BSc in Telecommunications Engineering

TEL3214

Computer Communication Networks

Lecture 06
The Internet Protocol

Eng Diarmuid O'Briain, CEng, CISSP

Department of Electrical and Computer Engineering,
College of Engineering, Design, Art and Technology,
Makerere University

Table of Contents

Illustration Index

This page is intentionally blank

1. Internet Protocol version 4 (IPv4)

IP version 4 (IPv4 or IP) was defined initially in 1980 and finalised in RFC 791 in 1981. It has been the mainstay of the Internet ever since, though the pressure on its limited address space of 4.3 billion addresses (2^{32}) is now telling in the era of the Internet of Things (IoT). This is forcing change to IPv6 with its $3.4x10^{38}$ addresses (2^{128}).

The IPv4 address defines the host at the Network/Internet layer and it has two sections or parts. The most significant part represents the network identification and the least significant part represents the node identification. The address is formatted as four numbers between 0-255, or 32 bits or four bytes, each byte represents either the network or the node in what is called dotted decimal notation.

There are some restrictions: The address 0.0.0.0/0 is reserved default routing and 127.0.0.1 is reserved for the local loopback or local host. 0's in the node part refers to this network i.e.192.168.0.0. All 1s (typically 255) is reserved for sending packets to all devices, this is known as broadcast i.e. 198.162.255.255).

1.1 Internet Protocol classes

There may be different types of networks or addresses in the different class assignments:

- ***Class A*** (network.host.host.host): *1.0.0.1* to *126.255.255.255* (126 networks, 16 million nodes) define the large networks. The binary standard is: 0 + 7 network bits + 24 node bits.
- ***Class B*** (network.network.host.host): *128.0.0.1* to *191.255.255.254* (16K networks, 65K nodes); (usually, the first node byte is used to identify subnets within an institution). The binary standard is 10 + 14 network bits + 16 node bits.
- ***Class C*** (net.net.net.host): *192.0.0.1* to *223.255.255.254* (2 million of networks, 254 nodes). The binary standard is 110 + 21 network bits + 8 node bits.
- ***Class D****: 224.0.0.1 to 239.255.255.255.* Reserved for Multicast use.
- ***Class E****: 240.0.0.1 to 255.255.255.255.* Reserved for experimental purposes.

1.2 Private addresses

Some address ranges have been reserved so that they do not correspond to public networks, and are considered to be private networks. These are nterconnected computers without external connection; messages between them will not be sent through Internet, but through an intranet. These address ranges are:

- **Class A** *10.0.0.0* to *10.255.255.255*
- **Class B** *172.16.0.0* to *172.31.0.0*
- **Class C** *192.168.0.0* to *192.168.255.0*.

1.3 Broadcast address

The broadcast address is special, because each node in a network listens to all the messages received with the broadcast destination address (as well as its own address of course). This address makes it possible to send datagrams (generally routing information and warning messages) to a network and all nodes on the network will be able to read them. For example, when ARP tries to find the Ethernet address corresponding to an IP, it uses a broadcast message, which is sent to all the machines on the network at the same time. Each node in the network reads this message and compares the IP that is being searched and sends back a message to the sender node if they match.

1.4 Routing

Routing, represents the mode in which the messages are sent through the networks. For example, there are three departments with Ethernet networks

1. Sales (subnet 192.168.2.0),
2. Clients (subnet 192.168.4.0),
3. Human Resources (HR), (subnet 192.168.6.0)
4. Backbone with GbE (subnet 192.168.1.0).

In order to route the packets between the hosts on the three networks, three gateway routers are required that will each have two network interfaces to switch between Ethernet and GbE plus a gateway to connect to the Internet. These would be:

1. SalesGW IPs: 192.168.2.1 and 192.168.1.2,
2. ClientsGW IPs: 192.168.4.1 and 192.168.1.3
3. HRGW IPs: 192.168.6.1 and 192.168.1.4, one IP on the subnet side and another on the backbone side.
4. InternetGW IP: 192.168.1.1 and the Internet side is determined by the companies Internet Service Provider (ISP).

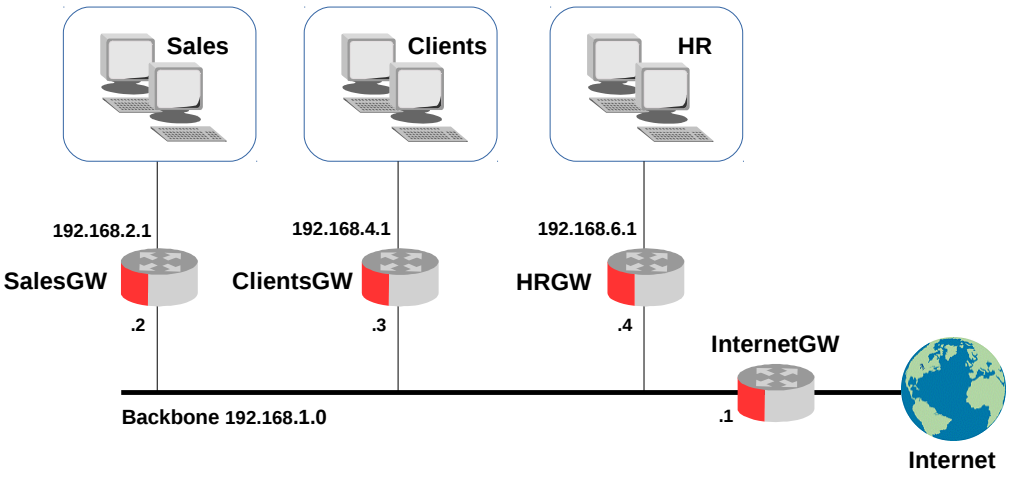

Illustration 1: Routed network

When messages are sent between devices in the Sales area, it is not necessary to leave the gateway, as the TCP/IP will find the destination device directly. The problem arises when the Sales device wishes to send a packet to a device on the HR network. The message must pass through the two respective routers. When Sales *identifies* that HR is on another network, it sends the packet through the SalesGW router, which in turn sends it to HRGW, which, in turn, sends it to the respective HR device.

TCP/IP routers use a table to route the packets between the different networks. A special route exists in the table to the network 0.0.0.0 which is a route of last resort. All the IP addresses on the Internet match with this route, as none of the 32 bits are necessary; they are sent through the default gateway router to the indicated network (assuming it can route to the required network).

In the SalesGW router, for example, the table would be:

Address	Mask	Gateway	Interface
192.168.1.0	255.255.255.0 /24	-	eth1
192.168.4.0	255.255.255.0 /24	192.168.1.2	eth1
192.168.6.0	255.255.255.0 /24	192.168.1.3	eth1
0.0.0.0	0.0.0.0	192.168.1.1	eth1
192.168.2.0	255.255.255.0 /24	-	eth0

Illustration 2: IP Address table

The '-' means that the machine is directly connected and does not need routing. The procedure for identifying whether routing is required or not consists of performing a very simple operation with the two logic ANDs (subnet AND mask and origin AND mask) and comparing the two results.

If they match, there is no routing required, if they are not a match then the packet is sent to the respective gateway router for onward forwarding. Each device must have its default gateway router pre-configured so it knows where to send such packets.

For example, a message from 192.168.2.4 to 192.168.2.6 would mean:

- 192.168.2.4 AND 255.255.255.0 = 192.168.2.0
- 192.168.2.6 AND 255.255.255.0 = 192.168.2.0

As the results are the same, there would be no routing required, simply a local ARP request to find the Medium Access Control (MAC) address of the destination device. On the other hand, for a packet from 192.168.2.4 to 192.168.6.6 routing will occur via the gateway router 192.168.2.1 with an interface change (eth0 to eth1) to 192.168.1.1 and from here to 192.168.1.2 with another interface change (eth1 to eth0) and then the packet will be forwarded to 192.168.6.6. Routes are matched in the routing table with the shortest mask first and the 0.0.0.0/0 default route is only used as a route of last resort.

In order to build the routing tables, the *ip route* command can be used to specify routes in the routes table, these are called *static* routes. However, for more complex networks such manual programming is unrealistic and dynamic building of the routing tables is necessary. For dynamic routing an Internal Gateway Protocol (IGP) like the Open Shortest Path First (OSPF) protocol or, between independent systems, an External Gateway Protocol (EGP) like Border Gateway Protocol (BGP) is used.

1.5 Quagga

The *quagga* package is the GNU/Linux routing daemon and it supports Border Gateway Protocol version 4 (BGP4), BGP4 plus (BGP4+), OSPF version 2 (OSPFv2), OSPF version 3 (OSPFv3), Intermediate System to Intermediate System (IS-IS), Routing Internet Protocol (RIP), RIP version 2 (RIPv2), and RIP Next Generation (RIPng).

To install a host on an existing network, it is necessary to have the following information, obtained from the network provider or the administrator:

- node IP address
- network IP address
- broadcast address
- netmask address / netmask length
- gateway router address
- DNS server address

2. Subnetting

As just mentioned additional bits can be added on to the self encoded subnet mask for a given class to further subnet a network. When a bitwise AND is performed between the subnet mask and IP address the result from the addition bits defines the resident subnet. However there are some restrictions on the subnet address. Network addresses of all 0's and all 1's are reserved for specifying this resident subnet (when a host does not know its resident subnet) and all hosts (broadcast address) respectively. This also applies to subnets. Therefore:

A subnet address cannot be all 0's or all 1's.

This also implies that a 1 bit subnet mask is not allowed.

Here is an example:

10010110.11010111.**0001**0001.01100010	150.215.017.098	IP Address
11111111.11111111.**1111**0000.00000000	255.255.240.000	Subnet Mask
---		
10010110.11010111.**0001**0000.00000000	150.215.016.000	Resident Subnet

← Network →← →← Host →
Subnet

Note: 255.255.240.000 can be written as /20 indicating the number of 1s in the mask.

2.1 Breakdown subnets using the rote method

Calculate the Network for an IP Address whose subnet is not /8 /16 /24

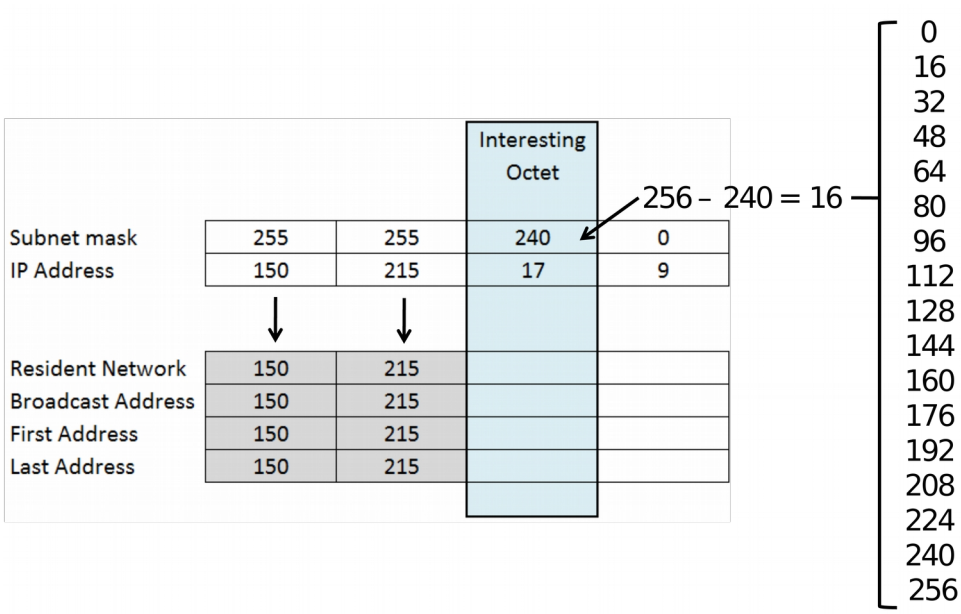

Illustration 3: Subnetting 150.215.17.9/20

- Determine the “interesting octet”, this is the octet where the subnet mask is not ‘0’ or ‘255’.
- Subtract the value in the interesting octet from 256, this is called the subnet differentiator and from zero (0) create a list in multiples of the value up to 256.
- Bring down the values on the left hand side from the IP Address as these will not change.

2.1.1 Calculate the 'Resident Network'

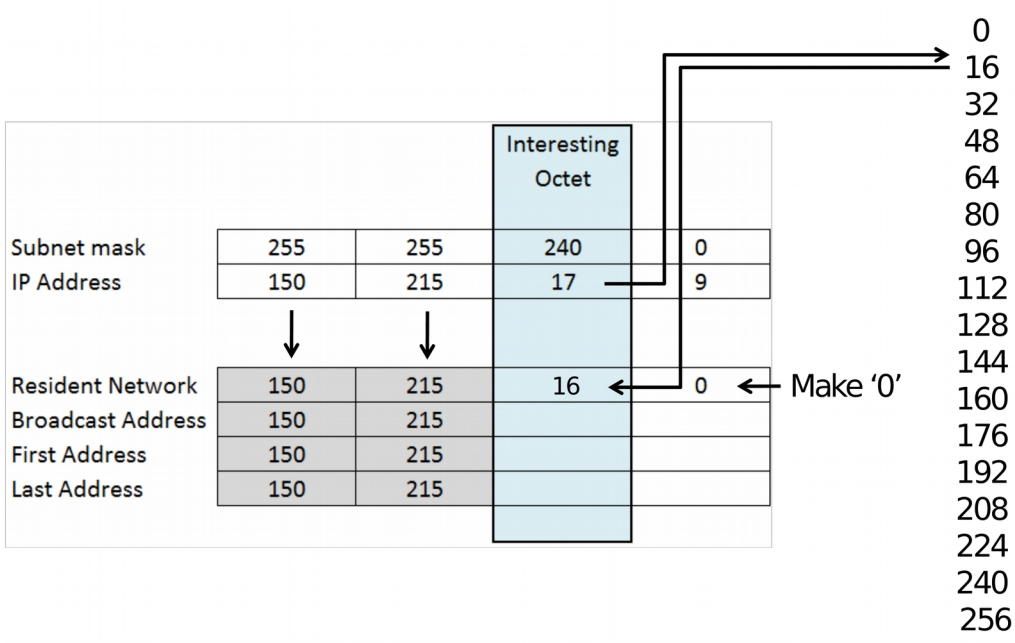

Illustration 4: Resident network of 150.215.17.9/20

- Make the right side of the resident network zero.
- Looking at the interesting octet for the IP Address select from the subnet list the nearest lower number from it and put that number in the interesting octet of the resident network.

2.1.2 Calculate the 'Broadcast Address'

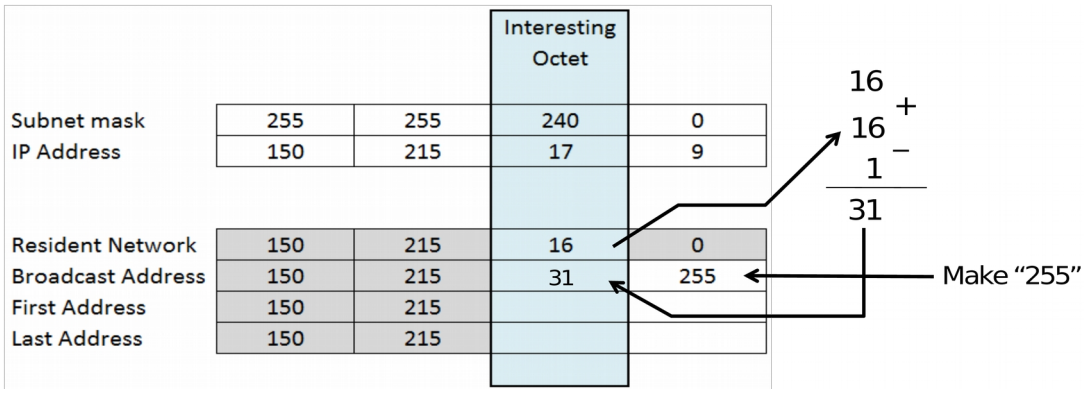

Illustration 5: Broadcast address of network containing 150.215.17.9/20

- Make the right side of the broadcast address '255'.
- Add the subnet differentiator and the interesting octet from the resident network and subtract '1', enter the answer as the interesting octet in the broadcast address.

2.1.3 Calculate the 'First Address'

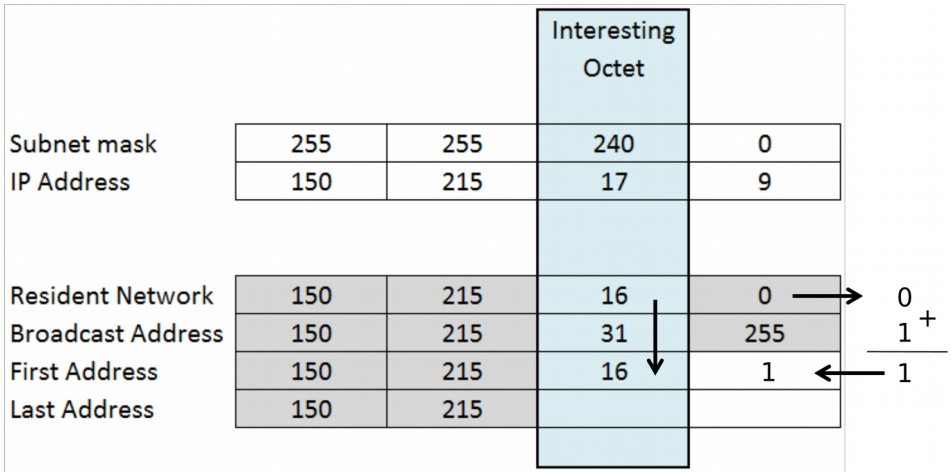

Illustration 6: First address of network containing 150.215.17.9/20

- This is basically the resident network plus '1'.
- Use the resident network interesting octet in the first address interesting octet field.
- Add '1' to the right hand octet of the resident network and place in the first address right hand octet field.

2.1.4 Calculate the 'Last Address'

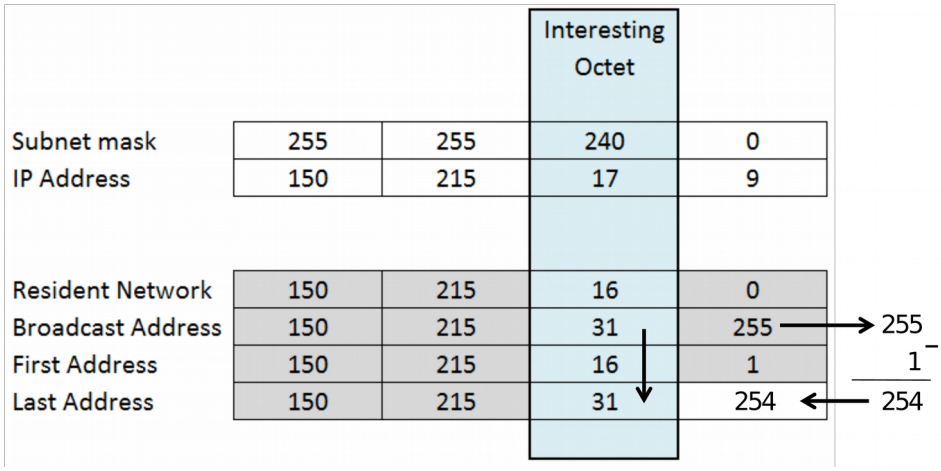

Illustration 7: Last address of network containing 150.215.17.9/20

- This is basically the broadcast address less '1'.
- Use the broadcast address interesting octet in the last address interesting octet field.
- Subtract '1' from the right hand octet of the broadcast address and place in the last address right hand octet field.

2.1.5 Network for an IP Address whose subnet is not /8 /16 /24

			Interesting Octet	
Subnet mask	255	255	240	0
IP Address	150	215	17	9
Resident Network	150	215	16	0
Broadcast Address	150	215	31	255
First Address	150	215	16	1
Last Address	150	215	31	254

Illustration 8: Table of subnetting details for 150.215.17.9/20

Here is a complete picture of the network into which 150.215.17.9/20 resides.

In the example a 4 bit subnet mask was used. The subnet in this case was 16. There are 14 subnets available with this mask (remember subnets with all 0's and all 1's are not allowed). Each subnet has 4,094 nodes (because of broadcast and network restrictions). This gives a total of 57,316 nodes for the entire class B address. Notice that this is less than the 65,534 nodes an unsubnetted class B address would have. Subnetting always reduces the number of possible nodes for a given network. To calculate the number of subnets or nodes use the following where n = number of bits in either the subnet or node field.

$$\text{Max nodes} = 2^n - 2$$

multiplying the number of subnets by the number of nodes available per subnet gives you the total number of nodes available for you class and subnet mask. Note that although subnet masks with non-contiguous mask bits are allowed they are not recommended.

2.2 Breakdown a subnet using the binary method

Subnetmask	255.255.240.0 (/20)
IP Address	150.215.17.9

Interesting octet is the 3rd octet, marked in blue. Convert interesting octet to binary and use | symbol to indicate break between network and host portions.

Subnetmask	255.255.	1111 | 0000	.0		
IP Address	150.215.	0001 | 0001	.9		
Resident network	150.215.	0001 | 0000	.0		
	150.215.	16	.0	→	150.215.16.0
First Address	150.215.	0001 | 0000	.1		
	150.215.	16	.1	→	150.215.16.1
Last Address	150.215.	0001 | 1111	.254		
	150.215.	31	.254	→	150.215.31.254
Broadcast	150.215.	0001 | 1111	.255		
	150.215.	31	.255	→	150.215. 31.255

2.2.1 Subnetting rote example

Breakdown the following Subnet

Breakdown 192.168.1.0/27 to show all its possible subnets

256 - 224 = 32

				Interesting Octet	
Subnet mask	255	255	255	224	
Network	192	168	1	0	Zero Subnet
Subnet 2	192	168	1	32	
Subnet 3	192	168	1	64	
Subnet 4	192	168	1	96	
Subnet 5	192	168	1	128	
Subnet 6	192	168	1	160	
Subnet 7	192	168	1	192	
Subnet 8	192	168	1	224	Broadcast Subnet
Subnet 9	192	168	1	256	

Illustration 9: Breakdown 192.168.1.0/27

Original Mask:	Class C	255.255.255.0	/24
Subnet bits	27 – 24 = 3		
No. of Subnets	$2^3 = 8$		
How many hosts/subnet:	$2^5 - 2 = 30$		
Valid Subnets	192.168.1.	0, 32, 64, 96, 128, 160, 192, 224	

2.2.2 Subnetting binary example

192.168.1.0 with subnetmask of 255.255.255.224 (/27)

Interesting octet is the last octet.

255.255.255.	111 | 00000
192.168.1.	000 | 00000

192.168.1.	001 | 00000	32
192.168.1.	010 | 00000	64
192.168.1.	011 | 00000	96
192.168.1.	100 | 00000	128
192.168.1.	101 | 00000	160
192.168.1.	110 | 00000	192

So there are 6 possible subnets available. As an example breakdown the network 192.168.1.96/27 some more.

Resident network	192.168.1.	011 | 00000		
	192.168.1.	96	→	192.168.1.96
First address	192.168.1.	011 | 00001		
	192.168.1.	97	→	192.168.1.97
Last address	192.168.1.	011 | 11110		
	192.168.1.	126	→	192.168.1.126
Broadcast	192.168.1.	011 | 11111		
	192.168.1.	127	→	192.168.1.127

2.2.3 Subnetting rote example

Breakdown the following Subnet

Breakdown 172.1.0.0/20 to show all its possible subnets

256 - 240 = 16

			Interesting Octet		
Subnet mask	255	255	240	0	
Network	172	1	0	0	Zero Subnet
Subnet 1	172	1	16	0	
Subnet 2	172	1	32	0	
Subnet 3	172	1	48	0	
Subnet 4	172	1	64	0	
Subnet 5	172	1	80	0	
Subnet 6	172	1	96	0	
Subnet 7	172	1	112	0	
Subnet 8	172	1	128	0	
Subnet 9	172	1	144	0	
Subnet 10	172	1	160	0	
Subnet 11	172	1	176	0	
Subnet 12	172	1	192	0	
Subnet 13	172	1	208	0	
Subnet 14	172	1	224	0	
	172	1	240	0	Broadcast Subnet
	172	1	256	0	

Illustration 10: Breakdown 172.1.0.0/20

Original Mask:	Class B	255.255.0.0	/16
Subnet bits	20 – 16 = 4		
No. of Subnets	$2^4 - 2 = 14$		
How many hosts/subnet:	$2^{12} - 2 = 4094$		
Valid Subnets	172.1.	16.0, 32.0, 48.0, 64.0, 80.0, 96.0, ……..	

2.2.4 Subnetting binary example

172.1.0.0 with a subnetmask of 255.255.240.0 (/20)
Interesting octet is again the 3rd octet.

255.255. 1111 | 0000 .0
172.1. 0000 | 0000 .0

172.1. 0001 | 0000 (16) .0
172.1. 0010 | 0000 (32) .0
172.1. 0011 | 0000 (48) .0
172.1. 0100 | 0000 (64) .0
172.1. 0101 | 0000 (80) .0
172.1. 0110 | 0000 (96) .0
172.1. 0111 | 0000 (112) .0
172.1. 1000 | 0000 (128) .0
172.1. 1001 | 0000 (144) .0
172.1. 1010 | 0000 (160) .0
172.1. 1011 | 0000 (176) .0
172.1. 1100 | 0000 (192) .0
172.1. 1101 | 0000 (208) .0
172.1. 1110 | 0000 (224) .0

As an example breakdown the network 172.1.096.0/20 some more.

Resident network	172.1.	0110 | 0000 (96)	.0		
	172.1.	96	.0	→	172.1.96.0
First address	172.1.	0110 | 0000 (96)	.1		
	172.1.	96	.1	→	172.1.96.1
Last address	172.1.	0110 | 1111 (111)	.254		
	172.1.	111	.254	→	172.1.111.254
Broadcast	172.1.	0110 | 1111 (111)	.255		
	172.1.	111	.255	→	172.1.111.255

Note: Some routers will allow the use of the top and bottom subnet. In reality however this is technically this is incorrect.

i.e. Taking the first example some routers will allow:

192.168.1.0255.255.255.224	→	192.168.1.0/27
Valid Subnets	192.168.1.	0, 32, 64, 96, 128, 160, 192, 224

3. Classless Inter-Domain Routing (CIDR)

Classless Inter-Domain Routing (CIDR) using variable length subnet masks (VLSM) was created to allow for greater flexibility with routed IP networks, to allow for the accelerating expansion of the Internet.

By 1990, the Internet was facing serious growth pains. The two most severe problems were the explosion of routing table size and the looming exhaustion of Class B networks. The wild popularity of the net had triggered a flood of new classful networks, and every one had to be included in the routing tables. The routers were running out of memory, and spending far too much time doing address lookups. Furthermore, it had become apparent that the pace of requests for new Class B networks would soon exhaust the available supply.

The Internet Engineering Task Force (IETF), recognising the urgency of these twin problems, assigned a group to develop a solution. That solution became known as Classless Inter-Domain Routing (CIDR) or supernetting and is the addressing scheme currently used in the Ipv4 parts of the Internet.

3.1 CIDR Blocks

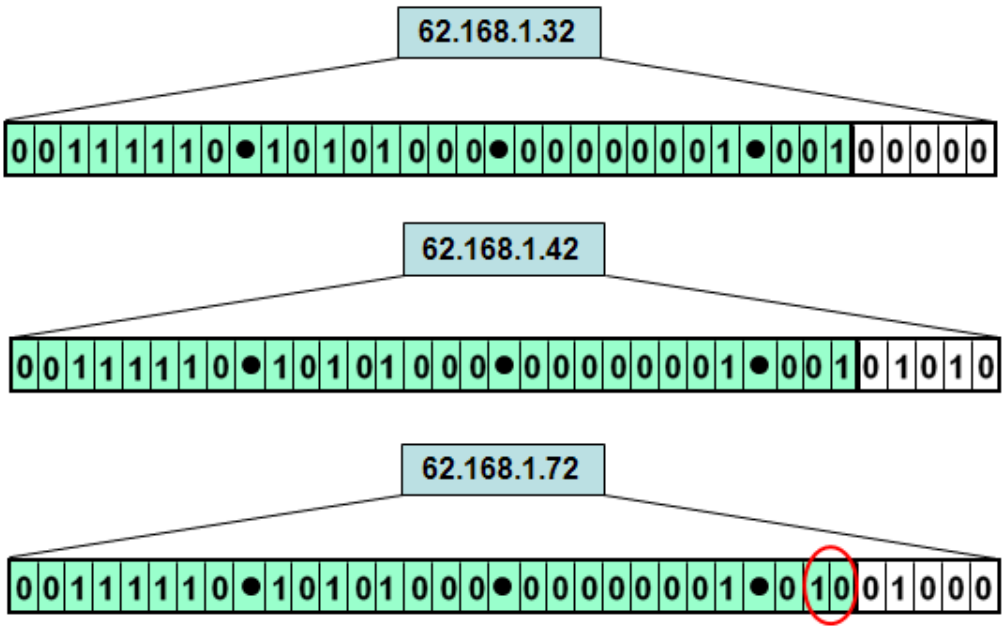

Illustration 11: CIDR Blocks

CIDR facilitates routing by allowing blocks of addresses to be grouped together into single routing table entries. These groups, commonly called CIDR blocks, share an initial sequence of bits in the binary representation of their IP addresses. IPv4 CIDR blocks are identified using syntax similar to that of IPv4 addresses: a four-part dotted-decimal address, followed by a slash, then a number from 0 to 32: A.B.C.D/N. The dotted decimal portion is interpreted, like an IPv4 address, as a 32-bit binary number that has been broken into four 8-bit bytes. The number following the slash is the prefix length, the number of shared initial bits, counting from the left-hand side of the address.

An IP address is part of a CIDR block, and is said to match the CIDR prefix if the initial N bits of the address and the CIDR prefix are the same. Thus, understanding CIDR requires that IP address be visualised in binary. Since the length of an IPv4 address is fixed at 32 bits, an N-bit CIDR prefix leaves 32 − N bits unmatched, and there are 2(32 − N) possible combinations of these bits, meaning that 2(32 − N) IPv4 addresses match a given N-bit CIDR prefix.

CIDR is also used with IPv6 addresses, where the prefix length can range from 0 to 128, due to the larger number of bits in the address. A similar syntax is used: the prefix is written as an IPv6 address, followed by a slash and the number of significant bits.

In the example in the diagram 62.168.1.42 in 62.168.1.32/27 but 62.168.1.72 is not.

3.2 Assignment of CIDR blocks

The Internet Assigned Numbers Authority (IANA) issues to Regional Internet Registries (RIR) large, short-prefix CIDR blocks. For example, 62.0.0.0/8, with over sixteen million addresses, is administered by Réseaux IP Européens Network Coordination Centre (RIPE NCC), the European RIR. The RIRs, each responsible for a single, large, geographic area (such as Europe or North America), then subdivide these blocks into smaller blocks and issue them publicly. This subdividing process can be repeated several times at different levels of delegation. Large Internet service providers (ISPs) typically obtain CIDR blocks from a RIR, then subdivide them into smaller CIDR blocks for their subscribers, sized according to the size of the subscriber's network.

Networks served by a single ISP are encouraged by IETF to obtain IP address space directly from their ISP. Networks served by multiple ISPs, on the other hand, will often obtain independent CIDR blocks directly from the appropriate RIR.

3.3 Variable Length Subnet Masks (VLSM)

A subnet mask is a bitmask that encodes the prefix length in a form similar to an IP address - 32 bits, starting with a number of 1 bits equal to the prefix length, ending with 0 bits, and encoded in four-part dotted-decimal format. A subnet mask encodes the same information as a prefix length, but predates the advent of CIDR.

CIDR uses variable length subnet masks (VLSM) to allocate IP addresses to subnets according to individual need, rather than some general network-wide rule. Thus the network/host division can occur at any bit boundary in the address. The process can be recursive, with a portion of the address space being further divided into even smaller portions, through the use of masks which cover more bits.

CIDR/VLSM network addresses are now used throughout the public Internet, although they are also used elsewhere, particularly in large private networks. An average desktop Local Area Network (LAN) user generally does not see them in practice, as their LAN is usually numbered using special private network addresses.

3.4 Prefix aggregation

Another benefit of CIDR is the possibility of routing prefix aggregation (also known as "supernetting" or "route summarisation"). For example, sixteen contiguous /24 networks could now be aggregated together, and advertised to the outside world as a single /20 route (if the first 20 bits of their network addresses match). Two aligned contiguous /20s could then be aggregated to a /19, and so forth. This allows a significant reduction in the number of routes that have to be advertised over the Internet, preventing 'routing table explosions' from overwhelming routers, and stopping the Internet from expanding further.

Nowadays most ISPs on the public Internet will not route anything smaller than a /19 prefix, effectively preventing small networks from obtaining full public Internet routing without going through a routing aggregator such as an ISP.

4. Internet Control Message Protocol (ICMP)

The Internet Control Message Protocol (ICMP) is a supporting protocol in the Internet Protocol suite. It is used by network devices, like routers, to send error messages and operational information indicating, for example, that a requested service is not available or that a host or router could not be reached. ICMP differs from transport protocols such as Transmission Control Protocol (TCP) and User Datagram Protocol (UDP) in that it is not typically used to exchange data between systems, nor is it regularly employed by end-user network applications with the exception of some diagnostic tools like ping and traceroute.

The most common ICMP messages are:

Type	Description
Echo Reply	used to ping
Echo request	used to pingDestination unreachable is generated by the host or its inbound gateway[6] to inform the client that the destination is unreachable for some reason. A Destination Unreachable message may be generated as a result of a TCP or UDP.
Destination Unreachable	Destination unreachable is generated by the host or its inbound gateway to inform the client that the destination is unreachable for some reason. A Destination Unreachable message may be generated as a result of a TCP or UDP.
Redirect Message	The message informs a host to update its routing information to send packets on an alternative route via another router on the network.
Router Advertisement	It is sent by a router on the LAN to announce its IP address as available for routing.
Router Solicitation	Router discovery/selection/solicitation. It is sent from a host to any routers on the LAN to request that they advertise their presence on the network.
Time Exceeded	TTL expired in transit. Time Exceeded is generated by a gateway to inform the source of a discarded datagram due to the time to live field reaching zero.
Parameter Problem	Bad IP header.
Timestamp	Timestamp is used for time synchronisation. The originating timestamp is set to the time in milliseconds since midnight the sender last touched the packet.
Timestamp Reply	Timestamp Reply replies to a Timestamp message. It consists of the originating timestamp sent by the sender of the Timestamp as well as a receive timestamp indicating when the Timestamp was received and a transmit timestamp indicating when the Timestamp reply was sent.

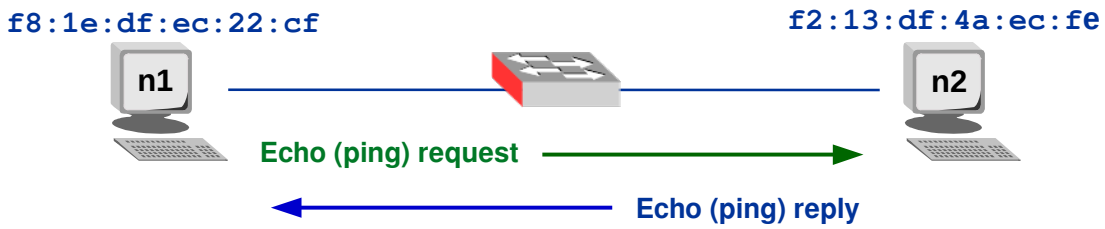

Illustration 12: ICMP Ping

Here is a ICMP Echo request packet.

```
Frame 1: 100 bytes on wire (800 bits)
Ethernet II, Src f8:1e:df:ec:22:cf, Dst f2:13:df:4a:ec:fe
Internet Protocol Version 4, Src: 192.168.10.4, Dst: 192.168.10.1
Internet Control Message Protocol
    Type: 8 (Echo (ping) request)
    Code: 0
    Checksum: 0x381e [correct]
    Identifier (BE): 19292 (0x4b5c)
    Identifier (LE): 23627 (0x5c4b)
    Sequence number (BE): 1 (0x0001)
    Sequence number (LE): 256 (0x0100)
    Data (48 bytes)

The device targetted responds with an ICMP Echo reply message.

Frame 2: 100 bytes on wire (800 bits)
Ethernet II, Src f2:13:df:4a:ec:fe, Dst f8:1e:df:ec:22:cf
Internet Protocol Version 4, Src: 192.168.10.1, Dst: 192.168.10.4
Internet Control Message Protocol
    Type: 0 (Echo (ping) reply)
    Code: 0
    Checksum: 0x401e [correct]
    Identifier (BE): 19292 (0x4b5c)
    Identifier (LE): 23627 (0x5c4b)
    Sequence number (BE): 1 (0x0001)
    Sequence number (LE): 256 (0x0100)
    Data (48 bytes)
```

This page is intentionally blank