# Internet Protocol

*This lecture describes the the most common network protocol, Internet Protocol (IP) version 4 (IPv4).*
*After this lecture you should be able to: differentiate between the Internet and IP; look up IP standards; interpret the values of the most common IP header fields; compute an IP checksum; determine the netmask for an IP network; determine if an IP address is in a particular network; determine if an IP address is public, private or link-local; decide which port a frame would be forwarded on based on the contents of a routing table; determine the effect on an ARP cache of receiving an 802.3-encapsulated IP frame; determine the IP source/destination addresses used on the public/private sides of a NAT router; list the recursive DNS queries used to resolve a domain name.*

## Introduction

IP was developed in the late 70's when different computer manufacturers, academic institutions and research groups were using incompatible data communication networks and protocols. IP was designed as a common protocol to link these networks together so they could exchange files, e-mail, terminal sessions, etc. It was thus an inter-network protocol or an "internet protocol".

As the usefulness of a universal networking protocol became clear, new system started using IP as their *native* networking protocol. Widespread adoption of IP has resulted in almost all networks using IP as their network-layer protocol.

The availability of IP, a widely-supported and freely-available protocol, facilitated the growth of a non-proprietary commercial data network using IP that is commonly called "The Internet".

**Exercise 1**: What is the difference between IP and "The Internet"? Does a network using IP have to be on the Internet? Does someone using the Internet have to use IP?

IP is defined in documents called Requests For Comment (RFCs) published by the Internet Engineering Task Force (IETF). IETF standards development is open to the public and adoption of proposals depends on technical merit.

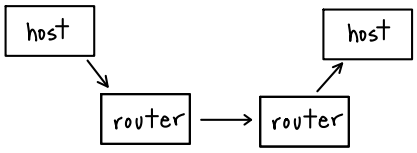
## IPv4

The current version of IP is known as IP version 4 (IPv4) and is defined in RFC 791, published in 1981.

IP Version 6 (IPv6) was originally designed to address the problem of address space exhaustion but introduced additional, unrelated changes. Simpler methods were developed to work around the address exhaustion problem and it remains to be seen if IPv6 will ever be widely used.

Not surprisingly, documentation for IP protocols is widely available on the Internet, for example, from www.ietf.org.

IP is a packet-switching protocol. Data is encapsulated in IP packets which are transferred in store-and-forward manner between routers from a source host to a destination host:




## IP Packet Format

Services provided by the IP layer are limited to routing and fragmentation. IP has "... no mechanisms to augment data reliability, flow control, sequencing, ..."[1].

An IP header consists of at least five 32-bit words (20 bytes). The following diagram is taken from RFC791[2] and shows the IP header:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|Version|  IHL  |Type of Service|          Total Length         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|         Identification        |Flags|      Fragment Offset    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Time to Live |    Protocol    |         Header Checksum       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       Source Address                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                    Destination Address                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                    Options                    |    Padding    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

                Example Internet Datagram Header
```

Figure 4.

[1] John Postel, Ed., RFC 791.
[2] RFCs were published in text format.

The most important fields are:

**Version** Protocol version number (4)

**IHL** IP Header Length; number of 32-bit words

**Type of Service** Priority. Not widely used.

**Total Length** Length of the IP packet in bytes.

**Identification/Flags/Fragment Offset** for fragmentation. Rarely used.

**Time to Live** A value that is decremented each time a packet is forwarded. Prevents packets traversing routing loops indefinitely.

**Protocol** the type of protocol embedded in the IP packet. 1 for ICMP, 6 for TCP, 17 for UDP. Assigned by IANA[3].

**Header Checksum** a one's-complement checksum for the header (see below)

**Source/Destination Address** the 32-bit source and destination IP addresses (see below).

**Options** Optional header components that are not normally used (security, source routing, route recording and timestamps).

**Exercise 2**: What is the value of the first byte of an IP packet that uses the shortest possible header? If first byte is 0x46, what is the length of the Options field in bytes?

### IPv4 Checksums

IPv4 checksums (used for IP, UDP, TCP, ...) are the bitwise complement of the one's-complement sum of the 16-bit values to be protected. The fields included in each checksum are defined in each specification.

A simple algorithm is to do 32-bit unsigned addition of the 16-bit fields and then add the overflow (in the MS 16 bits of the 32-bit sum) to the LS 16 bits. The checksum is the bitwise complement of this value. The receiver checks for errors by repeating the checksum calculation and verifying that the result is zero.

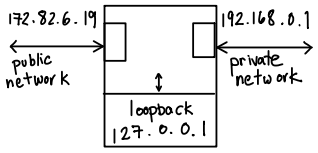
**Exercise 3**: A protocol header contains four 16-bit fields with decimal values 65535, 1, 2, and 3 that are to included in an IPv4 checksum. What is the value of the header checksum?

## IPv4 Addresses

### Host Addresses

Each IP (IPv4) network interface has a 32-bit (4 byte) address. Most hosts, and all routers, have more than one interface and thus more than one IP address.

For example, a typical home router will have two IP interfaces, one public and one private, as well as a software-based loopback interface:




IP addresses are usually written as a "dotted quad" of the decimal value of each byte separated by periods. The bytes are written in big-endian order. For example, 0xc0a80001 would be written 192.168.0.1.

Most devices with IP protocol stacks have a virtual network interface at address 127.0.0.1 (hostname `localhost`) which is used for communication between processes on the same device.

### Network Addressses

IP addresses composed of two parts. The most significant bits of the address, the network address, identify one of a few hundred thousand (up to perhaps $10^6$) IP networks in the world. The remaining bits identify a host within that network.

Originally networks were divided up into three classes: A, B and C. Class A networks could have up to $2^{24}$ host addresses. Class B addresses up to $2^{16}$ and Class C up to $2^8$.

This led to inefficient allocation of network addresses and today network address are "classless" and are composed of two parts: the value of the network prefix (e.g. 142.232.0.0) and the length of the network portion of the address in bits preceded by a slash (e.g. /16). The two values together are the (classless) network address. For example, the BCITNET2 network has an address of 142.232.0.0/16.

A netmask is a 32-bit value with 1's in the bits corresponding to the network address.

**Exercise 4**: What is the netmask in binary for a /24 network? What is it in decimal? How can the netmask be used to determine if one IP address is on the same network as another? Is the address 192.168.2.200 in the 192.168.2.0/25 network?

Network addresses are assigned by a non-profit organization called ICANN. The host whois.arin.net

[3] Internet Assigned Numbers Authority.

can be used to query for ownership of North American network addresses.

**Exercise 5**: Who "owns" the 24.80.0.0/13 network?

## IP Routing

IP networks operate in store-and-forward fashion. Routing is the process of getting a packet from source to destination.

Instead of Ethernet bridges connected by a spanning tree, IP networks use routers connected in a mesh which can, and often does, contain redundant links and loops.

Each IP packet includes a "time to live" field to protect against packets circulating in the network indefinitely as a result of misconfigured routers.

**Exercise 6**: Does the header checksum change each time a packet is forwarded? Why?

Each packet is routed independently. Thus each packet's header has to include the destination address.

Each device that forwards IP packets typically has multiple ports and is called a "router" because it decide on which port(s) to forward the packet. This decision is done by looking up the destination IP address in a "routing" table that defines the outgoing port for a network.

Each routing table entry can have a cost or "metric" associated with it. The routing algorithm selects the lowest cost route (port). Routing costs are local to each router. The metrics can be determined in many ways. For a simple host they may be based on the port data rate. Some routers exchange information with other routers to determine the best route to each network. The metrics can also be modified dynamically by factors such as delay. Routes can also be manually configured.

A routing table will usually have a default route containing the address of a "gateway" router. Packets for which there is no route are send with a destination L2 address of the gateway. The gateway receives these packets forwards them based on its own routing table.

Here is the routing table for a simple home wireless router:

| Destination | Gateway | Subnet Mask | Metric | Interface |
|---|---|---|---|---|
| 192.168.1.0 | * | 255.255.255.0 | 0 | br0 (LAN) |
| 204.191.0.0 | * | 255.255.0.0 | 0 | vlan1 (WAN) |
| 127.0.0.0 | * | 255.0.0.0 | 0 | lo |
| default | 204.191.1.1 | 0.0.0.0 | 0 | vlan1 (WAN) |

**Exercise 7**: For the routing table above, what port ("Interface") would be used by packets with the following destination IP addresses: 127.0.0.255? 192.168.1.1? 192.168.2.1? 204.191.10.32?

### Classes of Routers

Many routers like the one above contain only a few routes, often just for the directly connected network(s) and a default route pointing to an "upstream" router.

However, routers that connect multiple networks need to determine which port should be used to reach different networks. This can be done manually or using distributed algorithms that try to determine the best route from each router to each network. The most common of these algorithms is called OSPF (Open Shortest Path First). Most "enterprise" routers will configure their routes using OSPF.

Routers that connect different service providers are called "border" routers. Their routing decisions are more complex because they need to take into account "peering" agreements between service providers that determine which packets can be forwarded to which service providers. The protocol typically used to set up routing between border routers is BGP (Border Gateway Protocol). Border routers are typically found only at a one (or a few) data centers in each city called IXP (Inter-Exchange Points) where ISP interconnect their networks.

## ICMP

Internet Control Message Protocol (ICMP) is a simple protocol used to diagnose and report problems at IP routers. Typical ICMP packets include the echo request ("ping") packet that can be sent to another IP address to request a response to check connectivity and delay. Other ICMP packets carry diagnostic messages to the source address when router is unable to forward a packet towards its destination (e.g. "Destination Unreachable").

## ARP

The Address Resolution Protocol (ARP) allows hosts to discover the LAN address of another host on the same network segment. This is done by sending a broadcast ARP request with the desired IP address. A device that sees an ARP request that matches

its own IP address replies. Each host maintains an "ARP cache" of these responses. These entries are checked for consistency with incoming packets and are "aged" (removed when they are too old).

**Exercise 8**: What pairs of values are stored in an ARP cache? What addresses from a received packet need to be examined to validate an ARP cache entry?

## DHCP

The Dynamic Host Configuration Protocol (DHCP) is used by a host to configure its IP network stack including the IP address, DNS servers and gateway. When a device boots up it broadcasts a DHCP request. A DHCP server in the same network segment then responds to the request with a DHCP response packet containing the network configuration information for that particular host.

DHCP is an IP protocol so the broadcast destination is the IP broadcast address (255.255.255.255).

**Exercise 9**: When a host boots up, what must it send out first, an ARP request or a DHCP request?

Small ad-hoc networks may not have DHCP servers. In this case hosts may select a "link-local" address at random from the 169.254.0.0/16 network and use ARP to confirm that the address is not already in use.

### Private Addresses

In many cases networks and their hosts do not need to be reachable from the public internet. Certain network addresses are reserved for these networks. For example, the networks 10.0.0.0/8 and 192.168.0.0/16 are private and cannot be routed over the Internet.

## NAT

However, hosts in private networks often want to be able to reach other hosts on the internet (they want to "call out"). This is accomplished through a fairly complex network address translation (NAT) process at a router that has a public and a private interface.

A NAT router translates the destination address of packets coming in and the source address of packets going out of the private network. The requires the NAT router to have some way of determining which public-private IP address pairs which are "connected" even though although IP is normally considered a connectionless protocol. This requires that the router examine the contents of packets to try to guess the state of these connections. This requires looking at the details of the UDP and TCP layers to detect when connections are set up and the corresponding port numbers.

**Exercise 10**: A host with a (private) address 192.168.1.10 is behind a NAT router with an (public) address of 172.12.192.15. The host sends a packet to a host at address 74.125.225.113 requesting a web page. Show the source/destination address pairs of the request and response packets on the private and public sides of the router.

## Domain Name System

The Domain Name System (DNS) is a distributed database whose main purpose is to convert human-readable domain names into IP addresses. The database is arranged as a hierarchy that mirrors the hierarchical structure of domain names. For example, the DNS (or "name") server for the `.ca` domain holds the address of the DNS server for the `bcit.ca` DNS server, not the addresses of the hosts in the `bcit.ca` domain. The DNS server for `bcit.ca` will contain the IP address of the host `learn.bcit.ca` and (possibly) the address of the DNS server for the domain `learn.bcit.ca`.

When a host needs to look up the IP address corresponding to a domain name it sends a DNS query to its configured DNS server which the performs a recursive DNS lookup by travelling down the domain name hierarchy one level at a time. At each level it looks up the address of the responsible DNS server and queries it for either the IP address of the host or IP address of the DNS server for the next level down in the domain name hierarchy.

Caching of results greatly reduces DNS overhead.

The DNS servers for top-level domains (e.g. .com) are run by companies that have been given contracts to run these root DNS servers in exchange for collecting fees (about $10 per domain name per year) from entities registering domain names.

**Exercise 11**: Can a host's DNS server be configured using a host name? Why or why not? Assuming a host has an empty DNS cache, what queries would it generate to look up the IP address of the host mx.bcit.ca?