Understanding TCP/IP

TCP/IP, the ubiquitous network protocol, is actually a four-layer suite of protocols and is well worth gaining an understanding of, if only to ensure that you set it up in the most efficient way on your network.

By Julian Moss

Everyone knows that TCP/IP is a network protocol used on LANs, WANs and the Internet, but not everyone who uses it understands how it works. It's possible to use TCP/IP with little more than a knowledge of how to configure the protocol stack, but a better understanding will give you a clearer picture of what is going on in your network and why the protocol needs to be set up in a particular way.

The aim of this multi-part article is to explain the key concepts behind TCP/IP.

TCP/IP stands for Transmission Control Protocol/Internet Protocol. If this leads you to think that it is not just one protocol, you're right. In fact, it is not just two protocols, either. TCP/IP is a suite of protocols. We'll cover the most important ones in the course of this article.

Layered Protocol

Like most network protocols, TCP/IP is a layered protocol. Each layer builds upon the layer below it, adding new functionality. The lowest-level protocol is concerned purely with the business of sending and receiving data - any data - using specific network hardware. At the top are protocols designed specifically for tasks like transferring files or delivering email. In between are levels concerned with things like routing and reliability.

The benefit that the layered protocol stack gives you is that, if you invent a new network application or a new type of hardware, you only need to create a protocol for that application or that hardware: you don't have to re-write the whole stack.

Link Layer

TCP/IP is a four-layer protocol, as illustrated in Figure 1. The lowest level, the link layer, is implemented within the network adapter and its device driver. Like all the TCP/IP protocols, it is defined by standards. The standards for generic Ethernet-type networks are defined by the IEEE 802 Committee: for example, IEEE 802.3 for Ethernet networks, or IEEE 802.5 for Token Ring networks.

Other link layer protocols that could be used include Serial Line IP (SLIP) or Point-to-Point Protocol (PPP), which are used when connecting to a network over an asynchronous dial-up link.

Since Ethernet is the most common type of network, we will look at it in a bit more detail. The Ethernet protocol is designed for carrying blocks of data called frames. A frame consists of a header containing 48-bit hardware destination and source addresses (which identify specific network adapters), a 2-byte length field, and some control fields. There follows the data, and then a trailer which is simply a 32-bit cyclic redundancy check (CRC) field. The data portion of an Ethernet frame must be at least 38 bytes long, so filler bytes are inserted if necessary.

All this means that frames are at least 64 bytes long, even if they carry only one byte of user data: a significant overhead in some types of application.

Frames also have a maximum size. Less headers, the maximum size for an Ethernet frame is 1492 bytes, which is the maximum transmission unit (MTU) for Ethernet. All link layer protocols have an MTU. It is one hardware characteristic that the higher-level protocol needs to be aware of, because larger blocks of data must be fragmented into chunks that fit within the MTU and then reassembled on arrival at their destination.

Network Layer

The next layer up from the link layer is called the network layer. The most important protocol at this level is IP, the Internet Protocol. Its job is to send packets or datagrams - a term which basically means "blocks of data" - from one point to another. It uses the link layer protocol to achieve this.

Both the network layer and the link layer are concerned with getting data from point A to point B. However, whilst the network layer works in the world of TCP/IP, the link layer has to deal with the real world. Everything it does is geared towards the network hardware it uses.

An IP address is a "soft" address. It is a bit like calling your office block "Pan-Galactic House" instead of its real address, 2326 Western Boulevard. The former is no use to the postman

"A router examines every packet, and compares the destination address with a table of addresses that it holds in memory."

who has to deliver the letters, unless he can use it to find out the latter. The link-layer Ethernet protocol needs to know the unique hardware address of the specific network adapter it has to deliver the message to and, in case of an error, the address of the one it came from.

To make this possible, the TCP/IP protocol suite includes link-layer protocols which convert between IP and hardware addresses. The Address Resolution Protocol (ARP) finds out the physical address corresponding to an IP address. It does this by broadcasting an ARP request on the network. When a host recognises an ARP request containing its own IP address, it sends an ARP reply containing its hardware address. There is also a Reverse ARP (RARP) protocol. This is used by a host to find out its own IP address if it has no way of doing this except via the network.

Internet Protocol

IP is the bedrock protocol of TCP/IP. Every message and every piece of data sent over any TCP/IP network is sent as an IP packet.

IP's job is to enable data to be transmitted across and between networks. Hence the name: inter-net protocol. In a small LAN, it adds little to what could be achieved if the network applications talked directly to Ethernet. If every computer is connected to the same Ethernet cable, every message could be sent directly to the destination computer.

Once you start connecting networks together, however, direct Ethernet communication becomes impractical. At the application level you may address a message to a computer on the far side of the world, but your Ethernet card can't communicate with the Ethernet card on that computer. Physical Ethernet limitations would prevent it, for a start. It would, in any case, be undesirable for every computer in the world to be connected to one big network. Every message sent would have to be heard by every computer, which would be bedlam.

Instead, inter-net communications take place using one or more "hops". Your Ethernet card will communicate with another Ethernet device on the route to the final destination. Routing is the important capability that IP adds to a hardware network protocol. Before we come to it, we will look at some other features of IP.

Features Of IP

IP is a connectionless protocol. This means that it has no concept of a job or a session. Each packet is treated as an entity in itself. IP is rather like a postal worker sorting letters. He is not concerned with whether a packet is one of a batch. He simply routes packets, one at a time, to the next location on the delivery route.

IP is also unconcerned with whether a packet reaches its eventual destination, or whether packets arrive in the original order. There is no information in a packet to identify it as part of a sequence or as belonging to a particular job. Consequently, IP cannot tell if packets were lost or whether they were received out of order.

IP is an unreliable protocol. Any mechanisms for ensuring that data sent arrives correct and intact are provided by the higher-level protocols in the suite.

Packets

An IP packet consists of the IP header and data. The header includes a 4-bit protocol version number, a header length, a 16-bit total length, some control fields, a header checksum and the 32-bit source and destination IP addresses. This totals 20 bytes in all.

We won't go into the detail of all the IP control fields. However, the protocol field is important. It identifies which higher-level TCP/IP protocol sent the data. When data arrives at its destination (either the packet's destination address equals the host's own IP address, or it is a broadcast address) this field tells IP which protocol module to pass it on to.

One control field, the time-to-live (TTL) field, is interesting. It is initialised by the sender to a particular value, usually 64, and decremented by one (or the number of seconds it is held on to) by every router that the packet passes through. When it reaches zero the packet is discarded and the sender notified using the Internet Control Message Protocol (ICMP), a network-layer protocol for sending network-related messages.

The TTL field is a safety mechanism which prevents packets from travelling the Internet forever in routing loops. It is exploited in a novel way by the Traceroute diagnostic tool (see box).

"The TTL field is a safety mechanism which prevents packets from travelling the Internet forever in routing loops. It is exploited in a novel way by the Traceroute diagnostic tool."

Application layer:	FTP, SMTP, SNMP
Transport layer:	TCP, UDP
Network layer:	IP
Link layer:	IEEE 802.x, PPP, SLIP

Figure 1 - TCP/IP is a four-layer protocol, of which the link layer is the lowest layer.

TCP/IP

Although the total field length in the IP protocol header is 16 bits, IP packets are usually much smaller than the 64 KB maximum this implies. For one thing, the link layer will have to split this into smaller chunks anyway, so most of the efficiency advantages of sending data in large blocks is lost. For another, IP standards did not historically require a host to accept a packet of more than 576 bytes in length. Many TCP/IP applications limit themselves to using 512-byte blocks for this reason, though today most implementations of the protocol aren't so restricted.

"Like most network protocols, TCP/IP is a layered protocol. Each layer builds upon the layer below it, adding new functionality."

Internet Addressing

Internet protocol addresses, or IP addresses, uniquely identify every network or host on the Internet. To make sure they are unique, one body, called InterNIC, is responsible for issuing them.

If your network is connected to the Internet and the computers need to be addressable from the Internet you must use IP addresses issued by InterNIC. If you don't use InterNIC-issued addresses, you must set up the gateway between your network and the Internet so that packets containing the made-up addresses will never pass through it in either direction.

Internet addresses are 32 bits long, written as four bytes separated by periods (full stops). They can range from 1.0.0.1 to 223.255.255.255. It's worth noting that IP addresses are stored in big-endian format, with the most significant byte first, read left to right. This contrasts with the little-endian format used on Intel-based systems for storing 32-bit numbers. This minor point can cause a lot of trouble for PC programmers and others working with raw IP data if they forget.

IP addresses comprise two parts, the network ID and the host ID. An IP address can identify a network (if the host part is all zero) or an individual host. The dividing line between the network ID and the host ID is not constant. Instead, IP addresses are split into three classes which allow for a small number of very large networks, a medium number of medium-sized networks and a large number of small networks.

Class A addresses have a first byte in the range 1 to 126. The remaining three bytes can be used for unique host addresses. This allows for 126 networks each with up to 16m hosts.

Class B addresses can be distinguished by first byte values in the range 128.0.x.x to 191.255.x.x. In these addresses, the first two bytes are used for the net ID, and the last two for the host ID, giving addresses for 16,000 networks, each with up to 16,000 hosts.

Class C addresses are in the range 224.0.0.x to 239.255.255.x. Here, the first three bytes identify the network, leaving just one byte for the individual hosts. This provides for 2 million networks of up to 254 hosts each.

Although these addresses make it possible to uniquely identify quite a lot of networks and hosts, the number is not that large in relation to the current rate of expansion of the Internet. Consequently, a new addressing system has been devised which is part of Internet Protocol version 6 (IPv6). IPv6 won't come into use for a couple of years, and understanding it isn't essential to understanding how IP works in general, so we won't cover it here. [*For a full description of IPv6, see article C0655 in PCNA 83 - Ed.*]

IP addresses can be further divided to obtain a subnet ID. The main net ID identifies a network of networks. The subnet ID lets you address a specific network within that network. This system of addressing more accurately reflects how real-world large networks are connected together.

You decide how the subnet ID is arrived at by defining a 32-bit value called the subnet mask. This is logically ANDed with the IP address to obtain the subnet address. For example, if a subnet mask was 255.255.255.0 and an IP address was 128.124.14.5, 128.124 would identify the Class B network, 128.124.14 would identify the

Traceroute - How It Works

Traceroute, if you haven't used it before, is a diagnostic tool that lets you find out the route Internet traffic takes between you and any given destination. It exploits the fact that traffic between two points will usually follow the same route at any given time, and that a router will notify the sender using an ICMP message whenever it receives an IP packet containing a time-to-live (TTL) field of one.

Normally, the TTL field of an IP packet is set to the value 64. Traceroute starts by sending a UDP datagram to the destination you specify, setting the TTL field to 1. The first router that receives it discards it, and sends an ICMP "time-to-live equals 0" notification back. In the header of the ICMP message is the router's IP address, from which its name can be determined. Next, Traceroute sends the datagram with a TTL of 2. This gets as far as the second router before being discarded. Again, an ICMP message comes back.

This process is repeated with ever-increasing TTLs until the datagram reaches the destination. To create an error when the destination is reached, the UDP datagram is addressed to a non-existent port on the destination host. This causes the host to respond with an ICMP "destination port unreachable" message. Thus, Traceroute knows that the route has been completed.

"If you don't use InterNIC-issued addresses, you must set up the gateway between your network and the Internet so that packets containing the made-up addresses will never pass through it in either direction."

subnetwork, and 5 would identify the host on that subnetwork. [*An article which covers subnet masks and related topics in more detail is currently in preparation - Ed.*]

Special Meanings

A few IP addresses have special meanings. A network ID of 0 in an address means "this network", so for local communication only the host ID need be specified. A host ID of 0 means "this host".

A network ID of 127 denotes the loopback interface, which is another way of specifying "this host". The host ID part of the address can be anything in this case, though the address 127.0.0.1 is normally used. Packets sent to the loopback address will never appear on the network. It can be used by TCP/IP applications that run on the same machine and want to communicate with one another.

Addresses in the range 224.x.x.x to 239.x.x.x are Class D addresses, which are used for multi-casting. Addresses 240.x.x.x to 247.x.x.x are reserved for experimental purposes.

Net, subnet and host IDs of all binary ones (byte value 255) are used when an IP packet is to be broadcast. Mercifully, an address of 255.255.255.255 does not result in a broadcast to the entire Internet.

Three sets of addresses are reserved for private address space - networks of computers that do not need to be addressed from the Internet. There is one class A address (10.x.x.x), sixteen class B addresses (172.16.x.x to 172.31.x.x), and 256 class C addresses (192.168.0.x to 192.168.255.x). If you have equipment which uses IP addresses that have not been allocated by InterNIC then the addresses used should be within one of these ranges, as an extra precaution in case router misconfiguration allows packets to "leak" onto the Internet.

IP Routing

So how does an IP packet addressed to a computer on the other side of the world find its way to its destination? The basic mechanism is very simple.

On a LAN, every host sees every packet that is sent by every other host on that LAN. Normally, it will only do something with that packet if it is addressed to itself, or if the destination is a broadcast address.

A router is different. A router examines every packet, and compares the destination address with a table of addresses that it holds in memory. If it finds an exact match, it forwards the packet to an address associated with that entry in the table. This associated address may be the address of another network in a point-to-point link, or it may be the address of the next-hop router.

If the router doesn't find a match, it runs through the table again, this time looking for a match on just the network ID part of the address. Again, if a match is found, the packet is sent on to the address associated with that entry.

If a match still isn't found, the router looks to see if a default next-hop address is present. If so, the packet is sent there. If no default address is present, the router sends an ICMP "host unreachable" or "network unreachable" message back to the sender. If you see this message, it usually indicates a router failure at some point in the network.

The difficult part of a router's job is not how it routes packets, but how it builds up its table. In the simplest case, the router table is static: it is read in from a file at start-up. This is adequate for simple networks. You don't even need a dedicated piece of kit for this, because routing functionality is built into IP.

Dynamic routing is more complicated. A router builds up its table by broadcasting ICMP router solicitation messages, to which other routers respond. Routing protocols are used to discover the shortest path to a location. Routes are updated periodically in response to traffic conditions and availability of a route. However, the details of how this all works is beyond the scope of this article.

Click here for the second part of this article

PCNA

The Author

Julian Moss is a freelance IT writer and programmer, and developer of Visual DialogScript, a scripting and automation tool for Windows. He can be contacted as jmoss@cix.co.uk.

Recent Reviews from Tech Support Alert

Reviews of the Best Windows Backup Software

In this detailed comparative review, we checked out eighteen backup software utilities designed for home or SOHO use. Many of the products reviewed were disappointing. However 6 products passed our tests with flying colors and 2 of these were so impressive, they were awarded our "Editor's Choice."

Suppliers of Cheap Inkjet Printer Cartridges Reviewed and Rated

With hundreds of companies all claiming to have the *"cheapest and best inkjet printer cartridges*," our editors decided to put their claims to the test. Not unexpectedly, many suppliers flunked but we did manage to come up with a number of web sites that sell good quality inkjet printer cartridges at heavily discounted prices.

The Best Anti Trojan Software

Our editors took a close look at the 6 leading anti-trojan/trojan remover software utilities. Unfortunately, they found only 2 products that were effective in their ability to detect and remove dangerous modern polymorphic and process injecting trojans.

The 46 Best Ever Freeware Utilities

This is our Editor, Ian "Gizmo" Richards, personal selection of the best freeware utilities. He's hunted down some real gems, many of which perform better than expensive commercial products.