

<sup>1</sup><http://en.kioskea.net/contents/internet/ppp.php3>

## What is the Internet?

When IT began computers were developed, when they were capable of operating alone, some people had the idea of linking them in order that they could exchange data; this is the concept of a network. Therefore, it was not only necessary to develop physical links between computers so that information could circulate, but also a communication language so that they could have a real exchange; it was decided to name this language: a protocol.

Over the Internet, many protocols are used; they are part of a series of protocols which are called TCP/IP. TCP/IP is based on tagging each computer with an address called an IP address which makes it possible to convey data to the right address. Then these addresses were linked to domain names so that they could be remembered more easily.

Heterogeneous networks (of different kinds) were developed in the four corners of the globe; some people then decided to link these networks (universities or the army for example). Protocols were developed to allow all these networks to communicate and form the network of networks, little by little forming a giant spiders "web" then making the largest network containing all the networks which they called the **Internet!** There are different protocols on the internet (languages between computers) which enable different things to be done:

- IRC: live chat
- HTTP: look at web pages
- FTP: transfer files
- and many other things

Each of them is assigned a number (the port) which is sent during communication (transmission is carried out by small packets of information). It is therefore possible to know which program each small packet relates to:

- HTTP packets arrive on port 80 and are transmitted to the internet browser from which the page was requested
- IRC packets arrive on port 6667 (or another generally located around 7000) and are transmitted to an IRC client such as mIRC (or other)

## Connecting to the Internet

The network interface card is the part of the computer which makes it possible to connect to a network via specially provided lines for sending digital information. The modem makes it possible to connect to a network using telephone lines, which were not originally provided for this but which remain the most commonly used means of communication.

An IP address is associated to the network interface card making it possible to identify the computer on the network.

Connecting using a modem is totally different, a modem makes it possible to establish communication between two computers using a telephone line. You can however have access to a network (and thus by extension to the Internet) by contacting a computer linked ("on one side") to

one or several telephone lines (to receive the call) and ("on the other side") to a network using a network interface card.

This computer generally belongs to your Internet Service Provider (ISP). When it connects you via its intermediary, it borrows an IP address that the computer will retain during the connection. Each time it connects it arbitrarily allocates one of the free IP addresses that it has. If it is able to supply the same IP address for each connection, it is then called a "fixed IP address".

### **What is a protocol?**

A **protocol** is a standard method which enables communication between processes (potentially running on different machines), i.e. a collection of rules and procedures to be observed for issuing and receiving data over a network. There are several protocols according to how the communication is expected. Some protocols for example will specialise in the exchange of files (FTP), others may be used simply to manage the status of transmission and errors (as is the case with ICMP), ...

On the Internet, the protocols used belong to a suite of protocols, or a collection of linked protocols. This suite of protocols is called TCP/IP.

Among others, it contains the following protocols:

- HTTP
- FTP
- ARP
- ICMP
- IP
- TCP
- UDP
- SMTP
- Telnet
- NNTP

### **Connection oriented and connectionless protocols**

Generally protocols are classed in two categories depending on the level of data monitoring required:

- **Connection oriented protocols:** These protocols operate data transmission monitoring **during** a communication established between two machines. In such a schema, the recipient machine sends delivery acknowledgements during communication, so the originator machine is responsible for the validity of data that it sends. Data is therefore sent in the form of data flow. TCP is a connection oriented protocol
- **Connectionless protocols:** This is a communication method in which the originator machine sends data without warning the recipient machine, and the recipient machine receives the data without sending a delivery notification to the originator. Data is therefore sent in the form of blocks (datagrams). UDP is a connectionless protocol

### **Protocol and implementation**

A protocol uniquely defines the way in which machines must communicate, i.e. the format and sequence of data to be exchanged. Conversely, a protocol does not define the way that software is

programmed in such a way that it is compatible with the protocol. This is called **implementation** or the translation of a protocol into a programming language.

Protocol specifications are never exhaustive; also it is usual that implementations are subject to a certain interpretation of the specifications, which sometimes leads to the specificities of certain implementations or worse to incompatibility or flaws in security!

### What is an IP address?

Computers communicate over the Internet using the IP protocol (*Internet Protocol*), which uses numerical addresses, called **IP addresses**, made up of four whole numbers (4 bytes) between 0 and 255 and written in the format xxx.xxx.xxx.xxx. For example, *194.153.205.26* is an IP address given in technical format.

These addresses are used by networked computers to communicate, so each computer on a network has a unique IP address on that network.

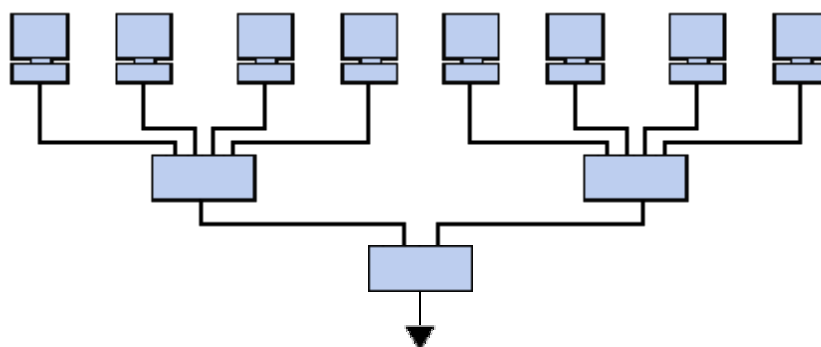
It is ICANN (*Internet Corporation for Assigned Names and Numbers*, replaced since 1998 by IANA, *Internet Assigned Numbers Agency*) which is responsible for allocating public IP addresses, i.e. IP addresses for computers directly connected to the public internet network.

### Decrypting an IP address

An **IP address** is a 32 bit address, generally written in the format of 4 whole numbers separated by dots. There are two distinct parts to an IP address:

- the numbers to the left indicate the network and are called the **netID**,
- the numbers to the right indicate the computers on this network and are called the **host-ID**.

Shown in the example below:



Note the network to the left *194.28.12.0*. It contains the following computers:

- 194.28.12.1 to 194.28.12.4

Note that of the right *178.12.0.0*. It includes the following computers:

- 178.12.77.1 to 178.12.77.6

In the case above, the networks are written *194.28.12* and *178.12.77*, then each computer making up the network is numbered incrementally.

### What is DNS?

Each computer directly connected to the Internet has at least one specific IP address. However, users do not want to work with numerical addresses such as *194.153.205.26* but with a domain name or more specifically addresses (called FQDN addresses) such as *www.commentcamarche.net*.

It is possible to associate names in normal language with numerical addresses thanks to a system called **DNS** (*Domain Name System*).

This correlation between the IP addresses and the associated domain name is called *domain name resolution* (or *address resolution*).

### What is a URL?

A **URL** (*Uniform Resource Locator*) is a universal naming format used to indicate a resource on the Internet. It is a printable ASCII character string which breaks down into five parts:

- **The name of the protocol**: i.e. in a way the language used to communicate over the network. The most widely used protocol is the HTTP protocol (*HyperText Transfer Protocol*), the protocol which makes it possible to change Web pages into HTML format. However, many other protocols can be used (FTP, News, Mailto, Gopher, ...).
- **Login and password**: enables the access parameters for a secure server to be specified. This option is unadvisable because the password is visible in the URL
- **The name of the server**: This is a domain name for the computer hosting the requested resource. It is worth noting that it is possible to use the server's IP address which conversely makes the URL less readable.
- **The number of the port**: this is a number related to a service allowing the server to know what type of resource is requested. The default port related to the protocol is port 80. So, when the Web service of the server is associated to port number 80, the port number is optional
- **The access path to the resource**: This last part allows the server to know where the resource is located, i.e. generally the site (directory) and the name of the file requested.

A URL therefore has the following structure:

<b>Protocol</b>	<b>Password (optional)</b>	<b>Server name</b>	<b>Port (optional if 80)</b>	<b>Path</b>
http://	user:password@	www.commentcamarche.net.	:80	/glossair/glossair.php3

For example, the following protocols may be used through URL:

### What does TCP/IP mean?

**TCP/IP** is a suite of protocols. The acronym TCP/IP means "**Transmission Control Protocol/Internet Protocol**" and is pronounced "T-C-P-I-P". It comes from the names of the two major protocols in the suite of protocols, i.e. the TCP and IP protocols).

In some ways, TCP/IP represents all communication rules for the internet and is based on the IP addressing notion, i.e. the idea of providing an IP address for each machine on the network so as to be able to route data packets. Given that the TCP/IP protocol suite was originally created with a military purpose, it is designed to respond to a certain number of criteria, including:

- Splitting messages into packets;
- Use of an address system;
- Routing data over the network;
- Error detection in data transmissions.

### The difference between standard and implementation

Overall TCP/IP brings together two notions:

- The notion of **standard**: TCP/IP represents the way in which communications are carried out over a network.
- The notion of **implementation**: the designation TCP/IP is often extended to software based on the TCP/IP protocol. TCP/IP is in fact a model which network application developers use. The applications are therefore implementations of the TCP/IP protocol.

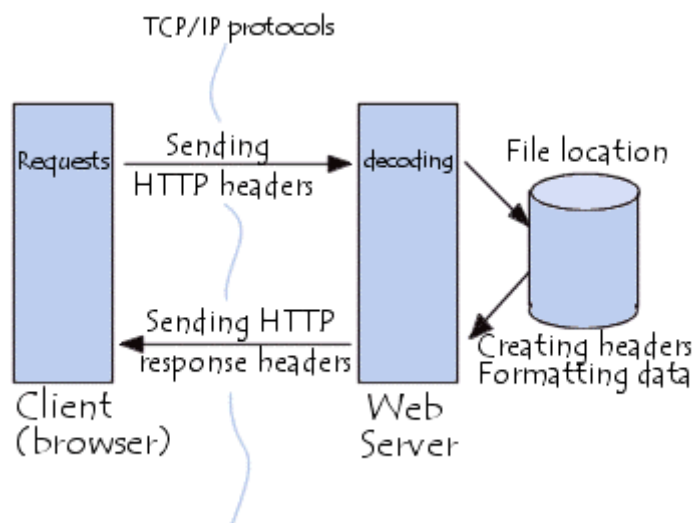
### Introduction to the HTTP protocol

Since 1990 HTTP protocol (HyperText Transfer Protocol) has been the most widely used protocol on the Internet. Version 0.9 was only intended to transfer data over the Internet (in particular Web pages written in HTML). Version 1.0 of the protocol (the most used) now allows the transfer of messages with headers describing the content of the message by using MIME type coding.

The aim of HTTP protocol is to allow transfer of files (essentially in HTML format) between a browser (the client) and a Web server (called among other things *httpd* on UNIX machines) located using a character string called a URL.

### Communication between browser and server

Communication between the browser and server takes place in two stages:



- The navigator makes a **HTTP request**
- The server processes the request then sends a **HTTP response**

In reality, the communication is conducted in more stages if you consider the processing of the request by the server. Given that we are only concerned with HTTP protocol, server side processing will not be explained as part of this article.

### What is a network?

The generic term "**network**" refers to a group of entities (objects, people, etc.) which are connected to one another. A network, therefore, allows material or immaterial elements to be circulated among all of these entities, based on well-defined rules.

- **network**: A group of computers and peripheral devices connected to each other. Note that the smallest possible network is two computers connected together.
- **networking**: Implementing tools and tasks for linking computers so that they can share resources over the network.

Depending on what kind of entity is involved, the term used will differ:

- **transportation network**: A combination of infrastructure and vehicles used for transporting people and goods between different geographic areas.
- **telephone network**: Infrastructure for transporting voice signals from one telephone station to another.
- **neural network**: A group of brain cells connected to each other
- **criminal network**: A group of con artists in cahoots (wherever there's one con artist, there's usually another!)
- **computer network**: A group of computers linked to each other with physical lines, exchanging information as digital data (binary values, i.e. values encoded as a signal which may represent either 0 or 1)

### What does "topology" mean?

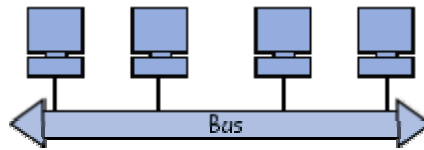
A computer network is made of computers which are linked to one another with communication lines (network cables, etc.) and hardware elements (network adapters, as well as other equipment for ensuring that data travels correctly). The physical arrangement — that is, the spatial configuration of the network — is called the **physical topology**. The different kinds of topology are:

- Bus topology
- Star topology
- Ring topology
- Tree topology
- Mesh topology

The **logical topology**, as opposed to the physical topology, refers to way that data travels along communication lines. The most common logical topologies are Ethernet, Token Ring and FDDI.

### Bus topology

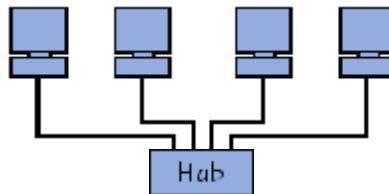
Bus topology is the simplest way a network can be organised. In bus topology, all computers are linked to the same transmission line by using a cable, usually coaxial. The word "bus" refers to the physical line that joins all the machines on the network.



The advantages of this topology are that it is easy to implement and functions easily; on the other hand, it is highly vulnerable, since if one of the connections is defective, the whole network is affected.

### Star topology

In **star topology**, the network computers are linked to a piece of hardware called a **hub**. This is a box which contains a certain number of sockets into which cables coming out of the computers can be plugged. Its role is to ensure communications between those sockets.

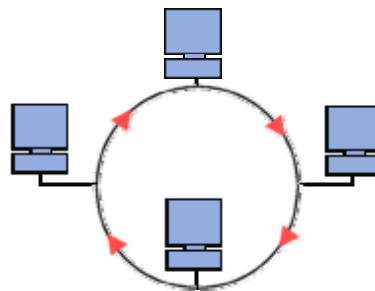


Unlike networks built with bus topology, networks which use star topology are much less vulnerable, as one of the connections can easily be removed by disconnecting it from the hub, without paralysing the rest of the network. The critical point in this network is the hub, as without it, communication between the computers on the network is no longer possible.

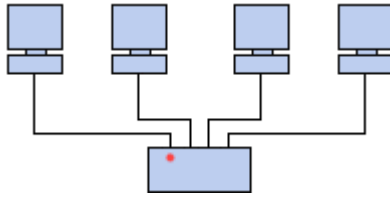
However, a star topology network is bulkier than a bus network, as additional hardware is required (the hub).

### Ring topology

In a **ring-topology** network, computers each take turns communicating, creating a loop of computers in which they each "have their turn to speak" one after another.



In reality, ring topology networks are not linked together in loops. They are actually linked to a distributor (called a MAU, **Multistation Access Unit**) which manages communication between the computers linked to it, by giving each of them time to "speak."



The two main logical topologies which use this physical topology are Token ring and FDDI.

### **Different types of networks**

Different types of (private) networks are distinguished based on their size (in terms of the number of machines), their data transfer speed, and their reach. Private networks are networks that belong to a single organisation. There are usually said to be three categories of networks:

- LAN (local area network)
- MAN (metropolitan area network)
- WAN (wide area network)

There are two other types of networks: TANs (Tiny Area Network), which are the same as LANs but smaller (2 to 3 machines), and CANs (Campus Area Networks), which are the same as MANs (with bandwidth limited between each of the network's LANs).

### **LAN**

*LAN* stands for *Local Area Network*. It's a group of computers which all belong to the same organisation, and which are linked within a small geographic area using a network, and often the same technology (the most widespread being Ethernet).

A local area network is a network in its simplest form. Data transfer speeds over a local area network can reach up to 10 Mbps (such as for an Ethernet network) and 1 Gbps (as with FDDI or Gigabit Ethernet). A local area network can reach as many as 100, or even 1000, users.

By expanding the definition of a LAN to the services that it provides, two different operating modes can be defined:

- In a "peer-to-peer" network, in which communication is carried out from one computer to another, without a central computer, and where each computer has the same role.
- in a "client/server" environment, in which a central computer provides network services to users.

with distance) and may be low.

WANs operate using routers, which can "choose" the most appropriate path for data to take to reach a network node.

The most well-known WAN is the Internet.