

The Internet Explained

The exponential growth of the Internet has been phenomenal. Or has it? Perhaps it is only to be expected when the cumulative acts of creation culminate in the proliferation of Mankind's greatest achievement: the ability to communicate – but globally and with astonishing, lightning speed. Once the preserve of the scientific and military communities, the Internet has now blossomed into a vehicle of expression and research for the common person with hundreds of thousands, if not millions, of new pages being added to the World Wide Web every day and tens of millions of searches being performed through our ubiquitous search engines, the likes of Google, Yahoo!, MSN and other portals to the Internet delivering results to queries in our incessant quest for information.

In the Beginning

Some 45 years ago the search for knowledge was no less insatiable but the storage, collation, selection and retrieval technologies were rudimentary and the expense enormous by today's standards. 65 years past, with WWII at an end and the might, energy and focused intellect of galvanised nations waning war, the first computers were being built along with man-machine interfaces. It is at this time that visionaries first hinted at the possibilities of extending human intellect by automating mundane, repetitive processes, devolving them to machines. One such man, Vannevar Bush, in his 1945 essay, 'As We May Think' envisaged a time when a machine called a 'memex' might enhance human memory by the storage and retrieval of documents linked by association, in much the same way as the cognitive processes of the brain link and enforce memories by association.

Post-War Development

Bush's contribution to computing science, although remarkable, was far less critical than his efforts to unite the military and scientific communities together with business leaders, resulting in the birth of the National Defence Research Committee (NDRC) which was later to become the Office of Scientific Research and Development (OSRD). In short, Bush galvanised research into technology as the key determinant in winning the Second World War and established respect for science within the military.

A few years after the war the National Science Foundation (NSF) was setup, paving the way for subsequent government backed scientific institutions and ensuring the American nation's commitment to scientific research. Then in 1958, perhaps in direct response to the Soviet launch of Sputnik, the Advanced Research Projects Agency (ARPA) was created, and, in 1962, employed a psychologist by the name of Joseph Licklider. He built upon Bush's contributions by presaging the development of the modern PC and computer networking, and was responsible for penning 'Man Computer Interface', <http://memex.org/licklider.pdf> paper on the symbiosis of man and machine.

Having acquired a computer from the US Air Force and heading up a couple of research teams, he initiated research contracts with leading computer institutions and companies who would later go on to form the ARPANET and lay down the foundations of the first networked computing group. Together they overcame problems associated with connecting computers delivered from different manufacturers whose disparate communications protocols meant direct communications was unsustainable, if not impossible.

It is interesting to note that Lick was not primarily a computer man; he was a psychologist interested in the functionality of human thought but his considerations on the working of the human mind brought him into the fold of computing as a natural extension of his interest.

The Internet Explained

Other Key Players

Another key player, Douglas Engelbart, entered web history at this point. After gaining his Ph.D. in electrical engineering and an Assistant Professorship at Berkeley, he setup a research laboratory – the Augmentation Research Center – to examine the human interface and storage and retrieval systems, producing NLS (oNLine System) with ARPA funding, the first system to use hypertext (coined by Ted Nelson in 1965) for collation of documents – and is credited as the developer of the first mouse or pointing device.

All the while visionary minds were laying the groundwork for the Internet, the hardware giants were consolidating their computing initiatives: Bell produced the first 300 baud commercial modem, the Bell 103, sold by ATT; DEC (Digital Equipment Corporation) released the PDP-8 mass-produced minicomputer; and the first live transatlantic TV broadcast took place via ATT's Telstar 1 satellite.

Credit must be afforded another thinker, Paul Baran, for conceiving the use of packets, small chunks of a message which could be reconstituted at destination, upon which current internet transmission and reception is based. Working at the RAND Corporation and with funding from government grants into Cold War technology, Baran examined the workings of data transmission systems, specifically, their survivability in the advent of nuclear attack. He turned to the idea of distributed networks comprising numerous interconnected nodes. Should one node fail the remainder of the network would still function. Across this network his packets of information would be routed and switched to take the optimum route and reconstructed at their destination into the original whole message. Modern day packet switching is controlled automatically by such routers.

ARPANET

As computer hardware became available the challenge of connecting them to make better use of the facilities became a focus for concern, ARPA engaged a young networking specialist, Larry Roberts, to lead a team responsible for linking computers via telephone lines. Four university and research sites would be connected and it was decided to build Interface Message Processors (IMPs, devised by Wesley Clark), smaller computers talking a common language dedicated to handle the interfacing between their hosts and the Network. Thus the first gateways were constructed and the precursor to the Internet was born under the name of the ARPANET in 1969.

The '70s saw the emergence of the first networks. As the ARPANET grew it adopted Network Control Protocol (NCP) on its host computers and File Transfer Protocol (FTP) is released by the Network Working Group as a user-transparent mechanism for sharing files between host computers.

And, significantly, the first Terminal Interface Processor (TIP) is implemented, permitting computer terminals to connect directly to ARPANET. Users at various sites could log on to the Network and request data from a number of host computers.

Communications Protocols

In 1972 Vinton Cerf is called to the chairmanship of the newly-formed Inter-Networking Group (INWG), a team setup to develop standards for the ARPANET. He and his team built upon their NCP communications system and devised TCP (Transmission-Control Protocol) in an effort to facilitate communications between the ever-growing number of networks now appearing – satellite, radio, ground-based like Ethernet, etc.

They conceived of a protocol that could be adopted by all gateway computers and hosts alike which

The Internet Explained

would eliminate the tedious process of developing specific interfaces to diverse systems. They envisaged an envelope of information, a 'datagram', whose contents would be immaterial to the transmission process, being processed and routed until they reached their destination and only then opened and read by the recipient host computer. In this way different networks could be linked together to form a network of networks.

By the late '70s the final protocol was developed - TCP/IP (Internet Protocol) - which would become the standard for internet communications.

Ethernet

One final piece of computer networking came together under Bob Metcalfe: Ethernet <http://www.digibarn.com/collections/diagrams/ethernet-original/> He submitted a dissertation on the ARPANET and packet switching networks for his Harvard graduate dissertation but was disappointed to have his paper junked. After taking a position at Xerox's Palo Alto Research Center (PARC) he read a paper on Alohanet, the university of Hawaii's radio network.

Alohanet was experiencing problems with packet collision (information was being lost due to the nature of radio broadcasting). Metcalfe examined the problem then refined the principles of packet collision, adopted cable as the communications medium, formed 3Com and marketed his invention as Ethernet. The take-up was almost immediate and the '80s witnessed the explosion of Local Area Networks (LANs). First educational establishments then businesses employed Ethernet as the business communications networking standard, and once connected through communications servers to the Internet, the World Wide Web was just an initiative away.

Birth of the Browser

In fact, it was ready and waiting in the wings. Tim Berners-Lee (now Sir Tim) wrote a program, 'Enquire-Within-Upon-Everything', in 1980 whilst contracted to CERN, the particle physics laboratory in Geneva. He needed some means to collate his own and his colleagues' information – notes, statistics, results, papers – the plethora of output generated by the mass of scientists both at the institution and located across the globe at various research centres. The seed was sown and upon his return to CERN after other research, he set to work to resolve the problems associated with diverse communities of scientists sharing data between themselves, especially as many were reluctant to take on the additional workload of structuring their output to accommodate CERN's document architecture format.

By 1989, the Internet was well established, LANs proliferated in business - especially with the introduction of personal computers (PC) - and the adoption of Microsoft's ubiquitous Window operating system meant a stable(-ish) platform for users to create, store and share information. Tim Berners-Lee submitted a paper to CERN's board for evaluation, 'Information Management: A Proposal', <http://www.funet.fi/index/FUNET/history/internet/w3c/proposal.html> wherein he detailed and encouraged the adoption of hypertext as the means to manage and collate the vast sum of information held by CERN and other scientific and business establishments. Sadly, it sparked little interest but he persevered and in 1990 wrote the Hypertext Transfer Protocol (HTTP) along with a way of identifying unique document Internet addresses, the URI or unique resource indicator. To view retrieved documents he wrote a browser, 'WorldWideWeb' and to store and transmit them, the first web server.

The Internet Explained

The World Wide Web

CERN remained diffident to his system so Berners-Lee took the next logical step: distribute web server and browser software on the Internet. The spontaneous take-up by computer enthusiasts was immediate and the World Wide Web came into being.

The browser he created was tied to a specific make of computer, the NeXT; what was required was a browser suited to different machines and operating systems like Unix, the PC and the Mac, specifically so that businesses and governments, who were increasingly using the Web to manage their public information, could guarantee their users could use it.

Soon browsers for different platforms started appearing, Erwise and Viola for Unix, Samba for Macintosh and ... Mosaic for Unix, Mac and PC, created by Marc Andreessen whilst at the National Center for Supercomputing Applications (NCSA).

Mosaic took off in popularity to such an extent that it made front page of the New York Times' technical section in late 1993, and soon CompuServe, AOL and Prodigy begin offering dial-up internet access.

Andreessen and Jim Clark (founder of Silicon Graphics Inc.) decided to form a new company, Mosaic Communications Corporation, to develop a successor to Mosaic. Since the original program belonged to the university of Illinois and was built with their time and money, they had to start from scratch. He and Clark set about assembling a team of developers drawn from NCSA. Netscape Navigator was born and by 1996 3-quarters of web surfers used it.

The Internet in Practice

So how does the Internet work? It is important to remember the Internet is a network of computer networks interconnected by communications lines of various compositions and speeds. Interspersed across this immense network are routers which either guide traffic to specific destinations or keep it within well defined areas. This vastness of scale can be distilled into two basic actions: requests for information and the servicing of such requests, which forms the relationship between the two types of computer using the Internet: clients and servers. Whether connected to a local area network (LAN) at a place of business or attached by cable modem from home, computers requesting information across a network or the Web are generally regarded as clients; machines supplying the information are servers. In practice the distinction is less polarised, with many computers both requesting and delivering information, but the premise forms the basis of the Internet.

Servers often perform specific duties: web servers hosting websites, email servers forwarding and collecting email, FTP (File Transfer Protocol) servers uploading and downloading files.

Web Access

Access to the Web for home users is achieved by dial-up modem, cable (broadband or ADSL) or wireless connection to their ISP (Internet Service Provider); business users will typically be connected to a local area network and gain access via a communications server or gateway, which is again linked through an ISP to the Web. ISPs themselves may be connected to larger ISPs, leasing high speed fibre-optic communications lines. Each of these forms a gateway to the Web with the largest maintaining the 'backbones' of the Web through which run the international 'pipes' connecting the world's networks.

The Internet Explained

Addressing the Web

TCP/IP (Transmission Control Protocol/Internet Protocol) is the governing set of protocols used for transmitting information over the Internet. It establishes and manages the connection between two computers and the packets of data sent between them.

Each computer connected to the Internet has a unique IP address assigned to it, either dynamically at the moment of connection or for a period of a day or so, or (for all intents and purposes) a fixed or static address like that assigned to a web or name server hosting websites. The current version of IP, version 4, allows for 4.3 billion unique addresses – thought more than adequate a few years ago but, as there are now only a billion left, no longer sufficient to address not only the volume of new users and hosts coming online but also the influx of new technologies demanding attendant IP addresses such as those associated with smart internet-enabled machines like auto-ordering fridges, Pepsi dispensers and media centres and now internet phones. However, the shortfall is being remedied with the emergence of IPv6 and its 340 billion billion address slots which not guarantees practically limitless web access but also offers intrinsic unbreakable security encryption levels.

ICANN, the Internet Corporation for Assigned Name and Numbers is the non-profit North American organisation responsible for Internet IP address space allocation and DNS management, among other technical management functions.

Most users have no need to know the unique identities of computers with which they communicate since software deals with this on their PC, they simply address their email to whomever or logon to their shared network drive and drill down folders to load a file to work on.

An IP address looks like 194.79.28.133, a cluster of four numbers known as octets. People don't think of addresses in such a way – although they have been forced to for some time with phone and cell numbers and their PINs for credit cards – but, as with email, use names as mnemonics. As the Internet grew, it became obvious users seeking specific machines would need some method of identifying and recalling computers quite apart from IP addresses.

Domain Name System

The Domain Name System (DNS) <http://www.netbsd.org/guide/en/chap-dns.html> was conceived in 1984, basically a lookup translation table converting machine readable IP addresses into human understandable names. Locating a website by its name www.yourbusiness.co.uk rather than entering 123.23.48.146 in the browser address bar makes eminently more sense. These translation tables – name servers - are dotted across the Internet and contain specific references to website/IP addresses on their own local list, pointers to other name servers who may be able to locate the desired computer should it not be found locally and a cache (temporary list) of recently requested domain names.

Name servers are maintained and updated on a daily basis as IP addresses change or are added when new websites come online. Millions of people and automated systems maintain this distributed naming system worldwide and it is accessed by billions of surfers each day, requesting not just websites but email addresses and FTP servers. It is the biggest and most active distributed database in the world. There are special name servers called root servers which hold the addresses of all the Top-Level Domains (TLDs like .com, .co.uk, .net, .org, etc.). These are frequently interrogated whenever an unknown domain name (website) is requested and point the requesting name server to the address of the server holding the translation table or map for the requested website.

The Internet Explained

Obviously a single name server holding all internet addresses would be immediately brought to its knees so there are several servers duplicating domain addresses at various levels of the system and hundreds of thousands worldwide which, as well as speeding up the process of web access, serve as a layer of inbuilt redundancy should local failure occur.

All name servers are not updated immediately - which is why a new website is not instantly visible across the Internet. Additions to name server lists take time to propagate around the world but are usually achieved within a day or two.

Domain Management

Various organizations are responsible for individual TLDs, ensuring duplicate domain names cannot exist. These often country-specific organisations employ registrars, businesses accredited to register and lease domain names to companies and individuals.

Nowadays the registration process is automated and remarkably simple. Choose a domain name, check it is not already registered, select the lease period (no, you don't actually own the domain but the right to use it for a period of time, a minimum of one or two years) ... and pay for it.

The domain is then added to the registrar's local domain name server and propagated to the world's root name servers. Whether a website exists for the domain is immaterial, its potential existence and location is described and forwarded. Web hosting companies may or may not be registrars which means a domain may be registered with one company but hosted – made visible to the Internet through a web server – by another. In this instance, the domain will be registered and a change must be made to the default name servers list to point to another set of name servers owned by the hosting company.

Website Development

Actually building a website is another matter. With the creation of HTML (HyperText Markup Language), Sir Tim Berners-Lee offered developers the opportunity to apply special tags to describe the structure and shape of documents – web pages. The initial minimal set has been supplemented to include about 90 tags <http://www.w3.org/TR/REC-html40/serving> different purposes such as presenting headings, titles and lists to embedded multimedia and graphical objects – though not without some disdain since Berners-Lee at one time was at odds with Netscape for introducing the image tag which he felt was making the Web frivolous.

Once a website or indeed any internet destined documents have been constructed they are invariably transferred to the host server by means of FTP (File Transfer Protocol) by opening a channel to the web server using either a browser connection string, for instance, `ftp.yourdomain.com` entered into a browser address bar or via a dedicated FTP client, a software program designed specifically for the bi-directional transfer of files. Some are standalone, some are an inbuilt feature of website design programs.

Driving the Web

Visionaries in the scientific, military and business industries contributed to the World Wide Web as we now know it. Certainly, many individuals observed altruistic notions of a communications medium for Mankind but there is no denying the United States and international war governments sponsored the research initiatives behind its initial development, funding their countries' respective war and Cold War

The Internet Explained

efforts and eventually recognising, perhaps reluctantly, the critical contributions of their peoples' academic and scientific communities.

The World Wide Web is not the Internet, it is a subset, designed specifically for the universal interchange and dissemination of information, although the terms have become synonymous to many users. To put it another way, all internet users have access to the Web but not conversely since some areas of the Internet are restricted access – many scientific, military, educational and business networks require privileged access to non public areas, areas often dedicated to research and development.

Internet2

One such area is Internet2, a subscription-only multidiscipline consortium of high-speed networks connected (at least in the United States) by an ultra high-speed backbone, formed for the investigation, creation and deployment of cutting edge internet technology. It links some 200+ United States universities in addition to scientific communities, governments and business, many of whom pay some \$30,000 in annual membership fees plus an annual connection fee (where a point-of-presence is available) to Abilene, the company providing a 10 Gigabit fibre-optic, high-speed router transmission infrastructure, which may amount to hundreds of thousands of dollars. It further extends to research centres in other countries via high-speed links.

One reason for the creation of such and similar ultra high-speed networks is a direct result of scientific research where, with regards to particle physics for instance, vast quantities of data are generated, data requiring many months or even years to transmit at conventional speeds. A far less extreme reason is for transmission of broadcast quality video and streaming multimedia files – big business a la video on demand.

Faster Communications

Bandwidth, the capacity of a network to carry information, depends on a number of factors which are predetermined and usually hardware limiting. The transport protocol managing movement of data (packets) around the Web, TCP/IP, is a variable factor. A user may well enjoy a high-speed broadband link to their ISP but from there outwards there is no guaranteeing the speed or capacity of subsequent internet connections to the streaming video server you subscribed to.

As mentioned earlier, TCP/IP was developed in the mid-'70s and governs all Internet communications. It has remained largely unchanged. It's strength – and weakness – lies in its ability to adjust data transmission to meet internet conditions, namely congestion, transmission urgency and quality. It does this by sending re-requests for information when it doesn't receive confirmation of receipt by a certain time but it doubles the wait time after each re-request in response to network congestion algorithms. This is often why file downloads may begin with a burst of activity then speed deteriorates to frustrating slowness.

In response to this a new approach has been developed to using TCP: FAST or Fast Active queue Scalable Transmission protocol. FAST dynamically adjusts transmission speeds in response to how quickly it receives acknowledgements of successful packet transmission and has managed spectacular sustained transfer speeds. This is not to say FAST increases bandwidth – largely fixed by physical hardware limitations (and, of course, set to maximums by lease costs) – but it does increase efficiency – from a typical 25% to upwards of 95%.

The Internet Explained

Big business players like Microsoft and Disney have shown keen interest in its development especially now that digital media has come of age. And the beauty of FAST is its implementation does not require specialised client-based software or hardware upgrades; existing computers will be able to make use of it immediately upon its release (although current versions do not support Windows-based servers).

Future Development

As penetration and uptake of high-speed internet connections reaches more and more homes and the efficiency of data transfer increases, so the Internet will subsume all digital broadcasting mediums. It will also transmit household and business utility readings, convey automated dairy produce replenishment requests initiated by intelligent fridges triggered by microchip-embedded product sell-by dates, blow-away conventional phone traffic and, once wireless or wi-fi hotspots proliferate, eliminate cell phone congestion.

Multiplayer real-time online games have recently taken off in a huge way now that broadband reception plus video compression techniques have reached a maturing plateau. Entire conceptual universes sprinkled with strange and wonderful planets populated by alien life forms occupy tens of thousands of people every month as they battle to win tokens, weapons and supremacy of these surreal landscapes.

The Politics of the Internet

But the enthusiasm and capability to deliver such technology, now practicable, may be blunted by the politics of the Internet. The United States effectively controls the Internet through ICANN who administer the IP addresses and root name servers for domains. Requests for an international domain governing body made at a three-day UN World Summit on the Information Society in November 2005 in Tunis went unresolved save for an agreement to form an Internet Governance Forum. There is no doubt the United States is reluctant to relinquish control of such a powerful medium as the Internet and perhaps distrustful of the competency of a mixed international body to administrate it efficiently or securely.

The coming decade will no doubt usher in the next generation of both ultra high-speed communications and the software and digital media able to exploit it. But for some it is a chilling thought that the Internet will become the communications platform of the world. Why? Because even though it is becoming ever more speedy and reliable, it is also prone to attack. Not as envisaged in the Cold War years from nuclear missiles whose electromagnetic pulses would render the router controlling microchips dead but by attacks from within, attacks by hackers, criminals or terrorist groups intent on crippling it.

The Internet does have an innate redundancy but the very speed with which files may be transmitted means it can take just milliseconds for thousands of computers to be infected by a virus. Firewalls and sophisticated anti-proliferation contingency firmware in routers go only so far to preventing such attacks.

Internet Identities

Perhaps the introduction of unique internet identities assigned to individuals would go some way to thwarting cyber criminals. Internet access would be granted once the user could be identified personally and all originating packet traffic for the session duration would hold their encrypted signature in much the same way as the TCP/IP envelope contains addressing information. Attempts

The Internet Explained

to commit malicious acts might then be traced to the individual and not the originating computer, and act as a deterrent. How contentious an issue internet identities is depends on an individual's stance on civil liberties; how effective it is depends on the ingenuity of implementation, ease of use ... and, as ever, human fallibility.