

## Internet Protocol Version 4 (IPv4)

### Primary Function

IPv4 facilitates global routing of data for higher-level protocols by tagging each message with "unique" addresses associated with the sender and designated recipient.

### Specification

RFC 791 (September 1981) provides the official specification for IPv4. Several updates, e.g., RFC 6864 are also applicable.

### PDU Structure

PDUs for IPv4 are more commonly referred to as packets. A legal IPv4 packet contains a header of 20 or more bytes (terminating on a 4-byte boundary). User data received from upper layer protocols follows immediately after the IP header.

0		7		8		15		16		23		24		31	
Version		IHL		Type of Service				Total Length							
Identification								Flags		Fragment Offset					
Time to Live				Protocol				Header Checksum							
Source Address															
Destination Address															
Options												Padding			

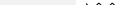
In conjunction with these addresses (described in more detail in the following section), IPv4 headers contain information that intermediate nodes and the final recipient will need to process the IPv4 header information and remaining data. In IPv4, this information includes the length of the packet and a pointer to the end of the header (which varies in length). Likewise, the IPv4 header contains a standard identifier that indicates the next layer of encapsulation associated with the message data.

Field Name	Function	Constraints
Version	Differentiates between protocol versions	4 (always)

Internet Header Length (IHL)	Enables the parser to identify the end of the IP header (and beginning of packet data)	Number of 4 byte words
Total Length	Enables the parser to identify the end of the packet	Number of bytes
Time to Live (TTL)	Ensures that all packets eventually time out	Number of hops
Protocol	Cues the receiving stack regarding which protocol to call next	Assigned Internet Protocol Numbers (managed by IANA)

### Network Identity

IPv4 network address are 32 bits in length. To facilitate routing between interconnected networks, each address is divided into a varying length network portion and a host portion. When writing addresses, each octet is notated in order from most to least significant with period separators. By convention, individual octets are written in decimal, e.g., 192.168.33.25. This basic structure is depicted here.

0	7	8	15	16	23	24	31
A		B		C		D	
n bits <i>Network Portion</i>			 (32 - n) bits <i>Host Portion</i>				

On a network using *n bits* of the address to identify the network, that value will be encoded in the leftmost bits. The rightmost  $32 - n$  bits are then used to identify the host relative to the network id. The network bits of an address distinguish Layer 1/2 network segments. Hosts with matching network bits *should* be able to reach each other using lower level services. When the network bits of two hosts differ, the hosts must rely on Layer-3 services to talk to one another.

### Packet Fragmentation

At times, the IPv4 stack encounters size restrictions associated with the next hop in a route and must break the current packet into smaller units in order for it to reach its destination. Using fragmentation, the network layer can split a packet into smaller units, each containing a multiple of 8 bytes (64 bits). Fragmentation happens at the source when the Transport Layer constructs

datagrams or segments that exceed the Maximum Transmission Unit (MTU) of the underlying network. Fragmentation may also occur at intermediate nodes, when a route traverses a link that with a small MTU. Bytes 4 - 7 of the IPv4 header enable this functionality.

Field Name	Function	Constraints
Identification	Identifies fragments associated with a single packet	RFC 6864
More Fragments Flag (MF)	Used to determine the last fragment associated with a packet	0 or 1
Do Not Fragment Flag (DF)	Prevents intermediate nodes from fragmenting a packet (network events may be triggered)	0 or 1
Fragment Offset	Provides an index into a reassembly buffer where the data from this fragment should be inserted	Number of 8-byte units

## Enrollment

Each host must be configured with an IP address and told how many bits are used to determine the network portion of an address. We can address hosts manually by assigning static values to each node or automate the process by enabling each host to configure itself in cooperation with other nodes on the same network segment.

## Managed Configuration

The *Dynamic Host Configuration Protocol (DHCP)* provides central control over this process at the cost of additional network infrastructure. *DHCP clients* (hosts) send broadcast messages on the local network to request initial configuration. *DHCP servers* within broadcast range respond to these requests, offering addresses out of a pre-defined pool along with other parameters defined in the configuration of the server.

## Mapping Layer 2 Addresses

A properly configured host further relies on the ability to connect destination IP addresses to hardware addresses in the local network segment. The Address Resolution Protocol (ARP) specified in RFC 826 provides this capability to IPv4 networks running over 802.3 (Ethernet) and close relatives like 802.11.

## Auto-Configuration

Absent a static address configuration or response from a DHCP server, hosts are permitted to configure themselves with addresses between 169.254.0.0 - 169.254.255.255, i.e., the *link-local address range*. Dynamic link-local address assignment for IPv4 is specified in RFC 3927. A host relying on RFC 3927 will select an address from the link-local range at random and proceed to check for address collisions by issuing a modified ARP request and waiting for a response before claiming the IPv4 address as its own. This mechanism is sufficient to provide network capabilities to hosts on simple networks without requiring any manual configuration or provisioning a DHCP server.

## IP variations

Basic IPv4 networks are relatively straightforward. Most data flows are unicast (point-to-point) in nature, traversing a path between nodes that maintain a stable presence on the network. As communications have evolved, IETF has accepted extensions to the basic network layer protocol to support additional use cases.

Multicast IP is defined initially in RFC 1112 and updated by later specifications. Unlike unicast IP, multicast traffic is sent from one host and routed to a dynamic group consisting of zero or more members. Multicast extensions to IPv4 enable streaming media applications and support numerous use cases that once relied on less precise broadcast mechanisms.

Likewise, hosts in basic networks do not roam between subnets and Layer 2 network segments. Mobile IPv4 is defined in RFC 5944 to accommodate the growing number of mobile devices that are apt to roam between subnets while trying to maintain active conversations.

## Special Addresses

Beyond the basic structure specified above, certain addresses and address ranges are reserved for special use. We have already encountered reserved address ranges in previous sections. Addresses between 169.254.0.0 and 169.254.255.255 are link local addresses that cannot be forwarded outside the local segment. Addresses between 224.0.0.0 and 240.255.255.255 are allocated as group addresses for multicast applications.

Addresses between 127.0.0.0 and 127.255.255.255 are reserved for loopback purposes, though 127.0.0.1 is typically thought of as *the* loopback address. Loopback addresses are assigned internally to hosts by the operating system and limited to intra-host communication. The packets sent via loopback mechanisms are isolated from external networks, thus ideal for restricting applications to local use when they otherwise rely on network services for interaction.

Within each possible network, one IPv4 address is reserved for broadcast communication. By setting the host portion of a destination IPv4 address to all ones, the address is interpreted as a broadcast target on the designated network, i.e., if broadcast traffic is supported, the message will be sent to all hosts on the segment.

## Additional Protocols

While the core packet delivery services are directly integrated within IPv4, the Internet Protocol Suite provides additional control over the Network Layer by way of the Internet Control Message Protocol (ICMP), which specified in RFC 792. The tools incorporated into ICMP provide crucial feedback to network devices, enabling these devices to remediate configuration issues and otherwise optimize network performance. Other uses of ICMP include network diagnostics and Mobile IP.

## Advanced IP Addressing

As mentioned previously, the network identity of a host is given by a variable number of bits ( $\leq 32$ ) which are counted from the most significant position. All remaining bits are allocated to the host identity.

For an n-bit network, a network mask is constructed by setting the leftmost bits of a 32-bit binary word to 1 while the remaining bits are 0, e.g., the subnet mask for  $n = 22$ :

```
11111111.11111111.11111100.00000000
255.255.252.0
```

From an address and a subnet mask, we can compute a network address by way of the bitwise AND:

```
192.168.32.0 = 192.168.33.25 & 255.255.252.0
```

Hosts and routers use each of these values to make decisions about how to forward packets destined for a different location. When configuring a router or host, all three values are given:

192.168.33.25	Host Address
192.168.32.0	Network Address
255.255.252.0	Subnet Mask

Alternative, we might combine the network address and subnet mask in a convenient shorthand known as CIDR notation:

192.168.33.25	Host Address
192.168.32.0/24	Network (CIDR)

We'll define CIDR later, but let's first make some observations about this configuration. We've defined one host in a network of  $1024 = 2^{10}$  potential hosts. As such, we determine that the 192.168.32.0/22 network includes IP addresses in the range from 192.168.32.0 .. 192.168.35.255. While the first and last address in this range are reserved, a host on the network is permitted to use any other value in the range. The reservations support the network address at 192.168.32.0 and a broadcast address at 192.168.35.255.

The 22-bit subnet mask can also be used to identify other networks. Let's explore the other networks that can be constructed in the 192.168/16 address range. In other words, let's look at network addresses that align with a 22-bit mask given a fixed value for first two octets. This constraint is common in the networking world. An organization will be given a network range large enough to be sub-divided into smaller networks by adjusting the subnet mask length accordingly. As you can see below, the networks follow share the first two octets followed by a multiple of four from 0 .. 255 in the third octet.

```
192.168.0.0 /22
192.168.4.0 /22
192.168.8.0 /22
...
192.168.32.0 /22
...
192.168.252.0 /22
```

This example is rather challenging given that the base-two network mask does not align with the octet boundaries and base-10 octet values. Until we acquaint ourselves with the valid networks and host ranges, it may not be immediately obvious that two hosts, e.g., 192.168.33.25 and 192.168.35.14, are members of the same network (thus reachable at Layer-2) while 192.168.36.40 is on the 192.168.36.0/22 network.

In common practice, unless network constraints dictate otherwise, we'll be more comfortable working with octet-aligned network masks, most often /16 and /24. In a /24 network, the network is defined by three octets while the host is defined by the last octet. Given the 172.20/16 range, we create networks 172.20.5.0/24 and 172.20.6.0/24. Each network can support up to  $254 = 2^8 - 2$  host addresses, i.e., 172.20.5.1 - 172.20.5.254 and 172.20.6.1 - 172.20.6.254.

### [Address Depletion](#)

The preceding discussion is driven in large part by the depletion of IPv4 addresses on the public Internet. Though IPv6 alleviates all concerns about address depletion, its adoption has proceeded slowly over more than a decade. In addition to the promise of IPv6, Internet governance bodies have adopted various and additional protocol complexity to prolong the life of IPv4.

### [Classless Inter-Domain Routing \(CIDR\)](#)

In the early days of the Internet, the IP address pool was divided up into five groups for management purposes. Three of these groups, identified as classes A - C, provide the framework for allocating address pools to individual entities. Classes D and E were reserved for Multicast IP and Experimental Use respectively.

Class	First Network ID	Last Network ID	Allocation
A	1.0.0.0	126.0.0.0	/8
B	128.0.0.0	191.255.0.0	/16
C	192.0.0.0	223.255.255.0	/24
D	224.0.0.0	-	/4
E	240.0.0.0	-	/4

In those days, routing between domains was simplified by the fact that allocation boundaries were known in advance. Any and every network allocated from Class A was a /8 range. Likewise, Class B allocations were /16 and Class C allocations were /24. The trouble with this approach, however, was that it resulted in a huge waste of usable addresses. Nearly half of the available address groups were /8 although few organizations would need this large a network.

The decision was eventually made to move away from the class-based routing model in favor of classless routing. Rather than specifying network masks as an implicit function of address class, allocations could be sized according to a variable-length mask. The underlying strategy described here is known as CIDR and is defined in RFC 4632.

CIDR is closely related to the concept of variable-length subnet masking (VLSM). Whereas CIDR specifies the broad approach to classless routing, VLSM gives us a tool to further divide network allocations into subnets that are suited to the application. VLSM was demonstrated in the Advanced IP Addressing section.

#### Private Addresses and Address Translation

RFC 1918 defines three address ranges within the Unicast address space that are restricted to private use, i.e., they can be routed on a LAN but not on the public Internet:

```
10.0.0.0      /24
172.16.0.0   /20
192.168.0.0  /16
```

Hosts deployed with RFC 1918 addresses can still access public networks as long as gateway nodes are provided to perform *address translation*, a process that rewrites egress packets to use public IP addresses and reverse the process on ingress. By leveraging RFC 1918 ranges and address translation gateways, most individuals and organizations can access the IPv4 Internet without relying on dedicated addresses in the depleted public range.

The two most common flavors of address translation are Port Address Translation (PAT) and Network Address Translation (NAT). PAT gateways map internal addresses to an external, public address by mapping internal hosts to specific TCP and UDP ports. Likewise, NAT gateways

maintain a one-to-one relationship between internal and external addresses that are drawn from a pool of public values. It is common practice for both techniques to be lumped under the generic designation of NAT.