IP SUBNETS

When we say that a device must figure out whether or not a packet's destination IP is local or not, what we really mean is that it must determine whether or not the destination IP is in the same subnet.

A subnet is a collection of similar IP addresses that share a single ethernet network. In other words, every device on a subnet must be able to send ethernet packets to every other device on the same subnet.

There is not a one-to-one mapping between ethernet networks and IP subnets though, so while a subnet can't be spread over multiple ethernet networks, one ethernet network can host multiple IP subnets. Having said that, in reality, the mapping usually is one-to-one, especially within our homes.

Note that when people talk about a Local Area Network, or LAN, they are often referring to an IP subnet. Do bear in mind though that many people use those terms very loosely, and often inaccurately.

In order to understand how a subnet is defined, we need a deeper understanding of IP addresses. We are accustomed to seeing them as so-called dotted quads (like192.168.0.1), but those dotted quads are just a human-friendly way of representing what an IP address really is – a 32bit binary number. The IP address 192.168.0.3 is really 11000000101010000000000000000011.

We create subnets of different sizes by choosing a dividing line somewhere inside this 32bit block, and saying that everything before our dividing line will be kept constant for this subnet. In other words, all IP addresses on a given subnet share the same prefix when expressed in binary form.

This means that to define a subnet we need two pieces of information – the first address in the subnet, and the location of the dividing line, or, to use the correct terms, we need a network address (or net address), and a netmask.

The size of the subnet (the number of IP addresses it contains) is determined by where you choose to place the divider. The closer to the front of the 32bits the bigger the subnet, the closer to the back, the smaller.

Netmasks are represented using 32bit binary numbers. All the bits before the chosen divide are set to 1, and all the bits after the divide are set to 0.

When written in binary, a netmask MUST be 32 bits long, and MUST consist of a series of1s followed by a series of 0s. The chosen dividing line represented by the netmask is the point where the 1s change to 0s.

Because netmasks are 32bit numbers, just like IP addresses, we can represented them in the same way, as dotted quads.

Source: https://www.bartbusschots.ie/s/2014/12/07/taming-the-terminal-part-25-of-n-ip-subnets/