address.

Switching Over

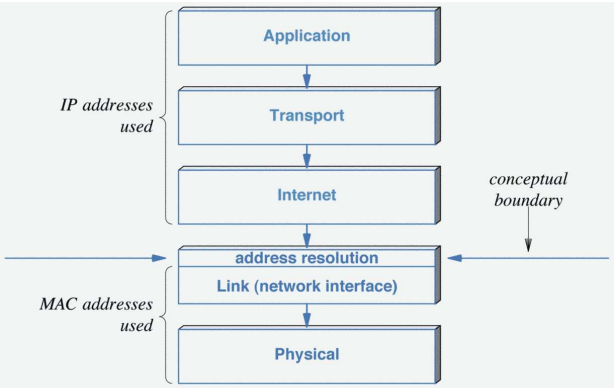


Figure 23.5 Illustration of the boundary between the use of IP addresses and MAC addresses.

Switching Over

The ARP layer is placed just above the hardware layer.

Higher levels use only IP addresses.

ARP And Routing

An IP sender needs to know the hardware address of any host to which it will transmit a packet.

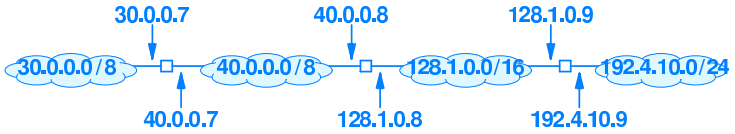
It may not need to know the hardware address of the packet's ultimate destination.

ARP use depends on the routing table.

If the action is "deliver to recipient," ARP the destination.

If the action is "forward to router," ARP the router.

For Instance



(a)

Destination	Mask	Next Hop
30.0.0.0	255.0.0.0	40.0.0.7
40.0.0.0	255.0.0.0	deliver direct
128.1.0.0	255.255.0.0	deliver direct
192.4.10.0	255.255.255.0	128.1.0.9

(b)

The middle router ARPs hosts in 40.0.0.0/8 and 128.1.0.0/16, including the near sides of the other routers.

It will not ARP any other hosts.

Control Messages

Report errors.

Pass control information.

Request changes in behavior.

Internet Control Message Protocol
ICMP

Things That Can Go Wrong

Packets must be dropped.

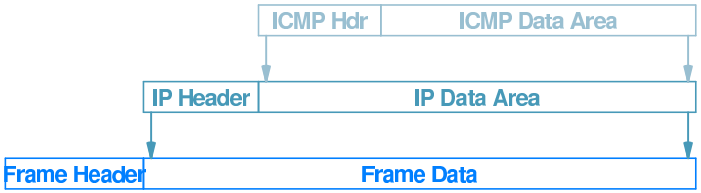
Routers get congested.

There is no route to that subnet.
Are you sure it exists?

There is no host at that address.

Etc.

ICMP Encoded In IP



ICMP messages are IP messages.

TYPE field in IP header set to 1 for ICMP.

Type of 0800 for IP set in the Frame Header.

ICMP Format

Varies with message type.

First byte is the type code
Second byte is a sub-code.

Next two bytes is a check sum.

Types sent in response to a regular datagram include the original IP header and first 64 data bits.

Some ICMP Types

3: Destination Unreachable. A node cannot get the packet to the destination.

Several sub-codes: 0: Net Unreachable.
1: Host Unreachable.
2: Protocol Unreachable.
3: Port Unreachable.
4: Fragmentation Needed but the Don't Fragment bit is Set.
5: Source Route Failed.

Some ICMP Types

5: Redirect.

Update your routing table.

Code tells you if its a host or a network that changed.

12: Parameter problem.

The data includes a "pointer," an offset which tells where the error was detected.

30: Response to traceroute request.

Some ICMP Types

11: Time exceeded.

Either the time-to-live was reduced to zero, or fragments were not all collected during the time limit.

Second byte tells which type of event.

0/8: Echo request and reply.

These are what the ping command uses.

Not Forever

A host may respond to a datagram with an ICMP error packet.

It may not produce an ICMP error packet in response to an ICMP error packet.

Traceroute

Send out ICMP echo requests with increasing TTL.

Each router along the way sends an ICMP Time Exceeded.
Traceroute gets the router address from the ICMP address.

Traceroute may use the recently-added traceroute request option.
Requests the router may respond with ICMP 30.
Still forwards the packet.

This allows traceroute to send just one transmission instead of one per router.

Dynamic Host Configuration Protocol (DHCP)

Each computer needs to know its IP address.

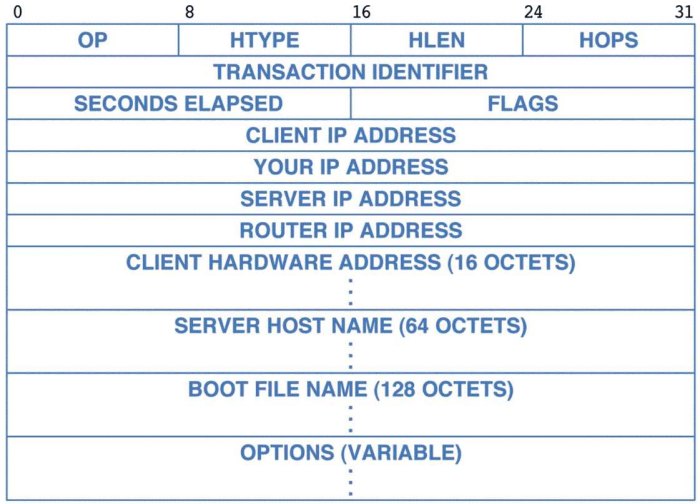
It's not much fun for the network admin to do this by hand for a large collection of workstations.

Need a way for a computer to ask a server for its IP when it boots: DHCP

Can also provide other useful info such as the default router and name server.

DHCP messages are IP messages.
Send to the limited broadcast address (255.255.255.255) when destination is not known.

Message Format



Message Format Fields

The OP is says if the packet is a request or a response.

The "DHCP message type" option specifies the exact operation.

HTYPE and HLEN: type and length of the client Hardware (MAC) address.

HOPS counts DHCP relay forwardings.
Relays will refuse if the count is too large.

Message Format Fields, Cont

Transaction Identifier is a random number used to associate requests with responses.

The server and boot file names are for remote booting.
If provided, the client boots the indicated file on the indicated server.

IP Address Fields

Client IP
The client's current IP address, if known, or zero.

Your IP
Address being provided by a server to a client. Set to zero in other contexts.

(Next) Server IP
Client should try here next. Used to distribute boot services over multiple servers.

Router IP
Actually a relay IP, filled in by a DHCP relay (see below).

Options

The server assigns many important options using the options section.

Host name.

DNS server names.

Default gateway and other routing information.

Operation

Booting machine broadcasts a DHCPDISCOVER request to 255.255.255.255:67.

Server responds with a DHCPOFFER to 255.255.255.255:68. Contains a "Your IP" address, and other parameters in the options section.

Client accepts by broadcasting DHCPREQUEST echoing the assignment.

Server responds with DHCPACK (okay) or DHCPNAK (no).

Operation, Cont.

A client wishing to reuse a previous address starts with the DHCPREQUEST.

On shutdown, the host sends DHCPRELEASE to surrender the address.

Address Allocation

Server may assign an address based on the requester's MAC address.
Typical in an office.

Server may assign an address randomly from a pool.
Typical at a public WiFi site.

A client may ask to use the address it had last time.
The server will allow or refuse.

Address Allocation (Cont)

Addresses are *leased*: Assigned for a limited period.

When the period is about to expire, the client asks to renew.
The request is usually granted.

The lease period allows the admin to reassign addresses without the old assignments enduring forever.

DHCP Relay

The DHCPDISCOVER cannot pass to a different subnet.

A DHCP relay agent can receive it and send to the server.

Forwards the response back.

IPv6 Auto-configuration

Multicast to discover the prefix used on the local network.

Set the suffix as with the link-local address.

Text mentions NAT in Chapter 23.

Sources

Comer, *Computer Networks and Internets*
(Our beloved textbook.)

Forouzan, *TCP/IP Protocol Suite*, McGraw-Hill, 2003.

RFC 1191

RFC 2131

http://www.netheaven.com/pmtu.html

Kevin R Fall and Richard Stevens, *TCP/IP Illustrated, Volume 1: The Protocols*, 2nd Ed, Addison-Wesley.

http://packetlife.net/blog/2008/aug/4/eui-64-ipv6/

Sources

http://www.tcpipguide.com/free/t_IPv6Datagram-MainHeaderFormat.htm

http://www.tutorialspoint.com/ipv6/ipv6_headers.htm