IP Networking fundamentals-3
The “clever” bits Part 1

Layer 3? “Three-layer hierarchy”? what’s the difference?

A LOT of this is in the world of marketing. For a long time “layer 3” was a “differentiator” for vendors in networking, Layer 3 being the Network layer – or to practical people the IP layer. This is where we start considering IP addresses and “routing”.

Down here IPv4 and IPv6 behave in similar ways. IPv6 has sensibly eliminated “broadcast” traffic (within its scope of course) because both have the need to “bind layer 2 (MAC Address) to IP address(es) IPv4 uses ARP Address Resolution protocol (Broadcast) IPv6 uses NDP (Neighbour Discovery Protocol) which is multicast, and only subscribed to when needed by the individual devices, otherwise it doesn’t take resources from them. The MAC address is “bound” to an IP address. In IPv4 this is a 1 to 1 binding. In IPv6 this can be several IPv6 addresses bound to a MAC address. This is essentially the beautiful decoupling of medium from content. In IPv6 the distinction of this conceptual difference between infrastructure and data is better defined and separated. This is indeed possible in IPv4, but we need to engage in a greater battery of techniques at different layers to approach the clean separation in IPv6.

By contrast the “three-layer hierarchical model” is about how the infrastructure is described, This is not directly related to the OSI “stack” although there are some crossovers. Both being “conceptual” models, lets add some concrete examples. The Three Layer Hierarchical Model is concerned with the Access, Distribution, and Core Layers. Broad tenets existed – 1. Traffic from the access layer should not traverse the core. Simply – Cores exist to connect to other cores. The role of the distribution layer is to facilitate connection between access devices. This is a sound idea – traditionally one might connect ones servers to the core (in a small to medium sized network). In a datacentre servers may be “demoted” to the access layer. This model was at its strongest when bandwidth was scarce, and processing capacity was limited, even to the deepest pockets. The successor to the Three Layer model is a reflection on today – we have huge bandwidth, and immense processing capacity (relatively) this suggests we are entering a phase of flattening. I will write a separate piece on spine/leaf topologies, how they are very similar, slightly differently considered, and how they aren’t incompatible with the three-layer hierarchical model.

OSI layer 3 – the Network Layer

So, for now, we will concentrate on OSI Layer 3, and what it means. The difference between IPv4 and IPv6 is quite marked. In IPv4 we begin to deal with the address space limitations. In IPv6, because we don’t have this problem, the main functional intervention in layer 3 in IPv4 routing is of little consequence. It is often thought that routing is hard. Conceptually all a “route” does is direct traffic in a general direction. Much like we might see signs on a more major road directing us to a city. As we get closer to the city we may need a street map. This could be considered analogous to the MAC address table and its interaction between the ARP or NDP records. This analogy holds ok if we

consider that a “subnet” is a single broadcast domain where MAC addresses become visible, just like a street map.

The huge difference is in how the IP addressing is done. IPv4 will typically use IETF RFC1918 “private” addresses. These are the addresses you will be familiar with 10.0.0.1, 192.168.1.1, 172.16.1.1, these are formally notated in CIDR (classless Inter Domain Routing) format, which shows the size of the “super net” which tells us the “range”, we can subnet within this and do bespoke routing. RFC 1918 says that the following “private” IPV4 addresses will not be forwarded over the public Internet. This stops the Internet from breaking! The RFC 1918 address ranges are: **10.0.0.0/8** (10.0.0.0-10.255.255.255) **172.16.0.0/12** (172.16.0.0- 172.31.255.255) and **192.168.0.0/16** (192.168.0.0-192.168.255.255) In ipv6 we (typically have a /64 subnet behind a 64 bit “prefix”.

Because this is huge the IP addresses can be directly understood by all devices. “routing” becomes in many ways a lot more like switching in that devices can be considered end to end connected. It is different to switching however, in that a firewall can ensure that incoming traffic arriving “uninvited” will be denied by default – the standard “default-deny” rule. We can of course create rules that allow source addresses from within our organisation to traverse a local router (and keep our network running). Note that this is typically done “to” layer 3 (ip addresses) “from” a device working at “layer 7” the application layer. These rules can also be applied at layer 7 on every device if so desired.

Where IPv4 gets really complicated!

We are familiar with using the IETF RFC1918 addresses (even if we didn’t know that there is a convention covering them. Unlike IPv6, where all IP addresses are globally unique, IPv4 is exhaustible (and has been exhausted!). RFC1918 was a pragmatic reaction to this exhaustion. OK we “should” have end to end routing, its just, that we haven’t got the resources to do this! So, to “compensate” we started using NAT, well NAT with PAT, or dynamic NAT with PAT. NAT is the simple process of “Network Address Translation” PAT is “Port Address Translation”. Now IPv6 can do this of course, there are fewer base practical uses, however it can be essential in certain large migration scenarios, and may be a obfuscatory tool in high security applications (but not a “first line control” necessarily.) Its certainly not used in the way it is in IPV4. Simply getting ones head around NAT with PAT is a bit of gymnastics, Whilst I sound like an IPv6 evangelist (I’m after an easy life, and I’m enjoying software creation more these days!) IPv6 is both easier and more secure!

The simple reason is that to connect to the Public Internet you first need to connect to your ISP. This will be using the IP address of your gateway router (normally supplied by your ISP) Now, your “IP Address” cannot be 192.168.1.10 (for example) and your “gateway” is technically 192.168.1.1 – this is good so that your traffic from your IP address (of your computer) which may be 192.168.1.10 or anyway within the 192.168.1.0/24 subnet (this is CIDR notation – It means that the last 8 bits are the hosts in your subnet ie 192.168.1.1-192.168.1.254. this (as we discussed) is NOT an address that can be routed over the Public internet – it is an RFC1918 “private” address). This means that your ISP will likely have given you a dynamic public IP address. This is fine (in the domestic context) they may have flogged you a “static IP address” normally for a fee they think they can get away with! The sad truth is they are banking that the number of people who are boring enough to understand how dhcp works (as more and more ISPs are using DHCP to assign IP addresses) is so small that they will likely be the same people who will pay for a “static” IP. I will cover DHCP in more detail later, as this subject is immensely frustrating! However your public IP address will be from the ISPs pool of IP addresses. Hopefully they have more addresses than customers, as if not on IPv4 your Internet will break! Nobody at scale does IP assignment manually – its just far too dangerous! There are potentially other

“tricks” that ISPs use, which are very cool for every growing network, and even on IPv4 can make the VLAN assignment myth and mystery a thing of the past.

So, your router has a public IP address – this means it can connect to the Internet. This is where IPv4 needs NAT with PAT. This is not a criticism, in fact, its actually admiration. Someone very inspired thought – ok, if a domestic customer has only 1 IP address, but tens of machines wanting to connect to the Public Internet…. How many of these will be public facing servers? Answer – probably 0! so why not use the 16 bits the layer up (in the transport layer) which has 65535 “ports” for TCP and the same for UDP… and lets sneakily “borrow” some. When your computer tries to access a web page, it starts a TCP connection to port 443 at the target IP address, when it negotiates the session it uses a different TCP port to receive the incoming data. Your ISP router “sees” this. It grabs the port number you said you would like the return data on, and manipulates that so that it may give an entirely different return port number to the target website. It remembers its lie (like all good sales people) and when the website replies with what your isp router said, your isp router sends that back to your computer. This means that your isp router has remembered your internal IP address, has stripped the session ports out, remembered them, used its IP address, and the spare port it found to talk to the web server, and “translates the Network address from your 192.168.1.10 to its public IP address, and then does “port address translation” to deliver the resultant data to your computer… as if it wasn’t there! This means that the connection appears to be “end to end” but it is far from it. This is fine, until your network gets big. At home, this is highly unlikely!

Beware! There are many “IT managers” who believe that “NAT” is a security feature. Sadly, these are often the same folks that think VLANs are a security feature too. Neither are, security through obscurity isn’t security. Whilst these seem like “big” numbers, and the process seems complex, it really isn’t! that is why one can buy a $8 box that does this… Security is a different thing altogether! Then we have the large palette of “routing protocols”. In the 1990’s/early 2000’s every network vendor had their own “secret sauce” way of doing routing, The fact is that the most brilliant routing protocol is BGP – Border Gateway Protocol. As I alluded earlier, routing protocols are a way for routers to share their street maps to other routers. BGP is and has always been the routing protocol of the Internet. It is a BEAST! It has a way to share thousands of routes to its peers, and it has a security layer, that is cryptographically protected, so if a “new” peer asks for its routing table, it doesn’t say anything. It can also select what it shares to its peers. This (when I first started in networking) was a tool for space adventurers. Nearly 30 years later, its coming into enterprise networks. This is because it is excellent, and implementations don’t cost the earth anymore, and are available in hardware that is affordable for most medium sized businesses.