The structure of the Internet.

The Internet is basically a hierarchy that allows any Internet connected device in one geographic location, talk to another Internet connected device in another geographic location. The way that the information is transmitted varies greatly, and in some countries, wireless ham radios are even used to transmit email. Keep in mind that the word “connected” is used very loosely here.

The Open Systems Interconnection (OSI) Model

The seven open systems interconnection layers (OSI) are a staple of most networking textbooks. The idea is that a network will work on many different levels, or “layers” each of which will perform a supporting function for the next layer.

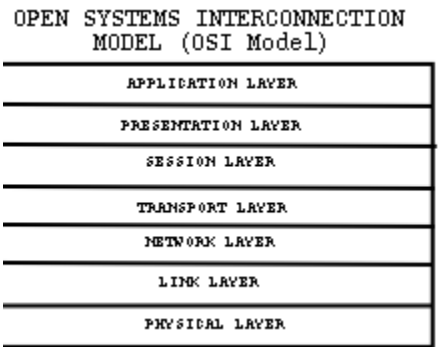

figure 1: open systems interconnection (OSI) model

The Network layers are the first three, being the physical link layer, the link layer, and the Network layer. Since the Internet is based on the Internet protocol which is in the Network Layer, and since the Internet can run on any number of different types of layers below that, we normally are not too concerned with the physical layers unless we are building an Ethernet cable, or transmitting an Internet signal through wireless means, and not to interested in the link layer unless we are registering a network card or router MAC address with our service provider.

The top three layers (session, presentation, and application layers) are for program communication, and are completely independant of the network so that the two communicating programs could even be on the same machine.

We also sometimes inlcude the transport layer when discussing the Internet, and often link the transport with the network layer as in the TCP/IP protocols. Transmission Control Protocol (TCP) is in the transport layer, and Internet Protocol (IP) is in the Network layer. Most Internet based functions such as the world wide web, and email, use TCP/IP, so this is a basic building block for the Internet.

The transport layer also makes sure that the top three layers are network independant.

Internet protocol

On the Internet, numbers, or Internet addresses, are used to allocate and address for each device on the network. These numbers are called Internet Protocol addresses, or IP addresses.

The current standard “IPV4” is usually separated into 4 tuples or bytes, such as 161.184.138.36, as a “dotted decimal”. What this means is that if we are to navigate the Internet to this IP address for some reason, then our

network ard will send out “packets” with that number in them, and a process called “routing” will send those packets to that Internet address.

These IP addresses are subdivided into three classes: a class A network, a class B network, and a class C network. The class A is the largest, class B is smaller, and class C is yet smaller. If we look at the four digit numbers in an IPV4 address, each network of class A will have different first number, and then it’s network will be addressed by the rest of the three numbers, or three bytes. Since a byte consists of 8 bits, a class a network can have 2 to the power of 24 (8 x 3), or 16777216 devices on it! Similarly, class B addresses have the first two bytes different, and has the authority over 2 to the power of 16, or 65536 devices on it, and class C addresses have one byte of address space, or 2 to the power of 8, or 256 devices on them

Not all addresses on the Internet are allowed, and some are reserved. In particular, the class A 10.0.0.0 addresses (from 10.0.0.0 – 10.255.255.255), the class B 172.16.0.0 (from 172.16.0.0 – 172.31.255.255), and the class C 192.168.0.0 (from 192.168.0.0 – 192.168.255.255) addresses are reserved fro private intranet use. If you had ever purchased a firewall router from your local business shop, and used it to expand your network in you home, then you may have noticed that it gives each network device an IP address that starts with 192.168.1.something.

These private IP’s have allowed the Internet to grow in the IPV4 address space without having to expand the number of addresses since you neighbour can have the same firewall router with the same internal IP addresses. What makes this work, is that when your Internet packets flow through the router, they are changed, or morphed, into the ip address that your access provider gives you.

IPV4

IPV4 is version 4 of the Internet Protocol.

Lets look at the structure of an IPV4 address. The class A addresses have the first bit zero, The next seven bits for the network id, or netuid, and the next 24 bits for the hostid. Similarly, class B addresses have the first two bits 10, and the next 14 bits as a netuid, and the next 16 bits as a hostid. A class C network has the first two bits 11, the next 21 bits for the netid, and the last 8 bits for the hostid. There is also something called a class D network that is reserved for broadcasting.

examples of IPV4 addresses are:
161.184.138.36 – 10100001 10111000 10001010 01001000 (class C public).
10.0.0.1 – 00001010 00000000 00000000 00000001 (class A private).
192.168.2.100 – 11000000 10101000 00000010 01100100 (class C private).
169.34.100.52 – 10101001 00100010 01100100 00110100 (class B public).
172.18.20.35 – 10101100 00010010 00010100 00100011 (class B private).

Internet packets, or Internet datagrams

An Internet protocol packet, or datagram, is a morcel of information that is sent out on the net. For example if you send an email to someone, then your computer will break that email message down into small peices and “encapsulate” them into packets with some destination address. This encapsulation means that your message, or more likely a small part of your message, will be put inside of an Internet packet.

Here is the whole picture with an IPV4 datagram sandwiched between the lower link layer. Remember that all data including the message to be sent is stored in sent in binary format (zero’s and ones):

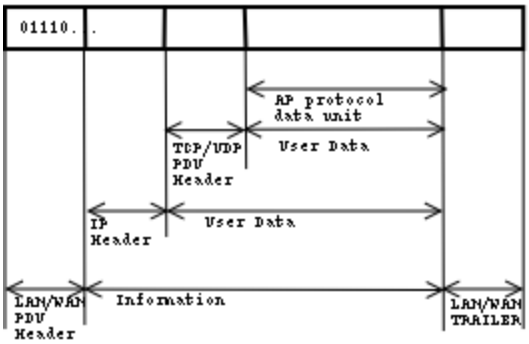

figure 2: IP datagram within a packet

Detail for this can be seen in Chapter 9.6.2 of “Data Communications Computer Networks, and Open Systems”, Third Edition by Fred Halsall published by Addison Wesley.

IPV6

IPV6 is version 6 of the Internet Protocol.

Based on RFC 2373, version 6 of the Internet Protocol has a vastly increased address space over the IPV4 version. IPV6 allows for 128 bit addressing, or 2 to the power of 128, or 3.40282e+38 devices on the network.

IPV6 notation is slightly different than IPV4 in that hex numbers (base 16) are used rather than decimal (base 10) numbers. As taken from the RFC Examples are:

FEDC:BA98:7654:3210:FEDC:BA98:7654:3210

1080:0:0:A0:8:800:200C:417A

As well any zeros in a row can be ignored as in:
1080:0:0:0:8:800:200C:417A a unicast address
FF01:0:0:0:0:0:0:101 a multicast address
0:0:0:0:0:0:0:1 the loopback address
0:0:0:0:0:0:0:0 the unspecified addresses

become
1080::8:800:200C:417A a unicast address
FF01::101 a multicast address
::1 the loopback address
:: the unspecified addresses

Please remember this notation as you may need to use it on the world wide web someday if your name services are not working.

Please go to RFC 2373 for more info.

Besides being able to handle many more than one device per person on earth, IPV6 should provide some new ways of doing things such as better encryption through the net, and complete source identity (ie. know where it is coming from) such as when it comes to sending emails since we can now bury the actual MAC address right into the datagrams.

What about IPV5? IPV5 was an experimental protocol called ST2 for delivering faster and more reliable services. It was abandoned in favour IPV6

routing

In reality, the Internet is really a collection of smaller networks linked together in many places. If we accept this model, then the process of routing, or sending Internet packets around the Internet is quite simple.

A host, or device, can only send messages within its own network. As the Internet grows, so do the networks, but remember that in the old Internet of the 1980's and early 1990's, there was only modems to communicate, so those acted as gateways. Even today if you are on, for example, the Telus network, then you must go through one of the Telus Internet gateways to see the rest of the Internet. Even to get out of your home network (hopefully protected by a firewall router), you must go through your router to get out of your house. This process of Internet datagram movement is called routing.

Each datagram has a source IP address, and a destination IP address in the IP header information. As a datagram is passed to the gateway (each host knows who its gateway is), it follows rules as to where it should go. Simply put, the router, or gateway looks at the packet and says "is this destination IP address in my network, or should I send it off to my gateway?". Obviously there is much more to it, but at the simplest level (recall our home firewall router example), that is really what happens in a "static routing table". There are, of course extremely complex rules automatically set by protocols such as border gate protocol (BDP) by higher level upstream providers (Telus, Sprint, etc.) since your destination IP address may not just be upstream, but may also be downstream somewhere too.

domains and the domain name system

All this talk about numbers is great for computers, but for us humans, we remember names better. After all it is much easier to remember a name like redhat.com, than 66.187.232.50. Names are also usefull incase we do something like change all the numbers on the Internet and make them REALLY hard to remember as will be the case as IPV6 is slowly introduced. IP addresses change, but names are supposed to be more static changing only when there is a human reason for it.

At first, in the 1970's, all the computers on ARPAnet knew about other computers through one text file called HOSTS.TXT. HOSTS.TXT, now /etc/hosts on UNIX and LINUX computers, held a name-to-address mapping of all the computers at the time. This was maintained by one Network Information Centre, or NIC for short. Changes were emailed to NIC when a new host was added, or one was deleted, and clients ftped the file to their own host to get the latest version.

When ARPAnet moved to TCP/IP, the population of the Internet exploded, and this primitive "hosts" scheme no longer worked.

To solve this, a distibuted database called the Domain Name System was created. The idea is to have a robust, and high performing database that is accessible to all devices on the Internet. Many use a phone book as an analogy for this system where your computer, or device has a number similar to a phone number, and the DNS system, similar to a phone book lets you look up that number to find another device on the Internet.

The stucture of the hierarchy is that there is one "root" point, and several nodes coming off of that root point called "top level domains", or tld's for short. We all have heard of "dot-com". Dot-com (.com) is a top level domain. So is .ca, .net, .org, .cc, .info, .uk, .edu, and a whole bunch of others. The website of the authority for the "root" of the Internet is at IANA.org. IANA is an acronym for the Internet Assigned Numbers Authority. The actual root point is not a single computer, but several core networks possibly separately corporately owned, but it is simpler to imagine it as a single point.

IANA does many things including distributing Internet numbers to providers. To do this they delegate to a non-profit organisation whose website is ICANN.org.

Back to the top level domains: there are currently three kinds: generic top level domains, or gTLD, sponsored top level domains, or sTLD, and country code top level domains or ccTLD. If you go to IANA.org's ccTLD whois information, and IANA.org's gTLD whois information, you can see all of the authorities or "sponsors" delegated for each top level domain. Figure 3 also shows a simple model of this.

figure 3: Internet structure

The stucture of the domain names below the top level domains is organised by the sponsors. For the ".uk", and ".au" domains, for example, the domain names are sub-divided into sub roots such as .co.uk, or .com.au whereas in the gTLD's we have all the names directly below the root name such as in compeng.net. In Canada we allow any Canadian legal entity to have a subdomain right below the tld such as computerengineering.ca. We also allow province level subdomains such as compusmart.ab.ca, and at one time allowed city level sub-domains such as joesgarage.edmonton.ab.ca.