Introduction to COMPUTER NETWORKING

The fundamentals Guide for beginners

Peter Aggarwal

Peter Aggarwal

Introduction to Computer Networking

The fundamentals Guide for beginners

Part I

The basic elements of the networks

The networks have the function to carry information in order to achieve of the services that can be found anywhere on the globe. A series of hardware equipment and software process are implemented to ensure this transport, since the terrestrial cables or radio waves in which circulate the data up to the protocols and rules to deal with them.

This first part of the book recalls the principles of operation of networks and presents in detail the hardware, software and protocol architectures on which they are based.

1

Introduction to Networks

The networks are born of the need to transport data from one computer to another computer. Since these data are put in the form of files, the base application of networks is called the transfer of files. A little later, the "transactional" appeared to allow a user to carry out transactions with a remote computer, for example to book a place of aircraft. It was called a session the whole of the transactions of a same user to perform a given task.

With the development of the Web, the transactional service has diversified in order to allow the search of information through links. These applications are called client-server, that is to say that a customer is directed to a server to obtain the information.

The next step of networks has been characterized by the peer-to-peer, or P2P (peer-to-peer), in which all the components connected to the network are equivalent and can be distributed in the network. The underlying applications are in fact very numerous, ranging from the telephony at the research of various information and varied, such as audio or video files on the Internet.

Without supplanting the applications for the transfer of files, that they are client-server or peer-to-peer, the new Internet service that develops from the years 2010 is the *cloud, or "cloud". Until the arrival of the Clouds, the Internet had the objective to carry data to achieve a remote service. The companies allowed for example mobile employees to connect to their servers through the Internet. They had to do this all the necessary elements, such as email, business applications or servers archiving, as well as the power of calculation required. Today, it is possible to achieve in the Cloud which was previously within the company: Calculation, storage, business application, messaging, telephony, etc. The advantages are many: the customer can access these services from anywhere; these can be secured by the redundancy; it is possible to instantly add new services, the computing power, storage space, etc., depending on the needs and by paying only what is used.*

This new generation is implementing the concept of virtualization, by which the resources which the company or the particular need can be found anywhere, or even move in function of the cost of the servers.

Before detailing more before these new generations of the networks, the sections that follow recall a few key elements of the evolution of networks and current architectures.

The packet

The base Entity of the networks is the packet. The latter brings together binary elements, suites of 0 and 1 corresponding to data from different types of information, such as the floor and the video, or of storage, calculations and applications. The packages contain in general between a few bytes (8 bits) of these binary elements and 1 500 bytes.

The objective of the networks is to transport the packets from a user, machine, of an object or

everything that can produce the information. In other words, a transmitter creates packets and sends the packets up to a receiver. As there is little chance that there is a direct line between the transmitter and the receiver, the packets pass through intermediate nodes that the transfer to the next node and so on up to the receiver.
This book details the different ways to perform the transport of packets of a transmitter to a receiver, such as Always follow the same path, represented by a succession of intermediate nodes, or take different routes depending on the nature of the packets to carry.
The first solution is to mark a path (path) and to send the packets. The packets will follow and arrive in the order of their issuance. In this case, even if it proves that another path is more short or request less time, the packages are still on the path that has been opened as long as it meets the quality of service required by the issuer. This solution of path has given birth to the so-called technical switching (switching), in which the nodes are called switches).
The second great solution is that of routing (Routing), in which the packets are routed to a node that can change with time in the function of the state of the network and trying at any moment to take the shortest route. The node that performs the referral is called a router. It has for this a routing table. In this case, packets can arrive in the disorder to the receiver.
The nodes that transfer the packets of a line of input to a line of output are called Transfer Nodes (forwarding nodes). They have long been established material elements, with little software to the interior to manage the passage of packets and detect, for example, of the anomalies. In the new generation of transfer nodes which is in the setup phase, physical nodes are replaced by nodes software, also called virtual nodes.
Go to a hardware architecture to a software architecture is fairly simple. It is sufficient to write a code in a language determined in order to describe and realize what a physical machine. This code is called a virtual machine or virtual machine (VM).
To run the virtual machines, physical infrastructure is obviously necessary. The latter must have enough computing power for offer the same performance that a transfer node hardware. Such a power is only available in data centers, or data centers, which bring together a large number of servers.
The new generation of networks which is put in place today is characterized by the whole of these elements: data centers that integrate transfer nodes and which are interconnected by links to very high speed, in general in optical fiber. The following chapter describes environments who deploy in these data centers: the cloud for the biggest of them, the MEC (Mobile Edge Computing) for those of average size, the fog for small and the skin for the very small.

Cloud, CME, fog and skin

A cloud is made possible by the exploitation by a network of the computing power and storage of a datacenter. In other words, it is the set of virtual machines deployed in one or multiple data centers with a view to meet the needs of users. The virtual machines, or VM, can be stored in three major categories: the VM storage, the vm of calculation and the VM network. They can add two new types of machines appeared on the market in 2017: The VM security and the VM management and control.
A VM of calculation is a set of hardware and software aggregating the computing power of one or more processors and memory to perform calculations that can prove to be extremely important. A VM of storage is essentially composed of a collection of memories to store the data. A VM network is an equipment virtual network, such as a router or a virtual switch, which is stored in the memory of a datacenter and thus has the power of calculation associated, which may vary in time. A VM of security can act as the authentication server virtual or virtual firewall in order to manage the security

of clients, of machinery or objects. Finally, a VM management and control may be office of Comptroller of or virtual Orchestrator to control the flow of packets or to put in place a set of virtual machines to deploy a service.
The Cloud generally benefits from the power of computing and storage of a large datacenter, possessing several hundreds or thousands of servers, or even up to a million servers, gathered in the same place. The current trend is however to install these data centers to more close to the user in order to offer the best response time possible. Closer to the user, these data centers serve less and less of clients and their size decreases. The Cloud in resulting is then called the Fog, that is to say a "fog" that surrounds the user very closely.
Two other environments can be defined: the skin, or skin, that is to say a cloud which is located at the contact of the user, at most a few meters. Called the corresponding datacenters of FEMTO-data center, or of home-datacenters, or even wall-datacenters. The size of these datacenters is that of a few servers, but with a great power of storage. The first generation of this type of submissions was represented by the NAS (Network Attached Storage), which were of the file servers autonomous. A femto-datacenter must also have a very strong capacity of calculation as well as a specific operating system supporting virtual machines. Chapter 3 describes in detail how to Monte such a virtualized environment.
Another environment of the cloud of average size is the MEC (Mobile Edge Computing). The latter covers a set of services provided by the data center average size that operators put essentially near the edge of networks, i.e. behind the large antennas relay 3G/4G. The MEC-data centers serve all customers located behind an antenna 3G or 4G and soon 5G. The Environment Guy is presented in Chapter 13. Being specific to the telecommunications operators, these data centers are not included in the series cloud, fog and skin.
To summarize, we can consolidate the different kinds of data centers in four broad categories, which define the four major architectures of the new generation of networks:

- Cloud, to the powers of calculation and storage very important, capable of simultaneously manage more than ten thousand users and can manage millions.
- MEC, able to manage from one thousand to ten thousand users.
- Fog, able to manage fifty to one thousand users.
- Skin, for less than fifty users.

These values are obviously related since the definitions of these different environments are not standardized. These classes of Datacenters however represent the major trends of the years 2020.

The environments of the Cloud

Cloud is a generic word designating all environments capable of exploiting of the calculation and the storage, regardless of the size of the data centers concerned.
A cloud can offer a very large number of services through the virtual machines that make it up. The corresponding environments are grouped in three main groups:

- Infrastructure as a Service (IaaS);
- PaaS (Platform as a Service);
- Software as a Service (SAAS).

The hierarchy of these environments is described in Figure 1.1.

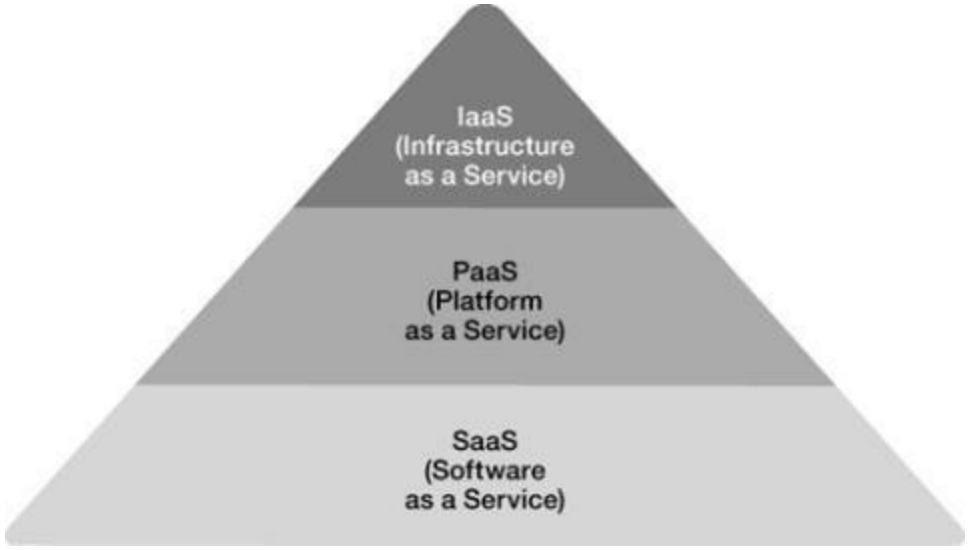


Figure 1.1

The three large environments of the Cloud

The features of the main environments of the Cloud are illustrated in Figure 1.2.

In the classic model, used before the arrival of the clouds, each user company should manage itself the whole computer system, since the network and the applications until the storage, in passing by the physical machines, the operating system, databases and possibly the virtualization, if it was already introduced in the system.

The IaaS allows the user to subcontract to the provider of the cloud the low part of the Environment, that is to say the network, storage, infrastructure, hardware and virtualization. The customer retains the operating system, the management of data and applications. In the case of a PaaS, the client sub-processes a little more to the provider of the Cloud, leaving him also the charge of the operating system and data. With the SaaS, the user sub-treats the whole of the system to the provider of the cloud, including its applications.

Regarding this which is of interest to this book, if each of the three environments has recourse to the cloud, the IaaS is the minimum version to treat the network portion.

Other environments, such as SECaaS (security as a Service), NaaS (Networking as a Service), XaaS (anything as a Service), etc., allow suppliers of cloud to provide specific services, such as security, of the network or full virtualization.

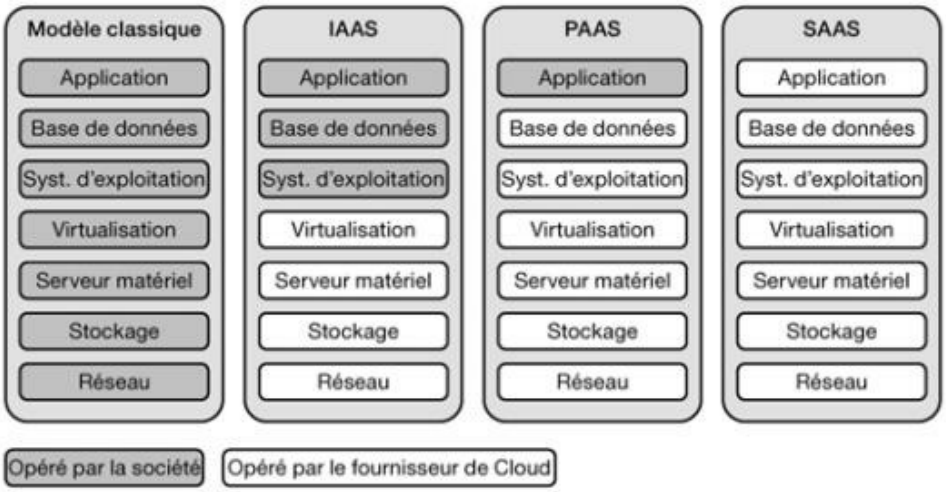


Figure 1.2

Features of the main environments of cloud providers

In summary, network environments of new generation are composed of data centers, large, medium, small or very small, that are running virtual machines in the guise of routers or switches. These datacenters are connected between them by the communication channels with very high flow, which

may be of the optical fiber in the major centers or microwave links to the periphery.
We can represent this new generation of networks in the form shown in figure 1.3, where the machines that transfer the packets are implemented in datacenters and where these datacenters are connected in a network. These are all the more small that they are toward the periphery. These networks are part of what is called the cloud networking.

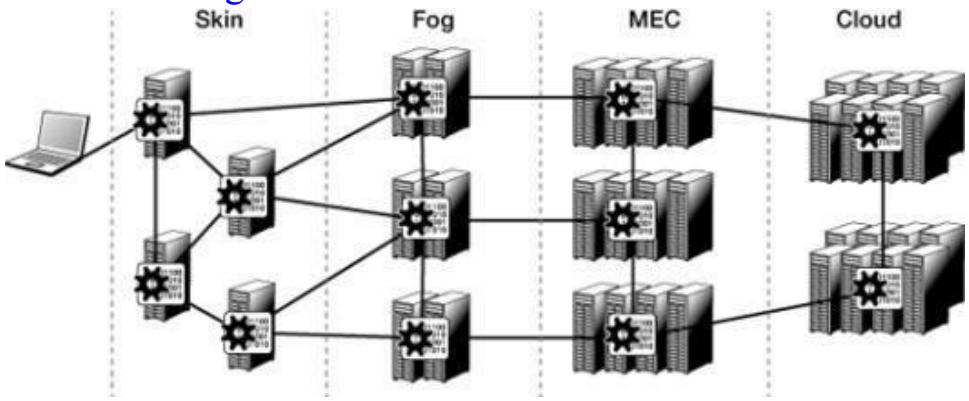


Figure 1.3
Architecture of networks on the basis of data centers

The new network architectures

From the general architecture described in Figure 1.3, one can deduce the four major network architectures that are in place today.
The first concerns cloud providers (Google, Amazon, Microsoft, Apple, etc.) who wish to manage the networks from very large central data centers.
In this architecture, the whole periphery with customers must send the data to the center, which then treats the whole of applications, either those of the clients or those associated with the control of the network.
In this case, the signal emitted by the user goes back up to the data center by the intermediary of a intermediate antenna or uses to do this the access network. In the first case, the signal goes back directly to the datacenter, and it is no longer necessary to put the signals into packets. This solution is called cloud-RAN (Radio Access Network), or C-ran. It is highly centralized and begins to deploy in developing countries due to its relatively low costs since there is an infrastructure in a star around the datacenter. All virtual machines are collected in the datacenter, which offers the best possible use of the resources of the datacenter.
This architecture is illustrated in Figure 1.4. It is examined in detail in Chapter 13, in the framework of the cloud networking.

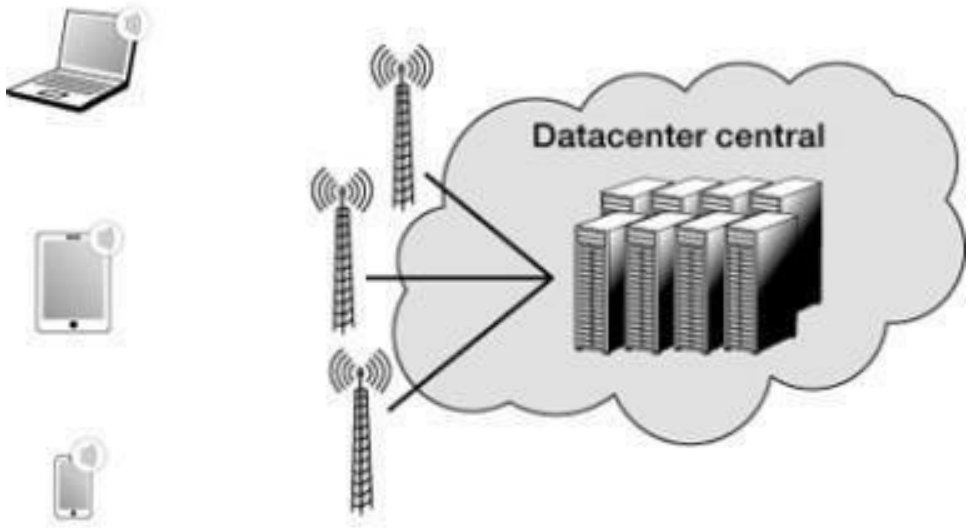


Figure 1.4
Cloud Architecture-ran of service providers

The second architecture is to distribute the central datacenter in datacenters MEC (Mobile Edge Computing) of smaller size. This solution is illustrated in Figure 1.5.

A datacenter guy takes in charge of centralized way any the periphery. In this solution, all equipment located between the client and the datacenter MEC are virtualized. For example, the Internet box disappears to become a simple VM in a server in the datacenter CME. The advantage of such an architecture is to provide a reaction time better than in the case of centralised the cloud architecture-ran.

This solution to the favor of telecommunications operators, who hope as well control and manage the whole periphery from their data centers. These must be connected between them to achieve the network heart, that is to say the central network allowing the interconnection of customers between them when they are not located in the area defined by the datacenter.

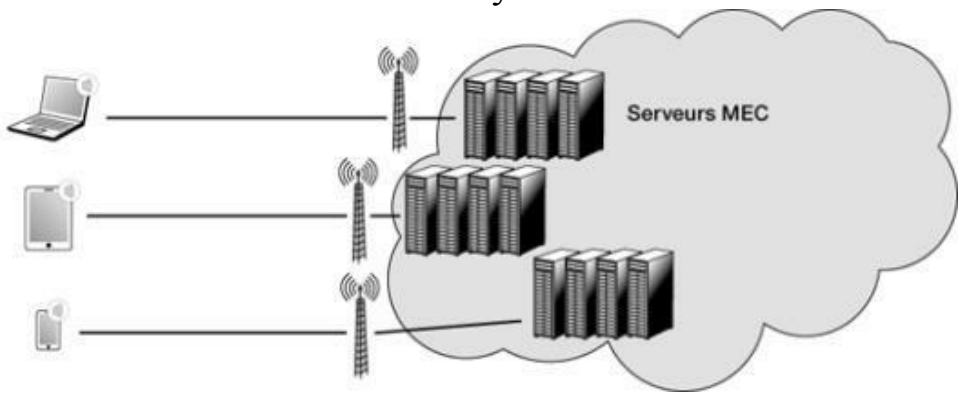


Figure 1.5
Cloud Architecture-MEC of operators

The third architecture corresponds to the needs of OEM network. This solution is more recent than the previous two and is being put in place since 2017 only. The OEMS network have lost much of their power because of the rise in power of the architectures described above, in which the decisions come from the datacenters, as much of the equipment which they do not control, at least with regard to the largest of them.

The idea of this solution is to restore primacy to network devices of type router or switch. However, as they are virtualized, it is necessary to integrate in any small data centers of the size of a network equipment classic.

This infrastructure generic hardware receiving of virtual machines is called FOG to indicate its proximity with the customer. In other words, a router is replaced by a physical machine to receive

virtual machines and, on the inside of the latter, a router, or a virtual switch. Of this fact, such a machine can be easily replaced by another or be supplemented by virtual machines, such as a firewall for security or a storage server or a server allowing to put in place a messaging service, for example.
This type of architecture is reconciled with the protocols that have been used for a long time in the framework of the Internet, which offers an interesting solution of continuity. In addition, this solution adds a central machine, always called a controller, in order to assist in the control and management of the network.
This central machine receives all measurements on the network. These measures, or "knowledge", are information contextualized from physical and virtual machines constituting the network. Such a controller has a complete view of all the users connected and the network to interconnect them. It therefore has a power of important control for the detection of anomalies thanks to his knowledge of everything that is happening in the network.
As soon as necessary, the Controller may take the relay of the routers and switches to control the periphery. It then forwards it to the virtual machines devices The power of control as soon as there is more danger or that the network returns to normal operation. This architecture is illustrated in Figure 1.6.

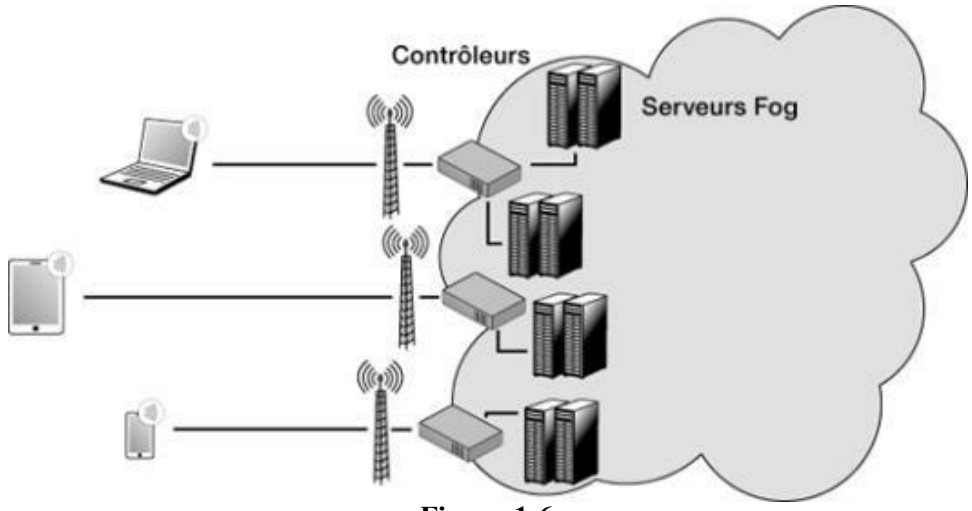


Figure 1.6
The Fog architecture of network equipment manufacturers

The fourth architecture of network of new generation can be regarded as a ubérisation of telecommunications. Indeed, the network is constructed from femto-Data Centers who are found among users and non-more in the area of the operator. The femto-datacenters are connected between them by Hertzian connections with very high flow to achieve mesh networks, called mesh.
The limits of such an architecture are derived from the support necessary to a operator of infrastructure to perform the maintenance and management of the network. As there is more of a telecommunications operator to speak of, the network ubérisé is no longer maintained and therefore cannot resolve the potential problems resulting from, for example, of failures or malfunctions. To install this architecture, it is prudent to await the emergence of solutions of self-control, self-healing, self-management, autosécurité, etc.
The ubérisée architecture of telecommunications is illustrated in Figure 1.7. On this figure, each home has a femto-datacenter still called home datacenter or home server or wall-datacenter. These femto-datacenters are interconnected to form a network mesh.

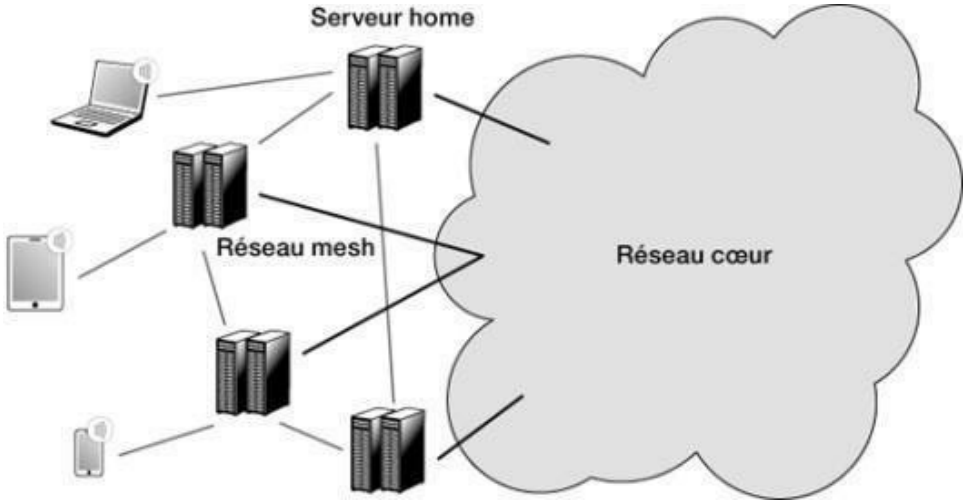


Figure 1.7
Ubérisée architecture of Telecommunications

These four solutions to the protocols and features at any point different are described in more detail in chapters dedicated. However, there are great chances that the final architecture of the World of Networks consists of a superposition of these architectures according to the size of the datacenter.

The Wireless Networks

The emergence of the technology package in the networks of mobile and wireless networks date of the beginning of the years 2000 with the UMTS (Universal Mobile Telecommunications System). The generation of before, GSM (Global System for Mobile Communications), was based on the circuit, i.e. a set of resources belonging only to the transmitter and the receiver and that no other person could use.

The circuit is different from the package in that the resources are assigned to the package that during its passage in a Transfer node.

The networks of mobile allow communication while moving. In the Wireless Networks, the communication is done through a antenna, with the result that the user must remain connected to a same antenna. The first radio networks had only flows very limited, but they quickly reached the performance almost identical to those of the terrestrial networks, at least on the metal pairs. The 5G will offer the flows almost identical to the optical fiber as early as 2020.

The radio networks are grouped into two categories, one from industrialists of telecommunications - and therefore of the Switching -, with a significant signalling and a high complexity to support all the problems of communication and the other from the Internet - and therefore of the routing -, with much less complexity, but an overall quality lower.

The fourth generation of networks of mobile, the 4G has become totally compatible with the Internet. It reaches the flows identical to those of the ADSL and is no longer far from the access optical fiber.

A standardization is therefore outline between the worlds of wired and wireless, so that a same network heart, the fiber optic network central, allows you to connect a fixed terminal as well as a mobile terminal. This convergence through a network single heart is called NGN (Next Generation Network). It is described in more detail in Chapter 16.

These wired or wireless networks are multimedia. A multimedia application uses at the same time the animated image, speech, data and various forms of assistance.

The characteristics of this convergence are the following:

- Flow rates are very important in the network core, in particular because of the increase of the power of the Terminal machines and the flow of each client to the

network core.
- The quality of service to achieve the constraints of each application.
- Securing the transport.
- The management of mobility and the connection to multiple networks simultaneously (multihoming).
- The virtualization of all network resources.

The 5G, which is studied in detail in Chapter 18, should be standardized to the end of 2020, and the first products appear on the market as soon as 2021. There are of course of pre-5G, which consist of strong proposals intended to influence the standardization. Many operators have announced the output of such proposals during the holding of the next Olympic Games in winter and summer.
The 5G stands on three feet:

- The massive connection objects, which should rise up to a 50 billion in 2020.
- The flows much more important in high mobility.
- The possibility of satisfying the applications requesting reaction time extremely short, that is called "mission critical".

The network architectures: distribution or centralisation?

The network architectures were up to this distributed. From 2015, the arrival of the Cloud has jumbled The gives the benefit of a centralization of services in data centers containing hundreds or even thousands of servers. At the beginning of 2018, the largest data centers exceed the million servers. These centers have powers of considerable calculation and gigantic memories.
The idea of this new generation of network architectures is to consolidate centrally all information available on the network, either on the part of the applications or infrastructure, and any check also centrally. The benefits are that the Center may accumulate a multitude of information which could not be treated simply to a distributed fashion.
In such architectures, this is the center which decides everything, in particular of the path or of the road to be followed by the packets. The choice of the path, which is the case the most classic of the centralized solution, is facilitated by the complete vision of the network at the disposal of the Center, which may as well book the necessary resources along the path to obtain the quality of service desired.
The Center may also choose a routing solution. In this case, it calculates the routing table, it distributes to the nodes in the network. It must however recalculate the routing table as soon as the state of the network changes and distribute it to all the nodes. The switching solution with establishment of path is in this regard preferable, because it does not request regular distribution of tables. In addition, once the path chosen, it is not necessary to change, even if the paths best can be found by the result, since the quality of the service that has been required is always available.
Called SDN (Software-Defined Networking), this new centralized architecture has been standardized by the NFB (Open Networking Foundation).
We could translate SDN by network software (Chapter 3 details the differences between Hardware networks and networks software). But the SDN architecture has a center, what the expression software network does not leave suspect. The present edition shows however that there are just as many networks software distributed. The Acronym SDN is therefore reserved for networks that have a control center called controller. The controller has an interface with the applications, called Interface north, and an interface with the nodes of the network, called Interface South.

Conclusion

The network architectures have continued to evolve since their birth, at the end of the 1960s and at the beginning of the 1970s. These developments are accelerating with the arrival of the CLOUD and its derivatives CME, fog and skin. Until the beginning of the years 2020, the trend will be for the centralization of the control, but will evolve then certainly again toward a distribution.

The other impressive developments of the time affect the world open source, which has not yet been introduced, but which is a major movement detailed in different places of this book, including Chapter 14, and the intelligence to manage and control networks in a manner much more automated, which is also introduced throughout the book.

2

The components of the networks

After a first chapter devoted to ongoing developments that affect the architecture of networks, this chapter details the operation of the transfer of packets (packet forwarding) and introduced its major principles, which have changed little over the past fifty years. It goes as well to review the transfer techniques, the switching and routing, and compares the solutions put in place in the framework of the Internet networks, of telecommunications networks and networks of the cable operators.

Transfer, switching and routing

Modern networks have emerged during the 1960s in favor of a completely new technology allowing to carry information from one machine to another. These machines were then computers of first generation, significantly less powerful that a smartphone today. The networks for telephony existed about them for a long time. They used the technology called *circuit switching* and the support of the physical lines connecting the all of the phones through switches. During a communication, these physical lines could be used only by the two users in contact. The signal that y transited was type analog.

The first revolution of networks has been made by the digital technology of codecs (encoders-decoders), which permitted to transform analog signals into digital signals, that is to say a result of 0 and 1. The fact translate any type of information in the form of 0 and 1 allowed to unify the networks. In this generation, circuit switching was still heavily used. The circuits are become digital, the question was asked to move simultaneously on a same circuit several waves, corresponding to different applications. That is the way it has been able, for example, have 1 byte (8 bits) of telephony, followed by 2 bits of transfer of files and then 8 bits of video application. This solution does, however, is virtually not developed and has left the place at the transfer of packets.

The transfer of packets has allowed to take into account the strong irregularity of the flow of the communication between two computers, alternating periods of significant flow and periods of silence, resulting from the fact that, for example, a computer must wait for the response of another computer.

In the switching of the circuit, the circuit remains unused during the periods of silence, inducing a significant waste of resources. Conversely, the transfer of packets only uses the resources of the network when the actual issuance of the packets. The idea is therefore done day of constitute blocks of information of variable length and send Transfer node in transfer node up to reach the destination. The resources of a connection between two nodes are not of the kind used that during the transfer of packets. The different packages from a same user and a same application form a *stream.* Once the packets of this stream managed to destination, it is possible to use the same connection and the resources of the network for the passage of other packets, from other waves.

Among the many solutions for the transfer of packets that have been proposed, two have withstood the test of time, the routing and switching. In the routing of packets, the packets are referred by each node of transfer in function of their destination. The chosen route may vary depending on the state of the network, so that two packages of a same stream can follow a different route. The routing *tables* are implemented in the nodes in order to optimize the transport of packets depending on the state of the network.

Outcome of the world of telecommunications, the switching of packets is to put in place, before sending the lesser package, a path between the entities in communication, path that all packets of a same stream must borrow. This path (path) has long been called virtual circuit because packets using different paths can use the same resources. There is therefore no dedicated resource.

Each of these techniques has advantages and disadvantages. Routing is a flexible technique. To the extent that each packet carries the address of the recipient, the road can vary, without the risk that the packet will be lost. On the other hand, it is very difficult to ensure a quality of service, i.e. to ensure that the transportation service will be able to comply with a performance determined. With the packet switching, the quality of service is more easily achieved, since all packets follow the same path and that it is possible to reserve the resources or to determine by calculation if a given stream has the possibility to cross the network without hindrance.

The main weakness of packet switching lies in the establishment of the path that will follow the different packages of a stream. This path is opened by a specific procedure, called *signage* : it is reported to the network the opening of a path, which must in addition be "labeled" so that the packets of flow can follow. This signaling requires significant resources, which makes the networks to packet switching significantly more expensive that the networks to routing packets.

Figure 2.1 illustrates these two branches of the transfer of packets, the routing and switching, as well as the main techniques that they use.

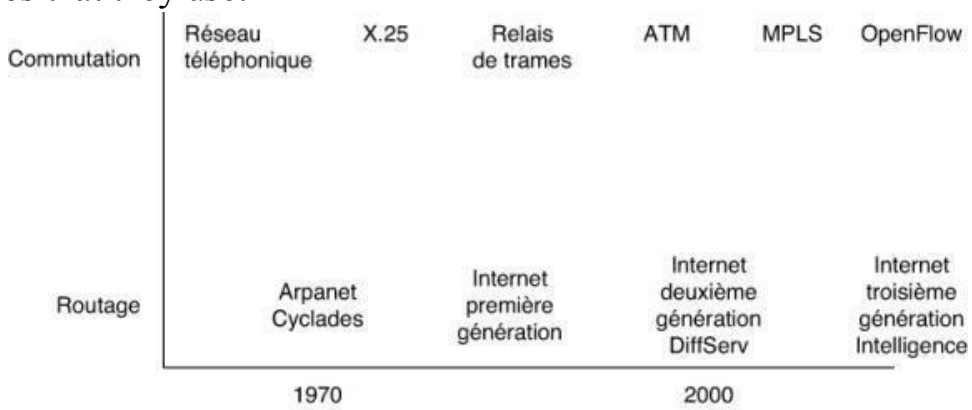


Figure 2.1

The two branches of the transfer of packets

These two categories of networks are developed in parallel. At the outset, there was very little competition between them, because they were directed to different worlds. With the time, the techniques for routing, linked to the Internet, are extended to the transport of synchronous applications such as telephony and the video. In parallel, the switching of packets in supported telephony and television. Today, both are in competition for the transport of multimedia applications. Their respective advantages and disadvantages would tend to do choose the switching of packets by the operators and the very large companies and the routing by small and medium-sized enterprises.

The techniques for routing have little changed. The Internet Protocol (IP) is the main deployment: The IP packet containing the full address of the recipient is routed in transfer nodes called routers.

Conversely, the protocols related to the switching has evolved a lot. The first major standard switching, X.25, has seen the light of day in the years 1980. This solution required important

operations to perform the switching: The path was mapped in the network by a set of indices, called *references, constituting as much "Colored Stones" on its entire length. The package had to do was to follow these stones up to destination. With the staggering increase in the number of waves, the "Colors" stones indicating the paths are become insufficient. A new signage has been introduced with the relay of frames and then with the technique Asynchronous Transfer Mode (ATM). Today, new solutions are being put in place with the technical SDN (Software-Defined Networking).*

Before going further, the sections that follow focus in more detail on the concept of package. A packet is not a block of data that can be send as is on a line of communication. For example, if the sending two packages glued to one another, the receiver is unable to distinguish between the end of the first package and the beginning of the second. To allow this operation of recognition, it must encapsulate each packet in a frame. The frame has a specific succession of binary elements allowing to recognize its beginning and end. To carry an IP packet, it can encapsulate it in a PPP frame (Point-to-Point Protocol); to carry a X.25 packet, it encapsulates the packet in a frame lap-B; to carry an IP packet in an Ethernet frame, it is necessary to add a suite long enough, named flag, containing a succession of 10 to finish at the end of 8 bytes per 11 (1010101010…11).

In the generations of the following networks, the full address of the recipient, or the reference, is postponed in the frame in order to simplify the recovery: it is of the so more necessary decapsulating the frame to retrieve the package and the information it contains. This solution, implementation in particular in Frame Relay and the technique MultiProtocol Label Switching (MPLS), greatly simplifies the work performed in the transfer nodes.

The transfer of packets

The technique used for the transport of data in digital form, i.e. in the form of 0 and 1, that it has adopted since the end of the 1960s called for the transfer of packets.

All the information to be transported are cut into packets to be routed from one end to another in the network. This technique is shown in Figure 2.2. The terminal equipment A wants to send a message to B. The message is divided into the three packages, which are issued from the terminal equipment to the first node in the network, which sends it to a second node, and so on, until they arrive at the terminal equipment B. In reality, an additional step is necessary: the encapsulation of the packet in a frame to achieve the transport on the lines of communication.

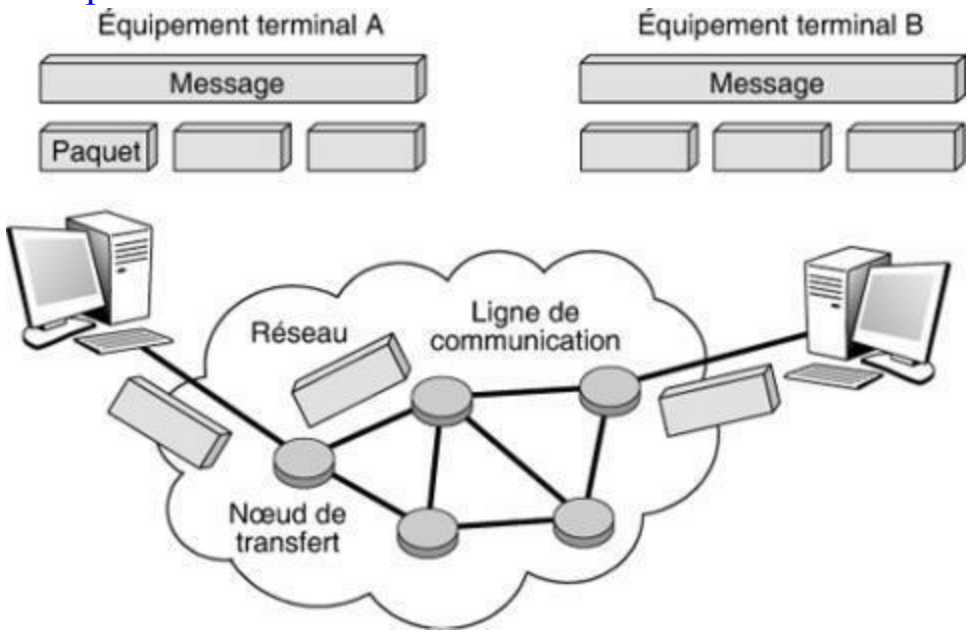


Figure 2.2
The transfer of packets

The packet can come from different sources. Figure 2.2 assumes that the source is a message prepared by the issuer, such as a page of text edited by means of a treatment of text. The term message is in fact much more vast and cuts across all forms in which the information may be present. This goes from a Web page to a stream of phone word representing a conversation.

In the word call, the information is grouped to be placed in a package, as shown in Figure 2.3. The telephone handset contains a equipment which transforms the analog speech in a suite of binary elements. These bits fulfill little by little the packet. As soon as the latter is full, it is sent to the recipient. Once the packet arrived at the terminal station, the reverse process is carried out, restoring the bits regularly from the packet to reconstitute the floor the phone.

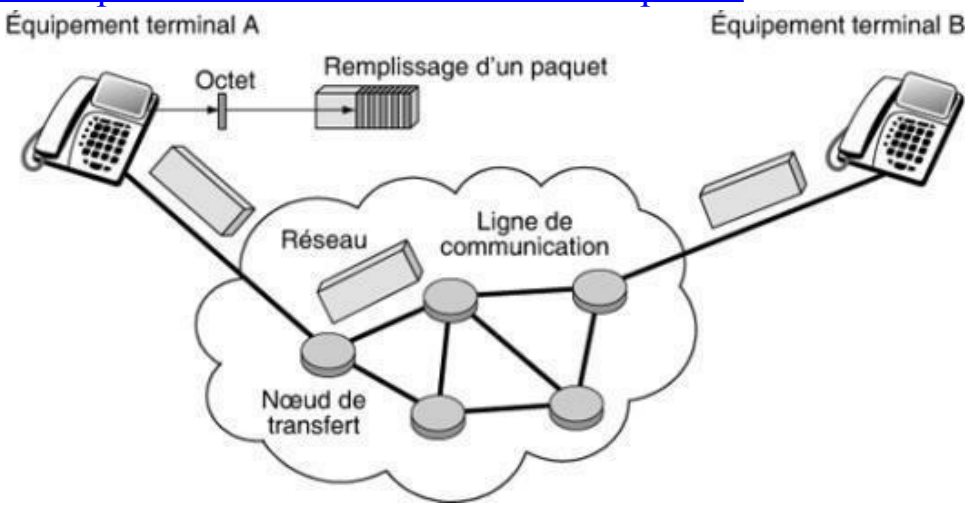


Figure 2.3
Stream of telephone packets

The network of transfer is composed of nodes, called transfer nodes, connected between them by lines of communication on which are issued the binary elements constituting the packets. The work of a transfer node is to receive packets and to determine to what next node these must be routed.

The package therefore form the base Entity, transferred from node in node up to reach the receiver. Following the case, this package can be grouped with others in order to reconstitute the information transmitted. The action of completing a packet with bytes is called the implementation package, or still the paquétisation, and the reverse action, consisting to find a stream of bytes from a package, the dépaquétisation.

The architecture of a network is defined primarily by the way in which the packets are transmitted from one end to another in the network. Many opportunities exist for this, such as those consisting to pass the packets always by the same route or, on the contrary, to transit through separate roads so as to minimize delays in the crossing.

The reference model

To correctly identify all of the components necessary for the proper functioning of a network to transfer of packets, a reference model has been developed. This model defines a partition of the architecture in seven levels, supports all of the functions necessary for the transportation and management of packets. These seven layers of protocols are not all indispensable, particularly to networks without referred generalist. Each level, or layer, offers a service at the higher level and uses the services of the lower level.

To provide these services, the layers have protocols, which apply the algorithms necessary to the good operation of the operations, as shown in the Figure 2.4, where the protocol architecture is

divided into seven levels, which is the case of the reference model.
It should be noted that if this model is no longer used in the reality, it nevertheless continues to define the terms and to serve as reference to indicate the network functions. The architectures used today are that of the Internet, represented by the acronym TCP/IP, which brings together the two main protocols used, and that of the NFB (Open Networking Foundation), which describes the SDN networks (Software-Defined Networking). These two architectures are not irreconcilable since the SDN uses IP and TCP.
Layer 3 of the reference model, or network layer, represents the level package, which defines the algorithms necessary to ensure that the entities in this layer, the packets, are routed correctly from the transmitter to the receiver. The Layer 7 corresponds to the application level. The role of the Protocol of the layer 7 is to successfully carry the entity of the application level, the user message, equipment transmitter to the receiver equipment.
The Layer 2, or Layer liaison, represents the level frame. This allows it to transfer the packet on one physical line. In effect, a packet containing not a delimiter, the receiver can determine the end or to identify the beginning of the next packet. To move a packet, it must therefore be put in a frame, which includes delimiters. It can also encapsulate a package in another package, itself encapsulated in a frame.
This book distinguishes the words "package" and "Frame" in such a way as to make clear the entities which are not portable, directly, as the IP packet, and the entities directly transportable by the physical layer, as the Ethernet frames or ATM.
The layered structure of the protocol architecture of networks greatly simplifies their overall understanding and facilitates their implementation. It is possible to replace a layer by another of the same level without having to touch the other layers. It may, for example, replace the Layer 3 by a layer 3 premium (3') without changing the layers 1, 2, 4, 5, 6 or 7. It does not modify the so that a part of the architecture, the layer 3, without touching the rest. The interfaces between the layers must be met to achieve these substitutions: The interface of the layer 3' with Layers 2 and 4 must ensure that these do not have to be changed.

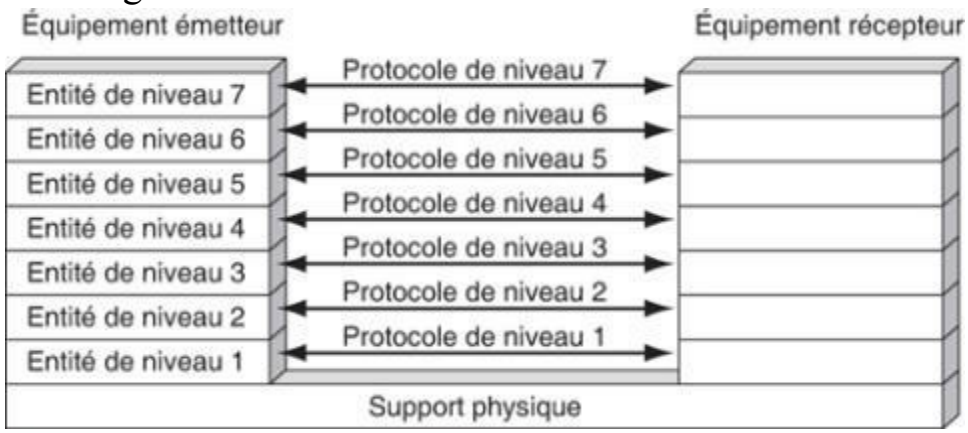


Figure 2.4
Protocol Architecture of a network to seven levels

The architecture shown in Figure 2.4 is used as a reference to all the network architectures, where his name of Reference Model. Another protocol architecture, architecture TCP/IP (Transmission Control Protocol/Internet Protocol), has been defined a little before the reference model. Its first role was to standardize the external interface of the different networks used by the U.S. Department of Defense to interconnect them easily. This is the architecture TCP/IP that has been adopted for the Internet network, which offered him a mass dissemination.
Another architecture from the use of the frame rather than of the package has been proposed by the

ITU-T (International Telecommunication Union-standardization of the telecommunications sector) for applications using both the data, telephony and the image. Coming mainly from the world of telecommunications, this architecture is well adapted to the transport of continuous flow, such as the floor the phone. It is the frame ATM that best represents this architecture. However, the latter has virtually disappeared, to be replaced by the Ethernet frame. In 2012, a new architecture has been introduced, the SDN, which uses the power of the Cloud to perform the calculations of roads or paths. The Revolution made by this architecture concerns the centralization of decisions, which were previously distributed. This new architecture is discussed in detail in Chapter 12, vested to networks SDN.

Routing and Switching

Under the concept of transfer of packets, two major techniques are competing for the Supremacy: the packet switching and routing of packets. In the routing, packets of the same client may take different routes, while in the switching, all packets of the same client follow a path is determined in advance. Many variants of these techniques have been proposed, that we describe in the following of the book.
Some applications, such as the floor telephone, face specific problems in transportation when they are forwarded in the form of packets. The difficulty lies in the recovery of the synchronism, the stream of speech before be reconstituted at the receiver with strong temporal constraints.
Assuming that a telephone conversation between two individuals accepts a delay of 150 milliseconds, it is possible to resynchronize the bytes to the output if the total time of paquétisation-dépaquétisation and crossing of the network is less than 150 milliseconds. Intelligent features implemented in the computer terminals allow this retiming. If a device does not have such an intelligence, the reconstruction of the synchronous flow is virtually impossible after the crossing of a network to transfer of packets a little complex. The networks of the Internet type have difficulty to take into account these constraints.

Computer networks

Computer networks are born of the need to connect remote devices to a central site and then the computers between them and finally of the terminal machines, such as workstations or servers. In a first time, these communications were intended for the carriage of computer data. Today, the integration of the telephone call and the video is widespread in the computer networks, even if this is not without difficulty.
Generally there are five categories of computer networks, differentiated by the maximum distance between the most remote points of the network:

- Personal networks, or Personal Area Network (PAN), which interconnect on a few meters of personal equipment such as mobile phone, laptops, pdas, etc., of a same user.
- Local networks, or Local Area Network (LAN), which correspond by their size to networks intra-company. They are used for the transport of all digital information of the company. As a general rule, buildings to wire extend over several hundreds of meters. The flow rates of these networks will today of a few megabits per second to several hundreds of megabits per second.
- Metropolitan-area networks, or MAN (metropolitan area network), which allow the interconnection of the companies or possibly of individuals on a specialized network at high speed which is managed at the scale of a metropolis. They must

be able to interconnect local networks of the various companies to give them the possibility to interact with the outside.

- The regional networks, or RAN (Regional Area Network), the aim of which was to cover a wide geographic area. In the case of wireless networks, the NAB can have a 50 kilometers radius, which allows, from a single antenna, to connect a very large number of users. This solution has benefited from the digital dividend, that is to say of the frequency bands of the analog television who have been released after the passage to the all-digital, at the end of 2011 in France.
- The WANS, or wide area network (WAN), which are intended to carry digital data on the distances to the scale of a country, or even a continent or of several continents. The network is either terrestrial, and it uses in this case of the infrastructure at the level of the soil, mainly large networks of optical fiber, or terrestrial, such as the satellite networks, but only for specific applications to flow is low.

Figure 2.5 briefly illustrates these broad categories of computer networks.

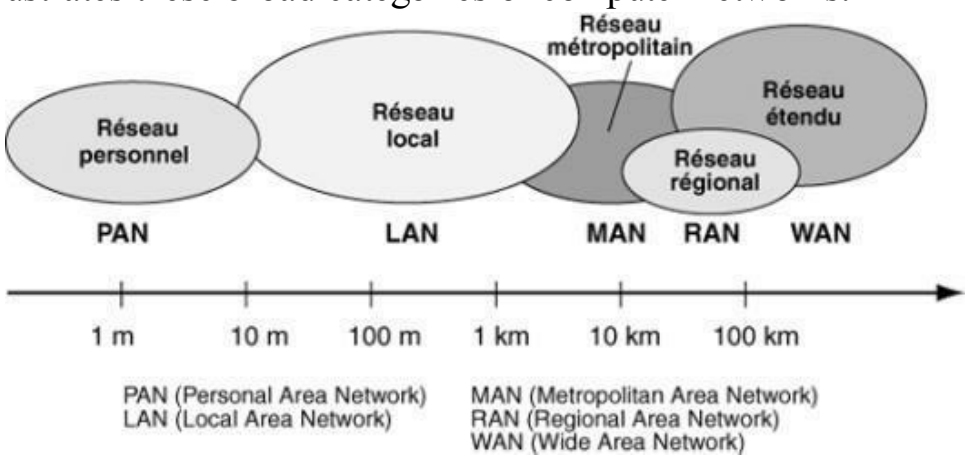


Figure 2.5

The major categories of computer networks

The techniques used by the computer networks are all from the transfer of packets IP (Internet Protocol) usually encapsulated in Ethernet frames. These techniques are studied throughout the book.

An essential characteristic of computer networks, which differentiates them from other categories of networks presented later in this chapter, is the management and control of the network, which are carried out in large part by the terminal equipment. For example, for that there is no bottling of packets in the network, the terminal equipment must be regulate itself. For this, the terminal equipment measure the response time go-return. If this response time grows too, the Device slows down its flow. This feature is made possible by the intelligence which is located in the terminal machines marketed by the computer industry.

Generally much more simple, the interior of the network is made up of transfer nodes elementary and lines of communication. The cost of the network is primarily supported by the terminal equipment, which possess all the power necessary to achieve, check and maintain communications.

Computer networks form a asynchronous environment. The data arrives at the receiver at moments that are not defined in advance, so that the packets can put a time more or less long to reach their recipient as a function of the saturation of the network. This characteristic explains the difficulty to pass the word call in this type of network, since this application synchronous strongly requires to return to the telephone handset of bytes to specific moments. The result of the publication focuses the means to regain this synchronization in an asynchronous network.

Today, the main computer network is the Internet. The Internet network carries packets called IP.

Rather than speak of Internet network, it is better to speak of IP network, which marks a greater generality. IP networks are networks that carry IP packets of a terminal machine to another. In a certain sense, the Internet is a particular IP network. Other networks, such as the intranet networks, also carry IP packets, but with different characteristics. The networks of type SDN (Software-Defined Networking) with routing are also networks using the TCP/IP protocol, but the calculations of the routing tables are made in a centralized way by a controller.

The telecommunications networks

The operators and the industrialists of telecommunications have a vision of the networks is different from that of IT specialists. Their basic application, the word phone, imposes severe constraints, such as synchronizing the ends or the time of crossing of the network, which must be limited. The opposite of computer networks, which depart from a asynchronous environment and must adapt to accept applications synchronous, telecommunications networks are based in essence on the passage of applications highly synchronized.

The floor is a real-time application, which requires that the signals are handed over to the receiver to specific moments in time. It is said that this application is for isochronous clarify this request of strong synchronization.

The solution that has been used almost since the inception of telecommunications to solve the problem of the synchronization is the switching of circuits. This technique is to put in place between the transmitter and the receiver a physical circuit not belonging to the two users in the relationship. The synchronization corresponds to the furnishing of a byte at regular interval. As was seen earlier, a device called a codec (coder-decoder) transforms the floor in byte to the transmitter and made the opposite approach to the receiver. The codec must receive the samples of a byte to specific moments. The loss of a sample from time to time is not catastrophic, since it is enough to replace the missing byte by the previous. On the other hand, if this process of loss is frequently repeated, the quality of the floor is deteriorating.

The telecommunications networks oriented to the transport of the floor Telephone are relatively simple and need not have a complex architecture. They use the switches of the circuits, or switches. Thirty years ago, when we began to imagine the networks integrating telephony and computing, the only solution proposed was based on circuits, one for the telephone speech and another to circulate the data packets.

Research conducted at the beginning of the 1980s have led industry and telecommunication operators to adopt the transfer of packets, but in adapting it to the integrated transport of information (telephone speech more computer data). Called Asynchronous Transfer Mode (ATM), or Asynchronous Transfer Mode, this technique is a transfer of packets very particular, in which all the packets have a length to the once fixed and very small. This package whose length is constant is called a frame. It is simple to find the beginning and the end since it is sufficient to count the number of bytes. With the adoption, in 1988, of the transfer of frames ATM, the world of telecommunications has experienced a real revolution.

The technique ATM has however been able to resist the mass arrival of the Internet and its IP packet. All Terminal machines from the computing world having adopted the IP interface, the problem of the transfer of packets is become that of IP packets. The telecommunications world admits, since the beginning of the years 2000, that networks must possess the IP interfaces. What is still the subject of debate, this is the way to carry IP packets from one terminal to another. The telecommunications world proposes, as the examines in detail the result of the book, to encapsulate the IP packet into a

frame and then to carry this frame and the decapsulating the arrival to find the IP packet.
Figure 2.6 illustrates the general case where the IP packet is encapsulated in a frame, classically the Ethernet frame, which is transported in the network of transfer. The case of the encapsulation in an ATM network request an additional step, consisting in cut the IP packet, because the frame ATM is much smaller than him.

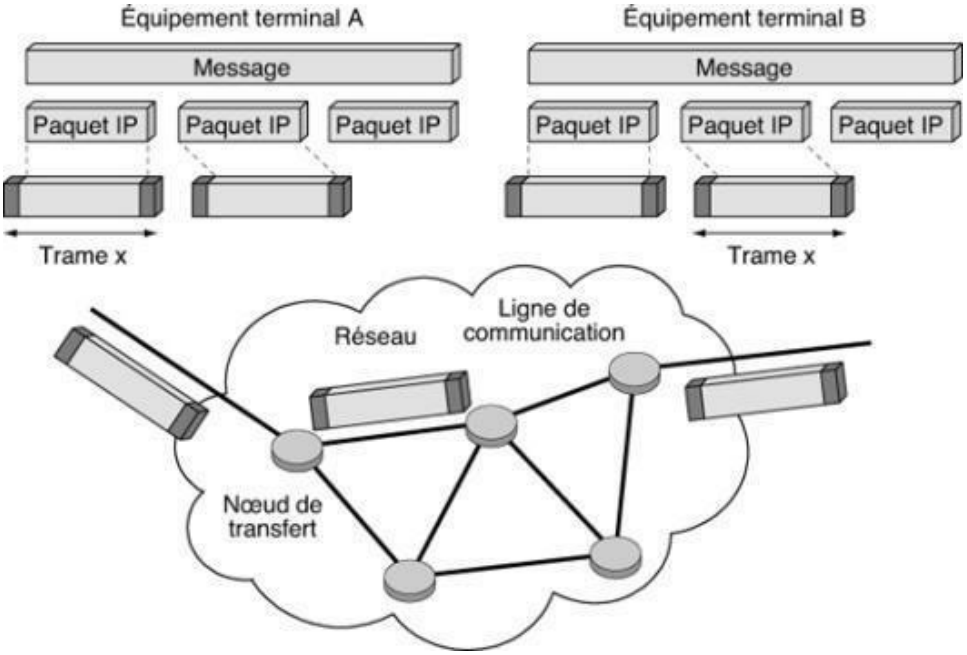


Figure 2.6
Encapsulation of the IP packet in a frame

The next generation networks, called the SDN, predominantly use the switching, but also support the routing. The controller has a thorough knowledge of clients and applications *via* its interface North and a complete vision of the network and its use *via* the interface South. Thanks to all of this knowledge, the controller can calculate the best path satisfying the request of the customer.
The signalling in the SDN is therefore centralized, unlike the previous techniques, all of which are distributed. The signalling system, carried out on the interface South, is called OpenFlow, it has been standardized by the NFB (Open Networking Foundation). This signaling however is not the only solution available (see Chapter 12).
In summary, the telecommunications networks are passed to a technology circuit to a technology package. Despite the success of the transfer ATM, optimized for multimedia, the use of the IP packet and its encapsulated in an Ethernet frame has become inescapable. The decisive question lies in the way of transporting the IP packet to ensure a high quality of service.

The networks of the cable operators

The video operators and cable operators have the mission to put in place of wired networks and loaded radio to transmit television images by land or air. This infrastructure of communication is transiting video channels to the end user. The depreciation of the wiring or radio relay passes by the provision of the users in many television channels.
The Radio Operators Provide since long years the dissemination of television channels. Their network was essentially analog until the beginning of the years 2000. Its scan is finalized since 2011 in France, as well for the satellite television that for DTT (Digital Terrestrial Television), through digital relay Terrestrial In this last case.
The quality of video images is a great variety, since the images jerky and of low definition until the animated images of very good quality. The classification of video applications, carried out according

to the level of quality of the images, is usually the following:

- **Videoconferencing.** A definition relatively low, its function is to show the face of the Correspondent. To win in flow, it decreases the number of images per second. The videoconference transports easily on a digital channel to 128 kbit/s using a simple compression to achieve. It can lower the flow up to 64 kbit/s, or even less, at the price of a quality degraded.
- **Television.** corresponds to a Channel 4 or 5 MHz of bandwidth in analog. The digitization of this channel allows to obtain a flow of more than 200 Mbit/s. Thanks to the compression, you can do down this flow rate to 2 Mbit/s, virtually without loss of quality, or even to a few hundreds of kilobits per second with a compression thrust, but at the price of a quality sometimes degraded. In addition, to such flows, the errors online becomes embarrassing, because they disturb the image at the time of the decompression. A compromise is to find between a high compression and an error rate of 10^{-9}, which does not destroy only a tiny fraction of the image and does not interfere with its vision. The transmission is done according to the standards DVB-T (Digital Video Broadcasting-Terrestrial), particularly in Europe, ISDB (Integrated Service Digital Broadcasting), in Japan and in South America and ATSC (Advanced Television Systems Committee) in North America. The two standards for the flow of transmission of a digital television channel today are H.262 / MPEG-2 and H.264 / MPEG-4.

H.264 or MPEG-4 AVC (Advanced Video Coding) is a format of coding for the recording and distribution of the video and audio in Full HD. This standard has been developed and is supported by the ITU-T VCEG (Video Coding Experts Group) and the ISO/IEC JTC1 MPEG (Moving Picture Experts Group).

The unending improvements made to the codecs should allow in a few years to move a television channel on a band even more restricted, while adding new features.

- **High definition television (HDTV).** request of transmissions to more than 500 Mbit/s if no compression is performed. After compression, you can go down to a value of the order of 8 Mbit/s. For example, on a channel DVB-T of 24 Mbit/s, it is possible to pass three channels of high-definition TV.
- **3D television.** This new generation requires flow rates even more important since it is to carry multiple images to obtain a single giving the impression to be in three dimensions. After compression, you can go down to a value of the order of 20 Mbit/s.
- **Tv ultra high definition(UHDTV).** It is a digital format of video whose main characteristic is a definition of the image containing four to sixteen times more pixels than high-definition television (HDTV). 4K is a format of digital images having a definition equal to or greater than 4 096 pixels wide.
- **Videoconferencing.** This generic term designates a conference between two or several users who is performed by means of telecommunications networks. We can also take as a definition of the Videoconferencing The very high quality. Close to the cinema, the quality videoconference requires considerable throughputs, of several tens of megabits per second. Taking account of these flows, this type of channel is only accessible with the use of fiber-optic cabling up to the user and using codecs (coder-decoder) specialty fairly expensive. A particular case, which has developed since the years 2010 In very large companies, concerns the walls of presence: it is projecting on a wall a high-quality video conference with a stereophonic sound and a transmission in real

time.

The cable operators are concerned about in the first place to disseminate moving images of type TV. The structures of wiring put in place to allow this to disseminate among the user many TV channels, which are today by the hundreds.

The Video Applications will of the remote monitoring to the video on demand, or VOD (Video On Demand), passing by the video messaging and the Home Media center for domestic video broadcast generalized to the scale of a home.

Wired networks used by broadcasters on the terminal part of the distribution network are called CATV (Community Antenna Television). The CATV uses a coaxial cable of 75 Ω, whose bandwidth exceeds 1 GHz. It is also used as the antenna cable television. It is a unidirectional support, which implies to send the signal to a center, which the rebroadcasts to all connected stations, unlike what happens, for example, in the Ethernet network, where the signal is distributed in the two directions of the physical media.

In a CATV network, the distribution of television channels is carried out easily from the center to the periphery. To add channels in the reverse direction, from the client to the head of the network, of Internet access for example, it divides the bandwidth in two: a party to go to the head of the network, the other serving users. We are talking about in this case of rising band and out of band down.

Since the cost price of the optical fiber and of the connectors associated with it has become competitive, it is used more and more to the place of the coaxial cable. The bandwidth of the optical fiber is much more important and permits to increase very strongly the flow of Internet access.

Wired networks have been exploited for a long time in analog and not in digital. The rates of flow are today sufficient to make transit of multimedia applications. However, the main difficulty is to pass several thousands of channels amounts from the terminal to the network on a shared channel of a limited capacity. Thousand clients issuing potentially at 1 Mbit/s represent a total throughput of 1 Gbit/s, which is generally significantly more than the available bandwidth on the rising part. It must therefore be very often a technique for sharing of the canal to arbitrate the access amounts of users.

In some countries such as the United States, the homes are for the cable operators a door of a simple entry to the end user. The wiring, which is one of the keys to the widespread dissemination of the information, was for many years the subject of all the lusts of telecommunications operators. The success of the *xDSL techniques, using the telephone wiring, however, has limited the impact of wired networks. The deployment of the optical fiber, of the TNT and networks of mobile 4G has greatly reduced the interest of these networks. This will be even more true with networks 5G.*

The main technique used by the cable operators to carry the channels of television is the multiplexing in frequency, which consists of a partition of the bandwidth in sub-bands. Each sub-band carries a television channel. This solution is illustrated in Figure 2.7.

The multiplexing in frequency of a large number of sub-bands has the disadvantage of requiring as many types of receptors that of channels to access. It must be a decoder for the TV, a cable modem for Internet and a telephone access for the floor Digital. The techniques of time-division multiplexing, in which the time is cut into small slices regularly assigned to users, are much more powerful since a same transmitter-receiver allows you to receive all the channels.

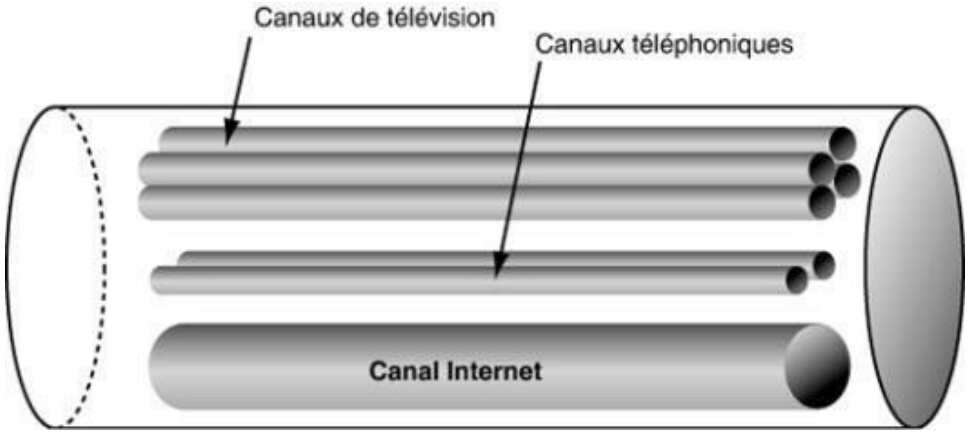


Figure 2.7

Multiplexing in frequency in the CATV

In conclusion, the technique employed by the cable operators allows integration into the CATV of a large number of applications using sub-different bands, adapted to different types of transmissions. Its main disadvantage is the multiplexing in frequency, which leads the cable operators to use a large number of bands in parallel. These bands may be considered as channels of communication that are independent of each other, so that there is no integration of flow: A customer can use in parallel a television channel, a channel of Internet access and a channel for the telephony. The television channel is connected to a cable operator, the channel of Internet access to an operator Internet and the channel of telephony at a telephone operator. The integration of these different networks in a same operator is today a reality highlighted throughout this book.

The integration of the networks

The previous sections have introduced briefly the three major categories of networks, computer, telecommunications and cable operators, who propose to carry the computer data, the floor telephone and video. Each of these networks today is trying to take charge of the three media simultaneously to move toward an integrated network. This section details the characteristics of such an integration of networks in a multimedia network, as well as the constraints that it must bear.

The world of telecommunications has adopted various solutions to equip its networks of switching solutions to obtain a quality of service satisfactory. The first of them was to use the paths associated with a class of service. The following packages this path were treated as a priority in the switch. This solution was then evolved toward the engineering of traffic. At the time of the opening of the path, the signaling packet note, in each switch, the characteristics of the traffic having been negotiated between the client and the network operator. This negotiation gives birth to a Service Level Agreement (SLA), presented in detail in Chapter 23. Thanks to these information, the nodes may decide to pass or not a signaling packet wishing to open a path. It is therefore relatively simple to negotiate a quality of service corresponding to the different applications of multimedia networks.

The networks of routing have even more difficulty to ensure this quality of service since there cannot be a reservation of resources and that it is not possible to determine in advance the routers by which must pass the packets to a same stream. A first solution to this problem is to overcommit the network for the packets to elapse in a fluid manner. If this solution was acceptable between 2000 and 2005 thanks to the significant capacity of transportation developed during the "Internet bubble", this is no longer the case today.

A new solution has been proposed consisting to introduce a classification of customers and not to overcommit than those of highest priority. This implies to discriminate against the customers, either by the payment of a higher subscription, or by restricting the number of clients of the application

considered. IP Telephony works thanks to this solution. Only IP packets out of IP phones are assigned the highest priority. In calculating the maximum number of telephone channels that can elapse on each link, we can deduce the capacity of the line for it to be seen as oversize.
The latest developments bring a new solution with the SDN by opening of the paths that are perfectly adapted to the waves, which must pass. The SDN should represent approximately 20 % of the market of networks in 2020 and much more in 2025.

Conclusion

This chapter has introduced the first concepts of networks. The convergence of these networks from different horizons, such as informatics, telecommunications and the video, to a single network is now completed.
The chapter is also stopped on the passage of networks carrying the information in analog form to networks carrying the information in digital form. Digital networks are developed in proposing several options, the routing and switching, and using as well of the physical media that terrestrial radio transmissions.
Today, the analog networks have virtually disappeared, except for very specific applications, such as telephony for the communication between an air traffic controller and a plane, for reasons of reliability and availability. But even in this case, the transition to digital will be in a few years. The reasons are the cost of the equipment and the simple reuse digital components which use only two values, 0 and 1.
The difficulties to be resolved for the perfect walking of the terrestrial networks and land are including the upgrade of the security, the management of the global network, the control of the mobility and the introduction of a complete virtualization.
The result of the work details of gradual way all the alternatives and examines the elements necessary to the construction of an end-to-end network.

3

Virtual networks and cloud

This chapter examines the network virtualization, which is becoming generalized after the storage virtualization and of the calculation. The Cloud is also a blade of substance in the application area, but also an environment highly important for the home of virtual machines. The power that we can get is such that the time of calculation to obtain a routing table or run an algorithm for the management of a handover or a control of security are minimal, which allows the reaction time particularly low.

The network virtualization

Virtualization is not a new technique, since it had been introduced on the first large computers in the 1960s, which used a virtual memory, that is to say a memory which, instead of being in RAM, was located on a hard disk. All the TIP was to bring the pages of memory of the hard disk to the RAM memory just before the central unit is needed. Subsequently, a number of mechanisms have been implemented with this solution so that the user has the impression that the services they need are located on a machine located in close proximity, while they are in reality on a distant machine.

The storage virtualization has had a tremendous success in allowing to consolidate several storage servers on a single machine, each user who yet the impression of having a separate server. The virtualization of network holds the same principle: multiple virtual networks share a same physical infrastructure.

The network virtualization has been launched by a U.S. project called geni (Global Environment for Network Innovations), in which Intel proposed to construct a router consists only of its hardware part, without any network operating system. The software publishers no longer had to implement their system on the machine. The advantage of virtualization lies precisely in the possibility to implement multiple network operating systems on the same machine with the help of a "hypervisor".

The architecture of a virtualization server to support multiple virtual machines on different operating systems is shown in Figure 3.1.

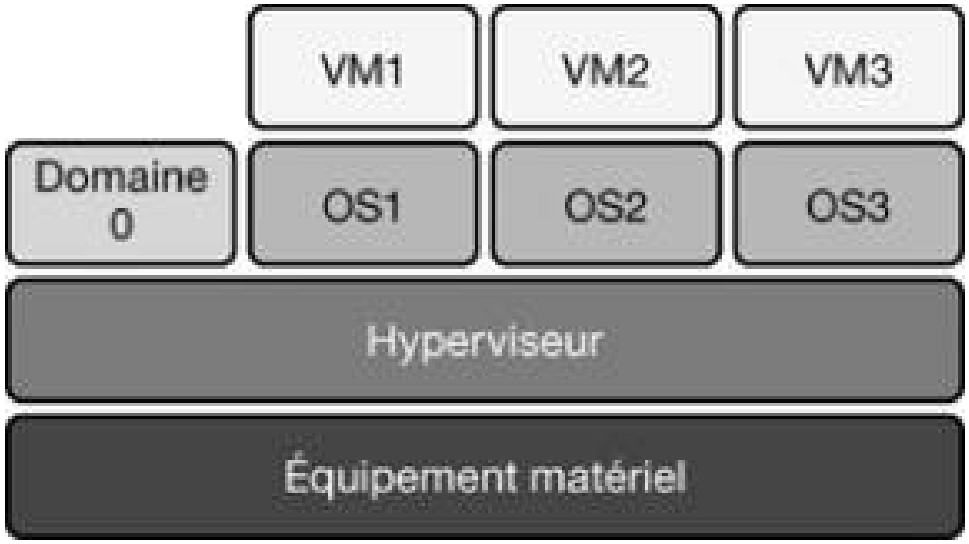


Figure 3.1
Architecture of a virtualization server

On this figure, a hardware equipment, for example a server in a datacenter, has a hypervisor. The hypervisor is a specific operating system capable of supporting multiple operating systems, which, themselves, support virtual machines, or Virtual Machine (VM).

Virtual machines can be very diverse, such as a router, a switch, a software for signal processing, an enclosure, etc. The number of virtual machines supported depends on the power of the physical machine, the number of central units and the capacity of the memoirs of storage. If the virtual machines spend little resources, several tens of them can run simultaneously.

A VM (Virtual Machine) specific, called Domain 0, allows to realize the inputs and outputs of the other virtual machines. It is detailed in the section of this chapter devoted to the Xen hypervisor.

The VM is itself a software which describes exactly what the physical machine that you want to virtualize. The code represents the VM. The immediate benefits of this virtualization are the agility and flexibility that it provides. For example, you can easily change of VM without touching the hardware part and replace a router by a switch in a few seconds. On the other hand, the energy consumption is increasing since it takes a lot more power on the part of the central unit when it virtualizes the work of a ASIC (Application Specific Integrated Circuit), that is to say of a material element specialized.

Another benefit of virtualization is to allow the grouping of the virtual machines on the same server when these are little used, such as routers during the night. These groupings are being carried out by the migration of a VM from one physical server to another physical server. One of the ways to regain the energy lost by a supplementary application for central unit is to consolidate virtual machines and standby servers that have no more virtual machines to run. As a general rule, it does not extinguish the physical machines, but are put to sleep in order to allow an immediate reboot or almost immediate when the system has again need more power.

A virtual router is as well as a logical instance of a physical router. Several virtual routers can run at the same time on the same physical server, with the inputs-outputs required to achieve completely the physical router. It then gets the opportunity to deploy multiple networks, that we can call virtual networks, on the same physical network. A virtual network is the grouping of a set of virtual machines compatible.

The routers or the virtual switches use the physical links to interconnect. We can even go much further by grouping of equipment of virtualized network using different protocols on a same physical infrastructure. Of this fact, we can create virtual networks of different types on this infrastructure, as a network of virtual routers IPv4 and a network with virtual switches to create a network MultiProtocol

Label Switching (MPLS).
Virtual networks that one decides to create on a same infrastructure can add network equipment virtual, as firewall (firewall), authentication servers, servers for manage network applications such as voice over IP, or Voice over IP (VoIP). It may also add other virtual machines storage and computing scattered or grouped in the data centers that form the backbone of the infrastructure.
Such an environment is illustrated in Figure 3.2, in which the networks A, B, C and D coexist on the same infrastructure servers and the same physical lines.
The virtual network has is realized from Virtual Machine (VM) with the operating system A. Similarly, the Virtual Network B is achieved with the Vm B and so on. Each VM on the same physical server may employ protocols completely different, and the networks also be totally different. Everyone can use the protocols related to a given application, such as VoIP, the transfer of files, e-mail, access to a security system, video broadcast, etc.
The benefits of this technique of virtualization are numerous. The first is to be able to rotate on a same physical infrastructure of virtual networks using technologies totally different. It can, for example, obtain a first network using the network operating system IOS from Cisco, a second using the Junos of juniper, a third and a fourth using the systems of Nokia and Ericsson, etc. It is also possible to run several different releases of the same operating system network. It is obvious that the previous examples have little chance of being achieved, because the OEMS do not wish to necessarily commercialize their software without their equipment. The decoupling hardware/software will only be available for open systems mature. Chapter 14 gives examples of such open source architectures, which will constitute the standards of the years 2020.

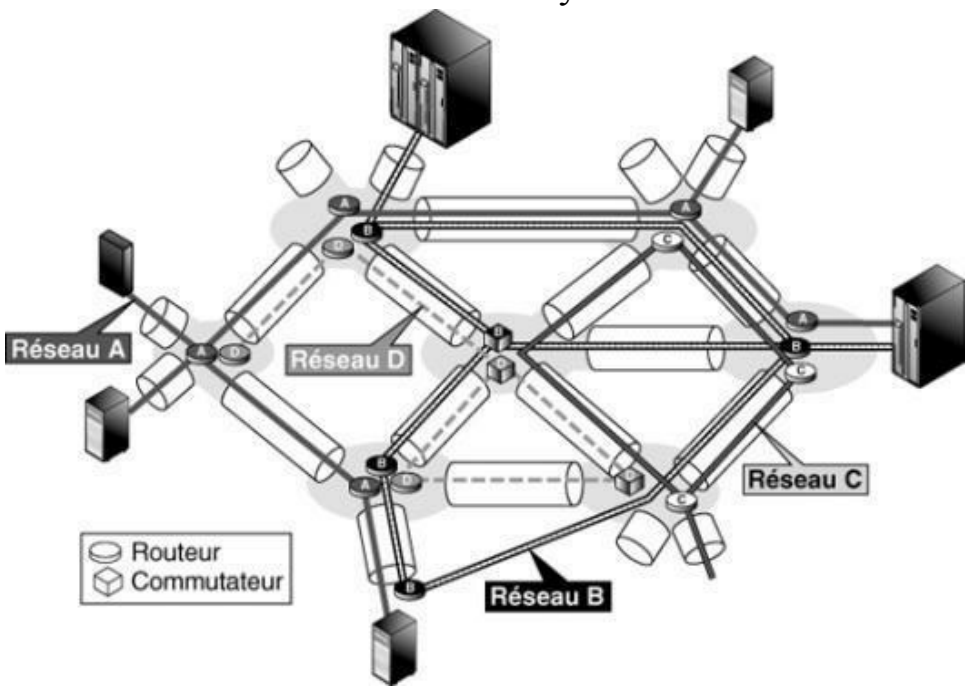


Figure 3.2
Virtualization of routers

A second benefit of virtualization concerns the use of resources. It is obvious that the resources can be much better used by running and Stopping Virtual Machines or still in the Moving: if one of them has not enough resources on the physical machine on which it is placed, for example, it is possible to move other physical servers.
In all cases, the resources of a same virtual network must be fully isolated from other virtual networks. The insulation prohibits a packet can pass a virtual network to another to avoid that a problem on a network can assign a different network. The passage could be allowed through external gateways particularly robust to maintain the insulation. This implies an excellent security in the

virtualization software, since it is necessary to avoid attacks on the hypervisors. If one of the virtual machines fails, not only must it not affect other networks, but they must not even notice it. If, of course, a physical router fails, all virtual machines are also in failure.

Technologies of network virtualization

As indicated previously, the technologies of network virtualization is based either on the hypervisors, either on containers (containers).

Hypervisors and containers

There are three major types of solutions to achieve virtual systems: paravirtualization, operating system virtualization and the insulation by container.

An example of a paravirtualization is described in Figure 3.3. Above the hardware of the infrastructure, the hypervisor enables you to support different operating systems that have been modified to ensure that the instructions to run directly on the processor of the infrastructure. In this category, there is the hypervisors Citrix Xen Server (open source), VMware vSphere, VMware ESX, Microsoft Hyper-V Server, bare metal and KVM (open source). The advantage of this solution is a quick execution, but to the condition to change a number of elements of the operating system.

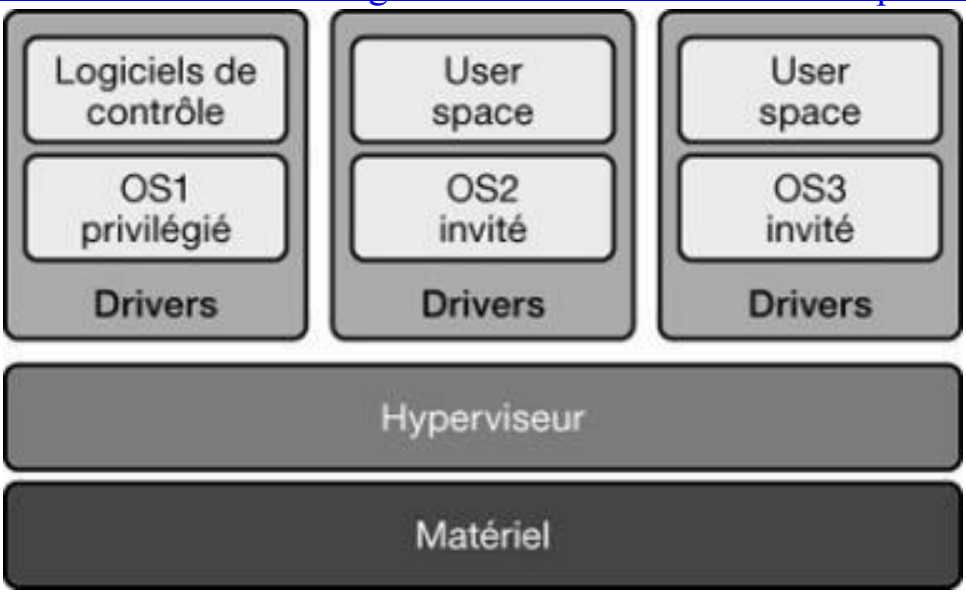


Figure 3.3

Example of paravirtualization with the help of a Hypervisor of type 1

In the second solution, called a hypervisor of type 2, the operating system that is supported by the hypervisor is not changed. Of this fact, some instructions can no longer run on the hypervisor, which in reality is the host operating system to the physical infrastructure. It should be added between the two, as shown in figure 3.4, an emulator that allows you to execute the instructions of the guest operating system on the host operating system. In this category, we can put Microsoft VirtualPC, Microsoft Virtual Server, Parallels Desktop, Parallels Server, Oracle VM VirtualBox (open source), VMware Fusion, VMware Player, VMware Server, VMware Workstation and QEMU (open source).

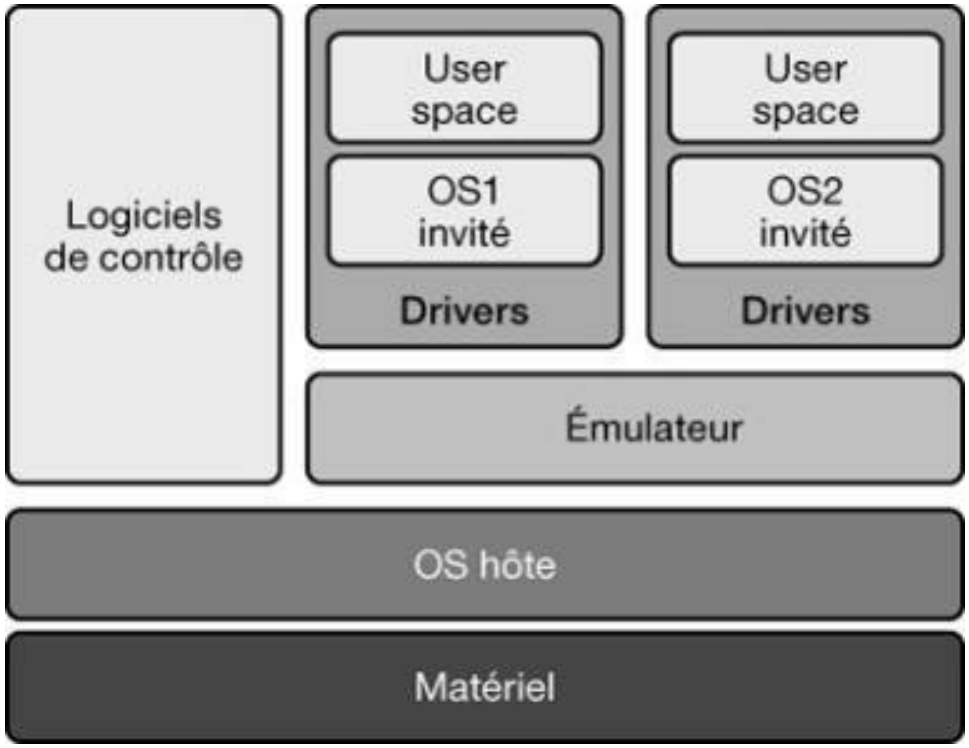


Figure 3.4
Example of operation of a Hypervisor of type 2

The third category, represented in Figure 3.5, supports virtual machines by putting them in containers, or isolators. The container contains in itself a runtime environment, complete with its drivers, its binary files, or libraries, as well as the application itself. The container allows you to isolate the execution of applications in what is still called contexts or areas of execution. In this category we can place Docker Linux-Vserver, chroot, BSD Jail and OpenVZ.

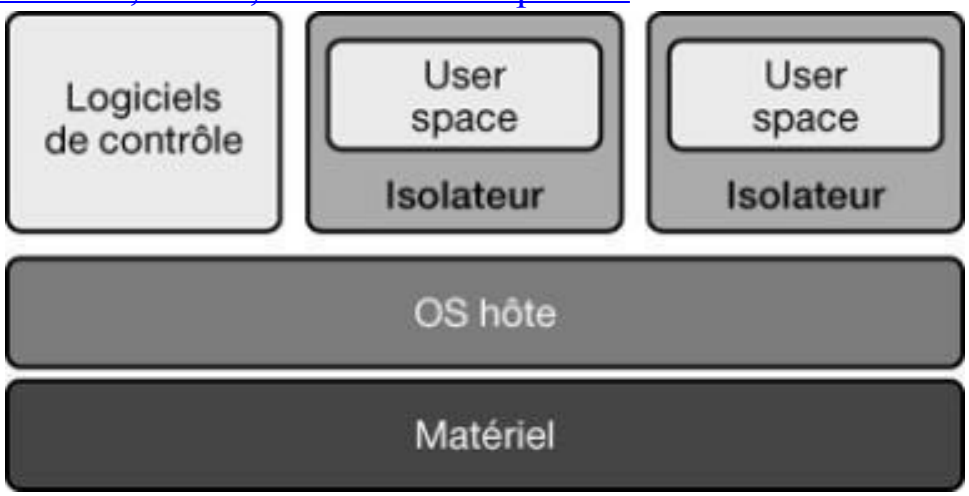


Figure 3.5
Example of supporting container of virtual machines

The insulation

The insulation is a fundamental principle of the virtualization of network because it is absolutely essential to avoid that a network impacts other networks.

A first simple solution, but little effective, would be to partition the resources between the different networks. In this case, there would be no large-thing to win, because the resources of a network not used at a moment *t* may not be used by other virtual networks. It is therefore necessary to find a solution for both partition the resources and to ensure that the resources not used to work for the benefit of the other.

You can obtain this property by the schedulers. The simplest way is to have in a node as many queues

as virtual networks. To each queue, it grants credits in function of the resources that have been allocated to the node. Each network uses a credit to send a packet. When a station has no more credit, it cannot issue. As long as there are other active queues, she is arrested. If, at a given moment, all queues are empty, a reallocation of appropriations is assigned to all the queues. However, this may lead to a blockage to the times of the Queues not having more credit and those with appropriations but not of packets to serve. To avoid this problem, queues blocked can create negative appropriations and continue to transmit. When the sum of the credits both positive and negative becomes equal to zero, it resets the system in the granting of positive tax credits to all nodes.

It may, from this example taken from Xen (Http://www.xen.org), modify parameters or add others, as to allow a queue to serve multiple packets, i.e. to use multiple tokens without interruption. It thus uses less energy, a same VM being to the work without interruption.

Xen

Xen is a hypervisor, that the is still called monitor of VM, or Virtual Machine Monitor (VMM). It is a open source software that runs on hardware platforms standards. In addition to the VMM, located on the physical hardware, the Xen architecture is composed of several areas turning simultaneously on the hypervisor, referred to as virtual machines (see figure 3.6). Each VM can have its own operating system and its applications. The VMM controls the access to the equipment multiple domains and manages the sharing of resources between the different areas. Thus, one of the main tasks of the VMM is to isolate the different virtual machines, so that the execution of a VM does not affect the performance of the other.

All the device drivers are kept in an isolated field which is reserved for them. Called domain 0 (dom0), it provides them with a support hardware reliable and efficient. The Field 0 has special privileges compared to other areas, called user domains (domU), and has, for example, of a total access to the hardware of the physical machine. The user domains have virtual drivers and operate as if they could access the hardware directly. However, these virtual drivers communicate with the dom0 in order to have access to the physical hardware.

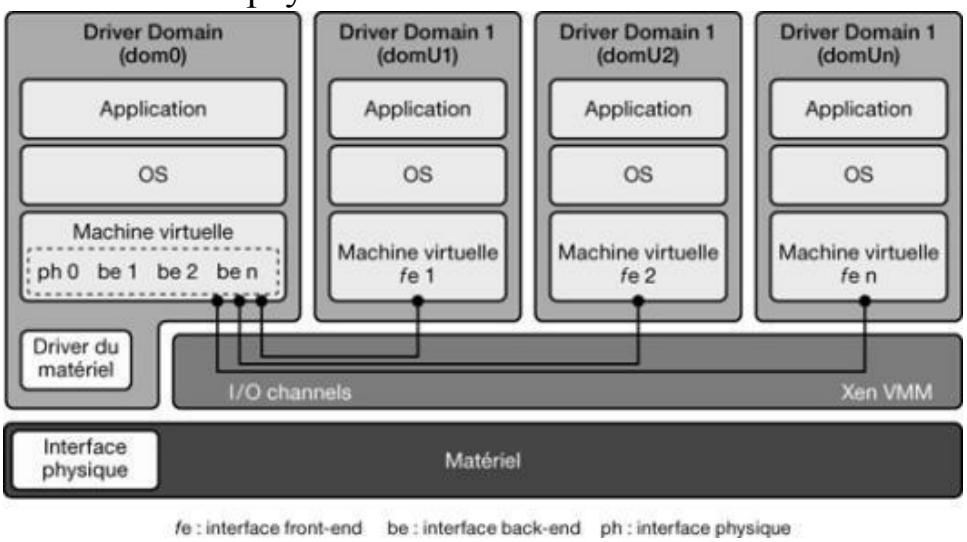


Figure 3.6

Xen hypervisor and virtual machines

Xen Virtualizes a physical network interface unique in démultiplexant incoming packets of the physical interface to the user domains and, in reverse way, Multiplexing the outgoing packets generated by these user domains. In this procedure, known as virtualization inputs and outputs of the network, the dom0 accesses directly to devices in the input-output by using its drivers of native devices and performs operations in the input-output on the part of the domU.

The user domains employ of the devices in the input-output controlled virtual by of virtual drivers in

order to ask the dom0 access to the device, as shown in Figure 3.7. Each domain user has its own virtual network interfaces, called interfaces of first plan, required for the communications of the network. The interfaces of the background are created in the dom0 corresponding to each interface of the first plan in a user domain and act as a proxy for the virtual interfaces in the dom0.
The interfaces of foreground and background are connected to each other *via* a channel of entry-exit which employs a mechanism the *zero-copy* to reset in correspondence the physical page containing the packet and the target domain. In this way, the packets are exchanged between the interfaces of the background and the foreground. The interfaces of the first plan are perceived by the operating systems running on the user domains as of the actual interfaces. However, the interfaces of the background in the dom0 are connected to the physical interface as well as to each other *via* a virtual network bridge. This is the architecture by default, called bridge mode, used by Xen. Thus, at the time the channel of inputs and outputs, and the Network Bridge Establish a communication path between the virtual interfaces created in the user domains and the physical interface.
Different elements of the virtual network can be implemented using Xen since it allows multiple virtual machines to run simultaneously on the same hardware, as shown in Figure 3.7. In this case, each VM is running a virtual router, which has its own plans of control and data, the virtualization layer of Xen being of low level.

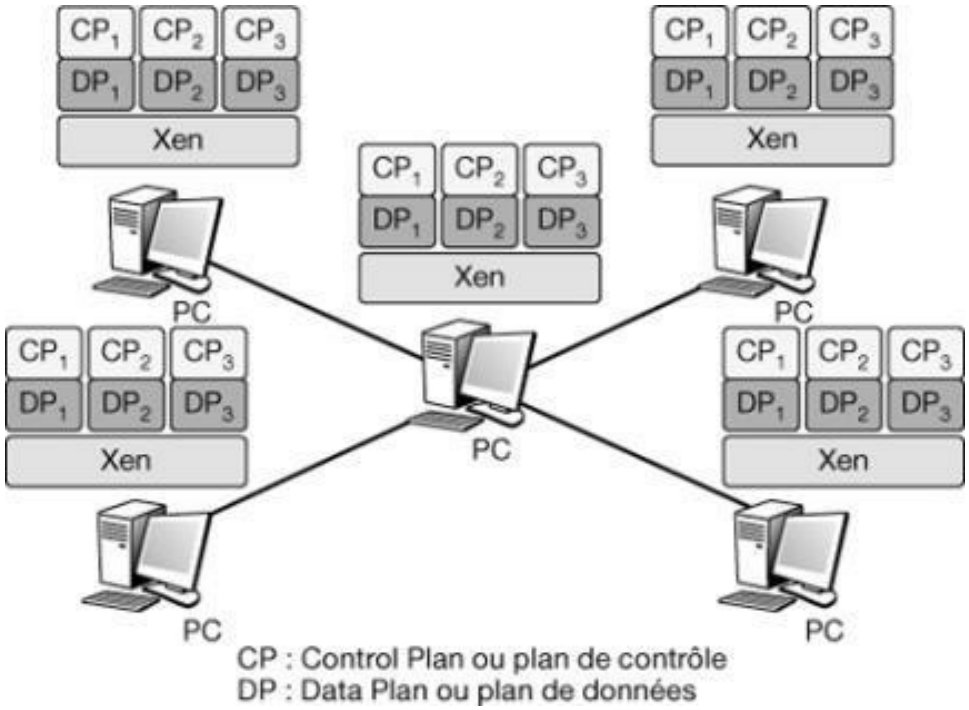


Figure 3.7
Virtualization with the Xen hypervisor

Use of the network virtualization

Two types of use may be made of the network virtualization:

- The creation of multiple virtual networks within the same enterprise to manage separately telephony, telemonitoring, videoconferencing in the company, etc.
- The sharing of the same virtual networks by different operators in order to reduce costs. When one of these wants to increase its capacity, it asks the operator of infrastructure to increase its resources.

Another use of virtualization continues to grow for the applications of type cloud. In simple terms, the principle of the Cloud is to place the resources of a company somewhere in the network in order to share them with other users. The Cloud offers resources of types server, compute, storage, software,

etc. The user can share these resources with other users, while taking into account the security of the information.
The virtual resources can migrate to optimize physical resources. Similarly, the customer can move on the Internet and have access to its resources from anywhere. For linking resources to the user, it must create networks. And they may not be as virtual since it would be impossible to create a new physical network to each moving or request for additional resources to a user.
Virtualization is finally a ideal technology to test new protocols and architectures without stop the operational network since it is sufficient to build a new virtual network with the new generations of protocols or architectures. The insulation is in this case a key property for that tests not to interfere with operational networks.
We can conclude that virtualization has many advantages. It is today regarded as an opportunity to bring the flexibility to the networks. In addition, it allows you to switch from one technology to another without too many complications.
The disadvantages of the virtualization are related to the overload provided by the hypervisor and the needs of memory and computing power, especially if one wants to completely separate the different areas.

Virtualization of the network equipment

It is possible to virtualize many network equipment. Virtualization in a Wi-Fi Controller, for example, allows you to manage points of access associated with each virtual controller. The access points associated can have their own protocol stack.
It can also virtualize firewalls. To access in the company, the waves are diverted to the virtual firewall General, who, in turn, may return the flow toward a specific firewall, always virtual. The firewalls can thus benefit from all the power necessary for carrying out the inspections to detect attacks.
The access points themselves can be virtualized. It is enough for that to introduce in the housing a hypervisor supporting several operating systems. Each virtual access point thus has its own software for the management and control and is independent of other points of access: it can therefore represent a particular operator. Similarly, it is possible to virtualise the antennas 2G, 3G and 4G (BTS, NodeB and e-NodeB) and share these virtual antennas between different operators.
Overall, any resource can be virtualized, except the sensors, who need a hardware component to measure a value. A virtual sensor is therefore a software associated with a physical sensor to perform calculations from the measured value. Of course, this software can be modified, where his name of virtual sensor. The future is thus clearly to the virtualization of the set of materials.

NFV (Network Functions Virtualization) and standardization of virtualization

The standardization of virtual machines is done to the ETSI (European Telecommunications Standards Institute). One might think that it only concerns the Europeans, but not at all: the ETSI is open at the global level, and all operators and equipment suppliers are involved in this standardization.
The term used, NFV (Network Functions virtualization), indicates the desire to define once and for all the different functions used in the field of networks. However, normalize the functions was not sufficient. The ETSI is gone further by proposing to develop in open source the different network functions, such as the routing of a router, the switching of a switch, the protection of a firewall, etc. from these virtual machines, you can construct a complete network in open source with multiple virtual networks simultaneously.

The standard architecture of the years 2020 In course of development has taken the name of OPNFV (Open Platform for NFV). It is described in detail in Chapter 14.

The most important of this standardization is however elsewhere: it is to allow the decoupling of the network function and of the physical equipment, a radical newness. In a router of the old generation, the routing function and the transfer of packets is carried out in the same physical equipment. In the new generation, the routing function can be performed in a remote server from the physical machine that performs the transfer of the packet. This dissociation allows you to consolidate the functions of routing in of the servers in the data center by optimizing the overall operation. The resources allocated to the virtual machines that perform these functions are perfectly adjusted. In addition, these resources can vary between the peak hours and the night, where those assigned to the virtual equipment are greatly reduced.

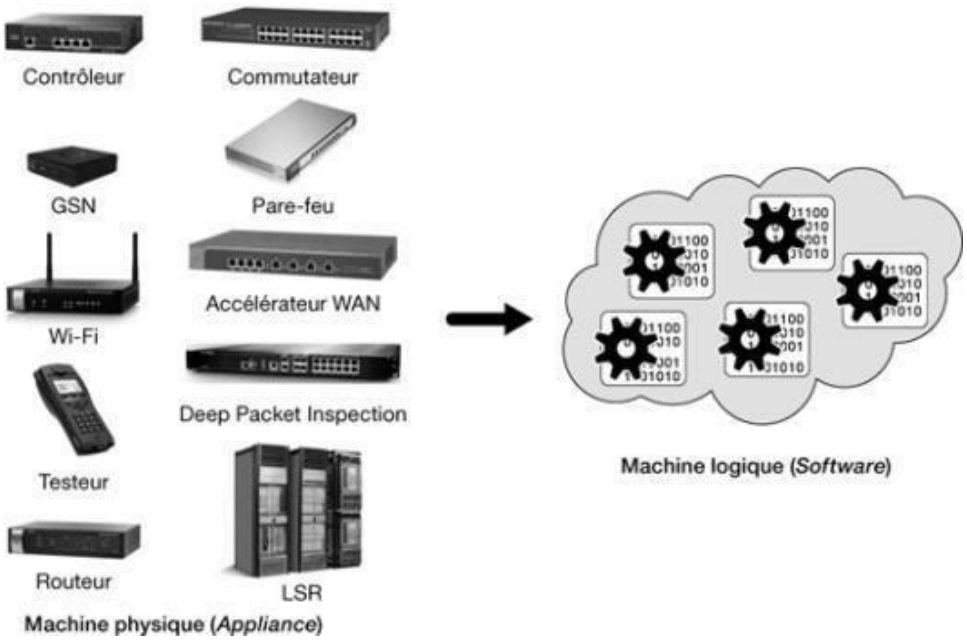


Figure 3.8

Passage of physical machines to logical machines with virtualization

Figure 3.8 illustrates this passage of physical to virtual machines corresponding to the network functions. Other Functions non-network, such as the storage, the calculation, security, etc., can be added just.

The virtualized networks

The characteristics of the networks detailed in previous sections are put to contribution in this section to introduce the principles of virtualization.

A network of new generation is described in Figure 3.9.

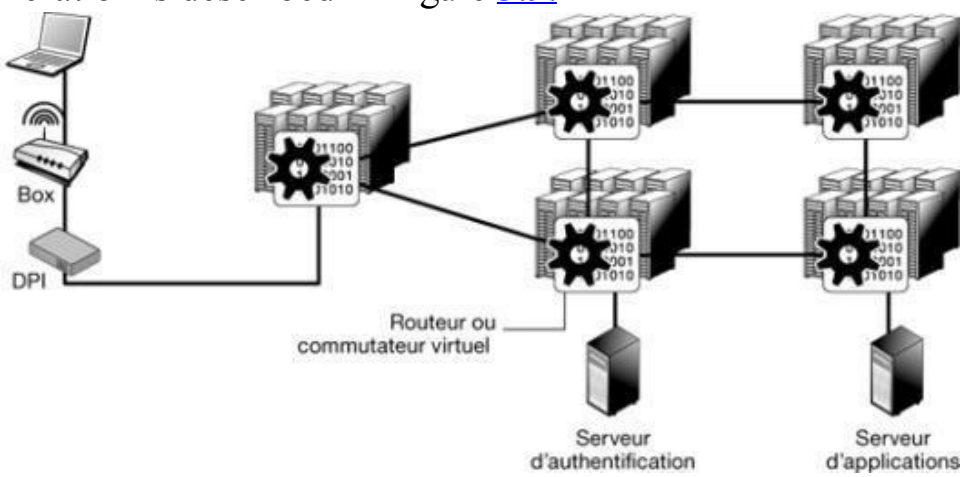


Figure 3.9

The infrastructure is carried out from datacenters, in which are positioned virtual machines, here routers or switches, and the enclosures various, such as box, authentication servers and applications, firewall and DPI (Deep Packet Inspection), this last inspecting the waves that pass through the network. These enclosures are also virtualized (see later).

As illustrated in Figure 3.10, it is possible to add a second virtual network, which uses for example of protocols completely different from the first one.

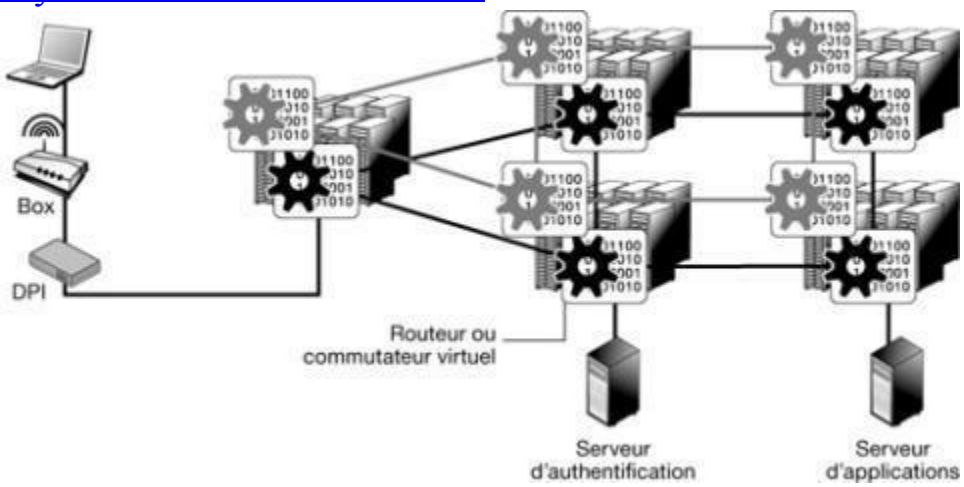


Figure 3.10

A network of new generation supporting two virtual networks

Always going a little later in the process, we can virtualize the different enclosures in data centers, as shown in the figure 3.11, where the physical enclosures have become virtual machines integrated into the data centers. It must of course retain physical form the elements that cannot be virtualized, as the antennas or the sensors.

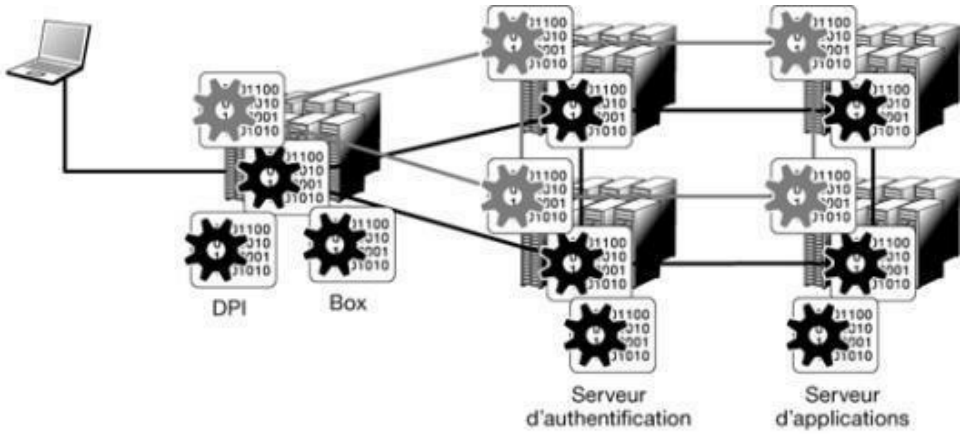


Figure 3.11

Network Environment fully virtualized

The optimization of the location of the VM is not simple. This process is called the urbanization of virtual machines. It should be noted that urbanization can change the whole according to the objective pursued. For example, urbanization to optimize the performance is totally different from the urbanization to optimize energy consumption. In the first case, it is necessary to distribute the VM a little everywhere in the network in order that they have each of the more than physical resources possible. In the second case, there is a need to aggregate by migration all virtual machines on servers common physical, so to be able to put it to sleep very many physical servers, the electrical consumption being highly proportional to the speed of the processor. The Urbanization is still quite different if one tries to optimize the safety or the availability of the network or its reliability.

Thanks to the NFV, you can gather all the functions in the same datacenter, as shown in the Figure

3.12. We can speak in this specific case of NFVaaS, that is to say that it was the network functions available in a cloud, represented here by a single datacenter. It can be seen that the network functions have migrated data centers that formed the physical infrastructure of the network toward a datacenter which may be very far from the nodes of this same network.

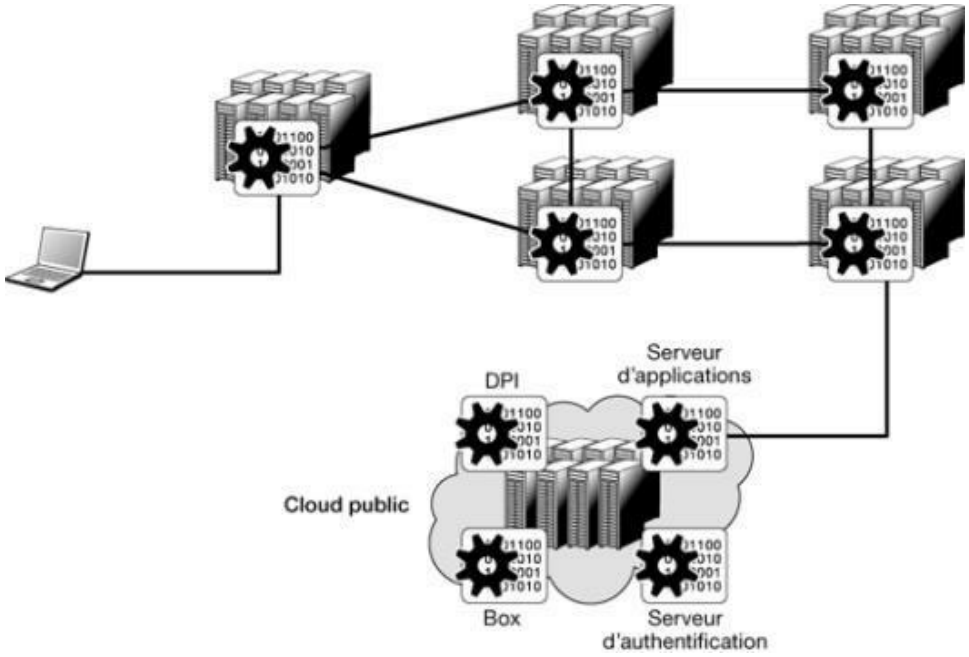


Figure 3.12

A managed network in NFVaaS NFV (as a Service)

The architecture of networks NFV can be summed up as shown in Figure 3.13. The virtual machines on the network functions are called VNF (Virtual Network Functions). These functions are installed on an infrastructure called NFVI NFV (infrastructure). They are worn by physical machines with hypervisor or container that virtualization. So that they can operate without problem, they must be orchestrated and managed by an environment called MANO (management and network orchestration). These orchestrateurs, examined in detail in Chapter 4, are bodies to create virtual machines, and then chaining, i.e. to determine the order in which it must cross, and finally to "urbanize", either to put them on physical machines in order that the whole works in the best possible way.

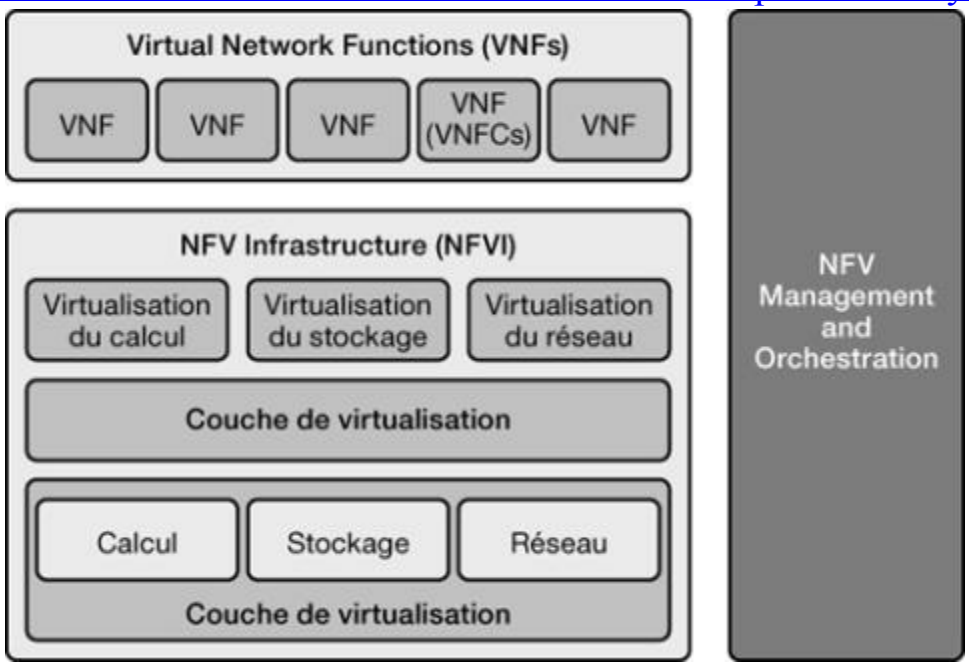


Figure 3.13

All of the elements involved in the network virtualization

Conclusion

Virtualization is an important technique, which allows the introduction of new technologies in the form of virtual networks individuals. It is thus possible, without risk, test on an operational network of new architectures. This solution also offers the networks the opportunity of a thought in the long term, because when the architecture of the future will be found, she will be able to introduce little by little in the world of networks through virtualization.

The virtualization of networks and, more broadly, that of resources constitute a turning point in the architecture of networks. In the years 2020, all network equipment will be virtualized, and the virtual equipment will be able to move according to the context in order to optimize the overall performance following criteria which remain to be defined. The complexity of these new environments is optimized by processes of more and more intelligent, often making call to the artificial intelligence, as detailed in the Chapter 4.

4

The intelligence in networks

The intelligence is a traditional term in computer science which simply refers to the ability to communicate, to reason and to decide. Until the beginning of the years 2000, the intelligence in networks was very low. The concepts of intelligent networks, which date back to the beginning of the years 1990, introduce a primary intelligence, whose role is to automatically adapt the components of the network to the requests of the users, but without reasoning and only by following the rules defined in advance.

The Intelligence has continued to increase since 2000 and will continue to do so in the future, to the point that we are talking now of "smart networking" by introducing in the management and control of the networks of intelligent agents, detailed a little later in this chapter, able to bring, for example, the diagnostics. We See Also appear the Autonomic networks, which tend to replace the programmable networks or the active networks. A autonomic Network is a network that is able to autoconfigure and whose nodes can become self-sufficient in case of failure or loss of communication.

Since the beginning of the years 2000, the real intelligence, that is to say on the basis of reasoning, appears in some components of the network to take control or management decisions. The bodies which take the decisions come from the field of artificial intelligence and Smart Objects. In particular, multi-agents systems propose to manage the security or failures.

The networks of today and of tomorrow will appeal to an intelligence much more important to really drive the network by equipment called orchestrateurs and controllers. The objective is to achieve automatic actuations of the network, in the manner of what is done in the aircraft. However, the difficulty is much greater, because the system is highly distributed. As explained in Chapter 12, devoted to the SDN, the solution adopted in a first time is to centralize the command.

This chapter introduces in a first time these orchestrateurs and controllers to achieve automatic actuations and then examines the transitional tools which make possible these complex processes of pilotage.

Orchestrateurs and controllers

Before addressing the tools from the artificial intelligence, this section examines ways and means to develop automation much more advanced than those of products marketed since a few years.

The controllers are bodies capable of controlling the network equipment, and more precisely to the configure so that they can transfer the packets from the entry of a node to the right output queue. The controllers also occupy the other network elements, such as the firewall or the dpi (Deep Packet Inspection). They must possess all the algorithms necessary to configure these elements of network, such as the gas struts of loads (load balancers), managers of VLAN (Virtual Local Area Network), but also of the firewall, etc.

As their name indicates, the orchestrateurs play the role of heads of orchestra for the network environment. Their first function is to create virtual machines, or VM, which will be necessary to achieve the network intended to support the flow of packets from the application that requests the opening of a communication.
After you have chosen the VM necessary, it must the chaining (chaining), i.e. describe the journey of the packets across the different VM. An example of chaining is described in Figure 4.1. Between the points of entry and exit of the network, packets must traverse of the VM that perform functions of network, or VNF (Virtual Network functions), for example a router followed by a firewall, followed by another router, then a housing giving specific functions, and then… There may also be referrals, with functions in parallel.

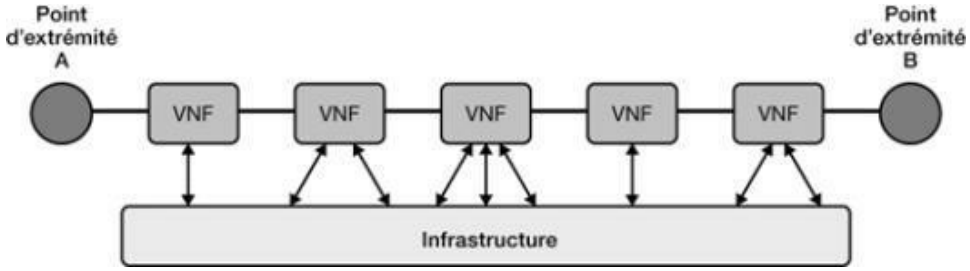


Figure 4.1
The chaining of virtual machines

After the chaining, the orchestrator must fulfill the function of urbanization, i.e. place the VM on physical machines. As shown in figure 4.1, a VM can have several physical instances of the virtual function located in different infrastructure servers, some of which may be located far enough away from the machine performing the work. Of duplications and triplications are also possible.
The result of this section gives some examples of controllers, before to look on the orchestrateurs that can acquire today.
There are two broad categories of controllers: those from the world open source and those, closed, from some OEMS. A lot of manufacturers have chosen to start from a strain open source and the supplement by specific modules. The controller used the most, by far, is ODL (OpenDaylight).
Among the other controllers open source, these include the following:

- NOX, the first controller compatible with the signaling protocol OpenFlow, examined in detail in Chapter 12.
- FlowVisor, realized in Java, which also uses the OpenFlow signaling and behaves as a transparent proxy between a switch and multiple controllers, all OpenFlow.
- POX, written in Python and always oriented OpenFlow, with a high-level interface for the SDN (Software-Defined Networking).
- Floodlight, written in Java and oriented OpenFlow.
- The ONOS (Open Network Operating System), a network operating system distributed in open source, equivalent of a controller distributed.
- OpenContrail, a Controller Open Source resumed by Juniper.

Many other controllers are also available, such loom, Ryu or Trema.
Let us look at more precisely the three controllers Open Source The more used today, which are ODL, OpenContrail and ONOS.
The architecture of the latest version of the controller ODL, called Oxygen, is described in Figure 4.2.
The controller contains three parts. The high part, which corresponds to the interface north (northbound Interface), contains the interfaces with the applications that require a communication to a

recipient. Through this interface, the controller retrieves all information necessary to allow for the opening of a path or route in the network in guaranteeing the quality of service and, more generally, the authorisation which has been concluded between the user and the manager of the network, or Service Level Agreement (SLA).
The second part corresponds to the lower part of the FIGURE 4.2 : The interface south (Southbound Interface), located between the controller and the network equipment, which can be virtual or not. This interface transports in a direction the configurations to apply in the nodes of the network and in the other all the information from the measurements made on the network, which will allow the controller to take the right decisions.
The third part, in the center of the controller, corresponds to all functions that can be exerted to make decisions, secure the network and, more broadly, check the functions dispersed in the network.

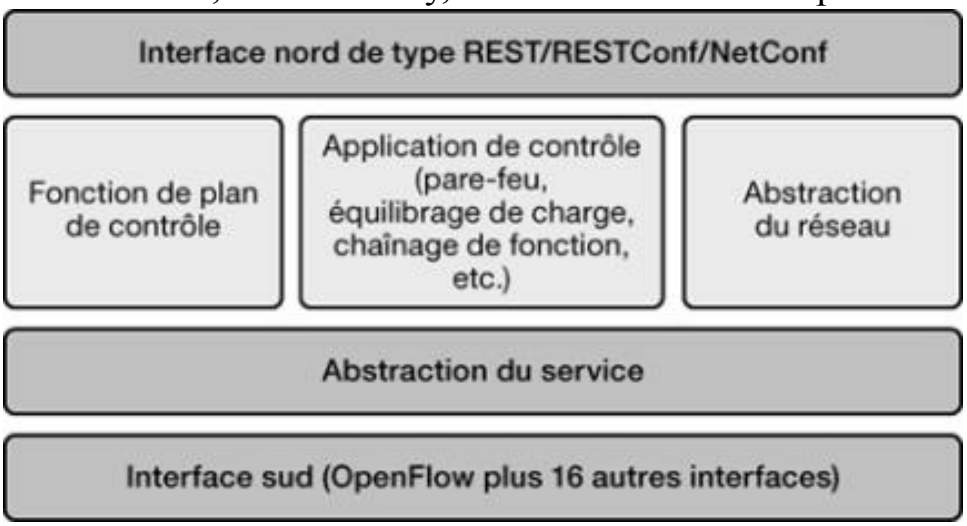


Figure 4.2
Architecture of oxygen, the last release of the ODL controller

Figure 4.3 shows the controller OpenContrail that comes from an open source project resumed in hand by Juniper. The interface components North and South More The control functions are found in this controller. The figure shows the elements from the intelligence embedded in the virtual appliance. It is the transformation of the data which comes from the interface North, for the application information, and the interface South, for the escalation of measurement information from network equipment, that they are virtual or not. This set of data is collected in a Big Data, whose tools called *analytics* are used to take the decisions of the opening of path or route in the network.

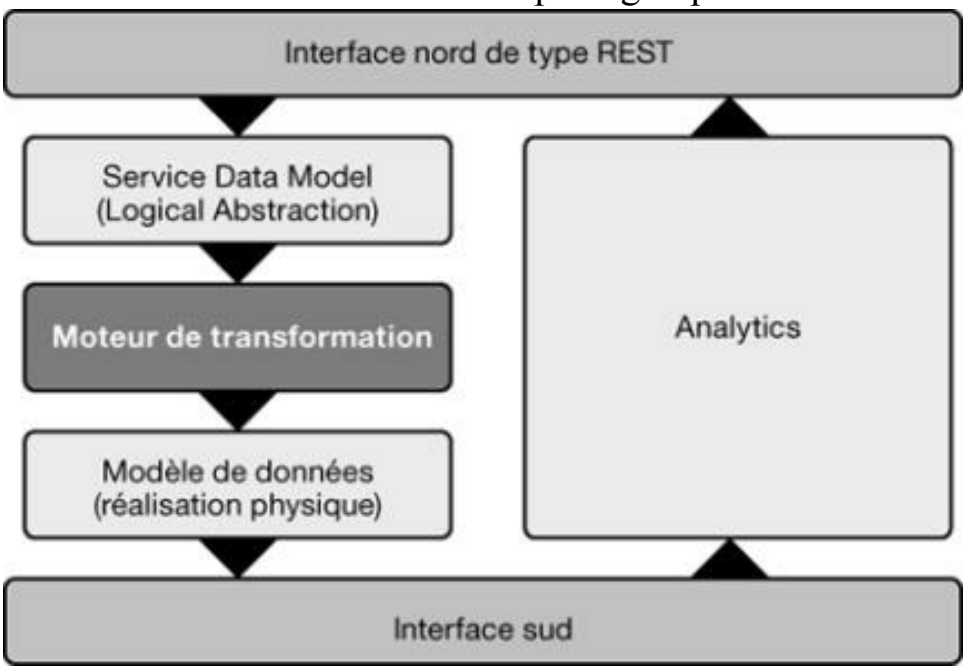


Figure 4.3
Architecture of the OpenContrail controller

The controller the onos (Open Network Operating System) uses the same concepts that the two precedents, but by introducing, important novelty, a distributed system, which allows you to operate the system using multiple controllers simultaneously. The controllers are interconnected by interfaces is and West which allow you to enlarge the network by removing the limitation introduced by a single controller.

Figure 4.4 describes this controller at any point to conform to the characteristics previously introduced, the only difference from the central system distributed, which allows you to assign the virtual machines to different controllers, causing the so primary and secondary connections in relief.

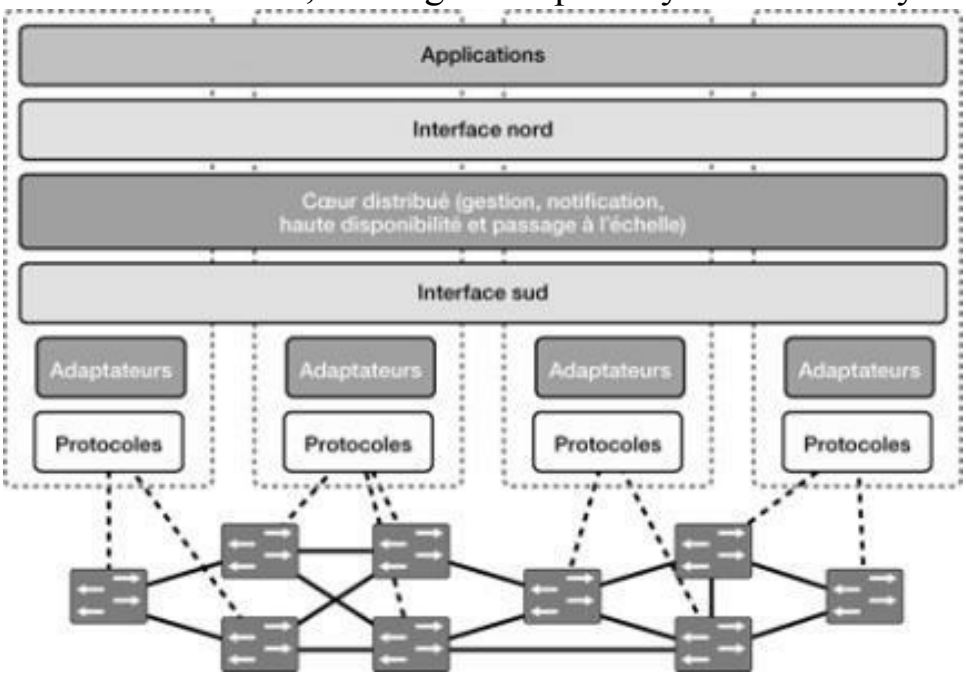


Figure 4.4
Controller architecture ONOS

Figure 4.5 still go a little further in the description of the functions that provide the intelligence to this controller. The internal functions include the management of the network using a graph from a database of knowledge, i.e. of contextual information.

Cassandra, a distributed database open source, is a management system designed to process very large databases and, in the present case, *in-memory, that is to say that the data are located in the RAM memory to achieve real-time functions. These data can be a huge size and divided on many servers.*

Cassandra provides a service highly available. The intelligence is provided by this knowledge base on which many treatments of type Big Data analytics can be exercised. This is the database which is distributed of the registers which gives the strong consistency of the global system

Zookeeper is an open source project to develop and maintain a server offering a distributed coordination highly reliable.

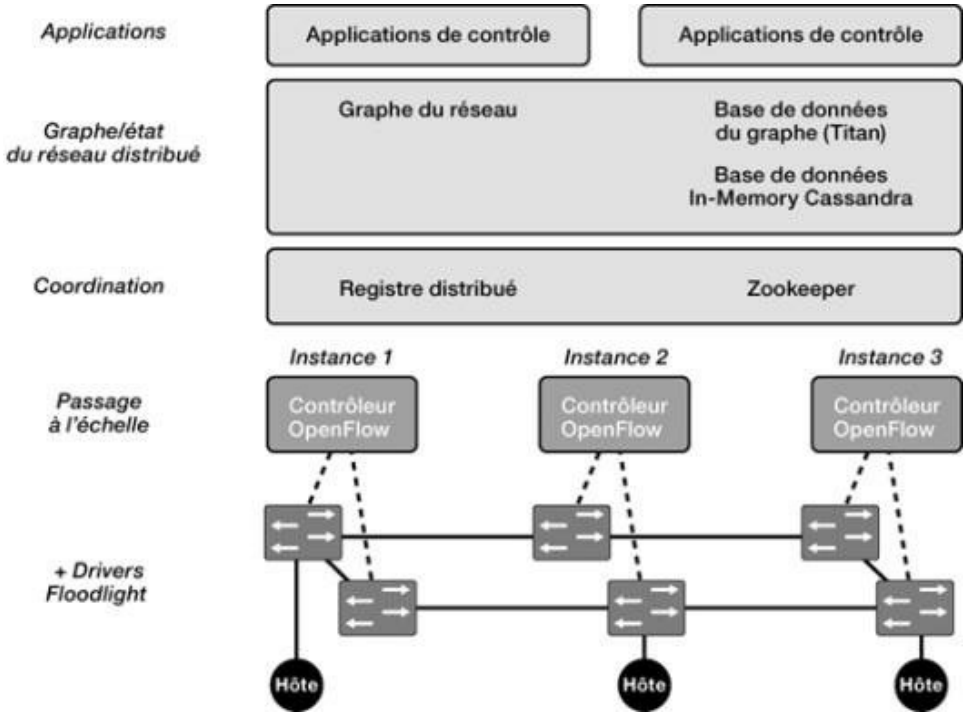


Figure 4.5

The modules providing the intelligence of the onos

The orchestrateurs provide the intelligence necessary to tasks for the establishment of a network and the allocation of resources to the VM in order to achieve a communication with the quality of service desired. Several open source solutions offer features more or less important. The result of this section briefly describes two projects, open-O (Open-Orchestration) and Napo (Open Network Automation Protocol).

Open-O is a collaborative project of the Linux Foundation Focused on the creation of an open source platform performing an orchestration of the carrier network. The main objective is to establish an open framework for organising the composite services end to end in the networks using the Emerging Infrastructure of type SDN and NFV.

Open-O has a consistent architecture to Mano (management and network orchestration) and adopts the modeling languages of services standards such as Yang or Tosca in order to ensure consistency and flexibility. Open-O consists of three functions of orchestration Main: GS-O (Global Services Orchestrator), SDN-O (SDN Orchestrator) NFV and-O (Network Functions Virtualization Orchestrator). The orchestrator provides micro services related to the different modules described in Figure 4.6.

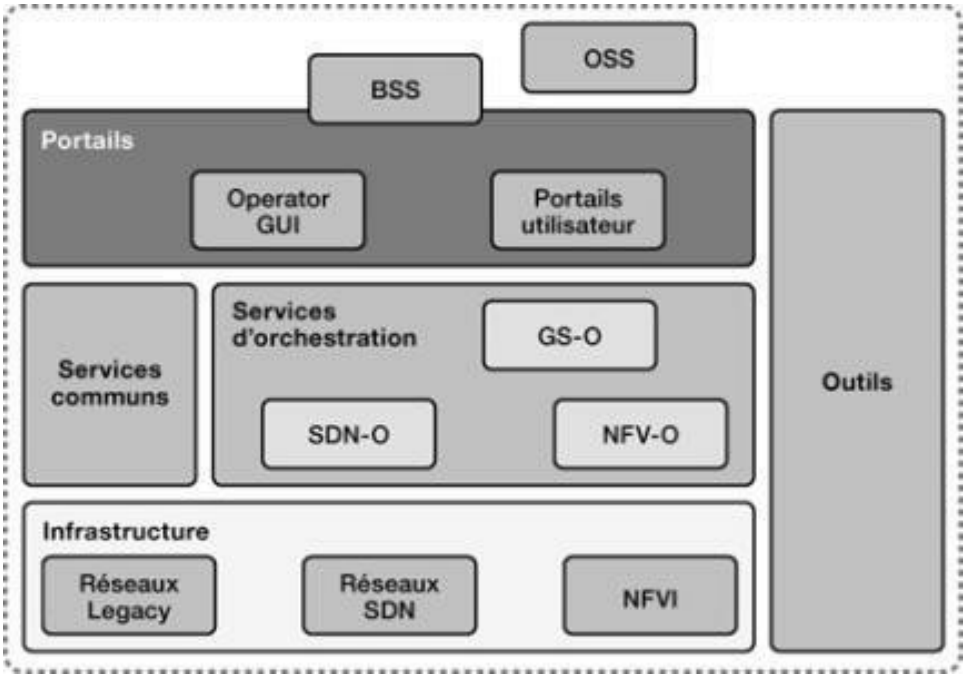


Figure 4.6

Architecture of open-O

Napo is an extension of the Open-O or, more exactly, the addition of open-O to another collaborative project devoted more specifically to the automation of the process of controls in the network. This second project is said ECOMP (enhanced control, orchestration, Management & Policy), a platform developed by AT&T.

ECOMP allows to put in place the services automatically, with their life cycle and their management. This platform consists of eight software covering the large architectures, a design environment, a platform of definition and a of execution and finally an environment to execute the programd logic. The design phase uses an approach in a closed loop based on the policies, which allows you to automate the process of the establishment of services and to manage by the suite.

The architecture of the Platform NAPO is illustrated in Figure 4.7. It includes a portal to manage the interface with the platform which has three major modules. The first, from ECOMP, carries out the design of the requested service thanks to a set of policies. The second module, from open-O, deals with the orchestration of the establishment of the elements to achieve the communication and control. The last module is involved in the life cycle of the Environment put in place. It contains modules of intelligence from data as well of the customers that the network.

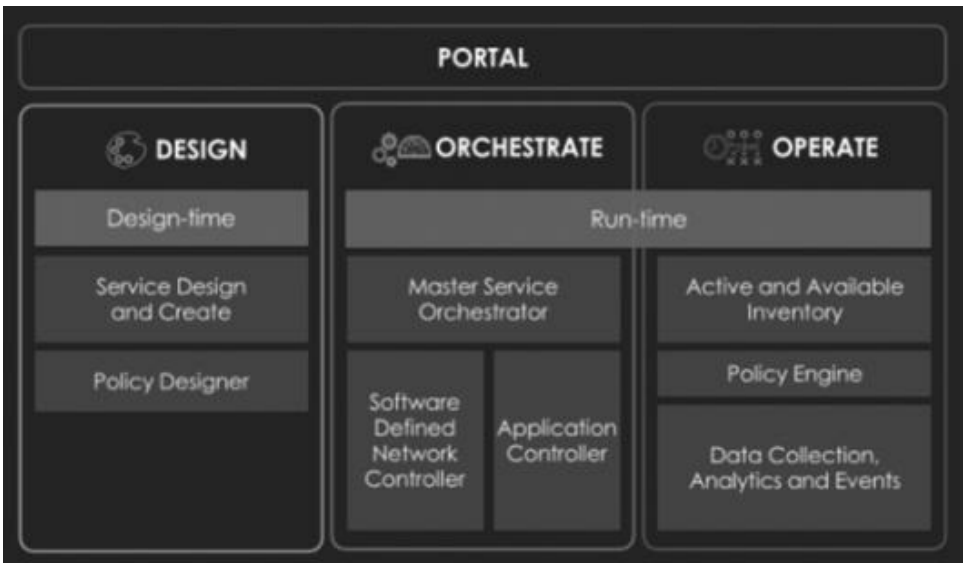


Figure 4.7

Architecture of the Platform NAPO (© Amdocs)

In conclusion of this section, one sees the arrival of a large number of open source software whose purpose is to automate the life of networks. This automation is done primarily by modules operating functions from the artificial intelligence, that it is the big data, of the machine learning, the deep learning or other solutions embedded in intelligent agents. The following sections examine in more detail the tools put in game to arrive at this automation.

Intelligent agents

Intelligent agents constitute a first category of tools including the introduction of large-scale modifies the environments of management, control and orchestration by making them more autonomous and more reactive.
The following section focuses on the reasons for this power and then on the way of constructing multi-agent systems.

Management of a complex environment

As networks become more complex, the management and control of these environments have become necessary for multiple reasons. In addition, network environments today are dynamic. Of the many and varied applications overlap and make the control of resources difficult. The gain statistics, that is to say what is gaining in processing the packets of statistical way in networks to transfer of packets, is undeniable. But if the flows exceed too strongly the capabilities of the network, a collapse of the performance is inevitable.
Network environments are by nature distributed, which complicates their control and management. In the case of small networks, less than a 50 nodes, the control is centralized in order to more easily access a intelligence thanks to the whole of the recovered data client sides and network and stored in the center.
The gigantism imposes a still controls more End. The sizing is a complex problem to apprehend, and we can say that there is no tool truly effective in this area, to the extent where the parameters to take into account in a network environment complex are difficult to assess. It was the choice between flow, response time, rate of use of the couplers of line and central units, the bit error rate, error rate package, rate of recovery and the rate of failure. In addition, the values of the mean, the variance and sometimes moments of higher order should be taken into account to have a real idea of the performance.
The engineering of the network design presents two major aspects, the qualitative and the quantitative. The qualitative often corresponds to a safety of operation, in the sense where it is possible to prove that the system is stable or that there is no State in which the network is blocked. The quantitative aspect refers to the values of the parameters listed in the previous paragraph, the purpose of a quantitative analysis being to show that these values are reasonable for a good functioning of the network.
Security is an important function for which can intervene in the intelligence. Today, a certain standardization allows you to better identify the problems, and a few large security classes have been defined, corresponding to the needs well expressed by users. One can easily imagine the contribution of intelligent tools in the world of security to recognize the anomalies, analyze, give a diagnosis, propose a solution and resolve the problem.
The management is also an area where intelligent agents can play a role engine. When a network is in a state of walking, it must be the administer, i.e. to be able to check all the operations that take place there in the network, since the faults up to the Accounting in passing by the security, the management

of performance and the management of the names of the users.
Several areas of specific administration already use of intelligent components, including the following:

- Configuration (Configuration Management);
- Security (Security Management);
- Faults (Fault Management);
- Audit of Performance (Performance Management);
- Accounting (Accounting Management).

The intelligence of the agents can come from different horizons. The diagram the more classical emanates from the Artificial Intelligence distributed, or IAD.
The artificial intelligence means that we put in place of a human being an intelligent agent to perform a task. The IAD is equivalent to a society of autonomous agents working in common to lead to a global objective. Among the many reasons that lead to the use of the IAD, these include the following:

- Taking into account different points of view. When the data are becoming more and more precise, inconsistencies can occur in the basis of rules. The ways of expressing the knowledge are different according to whether one is directed to the user, the developer or technician. In addition, two experts with the same expertise do not always arrive at the same result. The points of view are also often contradictory: one wants to focus on the costs, and therefore a cheaper system, while the other chooses to develop the advertising, and therefore a system more expensive. The use of the DST allows, by negotiation, to reach a compromise between different options.
- Adequacy In the real world. In a general way, it is always a group of experts, with the qualifications and the different specialties, who manages, by collaboration, to achieve an objective. In addition, if it seems easy to understand, and therefore to model, the behavior of individuals (the whole of their trade) thanks to the many sociological studies available, on the other hand, the functioning of the brain and the reasoning are less known.

For these reasons, the application of the Artificial Intelligence distributed is required little by little in the framework of the control of networks.

Multi-agent systems

An agent is an autonomous entity, capable of communicating with other agents, as well as to collect and to represent its environment. The whole of these agents in interaction form the multi-agent systems. We class these latest according to many criteria, such as the size of the agents, their number in interaction, the mechanisms and the types of communication, the behavior, the Organization and the control of each agent, the representation of the environment, etc.
From these criteria, there are two major categories of multi-agent systems:

- The systems of cognitive agents;
- The systems of agents reagents.

The cognitive agents have an explicit representation of the environment and other agents. They know take account of their past and operate according to a social mode of organization. The systems of this type have only a small number of agents. Several levels of complexity can be considered:

- The process in which the actors implement primitives of communication.
- The Communicating modules, which use of specialized communication protocols

(queries, commands).

- The co-operative agents, which involve notions of competence, mutual representation and allocation of tasks.
- The Intentional agents, who use of notions of intention, commitment and partial plans.
- The agents traders, which implement the resolution of conflict by negotiation.
- The agents organized, who act according to a regulation and social laws.

Agents communicate between them using a dedicated language and use of communication protocols. It is a communication intentional, which essentially comprises two types, the communication by the sharing of information and the communication by sending messages.

The communication between agents is carried out by the sharing of information when the current solution of the problem is centralized in a data structure overall, shared by all agents. This structure contains initially the data of the problem and is enriched during the resolution until the solution. It constitutes the only means of communication between the agents.

This type of communication is often designated by the model of table black, or *blackboard, developed in many publications. The agents settle and read information in a common data area, the black table, as shown in Figure 4.8.*

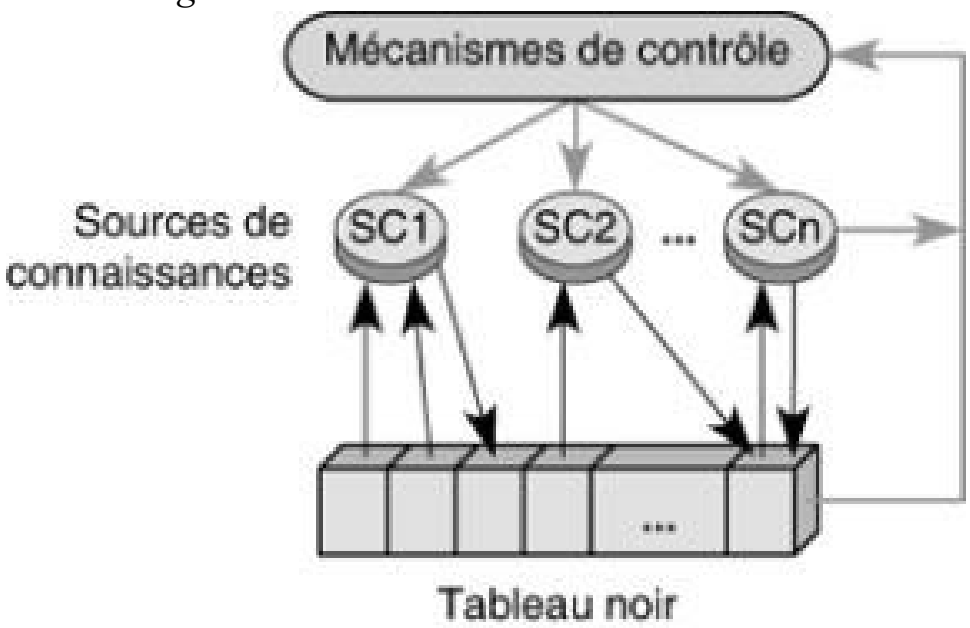


Figure 4.8
Operation of the black table

A system of black table contains the three basic elements:

- The black table itself, in which all the elements generated during the resolution are stored. This data structure is shared by the agents and is organized in a hierarchical manner, which allows to consider the solution in several levels of details.
- The agents, which create and store their assumptions in the table black. This are independent modules, called sources of knowledge. Their role is to cooperate to solve a given problem. The sources of knowledge are independent, since they are unaware of each other and do not react to events of change of table black.
- A control mechanism, which ensures the operation of the system in function of a certain strategy. Its role is, among other things, resolve conflicts of access to black table between the agents, the latter involved without being triggered. In the absence of centralized control, in effect, the sources of knowledge react

opportunistically, that is to say the best they can. This system of control works itself according to the model of table black.

The black tables have the advantage of offering a structuring and an automatic method (cutouts and hierarchy) in the way to approach a field of knowledge. They also show the interest of organizing the sets of rules in the systems to the rules of production. However, their lack of local memory does not authorize them not a real functioning multi-agent. As a general rule, multi-agents systems use a black table for each agent.

Multi-agent systems based on the communication by message are characterized by the total distribution at the time of the knowledge, partial results and methods used to achieve a result (see figure 4.9). Some languages of actors embody well this type of system. The communication can be done in point-to-point or by diffusion.

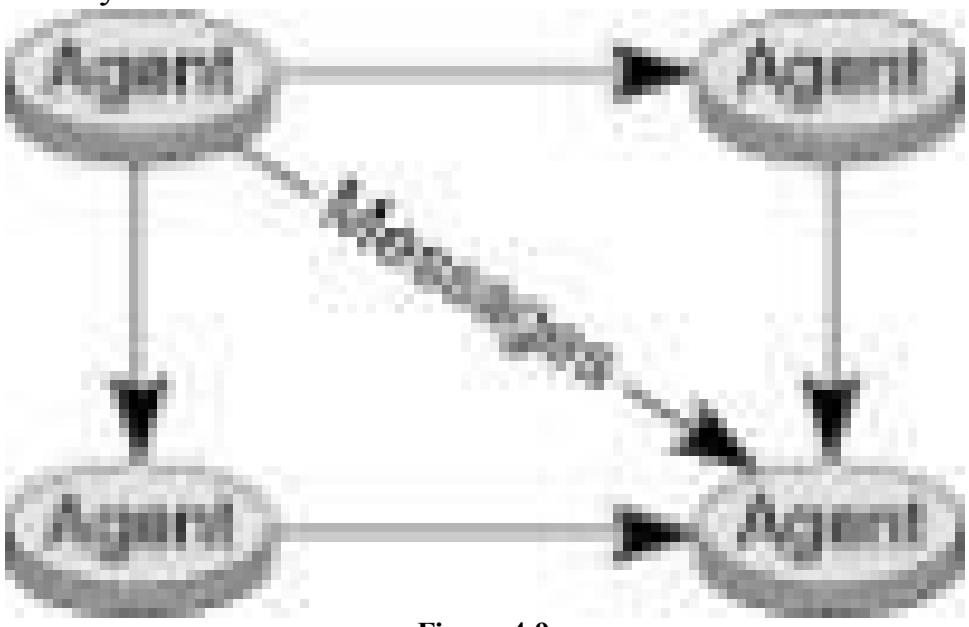


Figure 4.9
Operation of a multi-agent system

Such a system is structured around two components:

- **Local treatment.** to the reverse of the systems based on the table Black, knowledge are more concentrated in the same space, but divided between the different agents. An agent cannot manipulate that its base of local knowledge, send messages to other agents that he knows, that is called its closeness, and create new agents. At a given moment, the agents do not have a global vision of the system and only have a local perspective on the elements.
- **Sending messages with continuation.** When an agent sends a message, it specifies to which agent to the response to the message must be addressed. It may be the agent sender of the message, but also to another agent, specially created for the occasion.

The agents have a knowledge more or less precise of the other agents in the system. They must know and represent the skills of these agents, as well as the tasks that will be realized in a given moment, the intentions and commitments of the agents. This aspect of things poses the problem of the representation of this type of knowledge as well as that of its update.

The principle of the allocation of tasks constitutes one of the essential points of multi-agents systems cognitive impairment. A problem consists of a number of tasks carried out by agents which group together all partial solutions to obtain the overall solution (see Figure 4.10).

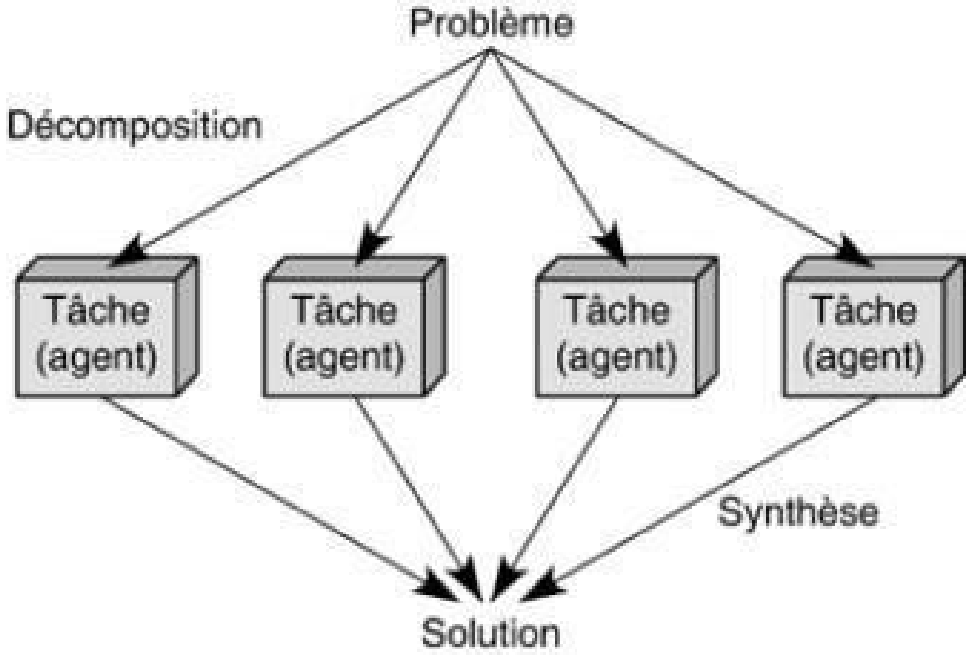


Figure 4.10
Resolution of Problems

To succeed the allocation of tasks, it is necessary to know the skills of each of the agents, decompose a problem in sub-problems, distribute the activities of resolution according to the agents and, if necessary, distribute to new these activities in a dynamic way.

The models for the allocation of tasks can be grouped into two categories:

- **Centralized allocation.** In this modular approach, an agent decomposes a problem in sub-problems and distributes them to the other agents who are under his authority. In this case, the actions are consistent, but there is a problem of reliability and extensibility. In addition, an agent is fully dedicated to the distribution of tasks. The system is therefore not used to its maximum capacity.
- **Decentralized allocation, or distributed.** Each agent is able to decompose its problem in sub-problems and thereby spread the associated tasks. All agents have the same weight in the decision-making. This type of allocation is suitable for applications having already a structure distributed. The reliability and the possibilities for expansion are better than in the previous model, but maintaining the coherence is more difficult to achieve.

The methods to decompose a problem can be of three kinds:

- **Static.** Each agent knowing the skills of other agents, the sub-problems can be attributed to agents the most qualified.
- **Dynamic.** The agents work together to distribute the sub-problems in a more effective manner.
- **Mixed.** Each agent is aware of the skills of the other agents, but this knowledge is updated on a periodic basis.

The degree of independence of Agents is based on the notion of intentionality. It can differentiate the intent in the action of the intention to perform an action in the future. In this last case, it is a purpose persistent. For that one agent has the intention to perform an action, it must be that it believes that the action is possible, that he intends to commit to achieve, that he believes that if certain conditions are fulfilled it can accomplish the action and finally that it is not seeking to realize all the consequences. However, it may be asked what happens when the action has been accomplished by another agent, when an agent has two intentions, or under what conditions an agent can abandon its intention.

Approaches to the cooperation between agents

There are currently two approaches for the cooperation between agents:

- **Voluntary.** The agents interact in a non-confrontational manner, that they have the same purpose or not.
- **Confrontational.** Agents can have similar goals, but divergent views, even opposite.

It is the voluntary cooperation which is of interest to this section because it reflects an interaction of agents for the distributed resolution of problems. One can achieve a planning for multiple agents or a distributed planning. In the first case, there is an agent coordinator who constitutes and realizes the plans and manages the conflict. As regards the planning distributed, each agent is able to produce its own plans, which will be merged for a overall result. We can also consider the distributed problem solving as a mutual assistance. An agent solves the problem, and another is critical of the solution obtained.

In the area of the resolution distributed, four modes of cooperation are possible when the concept of hierarchy is present:

- **Command.** An agent decomposes a problem in sub-problems, that it spreads between the other agents according to their skills. They resolve their sub-problem and return the partial solutions to the centralizing agent (see Figure 4.11).

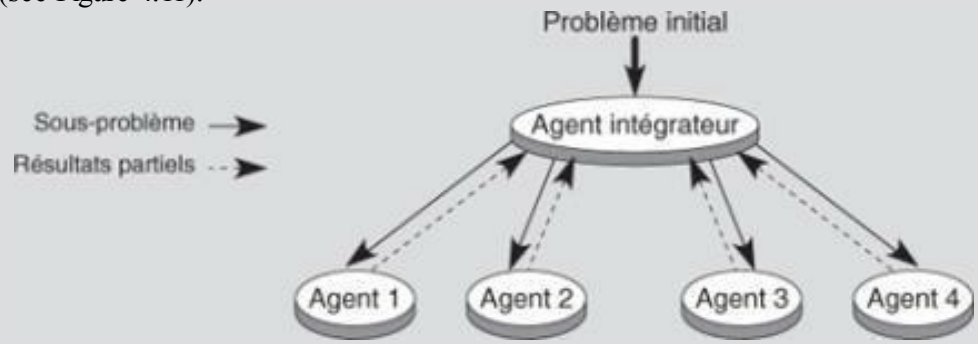


Figure 4.11

Operation of the command mode

- **Dissemination.** An agent still decomposes the problem but the interested agents by the resolution of the sub-problems decide whether they spend or not to the action. The results are then sent to all agents of the system.
- **Call for tenders.** An agent decomposes the problem in sub-problems, which it disseminates the list. The interested agents by a task Send an offer to the agent who has created this task. This last chooses a offers among those at its disposal. He then distributes the sub-problems to the officers that he has chosen (see Figure 4.12). The partial results are then returned, as in the command mode.

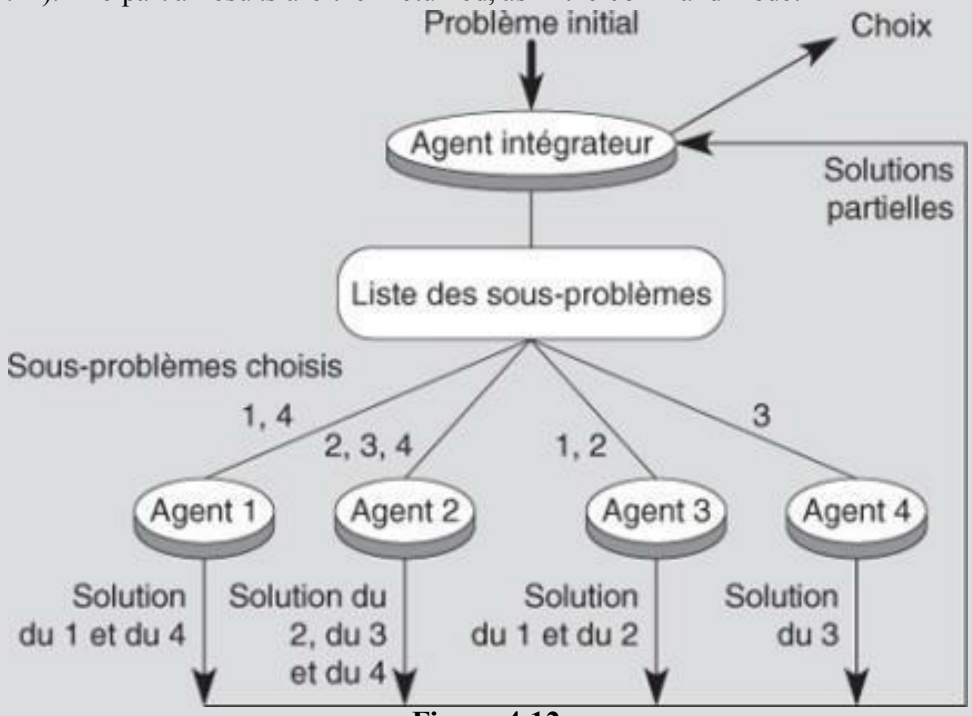


Figure 4.12

Operation of the mode call for tenders

- **Competition.** As in the other modes, the problem is decomposed into sub-problems by a central agent. Agents choose the task or tasks that they will achieve. The partial results are sent to the agent centralizing, who decides the solution among all results generated (see Figure 4.13).

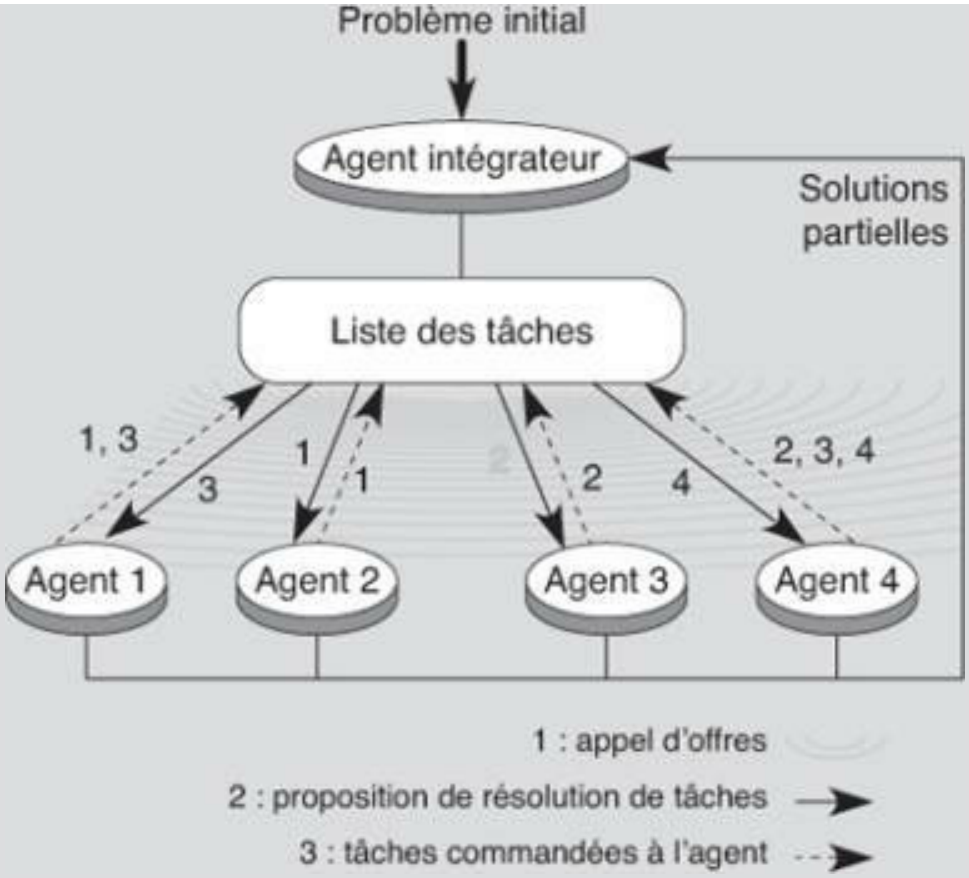


Figure 4.13

Operation of the competition mode

When the concept of hierarchy is not present, the agents can run according to a mode of sharing of tasks, by negotiation, for example, or exchange of partial results.

Talk about negotiation in the case of the sharing of tasks is to invoke a concept and a structure of interactions (the contractual network) falling heavily on the artificial intelligence distributed. The Contractual Network is composed of a set of agents, who can spend of contracts according to a fixed protocol. Agents must perform the tasks they can decompose. If they are not able to perform these sub-tasks, lack of skills, time or means, they can try to the sub-process by launching a call for tenders. The interested agents send then an offer to the requester agent, called *manager, who chooses the agent to which it assigns the task, and then a contract is established. This model is modeled on the protocols call for tenders for public contracts.*

The Contractual network represents an allocation of dynamic tasks and distributed by a call for tenders. This system has many advantages (distribution of decisions, extensibility, etc.), but remains marred by many problems (hierarchical design, high cost of communication, rigidity of the procedure, etc.).

Conflicts can be of several Orders: access to a resource, different solutions to a given problem, conflict of interest and goals, etc. to attempt to remedy these conflicts, we can:

- Establish a centralized control and therefore put in place an agent with a higher weight.
- Appoint an agent capable of establishing an arbitration between different points of view.
- Establish a negotiation between the different agents in conflict in order to lead to a situation that satisfies by a series of exchanges, of negotiations and compromise.

The systems of Reactive Agents

A reactive agent has not an explicit representation of its environment and may not take account of its past. Its mode of operation is simple and follows a set of pre-programd decisions, type stimulus-response. The organization is carried out according to a biological mode, and the number of agents present in such a system is very high. The communication is not intentional. The agents leave, for example, traces of their presence, or signals, which can be perceived by other agents. We then speak of communication by environment.

This type of agent is the result of the following postulate: of the interaction of a large number of simple agents may emerge complex organizations.

We can consider several levels of complexity for an Agent Reactive:

- Stimulus-response: simple reactions to the environment;
- Coordination of elementary actions: mechanisms of inhibition, relations between elementary actions;
- Reactive cooperation: mechanisms of recruitment between agents, aggregation of elementary agents;
- Reproduction: The mechanisms of reproduction of reactive agents;
- Organization of Reactive Agents.

The écorésolution is a technique for the resolution of problems based on the use of agents reagents. The resolution of problems is here considered as the result of a set of interactions. This design is opposed to the classical approaches in resolution of problems, such as the exploration of spaces of states, which poses problems of combinatorial explosion.

The approach to the resolution of problems distributed based on a design radically different: that of the appearance of the configurations as stable states stationary or a dynamic system, whose evolution is due to interactions from the behavior of small agents fairly simple.

In the conventional systems, all data are located in the statement, the system merely to find how to move from the initial configuration to the final state. On the contrary, in the phenomenon of the écorésolution, the determination is only local. The agents are characterized by behavior of satisfaction, aggression and leak. The problem itself is defined by a population of autonomous agents, which seek to meet. The final result is the consequence of an interaction non-deterministic. The model defines the combination of behaviors.

Behavior of agents

In a way almost general, we can characterize the behavior of the agents by a set of elements:

- A condition of local satisfaction, which describes a stable state for the agent, taking into account the information available to it. This condition is not based on objective criteria, but on the beliefs and the local characteristics of the agents.
- A local condition of rejection, i.e. a set of constraints describing the situations that temporarily rejects an agent (those that he seeks to escape).
- A survival reaction, i.e. a set of actions (satisfaction, aggression, leak) that are performed for the agent is able to approach its condition of satisfaction in avoiding the harmful situations.
- A function of energy cost, which allows you to choose the action of lower cost to achieve.
- A information.

This method avoids the combinatorial explosion Classic when the resolution of problems, but, in return, the solution is not necessarily the best. The écorésolution of problems can be used in a privileged manner for problems structurally distributed or in the framework of a universe scalable.

In conclusion of this section, it can be argued that the study of the learning, the real-time or distribution corresponds to a need in the area of the management network. However, it is the last point which seems to be the most interesting, in the measure where the multi-agents systems constitute a sort of generalization of the techniques of expert systems. Of this fact, they bring an added value to the conventional systems of artificial intelligence in proposing a new type of architecture, involving both the communication that the internal reasoning. The aspects of opening and distribution make it interesting for a network management system.

As regards the aspect of real time, the approach adopted by the Artificial Intelligence seems less clear. Yet, it is an essential step in the design of a system for the administration of the networks. In effect, the response time to a fault, for example, must be minimal, even if this is not of the same importance as in the systems of aid to the decision or the command.

The learning is still the Achilles heel of the systems to knowledge base. As long as these will not improve, increase and refine their knowledge through their own experience, they will depend on the good will and the availability of the experts and of the quality of a manual update.

The network agents

A network agent is a software component that uses either a service related to the network (e-mail, file transfer, Web, etc.), or the network itself, as in the case of mobile agents which are discussed in the next section.

A network agent can be defined in many ways. This section considers it as a software component that acts on behalf of its user. It is obvious that this definition is much narrower in scope than the one given by the researchers in artificial intelligence, who are considering the agent as a software component able to render a service, with a certain possibility of reasoning and communication.

The network agents generally provide simple tasks. For example, they filter e-mail messages by using a set of key-words. As well, the agent may be confused with a classical program. Some agents are much more developed. The user may request an agent to arrange a meeting, for example. To do this, the agent must break this task into sub-tasks and cooperate with agents of different participants.

There are three main categories of agents network:

- The agents the Internet;
- The intranet agents;
- The agents assistants, or desktop.

The following sections examine these three categories of agents, who can themselves be decomposed into sub-categories.

The Internet agents

The agents the Internet come mainly from the applications developed for this network. There is the following agents:

- Research officers of the Web: Provide a user Search Services on the Web.
- Server Agents of the Web: resident on a specific Web site to provide services.
- Agents of Information Filtering: filter the information according to the criteria specified by the user.
- Agents of Documentary research: Return a customized set of information corresponding to the request of the user.
- Agents of notification: indicate to a user of the events that may be of interest.
- Agents of Service: provide specialized services to users.
- Mobile Agents: move from one place to another in order to perform the tasks of a specific user.

The intranet agents

The intranet agents also come from the Internet applications, but are customized for a private environment. Four types of agents are recognized:

- Agents of customizing co-operative: allow the automation of the workflow to the inside of a company.
- Automation Agents: Automate the workflow of a business.
- Database Agents: Provide services agent to the user databases.
- Brokers agents of resources: Realize the allocation of resources in client-server architectures.

The agents assistants or desktop

The agents assistants are, as their name indicates, the software components able to provide assistance

to a user in the use of a product. These agents Assistants are often called desktop agents, because they meet essentially as agents of interface in the office suites marketed for the general public.
The three major categories of agents Assistants are the following:

- System Agents: provide assistance to the user so that the user can use the operating system.
- Agents of application: provide assistance to the user for that it employs correctly a particular application.
- Agents of Software Suite: provide assistance to the user for that it can work with applications correlated.

Mobile agents

A mobile agent is a software agent that can move between several points. This definition implies that a mobile agent is also characterized by a model of basic agent.
In addition to the base model, each agent software defines a life cycle model, a calculation model, a security model and a model of communication. A mobile agent is also characterized by a model of navigation. The mobile agent can be implemented using the mobile codes or the remote objects.
Examples of the first category come from the agents of TCL (Tool Command Language), a language agent, and Telescript. The second category is represented by the Aglets (see below).
To use the mobile agents, a system must incorporate a platform agent. This platform must allow all the features whose officers have need, in particular the model of navigation. For the cycle of life, it must define the services of creation, destruction, starting, suspension, stop, etc. The calculation model indicates the means of calculating the agent. The security model describes the way in which agents have the right to access the resources of the network and to devices. The communication model defines the modes of communication between agents and between an agent and another entity, such as the network. The navigational model will load all the transport of the agent between two entities of the network.
The integration of the Code by a machine Java is already very powerful, but of Java Chips could further strengthen the integration of the agents in the current systems. New software packages, as Jini, could further improve the power of mobile agents. The size of these agents depends on their functions. For example, mobile agents of control of the security may be very large and contain several thousands of lines of code. It should be noted that these mobile agents can increase their functions by loading the code *via* the network.
For that the mobile codes can impose, they must be standardized. The OMG has already proposed a first standard independent of the platform using strongly CORBA.
Another way to introduce the mobile agents is to define them as software processes that substitute for the user to interact in any freedom in the network with servers. The functional autonomy of the mobile agents, which is one of their strengths, solicits the minimum of communication resources. The telescript language, general Magic, constitutes one of the first tools for the development of mobile agents, to the sides of the Java language.
Mobile agents Telescript incorporate the instructions of the user concerning a specific task and move to places necessary to the execution of their task, for example the purchase of tickets. The sites which approve these mobile agents act as hosts and ensure the safety and security of the implementation of the programs of the agent.
Many other examples of the design and use of mobile agents in environments of telecommunications are mentioned in various publications. In particular, the virus can be considered to be of such agents.

Three targets are found in these jobs:

- The characteristics of the language to provide the needs of mobile agents;
- The implementation of these needs from extensions of the operating system to take advantage of the benefits of mobile agents;
- The system of mobile agents, considered a specialized application running above the operating system.

In summary, the mobile agent systems exist to support either on the Java classes, either on the scripting languages interpreted, either on the existing services of the operating system. There are constraints to limit the possibilities and to check the mobile agents. These constraints come from the following points: Security, identification, portability, mobility, communication, management of resources, control and management of data.

To compare the approach agent with the client-server approach, it should be recalled that, in the client-server architecture, specialized programs are implemented in order to meet the more customers possible. The client process runs the more often on a remote machine and communicates with the server to perform a task. This approach can generate a strong increase of the traffic on the network, which, depending on the type of network, can happen to clog the network. The concept of mobile agent proposes to bring the client of the source and to reduce the traffic generated by moving the requests of clients at the level of the servers.

The active networks

The active networks are similar to the other networks to transfer of packets and frames. The base unit transferred in this network is the packet. The nodes have the mission to examine the different fields in the packet, which have a location perfectly determined. In particular, the address field allows you to determine the output port. It is a VM (Virtual Machine), which interprets the different fields in the packet. It may be considered that this VM is an interface package, what is conventionally called a network API, or NAPI (Network Application Programming Interface). For an IP network, the API IP is the language defined by the syntax and semantics of the IP header. In networks that we know, the VM is fixed, and the language used primary.

We can say active networks that the nodes provide a Network API programmable. If one considers that, in an IP network, the header of the package provides the entries of the VM, we can define an active network as a network in which the nodes have a VM that runs the code contained in the header of the packet.

Many categories of active networks can be defined from the following attributes:

- Power of expression of the language, which determines the degree with which the network will be able to be programd. The language can go from simple orders to languages very sophisticated. The more the language is simple, more the treatment time is short. Conversely, the more the language is powerful, more of the on-measure can be implemented.
- Possibility to define a stable state from previous messages in the same stream, so that the speed of execution can be increased without that it is necessary to redefine a state of the VM.
- Granularity of control, which allows you to modify the behavior of a node for all the packets that pass through it, regardless of the stream to which he belongs, or, at the other extreme, not to modify the behavior of the node that for the single package in the course of the treatment. All intermediate cases are possible, in

particular the common behavior on a same stream or on a same set of waves.

- Means of giving the orders of programming: it is possible to consider that the orders to the active nodes are given by the specific packages, for example of signaling packets, and no longer by a order indicated by a language more or less evolved in the header contents of the packets.
- Architecture of the nodes, to look at what level of this architecture speakers the commands or, in other words, to what level is the programming interface. The architecture can influence the choice of the software and the hardware. In particular, she can use of reconfigurable processors, to levels more conceptual or less high.

The features of the nodes of the active networks are shared between the runtime environment and the operating system of the node. Figure 4.14 illustrates such an architecture of an active network.

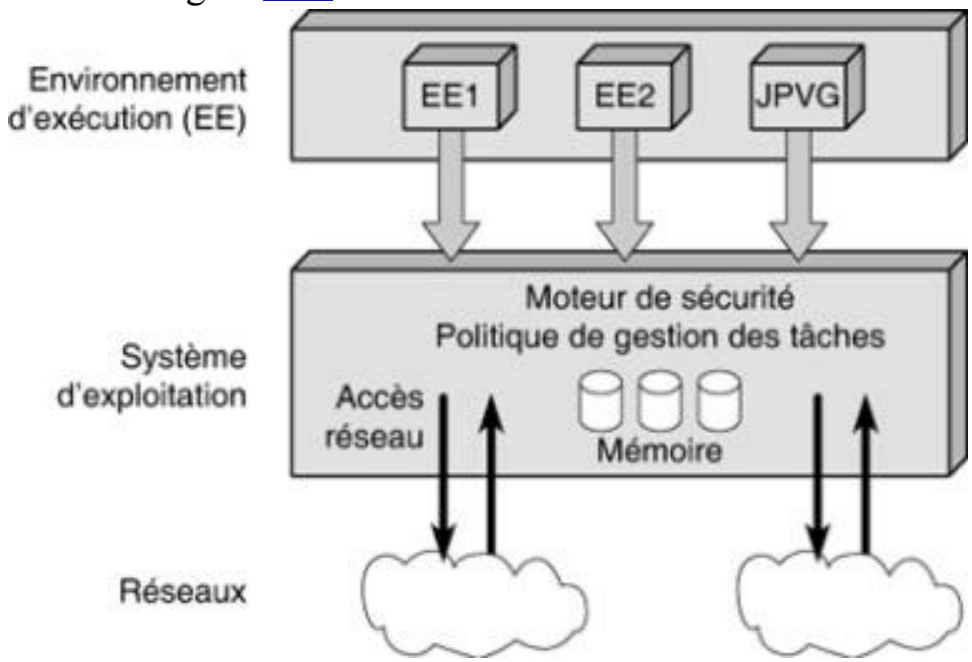


Figure 4.14
Architecture of active networks

It is possible to send commands on the Runtime Environment through an encapsulation protocol, called ANEP (Active Network encapsulation protocol). The header of a packet ANEP contains a field for the identification of the type of package. Several execution environments that can be present in an active node, the address of the node requires a complementary address.

The existing interfaces include:

- The interface to access the environment of execution;
- The interface between the runtime environment and the operating system of the node;
- The access interface to the operating system of the node.

Programmable networks

Programmable networks form a part of the active networks, whose role is to develop a set of software abstractions of the resources of the network allowing access to these resources by their abstraction.

The objective of these networks is to make the programmable nodes to adapt them to the demands of users and services. The programming commands, which can be carried out both by a network of signaling that by user packages containing control programs, can attack the nodes at different levels of

abstraction. The IEEE has launched a working group intended to normalize these interfaces.

The Reference Model P.1520

Figure 4.15 illustrates the different layers of the reference model P.1520, which is one of the proposals of the ISO to define the levels of programmable interfaces. In other words, the architecture P.1520 offers a programming of active nodes at the level of the four interfaces, named V, U, and CCM.

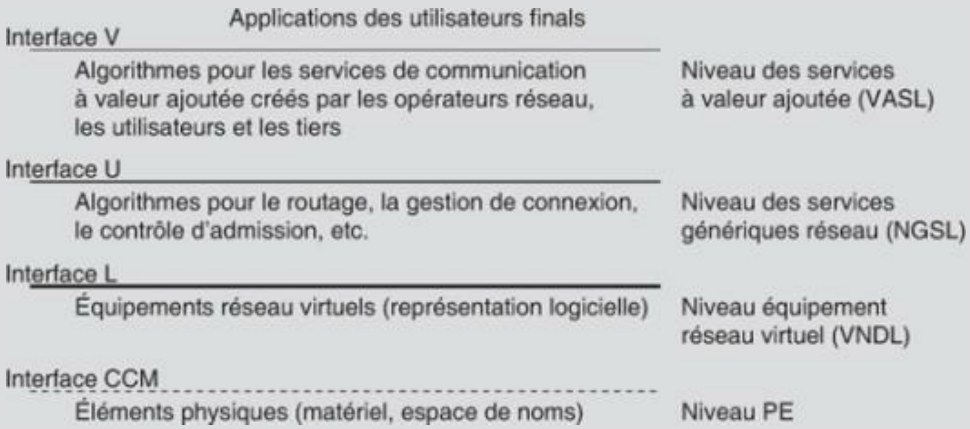


Figure 4.15

The Reference Model P.1520

At the top of the architecture is located the level VASL (Value-Added Services Level), whose entities are the algorithms of end-to-end which confer the added value to the services performed by the lower layers. These services can be of type management of waves in real time, synchronization of streams of different media, etc.

The underlying level, NGSL (Network Generic Services Level), contains the entities which are the algorithms working on the functions of the network layer, as the algorithms of establishment of circuits or virtual paths in an ATM network or the opening of a stream in an IP environment. This layer also must have a knowledge of sub-networks. In the same vein, the structuring in VPN is part of this layer.

The level VNDL (Virtual Network Device Level) Has of the entities which are logical representations of hardware and software resources of the components of the network.

The last layer is composed of the physical components of the network. This PE layer (Physical Element) manages the hardware and the space of physical address to achieve these physical elements. There is ATM switches and IP routers.

Associated with these different layers, four types of interfaces have been defined:

- The interface V, which allows an application to access the Layer VASL. It provides a large set of APIS to write software for top level allowing to add more value to the applications.
- The U interface (upper interface), which is located between the layers VASL and NGSL and which presents the features of the NGSL layer to layer VASL. The U interface works on the generic services networks and on the properties of the openings of the connections between applications or between clients and servers.
- The interface the (Lower Interface), which allows you to use the resources of the VNDL layer by layer NGSL. The interface The allows you to treat the member of network resources, such as a switching table ATM or an IP routing table.
- The interface CCM (Connection Control and Management), which enables access to the physical resources. This interface is not a programmable interface, to the difference of the previous, but a set of protocols for the exchange and the control of information to a low level of the architecture.

Figure 4.16 illustrates this architecture for an IP environment.

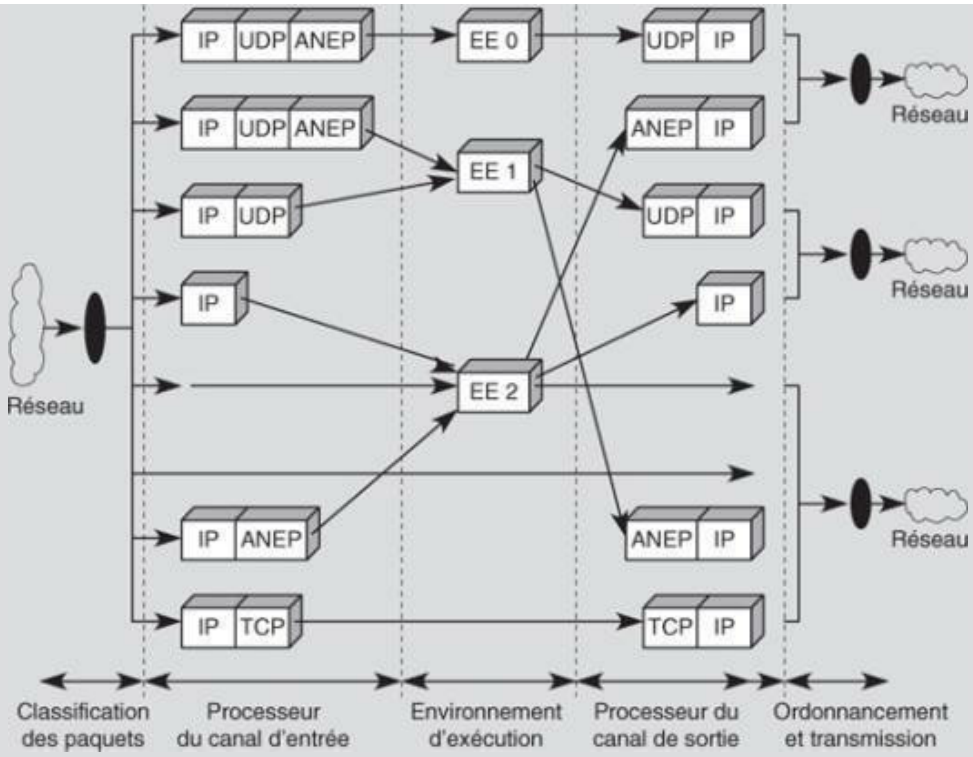


Figure 4.16

Application of the Model P.1520 to an IP network

The autonomous networks

The concept of active network and programmable is in the process of being replaced by the concept of autonomous network. An autonomous network is a network that does not need to center of management or control center to take its decisions. An autonomous network is therefore a network which can decide for itself how it behaves. It is a concept that has been introduced for the NGN (Next Generation Network), the objective of which is to replace all existing networks by integrating all the communication media.

An autonomous network must be able to self-manage, detect problems, repair itself the faults and to autocontrôler when no communication is not possible.

The elements of network must participate themselves in the construction of a stand-alone network with various properties, such as the optimization of resources, the automatic recognition of the environment (context aware) and the automatic organization of the security. The objective is to understand how to learn the good decisions, what influence have the different network elements and, more generally, how to optimize the behavior of the network. The tools to achieve this type of autonomous system come from the multi-agent systems, introduced earlier in this chapter.

The self-organization of an IP network is first by the need to have a global view of the network and to understand the consequences of an event that occurs in the network. Then, the network must be able to respond to the event.

The autonomous networks can develop on other networks that IP for objectives very specific, as networks with critical missions or interplanetary networks where the Control Center puts so much time to communicate with the probes that it is impossible to take decisions in real time.

The autonomic networks

The autonomic networks are, by definition, of autonomous networks and spontaneous. They corresponding to the networks defined previously, but by adding the property of spontaneity, i.e. to real time: The process is able to react independently and in an acceptable time frame for the process.

A first definition of autonomic networks is shown in Figure 4.17.

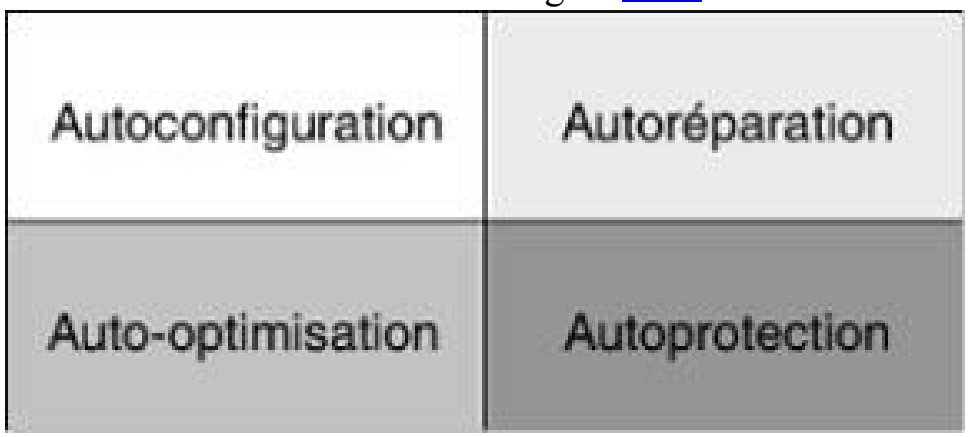


Figure 4.17

Definition of a network the Autonomic

The autonomic networks are able to autoconfigure to dynamically adapt to changes in the environment, self-optimize to offer a operational efficiency always optimized, autoréparer to acquire an important reliability and autoprotéger to secure the resources and the information passing through.

To achieve these different functions, the autonomic networks must have a number of attributes:

- Know their internal state (self aware);
- Know their environment (environnement aware);
- Be able to understand the characteristics of their performance (Self Monitoring);
- To be able to change their internal state (self adjusting).

To achieve these objectives, it must change the architecture of networks. The autonomic networks thus propose a new architecture to four plans, which adds a map of knowledge to the three usual plans that are the data plan, the plan for the control and the management plan.

Figure 4.18 illustrates the new architecture of autonomic networks. The objective of the plan of knowledge is to gather the knowledge of the network and to obtain for each point of the network a vision more or less comprehensive.

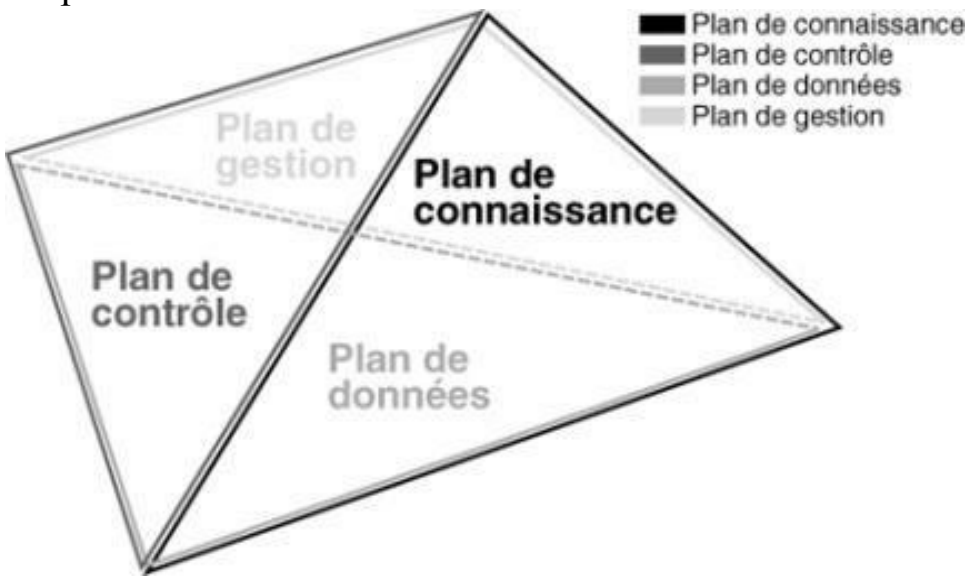


Figure 4.18

The architecture of autonomic networks

The plan of knowledge (knowledge Plane) has for objective to actuate control algorithms which are located in the Control Plan (Control Plane), which controls the data plan (Data Plane), which corresponds to the first four layers of the classical architecture of the networks. The management plan (management plane) is responsible for administering the other three layers.

The objective of the plan of knowledge is to make the network more intelligent allowing him to

understand his behavior, which has given rise to a new generation of protocols. Up to now, each control algorithm (routing, quality of service, security, reliability, etc.) should go search by itself the elements that he needed. For example, a routing algorithm such as OSPF research member of the upstream and downstream linkages up to the inputs and outputs of the network. These information are exploitable by other algorithms, as a control algorithm of the quality of service or the congestion or even of the admission in the network. Using a Plan of knowledge, this information is located in this plan, and each control algorithm can go at the moment where it is needed.

In the longer term, the standardized protocols should be modified to take account of this plan of knowledge. Another advantage of the plan of knowledge is the possibility to use in a control algorithm of the information which would not have been able to be taken into account in the normal algorithm.

View Located

It is obviously important not to make the network too heavy to force of transport of knowledge of any kind and in large quantity. For this, it is possible to use the "views located". These latest come from the world of artificial intelligence and indicate the taking into account of knowledge that are located in a view to determine. We can thus define a view located by the number of hops necessary to go looking for the information, for example to one or two hops, etc.

Figure 4.19 illustrates a view Located in a jump.

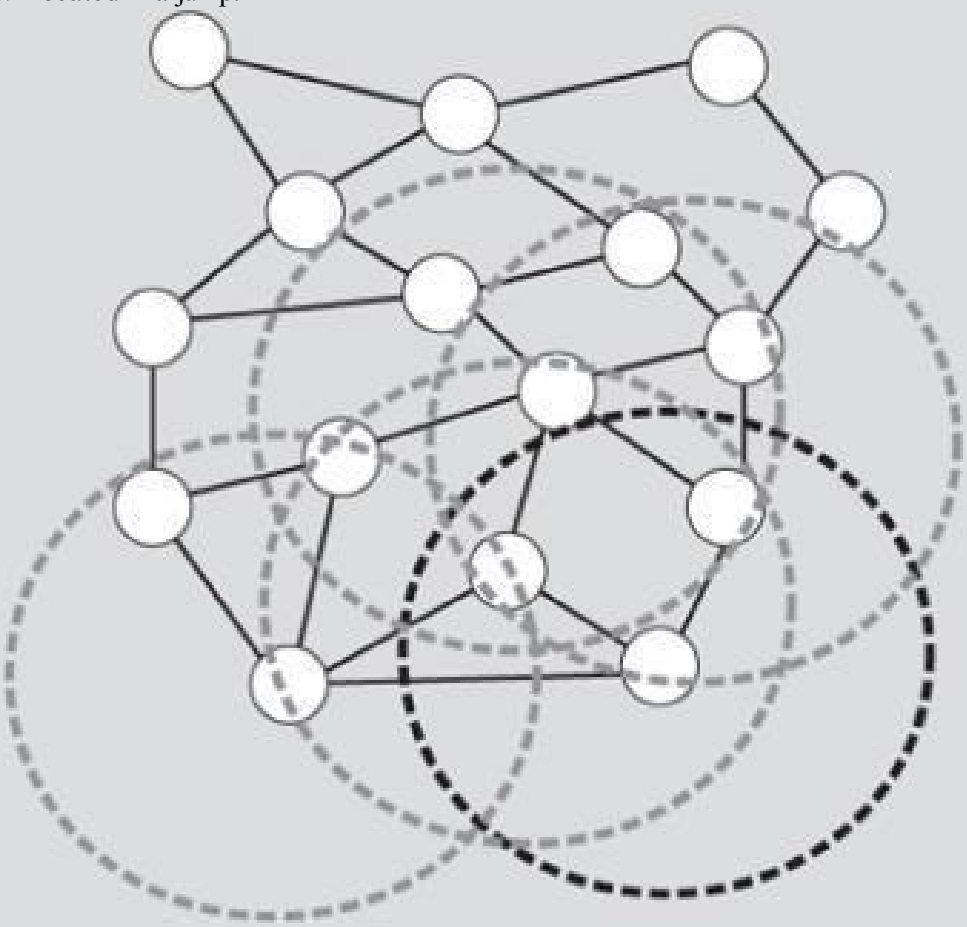


Figure 4.19

View located in a Jump

The view Located in a Jump presents the advantage of very well go to the scale since the transportation of knowledge is limited. This does not preclude the knowledge to disseminate in the network because a point receiving a knowledge integrates it into its own knowledge and distributes to its neighbors. This knowledge is however less fresh as if the view located was a little larger.

An optimization of the view Located must therefore be carried out on the basis of knowledge of which need control algorithms. It is necessary to respond to questions which? Where? And when? "Which" reference the necessary knowledge to the optimization of an algorithm; "where" concerns the scope of the view located; "when" indicates the refreshments necessary for that knowledge be exploitable. On the basis of these different parameters, it is possible to generate views located a little more complex than the only definition to a jump, the information links and all the seconds.

The autopilotés networks

Until 2008, the networks have been remarkably static, and a network engineer often had to be present to take things in hand as soon as a

problem arose. For very large networks, several dozens of network engineers could be necessary to ensure the maintenance and manage the various problems and varied that occurred.
The purpose of Networks autopilotés is to guarantee a automatic steering of the network through a software capable of managing control algorithms in a coordinated way and to optimize the operation of the network.
At the beginning of the years 2000, a first attempt has been to use of programmable networks and active networks. The research has not been totally conclusive for reasons of security and cost. A new generation has been launched from 2005 with the Autonomic networks presented in this chapter.

Conclusion

Slowly but surely, the intelligence arrives in the networks. This intelligence encompasses the communication, the reasoning and the decision. Multi-agent systems provide the main base of this intelligence, which allows you to take control or management decisions when it is time. We are still in the early stages, and yet the intelligence is omnipresent since the beginning of the years 2010. It starts to become well established with intelligent agents that implement the Big Data, machine learning and well of other components from the Artificial Intelligence.
Security is one of the first beneficiaries of this intelligence. For example, a Smart Component, Type neural network, is able to analyze the seizure of a person on a keyboard and to stop the communication in the case of seizure not recognized.
The intelligence in networks became a reality after having been very long a vast field of research. Automatic pilots of network develop from platforms such as NAPO or other solutions using, for example, the networks of neurons that are counted in millions.

Part II

The network protocols

This part examines in detail the first four levels of the architecture of networks as well as the infrastructure necessary to convey the data.
The infrastructure level contains the physical elements or which radio must convey the data packets, such as the phone cable, coaxial cable, optical fiber, or the relay antennas.
The Layer 1, or physical layer (item level binary), allows the transmission of a binary element from one machine to another. Depending on the desired flow, remoteness and many other physical characteristics, the choice of media must be weighed carefully.
The Layer 2, or layer connection (frame level), allows you to carry the binary element by placing it in a frame. The frames, detailed in Chapter 6, derived mainly from the Ethernet technology.
The Layer 3, or Layer Network (packet level), concerns the transport of packets from one end to the other end of the network. The Internet Protocol (IP) is imposed in this function and it has supplanted all other. It is described in detail in Chapter 7.

5

The physical level

The Physical layer determines how the binary elements are transported on a physical media. In a first time, the information to be transmitted are coded in a suite of 0 and 1. For the transmission to the receiver, these bits 0 and 1 are then introduced on the support under a specific form, recognizable of the receiver.
The first section is dedicated to the physical media itself, which is part of the infrastructure. Several components of the physical layer are defined in this level, such as modems, multiplexers, hubs, etc. This chapter explains these basic elements and introduced the architectures of the physical level which will be discussed later in the book.

The physical medium

By physical medium, it must hear all of the physical components to transmit the binary elements suites, 0 and 1, representing the data to transmit.
The nature of the applications conveyed by the network can have an influence on the choice of the support, some applications that require, for example, significant bandwidth and, by the same token, the adoption of the optical fiber. The coaxial cable allows you to also transfer of flows important binaries, even if these remain lower than those offered by the optical fiber.
Today, technological advances make the use of the pair of twisted wires well adapted to speeds of 10 to 100 Mbit/s, or 1 Gbit/s on shorter distances. Its ease of installation by report to the coaxial cable and its very low price make it both more attractive and more competitive.
The optical fiber is present in all wiring systems proposed by manufacturers, in particular on the connections between technical premises. It has the advantage of a small footprint, the space is very important required by the other physical media that can become binding. Another advantage of the optical fiber is its immunity to noise and electromagnetic interference. In some disturbed environments, transmission errors may indeed become unacceptable. Similarly, its natural protection against the listening makes it attractive in the sectors where confidentiality is important, such as the army or the bank.
There is a use of more and more frequent of the twisted pair. Technological advances have enabled him to push back its theoretical limits by the addition of electronic circuits and to achieve significant flows at prices significantly lower than those of the coaxial cable. The twisted pair is also more simple to install that the coaxial cable, as much as she can use the infrastructure put in place for a long time for the telephone wiring. The twisted pair allows you finally to reconfigure, to maintain or to change the network in a simple manner.

The pair of twisted wires

The pair of twisted wires is the support of transmission the most simple. As shown in the Figure 5.1, it is constituted of one or several pairs of electrical wires arranged in spiral. This type of support is appropriate to the transmission as well analog than digital.

The twisted pairs can be shielded, a metal sheath completely enveloping the metal pairs, or not shielded. They can also be "écrantées". In this case, a metal band surrounds the son.

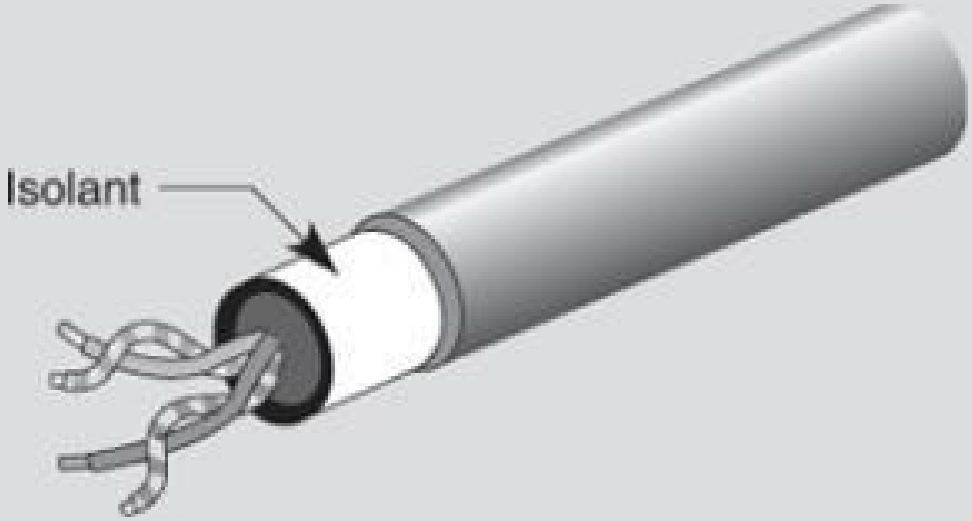


Figure 5.1
Twisted wire pairs

Very many debates have taken place on the advantages and disadvantages of the shielding of these cables. We can say, simplifying, that a shielded cable should be capable of better immunize the signals transported. The disadvantage of the Shielding however is that it requires the grounding of the whole of the equipment, since the physical media until the device. It is therefore necessary that any the connection string of land is correctly performed and maintained. In other words, a shielded network must be of very good quality, otherwise he may behave less well that a network without shielding, much less expensive.

The shielding may be global to the whole of the twisted pairs and the We then speak of screen or individually for each twisted pair.

The names of the types of twisted wires are introduced in the standard ISO/IEC 11801 who regularly evolves with amendments. They are indicated in the table 5.1.

Former name	New Name	The cable shield (screen)	Shielding of the individual pairs
UTP (Unshielded Twisted Pair) - pair torsade non-shielded	U/UTP	No	Non
Shielded Twisted Pair (STP) - Shielded Twisted Pair	U/FTP	No	Yes
FTP (foiled twisted pair) - Twisted Pair écrantée	F/UTP	Screen	Non
SFTP (Shielded and foiled twisted pair) - Twisted Pair écrantée and shielded	F/FTP	Screen	Yes
S-FTP (Shielded and foiled twisted pair) - Twisted Pair écrantée and shielded	SF/UTP	Screen and braid	Non
S-STP (Shielded and Shielded Twisted Pair) - Twisted Pair superblindée	S/FTP	Braid	Yes

Table 5.1 • Names of twisted cables

Metallic wires are particularly suited to the transmission of information over short distances. If the length of the wire is little important, of a few hundreds of meters to a few kilometers, speeds of several megabits per second can be achieved without error rates unacceptable. On shorter distances, we can obtain without difficulty of flows of several tens of megabits per second. On distances even more short, easily reached a few hundred megabits per second. A distance of the order of 100 meters allows to move the flow to several gigabits per second.

Standardization in the field of cables is carried out by the group ISO/IEC JTC1/SC25/WG3 at the international level and by national organizations such as the EIA/TIA (Electronic Industries Association/Telecommunications Industries Association), in the United States.

The main categories of cables defined are the following:

- Category 3 (for a Strip width up to 16 MHz). Used for Ethernet networks of base, this category is abandoned in favor of category 5.

- Category 5 (for a Strip width up to 100 MHz). Used for Ethernet networks at relatively high speed: 100BaseTX and 1000BaseT.
- Category 5E (Enhanced, for a bandwidth of 100 MHz, but with the technical characteristics which have evolved considerably).
- Category 6 (for a Strip width up to 250 MHz).
- Category 6A (for a Strip width up to 500 MHz). Used for the 10GBaseT on a hundred meters.
- Category 7A (for a Strip width up to 1 000 MHz). Used for the Ultimate versions of Ethernet to 10 Gbit/s and 100 Gb/s.

It is possible to compare the twisted pairs in function of their next, that is to say the loss of a part of the energy of the signal due to the proximity of another circuit and its weakening.

The coaxial cable

A Coaxial cable consists of two cylindrical conductors of the same axis, the soul and the braid, separated by an insulating material (see figure 5.2). This allows you to limit the disturbances due to the external noise. If the noise is important, a shield can be added. Although this support lost ground, including in relation to the optical fiber, it still remains very used.

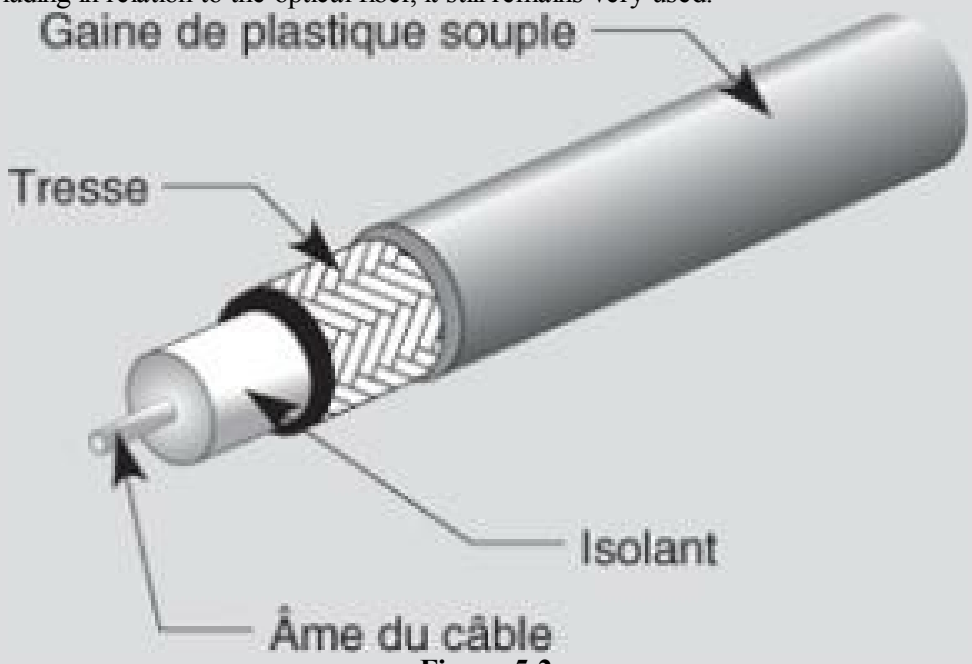


Figure 5.2

Cutting a coaxial cable

Electronic Technicians have demonstrated that the report of the diameters of the two drivers should be 3.6 mm. The different cables used are designated by the report in millimeters of the diameters of the soul and of the braid of the cable, the two most common being the 2,6/9.5 and 1,2/4,4.

As for the metal wires, the Binary flow obtained on a coaxial cable is inversely proportional to the distance to travel. On a coaxial cable of good quality with a length of one kilometer, higher flows to 100 Mbit/s can be achieved.

The main categories of coaxial cables available on the market are the following:

- Wire 50 Ω, type Ethernet;
- 75Ω cable, type CATV (Cable Television).

The optical fiber

The optical fiber is used in environments where a very high flow rate is requested, but also in the environments of poor quality. It includes the components end which will receive and transmit the light signals.

The main components transmitters are the following:

- Light emitting diode (LED) devoid of laser cavity, which emits light radiation when it is driven by an electric current.
- Laser Diode (DL), which emits a beam of coherent radiation in space and in time.
- Modulated laser.

The use of a laser transmitter decreases the phenomenon of dispersal, i.e. the deformation of the signal from a speed of propagation is slightly different following the frequencies. This gives an optical power superior to the DEL. The counterpart of these benefits is a cost more important and a life of the laser lower than that of a light emitting diode.

Figure 5.3 illustrates a fiber optic link. This figure consists of encoders and decoders that transform the electrical signals to signals which can be issued in the form of light in the optical fiber and *vice versa*. The transmitter is one of the three components end presented above and the receiver a photodetector capable to recover the light signals.

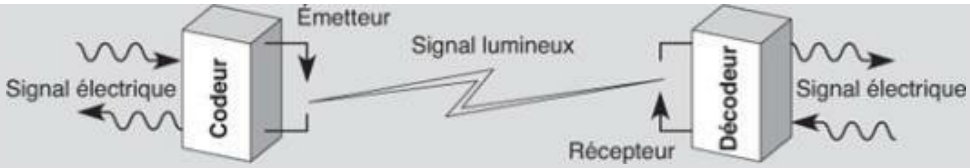


Figure 5.3

Fiber optic connection

The light beam is conveyed to the inside of the optical fiber, which is a guide cylindrical with a diameter ranging from a few microns to a few hundred microns, covered with insulation. The speed of propagation of the light in the optical fiber is of the order of 100,000 kilometers/second multimode and 250,000 kilometers/second in singlemode.

There are several types of fibers, including the following:

- The multimode fibers to jump of the index, which the bandwidth can reach 50 MHz on a kilometer.
- The multimode fibers to gradient index, whose bandwidth can reach 500 MHz on a kilometer.
- Singlemode fiber, very small diameter, which offer the greater capacity of potential information, of the order of 100 GHz/km, and the best flow rates. These are also the most complex to achieve.

It generally uses optical cables containing several fibers. The insulation surrounding the fibers avoids the problems of crosstalk, i.e. to disturbance of a signal by a signal neighbor, between the different fibers.

The transport capacity of the optical fiber continues to steadily increase thanks to the wavelength division multiplexing. At the same time, the flow of each wavelength is continually advancing.

It is estimated that it has been multiplied by two all years from 2000 to 2010, date on which it has reached almost 1 000 wavelengths. As, on the same wavelength, the capacity is increased for the same period of 2.5 to 40 Gbit/s and soon 160 Gbit/s, the capacity of several tens of terabits per second (Tbit/s, or 1012 bit/s) are today suffering on the optical fiber.

The wavelength division multiplexing, or WDM (Wavelength Division Multiplexing), is to issue simultaneously several wavelengths, that is to say several lights, on a same Heart of glass. This technique is heavily used in the hearts of network. It is called DWDM (Dense WDM) when the number of wavelengths becomes very large.

The main advantages of the optical fiber are the following:

- Very wide bandwidth, of the order of 1 GHz for a kilometer;
- Small footprint;
- Great lightness;
- Very low attenuation;
- Very good quality of transmission;
- Good resistance to heat and cold;
- Raw material good market (silica);
- Absence of radiation.

The radio media

The success of the GSM and the arrival of the mobile devices that can connect to wireless local area networks have made very popular radio media. This success is further amplified by the interconnection of personal equipment (Smartphone, Tablet, notebook PC, etc.).

All of the equipment mobile terminals use the Hertzian waves to communicate. The transmission is carried out through a cellular network, a cell being a geographical area whose all points can be achieved from a same antenna. Among the cellular networks, we distinguish the networks of mobile, the satellite networks and wireless networks. The networks of mobile allow devices to move from one cell to another without cut-off of the communication, which, as a general rule, is not the case of wireless networks. The satellite networks are of a different kind, because they ask propagation delay far longer than the terrestrial networks.

In a network of mobile, when a user moves from one cell to another, the flow of information must be amended to take account of this movement. This modification is called a change intercellular, or handover, or handoff. The management of these handovers is often difficult since it must find a new route to the communication, without however the interrupt.

Each cell has a base station, or BTS (base transceiver station) in the case of the 2G (GSM) or NODEB in the case of the 3G (UMTS) ENodeB or in the case of the 4G LTE (Advanced) or Access Point (Access Point) in the case of Wi-Fi, i.e. an antenna ensuring the radio coverage of the cell.

Is described in the following the classical case of the 2G (GSM), but there are similar situations in other radio networks; only the names are in part modified. All this is presented in detail in chapters 16, 17 and 18.

A base station has several frequencies to serve both the traffic channels of users, a channel of dissemination, a channel of common control and signaling channels. Each base station is connected by a physical support of type metal cable to a Base Station Controller, or BSC (Base Station Controller). The controller BSC and the whole of the BTS antennas which are connected constitute a sub-radio system, or BSS (Base Station Subsystem). The BSC are all connected to the switches of the mobile service, or MSC (mobile service switching center).

The architecture of a network of mobile 2G is shown in Figure 5.4.

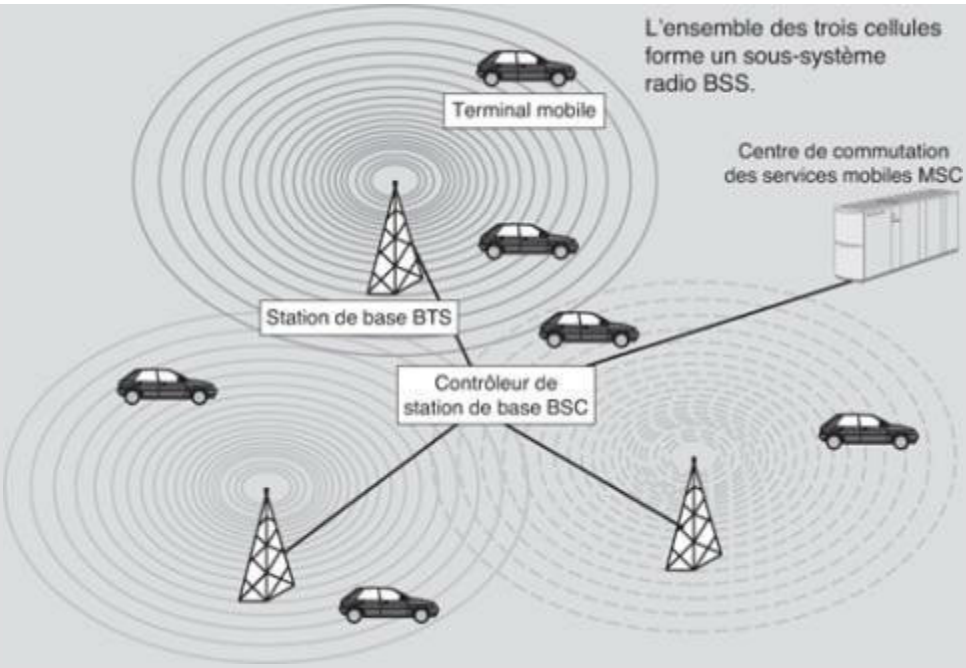


Figure 5.4.
Architecture of a network of mobile

The equipment

The equipment are of course of the elements essential to manage the transmission of signals from a transmitter to a receiver. These facilities are the following:

- The physical media for interconnection, which allow the routing of signals carrying the information.
- The taken (in English *tap), which provide the connection on the support.*
- The adaptors (in English *transceiver), which will load including the processing of signals to transmit (coding, serialization, etc.).*
- The couplers, also called communicators or maps of transmission, which support the functions of communication.

The user interfaces provide the link between the equipment to connect to and the coupler. The data that the user wishes to issue transit through this interface at a speed that depends on the chosen standard. As a general rule, the interface follows the specifications of the bus of the machine to connect to the network.

The connector

The connector performs the mechanical connection. It allows the connection on the support. The type of connector used obviously depends on the physical media.

The optical fiber poses problems of connection. The heart of the fiber being very end, of the order of a few microns, a delicate operation is necessary to attach a socket. The difficulty of the connection on optical fiber is however an asset for security, to the extent that this in fact a support difficult to spy, to the difference of the coaxial cable.

The advantage of the metal wire is that it allows you to use a telephone jack classic, which offers a great ease of connection of the coupler on the physical media. The RJ-45 jack to 8 contacts in is an example. It is the decision that we encounter now in all the companies to carry out the communication networks low current.

The adapter

The adapter (transceiver or transmitter) is responsible for the electrical connection. It is a component that is located on the map that manages the interface between the equipment and the physical media. It is responsible for the implementation of the series of bytes, that is to say of the transmission of bits one after the other, unlike what happens at the interface between the communication card and the Machine Terminal, where there is a parallelism on 8, 16 or 32 bits. The adapter performs the serialization and deserialization of packets, as well as the transformation of logical signals in communicable signals on the bracket and then their issuance and their reception.

According to the access method used, additional functions can be assigned to the adapter. For example, it may be responsible for the detection of occupation of the cable or the detection of collisions of signals. It can also play a security role in ensuring the limitation of occupation of the media by a transmitter. The adapter is now more and more integrated into the coupler.

The coupler

The body called coupler, or network card or access card (an Ethernet card, for example), is in charge of Check the driveshafts on the cable (see figure 5.5). The coupler ensures the formatting and the déformatage blocks of data to be transmitted as well as the detection of error, but very rarely in the occasions on error when an error is discovered. It is also responsible for managing the resources such as the memory areas as well as the interface with the outside.

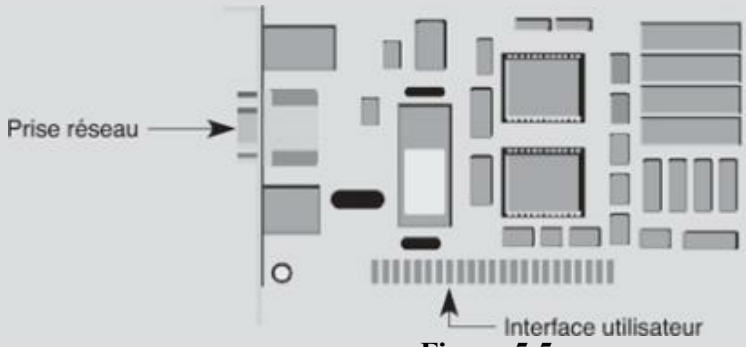


Figure 5.5

Carte coupleur

Figure 5.5

Coupler card

The flow of a coupler must adjust to the flow rate allowed by the cable. For example, on an Ethernet network with a physical media whose capacity is 100 Mbit/s, the coupler must issue at this same speed of 100 Mbit/s.

The coding and the transmission

The data networks are based on the digitization of information, i.e. the representation of the data by suites of 0 and 1. To transform the information into binary suites, it uses of the codes, which are correspond to each character a specific suite of binary elements. The number of bits used to represent one character corresponds to the number of moments of a code. A code to *n* moments allows to represent 2n separate characters.

Several codes have been standardized to facilitate exchanges between computer equipment. The number of times used increases with the size of the alphabet, which is none other than the list of characters that must be encoded. The Alphabet may not be composed of digits. We can add the lowercase and uppercase letters, punctuation, arithmetic operators, but also specific commands.

The main codes used are the following:

- Telegraphic Code, 5 times. The alphabet can be 32 characters long, of which only 31 are used.
- ASCII code, 7 times, or 128 characters available.
- EBCDIC code to 8 times, which allows up to 256 characters.
- Unicode, to 16 times, which resumed in a manner slightly simplified the specifications of the ISO code 10646 UCS (Universal Character Set), to 32 times. This unique code allows to take into account all the languages of the world.

After the step of the Coding intervenes the transmission itself, i.e. the sending of binary suites of characters to the end user. This transport can be performed in parallel or in series.

In the transmission in parallel, the bits of a same character are sent on metallic wires distinct to arrive together at destination. There may be 8, 16, 32 or 64 parallel wires, or even more in specific cases. This method however, poses problems of synchronization, which lead to use it only on very short distances, the bus of a computer, for example.

In serial transmission, the bits are sent one behind the other. The succession of characters can be asynchronous or synchronous. The asynchronous mode indicates that there is no predetermined relationship between the transmitter and the receiver. The bits of a same character are supervised from two signals, one indicating the beginning of the character, the other The End. This are the bits Start and Stop. The beginning of a transmission can take place at any point in time in time, as shown

in Figure 5.6.

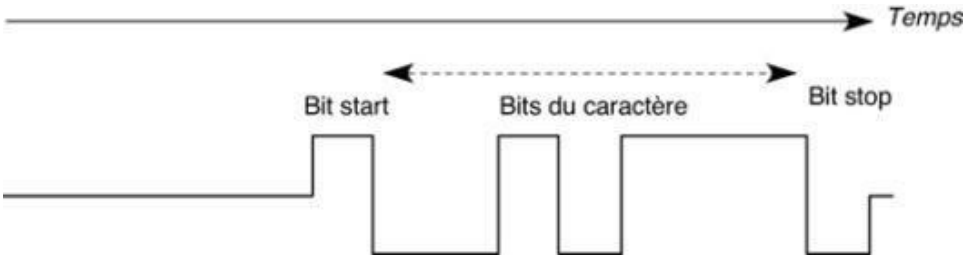


Figure 5.6

Transmission of a character in the asynchronous mode

In the synchronous mode, the transmitter and the receiver to agree on a constant interval, which repeats without judgment in time. The bits of a character are sent the one behind the other and are synchronized with the beginning of the time intervals. In this type of transmission, the characters are issued in sequence, without any separation. This mode is used for very high flow rates.

In all cases, the signal emitted is synchronized on a clock during the transmission of a binary element. The speed of the clock provides the throughput of the line expressed in baud, that is to say the number of clock ticks per second. For example, a communication line that operates at 50 baud indicates that there are 50 intervals of time basic in a second. On a basic interval, it usually issues a bit, that is to say a signal to 1 or 0. Nothing prevents to transmit four types of distinct signals, which would have as meaning 0, 1, 2 and 3. We are told, in the latter case, that the signal has a valence of 2. A signal has a valence of *n* If the number of levels transported in a elementary time interval is equal to 2n. The transmission capacity of the line in number of bits transported per second is worth n multiplied by the speed expressed in baud. It expresses this capacity in bits per second. For example, a line of a speed of 50 Baud who has a valence of 2 has a capacity of 100 bits per second (100 bit/s).

During the transmission of a signal, disturbances of the physical line by what is called the outside noise may occur. If we know the level of this noise, we can calculate the maximum capacity of the line. In more specific terms, the noise may have for the origin of the poor quality of the line itself, which amends the signals which spread, as well as intermediate elements, such as such as modems and multiplexers, which do not always send exactly the signals requested, or of external events, such as the electromagnetic waves.

The noise is considered as a random process described by a function *B(t)*. If *S(t)* is the transmitted signal, the signal arriving at the receiver is written *S(t) + b(t)*. The signal to noise ratio is a characteristic of a channel: this is the report of the energy of the signal on the energy of the noise. This report varies in time, since the noise is not uniform. However, it is estimated by an average value on a time interval. It is expressed in decibel (dB). This report is written S/B.

The Theorem Shannon gives the maximum capacity of a channel subject to a noise:

$$C = Wlog2(1 + s/B),$$

Where *C* is the maximum capacity in bit per second and *w* the bandwidth in hertz.

On a phone line in which bandwidth is 3 200 Hz, for a signal to noise ratio of 10 dB, you can theoretically reach a capacity of 10 kbit/s.

To finish with this brief overview of the transmission techniques, let us look at the different possibilities of transmission between two points. Unidirectional links, or simplex, were always held in the same direction, from the transmitter to the receiver. The bidirectional connections, full duplex or half duplex, or even half-duplex, can transform the transmitter in receiver and *vice versa, the changing communication of meaning in turn. The bidirectional links simultaneous, or duplex, or even full-duplex, allow a simultaneous transmission in both directions. These various cases are illustrated in Figure 5.7.*

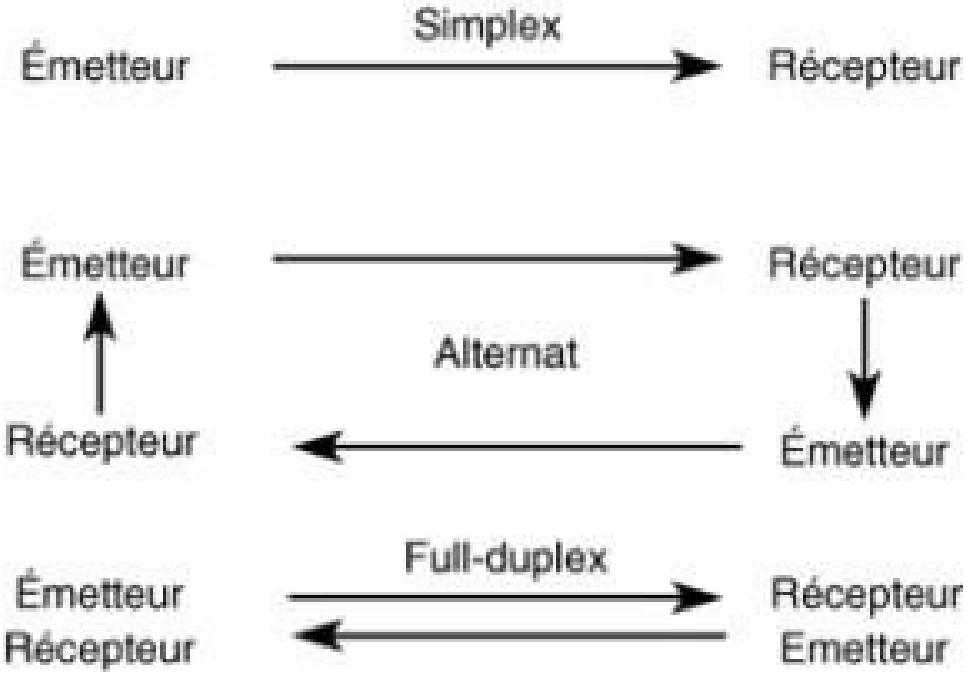


Figure 5.7
Transmission Direction

The transmission in basic band

Let us now examine the transmission techniques used, that is to say how a transmitter can send a signal that the receiver will recognize as being a 1 or A 0.

The simplest method is to issue on the line of the different currents, zero current indicating a 0 and a current a positive 1. It obtains from the so a representation of the Bits of the character to transmit in the form of niches, as shown in Figure 5.8.

This method is called Transmission in basic band. The exact fulfillment of these slots is very complex, the fact that it is often difficult to get continuous current between two stations. The same difficulty is found in the NRZ (Non Return to Zero), also shown in Figure 5.3. The Bipolar coding is an encoding all-or-nothing, in which the bit 1 is indicated by a current positive or negative to the tower of role, so as to avoid the currents. This Code leaves the bit 0 defined by a current zero.

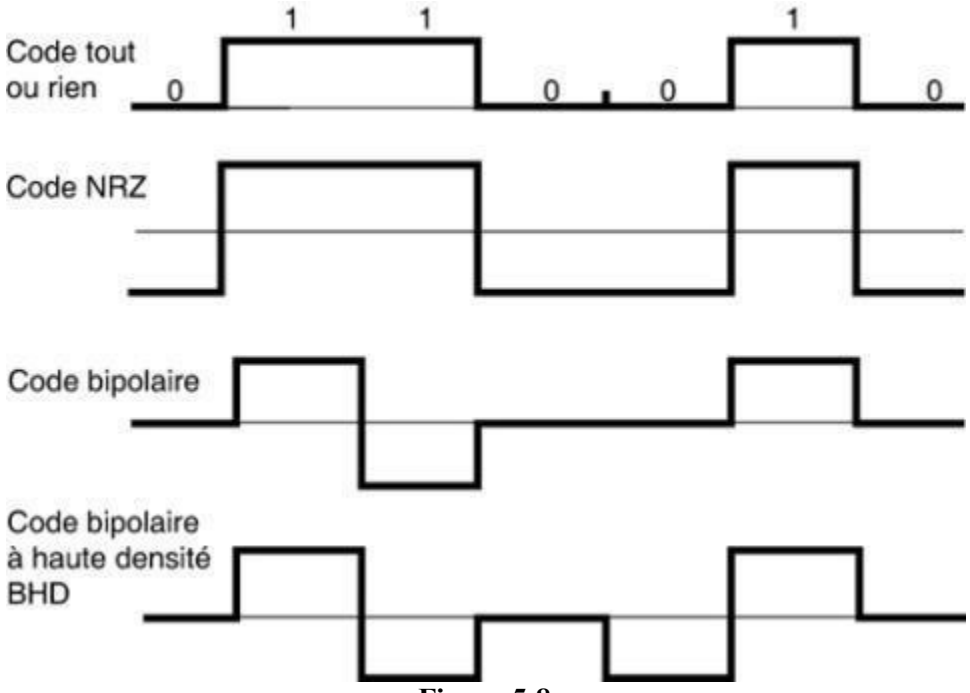


Figure 5.8
The encodings in basic band

The Bipolar coding at high density allows you not to leave the current zero during the suites of 0. Special suites of filling (current negative, zero, or positive) are then inserted in place of these zeros. A new 1 is indicated by a current positive or negative, in violation with the suite of filling.
Many other encodings in basic band have been developed at the discretion of the application to improve such or such a characteristic of the signal. Figure 5.9 shows the encodings RZ (Return to Zero), Miller, Manchester and biphase-M and S.
The rapid degradation of the signals as the distance travelled is the main problem of the transmission in basic band. If the signal is not regenerated very often, it takes a form, that the receiver is unable to understand. This transmission method can therefore not be used on short distances of less than 5 kilometers. Over longer distances, it uses a signal of sine-wave form. This type of signal, even weakened, may very well be decoded by the receiver.

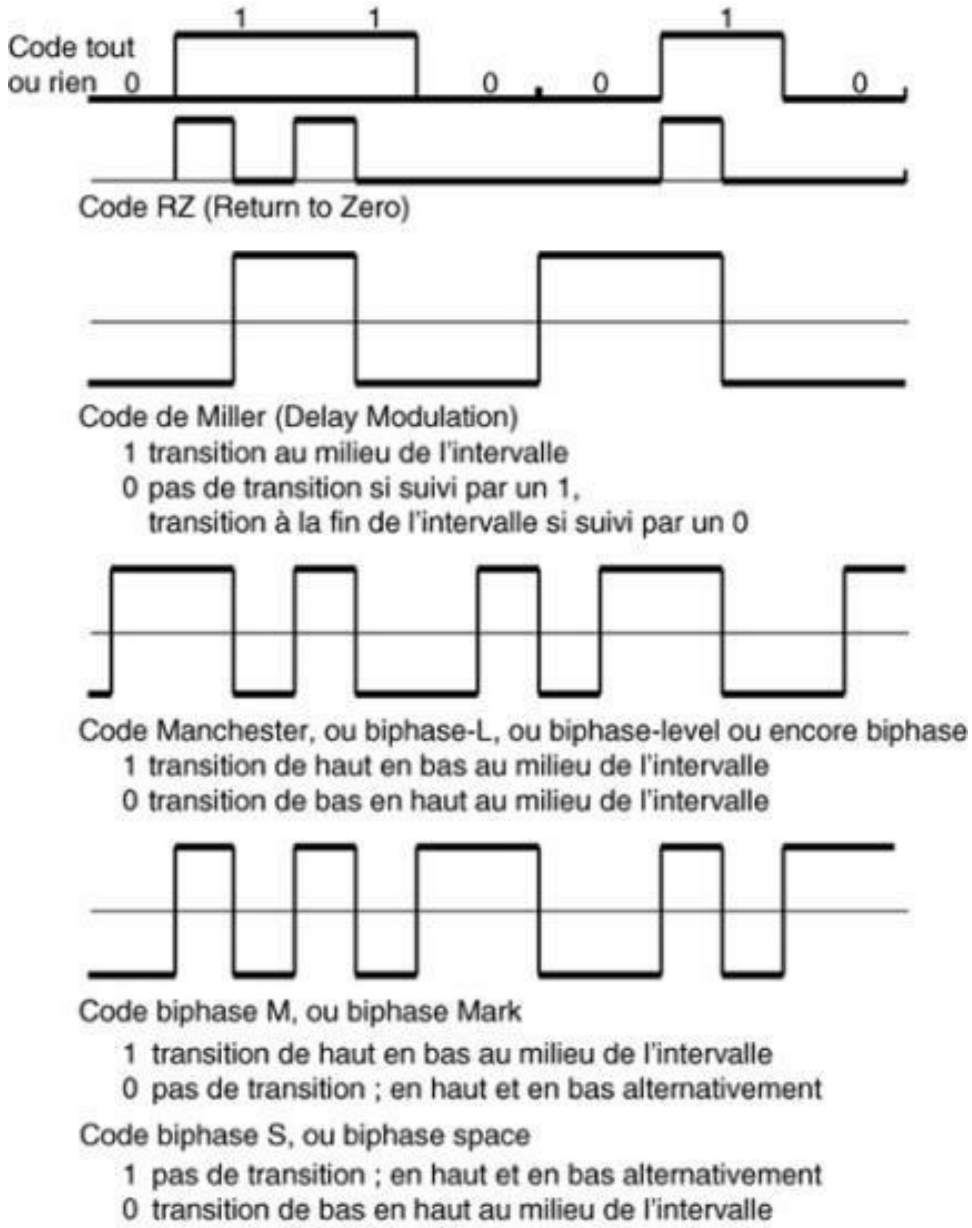


Figure 5.9
A few encodings in basic band

The modulation

As explained earlier, to transmit a binary element, it must emit a signal very particular to recognize if its value is equal to 0 or 1. The techniques in basic band in the form of niche are not reliable as soon as the distance exceeds a few hundreds of meters. To have a signal that we can recover properly, we

must give it a special form in the modulating.
There are three broad categories of following modulation:

- Amplitude modulation, or ask (Amplitude-Shift Keying);
- Phase modulation, or PSK (Phase-Shift Keying);
- Frequency modulation, or FSK (Frequency Shift Keying).

An intermediate material, the modem (modulator-demodulator), is necessary to modulate the signal in a form sinusoidal. The modem receives a signal in basic band and the module, i.e. Assigns an analog form sinusoidal. The fact of not having more fronts amounts nor descendants protects much better the signal of the damage caused by the distance travelled by the signal in the cable since the signal is continuous and not more discreet.
As soon as a device located at a distance of a few important must be reached, a modem is required to ensure that the rate of error is acceptable. The distance depends very heavily on the cable used and the speed of transmission. Classically, from a few hundreds of meters for the very high speeds and a few kilometers for the lower flow rates, it is necessary to appeal to a modem.

The modulation of amplitude

In the amplitude modulation, the distinction between the 0 and the 1 is obtained by a difference in the amplitude of the signal, as shown in Figure 5.10.

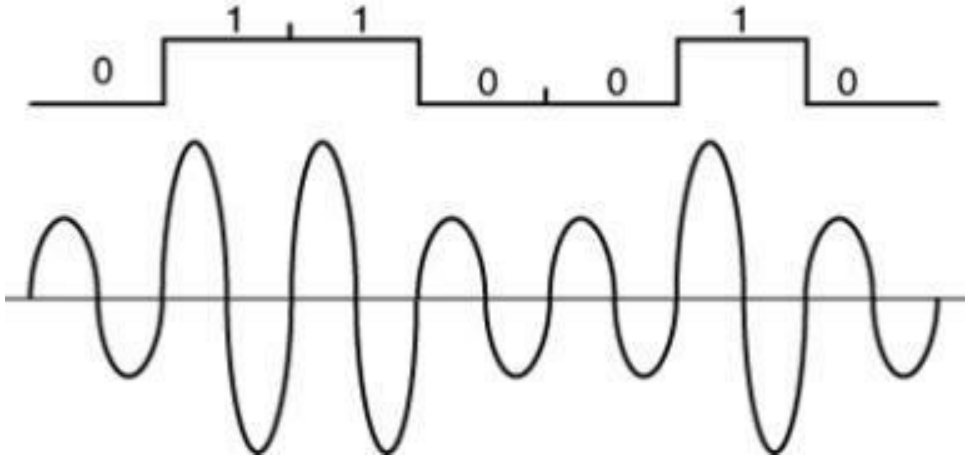


Figure 5.10
Amplitude modulation

The phase modulation

For the phase modulation, the distinction between 0 and 1 is performed by a signal that starts at different locations of the sinusoid, called phases. In Figure 5.11, the values 0 and 1 are represented by the respective phases of 0° and 180°.

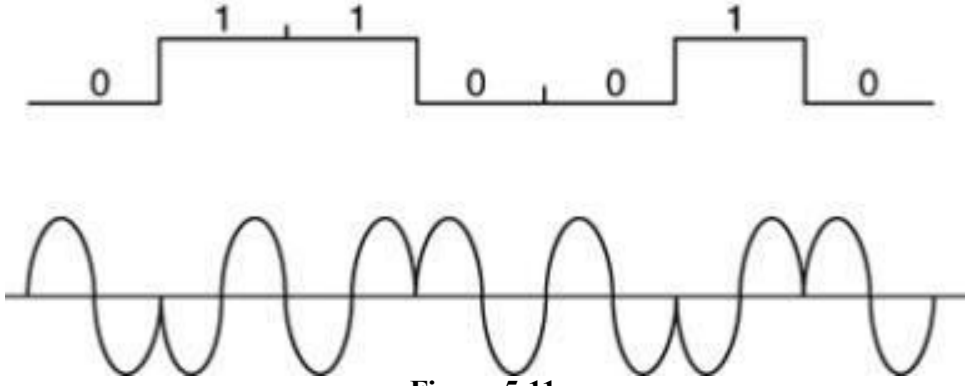


Figure 5.11
Phase Modulation

The frequency modulation

In Frequency modulation, the transmitter has the possibility to change the frequency of the sending of the following signals that the binary element to issue is 0 or 1, as shown in the Figure 5.12.

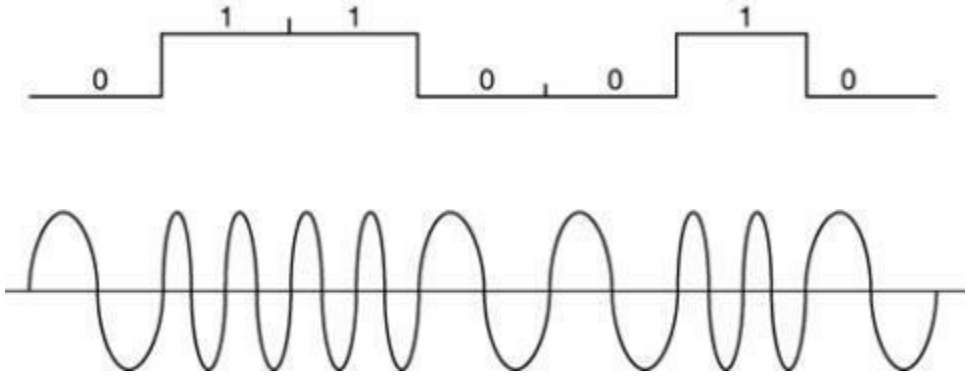


Figure 5.12
Modulation of Frequency

In the previous presentations of the modulation techniques, the physical size used for the amplitude, phase or frequency represents only two possible states. If one arrives at issue and to detect the arrival of more than two States of the same magnitude, you can give to each State a meaning for coding 2 or several bits. For example, using 4 frequencies, 4 phases or 4 amplitudes, it can encode 2 bits to each State. Figure 5.13 illustrates a possibility to encode 2 bits in phase modulation.

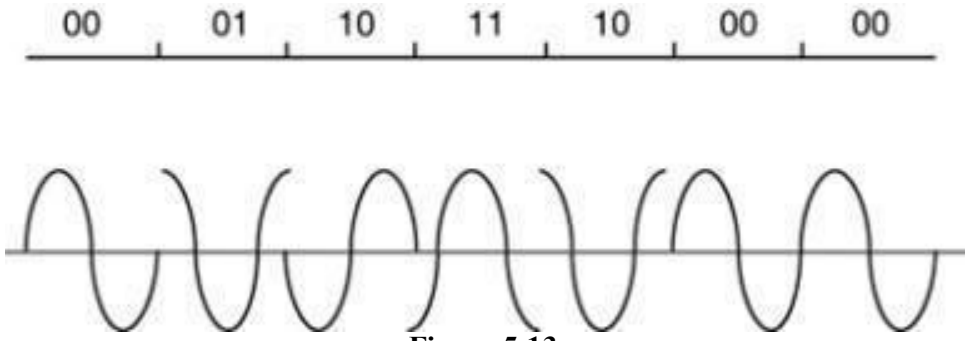


Figure 5.13
Phase Modulation in four moments

The modems

The modems can transform the binary signals in basic band In analog signals indicating specific a numeric value. The signal is presented in a sine-wave form.

The modems can adapt to the different types of support, which can be:

- 2 wires of normal quality;
- 4 wires of normal quality;
- 4 wires of superior quality in accordance with the opinion M.1020 of the ITU-T;
- 2 wires in the basic band;
- 4 wires in the basic band;
- The primary groups, the secondary groups, etc.

Table 5.2 lists the opinion of the ITU-T concerning the conventional modems. The ADSL modems are presented in detail in Chapter 15. They require the functions of multiplexing, which are introduced in the following sections.

Opinion CCITT	Flow rate in bit/s	Modulation Type	Modulation rate	Operating System
V.21	300	Frequency	300	Full-duplex (FD)
V.22	600/1 200	Phase	600	FD

V.22bis	1 200/2 400	Phase	600	FD
V.23	600/1 200	Frequency	600/1 200	Half Duplex (HD)
V.23	1 200/75	Frequency	1 200/75	FD
V.26	2 400	Phase	1 200	FD
V.26bis	1 200/2 400	Phase	1 200	HD
V.26ter	1 200/2 400	Phase	1 200	FD
V.27	4 800	Phase	1 600	FD or HD
V.27bis	2 400/4 800	Phase	1 200/1 600	FD or HD
V.27ter	4 800	Phase	1 200/1 600	HD
V.29	4 800/9 600	Phase + amplitude	4 800/9 600	FD
V.32	4 800/9 600	Phase + amplitude	2 400	FD
V.32bis	Up to 14 400	Phase + amplitude	3 200	FD
V.34	Up to 28 800	Phase + amplitude	3 200	FD
V.34+	Up to 33 600	Phase + amplitude	3 200	FD
V.90	Up to 56 000/33 600	Phase + amplitude	3 200	FD
V.92	Up to 56 000/48 000	Phase + amplitude	3 200	FD

Table 5.2 • Standardized modems

It happens sometimes that the additional features are implemented in the modems. An important feature is the compression: rather than to increase the speed, it compresses the data. The Protocol MNP (Microcom Networking Protocol) is a good example of proposal of compression and error correction. This protocol, developed by the American manufacturer Microcom, is standardized by the Notice V.42bis of the ITU-T.

Multiplexers

On a communication line forming a connection between two remote points, it may be interesting to pass data at the same time of several customers. Rather than each client has its own infrastructure, it is more economical to have only a connection shared by several users. A multiplexer function is to receive data from several terminals through specific connections, called ways low speed, and transmit all together on a single link, track high speed.

At the other end of the connection, it is necessary to carry out the opposite approach, i.e. recover, from information arriving on track high speed, the data of different users and send them on the good output channels. This task is the responsibility of the demuxer. The machine that performs the multiplexing and demultiplexing is called a MUX.

There are a large number of possibilities of multiplexing. The sections that follow present the main.

Frequency Multiplexages temporal and

In the Multiplexing in frequency, each track Low Speed has its own bandwidth on the track high speed. In this case, the track high speed must have the necessary capacity to absorb all the frames that come of terminal equipment connected.

Time Division Multiplexing follows the same mechanism, but instead that the track High Speed is divided into separate frequencies, it is cut into slices of time, which are assigned regularly to each track low speed. We understand that time division multiplexing is more effective than the previous since it makes better use of the bandwidth. A problem arises however: when a frame is present at the input of the multiplexer and that the time slice that is assigned to this device is not exactly to its beginning, we must remember the information up at the appropriate time.

A temporal multiplexer must therefore be equipped with memory buffers for storing the binary elements that arise between the two slices of time. It is very simple to calculate the size of this memory, since it corresponds to the maximum number of bits presenting themselves between the two slices of time assigned to the device. This expectation is not always negligible compared to the propagation time of the signal on a line of communication.

Statistical multiplexing

In the two types of previous multiplexing frequency, and temporal, there can be no problem of flow, the track high speed with a capacity equal to the sum of the capacities of tracks Low Speed connected. As a general rule, this leads to a waste of bandwidth, since the tracks low speed do not transmit continuously. To optimize the capacity of the track high speed, it is possible to play on the average of the flow rates of the tracks at low speed. This is what is called the statistical multiplexing. In this case, the sum of the average flows of tracks low speed should be slightly less than the flow rate of the track high speed. If, during a period of time, there are more of arrivals that cannot bear the connection, briefs take additional the relay in the Multiplexer.

Operation of the statistical multiplexing

Statistical multiplexing is based on a statistical calculation of arrivals and not on the flows means. For example, if ten tracks low speed of a flow rate of 64 kbit/s arrive on a multiplexer statistics, the total flow rate can reach 640 kbit/s. This value corresponds to the maximum value when the machines debiting on tracks low speed work without no judgment.

In the facts, it is rare to exceed 50% in the use of the line, that is to say in our example 32 kbit/s per line. Playing statistically, it can take a high-speed link of a flow equal to 320 kbit/s. However, nothing prevents that all stations are active at a given time. In this case, a capacity of 640 kbit/s is presented to the Multiplexer, which can not debit that 320 kbit/s. A significant memory must therefore dab pending data transmission on the high line speed. If the statistical calculation is not performed correctly, the losses are to be expected.

Figure 5.14 gives a representation of the statistical multiplexing.

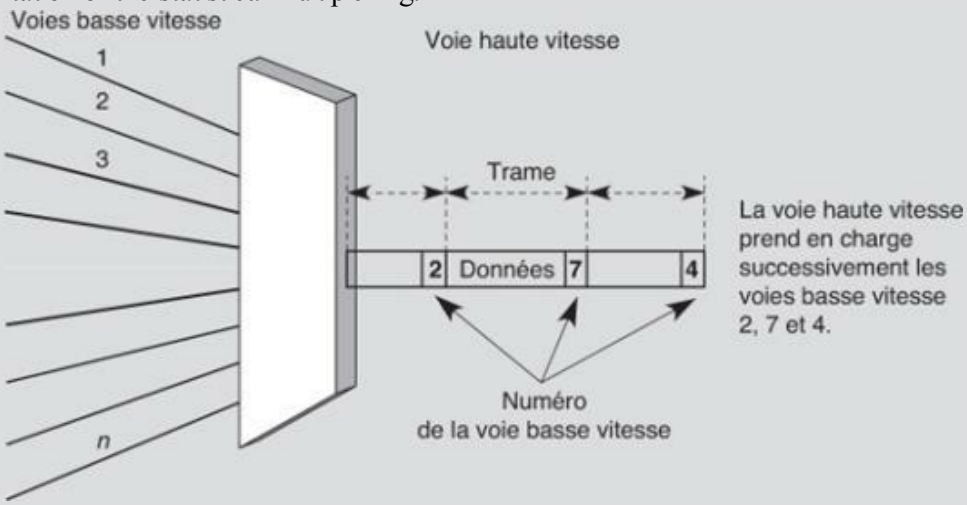


Figure 5.14

Statistical multiplexing

In this diagram, it notes that the information in the track low speed are transported in a frame. This last must contain a number in the header for that track low speed to be recognized in the demuxer.

A hub is a multiplexer statistics which has additional features, such as higher-level protocols to the physical layer.

The transmission

Before transmitting the information on a transmission medium, it must encode adequately. The networks must allow very high throughput rates over distances more or less long. In this context, three approaches are possible for the coding of the binary elements from the applications:

- The information is conveyed directly in basic band, which allows to obtain the flows to cash in Gigabits per second on a few tens of meters. On a few hundreds of meters, one can reach several tens of megabits per second. Finally, over

distances of a few kilometers, we reached on pairs of metal wires telephone-type flow rates of the order of a few hundreds of kilobits per second.

- The information is modulated according to the principles listed earlier in this chapter. The speeds are much smaller, but the distances are much greater. To increase the speed, it must be able to carry a large number of binary elements by interval of time.
- The digital signals are modulated on a carrier, and each type of information is allocated a bandwidth depending on its needs. It is the approach of broadband.

The transmission in basic band

The transmission in basic band is the most simple, since no modulation is necessary. The binary result representing the information is transmitted directly on the support by the introduced changes in the signals representing the information in the form of voltage transitions, or light pulses if one uses the optical fiber.

The signals in the basic band are subject to a mitigation, the importance of which depends on the media used. They must therefore be regenerated periodically on a long distance. This regeneration is carried out with the aid of repeaters, which receive the signals and store a fraction of a second before the rebroadcast on the outgoing line.

The broadband

This method uses the Multiplexing in frequency. Different channels are created, resulting from the division of the bandwidth of the media in several sub-bands of frequencies. This technique has the advantage of allowing simultaneous transmissions independent.

Each device on the cable is equipped with a modem particular. This allows you to choose the mode of transmission, digital or analog, the better adapted and more effective for the type of information to transmit. For example, computer data are issued on a digital tape, and the voice and the image on a analog band. The transmission broadband grows however the cost of connection by report to a network in basic band, more simple to install and generally less expensive.

The digitalisation of signals

How to encode the digital signal is an important function of the coupler of communication. This operation has the principal function to adapt the signals to the transmission channel. In the case of local networks, the transmission speed is several tens or hundreds of megabits per second. Of this fact, the choice of the physical representation of the data is important. To perform the synchronization bit, that is to say to ensure that each bit is read at the right time, there must be a minimum of transitions are being carried out to extract the clock signal.

The coding used in most of the local networks, and in particular in Ethernet networks, is the Manchester encoding, or his version Manchester Differential. The Manchester encoding, also said biphase-The (biphase-level), is shown in Figure 5.10. There is always a transition by binary element, so that the value of the signal passes without stopping a positive value to a negative value. This transition is carried out in the middle of the interval.

In Figure 5.15, the 0 is indicated by a transition from top to bottom, while the 1 is indicated by a transition from the bottom to the top. The figure shows the following 100110 coded in Manchester. The signal begins by a negative polarity then goes to a positive polarity in the middle of the interval. This passage of a negative polarity to a positive polarity is called a FRONT amount. The 1 is represented by a front amount and the 0 by a falling edge.

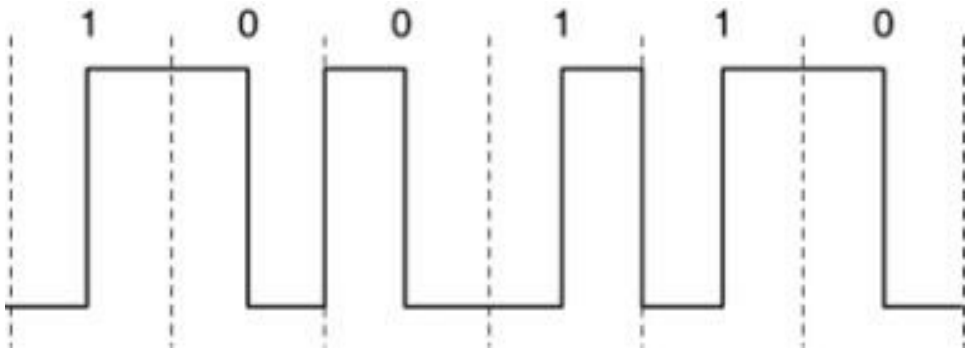


Figure 5.15

The Manchester Encoding differential assembly takes into account the ILO precedent, as shown in the Figure 5.16. Bit 0 is represented by a polarity change at the beginning of a time bit. Bit 1 is characterized by the absence of polarity change at the beginning of a time bit. This encoding has the advantage of being independent of the polarity.

The coding by blocks is another widely used method, in particular in the local networks. The general principle of this coding is to transform a word of *n* bits in a word of *M* bits, where his other encoding name *nB/MB. Due to technological constraints, the values of N are* usually chosen are 1, 4, or 8.

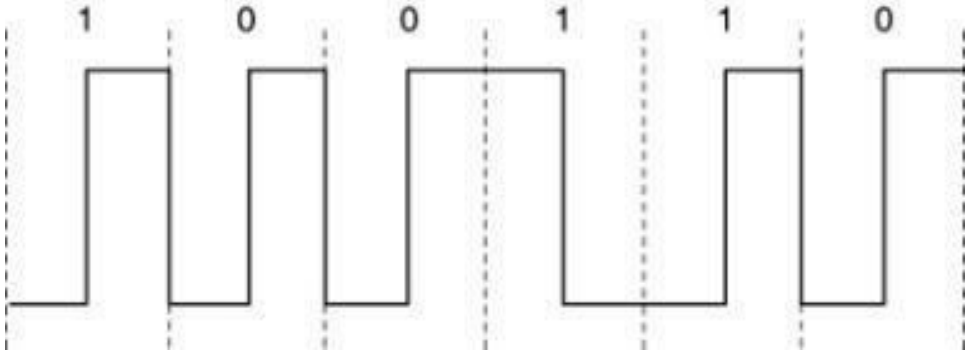


Figure 5.16

Manchester Encoding differential assembly

The principle of the Coding 1B/2B, or coding IJC (Mode Codec indication), is shown in Figure 5.17. A signal is issued on two clock time. The 1 is indicated by a continuous level top and then a continuous level bottom in turn. The value 0 starts by a continuous level down on the first signal of clock and then continues by a continuous level top on the second clock signal. The Suite 1001 represented on the figure therefore request eight Clock Time to encode the four bits. The first two clock time carry the bit 1, which corresponds to a continuous level high. After the two bits 00, the value 1 is transported by a continuous level low. The following 1 is transported by a continuous level high. This coding is easy to implement, but has the disadvantage that the signal occupies a double band width.

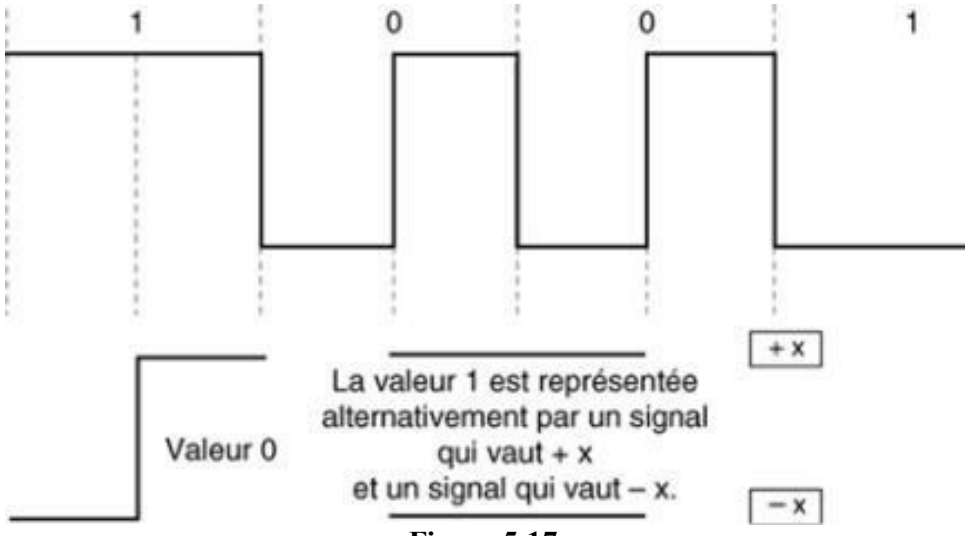


Figure 5.17

Ijc coding

The codes *nB/MB have interesting properties. You can use their particularities to detect transmission errors, get banned words or represent specific sequences, as delimiters of frames or tokens. For example, in putting two levels Senior of suite in the previous example we would not represent the Suite 11, but a forbidden word. The controls of error can be performed by verifying that the coding rules are not violated. You can recognize a particular sequence by violation of polarity, that is to say by the non-compliance of the alternating high level-low level. Figure*

5.18 provides an example.

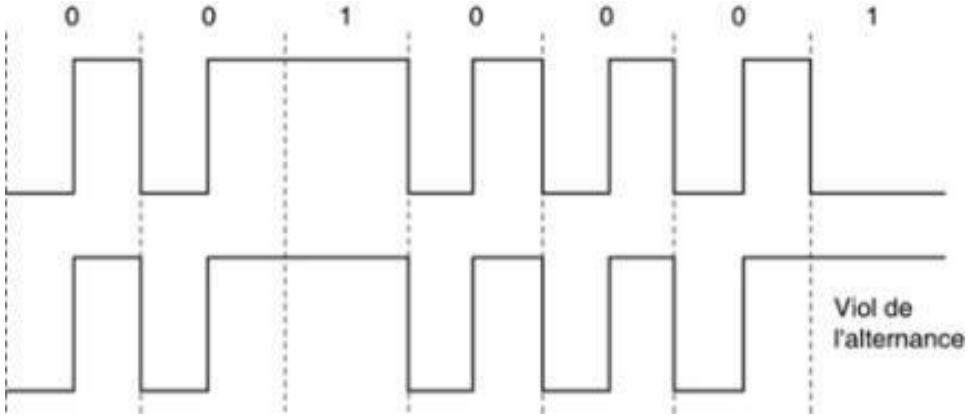


Figure 5.18
Principle of the rape of polarity

Digitization of analog signals

Now, the vast majority of transport of information is carried out in digital. The analog signals must therefore be transformed into a suite of binary elements. The value of the bit rate obtained by scanning of the signal requires that the bandwidth of the physical media is sometimes greater than that necessary for the transport of the analog signal. For example, the telephone speech uncompressed, which requests a analog bandwidth of 3 200 Hz, requires a digital flow of 64 000 bit/s, flow rate which may in no case be absorbed by a physical support to 3 200 Hz bandwidth. Indeed, as we have seen previously, the maximum flow routed on a band of *W* Hz is obtained by the Theorem Shannon:

$$D = W\log2(1 + s/B)$$

Where s/B is the signal to noise ratio expressed in decibels. For a report of 10, which is relatively important, one obtains a maximum bitrate of 10 000 bit/s.

Three successive operations must be performed to arrive at this scan, sampling, the quantification and the coding:

1.Sampling. Is to take the points of the analog signal as it unfolds. More bandwidth is important, the more it is necessary to take samples per second. It is the sampling theorem which gives the solution: If a signal f(t) is sampled at regular interval in time and at a rate higher than the double of the significant frequency the Most High, the samples contain all the information of the original signal. In particular, the function *f(t)* can be reconstructed from the samples. This phase is shown in Figure 5.19.

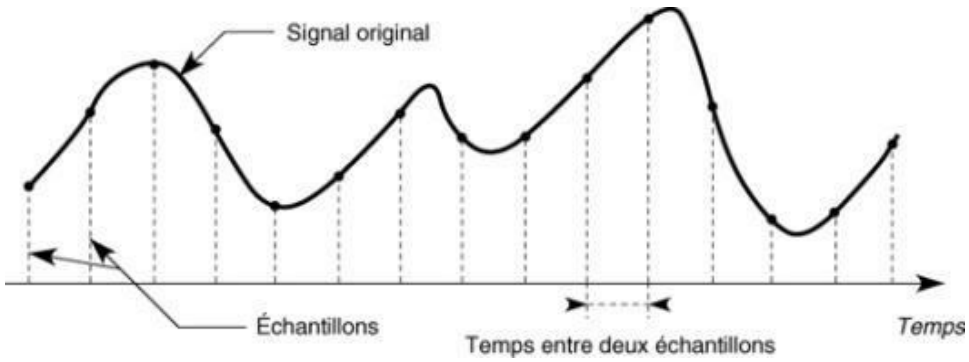


Figure 5.19
Sampling

If we take a signal with the width of the bandwidth is 10 000 Hz, it must be sampled at least 20 000 times per second.

2.quantification. Is to represent a sample by a numeric value in the middle of a law of correspondence. It should be to find this act of correspondence so that the value of the signals has the

most possible significance. If all the samples are almost equal, it must try, in this delicate area, to have more opportunities for coding that in areas where there is little of samples. To obtain a correspondence between the value of the sample and the number the representative, one usually uses two laws, the act has in Europe and the Act Mu in North America. These two laws are of type semi-logarithmic, guaranteeing a precision almost constant. This phase is shown in Figure 5.20.

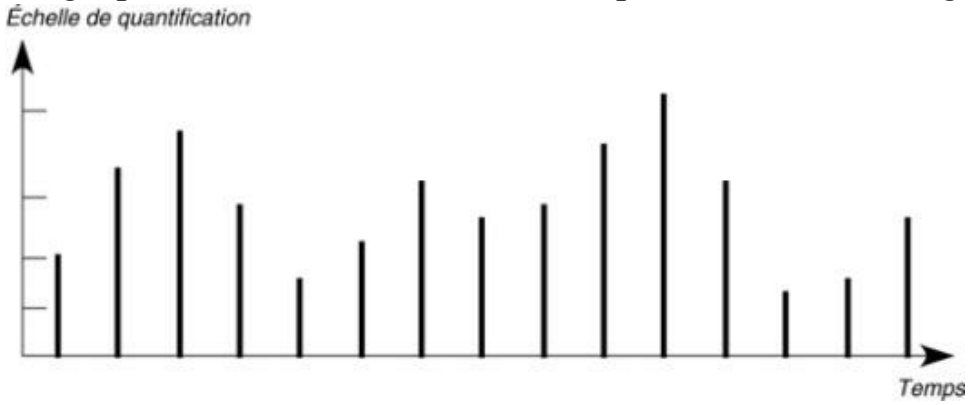


Figure 5.20
Quantification of a signal is sampled

3.coding. is to assign a numeric value to the samples obtained during the first phase. These are the values which are transported in the digital signal. This phase is shown in Figure 5.21.

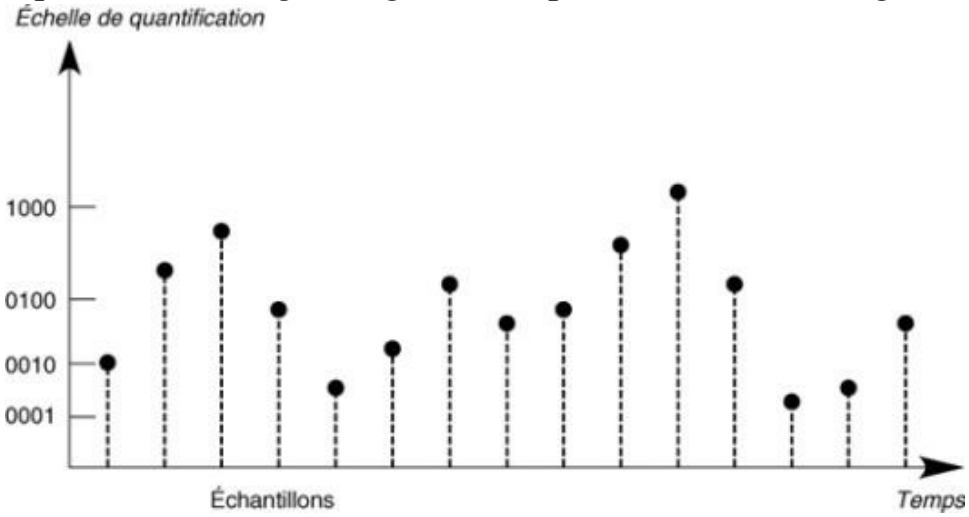


Figure 5.21
Coding

Digitization of the telephone speech

The digitization of the telephone speech is generally carried out by means of the conventional methods PCM (Pulse Code Modulation) in North America and the MIC (pulse width modulation and coding) in Europe. These methods have slight differences, including the most visible affects the flow of output, which is 56 kbit/s in North America and 64 kbit/s in Europe.

The bandwidth of the floor analog telephone is 3 200 Hz. To scan this signal correctly without loss of quality, already relatively low, must be sampled at least 6 400 times per second. In the normalization, we adopted the value of 8 000 times per second. The quantification is performed by laws semi-logarithmic. The maximum amplitude permitted is divided into 128 positive levels for the PCM version, to which we must add 128 negative levels in the European version MIC. The coding is done either on 128 values, either on 256 values, which request, in binary, 7 or 8 bits of encoding.

The total value of the flow of the digitization of the telephone speech is obtained by multiplying the number of samples by the number of levels. This gives:

- 8 000 × 7 bit/s = 56 kbit/s in North America and Japan;
- 8 000 × 8 bit/s = 64 kbit/s in Europe.

The sampling has place all 125 microseconds, value which are very often to be found in the result of this book.

Any type of analog signal can be scanned by the general method described above. We see that more bandwidth is important, the greater the quantity of binary elements to transmit increases. For normal speech, limited the more often to 10 000 Hz bandwidth, there must be a flow of 320 kbit/s if the encoding is performed on 16 bits.

Other techniques of digitization of the floor are also used. They work in real time or delayed. In the first case, the algorithm that allows to translate the intermediate act of quantification is executed in real time, and the binary elements obtained are not compressed, or very little. In the second case, the floor can be stored on volumes much lower, but the time required to perform the decompression is too long

to regenerate a synchronous stream of bytes and therefore the analog output signal. That is why it must be a intermediate storage which removes the time aspect of the actual floor. For the digital courier, a compression is almost always performed in order that storage capacity required are not too important. In this case, it descends at flow rates of less than 2 kbit/s.
We can still quote in real time techniques the methods Δ (Delta) or ΔM (Delta modulation), which rely on the coding of a sample in relationship with the previous one. For example, we can define the sampling point $k + 1$ by the slope of the line connecting the samples k and $k + 1$, as shown in Figure 5.22. It sends the exact value of the first sample, then transmits only the slopes. Given that the slope of the right gives only an approximation of the following point, it must be periodically issue a new sample with its exact value.

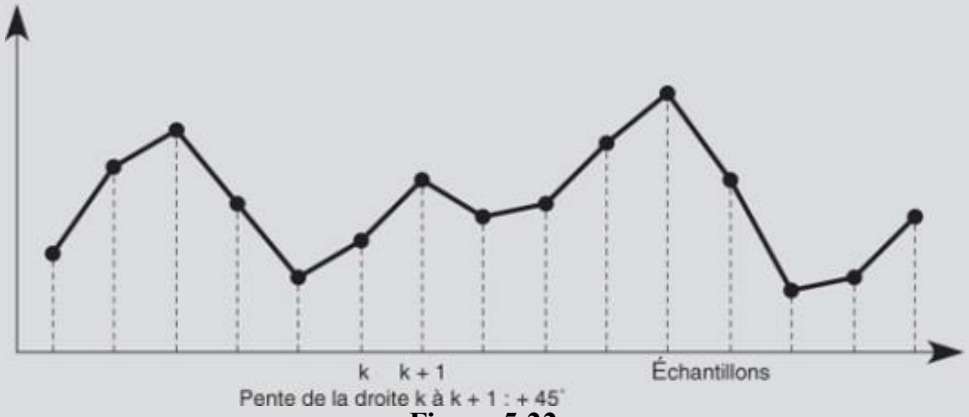


Figure 5.22
Scan by a delta method

Thanks to these methods, the flow rate of the digital word can descend to 32 or 16 kbit/s, or even less. You can go up to 2 kbit/s, but it then gets a synthetic speech of poor quality. This section had referred only to the floor on the phone. It goes without saying that all analog information can be scanned in the same way.
The scan of the animated image follows a similar process, the image being decomposed in basic points, called pixels, and each pixel being coded on several bits or even on several bytes, if the number of colors in the image is high. Table 5.3 identifies some values of digital flows necessary to transport of analog signals scanned.

Type of Information	Flow of the scanned signal	Flow rate after compression
Its	64 Kbit/s	1.2 to 9.6 kbit/s
Animated images black and white/video-conferencing	16 Mbit/s	64 Kbit/s to 1 Mbit/s
Animated images color/video-conferencing	100 Mbit/s	128 kbit/s to 2 Mbit/s
Images color television	204 Mbit/s	512 kbit/s to 4 Mbit/s
Videoconferencing Images	500 Mbit/s	2 Mbit/s to 16 Mbit/s

Table 5.3 • Flow rates of a few scanned signals

Great progress has been made in recent years in the area of compression, which allow to significantly reduce the flow of the waves to route. The experiences of transport of animated images is carried out, for example, on Channels to 64 kbit/s. Terminal equipment however, returned still expensive for the very strong cuts. Coding Techniques continuing to move forward very quickly, they represent a solution for the transit of the video on networks of mobile such as UMTS, in which the available bandwidth is low compared to flows envisaged. The main of them are presented in Chapter 16.
The encoders that perform the passage of the analog signal to the digital signal are called codecs (coder-decoder). Simple to perform, the codec MIC is today good market. In contrast, the codecs for analog signals for the analog signals to very wide bandwidth return again dear, because of the technology they employ.

Error detection and correction

The detection and correction of errors have long been automated to level frame from the fact that the quality of the physical lines was insufficient to obtain the error rates acceptable for the applications running on the network. Today, the problem is somewhat different, and this for two reasons:

- The rate of error in line is become satisfactory, descendant often under the bar of 10^{-9}, and the few errors which remain does not alter the quality of the application. This comes from encoding techniques more powerful and of the use of physical media such as optical fiber.
- The applications forwarded are type of multimedia and do not tolerate the loss of

time associated with the times on error. The correction of the errors do not affect the quality of the image or sound. For as much as the number of errors is not too important, the eye or the ear cannot detect minor modifications of the image or sound. The retransmission is therefore a loss of time useless. For example, the floor phone, who request a transportation time from 150 milliseconds to the maximum, does not authorize the expectation of retransmissions. In addition, the correction of a bit here and there by changes practically nothing to the quality of the floor.

The detection and correction of errors are essential on the physical media of poor quality or for applications that require the transport of valuable data. In this case, automation at the level frame or a particular recovery at the message level can be performed for a particular application. It is always possible to add at the semantic level, layer 7, or application, the reference architecture, a process of correction of errors.

The two major opportunities for error recovery are sending the information in redundancy, which can detect and correct errors in a the same time, or the use only of a detector code of error, for identifying the frames in error and to request their retransmission.

A code to the times detector and spell requires to send on average half of the information transported in more. To send 1 000-bit security in to the receiver, it must therefore issue 1 500 bits. The detector code of error request a zone of 16 bits, sometimes of 32 bits. Each time an error is detected, it retransmits the whole of the frame. For frames with a length of 1 000-bit to 10 000 bits, a bit error rate of the order of 10^{-4} constitutes the limit between the two methods. A lower rate to 10^{-4} makes the technique for detection and application of retransmission more efficient that the correction of error only. As most of the lines of communication have a bit error rate of less than 10^{-4}, it is virtually always the detection method and resumption of frames or erroneous messages which is used.

Specific cases, such as the transmission through a satellite, can be optimized by a method for the detection and correction immediate. The time of the round-trip between the transmitter and the receiver being very long (more than 0.25 s), the negative acquittals claiming the retransmission take 0.5 s after the departure of the frame. If the flow is 10 Mbit/s, this means that 5 Mbit data have already been transmitted, which implies a significant management of the buffers of the transmitter and the receiver. Even in the case of a satellite, a optimization is usually obtained by for retransmission requests.

Before addressing the techniques for the detection and correction itself, the result of this section focuses on the functioning of a protocol of connection to show the solutions developed for the resumption on error. The transmitter formats the frames in adding to the data of the fields of supervision, address, data and error detection, and then it passes by maintaining a copy (see Figure 5.23), which is located in a memory.

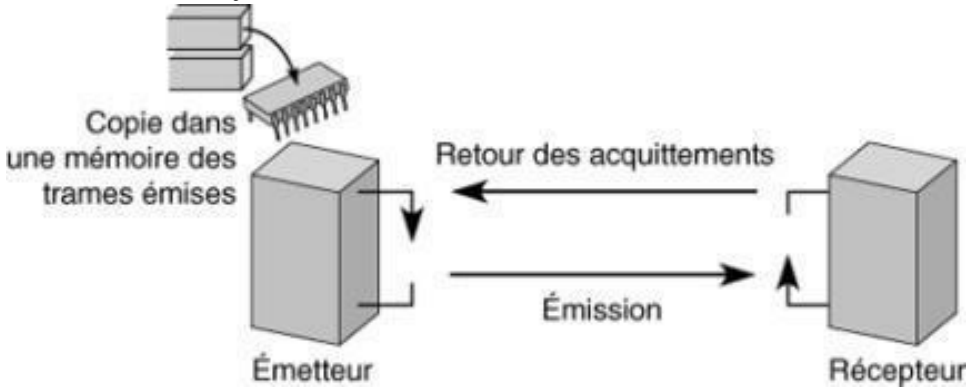


Figure 5.23

The receiver emits the acquittals positive or negative. At each positive acknowledgment, the corresponding data block is destroyed. To each negative acknowledgment, the frame which has not been correctly received is retransmitted.
The policies of acquittal and retransmission The most used are the following:

- The acquittals are sent to each frame received.
- The acquittals are accumulated to be sent all at the same time.
- The acquittals are issued in the frames transmitted in the opposite direction.
- Only the erroneous frame is retransmitted.
- All frames, from the frame wrong, are retransmitted.

Error correction

The error detection followed by a retransmission is the solution most used in terms of the transmission of information. Another solution can be implemented, however, consisting to detect and correct directly the errors. To correct, algorithms quite complex are necessary. More importantly, it must be put in the message to carry of the redundant information, which request a flow much more important. Among the techniques of correction, there are simple, which consist, for example, to send three times the same information and to choose the receiver information the more likely: to the receipt of a bit which is two times equal to 1 and once in 0, it is assumed that the bit is equal to 1.
There are more complex techniques, such as FEC (Forward Error Correction), which is to add to each block of kilobit additional bits to arrive at a total of *n* bits. We call such a code *k/n. The codes classically used correspond to the 2/3 or the 1/2, which indicates an addition respectively of 50 and 100 per cent of additional information. The complex encodings used resort to many algorithms, which are described in the suite that summarily. The Algorithms The most famous are those of Bose-Chaudhuri-Hocquenghem (BCH), Reed-Solomon and turbocodes.*
Suppose that the information can be decomposed into characters and that a character consists of 8 bits. You cut the DATA 8 bits by 8 bits. To be able to correct the errors, it must be able to distinguish between the different characters issued, even when the latter have a bit wrong. Either an alphabet composed of four characters, 00, 01, 10 and 11.
If an error occurs, a character is transformed into another character, and the error may not be detected. It is therefore necessary to add the information for that the characters are different from each other. For example, you can replace the four characters in base by four new characters:

- 0000000
- 0101111
- 1010110
- 1111001

So, if an error occurs on a bit, we can compare the data transmitted with the four characters above and decide that the good character is the one most closely resembling. If the receiver receives the value 0010000, it is immediately apparent that the character which would have had to be received is 0000000, since the other character the most closely resembling that we would have been able to receive is 0101111, which is much more differences with the value received.
If two errors occur on the same character, it is impossible in the context described to retrieve the exact value. These are errors residual say, which remain after the correction.
Formalize the method described previously. Either *(x,y)* the distance between the two characters *X* and *Y* defined by:

$$d(x,y) = \sum_{i=1}^{N} (x_i - y_i) \bmod 2$$

Where *N* is the number of bits of the character and *xi, yi* are the different values of *n* bits (i = 1… *N).*
The hamming distance is defined by the formula:

$$DH = \text{Inf}\, d(x,y)$$

Where the bottom terminal applies to the whole of the characters of the alphabet.
In the first example, the hamming distance is equal to 1. In the new alphabet, which has been determined to correct the errors, *DH* = 3. To be able to detect and correct a single error, it must be that the different characters of the same alphabet to meet *DH* = 3 so that, in case of error, the distance between the incorrect character and the exact nature of either 1. It deducted the character assumed correct in seeking the character whose distance is 1 by report to the incorrect character. It is understood that if the Hamming distance is 2 for an alphabet, we cannot decide the character the closest.
If one wants to correct two errors at the same time, it is necessary to have a Hamming distance equal to 5. Taking the previous example, you can replace the four characters by:

- 0000000000
- 0101111011
- 1010110101
- 1111001110

With this new alphabet, the hamming distance is 5. If the 10001010 character is received, it is deduced that the character is issued 11001110 since that *d(10001010,11221112) = 2 and that d(10001010,x) > 2 if x* ≠ 11001110.
These examples give the impression that it adds a lot of information for each character. This is true when the alphabet is composed of a few characters and that they are short. If there are a lot of characters, the number of information to add is proportionally much lower.

The error detection

There are many techniques for error detection. The parity bit, for example, is an extra bit added to the character positioned in such a way that the sum of the binary elements modulo 2 is equal to 0 or 1. This parity bit is determined from a character - it often takes a byte - composed either of successive bits, either of bits that determines in a specific way. This protection is fairly low performing, since it is necessary to add 1 bit all 8 bit, if the chosen character is a byte, and that two errors on a same byte are not detected.
The most common methods used are carried out from a division of polynomials. Let us suppose that both ends of the connection have in common a polynomial of degree 16, for example *x16 + x8 + x7 +* 1. From the binary elements of the frame noted I,i = 0…, *M* - 1, where *M* is the number of bits forming the frame, it constitutes the polynomial of degree *m* - 1 Next:

$$M - 1: p(x) = a0 + a1x + \ldots + am-1xM-1$$

This polynomial is divided in the transmitter by the generator polynomial of degree 16. The rest of this Division is of a maximum degree of 15, which is written in the following form:

$$R(X) = r0 + R1x + \ldots + r15x15$$

The binary values *r0, R1… R15* are placed in the frame, in the area of detection of error. On arrival, the receiver performs the same algorithm as the transmitter in defining a polynomial formed by the

binary elements received and of degree *m* - 1. It performs the division by the polynomial generator and finds a rest of level 15, which is compared to that which appears in the control zone of error. If the two offcuts are identical, the receiver in the deduced that the transmission is well placed. On the other hand, if the two offcuts are different, the receiver deducted an error in the transmission and request the retransmission of the frame erroneous.
This method allows to find virtually all errors that occur on the physical media. However, if an error slips in the area of detection of error, it is concluded to an error, even if the data area has been properly transported, since the rest calculated by the receiver is different from that carried in the frame. If the frame is 16 000-bit, that is to say if it is a thousand times longer than the area of error detection, it adds a time on 1 000 An error due to the technique of detecting itself.
The effectiveness of the method described depends on many criteria, such as the length of the data area to protect or of the control zone of error, the polynomial generator, etc. It is estimated that at least 999 errors on 1 000 are thus corrected. If the error rate on the medium is of 10^{-6}, it becomes 10^{-9} after the passage by the correction algorithm, which can be considered as a rate of residual error negligible.
The area of detection of error is sometimes called CRC (Cyclic Redundancy Check), the generic name of the method described above, and sometimes the Frame Check Sequence (FCS), or sequence of binary elements generated by the content of the frame.
The turbocodes are a class of solutions to detect and correct errors online in simultaneously using two codes which, individually, do not give extraordinary result. The turbocodes thus make a new method of encoding to two dimensions extremely effective for the correction of an error.

Architecture of the Routers

The routers are network equipment capable of router the blocks of the information which they arrive. These blocks of information can be packets (with regard to the level 3) or frames (for Level 2). The routers are the best known IP routers, since an IP packet has the full address of the recipient of the packet. There is a tendency to make the amalgam between router and IP router since the IP packet is imposed as the single standard for level 3. The appearance any recent of the routers to Level 2 aims to improve performance by dealing directly with the frame. The protocol architecture of a router of level 3 is shown in Figure 5.24.

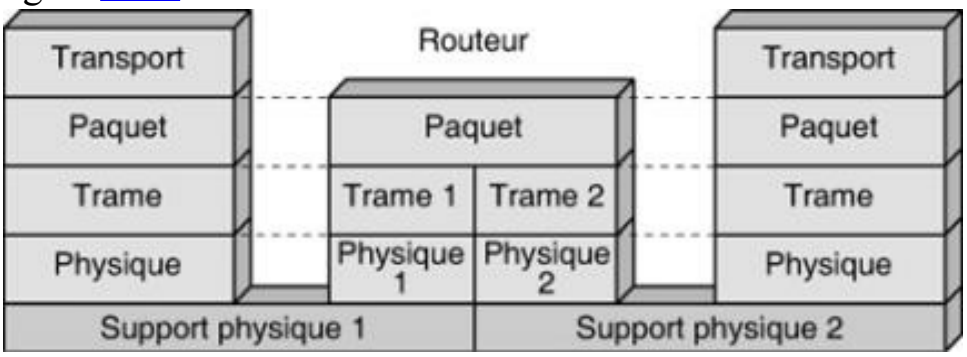


Figure 5.24 Figure 5.24
Protocol Architecture of a router

IP routers are differentiated from the switches by the treatment of the address of the recipient. In a router, the treatment is exercised on the full address and the consultation of the routing table. In a switch, the processing is performed on the reference and uses the switching table.
In the first routers, the search of the port of output is performed by software, which slowed dramatically the transfers. The overall speed of the router was in fact limited by the processing power of the Protocol and addresses. For a software router, ports reaching a flow rate of 100 Mbit/s are possible. Hardware routers use ASIC (Application Specific Integrated Circuit), microprocessors

to fast memories, which offer them a level of performance without common measure with that of software routers.

A router is composed of interfaces to access and output, of one or more processors, memory modules and a unit of interconnection. This last fact often appeal to a Switching technology which is detailed in the following chapter. The packet is first treated by the input interface, then the processor for the treatment of the routing table chooses the exit interface, the packet is stored in a memory module. The packet is then transferred to the exit interface by the unit of the interconnection. Figure 5.25 illustrates the internal architecture of a router.

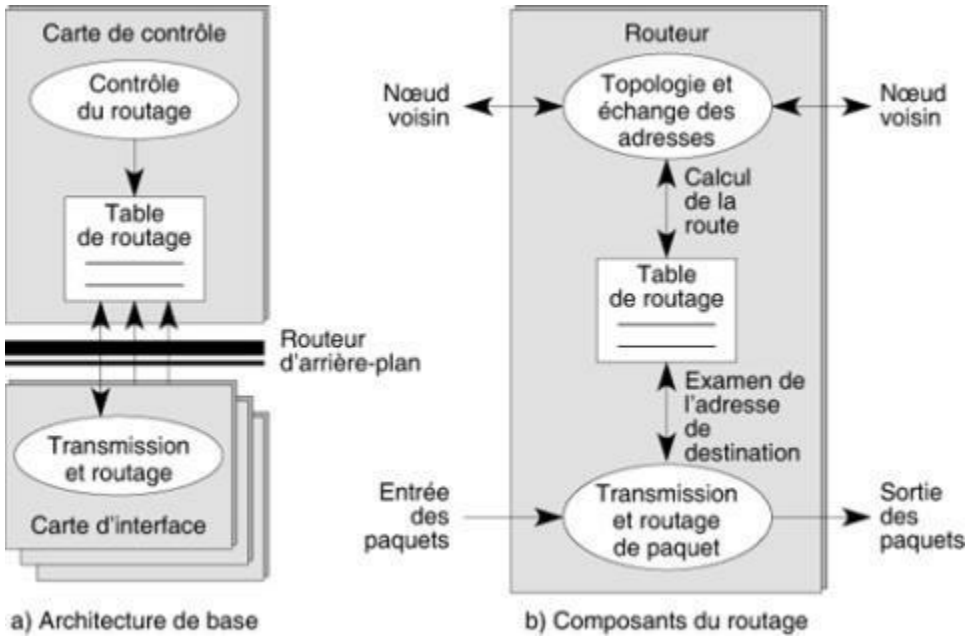


Figure 5.25

Internal architecture of a router

The determination of the output port is the responsibility of the processor managing the table, which must find the best route to go to the destination address of the packet. In an IP environment, it uses for this routing protocols like RIP (Routing Information Protocol) or Open Shortest Path First (OSPF) (see Chapter 10). This requires a knowledge of the topology of the network and possibly of the failures of the connections and strong congestion.

To route an IP packet, it must, in order, validate the packet and its different fields then search the output port, which may be local, remote, or multicast. In the latter case, it must eventually, after treatment, duplicate the package on several output ports. It is then necessary to check the time of life of the package (value of the TTL field, or Time To Live). In addition, it is necessary that it be years IPv4 or IPv6, recalculate the area of error detection. Eventually, it must fragment or reassemble the packets to make them compatible with the packages or the frames of the output line.

To accelerate the speed of treatment, it is possible to use a routing cache, as shown in Figure 5.26.

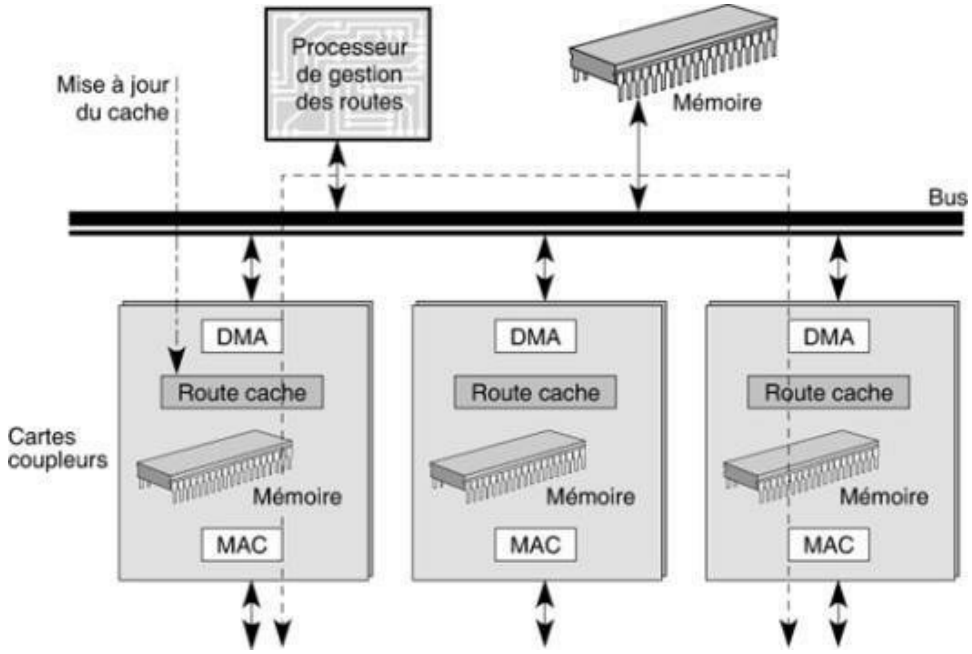


Figure 5.26

Routers with memory cache for the treatment of the address

Architecture of the switches

Architectures of the switches are quite similar in most of the switching technologies such as ATM or Ethernet switched as well as in the heart of switching of routers, that is to say in the central element of the router that allows for the transfer of a packet on one interface of entry to an exit interface once the next node selected. As most of the switches are at level 2, there is talk of switch of frames, even if it is in a way underlying switch with packets.

Many categories of architectures of switches have been proposed and types very varied OF VLSI circuits (Very Large Scale Integration), the famous silicon chips, have been developed in the research laboratories and among industrial sectors. The various publications resulting from these work distinguish three main types of architecture, as indicated in Chapter 2 : to shared memory (shared-memory), to shared media (shared-medium) and spatial division (space-division).

Given the high speeds of transmission lines modern, the switches must be able to transfer the frames at rates of extremely rapid, on the order of one hundred thousand to one million frames per second and by entry line. The realization of such switches request of high performance components.

The switches must be capable of withstanding trafficking both homogeneous than sporadic. In addition, the quality of service provided by the network being affected by the transfer time from end to end, and the probability of loss of frames, a differentiation of services is necessary. The objectives and performance criteria in this differentiation may however be opposed, such that the loss of no frame, but with a latency time important, or on the contrary the loss of frame, but with a reduced crossing time. A service of switching with priority is therefore essential for the different classes of services can coexist within the same switch.

A switch is a component with N inputs and N outputs that routes packets arriving on the entries to their destination of output.

The role of a switch is to ensure the three essential functions:

- Analysis of the header of the frame and its translation;
- Spatial switching, or routing;
- Multiplexing of frames on the required output.

The data of the users are transported in the data field of the frames and transferred asynchronously.

The fact of its statistical behavior and because a significant number of waves can share the same link, the switch must synchronize on the moments of entry of the frames in the node. The switch examines the header of each frame to identify the exit door of the frame. This identification is carried out either by the intermediary of the reference which determines the path, either by the full address of the recipient in the case of a routing. It converts the area of supervision in a new header for the switching node next, manages the routing and sends information control and management in the associated networks.

In an ATM switch, the switching is performed from the VCI (Virtual Channel Identifier) or the VPI (Virtual Path Identifier) contained in the header of the cell. The mechanisms of control of collision allow frames from different inputs to access the queue of a same multiplex, which is other than a communication channel that supports multiple streams simultaneously. Frames are switched individually, the internal clock of the switch working at a rate corresponding to the time of the transmission of a frame ATM. For example, if the line of communication The faster has a flow rate of 10 Gbit/s, the duration of the transmission of a frame ATM is 42.4 ns. In this case, the switch is punctuated at the rate of a Decision All 42.4 ns.

The Ethernet switches have to take charge of the frames are a little longer, 64 bytes to 1 500 bytes. The processing time is the same as the frame either short or long, recognizes the performance of a switch by the number of frames issued per second. Of course, it must take account of the average length of the frames to determine the speed of lines of output.

The achievement of a switch can be performed in various ways. In all cases, it is necessary to create a storage function, which can be found in the entry, exit or along the chain of switching. To the inside of the switch, various techniques of routing can be implemented: virtual circuit, autorouting or datagram. Two of the main functions provided by the switch correspond to routing and the storage of the frames. Optional features, such as the recovery of error or flow control, may possibly be implemented in the switches.

A switch must meet many constraints, including the following:

- Very high flow rate;
- Low switching delay time;
- Very low rate of lost frames;
- Management of multicast applications (multipoint communication);
- Modularity and scalability;
- Low implementation cost.

In addition, a modern switch must be provided of functions for the distribution and management of priorities.

The Gateways

We can no longer conceive a Network Without a passage to the outside. It must interconnect networks so that they can exchange information. The node which plays the role of intermediary is called a gateway, or *gateway* (generic term). This intermediate node may be more or less complex, following the similarity or the dissimilarity of two networks to interconnect. If the two networks are identical, the gateway is extremely simple. Conversely, if the two architectures to interconnect are dissimilar, the means to implement quickly become heavy and complex.

To thwart the development a little anarchic and the proliferation of Network Solutions Manufacturer side, the reference model has had for objective the standardization of network architectures. The objective was to avoid to move from an architecture to another through gateways, always costly and

complex to implement.
The IP interfaces have solved the problem for a large part of the heterogeneity of the network infrastructure. However, the solutions for transporting an IP packet remain very diverse, that this is to the inside of a company or in a network of operator. In addition, interconnection must also be done in the senior levels of the architecture.
This explains why the interconnection of different technologies makes necessary the use of gateways to connect different categories of networks. With the multiplication of networks, the Internet, mobile, wireless, etc., to which we have been witnessing for a few years, these gateways are have become indispensable, for both manufacturers and users.
In addition, the intermediate equipment that we encounter along a path to solve specific problems are also in full development, such as the firewall and the appliances in any kind to ensure the control of the traffic or the distribution of load.
The convergence of fixed/mobile is another important reason of the multiplication of the gateways, even if the systems that we put in place since several years often allow you to run the same applications on a mobile terminal and on a fixed terminal. It is in particular the objective of the IMS (IP Multimedia Subsystem). The new wireless networks require addition of machines Intermediaries to Manage cells, to the image of the controllers, which are detailed later in this chapter.
If it is strictly to the definition of a gateway, one can achieve an interconnection of networks at any level of the architecture of the reference model. However, the general rule is the following: the use of a gateway of level *N* is necessary when the lower layers to *N* are different, but that all the layers from the Layer $n + 1$, are identical.
The three categories of gateways are the most common bridges, routers and relay. It also distinguishes the bridges-routers (bridge-routers), which, although not standardized, are widely used.
A hierarchy of names has been defined to take into account the level of the interconnection in referring to the reference model. These different levels are illustrated in Figure 5.27. A repeater is a gateway of level 1, or physical; a bridge is a gateway of level 2, or frame; a relay is a gateway of level 3, or package; a transport relay is a gateway of level 4, or message, etc.

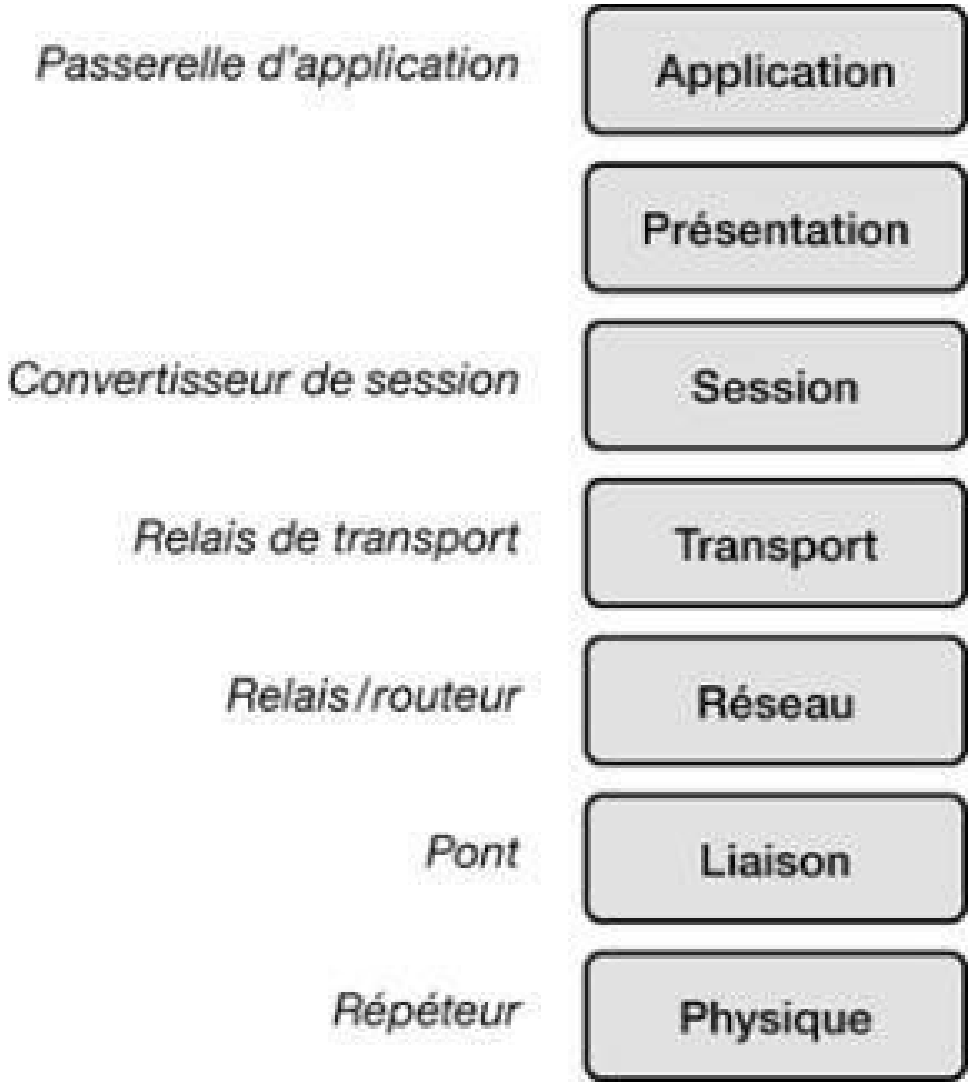


Figure 5.27 Figure 5.27
Hierarchy of gateways

The terms "switch" and "router" are not related to a level. A switch is a body of bridge type when the switching is performed at level 2 and relay type when it is carried out at the level 3. For example, an Ethernet switch is to bridge type while a switch X.25 is of type relay. Similarly, a router is to bridge type when the routing is carried out at the level 2 and type relay when the routing is carried out at the level 3. The term "router" has been so associated with the Layer 3 IP routing that it seems natural to use it to indicate a level relay packet. It is however true that for the world IP, which represents all the same almost 99 per cent of level relay 3.

The repeaters

A repeater is a gateway of the physical level between two networks with a level common frame. For example, an Ethernet repeater is a device that automatically repeat the frames of a strand Ethernet to another Ethernet strand.

The role of the repeater is to send a frame further than does the allows a simple cable, whose length is limited by the attenuation of the signal. In the case of Ethernet to 10 Mbit/s, a shielded coaxial cable may not exceed a length of 500 meters under penalty to see the error rate become unacceptable.

Let us look at more specifically the case of the Ethernet network. We know that the maximum coverage of an Ethernet network to 10 Mbit/s is limited to 2.5 kilometers, since the time of propagation from one end to the other of the physical media may not exceed 51.2 microseconds. The question is how to achieve these 2.5 kilometers if the maximum length of a strand may not exceed 500 meters. The answer is simple: it is sufficient to connect of the strands to each other by using

repeaters.
The repeaters do not prevent collisions, but make it difficult their repetition on the strand next. In effect, a repeater is not something other than a shift register, i.e. a set of records in which the information in the form of 0 and 1 come to store and move for leave to enter a new element binary. The Register of entry expects to receive a 0 or A 1 and not a signal from an overlay. It is therefore very difficult to repeat of signals which are neither 0 nor 1. This is the reason for which the repeaters to replace the elements in collision by a series of bits specific enabling to other stations to detect the collision.
The repeaters can possibly change of physical media while respecting the structure of the frame in the course of shipment. For example, you can pass a metal bracket to an optical fiber or a terrestrial support of a wireless network. This is the reason for which it is possible to achieve of Ethernet networks with metal parts, optical and microwave links.
In summary, a repeater is a body which foolish allows to extend the length of the physical media, to the contrary of a bridge, which filter messages on their destination address.

The bridges

The bridge, or *bridge, is a gateway of level 2. This network equipment is fairly simple to implement has evolved a lot since the onset of the first Ethernet networks.*
A bridge unit nearby networks or in remote locations by dating back up to level frame. It receives a frame and calculates the output line thanks to a routing algorithm or to the switching table. It filters the received frames by examining the address of level 2 and not leaving go that the frames destined for the outside.
The architecture of a bridge is shown in Figure 5.28. The bridge creates a virtual network from a set of sub-networks, ignoring the protocols of the higher layers. The Layer 2 is in fact divided into two sub-layers: the layer MAC (Medium Access Control) and the Layer LLC (Logical Link Control). The bridge can accept controls to access different, but must have the same protocol of connection.

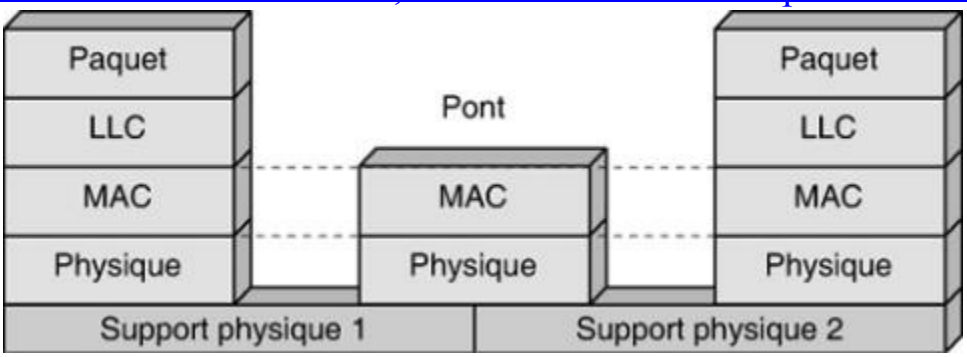


Figure 5.28 Figure 5.28
Architecture of a bridge

The bridge saves in internal tables the addresses of all stations connected to the network. If a station is added or removed, the system must be reconfigured. This is the reason for which the bridges cannot *a priori* be used that in environments well localized. As soon as the number of stations is important, the addresses management becomes very complex.
The interconnection of sub-networks by bridges allows high flows, since the number of levels to cross is small and that it only goes as a level to arrive at the level frame. The gateways of packet level, or relay, are less powerful because, each passage of a relay, it must cross the levels 1 and then 2 to arrive at level 3.
Two major routing protocols level bridge, spanning tree and Source-Routing, have been developed respectively for Ethernet networks and Token-Ring. As the solution Token-Ring has almost

disappeared, it remains in the facts more than almost the Spanning Tree. However, the Source-Routing being used in other circumstances, it is also described briefly in the suite.

The Spanning Tree Protocol

Normalized in 1990 by the IEEE committee 802.1, in the working group IEEE 802.1D, the Spanning Tree Protocol (STP) is planned for the interconnection of any type of network. It consists of the Constitution, from any topology, a shaft which covers perfectly the network and in which, from any sheet of the shaft, any point of the network is accessible.

For the proper functioning of the Protocol, the network must meet the following conditions:

- A unique identification (ID) must be associated with each bridge in the network.
- The bridge with the lowest ID must be selected as the root of the tree.

Bridges exchange of messages called "Hello", in which they indicate their ID as well as the ID of the bridge that they consider as the root of the tree by which must pass their frames. When they receive an ID lower than that designated as their root bridge, they rectify the ID of the bridge which serves as their root for take the new value. In other words, they determine a new root bridge. With the time, each bridge ends by determine the root of the tree, that is to say the root bridge. Then, each bridge calculates the distance which separates it from the root. This distance is calculated in the vicinity: at each bridge crossed, the distances are incremented by 1.

On each physical network, a bridge is chosen as the closest to the root. If two bridges in the same network are the same distance from the root, the smallest ID is chosen. All the traffic from this network and to a destination in another physical network passes through this bridge, called a bridge elected. Thanks to this Protocol, any physical network is similar to a virtual tree, and there is no loop in the network.

We can blame this protocol of possibly performance dependent on the topology of the network. In addition, if the ids of the bridges are not defined by the manager of the network, but by the manufacturer, the bridge elected as root is independent of the will of the manager and may constitute a bottleneck.

The spanning-tree algorithm has variants of which the most interesting is the Rapid Spanning Tree Protocol (RSTP), standardized in 1998 by the group IEEE 802.1w. This algorithm allows you to bring the convergence of the Protocol of 30 seconds on average to 6 seconds on average.

The Protocol Source-Routing

Standardized by the IEEE committee 802.5, the Source-Routing protocol was used initially for the interconnection of networks Token-Ring. This protocol is still used in other contexts, as well from the IP networks that local networks. An important example studied in Chapter 17 comes from standardized protocols for ad-hoc networks.

When a station X wants to send information to a station Y, it sends in diffusion a frame of discovery of the path. A bridge which sees arrive a frame of this type Y adds its own address and transmits this frame to all networks, with the exception of the one by which the frame arrived. The destination station is therefore seen as arrive one or several frames and returns to X all received frames using the routing information found in each. Then, X can use the roads that the protocol has allowed him to discover. His choice is guided by various parameters, such as delivery times, number of bridges that are crossed, length of frame permitted, etc.

The frames constituted by each station have the following structure: they begin by the destination address, followed by the source address, routing information, address DSAP (destination service access point), the address SSAP (Source Service Access Point), control data and finally the data to Transport, to finish by a zone FCS (Frame Check Sequence). This suite is expressed by the sequence:

@Dest.~@Source~info-routing~DSAP~SSAP~Control Data~~FSC.

The length of the destination address and source is of 2 or 6 bytes, and that of each element of the routing information of 2 bytes.

The relay-routers

As indicated at the beginning of this Chapter, "relay" is the term standardized for indicating a gateway of level 3, or packet. As the world IP has now almost the exclusivity of the Level 3, the trend is to use the term "router" to express a level relay IP packet. Unfortunately, this term may be confusing when one speaks of a router to level frame. The concept of router is not linked to a level, but to a technology. Talk about router, it is therefore the most often talk about IP router, what makes this chapter. However, it must be remembered that a router of level 2 is imaginable, even if this is not a classic case, if frames contain the full address of the recipient.

The multiprotocol routers

The multiprotocol routers are distinguished by the range of network protocols managed as well as by the number and type of network interfaces supported. These products are relatively complex, which explains that a low number of companies have specialized in these routers.

A multiprotocol router has multiple interfaces of frame level and several protocols for packet level. When the frame is present in the router, it is décapsulée so that the packet to be recovered. After a review of the address area of the packet, the latter can be transcoded into the format package of another protocol before to be encapsulated in a new frame structure.
The multiprotocol routers can withstand a bridge-router, or bridge-router (see later). The node can in this case recognize the reference or the address of frame level and router or switch the frame without going back to the level package. If the reference or the address of frame level is not recognized, it goes at the level package to route the packet on the address of level 3.
The difference of a bridge, a router can isolate some segments of the network and create fields. It allows you to offer a good isolation between each connected network, thus avoiding the spread of signals emitted in broadcast. Currently, the speeds reached by the enterprise routers are 10 000 to 15 000 packets/s and are close to often the 100,000 packets/s. The fact of the constant increase in the flows of applications, it took, at the end of the years 1990, focus on the design of routers much more powerful, in particular for the operators, capable of router A to one thousand million packets per second. They are detailed in the next section.

The gigarouteurs

The generation of broadband routers, called gigarouteurs or térarouteurs, is based on a distribution of the routing table and processing the packet in the access interface and then on the use of a switch to transport the packet to a port of entry to a port of exit.
Figure 5.29 gives an idea of the architecture of a gigarouteur. The gigarouteurs allow you to access ports to reach speeds of 10 and 100 Gbit/s. The transmission of IP packets at these speeds is operated today by the Techniques IP on Synchronous Optical Network (SONET) and MPLS.

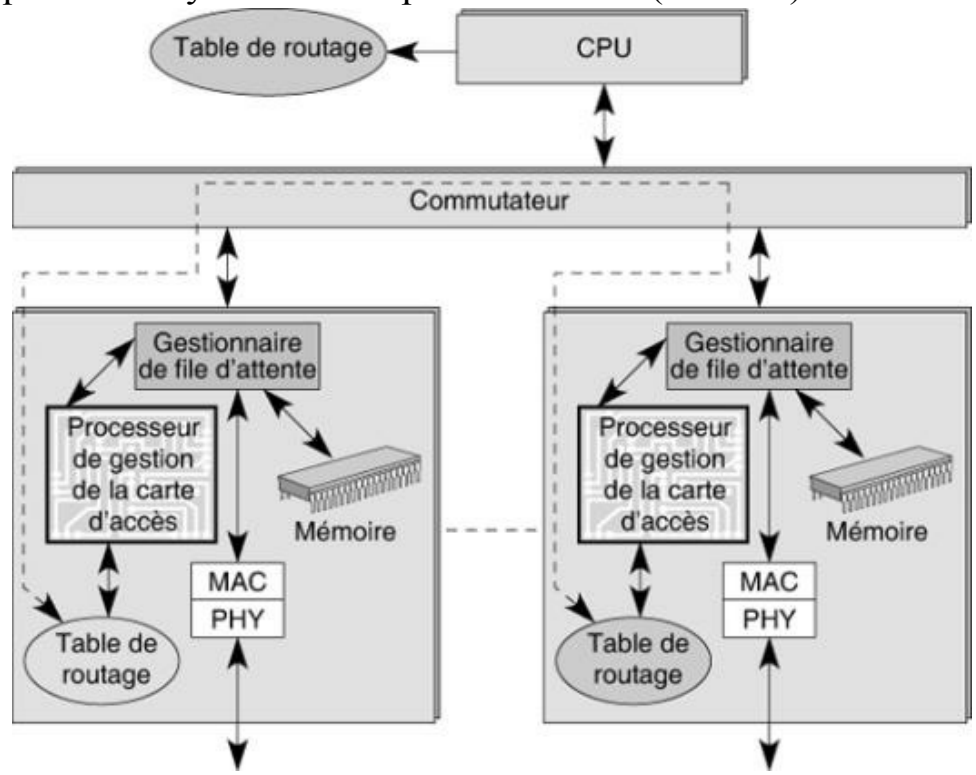


Figure 5.29
Architecture of a gigarouteur

The Switches form the heart of the Routers very high flow rate to allow access to several hundreds of megabits and even gigabits per second.

The bridge-routers

The bridge-routers, also called B-routers or bridges-routers, have a role to play to combine the best of both technologies. They incorporate, depending on the products, the three lower layers, physical, liaison and package, and are trying to act on the level link when they have the possibility, in the absence of what they date back to the packet to treat the address of packet level. In other words, the bridge-routers receive a frame that is treated as if the body was a bridge. If the bridge frame level does not determine the direction in which he must send the frame, it is décapsulée to retrieve the package that she carries in its area of data. The gateway has now a packet to examine, and it plays the role of gateway of level 3 which is usually a router IP.
The bridge-router is a body quite complex since it requires management tables of frame level and level package. On the other hand, it is a piece of equipment very efficient from the point of view of the possibilities of treatment addresses and references.

The techniques of tunneling

The techniques of interconnection encountered up to now concern only the translation of the information in a frame to another frame or a packet to another package. Another method, completely different, called encapsulation, is to place a frame to the inside of another frame or a package to the inside of another package.
For example, the interconnection of an IPv6 network with an IPv4 network can be done in the following way. Suppose a customer IPv6 wants to transmit an IPv6 packet to a customer who works on a machine terminal end IPv6. Suppose also that the only network that interconnects these two machines either the Internet of type IPv4. A first solution would be to make a translation, that is to say the transfer of the inside of the IPv6 packet in the IPv4 packet and, on arrival, transfer again the content of the IPv4 packet in an IPv6 packet. This solution is possible, but complex, because it must redefine completely the areas of supervision of the packets transferred. This is the reason for which it prefers to use another method: In the Machine terminal end of the transmitter, it encapsulates the IPv6 packet to the inside of an IPv4 packet. The IPv4 packet is transported on the Internet and, on arrival, decapsulates the IPv4 packet to find the IPv6 packet. It was in fact used the IPv4 network as a tunnel.
To interconnect two networks without resorting to a gateway, the use of a tunnel is classic. This is what is called make the tunneling.

Translation and encapsulation

The two main levels of interconnection are, as indicated previously:

- The level frame, with bridges;
- The level package, with routers.

If it remains at the level bridge, two solutions are possible: the translation and the encapsulation.
In the translation, the source and destination addresses of the terminal stations are conveyed in the headers. In the encapsulation, a complete frame coming from the local network is included in the frame of the network that will serve as a tunnel. This method does not request processing of the frame, but as it is not standardized, it has the disadvantage to restrict itself to a world homogeneous, that is to say to go to a terminal station with a X protocol to a terminal station using the same protocol X.
Figure 5.30 illustrates the architecture of a technique for the encapsulation packet level. It is assumed that a terminal machine of a company uses the IPv6 protocol and that it wants to connect to the local network IPv6. The customer is represented by the left battery and the company by the right stack. The protocol specified with the value 3' is therefore IPv6.
To interconnect this station and the local network, only the network the IPv4 Internet is available. IPv4 is represented by the protocol indicated by the value 3. The terminal station encapsulates its IPv6 packet (Protocol 3') in an IPv4 packet (Protocol 3). This IPv4 packet is transported on the Internet until the router to access of the company, which is symbolized by the protocol stack in the middle. In this router, the IPv4 packet (Protocol 3) is decapsulated to find the IPv6 packet (Protocol 3'). This IPv6 packet is then transported in IPv6 in the local network, represented by the right part of the diagram.

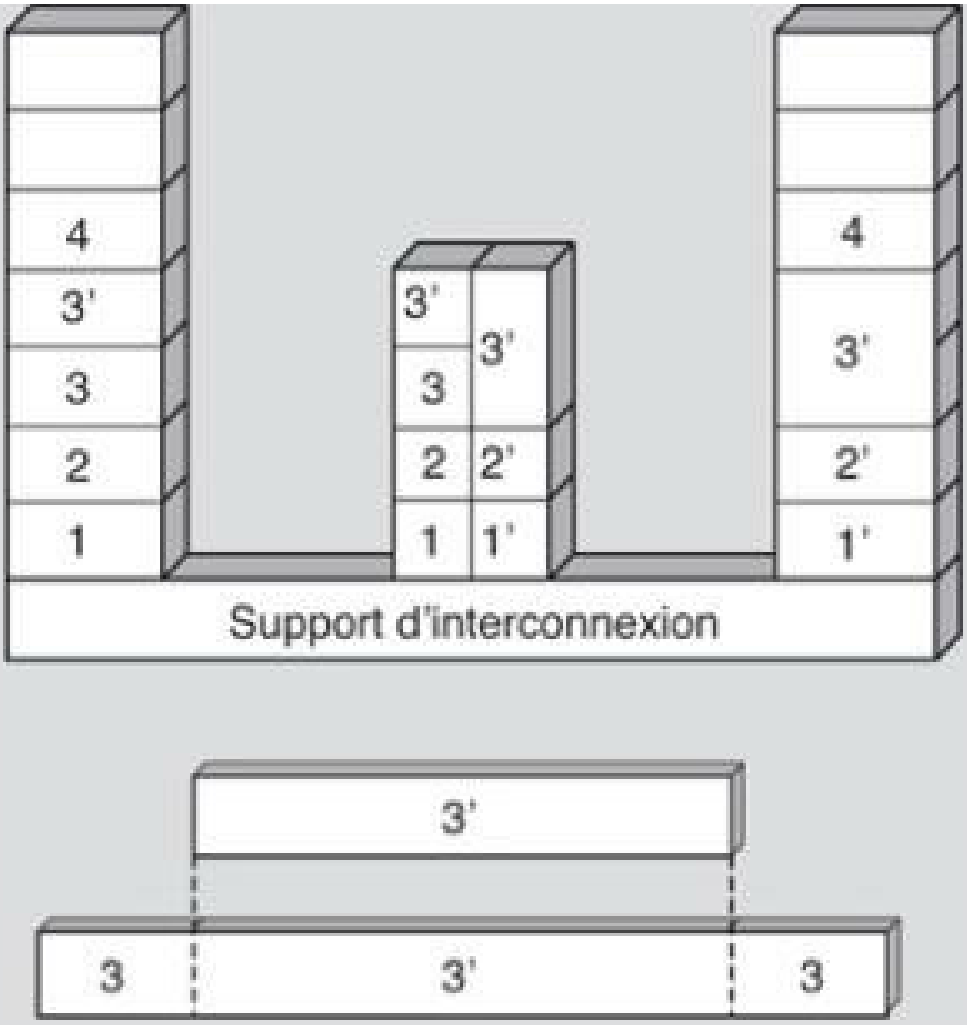


Figure 5.30

Encapsulation of packet level

The same solution is offered to the network designer to interconnect machines IPv4 by the intermediary of an IPv6 network. It is enough to encapsulate the IPv4 packet in the IPv6 packet, then, at the arrival, decapsulating the IPv6 packet to find the IPv4 packet.

The two solutions of encapsulation are comparable. The one that will be the most practiced will depend on the way to move from IPv4 to IPv6. A first solution is to assume that an operator decides to propose a network IPv6 to perform the transfer of packets for the simple reason that with IPv6 it will be able to offer more services to these customers that with IPv4. The clients will remain surely yet some time in IPv4 before go in IPv6. It will be then to encapsulate the IPv4 packets into IPv6 packets of the operator. Now, if this are the customers who decide to switch to IPv6 - because they can indicate more information in their areas of supervision -, but that the operators remain in IPv4, there will be different encapsulations of IPv6 packets in IPv4 packets.

The firewalls

The features of the firewall are analyzed in detail in Chapter 24. This chapter introduces these network devices, because they become more and more necessary in the networks of today, even for a particular As soon at it relates to the Internet.

A firewall, or firewall or *firewall, is, as its name indicates, an equipment whose objective is to separate the world outside of the inner world to protect. Its role is not to let enter that packets whose company is sure that they do not pose a problem.*

Firewalls provide many functions, whose main is to sort what is in and what fate and decide on an action when the recognition has been performed. The actions can go of the rejection of the packet to its compression-decompression, in passing by its review by an antivirus, its slowdown, its acceleration, etc.

Various means are being implemented to recognize a packet and more generally the stream, such as the recognition of the application which passes through the firewall, the address of the destination or source address, the machine and the application on which Remote The wants to connect, etc.

The firewalls are distinguished by the level at which they work. As a general rule, they are level 4, or Message: we are trying to find the message in TCP level a way to recognize the application. The users are differentiated by their source and destination addresses, but especially, in the first generation, by the port number, which indicates the current application. For example, port 80 indicates an http application. However, the port numbers are less and less reliable, because the attackers used the open ports and often the port 80 using the HTTP protocol as a capsule in which they integrate their message.
The use of port numbers is quite restrictive, in the measure where more and more applications have the dynamic ports, such as FTP, most applications P2P (peer-to-peer) or the telephone signalling. In addition, two customers can determine between themselves a port number on which they wish to communicate.
The evolution of the firewall has been to fit in the layers of protocols to achieve the application layer in order to be able to determine the current application. Is called application-layer firewall, or firewall to level 7, the firewall that are able to distinguish clearly between the applications.
If the reproach long sent to the firewall was to take a lot of time and not be able to determine the applications to the wire of the water, this is no longer true today. The products of firewall applications introduced on the market since some time do not take more time than most of the network equipment encountered in the IP world. Such is the case of the enclosure QoSMOS, which is capable of filtering and determine the applications in a very short period of time, so that the output of packets is delayed as a maximum time equal to the time of crossing a current router.
The firewall is often installs in a dedicated unit to simplify its implementation, but it can also be found in different points of the network, ranging from the router to the switch, passing by a specialized server or client.

The proxy

The Enable Proxy to break with the classic model client-server in a communication in prohibiting a direct connection from the client to the server. There are two main types of proxy, proxy of application type and the proxy type of circuit.
The application proxy speakers at level 7, or application, with the aim of breaking the client-server model to go to model client to client. The latter do not allow a TCP connection of end-to-end and are rather intended to outgoing traffic of authenticated users.

The Application Proxy

As explained earlier, the Application Proxy speakers at the application level of the reference model. Their objective is to break with the client-server model classic in the replacement by a double client-server model, as shown in Figure 5.31. The direct relationship is cut off to be replaced by two relationships with the proxy making the transition between the two relationships that is to say between the proxy playing the role of the server and the proxy playing the role of client. In other words, a TCP connection from end to end is replaced by two connections placed end to end through the proxy.
This solution provides a good security since it must run the application in the proxy, which allows you to check that the flow of packets does not form an attack. You can achieve firewall proxy type which are equivalent to the firewall of level 7. The major drawback of this solution is the cumbersome nature and the difficulty of obtaining good performance.

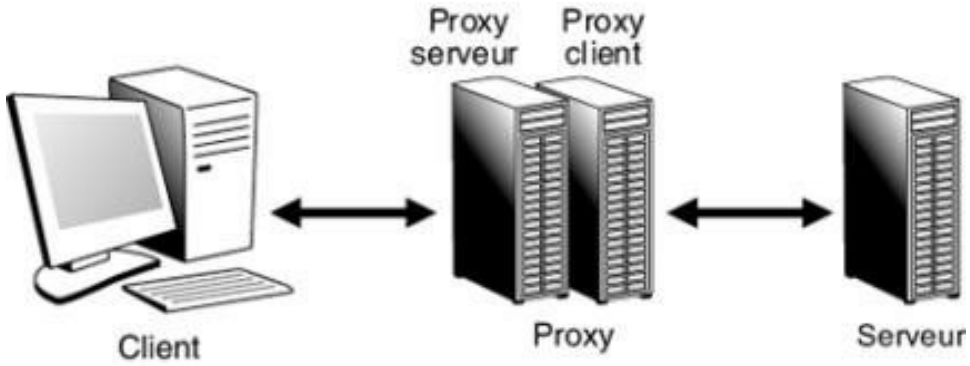


Figure 5.31
Application Proxy

The circuit proxy

The proxy to circuit type have the objective to check that the result of packets on a path, or virtual circuit, is consistent with the corresponding RFC. Indeed, many of attacks are carried out by inserting in the normal flow of packets of packets of attack. With a proxy circuit, the various fields of the packets are verified to ensure that no package bears a attack.

This solution also offers a good security by requesting an authentication of the user who will use the path, at the beginning of its connection.

The appliances

The appliances are the enclosures that have one or several functions well determined and which fit easily into the network. The advantage of these enclosures is generally to be able to start a new functionality without having to program or to adapt the existing software. The appliances can be used to the security and, therefore, integrate a firewall, but also, through specific features, to the management of the quality of service.

This section is primarily devoted to appliances to carry out the monitoring of the quality of service and the acceleration of IP flow.

The appliances on the monitoring of flows are used to determine the different flows transiting on the Internet and, after recognition, to treat them. The treatments can be extremely diverse following the enclosures (loss, compression, hold, acceleration, etc.).

For the recognition of flows, many opportunities are available today, the more traditional to use the port numbers. However, as the applications the most modern use dynamic ports, this solution is sometimes disastrous from the point of view of the recognition of the waves and therefore of the security or the management of flows of IP packets. One solution is to recognize the waves by their grammar, that is to say all the rules to follow to achieve the writing of application messages. As the grammar is unique for each application, it is possible to recognize a stream, even if it is encapsulated in other waves, as in L2TP tunnels. Once the flow recognized, the enclosure can perform a function decided by the Network Manager and programd in advance.

We can rank among the appliances switches or routers of Level 4 or Level 7, that is to say capable of switch or router as a function of information collected at the message level or application. For example, in function of a port number or a recognition of the application, the routing decision or control may differ.

The accelerators of IP streams can also be stored in the appliances. These incorporate a means to send to the issuer an answer more quickly or to perform a data transfer, a point to another, in less time than without accelerator.

The accelerations may be carried out at the different levels of the architecture. As a general rule, the

higher the level is low, the more the global acceleration is important. Conversely, the more the level is high, more the acceleration is slow and intended for specific applications. For example, it is possible to compress the flow of level 1, and thus reduce the number of packets to transmit, or to reduce their size, which leads to a lower load to the inside of the network and therefore a better time of transit. At level 2, one can conceive of the accelerators for the correction of error when the error rate is important. At level 3, you can play on the IP addresses and on the content of the headers of the IP packets. Finally, at senior levels, we can work on specific applications rather than on all applications simultaneously, such as at levels 1, 2 and 3.
The appliances also concern the acceleration by the storing of information in intermediate caches, i.e. in buffers that are located relatively close to the inputs of the network operators. It puts in the cache to be whole pages of information, if the latter are strongly requested, so that it is not necessary to go and look for the page on the origin server, which can be located at the other end of the earth. You can also save a part of the page and not seek that additional information. For example, for a Web page that has a bottom gourmand enough in bytes, only the background is kept in a cache in close proximity to the customer, and only are requested to the server the type information text to update. The flow rates measured in this solution represent only 5 to 20 per cent of the total flow necessary for the transport of the full page.

Conclusion

The convergence at the level packet to the IP technology does not prevent a persistence of techniques for the interconnection of networks. Indeed, at the level frame, a high diversity still exists between the frames ATM and the different Ethernet frames. Similarly, at the level package, the breakthrough of IPv6 will ask of the interconnections IPv4-IPv6 during a certain time yet.
The trend of the major operators is to converge all their networks heart (telephone network, data networks, heart Networks Networks of mobile, etc.) to a single network for the transport of IP packets. To route these data, IP packets are routed either in routers, either encapsulated in frames, to be the most often switched. To allow for a safe transport of these packets, many solutions are marketed with more or less power and success.
The appliances offer various functions while remaining generally simple to implement. Their main role is to improve the performance of the network by means of extremely diverse.

6

The frame level

As explained in Chapter 2, the level frame is the level where flows the entity called the frame. A frame is a suite of binary elements that have been gathered together to form a block. This block must be transmitted to the next node so that the receiver will be able to recognize its beginning and end. In summary, the role of the frame level is to carry information on a physical media between a transmitter and a receiver. Its main function is to detect the beginnings and the purposes of the frames.
The level frame is fundamentally different next that one is dealing with a network of frame level, i.e. a network that only goes in the intermediate nodes to the Layer 2, or Layer liaison, for router or switch, or packet level, i.e. a network that should go back to the Layer 3, or Layer network, to perform the transfer. In an architecture of frame level, the header contains the areas which are used to route the frame to the recipient of the message. In an architecture of packet level, the data area of the frame contains a packet. The necessary information to the routing of the message are located in the areas of supervision of this package.
This chapter presents the main frames that allow you to carry packets in an architecture of frame level, PPP, Ethernet and ATM. The HDLC protocol, which has been very popular during the period of domination of X.25 and which has almost disappeared today, is detailed in Annex D. This standard HDLC is important, however, because it remains a model to understand all the features of this level.
The frame the more important is the Ethernet frame. It is she who dominates the world of small networks, corporate networks and networks of operators. It has also taken a place without sharing in the connections of ADSL type and in wireless networks. It is finally the frame used, slightly modified, but always compatible, to the inside of the datacenters and between data centers. In particular, the Protocol trill (Transparent Interconnection of Lots of links) the uses of a particular way with RBridges (Routing Bridge) or switches trill, that could also call of the Routers-bridges.
The Annex D also focuses on the old generation of frame level, called connection level, whose function was to the times to play the role of the frame level and correct errors in line in order to make the acceptable error rate for the upper layers. In this latter case, a detection zone of error is added to the end of the frame in order to detect if an error occurred during the transfer. If this is the case, two methods can be implemented to carry out the repairs. In the first, control bits added by the transmitter allow to detect and then correct the errors. In the second, a retransmission is requested to the previous node, which must have kept a copy of the frame.

The architecture of the frame level

The level frame (layer 2) has the function to make a service at just the level higher, i.e. at the packet level (layer 3). This service relates to the transport of packets from node to node. More specifically, its role is to carry a packet of the Layer 3 or a fragment of a message from the Layer 4 from one node

to another node. For this, the frame level requested in turn at the level just below, the physical level, a service, consisting of carry the bits of the frame from one node to another node. This section presents the functions necessary to the achievement of all of these actions.

The features of the frame level

The specificity of the level frame is to transmit the information as quickly as possible between two nodes. The important part of the protocols of frame level lies in the structure of the frame and in the way the recognize and deal with it in the shortest possible time.

The first feature to implement concerns the recognition of the beginning and the end of the frame when the flow of binary information arrives at the receiver. How to recognize the last bit of a frame and the first bit of the next frame? Several generations of protocols are the successor to try to bring the best response to this problem:

- The first generation of protocols of frame level involved a recognition by flag: the beginning and the end of the frame were recognizable by the presence of a suite of binary elements, which had to be unique. To this effect, insertion techniques were used to break the Suites which would resemble a field of beginning or end, called flag or pennant (flag). The insertion of additional bits, however, presents a disadvantage, since the total length is no longer known in advance and that specific mechanisms are required to add these bits to the transmitter and then cut to the receiver.
- The second generation has tried to find other modes of recognition, mainly violations of codes or systems using keys to locate the beginning and the end of a frame. The advantage of these techniques is to confer on the frame a determined length and not lose any lag time to his arrival.
- The trend of the current generation is to go back to the first solution, but with a flag long enough that the likelihood of finding the same suite of binary elements is almost zero. The advantage of this solution is to allow a recognition very simple of the flag without having to add bits or to calculate the value of a key. The Ethernet frame corresponds to this definition with a flag, called Preamble, of a length of 64 bits.

The features of the frame level have been significantly modified since the beginning of the years 1990. It was done, for example, down in this level the functions of Layer 3 (Network), or packet level, with a view to simplify the architecture and increase performance. Figure 6.1 illustrates the potential roles of the frame level compared to the reference model.

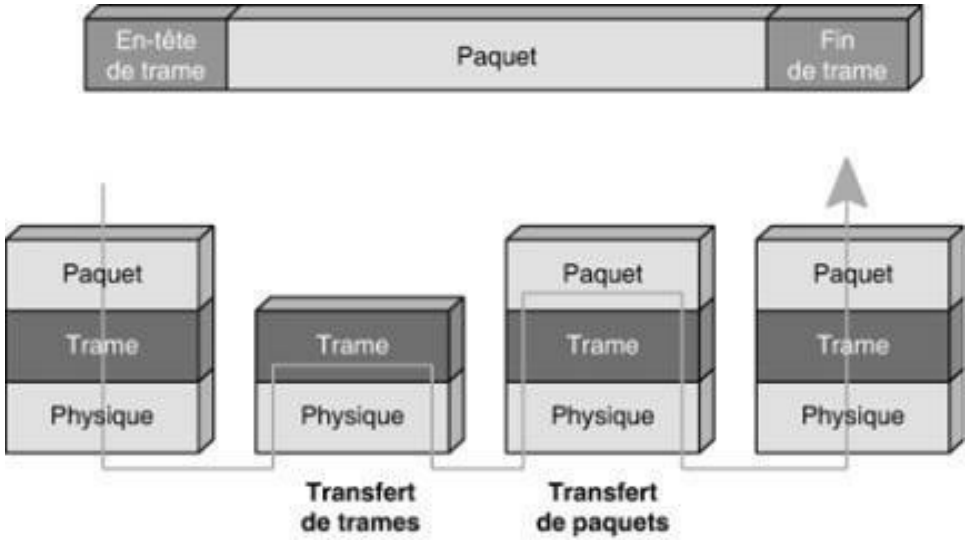


Figure 6.1

Potential roles of the frame level compared to the reference model

The main protocols used by the frame level are the following:

- The protocols from the international standardization in the years 1980 or taking birth, such as the basic protocol HDLC and its derivatives lap-B (Link Access Procedure-Balanced), lap-D (Link Access Procedure for the D-channel) and Lap-F.
- The Protocol PPP (Point-to-Point Protocol), which allows to carry an IP packet from one machine to another machine. This protocol is strongly used in the framework of personal networks to connect for example two PC between them.
- The protocols from the switching of Ethernet frames: several protocols are used in this framework on the basis of what is sought. Overall, there is the Ethernet protocol shared and switched Ethernet.
- The ATM frame is still used in networks of operators, but it is in sharp fall by the omnipresence of the Ethernet frame.
- The "Label Switching" the Protocol MultiProtocol Label Switching (MPLS), which uses in fact previous frames, but with a IP signalling.
- The protocols for travel with very high flow inside and between data centers.
- The protocols from local networks virtual, or VLAN (Virtual Local Area Network), adapted to the major networks of operators. In this framework, we can cite the Ethernet networks Carrier Grade, capable of a high level of service from end to end.

The level frame may have for additional function the management of access to the physical media, such as in networks with communication support shared. We then speak of function MAC (Medium Access Control).

The addressing of Frame Level

As indicated previously, the level frame is derived from the Layer 2 (link) of the reference model. The addressing has therefore not been particularly studied, since it had to be taken into account by the Layer 3 (network). As a general rule, the addresses used are very simple: one for the transmitter and one for the receiver. We can make this more complex model by introducing a connection not more point-to-point, multipoint but, in which several machines to share a same media.

The first address for multipoint networks emerged in the frame level comes from the Ethernet environment and more particularly of Ethernet networks shared. As all clients of a local network

shared Ethernet Connect on a same cable, when a station emits, each station can receive a copy. This is what is called the Broadcast feature. The addresses are emerged in this framework, because it was necessary that a station can distinguish a frame destined for him of a frame destined for another user.
The Ethernet address is thus born of local concerns. We often talk of the physical addressing because the address is located in the map in connection. It was necessary to avoid that two cards, even coming from different manufacturers, have a same address. As indicated in Chapter 9, devoted to Ethernet networks, the address has taken a flat structure, or absolute. Each manufacturer of Ethernet card has a number manufacturer on 3 bytes. This value, it is sufficient to add a serial number, on 3 bytes also, to obtain a unique address. Since there is broadcast on the local network shared, it is easy to determine where is physically located the address of the receiver.
New problems have arisen with the arrival of the new generation Ethernet replacing the diffusion mode by a function of transfer of frames, called Ethernet switching. The reference used for this switching is simply the address on 6 bytes of the card of the recipient. An additional area in the Ethernet frame has been added in order to have a true reference, independent of the Ethernet address. This area with the *shim-label* is used especially in the MPLS networks.
The ATM cell is also place at the level FRAME: it detects the beginning of a cell by counting the number of the received bits since the ATM cell has a constant length. In the event of loss of synchronization, it is always possible to find the first bit of a cell through the field of error detection and correction, which is located in the header. ATM using a dial-up mode, the header contains a reference.

The protocols of Frame Level

The protocols define the rules for that two entities can communicate in a coordinated manner. For this, it is necessary that the two interconnecting entities use the same protocol. To simplify the communications of frame level, many of the protocols have been standardized. The oldest, HDLC, is hardly ever used, but remains a good example procedure for frame level.
A protocol of frame level that is often used without the knowledge is PPP, which allows you to connect two PC between them. It is described briefly in the suite. The ATM protocol, which has strongly decreased, is also introduced, as well as different solutions for the use of the Ethernet frame, promised to become the reference frame.

The Protocol PPP (Point-to-Point Protocol)

PPP is used in the Connections to access the Internet network or on a connection between the two devices, that they are of personal computers or network nodes. Its role is essentially to encapsulate an IP packet in order to transport it to the next node.
While strongly inspired by the HDLC protocol, its function is to indicate the type of information carried in the data field of the frame. The network Internet being multiprotocol, it is important to know how to detect, by a specific field of frame level, the application that is transported in a way to be able to send it to the good output door.
The frame of the PPP protocol is similar to that of HDLC. A field identifying the higher level protocol comes to add just behind the field of supervision. Figure 6.2 illustrates the PPP frame.

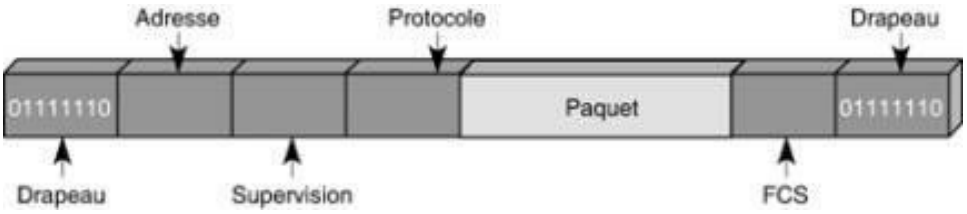


Figure 6.2
Structure of the PPP frame

The values of the most traditional of the protocol field are the following:

- 0x0021: IPv4 protocol;
- 0x002B: protocol IPX (Internetwork Packet eXchange);
- 0x002d: TCP/IP compressed header;
- 0X800F: IPv6 protocol.

The features of PPP are very similar to those of the HDLC protocol (see Annex D).

The ATM protocol

The idea of achieving a network extremely powerful with an architecture of frame level (layer 2), likely to take charge of the multimedia applications, has seen the day toward the middle of the 1980s. From there was born the ATM protocol and its frame, a constant length of 53 bytes, as shown in Figure 6.3.

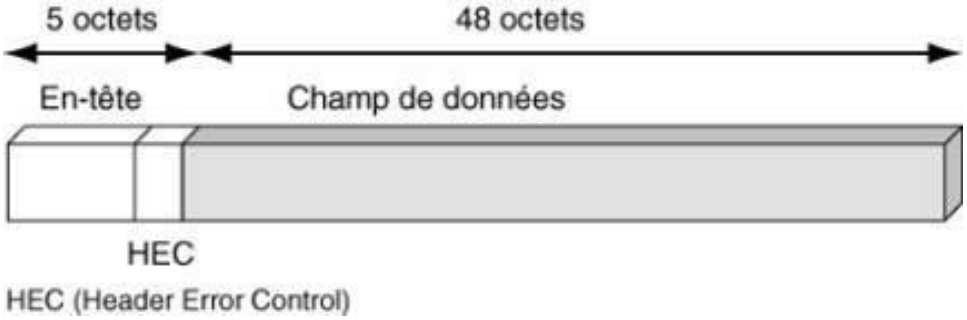


Figure 6.3
Structure of the ATM frame

This constant length of 424 bits allows to discover the beginnings and the purposes of frame to simply count the number of the received bits. In the event of loss of the synchronization frame, it is possible to discover the beginning of a frame ATM in using the area HEC (Header error control), the fifth byte of the header area, which allows in addition to correct a possible error in the header.

The area HEC door a key on 8 bit, which is quite complex to handle. The loss of synchronization results of this fact a important work. The arrival of each element binary, it must in fact perform a polynomial Division to verify if the rest corresponds to the value indicated in the area HEC. In spite of the current increase in flow, it is for this reason difficult to exceed 1 Gbit/s on an ATM link.

The header of the frame ATM includes a reference, which allows you to switch the frames of node in the node. At the ends of the network, it must encapsulate the data user, who come from various applications, ranging from the floor phone in the transfer of data in the data field of 48 bytes. This decomposition of the message in fragments of 48 bytes is done in the AAL layer (ATM Adaptation Layer), so that the message, once cut, either switched quickly on the virtual circuit up to the remote equipment.

The signalling system is the last element of the switched environment. It allows you to open and close the virtual circuit. From extensions of the system of signalling of telephony world, the virtual circuit of the transfer technique ATM is specific to this technology (see Annex G).

The header of the frame ATM

The 5 bytes of supervision of the ATM frame forming the header (header) are illustrated in Figure 6.4. Its features are detailed later in this chapter.

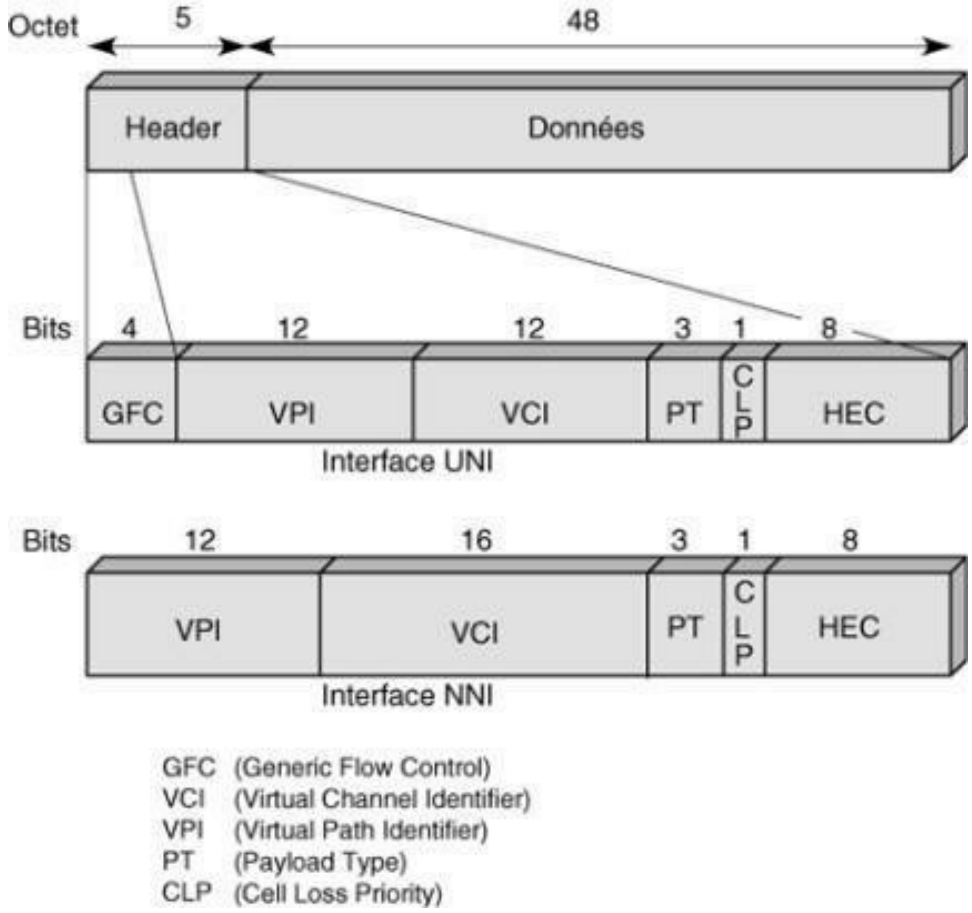


Figure 6.4

Format of the header of the ATM cell

The GFC bit (Generic flow control) are used to access control and flow control on the terminal part, between the user and the network. When multiple users want to enter in the ATM network by a same point of entry, it must order their applications. This control is simultaneously a technique of access, as in local networks, and a flow control on what between in the network. Unfortunately for the world ATM, this area has never been standardized, what constitutes a strong handicap for the user interfaces. In the absence of a standard on the terminal interfaces, it has not been possible to the ATM to compete with the IP interface, which has finished by s impose everywhere.

In the field of control, 3 bits PT (payload type) define the type of information carried in the cell, particularly for the management and control of the network. The eight possibilities for this field are the following:

- 000: cell of user data, no congestion; indication of a user level in the ATM network to another user of the ATM network = 0;
- 001: cell of user data, no congestion; indication of a user level in the ATM network to another user of the ATM network = 1;
- 010: cell of user data, congestion; indication of a user level in the ATM network to another user of the ATM network = 0;
- 011: cell of user data, congestion; indication of a user level in the ATM network to another user of the ATM network = 1;
- 100: cell of management for the flow OAM F5 of segment;
- 101: cell of management for the flow OAM F5 end-to-end;

- 110: cell for the management of resources;
- 111: Reserved for future functions.

Then comes the CLP bit (cell loss priority), which indicates if the cell may be lost (CLP = 1) or if, on the contrary, it is important (CLP = 0). This bit has the function to assist in the control of flow. Before issuing a cell in the network, it is necessary to respect a rate of entry, negotiated at the time of the opening of the virtual circuit. It is always possible to do enter cells in excess, but it must bring an indicator identifying them in relation to the basic data. The operator of the ATM network can lose these data in excess to allow information entries in the framework of the flow control of transit without problem.

The last part of the control zone, the HEC (Header error control), is reserved for the protection of the header. This field allows to detect and correct an error in standard mode. When a header in error is detected and that a correction is not possible, the cell is destroyed. This point is taken up in more detail a little later in this chapter to describe the procedure used and show the use of this field to retrieve the synchronization when the latter is lost.

As explained at the beginning of the chapter, two interfaces have been defined in the ATM: the interface UNI of entry and exit of the network and the interface NNI between two nodes to the interior of the network. The structure of the ATM cell is not exactly the same on the two interfaces. The structure of the ATM cell on the interface UNI is illustrated in Figure 6.5 and the one on the NNI interface in figure 6.6.

The GFC field allows you to control the flow of incoming cells in the network, the multiplexer and decrease the periods of congestion in the network of the end user, called CPN (Customer Premises network). The GFC guarantees the performance required by the user, as the bandwidth allocated or the rate of traffic negotiated. The ITU-T has defined in the recommendation I.361 Two sets of procedures for the GFC, the transmission procedures controlled and those not controlled. For the procedures for transmission not controlled, the code 0000 is placed in the field GFC. In this case, the GFC plays no role.

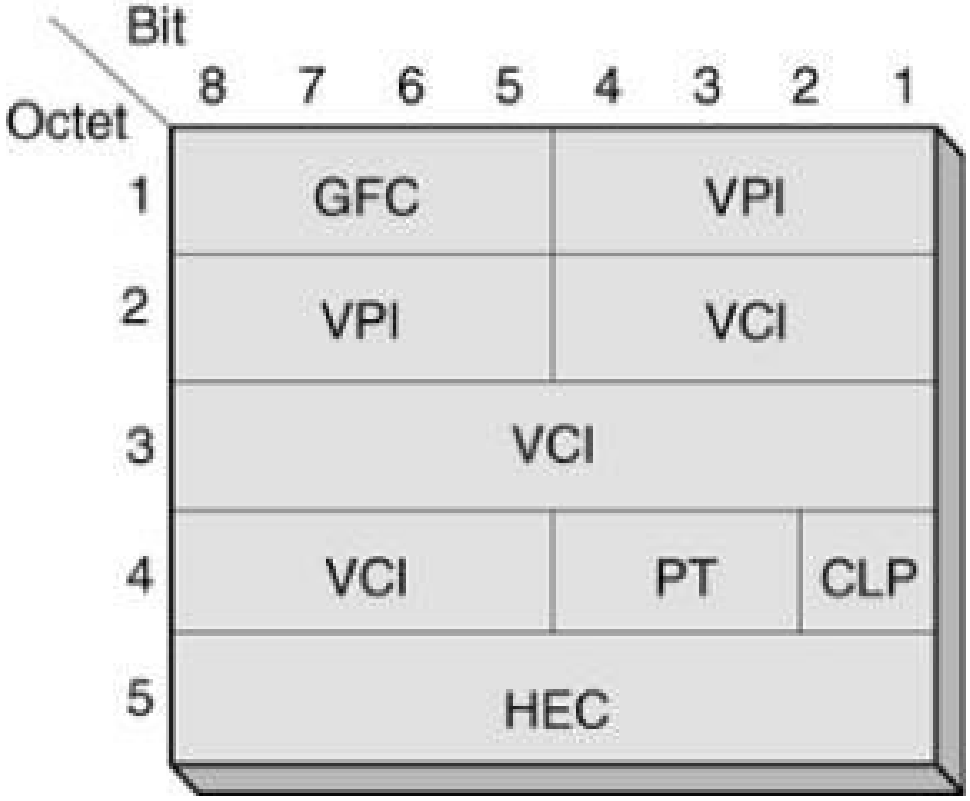


Figure 6.5
Structure of the ATM cell on the interface UNI

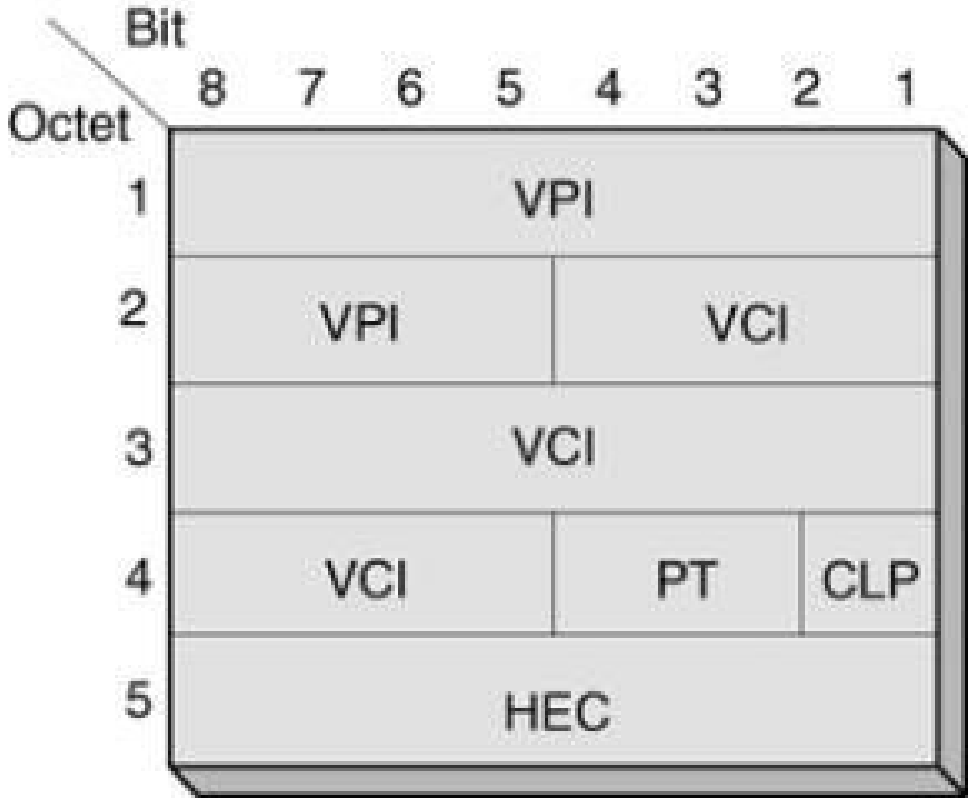


Figure 6.6

Structure of the ATM cell on the NNI interface

In summary, the two main functions carried out by the GFC are:

- The flow control in the short term;
- The control of the quality of service in the network of the final user.

The GFC field only exists on the UNI. The four bits of this field GFC are replaced on the inside of the network on the NNI interfaces by four other bits, who come lengthen the reference. When a user positions the four GFC bit on its interface, these four bits are cleared in the network to be replaced by a complement of the reference number and do not arrive therefore never to the recipient. In other words, these four bits cannot be used to a transmission end-to-end information, but only locally on the input interface in the network.

The Ethernet Frame

The Ethernet frame has been designed for the transport of packets in enterprise networks by means of an original method of dissemination on a local network. This solution has given birth to Ethernet networks shared, in which the frame is emitted in dissemination and where only the station that recognizes itself has the right to copy the information. To this solution of dissemination has added the Ethernet switching.

Before to look on the various types of Ethernet switching, indicate that Ethernet uses well a frame since the Ethernet block is preceded by a succession of 8 bytes starting by 10101010101010101, and so on until the end of the eighth byte, which ends by 11. This preamble is long enough to ensure that it is not possible to find the same succession between two preambles, the likelihood of finding this succession being 1/264.

The structure of the Ethernet frame has been standardized by the IEEE (Institute of Electrical and Electronics Engineers), after having been defined at the origin by the triumvirate of industrialists Xerox, digital and Intel. Two Ethernet frames therefore coexist, the primitive version of the Triumvirate Founder and that of the standardization by the IEEE. The format of these two frames is shown in Figure 6.7.

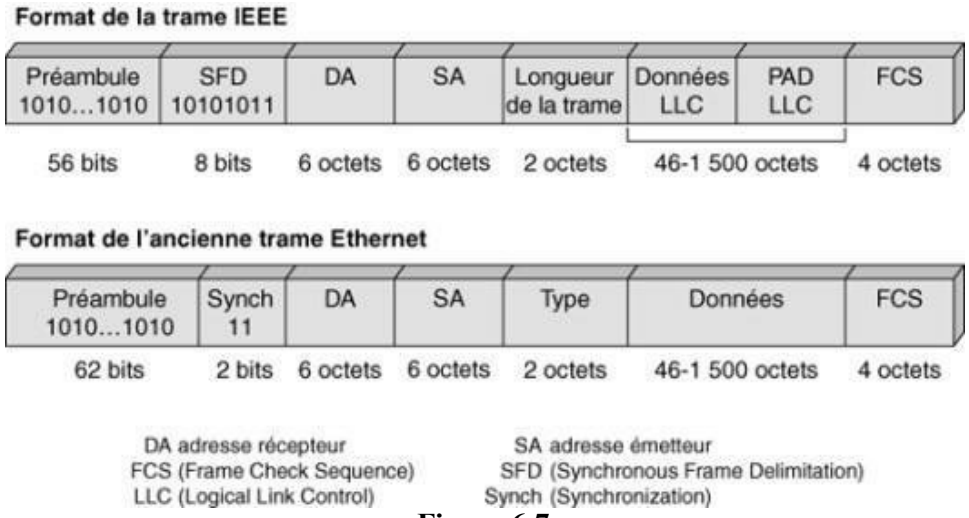


Figure 6.7

Format of the two types of Ethernet frames

In the case of the frame IEEE, the preamble is followed of a zone of beginning of message, called SFD (Start Frame Delimiter), whose value is 10101011. In the old frame, it is followed by 2 bits of synchronization. These two sequences are in fact identical, and only the presentation differs from a frame to another.

The frame contains the address of the transmitter and receiver, each on 6 bytes. These addresses are equipped with a specific form of the world Ethernet, designed so that there are not two couplers in the world who have the same address. In this addressing, said flat, the first three bytes correspond to a number A number of manufacturer, and the three following to a serial number. In the first three bytes, the two initial bits have a special meaning. Positioned at 1, the first bit indicates a group address. If the second bit is also value to 1, this indicates that the address does not follow the standard structure.

Let us look in a first time the result of the IEEE frame. The Area length (length) indicates the length of the data field from the upper layer. The frame then encapsulates the block of frame level itself, or Frame LLC (Logical Link Control). This encapsulated frame contains a zone pad, which allows you to fill in the data field of a manner to achieve the value of 46 bytes, which is the minimum length that must reach this area for the total frame make 64 bytes by including areas of preamble and delimitation.

The old Ethernet frame has in addition a type, which indicates how to present the data area (data). For example, if the value of this area is 0800 in hexadecimal, this means that the Ethernet frame carries a IP packet.

The detection of errors is ensured by means of a polynomial generator *g(x)* according to the formula:

$$G(x) = x32 + x26 + x22 + x16 + x12 + x11 + x10 + x8 + x7 + x5 + x4 + x2 + x1$$

This polynomial gives birth to a sequence of control (CRC) on 4 bytes.

To connect to the Ethernet network, a machine uses a coupler, i.e. a card that is inserted into the machine and which supports the software and the network hardware necessary for the connection.

As indicated previously, the Ethernet frame contains a preamble. The latter allows the receiver to synchronize its clock and its various physical circuits with the transmitter, so as to receive correctly the frame. In the case of an Ethernet network shared, all couplers of the network save the frame as and to the extent of its passage. The electronic component responsible for the extraction of the data included in the frame verifies the concordance between the address of the destination station scope in the frame and the address of the coupler. If there is a match, the packet is transferred to the user after verification of the conformity of the frame through the sequence of control.

The Ethernet switching

Several kinds of Ethernet switches have succeeded. The first, which is very widespread in the businesses, is to use the MAC address on 6 bytes as a single reference on all the way which leads to the card Ethernet coupler. All frames that use the same reference go to the same place, i.e. to the card coupler that has the MAC address specified. This means that all the switches in the network must have a switching table, called " *Lookup Table* ", with as many lines as Ethernet cards to achieve.
The updates to this table are complex, because we need to add or remove lines on the whole of the switches in the network for all Ethernet cards which are activated or deactivated. The recognition of addresses is by learning. When a frame enters a node and that this node does not have the source address of the frame in its switching table, the node adds a line to its table, indicating the new reference and the direction of where comes the frame.
When a frame arriving in a node does not find in the switching table the address of the recipient, there are several possible solutions. A first solution is to issue the frame in dissemination of so that the receiver will eventually receive it. Another solution is to send a frame of signage in dissemination requesting the address of the recipient. The latter is made known when it receives the frame.
Despite these solutions automatic learning, the Ethernet switching has been able to develop on the very large networks. In effect, the management of the switching tables was fast becoming too restrictive. There is the example of a switched network without genuine signalling system. This solution is strongly developed in the enterprise networks, in structuring the network in several sub-networks so as to avoid the flooding of frames when the switching table is incomplete.
The second solution of switching has been made by the SPLM. It is presented in detail in Chapter 11. It is to introduce into the Ethernet frame a specific reference, the shim-label, or MPLS shim, in a new area added to the Ethernet frame behind the MAC address.
The Ethernet frame is switched in a classical way by using the input line and the reference. For this, it must put in place the references all along the path. The Frames occur in remaining in the order of issue. That is why a explicit signaling is essential in this type of network. MPLS has adopted the IP network as a signalling system. This solution has been extended by the technique called *Label Switching, detailed a little further.*
A third switching solution comes from the VLAN techniques (Virtual LAN), explained in Chapter 22. A VLAN is a grouping of machines geographically dispersed to enable them to communicate as if they were in the same local network. A specific field in the Ethernet frame has been added to bear a reference of VLAN. This area allows you to define the 4 096 VLAN, a value very insufficient for the large networks of operators.
From this solution to basis of VLAN, operators have proposed many improvements, to begin by an extension of the area of the numbering of the VLAN so to achieve several millions of possibilities. A VLAN to two machines determines a path which is defined by the result of references corresponding to the number of vlans.
Since 2012, another solution expands to the transport intra- and inter-datacenters, as explained at the beginning of the chapter with the RBridge. They communicate between them in the dissemination by the intermediary of a routing protocol of Ethernet frames, therefore of Level 2 (IS-IS in the case of the Protocol TRILL). This protocol is studied in Chapter 13.
The transport of virtual machines, or VM (which is also examined in Chapter 13), between datacenters at level 3 is based on a completely different idea: instead of changing the IP address of the VM (Virtual Machine) moved, we dissociate the IP address classic in two distinct IP addresses, one used to identify the customer, the other to routing. Where the name of the Protocol in question: LISP (Locator/Identify separation Protocol). This protocol is detailed in Chapter 10.

It may be noted to finish the introduction of a new technique for routing or switching of Level 2 in a centralized fashion. The calculations of the routing tables or switching is performed in a centralized machine, called the controller. This controller can be a physical machine or a VM Located in a cloud. The advantage of this solution, which breaks completely with the transfer protocols classics, who are mainly distributed, comes from the computing power available.
This power allows to calculate multiple tables of transfer (forwarding table), even to propose a table by user. It is required that the central machine retrieves a large number of information from the set of users, which would pose many problems in a distributed configuration. The basic method for this type of transfer, called SDN (Software-Defined Networking), décorrèle the control associated with the calculation of the table of transfer and the transfer itself. The calculations can be carried out in a different machine (the controller) of the transfer machine (router or switch).
Most of the manufacturers are launching in the definition of protocols for the SDN, even if the first signaling protocol to have seen the light of day, OpenFlow, remains the most used. This solution is presented in detail in Chapter 12.

The Label Switching

The switching the more widespread among the operators comes from the Technology MultiProtocol Label Switching (MPLS), which is detailed in Chapter 11.
The dialup technologies call for a reference (label) to enable data blocks, that this is of frames, packets or other entities, to move forward in the network. The whole of these techniques is called Today Label Switching, or switching of references. Are part of the Frames ATM and Ethernet, which use a switching to a reference, as well as all the techniques which can manage a reference or to which one can add a reference.
The introduction of references in the Label Switching is illustrated in Figure 6.8.
The reference is located in a field called MPLS shim, shim or label, presented in detail in Chapter 11. This field contains the reference itself as well as a 3-bit field called experimental and intended for OEMS, a bit called stacking, which allows you to stack the references, i.e. to put several MPLS shim between the header frame level (layer 2) and the level package (layer 3), and a last field, said TTL (Time To Live), on 8 bit, which defines the time after which the package will be destroyed.
Other types of references can be introduced, such as the number of the wavelength of an optical fiber in a system to wavelength division multiplexing or number of an optical fiber or a metal cable in a beam of several tens or hundreds of cables. These latest solutions are explained in the section of the Chapter 11 devoted to GMPLS.

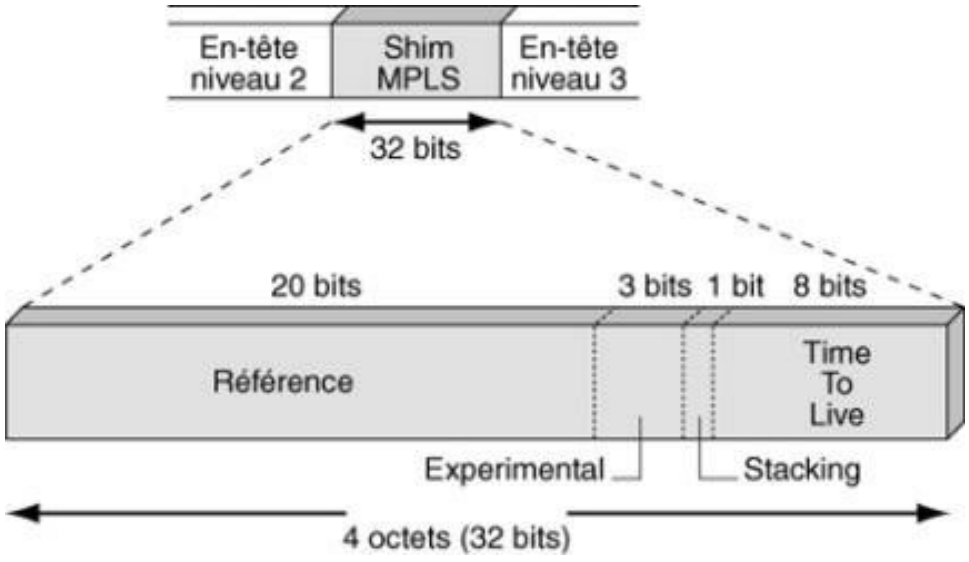


Figure 6.8
Introduction of references in the Label Switching

Conclusion

As the physical level, the level frame is essential for the transport of the information in a network of transfer. Fifteen years ago, this level frame was only an intermediary to the level package and was essentially dealing to detect errors and to request a retransmission in case of error. All the work of routing and switching is performed at the top level, the level package.

Today, the frame level of network architectures is no longer of error detection, since there is more than a negligible number and that the multimedia applications to are largely enough. On the other hand, the Processed features formerly in the Layer 3 have been lowered in the Layer 2. The architectures of frame level have become the standard in the networks of operators and providers of services.

7

The levels packet and message

The role of the level packet is to transport from one end to the other end of the network of blocks of data from a fragmentation of the messages of the higher level, the transport level.
The packet is the entity of the layer 3 which has the address of the consignee or the necessary reference to its routing in the network. The level package is also responsible for the flow control, which, if it is well designed, avoids congestion in the network nodes. As indicated previously, the features of the packet level can be found at the level frame.
It is to be noted that a packet cannot be forwarded directly on a physical media, because the receiver would be unable to recognize the beginnings and ends of package. This is the reason for which the packet is encapsulated in a frame.
The whole of the packages ranging from a same transmitter to the same consignee is called a stream. The latter may be long or short, depending on the nature of the service which has created. If a large file gives birth to a stream important, a transaction produces only a stream very short, one to a few packets. The level packet may make call or not to a connection (see chapter 2) to negotiate a high quality of service with the recipient of the stream.
Before addressing in detail the features of level package, the following sections are reminiscent of the characteristics of this level, defined in the framework of the reference model. The role of the layer 3 (Network) is to transport the packets from one end to the other of the network.

The packet level

The level package, or network layer, is located at the third level of the hierarchy of the architecture of the reference model, as shown in Figure 7.1.

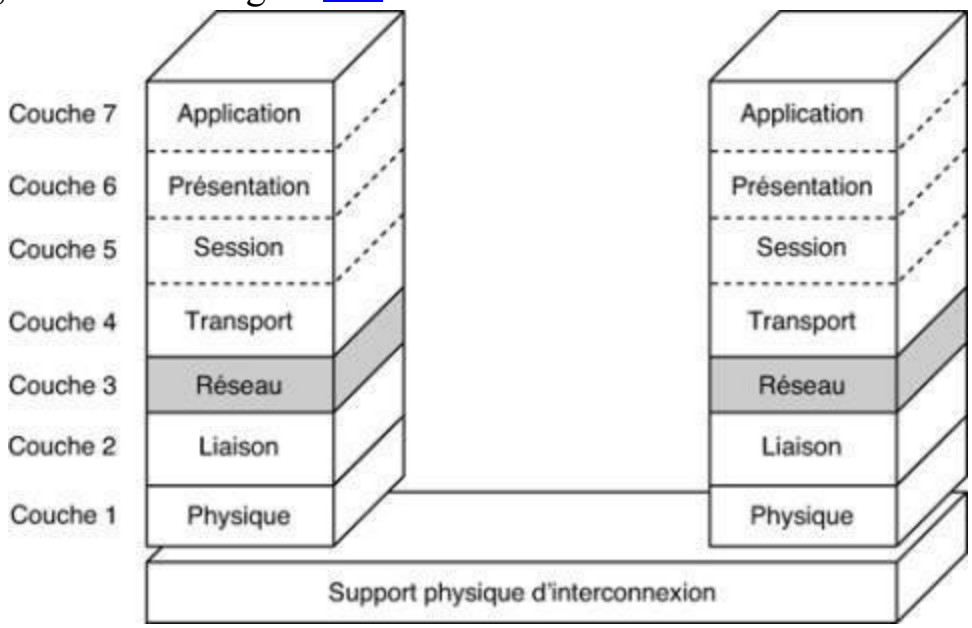


Figure 7.1

The level package, or network layer

This layer exists when the network uses a level package. The functions defined by the international standardisation in the framework of the reference model are the following:

- Establishment of network connections to route packets. This function is however not an obligation in the level package, and the IP protocol works in mode without connection. In contrast, the technique X.25, which has disappeared, used a mode with connection (see Annex E).
- Support for the routing of packets and the problems of gateways to reach another network.
- Multiplexing of network connections.
- Support for the segmentation and the grouping.
- Detection of packet loss and recovery of lost packets. This function is not mandatory at the level package. IP, for example, packet level protocol, has no means to recover the lost packets.
- Maintenance in sequence of data submitted to the upper layer, without that it is an obligation.
- Flow control, so that there is no overflow of the briefs in charge of the transmission of packets.
- Transfer of Data Express (Non-mandatory).
- Resetting the Network Connection (non-mandatory).
- Quality of Service (non-mandatory).
- Management of the network layer.

The Modes With and Without connection

In the mode with connection, a connection must be established between the two ends before issuing the packets from the user to enable two interconnecting entities to exchange information of control. In the mode without connection, on the contrary, the packets can be sent by the transmitter without consultation with the receiver. We do not know if the recipient entity is ready to receive the packets. It is to be noted that the protocols in mode without connection are much more simple than the protocols in mode with connection.

In mode with connection, the network usually employs a technique of switching. Leaves to send a packet of supervision to ask for the opening of the connection, as much to use the crossing of the network by this package of supervision, also known as the package of appeal, to put in place of references, which will issue to very high flow on the path as well implemented. In mode without connection, the network managers prefer the routing since there is no signage. It should nevertheless be noted that a network with connection can be satisfied with a technique of routing and that a network without a connection can use a switching.

In mode without a connection, a network entity emits a package without having to worry about the state nor of the desires of the receiver. As for the whole of the modes without a connection, the other end must be present, or at least represented. For this, a connection is put in place at a higher level, generally the level session. The mode without connection is much more flexible, since it does not take account of what is happening at the level of the receiver.

The two modes have advantages and disadvantages.

The main advantages of the mode with connection are the following:

- Security of the transmission;
- Sequencing of packets on the connection;
- Easy adjustment of the parameters of the network protocol.

Its main disadvantages are the following:

- Heaviness of the Protocol is implemented, in particular for small packet size;
- Difficulties to reach the stations in multipoint or dissemination, the fact of the need to open as many connections as there are points to achieve;
- Relatively low throughput routed on the connection.

The benefits of the mode without connection are the following:

- Dissemination and issuance in multipoint greatly facilitated;
- Simplicity of the Protocol, allowing the performance high enough.

Its disadvantages are the following:

- Low guarantee of the safety of the transport of packages;
- Adjustment more complex of parameters in order to achieve desired performance.

The main protocols of packet level

The packet level protocol exclusive almost is IP. It is also one of the oldest since the first studies from the Ministry of Defense in the United States and of the Cyclades project in France have started at the end of the 1960s. The IP protocol became stable at the very beginning of the years 1980.
IP is a de facto standard of the Internet Engineering Task Force (IETF). This body, which has no power of law offers protocols, some of which eventually to impose by the number of industrialists who the choose.
A second protocol of the level package has known its hour of glory between the years 1980 and 2000. It has been standardized by the ISO (International Organization for Standardization) and ITU-T, the two standardization bodies of law, since dependent on the member and representative users and industry of telecommunications. This protocol is known by its number of recommendation, X.25.3, or X.25 PLP (Packet Level Protocol), and by its number of standard, ISO 8208.

The major features of the packet level

As explained previously, the role of the packet level is to take in charge the packets and to transport from one end to the other end of the network toward the good point of destination and in the best conditions possible. There are two ways to proceed: put in place a path, or virtual circuit, between the transmitter and the receiver or use the mode without connection. The word path overrides the expression virtual circuit, the IP world did not appreciate much this last expression, which recalls the old technologies of switched telephony. The English expression *path* is heavily used. When the path uses a technique of switching, the devoted expression is *label-switched path, or switched path.*
In the mode the path the packets travel in an orderly way to arrive in the order in which they were issued. To open the path, it is necessary to use a signalling. It must remove as and to the extent of its progress in the nodes of the network the references that will be used by the data packets, as shown in Figure 7.2.

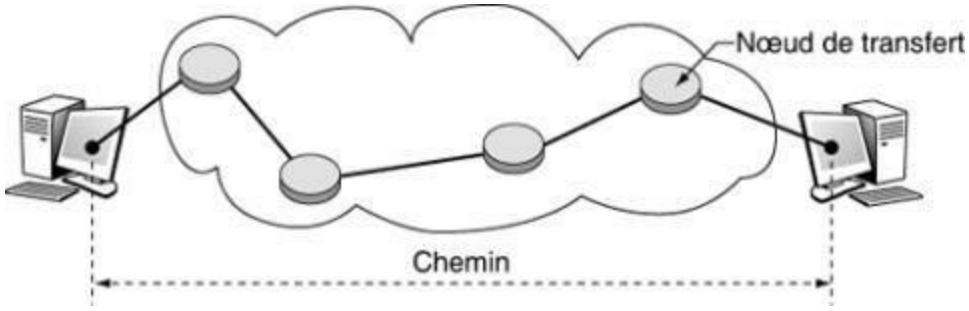


Figure 7.2

Install references in the nodes of a network

In the mode without connection, also said Datagram mode, each packet is considered independent of the other, even if all the packets belong to the same stream. The packets may take different paths and arrive in any order at the receiver, contrary to what occurs in the mode path, where the packets arrive always in the order of issue. The control of the different packages isolated request specific algorithms, which are presented later in the sections devoted to the controls of flow and congestion.

Nothing prevents to achieve a network with connection using an internal Datagram mode for the transport of packets, as shown in Figure 7.3. A package of supervision is routed to the receiver to establish the connection. The opening of the connection can take place without the existence of a path.

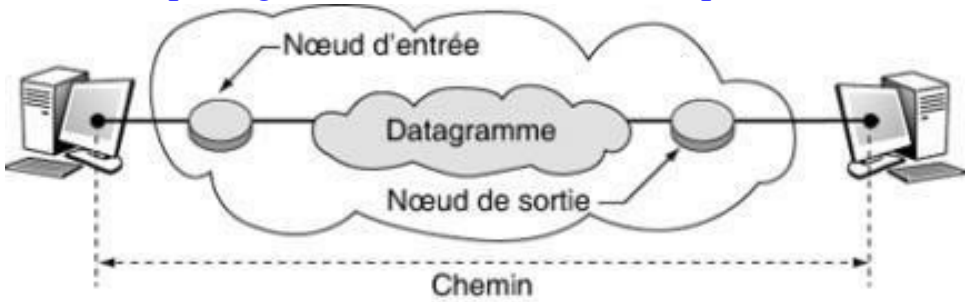


Figure 7.3

A path above of a datagram

To the contrary, nothing is opposed to build a network without a connection using paths, but it does strictly nothing. It is sufficient to send a packet of control, which remove the references without asking for connection to the destination host.

The three main features supported by a Packet level protocol are the flow control, i.e. the means to avoid that the feed does not grow too much in relation to the resources of the network, the management of addresses or references and the algorithms related to routing.

The flow control

Flow control is the first requested functionality at the level package. It is to manage the packages so that they arrive at the receiver in the period of time the more short and, especially, to avoid losses by crushing in the buffers of the intermediate nodes in case of overload. The networks to transfer of packets are as motorways: if there is too many packages, nobody can move forward. The regulation of the flow is however a complex problem.

Very many methods have been tested in specific contexts. In all cases, the check is carried out by a constraint on the number of packets flowing in the network. This limitation is exercised either on the number of packets in transit between one input and one output or on the whole of the network, either on the number of packets that are left to enter inside the network by unit of time. To these controls can be added techniques for allocation of resources in order to avoid any congestion. Some of these checks of flows are detailed below.

The control by Credit

In the Control by credit, there are a number *n* of appropriations that are circulating in the network. For a packet between, it must acquire a credit, which is released once the destination is reached. The total

number of packet circulating in the network is obviously limited to *N*. credits can be unmarked or dedicated. The method isarythmique manages the appropriations totally trivialized. The difficulty is to distribute the appropriations to the right doors of entry in such a way as to offer a maximum flow. This technique is very difficult to control, and its performance have not been proven as optimal.

A first improvement to control by credit has been to define of appropriations dedicated to a node entry in the network. A queue of appropriations, associated with the entry node, allows packets to enter the network. Once the packet arrived at the destination node, the credit used is released and redirected, with the acquittal for example, toward the transmitter. Again, the control is quite delicate since it is done only locally and not to the interior of the network.

Most often uses of appropriations dedicated to a user, or at least a path. This method is known under the name of flow control by credit. Figure 7.4 gives an illustration.

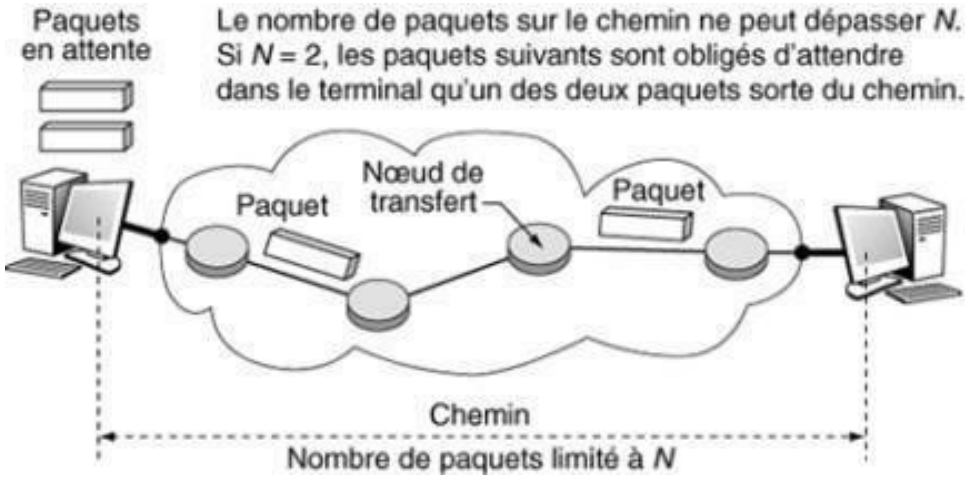


Figure 7.4

The flow control by Credit

The World IP uses this type of control. Each connection is controlled by a window size of variable, which adjusts to any time to avoid that the network is too overloaded. This solution is clever for a network such as the Internet, where the control is carried out by the hundreds of millions of Terminal Machines. The window by a value of 1 and increases exponentially: 2, 4, 8, 16, 32, etc. As soon as the return time of acquittals increases, the PC believes that the flow generated is too important, and it restarts with a window of 1. This flow control is detailed in chapters 10 and 23.

The control by threshold

Another great policy of flow control is to use thresholds of entry into the network. A switch located at the entrance of the network opens more or less to allow more or less of packets, following indications that are provided by the manager of the network.

Several implemented control of threshold can be performed, including the following:

- Packets of the Management bring to input nodes of the network the information necessary to position the switches to the correct value. This method, which is one of the ones that give the best results, has the disadvantage that the network risk of collapse if the control is not carried out quickly enough, to the suite, for example, a failure of a link or a node. In effect, control packets are shipped at approximately the same speed as the other and can request a time too long during a effective congestion of a point in the network.
- The entry of the network is controlled by a window. In this case, packets must be paid locally to allow the window to open to new. In the event of a problem, the network manager can not send the acquittals, which has the effect of blocking the emissions by closing the switch (see figure 7.5).

If the maximum window is equal to *n, between the transmitter and the entry node of the network, there can be more than n* packets.

If the manager of the network no longer sends of acquittal, the window may become equal to 0, which blocks the communication.

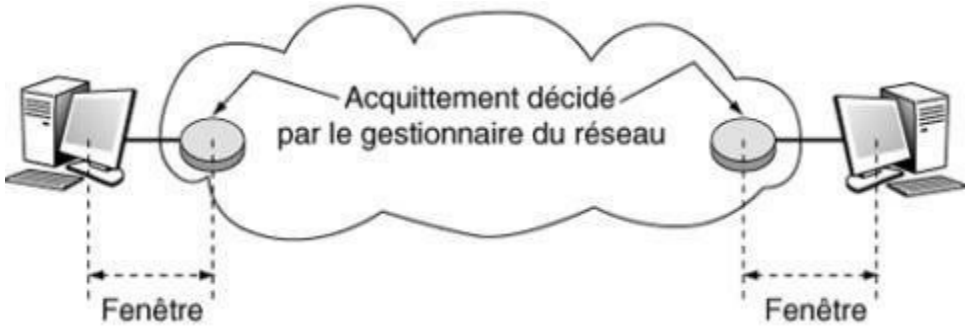


Figure 7.5

The flow control by threshold

The control by allocation of resources

The policies of allowance or preallocation of resources are a third large category of flow control. These policies are essentially adapted to the Dial-up mode with connection, in which a packet of appeal is necessary to the establishment of references and the connection. This package reserve of intermediate resources in the different nodes traversed by the path.

The algorithm for allocation of resources takes forms very different depending on the network. It may in particular superimpose a end-to-end flow control on a path and a method of preallocation. For example, if *n* is the number of credits dedicated to the connection and that the package of appeal reserve exactly the place of *N* packets in its buffers, flow control is perfect, and no packet is lost.

Unfortunately, this control is extremely expensive to put in place, because it is necessary to have a quantity of resources far superior to that which exists in implementations carried out. As shown in figure 7.6, the total number of submissions reserved in the network is worth n × *M, M* is the number of nodes traversed. In addition, on a path, the probability that there are indeed *N* Packet in transit is very low, and this for many reasons: return of acquittals, user inactive or active little, the establishment of the connection, etc.

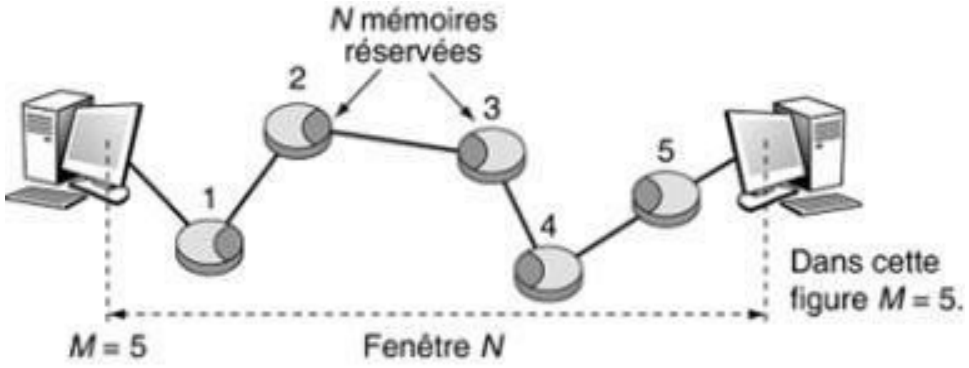


Figure 7.6

The flow control by allocation of resources

To minimize the cost of implementation of such a solution, it is possible to perform a oversubscription. The oversubscription is not to give to a package of appeal which between In a switching node that a part of this that it request. It is hoped that, statistically, if there is more packages than expected on a connection, there will be less on another. Either *k, 0* <*K* ≤1, the factor of oversubscription. If *N* is always the window of end-to-end control, the intermediate node that owns a factor of oversubscription of *K* reserve *kN the* buffers. The value of *k* depends in large part to the occupancy rate of the paths in the network. The traditional values are very low, the rate of use of a virtual circuit often being less than 10%, and factors of oversubscription of 0.2 are quite common.

The oversubscription allows, at a cost low enough, to significantly increase the number of paths that can go by a node. If all buffers are allocated, the package of appeal is denied. It then increases by a

factor of 1/k the number of paths open and, of this fact, the overall throughput of the network. It is obvious that there is a risk of malfunction if, for any reason, the rate of use paths comes to increase. The risk grows again if the average number of packets in the Paths exceeds the limit of oversubscription.

We can draw the curve classic of oversubscription in function of the rate of use of paths for a number *M* of nodes to cross and a number *K* of memory available, so that the probability of packet loss remains to a value ε = 10- 7 (see figure 7.7).

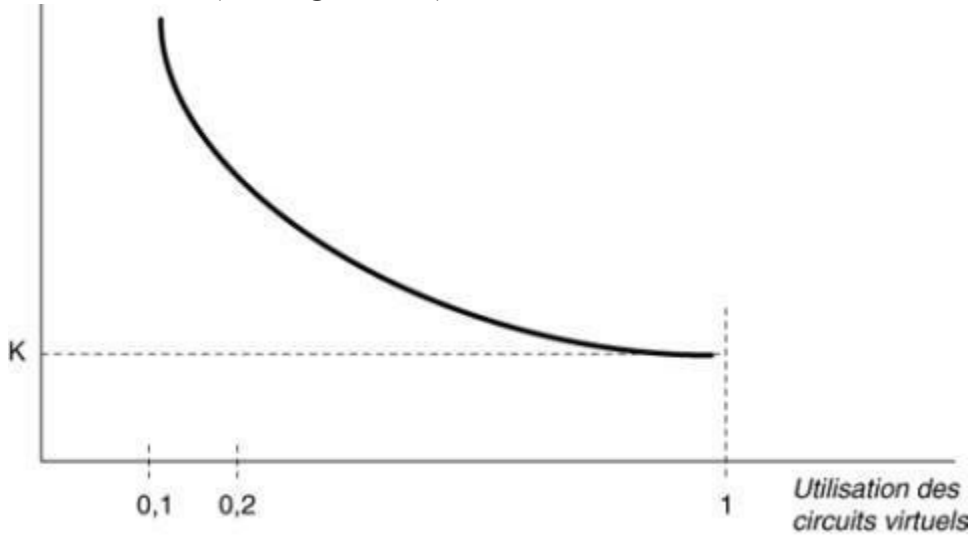


Figure 7.7

Oversubscription of the buffers

One sees that in the vicinity of a utilization rate of 0.1 or 0.2, the oversubscription possible change dramatically. It is therefore necessary to check that the rate of use does not vary too much. To this effect, it is preferable to release of paths rather than lose packets in an uncontrolled manner.

Another possibility to control the flow in a network, always for the method by allocation of resources, is to allocate portions of the bandwidth to each packet of appeal. For a *K coefficient'* of oversubscription, if the flow of a link is *Di,* the path reserves a flow of *k'Di*. Once the entire flow *of* available assigned, the node can no longer accept new packages of appeal, and therefore new openings of paths.

With a few exceptions, these flow control techniques have the default, at very high cost, not work correctly in some cases of the figure, where it occurs a congestion which he must exit. The methods of congestion control presented below allow to cope with the problems in the network, even if No is really effective.

The Congestion Control

The Congestion Control refers to the means implemented to get out of a state of congestion. The controls of flows are here to avoid to enter in the states of congestion, but it is obvious that, despite the efforts to control the flow, of states of congestion remain possible.

A method of congestion control quite used is to keep in reserve in the switching nodes of the place memory not taken into account in the allowances. When the buffers are all filled, it opens the extra space. For little convincing evidence that it may seem, this method has an interest. When two packets, or two frames at the level liaison, must be exchanged on a link, the fact to keep in memory the information pending the acquittal may lead to a blockage, or *deadlock. With an additional place we can resolve the problem (see figure 7.8).*

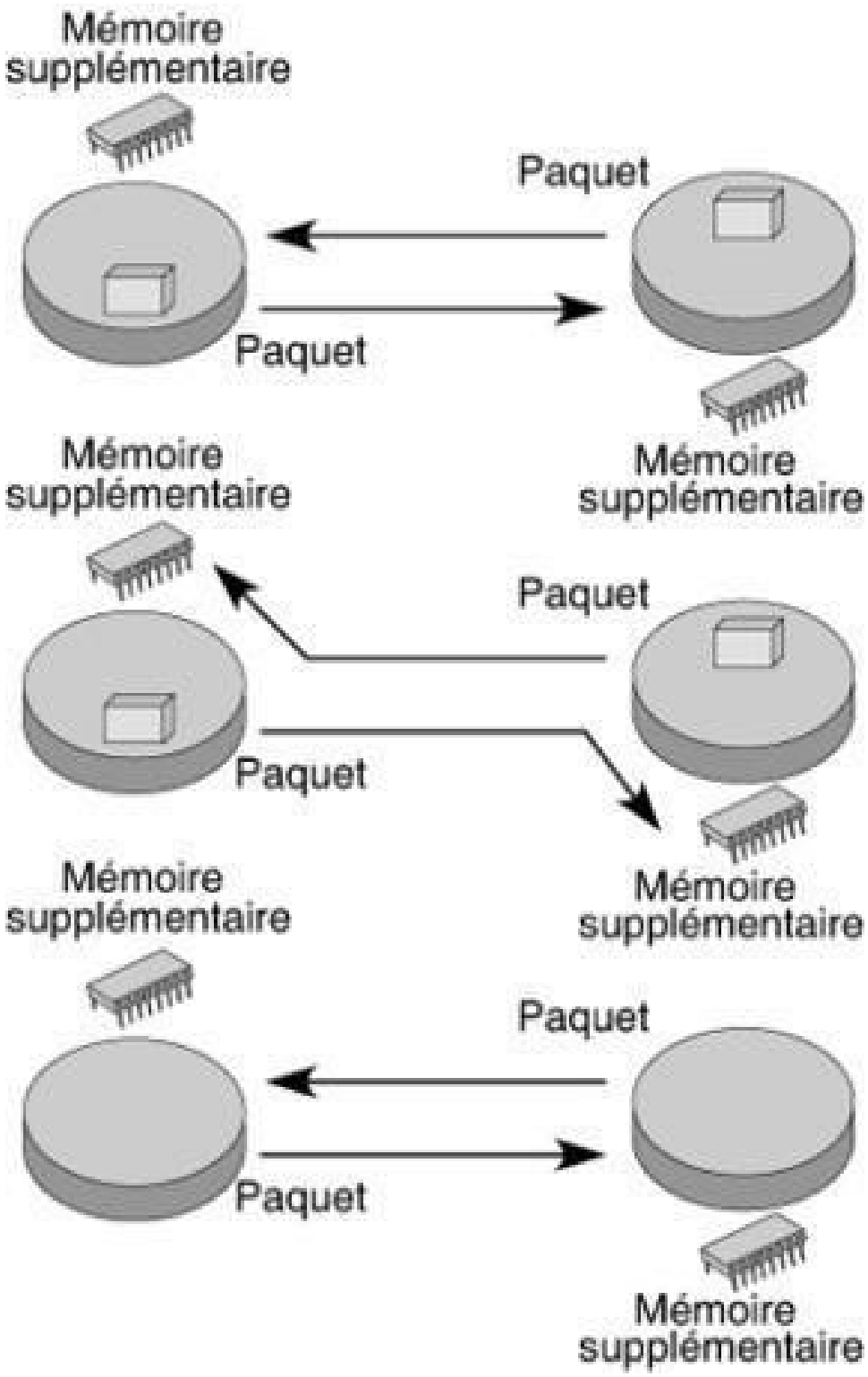


Figure 7.8

Resolution of a blockage by allocation of additional memory

The use of a maximum time of residence in the network is a second possibility of congestion control. We place in the incoming packet the value of a clock common to the whole of the network. This method, called timer, allows you to control the amount of time spent in the network and destroy blocked packets in a node. It helps in addition to delete the packets lost as a result of a false address or a routing error. It is however quite difficult to implement, since it requires a common clock and comparators of time. Most of the protocols that the implement, among which the IP, greatly simplify the algorithm to follow: in the area reserved for the maximum time figure a number that is decremented at each crossing of node.

The routing

In a mesh network, the routing of packets is part of a complex algorithmic, by the distribution of

decisions to take, which fall within the scope of both the space and the time. A node should know the state of all of the other nodes before deciding where to send a packet, which is impossible to achieve. In a first time, let us look at the components necessary for the establishment of a routing. It must be first of all a routing table, which is shown in Figure 7.9.

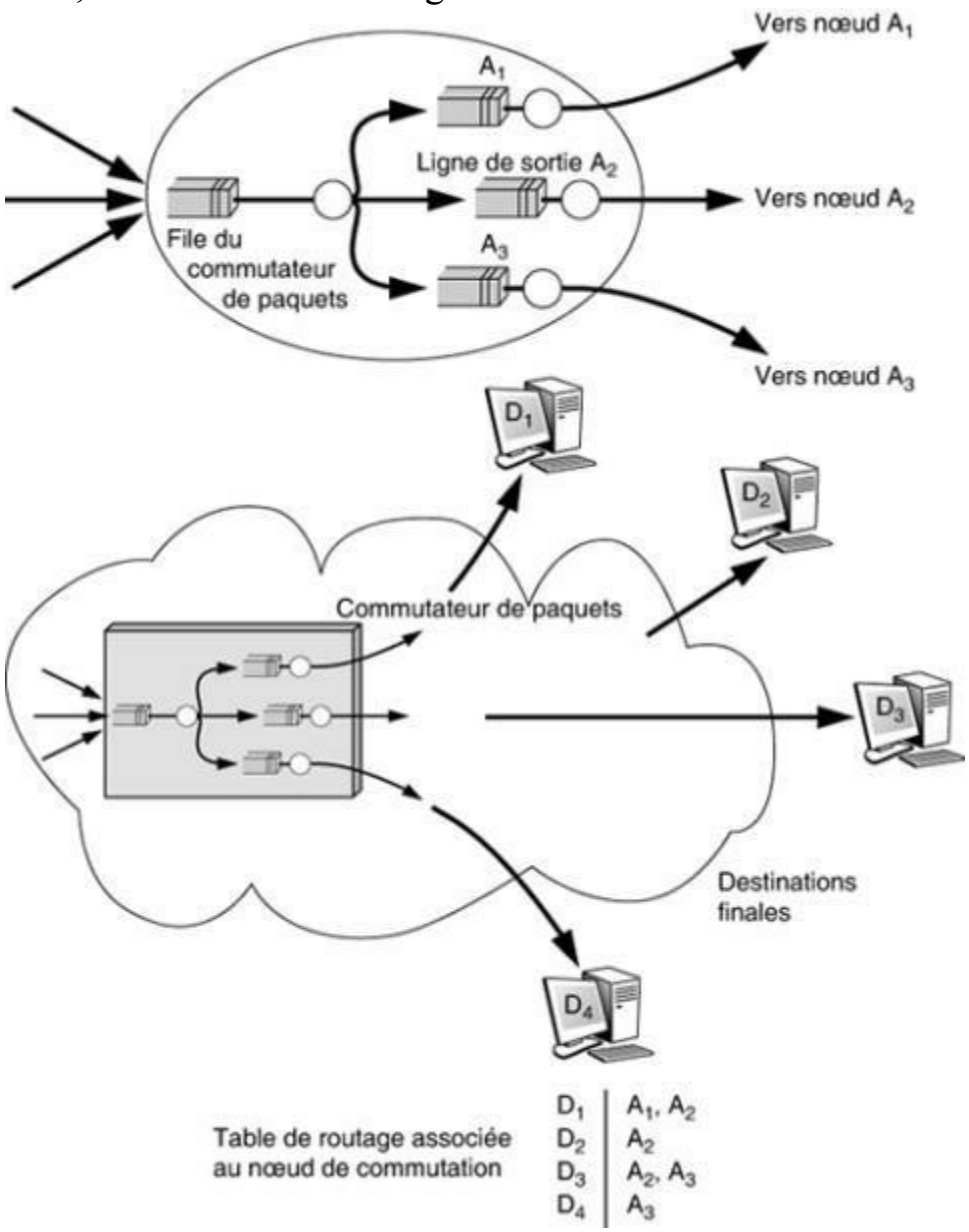


Figure 7.9

Operation of a routing table

One sees that a transfer node is formed of lines of output, which emit of frames obtained from packets. The Packets are routed by the node to a line of output thanks to the routing table. If a packet is present at the node with, for final destination, the node D1, the node may send this packet to the output line A1 or toward the output line A2. The decision is carried out on local criteria in the case considered. For example, it sends the packet on the queue the shorter. If the final destination is D2, the package is placed in the queue A2.

The centralized routing

The centralized routing is characterized by the existence of a center, which takes the decisions with respect to the definition of a new table and the sending of this table to the whole of the transfer nodes of the network. This central node receives the information on the part of all the components of the network, and it designs its routing table after the algorithms determined in advance. This solution of centralized routing has virtually never been used. However, since 2012, it becomes to fashion thanks to the potential of the Cloud. The centralized controller can be a server powerful enough or a VM

(Virtual Machine) in the cloud which can have as much power as needed to calculate transfer tables necessary. There may even be a table by Client in order to optimize the roads or paths for each customer. For this, the controller must receive these customers a large number of information that can be obtained by the controllers to access.

The main considerations to take into account to determine the best routes in a network, either in routing or for the opening of a path, are the following:

- Cost of connections;
- Cost of the passage in a node;
- Requested flow;
- Transit time requested;
- Number of nodes to cross;
- Safety of the transport of certain classes of packages;
- Energy cost;
- Occupation of the briefs of the switching nodes;
- Occupation of the couplers in line.

Routing algorithms use most of the time the criteria of cost. There is, for example, the algorithm of the lowest cost, which, as its name indicates, is to find the path that optimizes the price. The most simple of algorithms, and often the more efficient, gives a cost of 1 for each passage in a node. This is the algorithm of the shortest route. Contrary to what one might think, it is a good way to proceed. You can easily add through to take into account the occupation of intermediate memories, the use of the output lines, etc.

The fixed routing is another technique particularly simple since the table does not vary in time. Each time a packet between in a node, it is sent in the same direction, which corresponds, in almost all cases, to the algorithm of the shortest route. However it cannot be talk of routing algorithm In this case, since the routing is fixed and does not require the update. The fixed routing goes hand in hand with a control center, which manages the serious faults and generates a new table when a node fails or that a line of communication is broken. This is called fixed routing between the updates.

It can improve the routing fixed taking into account of events indicated by the network, such congestion or occupations of lines or briefs too full. All ten seconds, all the nodes in the network send a packet of control indicating their situation. From these accounts rendered, the central node is developing a new routing table, which is disseminated.

The sending of routing tables of a asynchronously is a technique that more developed. The central node diffuses to all of the nodes a new routing table as soon as this table has changed sufficiently compared to that in force. In other words, the Control Center provides the routing tables as the arrival of new information and then sends to all nodes the first routing table which seemed to him to be sufficiently different from the previous one. The adaptation here is asynchronous and not synchronous, as previously.

The performance of this centralized routing depend on the architecture and the topology of the network. Indeed, the main problem of routing and adaptation is that they must perform in real time. Between the time when a node sends a report involving a new routing and the one where the new routing table arrives, there should be no substantial change in the status of the system. This condition is very poorly done if the network is important and the arteries overloaded, control packets being low priority compared to packets carrying information.

The quality of routing corresponds to first view to an adaptation of more and more sophisticated. It is here that arises the second major problem concerning the performance, moreover linked to the first:

the sophistication causes overloading of the network by control packets, which can prevent a operation in real time.
One sees that a routing algorithm given has not the same efficiency for a network to three nodes, for example, that for a network to twenty knots. The first conclusion that it is possible to draw is that it does not exist of algorithm better than another, even for a network well determined, since everything depends on the traffic. In addition, it seems that there is an optimum in the complexity of the algorithm of adaptation for Do not overload the network unnecessarily.
The distributed routing
The most simple techniques for distributed routing, the flood, is not adaptive. When a packet is received in a node, it is retransmitted to all destinations possible. This effective routing is however penalizing in terms of flow and may be adopted only in specific cases, such as the networks in which the real time is essential and the low traffic.
In the algorithms a little more complex, the Adaptability begins to appear. It concerns only a dimension, the time. For a package in transit in the node I and is directing toward the node I, several lines of output can be selected. In the routing method called *hot-potato, we are trying to get rid of the package as quickly as possible in the transmitting on the first line of empty output. In reality, we never used a method hot-potato pure. We prefer more elaborate techniques, in which the coefficients are assigned to different lines of output for a given destination (see Figure 7.10).*

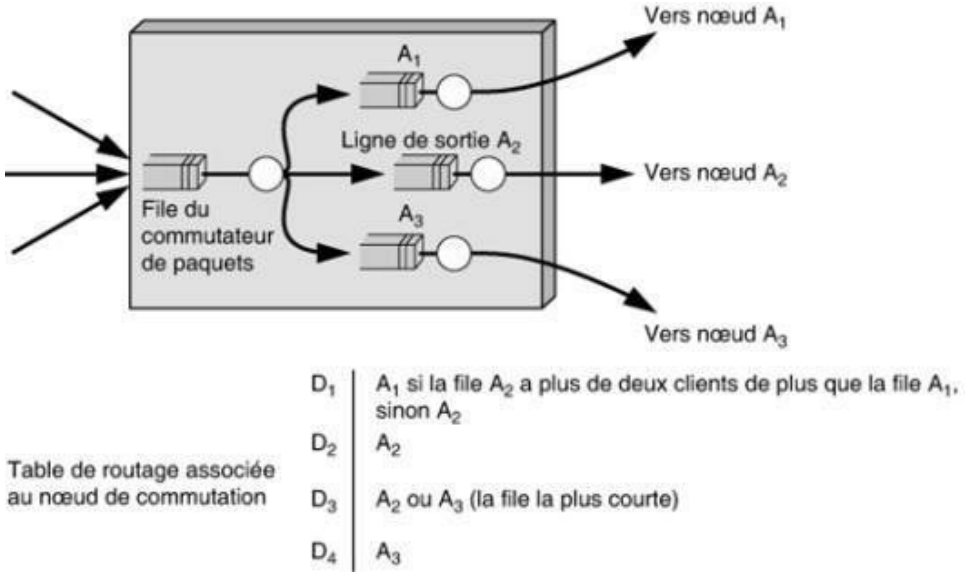


Figure 7.10
Hot routing-potato with through

Example of distributed routing
Either the node shown in figure 7.11, with three neighbors: N1, N2, N3. The processor of this switch is able to know the response time of its three queues of output, noted respectively W1, W2 and W3. This response time is obtained by counting the number of bytes waiting in the Coupler Lines and multiplying by the speed of transmission on the physical media.

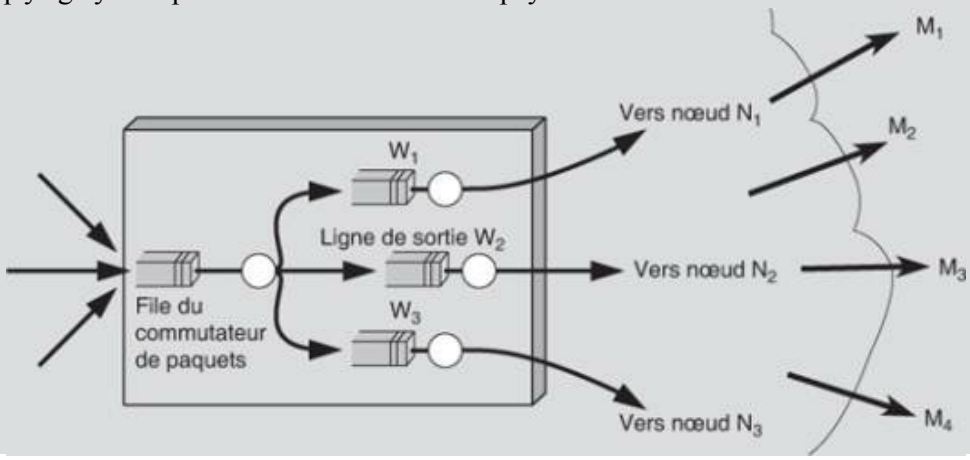


Figure 7.11

Distributed routing in a switching node

Let us take a look at four possible recipients: M1, M2, M3 and M4, and suppose that the Nodes N1, N2 and N3 are able to know the time of optimal routing of an incoming packet in their switch to achieve a device node. These deadlines are identified in Table 7.1.

	M1	M2	M3	M4
N1	160	260	180	218
N2	140	255	175	200
N3	100	247	140	220

Table 7.1 • Time to route a packet

The intersection of the column M3 and line n2 indicates the time of optimal routing to the destination M3 for an incoming packet in the node N2. Let us take the values shown in Table 7.2 as response time of files node's output studied.

W1	W2	W3
26	40	64

Table 7.2 • Response time of the queues of exit

It is possible to calculate, for each destination, delivery times from the node studied in passing through each of the neighbors of this node. These deadlines are listed in Table 7.3.

	To M1	To M2	To M3	To M4
By N1	160 + 26	260 + 26	180 + 26	218 + 26
By N2	140 + 40	255 + 40	175 + 40	200 + 40
By N3	100 + 64	247 + 64	140 + 64	220 + 64

Table 7.3 • total delay for routing a packet

We infer from the routing table of the node studied (see Table 7.4).

	M1	M2	M3	M1
Final destination	N3	N1	N3	N2
Output queue	140	255	175	200
The propagation time of	100	247	140	220

Table 7.4 • Routing Table deducted

The hypothesis of a node is capable of knowing the time of delivery to a recipient has no more place to be since the knowledge of these deadlines for each node allows to obtain not to not. To do this, each node must send to all its neighbors its table of deadlines. The time of propagation for these various transmissions are added. Of this fact, the information in the possession of a node may not be up to date. The problems of overload due to control packets and the accuracy of the information are the two generators of disorder in the routing policies distributed who want to take account of the whole of the resources of transport.

One of the crucial problems posed by the routing distributed, that it is adaptive or not, is that of the loopback, a package that can iron several times by the same node. A characteristic example is provided by the Network ARPANET, the first version of the Internet network, which has experienced the last algorithm described and for which the measurements showed a large number of rebouclages.

The Addressing

The data located in the user or on servers in a network can be achieved only through a specifying addressing the exit interface or by a reference to route the packet up to the interface sought. In this last case, it is necessary to use the address of the recipient for the signaling packet opens a path. The sections that follow are not interested in a first time as to the addressing and then return on systems using references.

The addressing can be physical or logical. A physical address corresponds to a physical connection point to which is connected to a terminal equipment. A logical address corresponds to a user, a

device or a user program that can move geographically. The telephone network offers a prime example of physical addressing: to a number matches a user, or more exactly a junction. In this network, the addressing is hierarchical. It uses a different code for the country, the region and the switch, the last four digits indicating the subscriber. If the Subscriber moves, it changes the number. The temporal switches can confuse the call to another number at the request of the subscriber, but the addressing is not retained.
A second example is proposed by the Ethernet network and more generally by the local networks. It is a addressing of frame level and non-PACKET level. It is introduced in this chapter as an example, because it could very well be implemented in a package.
By the intermediary of the IEEE (Institute of Electrical and Electronics Engineers), to each coupler is assigned a unique number (see Figure 7.12). There is therefore no two couplers with the same address. If the party bearing the address cannot be moved, addressing is physical. On the other hand, if the user can leave with its device and its interface and reconnect elsewhere, addressing becomes logical. In this last case, the routing in large networks is particularly complex.
In the case of the Ethernet network, addressing is absolute. So there is no relationship between the addresses located on sites close to one another. As indicated in figure 7.12, the first bit of the Ethernet addressing specifies if the address corresponds to a single coupler (single address) or if it is shared by other couplers for allow communications in multipoint or dissemination. The second bit indicates if the addressing used is that defined by the IEEE, that is to say if the address fields have the Ethernet address of the coupler or if the user has replaced the two fields by a specific address. It is strongly advised to keep the IEEE address of origin of the coupler to prevent any collision with another IEEE address.

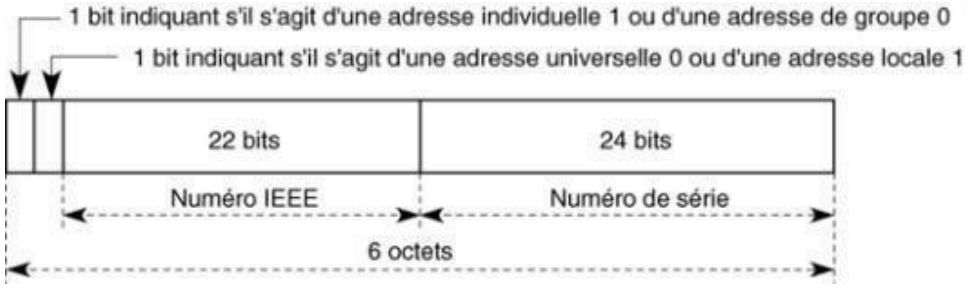


Figure 7.12
The Ethernet addressing

In the Ethernet addressing described above, the geographical situation of the subscriber is impossible to know, and the delicate routing to implement. This is the reason for which the routing of Frame Level has never developed. It would have taken to this develop a hierarchical addressing.
The network Internet gives a good example of hierarchical addressing logic. The address is decomposed into two Parties giving two levels of hierarchy. The first identifies a terminal machine on a network is determined, and the second the number of this network. It can be seen that the address is not necessarily geographic, since a network can contain five PC, one in each continent. It is necessary to find a route to go in the network of belonging and then look at the inside of the network the road going to the recipient. The IP addresses are discussed in detail in Chapter 10.

Other features of the packet level

As explained at the beginning of this chapter, the level package (layer 3), also called network layer, offers a service at the message level, or transport layer. This service must be provided by the network protocol taking into account the service that is offered by the lower layer, as shown in Figure 7.13.

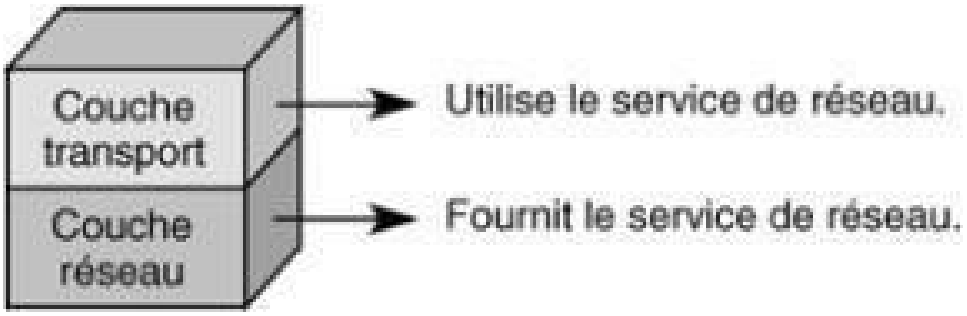


Figure 7.13
Relationship between the network layer and transport

The services that must be rendered by the packet level must meet the following five criteria:

- **Independence in relation to the transmission media underlying it.** The network service frees its users of all of the concerns related to the way in which are used the various sub-networks to ensure the network service. It mask to the user of the Service Network the way in which its packets are transported. In effect, the latter do not generally wishes to not know the protocols used for the transportation of its packets, but just want to be sure that its packets arrive to the recipient with a defined quality.
- **Transfer of end-to-end.** The network service ensures the transfer from one end to the other of the user data. All the routing functions and relays are provided by the supplier of the service network, including in the case where various resources of transmission, similar or different, are used in tandem or in parallel.
- **Transparency of information transferred.** The service network ensures a transparent transfer of information in the form of a suite of bytes of user data or control information. It imposes no restriction as to the content, in the format or encoding information and did not need to interpret their structure or their meaning.
- **Choice of the quality of service.** The service network offers users the possibility to request or accept the quality of service provided for the transfer of user data. The quality of service is specified by QoS settings expressing characteristics such that the flow, the transit time, the accuracy and reliability.
- **Addressing of the user of the service network.** The network service uses an addressing system which enables each of its users to unambiguously identify other users of the network service.

The quality of service

The concept of quality of service, or QoS, is with regard to some of the characteristics of a network connection under the sole responsibility of the service provider network.

A value of QoS applies to the whole of a network connection. It must be identical at both ends of the connection, even if this last is supported by several sub-interconnected networks offering each of the different services.

The QoS is described with the help of parameters. The definition of a parameter of QoS indicates the way to measure or determine its value, indicating the need The events specified by the primitives of the network service.

Two types of QoS parameters have been defined:

- Those whose values are transmitted between users peers through the network service during the establishment phase of the network connection. In the course of

this transmission, a tripartite negotiation can take place between the users and the provider of the service network in order to define a value for these QoS parameters.

- Those whose values are neither forwarded nor negotiated between users and the service provider network. For these QoS parameters, however, it is possible to obtain, through local means, the information relating to the useful values to the supplier and to each of the users of the network service.

The main QoS parameters are the following:

- **Time limit for the establishment of the network connection.** corresponds to the time that elapses between an application for network connection and the confirmation of the connection. This parameter of QoS indicates the maximum time acceptable by the user.
- **Probability of Failure of the establishment of the network connection.** This probability is established on the basis of the requests that have not been met within the normal time limit for the establishment of the connection.
- **Flow rate of the transfer of data.** The flow defines the number of bytes transported on a network connection in a reasonably long (a few minutes, a few hours or a few days). The difficulty in determining the flow of a network connection comes from the asynchrony of the transport of packets. To obtain an acceptable value, it must observe the network on a suite of several packages and consider the number of bytes of data transported by taking into account the time elapsed since the request or the indication of transfer of data.
- **Time to transitlors transfer of data.** The transit time of the corresponds to the time elapsed between a request for data transfer and the indication of the data transfer. This time of transit is difficult to calculate because of the geographical distribution of the ends. The satisfaction of a quality of service on the transit time can furthermore enter in contradiction with a flow control.
- **Rate of residual error.** is calculated from the number of packets that arrive erroneous, lost or double on the total number of transmitted packets. It is therefore an error rate by package. Also refers to the probability that a packet is not correctly to the receiver.
- **Probability of incident of transfer.** is obtained by the report of the number of incident listed on the total number of transfer is carried out. To have a proper estimate of this probability, it is sufficient to consider the number of disconnections of the network by report to the number of transfer is carried out.
- **Probability of Failure of the network connection.** is calculated from the number of liberation and reset of a network connection in relation to the number of transfer is carried out.
- **Deadline for release of the network connection.** C Is The Maximum Acceptable Delay between a disconnect request and the actual release.
- **Probability of failure during the liberation of the network connection.** C is the number of failed release requested by report to the total number of liberation requested.

The three additional parameters The following are used to characterize the quality of service:

- **Protection of the network connection.** determines the probability that the network connection is in a state of walking during the entire period where it is opened by the user. There are several ways to protect a connection in the duplicating or having a backup connection ready to be opened in the event of cut-off. The value for a telephone network is of 99.999%, that is called the five nine, which is equivalent to a few minutes of downtime per year. The protection is much lower for an IP network, with a value of the order of 99.9%, or three nine. This value also presents a problem for IP telephony, which demand greater protection of telephone connections.
- **The priority of the network connection.** determines the priority of access to a network connection, the priority of maintaining a network connection and the priority of the data on the connection.
- **Maximum acceptable cost.** Determines if the network connection is tolerable or not. The definition of the cost is fairly complex since it depends on the use of the necessary resources for the establishment, maintenance and to the release of the network connection.

IP (Internet Protocol)

The basic protocol of the Internet network is called IP, for Internet Protocol. The objective of departure assigned to this Protocol is to interconnect networks not having the same protocols of frame level or level package. The Acronym The Internet Comes of inter-networking and corresponds to a mode of interconnection: Each independent network must carry in its frame or in the data area of its package A package IP, as shown in Figure 7.14.

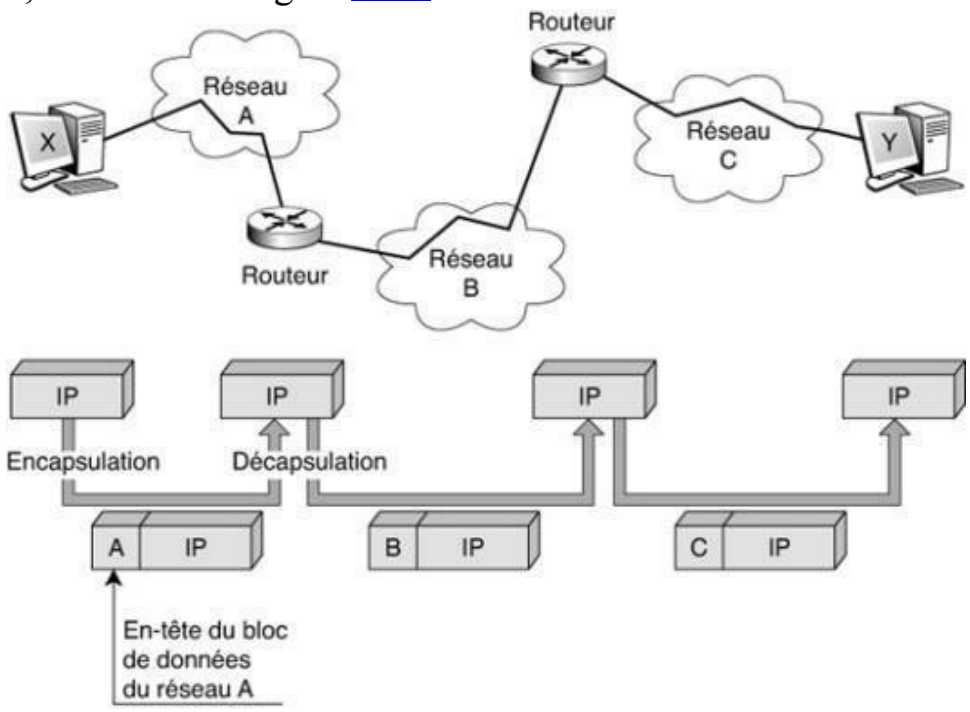


Figure 7.14
Network interconnection

There are two generations of IP packets, called IP version 4 (IPv4) and IPv6 (IP version 6). IPv4 has been role up to now. The transition to IPv6 could accelerate the fact of its adoption in many Asian countries. The transition is however difficult and will last many years.

The IPv4 and IPv6 protocols

The first generation of the IP protocol, IPv4 is implemented in all stations connected to the Internet

network. The fundamental difference with IPv6 resides in a package provided with little functionality, since it is seen as a common syntax for the exchange of information. With the second generation IPv6, which started to be implemented, the change vision is unequivocal, since the IP packet becomes a true package, with all the features necessary for it to be processed and controlled in the nodes of the network.

As indicated on several occasions, IP provides a service without connection. This mode without connection explains expectations long enough during the interrogation of servers very frequented. Even overloaded, the latter may not refuse the arrival of new packages since the transmitter does not request any connection, that is to say is not concerned to know if the server accepts to serve.

The packages of a same stream, hence of a machine and going to another, can use different routes, Internet loading of the routing of IP packets independently of each other. The IP protocol defines the unit of data as well as the format of all the data that travels in the network. It also includes a set of rules that define how to treat the packets, manage the routing function and respond to some of the types of errors.

There is an analogy between the physical network and the logical network in which IP. In a physical network, the unit is transferred is the frame - in reality a packet or frame - the sub-network traversed. This frame contains a header and data, the latter being included in the IP packet. The header contains the information of supervision necessary to route the frame.

In the logical IP network, the base unit to transfer is the IP packet, that is called IP datagram. The datagrams can be of any length. As they must transit from router to router, they can be split, so to adapt to the structure of the frame underlying it. This concept is called the encapsulation. For a sub-network, a datagram is a given as another. In the best of cases, the datagram is content in a single frame, which makes the transmission more efficient.

The sections that follow examine the structure of IPv4 and IPv6. The chapter 10 is devoted to IP networks in general.

IPv4

The service rendered by the IPv4 protocol is based on a system of delivery of packets not reliable, that is called service best-effort, that is to say "at best" and without connection. The service is said to be non-reliable, because the discount presents no warranty. A packet may be lost, duplicated or given out of sequence, without that the Internet does the detects nor shall inform the transmitter or receiver.

Figure 7.15 illustrates the format of the IPv4 packet. After the value 4, for the version number, is indicated the length of the header, which allows you to know the location of the beginning of the data in the IP fragment. The next field, TOS (Type of Service), specifies the type of service of the information transported in the body of the packet. This field has never actually been used before the arrival of the new management protocols relating to the quality of service, as Differentiated Services (DiffServ), which are presented in Chapter 10. Then comes the total length (length). The next field (Identification) identifies the message to which belongs the packet: The message has been broken into packets and it must be able to the receiver to know what message belongs to the packet.

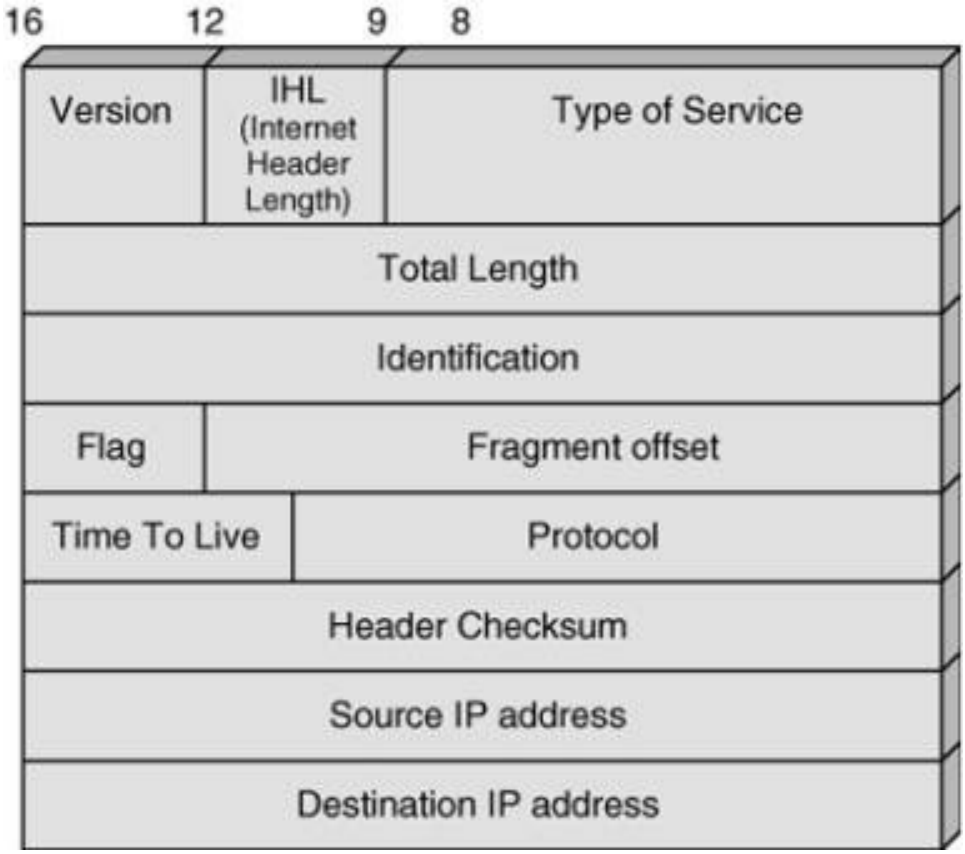


Figure 7.15

Format of the IPv4 packet

The flag (flag) door several notifications. It specifies, in particular, if a segmentation has been performed. If yes, the place of the segment, from the segmentation of the message of level 4, is indicated in the field offset, or location of the segment. The field TTL (Time To Live), or the time of life, specifies the time after which the packet is destroyed. If the package does not find the more its path or conducts round-trips, it is eliminated at the end of a certain time. In reality, this area contains an integer value, indicating the number of nodes that can be crossed before destruction of the packet. The value 16 is used on the Internet to indicate that an IP packet that crosses more than 15 routers is destroyed.

The protocol number indicates the protocol which has been encapsulated inside the package. The area of error detection allows you to determine if the transmission of the packet is performed correctly or not. Finally, the addresses of the transmitter and the receiver are specified in the last part of the header. They take a place of 4 bytes each.

As the Internet is a network of networks, the addressing is particularly important. The machines connected to the Internet have an IPv4 address represented on a 32-bit integer. The address consists of two parts: a network identifier and an identifier of machine for this network. There are four classes of addresses, each for coding a different number of networks and machines:

- Class A, 128 Networks and 16 777 216 hosts (7 bits for the networks and 24 for the hosts);
- Class B, 16,384 networks and 65,535 hosts (14 bits for the networks and 16 for the hosts);
- Class C, 2 097 152 networks and 256 hosts (21 bits for the networks and 8 for the hosts);
- Class D, addresses of groups (28 bits for the hosts belonging to a same group).

These addresses are illustrated in Figure 7.16.

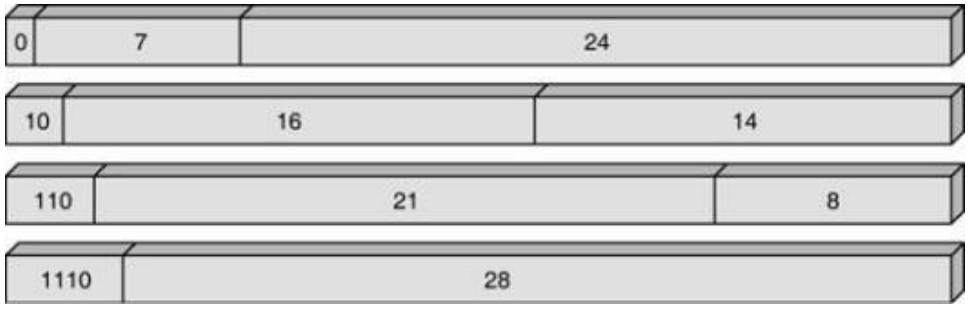


Figure 7.16

Classes of addresses of IPv4

The IP addresses have been defined to be processed quickly. The routers that perform the routing based on the number of network are dependent on this structure. A host connected to multiple networks has multiple IP addresses. In fact, an address does not identify just a machine, but a connection to a network.

IPv6

IPv6, sometimes called IPng (Next Generation), is a protocol completely redesigned, which belongs to the packet level. The format of the IPv6 packet is shown in Figure 7.17.

To improve the performance, IPv6 prohibits the fragmentation and reassembly in the intermediate routers. The protocol must therefore choose the right value of length of the datagram so that it can be directly encapsulate in the different frames, or packets encountered. If, in an environment IPv6, a datagram is present at the input of a sub-network with a size not acceptable, it is destroyed. As explained previously, the level package represented by IP is considered as a logical level of interconnection between sub-networks. This level IP can become a packet level protocol self-sufficient, usable to carry the information on a network. That is exactly the role played by IPv6.

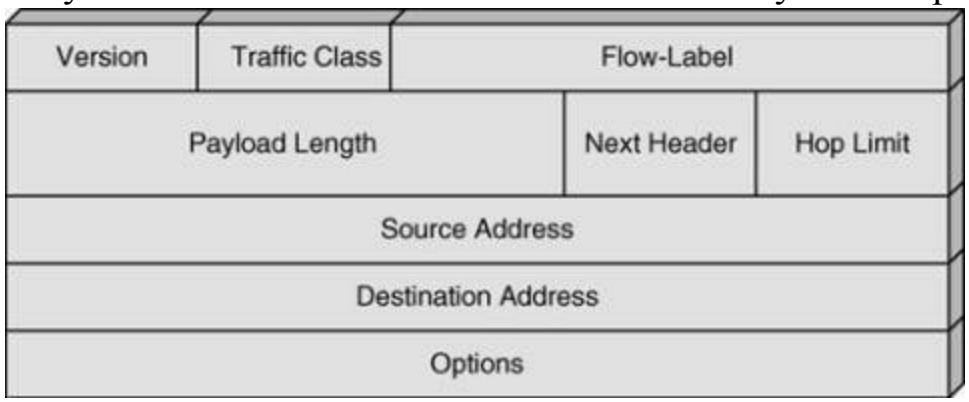


Figure 7.17

Format of the IPv6 packet

The IPv6 packet contains the following fields:

- **Version.** bears the number 6.
- **Priority (Priority).** Indicates a level of priority, which allows the processing of packets more or less quickly in the nodes of the network.
- **Flow-Label** (reference of flow). Also new, this field allows you to carry a reference (label) able to specify the stream to which belongs the packet and therefore indicate the quality of service required by the information transported. This reference allows routers to make appropriate decisions. Thanks to this new field, the router can handle in a customized way IPv6 packets, thus allowing the taking into account of various constraints.
- **Length** (length). Indicates the total length of the datagram in byte, without taking account of the header. This field is 2 bytes, the maximum length of the datagram is 64 kb.

- **Next-Header** (following header). Indicates the protocol encapsulated in the data area of the packet. This process is illustrated in Figure 7.18. The options of the most traditional for the value of this field are 0 for hop-by-hop Header Option, 4 for IP, 6 for TCP and 17 for UDP (see the box below).

The field values Next-Header

- 0hop-by-hop Header Option
- 4IP
- 6TCP
- 17UDP
- 43Routing Header
- 44Fragment Header
- 45IRP (Interdomain routing protocol)
- 4RSVP Resource Reservation Protocol ()
- 50Encapsulating Security Payload (ESP)
- 51Authentication Header
- 58ICMP
- 59No. Next-Header
- 60Destination Header Options

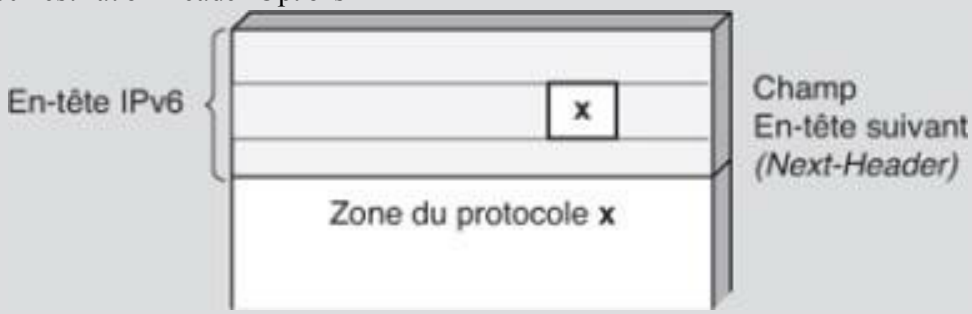


Figure 7.18
Next Header field (Next-Header)

- **Hop Limit** (Maximum number of nodes crossed). Specifies after how many nodes the packet is destroyed.
- **Address.** The address area is often presented as the reason for the new version of IP. In fact, it is only one reason among others. The IPv6 address takes on 16 bytes. The difficulty lies in the representation and the rational use of these 128 bits. The number of potential addresses exceeds 1023 for each square meter of the surface of the earth. The representation is carried out by a group of 16-bit and is in the form 123:FCBA:1024:AB23:0:0:24:FEDC. Of the series of addresses equal to 0 can be abbreviated by the sign ::, which can only appear once in the address, as in the example 123:FCBA:1024:AB23::24:FEDC. This sign does not indicate the number of successive 0. To deduct this number, the other series can only be abbreviated. If there were two series abbreviated, it would be impossible to deduce the respective length of each. IPv6 addressing is hierarchical. An allocation of addresses has been proposed, the details of which are given in Table 7.5.
- **Options.** The header of the IPv6 packet ends by a field of options that allows the addition of new features, in particular concerning the security. Figure 7.19 illustrates the operation of this field of options. Each area of option begins by a field bearing a number corresponding to the type of option. In this field of options, the different zones are in a predetermined order, which is dictated by their potential use in the intermediate nodes. If an intermediate node cannot support an option, several cases of the figure are: destruction of the

packet, emission without treatment, issuance of a signage or waiting for a response to take a decision. Figure 7.20 gives an idea of the order of processing.

Address	The first few bits of the address	Characteristics
0:: /8	0000 0000	Reserved
100 :: /8	0000 0001	Not assigned
200 :: /7	0000 0001	ISO Address
400:: /7)	0000 010	Novell address (IPX)
600 :: /7	0000 011	Not assigned
800 :: /5	0000 1	Not assigned
1000 :: /4	0001	Not assigned
2000 :: /3	001	Not assigned
4000 :: /3	010	Address of Service Provider
6000 :: /3	011	Not assigned
8000:: /3	100	Geographical address of user
A000 :: /3	101	Not assigned
C000::/3	110	Not assigned
E000 :: /4	1110	Not assigned
F000:: /5	1111 0	Not assigned
F800:: /6	1111 10	Not assigned
FC00 :: /7	1111 110	Not assigned
FE00:: /9	1111 1110 0	Not assigned
FE80:: /10	1111 1110 10	Link-local address
FEC0:: /10	1111 1110 11	Address of local site
FF00:: /8	1111 1111	Address of multipoint

Table 7.5 • Addresses of IPv6

Figure 7.19
The fields of options in the IPv6 packet

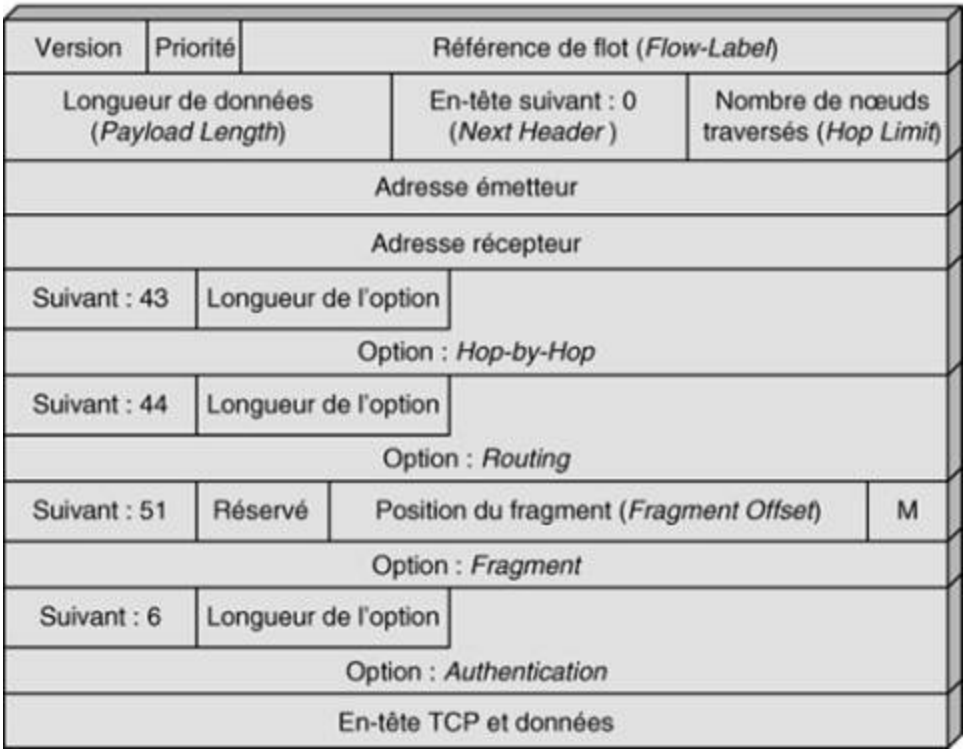


Figure 7.20
Treatment Order of expansion options.

The message level

The message level relates to the transfer of end-to-end data from one end to another in a network. The data of the user are grouped into messages, although this entity is not well-defined. These messages must be transported from the transmitter to the receiver. This is the reason for which this level is also called Transport Layer. It is also the term that is found in the nomenclature of the reference model.
The message is a logical entity of the user Transmitter, its length is not determined in advance. The message level is based on features capable of route information from one end to the other of the network. It corresponds to an end-to-end protocol. Its definition is precise: Ensure the routing of the message from the transmitter to the receiver, possibly through several networks. In comparison, the level package has for ambition that to do the necessary to ensure the crossing of a network. By deduction, no message level must not be crossed before reaching the terminal equipment of destination, otherwise the transmission would not be of end-to-End.
After having examined the features that are found in all the message level, this chapter presents fairly succinctly the main protocols from the OSI architectures, ATM, and the Internet.

The features of the message level

The message level is directly linked to the features of the layer 4 (Transport) of the reference model, but it also takes into account the equivalent levels of other architectures, TCP in particular. Its role can be described quite formally by the three properties defined for the transport layer, namely the transport of end-to-end, the selection of a quality of service and transparency.
The Transport layer must allow communication between two users located in different systems, regardless of the characteristics of the sub-networks on which the transfer is made of the data. A user of the transport service has not the possibility to know if one or several networks underlying are put into play in the communication which the interested.
Figure 7.21 illustrates a transport connection involving several networks put end-to-end, or concatenated. The role of the transport layer is to achieve the operation of the transport connection

established between X and Y. The problems inherent to routing and the concatenation of the network connections are taken into account by the Layer 3. In a transport connection, the information must be issued in the order, without loss or duplication.

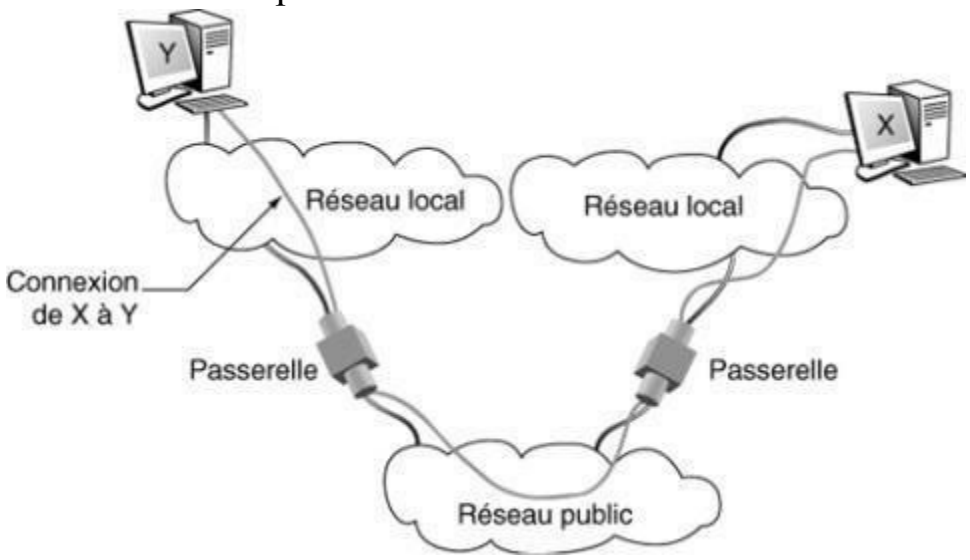


Figure 7.21
Example of End-to-end connection

To achieve the desired quality of service by the users of the transport service, the transport protocol must optimize the available network resources, and at the lowest cost. The concept of quality of service is defined by the value of some of the parameters, including the ISO has established an exhaustive list:

- Time limit for the establishment of a transport connection;
- Probability of Failure of the establishment of a connection;
- Flow of information on a connection of transport;
- Time of crossing of the connection: Time elapsed between the issuance of the data and their reception at the other end of the connection;
- Rate of residual error: rate of uncorrected errors encountered on a connection;
- Probability of Failure: The probability to stop a transport connection not desired by the users;
- Disconnect Timeout: maximum time acceptable to determine properly a transport connection;
- Probability of failure of disconnection: probability of a disconnection not coordinated between users, often leading to a loss of information;
- Protection of connections of transport, within the meaning of maintenance of a security, to avoid unauthorized manipulation of data circulating on a connection of transport;
- Priority of connections: relative importance to use different connections in the event of a major problem at the risk to degrade the quality of service of a link, or even of the complete;
- The fragility of a connection: probability of accidental cut of a transport connection already established.

The information exchanged on a transport connection the are regardless of their format, their coding or of their meaning. This is what is called a transparent mode. This transparency is achieved mainly by the protocols of lowest level, which must also work in a transparent mode.

The transport layer is based on a method of addressing independent of the conventions used in the lower layers. Its role is to achieve a correspondence between the transport address of a given user

and an address of a network to be able to initialize the communication.

The characteristics of the message level

It is interesting to note the parallel that exists between the specifications of the message level, or Layer 4 (transport) and those of the packet level, or Layer 3 (network). In a certain sense, these two levels must achieve efficient transport through several nodes. The fundamental differences between the two levels are however to consider:

- The problems of addressing are more simple at the packet level that at the message level.
- The concept of the establishment of a virtual circuit at the level package has a sense more practical: it must open a path that the different packages will follow.
- The connection is established at the message level is characterized by a quantity of important information in the course of transmission to the interior of the network. This property comes from the potential crossing of multiple Layer 3 networks (see Figure 7.21). At the level package, the memory of the transmission medium is often much more low, with the exception of satellite networks.

The concept of transparency, previously stated as a property of the transport layer, implies the possibility of route information of any size. This is the transparency vis-a-vis the format of the data. In practice, this means that the achievement of an entity to transport requires a complex management of submissions for the storage of information. In practice, it must be able to manage buffers at the same store TPDU of a size ranging between 128 and 8 192 bytes in the standardized protocol and up to 64 kb in the Internet.

Address and Data paths

Figure 7.22 illustrates the connections between several users located on two remote machines assuming that the communication network uses virtual circuits.

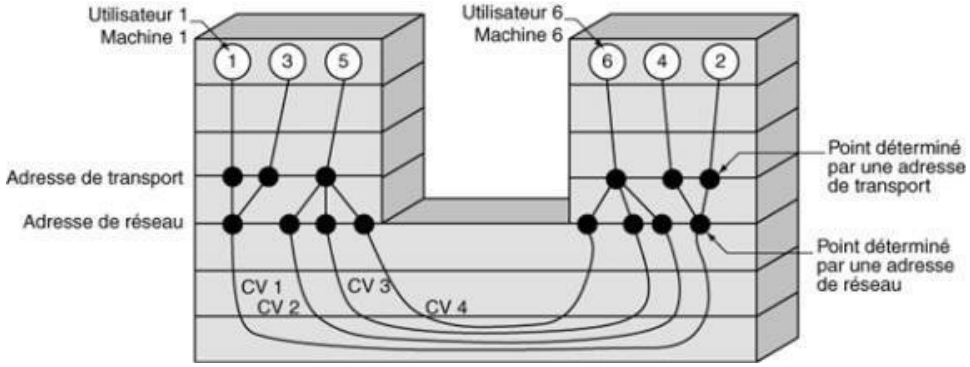


Figure 7.22

Addressing and data path

The transport addresses shown are the points of access to the service of transport visible by the protocol layer above, i.e. the session layer. This example shows the optimization that supports the message level thanks to the multiplexing and to bursting. The multiplexing corresponds to the sharing of a network connection (CV, or virtual circuit, 1) by multiple connections of Transport (1-2 and 3-4). The bursting corresponds to the use of multiple network connections (CV 2, resume 3 and resume 4) By the same transport connection (5-6). These two mechanisms to optimize the cost and performance of a connection. It has recourse to the multiplexing when, for example, connections of transport in low flow are necessary in large number. It uses the burst to maximize the throughput of a transport connection which should be established on a network with low performance.

Location of layers of protocols in a computer system

It is interesting to situate now the different layers of communication protocols in a computer system by network context.

Figure 7.23 shows how to distribute the more often the layers of protocols between a computer and the frontal responsible for communications. The sharing of functions highlights the separation between layers High (5, 6 and 7) and the lower layers (1, 2, and 3) of the ISO - the first are oriented treatment, and the seconds communication -, as well as the pivotal role played by the message level.
The personal computer has however modified this distribution. The personal computer manages in effect the layers high, the message level and the level IP packet. The network adapter that is added is responsible for its share of the lower layers (1 and 2). As shown in the Figure 8.3, Layers 3 and 4 which were found in the operator on the frontal machine are found here on the side of the computer, become a micro-computer. The card coupler that has taken the place of the frontal no longer takes into account that the two lowest layers.

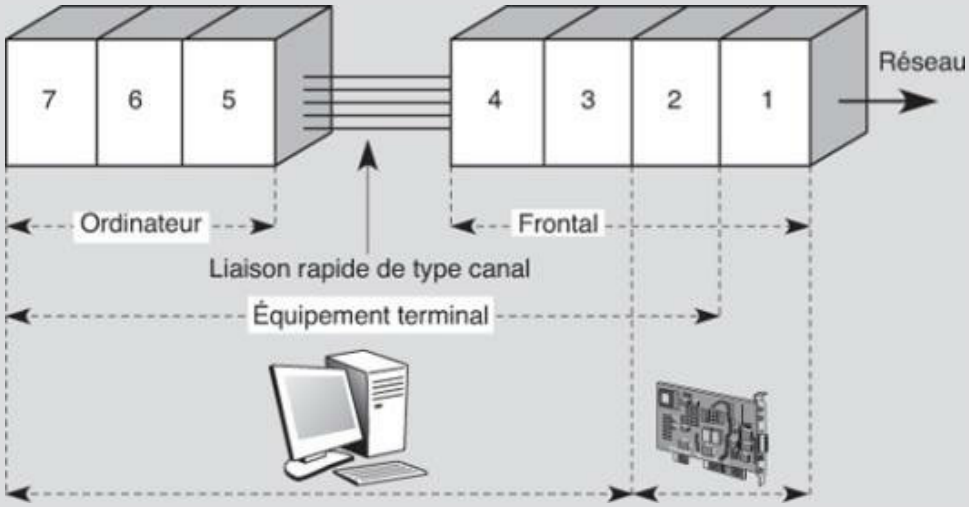


Figure 7.23
Distribution of layers of protocols in a terminal device

Negotiation of a quality of service

As indicated previously, the selection of a quality of service figure among the possibilities of the message level. It may be asked who selects this quality of service. The answer is relatively simple: the user of the transport service expresses its desire in terms of quality of service, and the provider of the service shows him in response if it may or may not meet its requirements. If it cannot, it specifies the quality of service it can offer in relation to initial applications. This process is called the negotiation. It applies to each parameter negotiable of a transport connection.
Figure 7.24 illustrates the flow of such a negotiation for the choice of the actual flow of information on a connection in the course of the establishment. In a real situation, all parameters are negotiated in this way, either successively or simultaneously. In some cases, explanations can be provided by the provider of the transportation service indicating the reasons for the final choice.

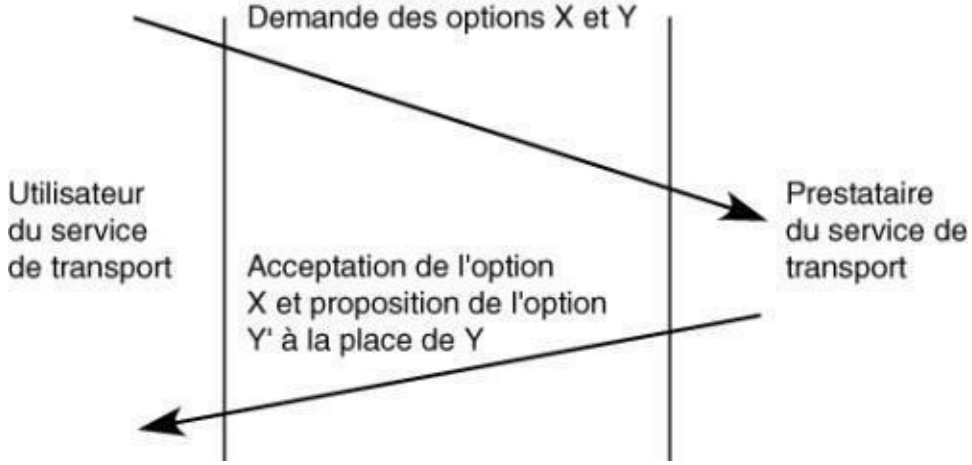


Figure 7.24
Negotiation of a connection parameter

The protocols of the message level

This section focuses on the protocols of the message level from the Internet architectures, Ethernet and ATM. The case of the OSI architecture is examined in Appendix E.

TCP and UDP, the two Protocols to the message level the most important, come from the World Internet. They are studied in the first, the protocols of the message level associated with the worlds Ethernet and ATM being introduced more succinctly.

The TCP protocol

The Internet network uses the IP protocol at the packet level. The message level offers two other possibilities: the Transmission Control Protocol (TCP), which introduced several features that guarantee a certain quality of the transport service, and the Protocol UDP (User Datagram Protocol), much more simple, but not giving any guarantee on the transport of messages. The simplicity of UDP offers in return for higher data rates.

TCP provides a service of reliable transportation. The data exchanged are considered as a stream of bits divided in bytes, the latter to be received in the order in which they are sent. The data transfer may only begin after the establishment of a connection between two machines. This property is shown in Figure 7.25. During the transfer, the two machines continue to check that the data passes correctly.

The application programs send their data by passing them regularly to the operating system of the machine. Each application chooses the size of data that suits him. The transfer can be, for example, of one byte at a time. The TCP implementation is free to cut the data in packets of a different size than the size of the blocks received from the application. To make the transfer more efficient, the TCP implementation expects to have sufficient data before completing a datagram and send on the sub-network.

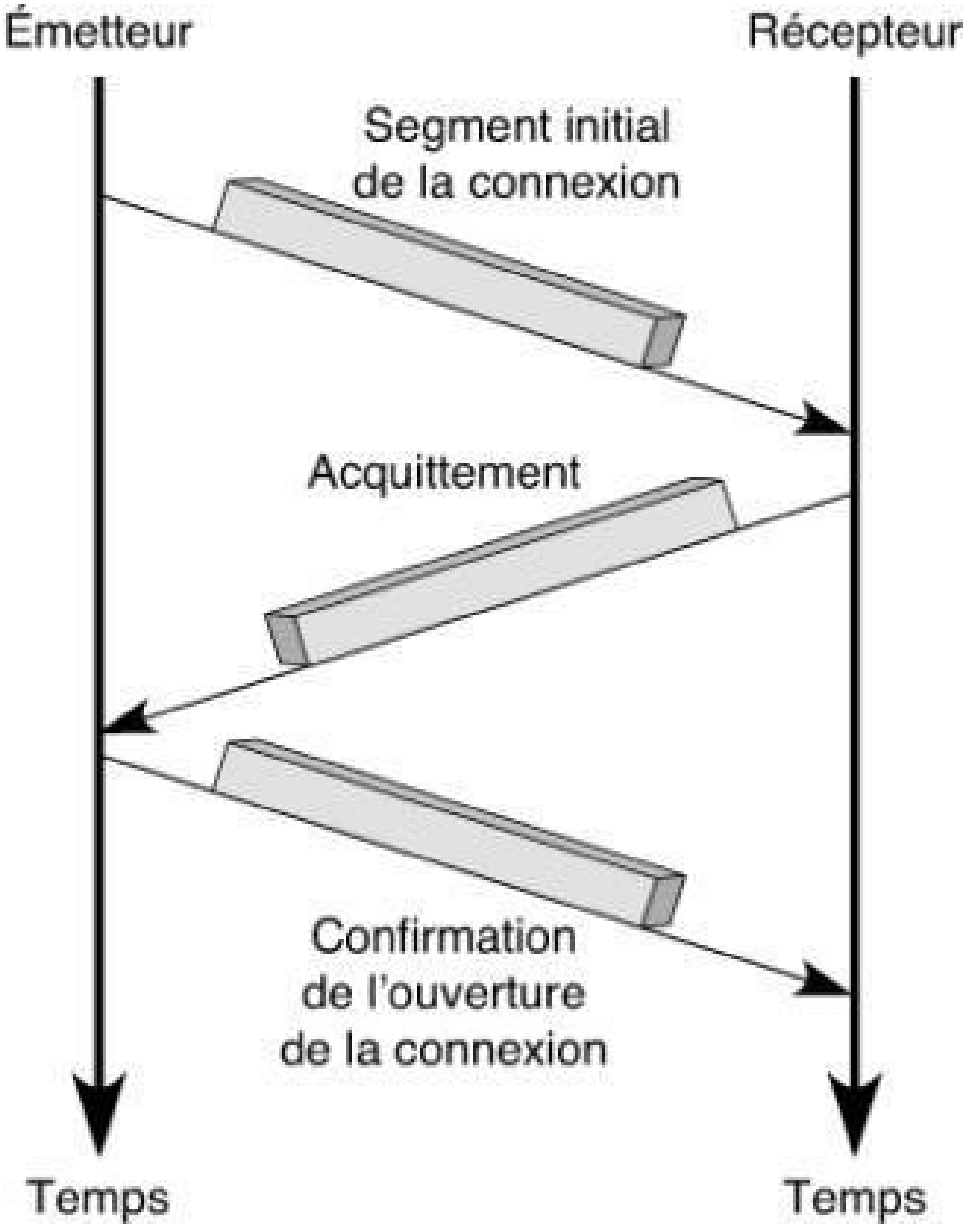


Figure 7.25
Establishment of a TCP connection

Open in both transmission directions at once, the connection guarantees a bi-directional data transfer, with two streams of data inverse, without apparent interaction. It is possible to complete the shipment in a direction without stop the one in the other direction. This allows the sending of acquittals in a direction of transmission at the same time that data in the other direction.

The TCP protocol defines the structure of the data and acquittals exchanged, as well as the mechanisms to make the reliable transportation. It specifies how to distinguish multiple connections on the same machine and how to detect lost packets or duplicated and remedy this situation. It also defines the way to establish a connection and the finish. TCP permits several programs to establish a simultaneous connection and for multiplexing the data received from different applications. To do this it uses the abstract notion of port, which identifies a particular destination in a machine.

TCP is a protocol in mode with connection. It has meaning only between two points end of a connection. The program of one end performs a opening of passive connection, i.e. that it accepts an incoming connection by assigning him a port number. The other application program executes a opening of active connection. Once the connection is established, the data transfer can begin. The concept of port is shown in Figure 7.26.

For the protocol TCP, a data stream is a suite of bytes grouped into fragments. The fragments usually give birth to an IP packet. The TCP protocol uses a mechanism of window to ensure a transmission

performance and a flow control. The mechanism of window allows the anticipation, i.e. the sending of several fragments without waiting of acquittal. The flow is improved.

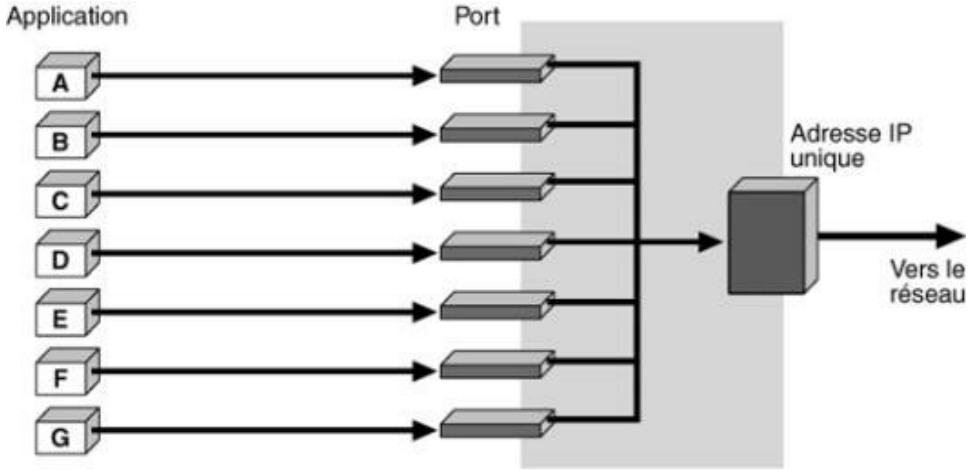


Figure 7.26

Connection of multiple applications on the same IP address

The window also allows you to achieve a end-to-end flow control, allowing the receiver to restrict the sending of data as long as it has not the place of the receive in his memoirs. The mechanism of window operates at the level of the byte and not of the fragment. The bytes to be transmitted are sequentially numbered. The transmitter manages three pointers for each window. In the same way, the receiver must maintain a window in reception, which indicates the number of the next expected byte, as well as the extreme value which can be received. The difference between these two quantities indicates the value of the credit is accepted by the receiver, value which generally corresponds to the buffer available for this connection. TCP flow control is shown in Figure 7.27. It is presented in more detail a little later in this chapter.

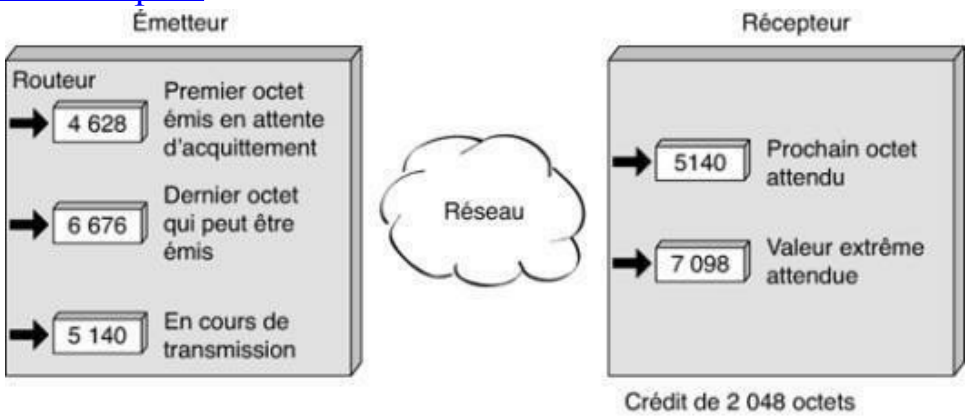


Figure 7.27

TCP flow control

During a connection, it is possible to exchange data in each direction, each end of the connection before in this case maintain two windows, one in emission, the other in reception.

The fact that the size of the window varies in time constitutes an important difference compared to a mechanism of classic window. Each acquittal, specifying how many bytes have been received, contains an information window size to the number of additional bytes that the receiver is able to accept. The size of the window can be considered as the space available in the memory of the receiver. The latter may not reduce the window below a certain value, that he has accepted previously. The fact that the window size can vary in time constitutes an important difference compared to a mechanism of classic window.

The unit of protocol of TCP is the fragment, fragments are exchanged to establish the connection, transfer data, change the size of the window, close a connection and issue of the acquittals. Each fragment is composed of two parts: the header and the data. The format of a fragment is shown in

Figure 7.28. The information flow control can be transported in the data stream reverse.

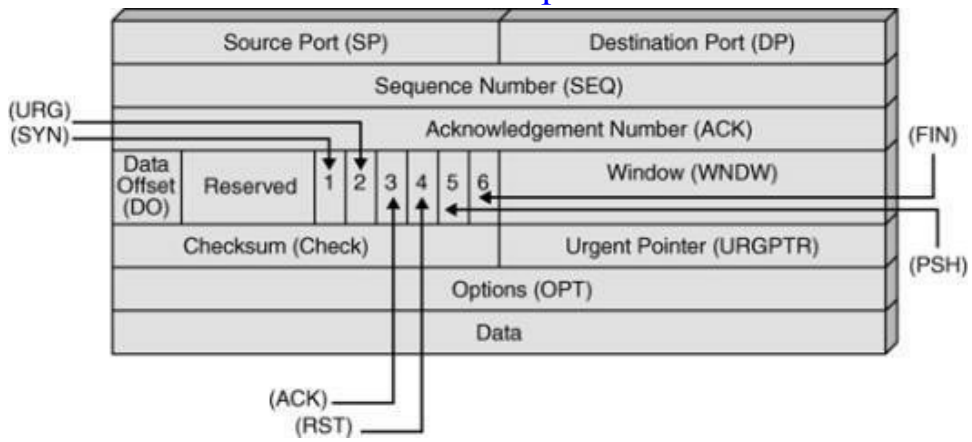


Figure 7.28
Format of a TCP fragment

The TCP fragment contains the following areas:

- **SP (Source Port),** or source port. 16 bit field containing the address of the port of entry. Associated with the IP address, this value provides a unique identifier, called socket, formed from the concatenation of the IP address and the port number. The identifier is used to determine an application running on a machine terminal.
- **DP (Destination Port),** or port of destination. 16 bit field, whose function is identical to the previous one, but for the destination address.
- **SEQ (Sequence Number),** or sequence number. Field on 32 bits indicating the number of the first byte worn by the fragment.
- **ACK (Acknowledgment number),** or number of acquittal. Field on 32 bits indicating the Seq Number of the next expected fragment and corresponding to the acquittal of all the bytes received previously. The value ACK Indicates the number of the first octet expected, either the number of the last received byte + 1.
- **DO (data offset),** or the length of the header. 4-bit field indicating the length of the header by a multiple of 32 bits. If the value 8 is located in this field, the total length of the header is 8 × 32-bit. This value is required of the fact that the zone of option may have any size. It is deduced that the length of the header may not exceed 15 × 32-bit, or 60 bytes.
- The next area is reserved for later use. This field must be filled with 0.
- **URG (Urgent pointer),** or pointer of emergency. Field On 1 bit, numbered 1 in Figure 7.28. If this bit is set to a value of 1, the urgent field point located in the result of the header has a significant value.
- **ACK (acknowledgment),** or acquittal. Field On 1 bit, numbered 3 in Figure 7.28. If ACK = 1, the field Acknowledgment number located in the header contains a significant value, to take into account by the receiver.
- **Hsp (Push function),** or function of push. Field On 1 bit, numbered 5 in Figure 7.28. If HSP = 1, the issuer wishes that the data of this fragment are issued as soon as possible to the recipient.
- **RST (Reset),** or restart. Field On 1 bit, numbered 4 in Figure 7.28. If RST = 1, the issuer asks that the TCP connection restarts.
- **SYN (synchronization),** or synchronization. Field On 1 bit, numbered 2 in Figure 7.28. SYN = 1 means a request for opening of the connection. In this case,

the sequence number is the number for the first byte of the stream.

- **End (Terminate),** or closure. Field On 1 bit, numbered 6 in Figure 7.28. End = 1 means that the issuer wishes to close the connection.
- **WNDW (window),** or window. 16 bit field indicating the number of bytes that the receiver accepts to receive. More exactly, the value of WNDW contains the ultimate byte number that the transmitter of the fragment accepts to receive. By subtracting the number indicated in the value of the ack field, it gets the number of bytes that the receiver accepts to receive.
- **CHECK (checksum).** 16 Bit field to detect errors in the header and the body of the fragment. The protected data is not limited to the fragment TCP. The Checksum also takes account of the IP header of the source address, called *pseudo-header, to protect these sensitive data. A pseudo-header is a modified heading, including some fields have been removed and other added, that the area of Error Detection takes into account in its calculation.*
- **URGPTR (Urgent pointer),** or pointer of emergency. 16 bit field specifying the last octet of an urgent message.
- **OPT (Options),** or options. Area containing the different options in the TCP protocol. If the value of the field do (data offset), indicating the length of the header, is greater than 5, it is that there is an option field. To determine the length of the option field, simply subtract 5 from the value of DO. Two formats are working simultaneously. In one case, the first byte indicates the type of the option, which implicitly defines its length, the following bytes giving the value of the option setting. In the other case, the first byte always indicates the type of the option, and the second the value of the length of the option. The main options relate to the size of the fragment, that of Windows and timers, as well as the constraints of routing.

The fragment ends by the transported data.

The fragments are variable in size, the acquittals relate to a number of particular byte in the data stream. Each acknowledgment message specifies the number of the next byte to be transmitted and pays the precedents.

The acquittals TCP is cumulative, they repeat themselves and to accumulate to specify up to what byte the flow has been well received. For example, the receiver can receive a first acquittal of flow up to byte 43 568 and then a second up to byte 44 278 and a third up to byte 44 988, indicating three times that up to byte 43 568 everything has been well received. This cumulative principle allows you to lose the first two acquittals without consequence.

This process has advantages, but also disadvantages. A first advantage is to have the acquittals simple to generate and not ambiguous. Another advantage is that the loss of an acquittal does not necessarily require the retransmission. In contrast, the transmitter does not receive the acquittals of all successful transmissions, but only the position in the stream of data that have been received. This process is illustrated in Figure 7.29, which, to make simple, indicates the numbers of packets while, in reality, this are numbers of bytes that are transmitted.

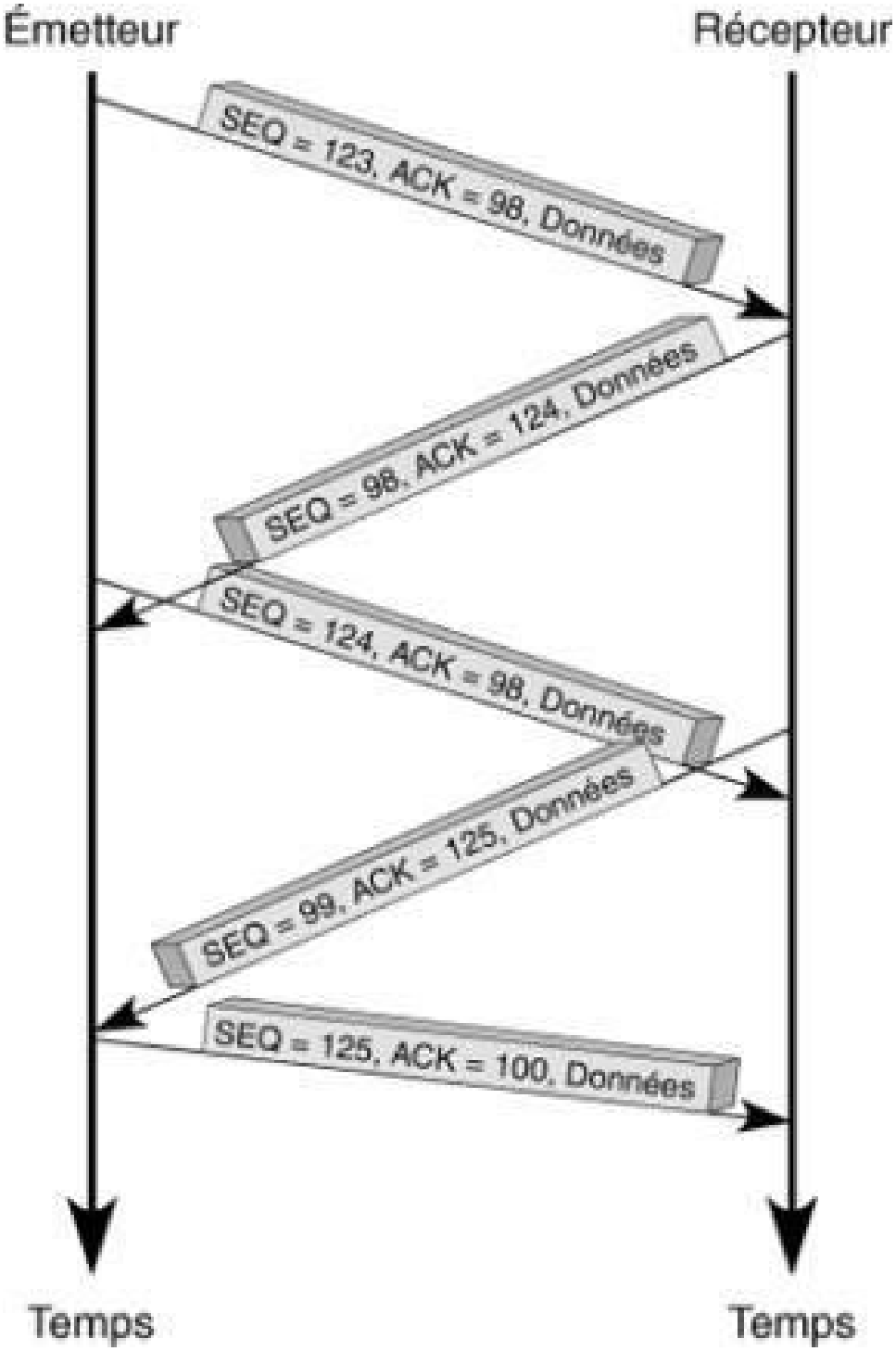


Figure 7.29
Process of acquittals in TCP

How to manage the timers and the acquittals constitutes one of the essential characteristics of the TCP protocol, which is based on the principle of the acquittals positive. Each time a fragment is issued, a timer is triggered, waiting for the acquittal. If the acknowledgment arrives before the timer reaches maturity, the timer is stopped.

If the timer expires before the data of the fragment have been paid, TCP assumes that the fragment is lost and the retransmits. This process is illustrated in Figure 7.30.

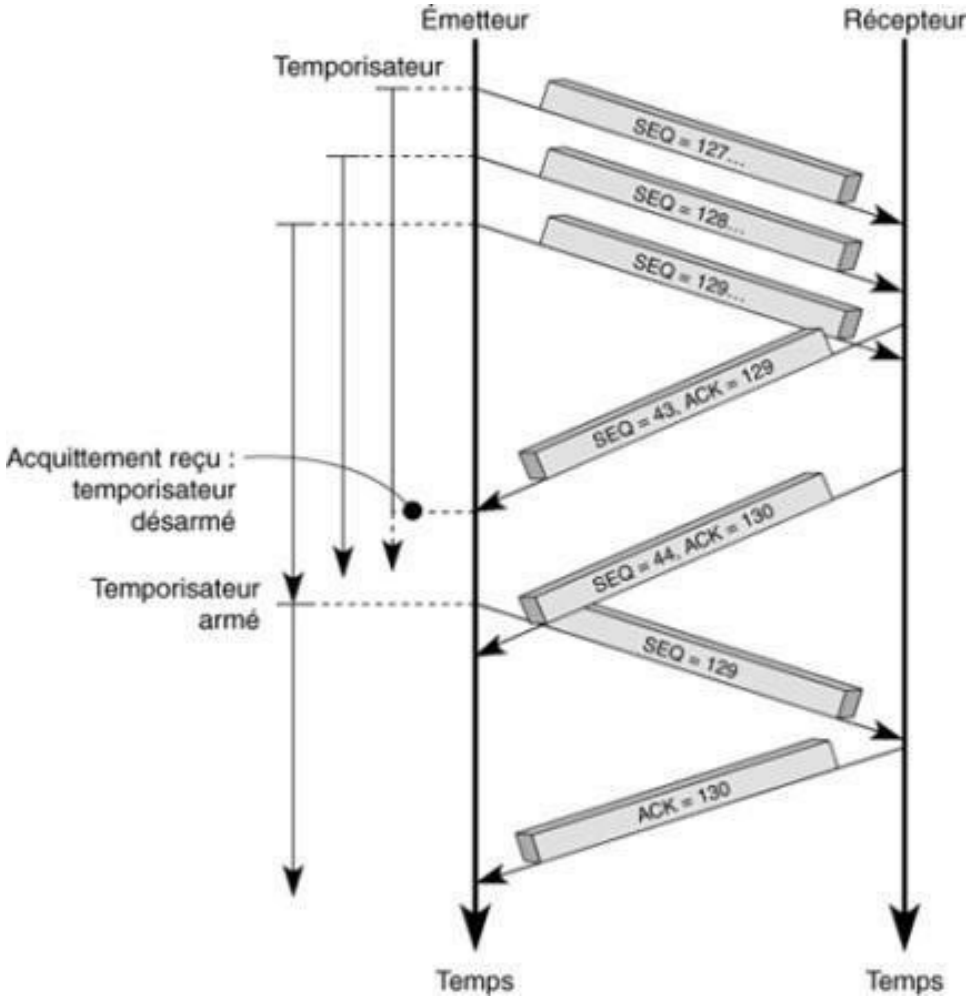


Figure 7.30

Process to delay and recovery in TCP

Operation of timer of recovery

TCP does doing no assumption on the time of transit in the networks traversed, it is impossible to know *a priori* the moment of arrival of an acquittal. In addition, the time of crossing of the Routers and Gateways depends on the network load, which varies itself in time. TCP uses an adaptive algorithm to take into account these variations. It saves for this the time to which he sent the fragment and the time at which it receives the corresponding acknowledgment. After several measures of this type, the transmitter makes an estimate of the time necessary for the reception of the acquittal. This estimate allows him to determine a duration for the timer of recovery.

During congestion, TCP reacts by reducing the flow rate of the connection. The Protocol has the possibility to measure the importance of the problem by observing the increase in the response time. If the protocol does not react to congestion, the number of retransmissions can continue to increase and worsen as well the congestion. This is the reason for which a control algorithm reduces the flow in the event of congestion. This algorithm, called *slow-start and collision avoidance, literally "slow start and collision avoidance," must be fully distributed since there is no central system of control in TCP. Its principle is to begin a window size of 1, and to double the size of the window each time that the whole of the packets in the window has been well received before the end of the timers of respective recovery. When a fragment arrives late, that is to say after the timer has expired, it is retransmitted in restarting of a window of 1.*

During the second phase of the algorithm, collision avoidance, where a delay is detected, which requires a restart on a window of 1, the size of the window *N* that has caused the delay is divided by 2 (N/2). From the value of the size 1 restart, the double size until the size of the window exceeds n/2. At this time, we are back to the previous size, which was less than n/2, and, instead of doubling, it only adds 1 to the size of the window. This process of adding 1 continues until a delay of acquittal restarts the process to the window size of 1. The new value that triggers the part collision avoidance is calculated from the window reached divided by two.

An example of a behavior of this algorithm is shown in Figure 7.31.

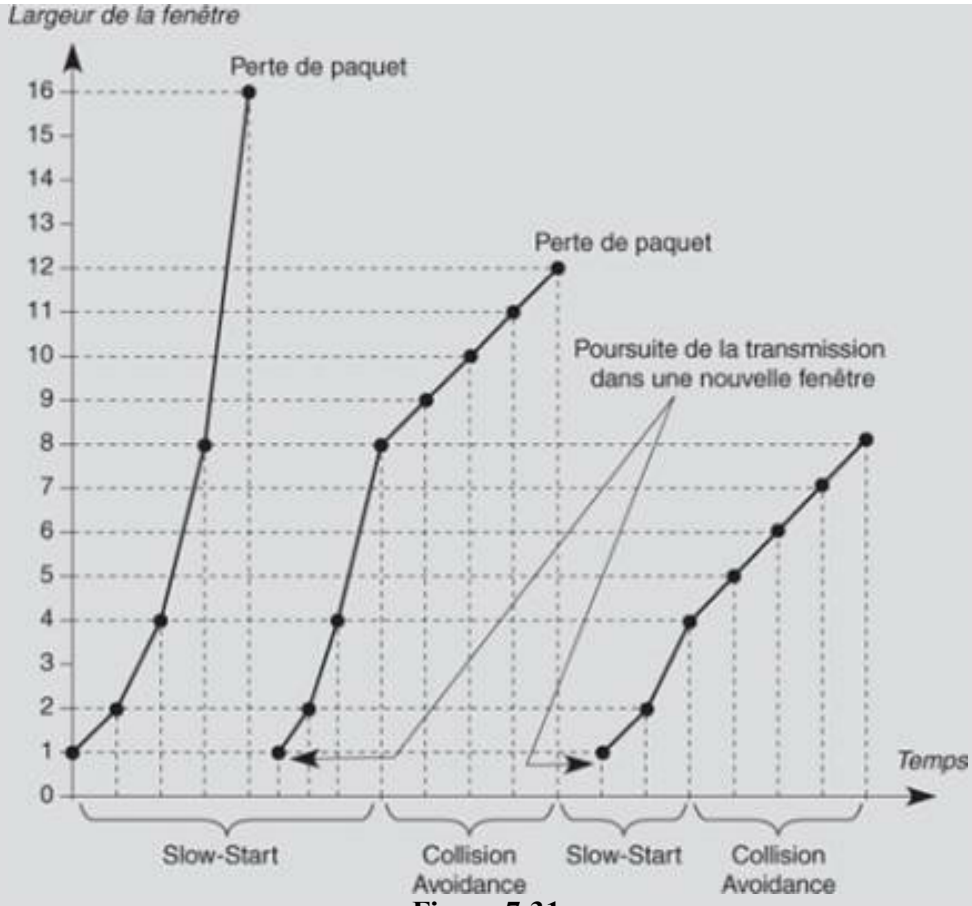


Figure 7.31

Operation of the algorithm slow-start and collision avoidance

One could believe that the system is never stable, but it retains a means to stabilize. When the acquittals arrive before the window is reached, this indicates that the maximum flow rate of the connection is reached and that it is no longer necessary to increase the value of the window.

The UDP protocol

The UDP protocol allows applications to exchange datagrams. To do this it uses the concept of port, which allows to distinguish between the different applications that run on a machine. In addition to the datagram and its data, a UDP message contains a source port number and a destination port number.

The UDP protocol provides a service in mode without connection and without error recovery. It uses no acquittal, do not reséquence messages and does puts in place no flow control. It may be therefore that the UDP messages that are lost are duplicated, handed out of sequence or they arrive too early to be treated at the time of their receipt. As explained previously, UDP is a protocol particularly simple message level of the architecture of the reference model. It presents the advantage of fast execution, taking into account real time constraints or a limitation of the spot on a processor. These constraints or limitations do not always allow the use of protocols more heavy, as TCP.

Applications that do not have the need for a strong security level transmission, and they are many, as well as the software of management, which require quick questions of resources, prefer to use UDP. Research Requests in the Directories pass through UDP, for example.

To identify the different applications, UDP imposes to place in each fragment a reference which plays the role of port. Figure 7.32 illustrates the structure of a fragment UDP. A reference identifies, a little like the Next Header field in the IPv6, which is transported in the body of the fragment. The most significant applications that use the UDP protocol correspond to the following port numbers:

- 7: ECHO service;
- 9: Service of rejection;
- 53: Domain Name Server Domain Name Server (DNS);

- 67: DHCP configuration server;
- 68: configuration client DHCP.

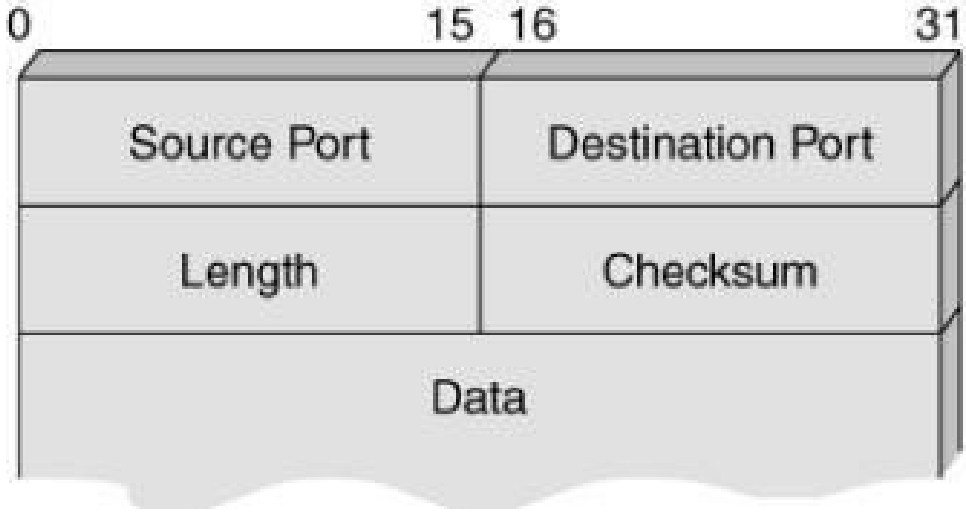


Figure 7.32
Structure of a UDP fragment

Conclusion

The IP protocol has taken the front of the scene since the 1990s. It is a standard of fact, Old of more than forty years, well known today and which has the great merit of be conceptually simple. In contrast, the engineering of IP networks remains a complex area, not yet totally controlled.

The role of the IPv6 version is to propose a protocol much more manageable, thanks to new fields to introduce an address better built, an area of identification of flows and many options, in particular in the field of security. However, all the new features introduced in IPv6 have been added to IPv4.

The fact remains that the cost of transition to IPv6 is far from being negligible. The western world is often considered that this cost is disproportionate in comparison to the actual intake of IPv6 in relation to the latest versions of IPv4.

Be that as it may, IPv6 replaces in small IPv4. The mass arrival of mobile terminals and the direct addressing of all terminal stations involve the adoption of the address of the new generation. The recognition of the flows will assist in the introduction of new security features, quality of service and of the control of mobility.

The message level is concerned primarily with cut the information into segments exploitable by the lower levels. It can provide additional functionality, such as a request for the retransmission of a message if an error is detected or a flow control, as in TCP.

There are virtually no more that two protocols of transport in the networks, TCP and UDP, both from the IP world.

Part III

The Network Solutions

The network solutions are interested in how the messages are transported from one end to the other end of the network through the different layers of the Protocol. The architecture is strongly dependent on the level at which it must go back to do this in the intermediate nodes.
Chapter 8 is dedicated to the networks of the physical level and the Chapter 9 to those of frame level, represented by the Ethernet networks. Chapters 10 and 11 spend at the top layer, with IP networks and then MPLS networks and their extension GMPLS, which require both the level 2 and Level 3.

8

The networks of the physical level

This chapter examines the transport techniques of signals on the optical networks, as well as the wavelength division multiplexing, which brings a very substantial increase in the capacity of an optical fiber, and the techniques of transport of frames.

Optical networks

Optical networks allow you to carry signals under An optical form and non-electrical, to the difference of the conventional networks. The benefits of the optic are many, including because the signals are better preserved, since they are not disturbed by the electromagnetic noises, and that the speeds are very important.

The optical fiber

Considered as the support allowing the highest flow rates, the optical fiber is a technology today completely under control. In the metallic wires, it transmits the information through an electrical current modulated. With the optical fiber, one uses a light beam modulated. It was necessary to wait for the 1960s and the invention of the laser for this type of transmission develops.

One optical connection requires a transmitter and a receiver. Different types of components are possible. Figure 8.1 illustrates the structure of a fiber optic link. The digital information are modulated by a transmitter of light, which may be:

- A light emitting diode or LED (Light Emitting Diode) which does not contain a laser cavity.
- A diode to infrared.
- A laser for singlemode fiber.

The phenomenon of dispersal is less important if it uses a laser, which offers an optical power superior to LEDS, but at a cost more important. In addition, the life of a laser is less than that of a light emitting diode.

The light beam is conveyed to the inside of an optical fiber. The latter consists of a cylindrical guide with a diameter between 100 and 300 microns (µm), covered with insulating material for the multimode fibers and between 50 and 62.5 µm for singlemode fiber.

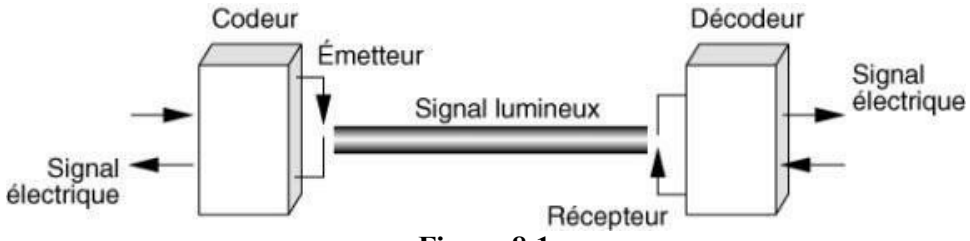


Figure 8.1

There are two types of receptors:

- The PIN photodiodes;
- The avalanche photodiodes.

The components end, transmitters and receivers, currently limit the speeds that can be achieved on the fibers.

The main benefits of the optical fiber are the following:

- Very wide bandwidth, of the order of 1 GHz for one kilometer, which allows the Multiplexing on a same media very many channels such as the telephone, television, etc.
- Small footprint.
- Great lightness, the weight of an optical cable per unit of length, of the order of a few grams per kilometer, being approximately nine times lower than that of a conventional cable. The radius of curvature can descend below 1 cm.
- Very low attenuation, which allows to consider a spacing important from the point of regeneration of transmitted signals. The no regeneration is greater than 10 kilometers, while on the coaxial cable, it is of the order of 2 to 3 kilometers. A system in optical fiber debiting several gigabits per second on a wavelength of 0.85 μm presents a weakening of 3 dB/km, which gives a no regeneration of nearly 50 kilometers. We can today, thanks to the optical amplifiers to Doped Fiber to the erbium, reach thousands of kilometers.
- Excellent quality of the transmission. A connection by light beam is, for example, insensitive to the thunderstorms, to Sparks and to electromagnetic noise. This immunity to noise is one of the main advantages of the optical fiber, which is particularly recommended in a bad electromagnetic environment. The wiring of the workshops and industrial environments can thus be carried out in optical fiber.
- Good resistance to heat and cold.
- Raw material good market, the silica.
- Absence of radiation, which makes his employment particularly interesting for military applications. An intrusion attempt on the optical fiber can be easily detected by the weakening of the light energy in reception.

The optical fiber, however, presents a few employment difficulties, including the following:

- The difficulties of connection as well as between two fibers that between a Fiber and the module of the emission or reception. You can achieve connections for which the losses are less than 0.2 dB. On the ground, it is necessary to appeal to removable connectors which require a precise adjustment and cause losses greater than 1 dB. Of this fact, when one wants to add a connection to a fiber optic media, it is necessary to cut the optical fiber and add of the connectors very sensitive to place. The passage of light electrical (see figure 8.2) that the one adds fact lose the benefits of lower attenuation and good quality of the transmission.

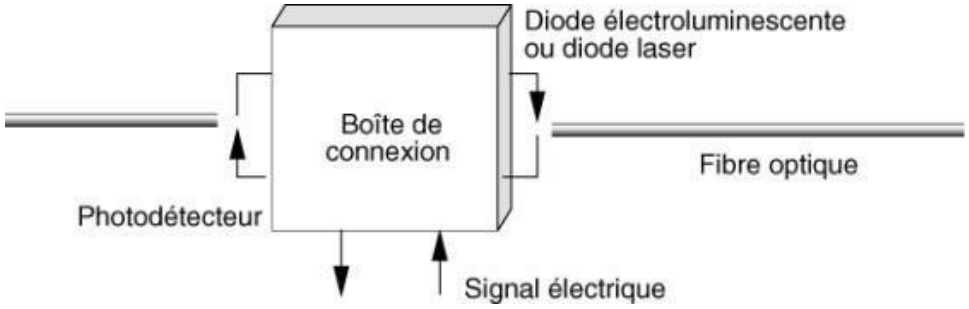


Figure 8.2

Connections to an optical fiber

- Leads difficult to achieve, the weakening that derives from often exceeding 5 dB. These leads are necessary since the components end of the optical access are most often assets and generate a final failure of the network in the event of failure (see figure 8.3).

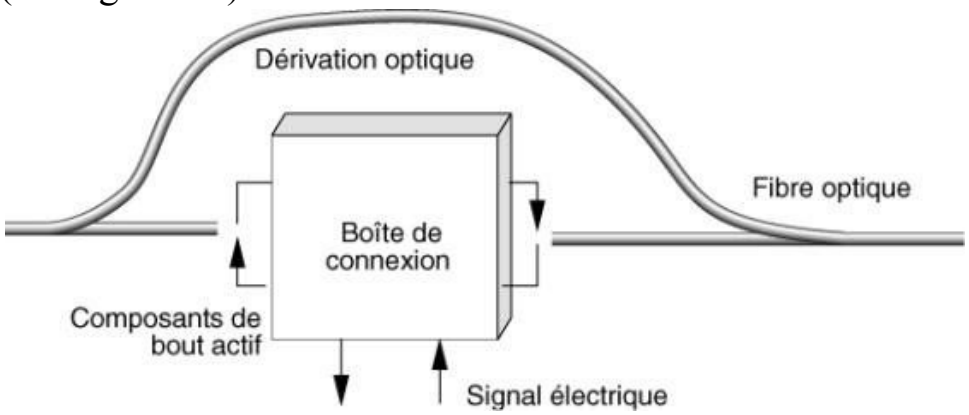


Figure 8.3

Bypass for optical fiber

- The wavelength division multiplexing, which is to make in transit in a same fiber several wavelengths in parallel or even what we can call several colors at the same time. It is easily found on the market of multiplexages in wavelength up to a hundred of colors. The current limit is of the order of a thousand wavelengths on the same optical fiber. With a thousand wavelengths, the fiber is completely filled, and new progress can only be made if a significant discovery is performed to further increase the number of wavelengths.

There are several types of fibers:

- The multimode fibers to jump of index;
- The multimode fibers to gradient index;
- Singlemode fiber, diameter, very small.

The multimode fibers to jump of the index have a bandwidth of up to 100 MHz on one kilometer and those to gradient index up to 1 GHz on a kilometer. Singlemode fiber offer the greater capacity of potential information, of the order of 100 GHz/km, and the best flows, but these are also the most complex to achieve. It generally uses optical cables containing several fibers. The insulation surrounding the fibers avoids the problems of crosstalk between the different fibers.

The optical fiber is particularly adapted to the point-to-point connections Digital. You can achieve multipoint connections using optical couplers or stars optics.

The world of the optical fiber is always in full evolution. Many of the research are underway, some of which have already resulted. One realizes including switches optical packets, the use of which could prove to be particularly interesting in the future to further increase the flexibility of switching in the optical fiber.

The costs of equipment and installation hinder however its employment in the access networks, even

if it has become the physical medium the more widespread in the hearts of network. Its insensitivity to electrical disturbances makes its use necessary in some environments strongly disturbed or in specific situations, such as the wiring of a company wishing to that no radiation can be detected at a distance from the cables. In effect, it is possible to detect the signals in a metal cable using equipment *ad hoc, the fact of the radiation of electrical waves that are circulating in the cable.*

The wavelength division multiplexing

On the metal cables, and including the coaxial cable, it uses more and more a multiplexing in frequency to pass multiple channels in parallel on different frequencies. If one wants to resume this idea in the optical fiber and achieve the passage of several light signals simultaneously, it is necessary to appeal to a wavelength division multiplexing. Today, many components adapted end offer this possibility.

The flow rates can reach of the so 10 Gbit/s on a single wavelength, with a rise in power anticipated soon to 40 Gbit/s, or 160 Gbit/s. Speeds of even larger have already been obtained in the laboratory. A fiber to 128 wavelengths of a flow of 10 Gbit/s provides a total flow of 1.28 Tbit/s. This represents approximately 60 million telephone voice in transit to the same time on the physical media. In other words, the whole of the French population can call to 60 million people at the same time in using only one optical fiber.

Architecture of optical networks

Optical networks rely on the wavelength division multiplexing, which consists, as explained in the previous section, to divide the optical spectrum in several sub-channels, each sub-channel is being associated with a wavelength. This technique is also called WDM (Wavelength Division Multiplexing), DWDM (Dense WDM), when the number of wavelengths exceeds the 20, then U-DWDM (ultra-Dense WDM) for more than two hundred wavelengths.

A particular case has been developed with CWDM (Coarse WDM) to reduce the cost of the wavelength division multiplexing with a separation between the wavelengths much more important. This solution greatly reduced the cost of optical components, but also decreases the number of wavelengths to less than twenty, which is largely sufficient in the metropolitan environments.

Figure 8.4 illustrates a network of communication using the wavelength division multiplexing. The path to a node to another can be fully optical or go by switches optoelectronics.

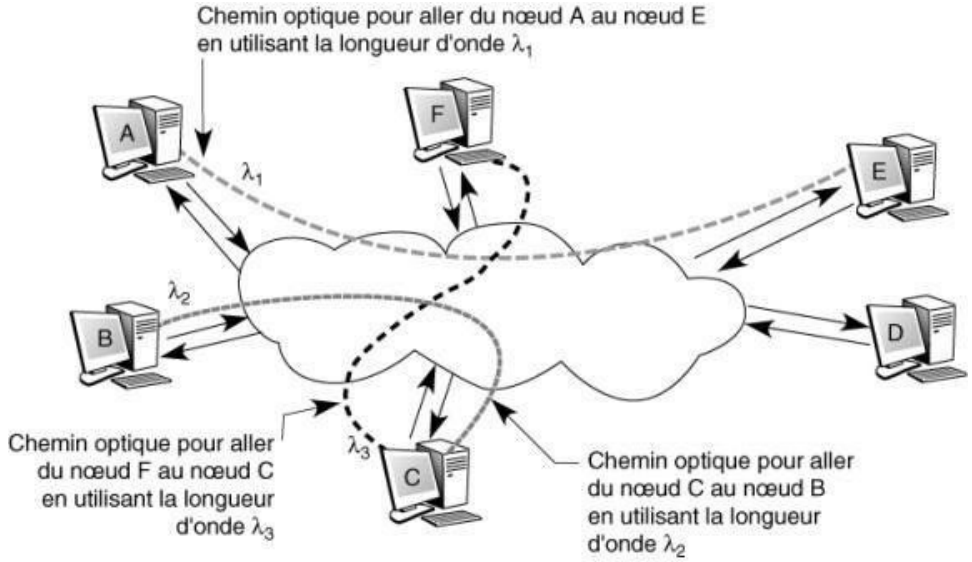


Figure 8.4

Wavelength Division Multiplexing In an optical network

On each wavelength, another level of optoelectronics can be used, either by a multiplexing in frequency, and in this case the bandwidth is again subdivided between several stations (Subcarrier Multiplexing), either by a TDM, or TDM (Time Division Multiplexing).
Optical networks to wavelength division multiplexing can be grouped into two sub-categories:

- The networks to dissemination;
- The networks to routing in wavelength.

Each of these sub-categories may be to jump (single-hop) or multiple jump (multi-hop).

The networks to dissemination

In the networks to broadcast, each receiving station receives the whole of the signals sent by the issuers. The routing of the signals is carried out in a passive manner. Each station can issue on a wavelength distinct. The receiver receives the desired signal by placing themselves on the correct wavelength. Both topologies more conventional are the star and the bus, as shown in Figures 8.5 and 8.6. In both cases, each station emits toward the center, which performs a wavelength division multiplexing of the whole of the waves which reaches it.

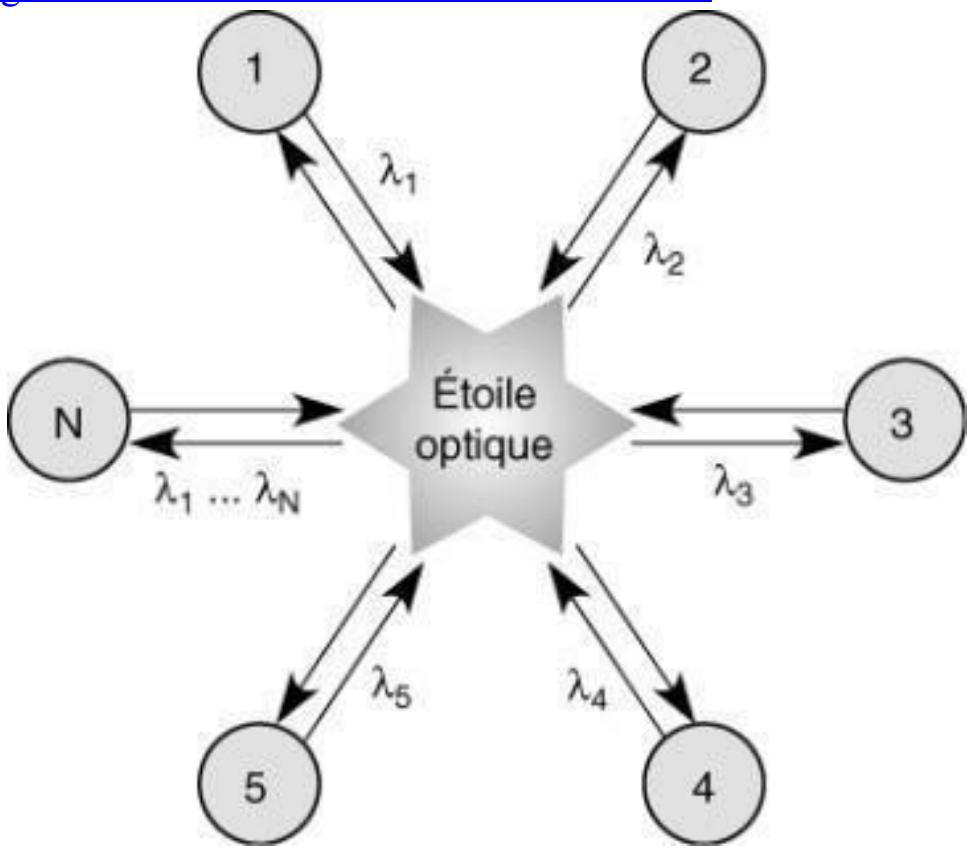


Figure 8.5
Star Topology

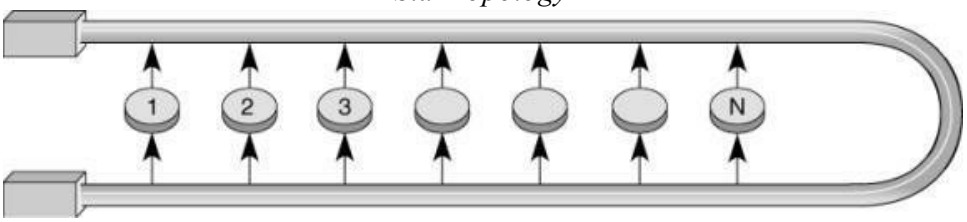


Figure 8.6
Bus topology

When the set of signals comes directly to all of the stations without ironing by electrical forms, the network is said to leap (single-hop). It is the case of the two structures illustrated in Figures 8.5 and 8.6. If it is necessary to go through the intermediate steps to perform a routing, networks to multiple jumps (multi-hop), such as those described in Figures 8.7 and 8.8, are necessary.

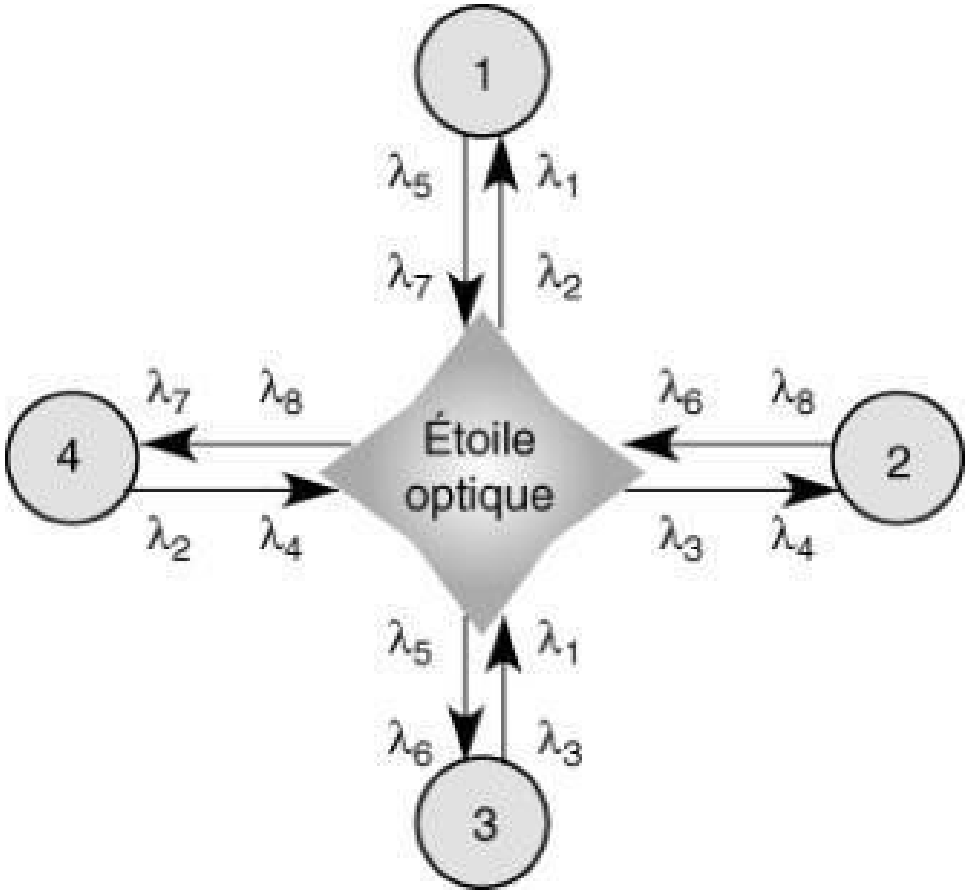


Figure 8.7

A network architecture in star to Multiple Jumps

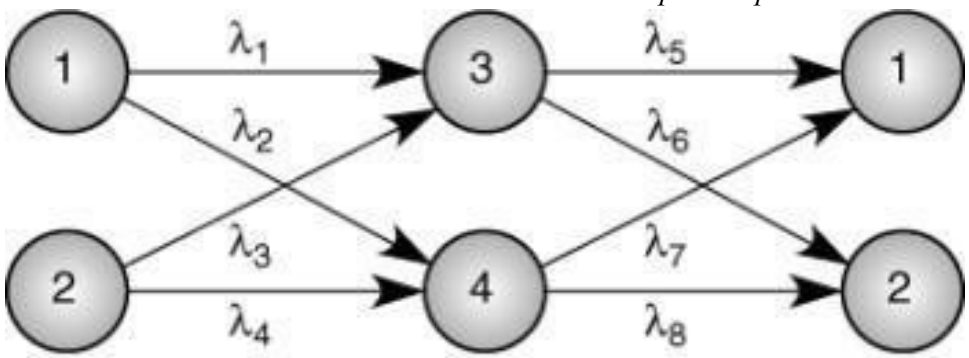


Figure 8.8

Network architecture to ShuffleNet Multiple Jumps

The Lambdanet of Bellcore is an example of a network to dissemination and single jump. The difficulty of this type of network is to dispose of wavelengths in sufficient number and receivers equipped of components capable of adapting to rapid variations of wavelength of optical signals. Taking account of this major difficulty, networks to dissemination and multiple jumps have been developed by several companies. In these networks, the transmitter and the receiver does not generally have only two wavelengths. To go to a port of entry to a port of exit, the information is routed in the form of a data packet. As the switching is performed in an intermediate node, there was a transition by an electronic component, which constitutes a delicate point to secure.

Figure 8.8 shows that, to move from node 1 to node 2, it must issue, for example, on the wavelength 2 to node 4, which rebroadcasts on the wavelength 8 to node 2, or issue on the wavelength 1 to node 3, which forwards it to the station 2 on the wavelength 6. It is seen that two paths are possible, which secures the communication process.

The networks to routing by wavelength

The idea at the basis of networks to routing in wavelength is to reuse at the maximum the same wavelengths. Figure 8.9 illustrates a node of a network to routing by wavelength in which the same

wavelengths are used on several occasions.

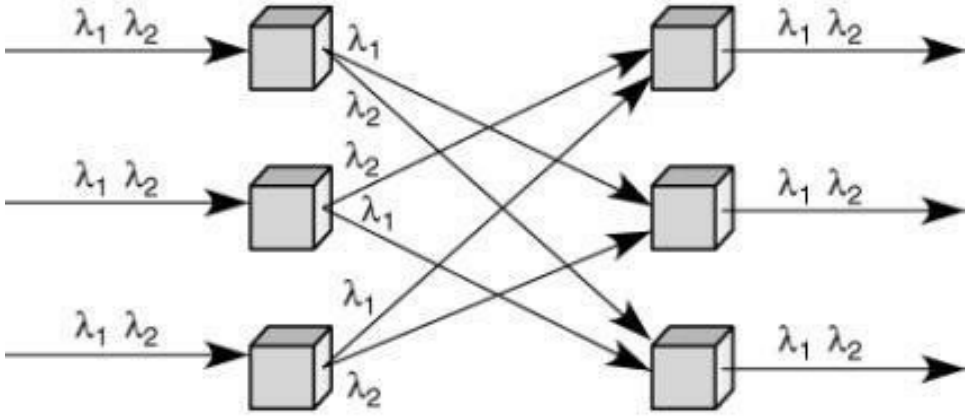


Figure 8.9
Routing by wavelength

This architecture corresponds to a fixed routing on the wavelengths. It can also develop networks to routing in wavelength with dynamic routes in the time. To this effect, it is necessary to insert the optical switches optoelectronic or, depending on the technology used, between the ports of emission and reception. An example of this technique is shown in Figure 8.10.

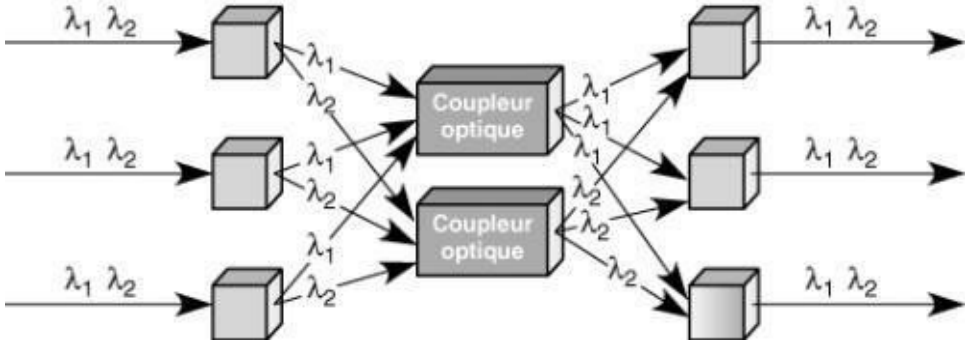


Figure 8.10
Dynamic Routing

Many research have still take place in the field of optics to optimize the use of wavelengths. This technique allows you to reach the flows is particularly high, who rely in terabits per second. The difficulties come from the costs still high in wavelength division multiplexing and especially of optical switches. When we want to minimize the cost or increase the scope, it must use switches optoelectronics. A certain fragility is then visible to each passage of a bright environment to an electrical environment. Considerable progress must still be made to réamplifier the signals of optical way and adjust almost instantaneously the couplers of emission or reception on the correct wavelength.

The optical switches

The optical switches are used to interconnect fiber-optic links between them. To optical fibers match incoming optical fibers outbound connects. If the switch uses an electrical part, the switch is said optoelectronics and not more optical only. These switches are based on the interconnection of elementary switches, i.e. of switches that have two entry doors and two exit doors, as shown in Figure 8.11. Mounted in series, these basic switches allow you to perform large switches. The design of this equipment, however, raises many problems.

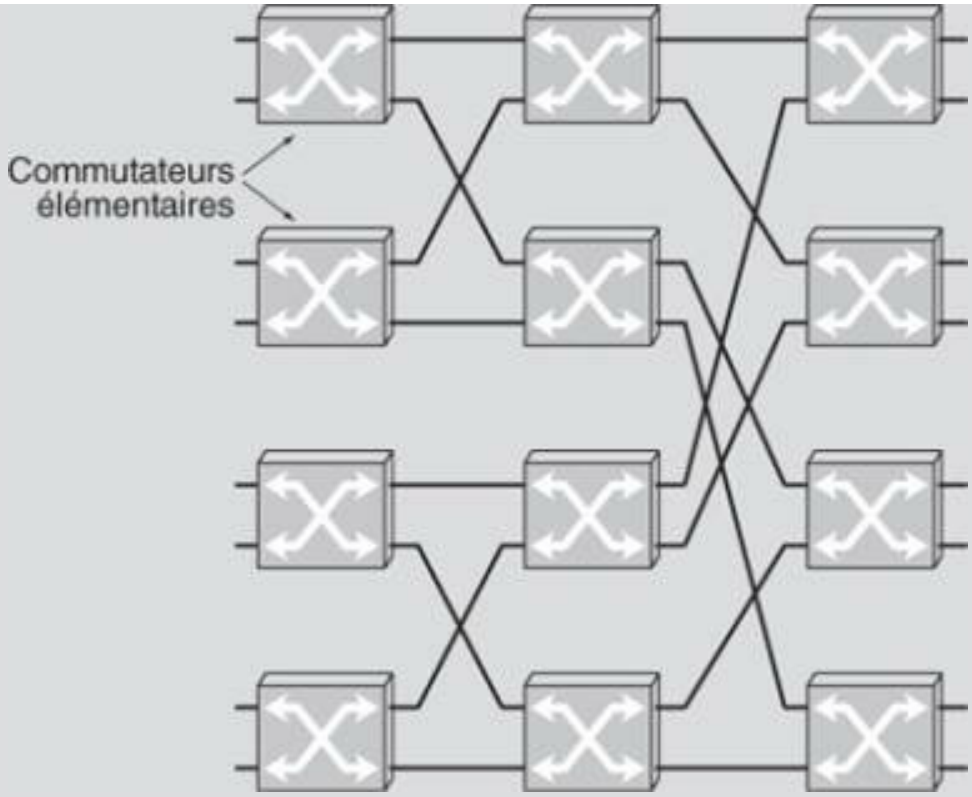


Figure 8.11

Operation of a switch optical or Optoelectronics

These switches, or Min (multistage Interconnection Network), can be of two types: either the signal is transformed into an electrical signal, either the signal is switched in optics. The first case is presented in detail in Chapter 5, dealing with the routers and switches. In the second category, it distinguishes between the following two techniques:

- Control decisions and routing electrically performed;
- All switching optical.

Circuit switching remains more easy to implement that the switching of packets, the decisions of control and routing being much more simple in this case. In a packet switching, to the arrival of each packet, it must take several decisions of control and routing, which require several hundreds of picoseconds in the best of cases. Currently, we know manage potential collisions of optical signals when a crossing of paths is necessary. The technique Optical everything today still remains experimental.

Another technology develops. Located between the circuit switching and packet switching, it comes from the switching of the *bursts* (burst-switching). This switching is to switch a set of packets issued the one behind the other, without loss of time between each packet. This issue returns to put in place a Circuit the time of peak traffic. This time can go to a fraction of a second to a few seconds. The idea is obviously to simplify the switching of packets using the equivalent of a long Packet, consisting of all the packages of a burst, but also to gain use of resources by report to a circuit switching, in which the circuit is sometimes used incorrectly.

A sensitive point of this system concerns the control of switches. How to configure them to deal with the different streams of a differentiated way, or even how to apply to each flow the priority or security which has been negotiated by the user with the operator of the network at the start of the connection? For this, a network of signs must be added to the fiber optic network. This network of signalling, said out of band, i.e. using a transportation capacity separate from the devolved to user streams, is more and more often consisting of an IP network. To each optical switch corresponds a Router IP, by which pass commands arriving in IP packets.

Signage and GMPLS

For the moment, the solution chosen by many large operators for their network and more particularly their heart of optical network is GMPLS (Generalized MultiProtocol Label Switching), which is studied in detail in chapter 11 with the MPLS protocol. The main idea at the base of this adoption is the right granularity of the duct before transporting the packets of a same communication of an entry to an output of the network and the performance of the plans for monitoring and management associated. A plan of control or management is in fact a network specialized in the transport of data of control or management. The Control Plan is also called network of signalling. The protocols associated with the signage are detailed in Chapter 23.

In the framework of the network GMPLS, of conventional protocols are used, such as the Routes

OSPF, IS-IS, and the RSVP Reservation. But these protocols are complemented by options often called TE (Traffic Engineering), which allow you to open the best possible paths and to make reservations which are calculated by the engineering of traffic.
Figure 8.12 illustrates the developments of the transport plan and the plan of control/management since the 1970s. Up to the year 2030, the Control Plan/management was centralized around a central node for become more and more distributed. GMPLS-describes a management plan to define the granularités ducts and a control plan using the IP protocols and techniques of control by policy (see Chapter 23).

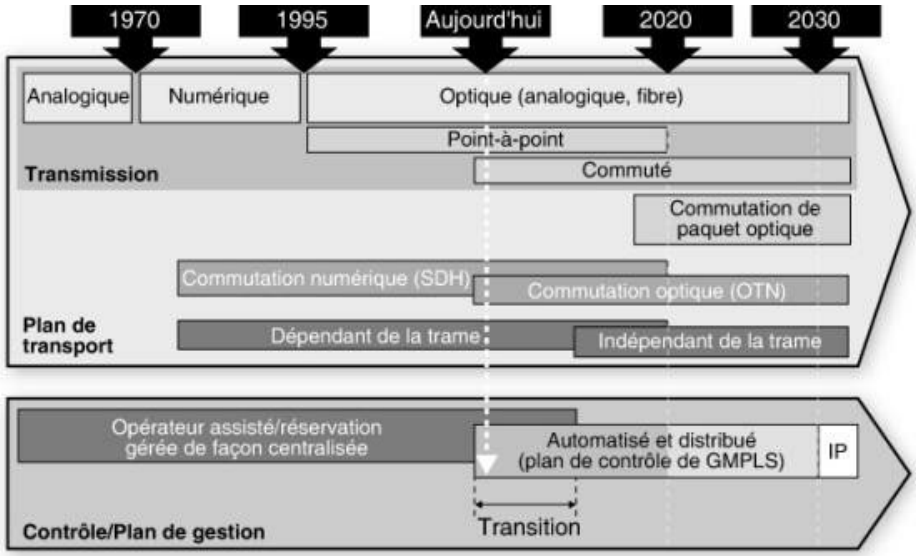


Figure 8.12
Stages of introduction of the control and management in the optical networks

GMPLS-is in fact the signaling of WSON (Wavelength Switched Optical Network), which allows you to achieve the networks at a very high speed by switching of a wavelength to another wavelength.

The interfaces of physical Levels

An interface refers to a point located between two devices. Two broad categories of interfaces are classically defined: the UNI interfaces (User Network Interface), located between the user equipment and the network, and interfaces NNI (Network Node Interface), located between two nodes in the same network or separate networks. The interface determines how the data passes through the point of passage between the two devices.
This section focuses on the interfaces of the networks of the physical level and details including the following:

- Synchronous Optical Network (SONET) and Synchronous Digital Hierarchy (SDH), defined at the outset to carry a large number of telephone communications on a same physical media.
- Packet over SONET (PoS), where the telephone communications are replaced by packets.
- G.709 OTN (Optical transport network), which determines the last generation of interfaces for access on the optical networks.
- RPR (Resilient Packet Ring), which defines a new generation of networks of the physical level and an interface, of the same type as SONET/SDH, associated with the Ethernet world.
- MPLS-TP (Transport Profile), which replaces the synchronous interfaces of the SONET type by an asynchronous interface.

Communication equipment have accompanied the mass dissemination of the phone. With the emergence of the Plesiochronous hierarchy, defining the capacity of physical lines to carry a large number of simultaneous phone calls, it was necessary to determine the techniques of access Defining the multiplexing of communications. These techniques, which are almost more used, are described in Annex F. In contrast, the Hierarchy SONET Synchronous, improved by SDH, which has the same function, but of a more modern way, is introduced in this chapter. The transport of packets is performed on this hierarchy thanks to the technology Packet over SONET (POS). This hierarchy is always very used, because SONET/SDH technology allows a reconfiguration of the network in the event of a failure.
An important standard has been defined by the ITU-T with OTN (Optical transport network), MPLS-TP (transportation Profile) and BPR (Resilient Packet Ring), for the use of a frame similar to that of Ethernet in the framework of the IEEE Standard 802.17 (see the section of the annex F dedicated to BPR networks).
The solution proposed in the framework of MPLS-TP, which proposes to replace the standard SONET Synchronous by a system completely asynchronous, but able to regain the synchronization at the end of the race, is studied at the end of the chapter.

The interfaces with the physical level

The physical level represents the first level of the hierarchy of the reference model. This level is responsible for transporting the binary elements on physical media varied. To access a media, it is necessary to use an interface to access.
Figure 8.13 illustrates the whole of the physical layer interfaces with a physical media, who is here of the optical fiber, for the transit of the IP packets. One sees on the right side of the figure that can encapsulate an IP packet in a frame ATM, or at least a fragment of the IP packet because the frame ATM is of limited size. Then, the atm frame can be issued directly on an optical fiber, if the speed is low enough for that there is no need to synchronize the clock, or encapsulated in a frame SONET/SDH prior to be routed on the optical fiber. This solution has been heavily used at the beginning of the years 2000, but today, the atm frame is replaced by the Ethernet frame (see later).
A second solution, also endangered, is to use the HDLC frame or the frame PPP (Point-to-Point Protocol) to achieve this encapsulation. To allow for the synchronization of clocks, the frames HDLC and PPP are themselves encapsulated in a frame SONET/SDH prior to be transmitted on the optical fiber.
Going to the left of the figure, one sees the technique the more classical today: the IP packet is encapsulated in an Ethernet frame, GbE (1 Gbit/s), 10GbE (10 Gbit/s) and 100GbE (100 Gbit/s). These frames are either conveyed directly on the optical fiber, is transmitted by the intermediary of SONET/SDH following the speed, the remoteness of the stations, is still using the Profile transport of MPLS-TP.

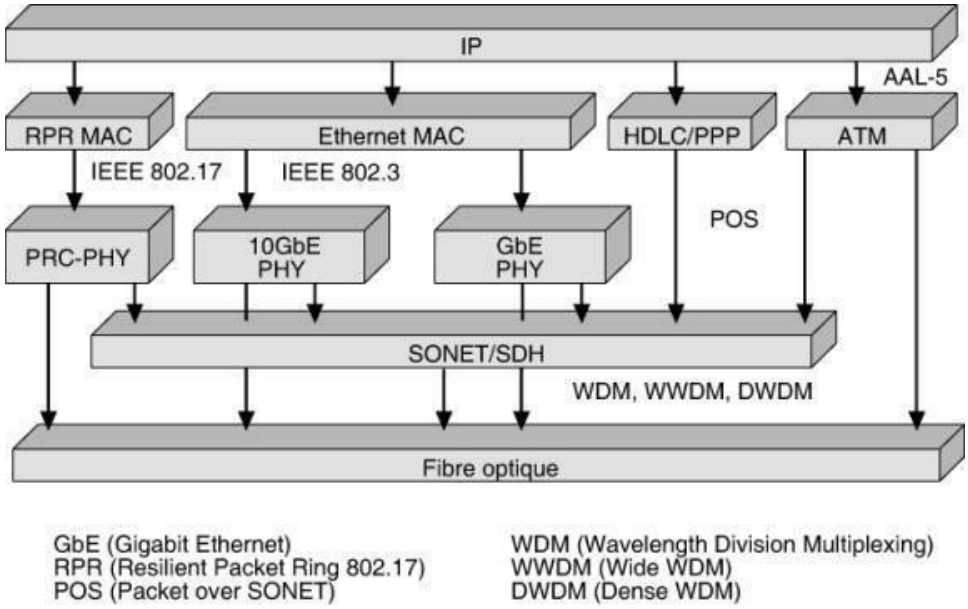


Figure 8.13

Interfaces on fiber optics

The left part of the figure shows the technique compatible with Ethernet, in which the preambles of Ethernet frames are increased to give the time to perform the synchronization of clocks. This solution, standardized under the name IEEE 802.17, is an excellent solution on the metropolitan networks. This standard is described in Annex F.

The three frames used on the physical interface are PPP, ATM, and Ethernet. When one examines the standards, we note often, even if this is not an obligation, that the physical level is itself divided into two: the level PM (physical medium), the sub-the lowest level of the physical level, which is responsible for the transport of the information on the physical support and the synchronization bit, and the level TC (Transmission convergence), a sub-Level the top of the physical level, which is responsible for the adaptation of the physical support of binary elements from the frames.

The functions of the interface to access the sub-level TC are including the following:

- Adjustment of flow rate;
- Protection of the header;
- Delimitation of the frames;
- Adaptation to the systems of transmission;
- Generation and Recovery of the frame.

The rate adaptation is to adjust the different flows of information to the bandwidth of the physical connection. For this, we can add blank frames, say of jam on a system of type SONET. Another solution is to insert bytes of jam when no transmission of frame is in progress. The number of bytes of jam depends on the time interval between two frames.

Synchronous Optical Network (SONET)

Following a proposal of Bellcore (Bell Communication Research), SONET, is a technique of transportation between two nodes, which defines the interface adopted for the NNI (Network Node Interface). It was concerned at the outset that the interconnection of telephone networks of the major operators, PTT, *carrier,* etc. All the difficulty of the standardization has been to find a compromise between the interests of Americans, Europeans and Japanese to allow for the interconnection of different networks of operators and national networks.

The Hierarchy of flows being different on the three continents, it was necessary to agree on a basic level. It is ultimately the flow of 51.84 Mbps which has been retained and which form the first level, called STS-1 (Synchronous Transport Signal, level 1). The levels located above the level 1, called

STS-n, are a multiple of the basic level.
SONET describes the composition of a Synchronous frame issued every 125 microseconds. The length of this frame depends on the speed of the interface. Its various values are summarized in Table 8.1 depending on the speed of the optical media, or OC (Optical Carrier).

OC-1	51.84 Mbit/s	OC-24	1 244,16 Mbit/s
OC-3	155.52 Mbit/s	OC-36	1 866,24 Mbit/s
OC-9	466,56 Mbit/s	OC-48	488,32 2 Mbit/s
OC-12	622,08 Mbit/s	OC-96	976,64 4 Mbit/s
OC-18	933,12 Mbit/s	OC-192	953,28 9 Mbit/s

Table 8.1 • Values of the SONET frame depending on the speed of the optical media

As shown in figure 8.14, the SONET frame includes in the first three bytes of each row of synchronization information and supervision. The cells are issued in the frame. The moment of the beginning of the sending of a cell does not necessarily correspond to the beginning of the frame, but can be located anywhere in the frame. Of the bits of supervision precede this early so that it does not lose time for the issuance of a cell.
When the signals to carry arrive in the SONET coupler, they are not copied directly such what, but included in a virtual container (virtual container). This filling is called *adaptation. The Frames SONET and SDH contain several types of virtual containers, called VC-N (Virtual Container of level N). To these containers, it must add information of supervision located in the bytes of the beginning of each row. By adding these additional information, it defines a administrative unit, or at-N (Administrative Unit-N).*
The higher levels are always nine rows, but there is *N* times 90 bytes per row for the level *N. The frame of the Level N* of the SONET hierarchy is illustrated in Figure 8.15.

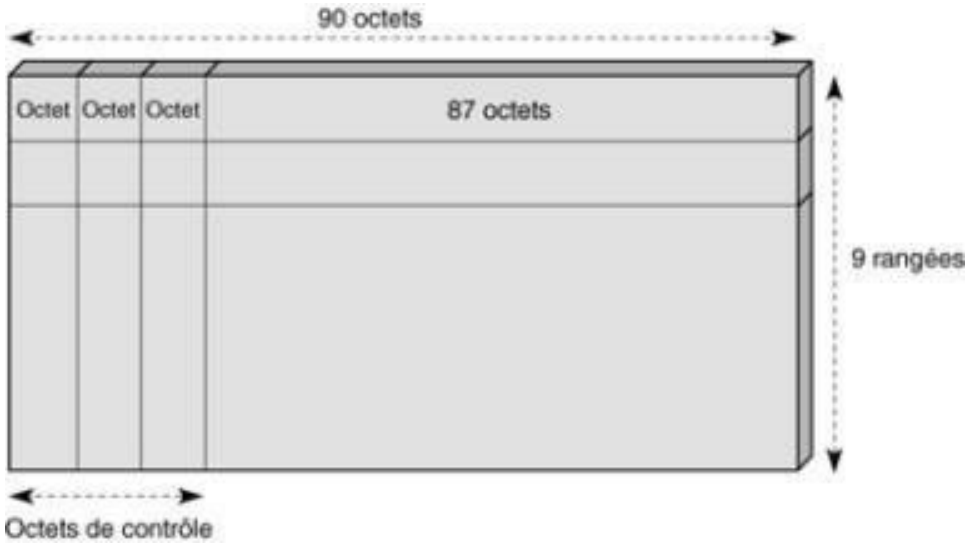


Figure 8.14
SONET frame of basis

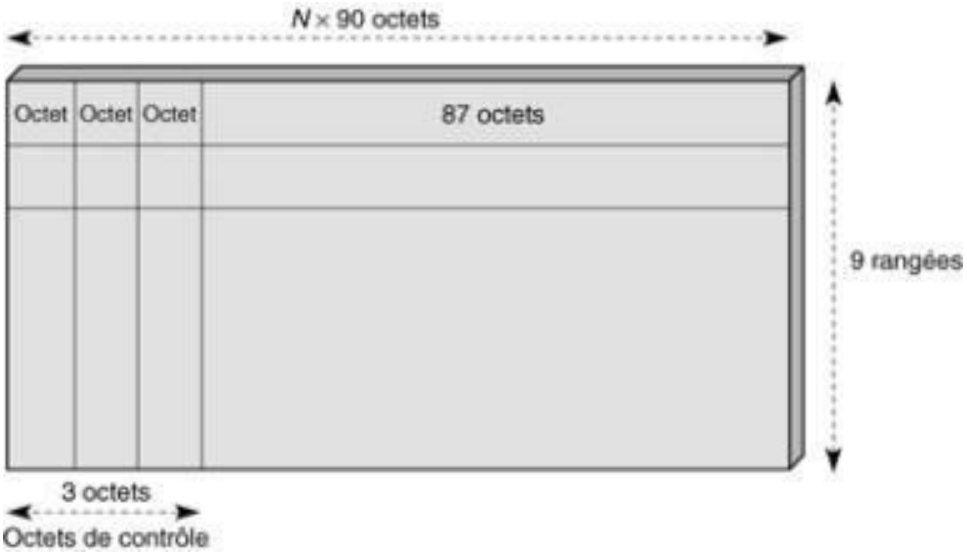


Figure 8.15
SONET frame STS-n

The SONET standard is used to carry frames (level frame), or even of the packets (level package), but, of course, encapsulated in a frame at a very high speed. The SONET frame has the flows of, respectively, 155 Mbit/s, 622 Mbit/s, 2,488 Gbit/s and 9,953 Gbit/s for the OC3, OC12, the OC48 and the OC192. This allows you to carry any type of frames to high speed, that this is ATM, Ethernet, IP, encapsulated in a frame, or any other entity.

A very important feature of SONET is to ensure reliable communication in case of breakage or failure of one of its components. The SONET networks that are found in the metropolises have a loop topology. Two paths are as well available to go from one point to another, in particular of the user to the network heart of the operator. SONET allows modification of this path in 50 milliseconds. During a break in the communication in a sense of the loop, the reconfiguration can be performed in a time causing a cut virtually undetectable to the two persons in train to talk. This ability to reconfiguration is one of the major strengths of the structures SONET/SDH. The topologies in loop of SONET/SDH are compared at the end of the Chapter to those coming from the world Ethernet.

Synchronous Digital Hierarchy (SDH)

The SDH recommendation has been standardized by the ITU-T (G.707 and G.708):

- G.707: Synchronous Digital Bit rate;
- G.708: Network Node Interface for the Synchronous Digital Hierarchy.

Found in SDH the flows to 155, 622, 2 488 Mbit/s and 9 953 Mbit/s of SONET.

The time basis always corresponds to 8 000 frames per second, each frame is composed of nine times 270 bytes. In total, this fact 155,520 Mbit/s. The structure of the frame SDH Synchronous is illustrated in Figure 8.16.

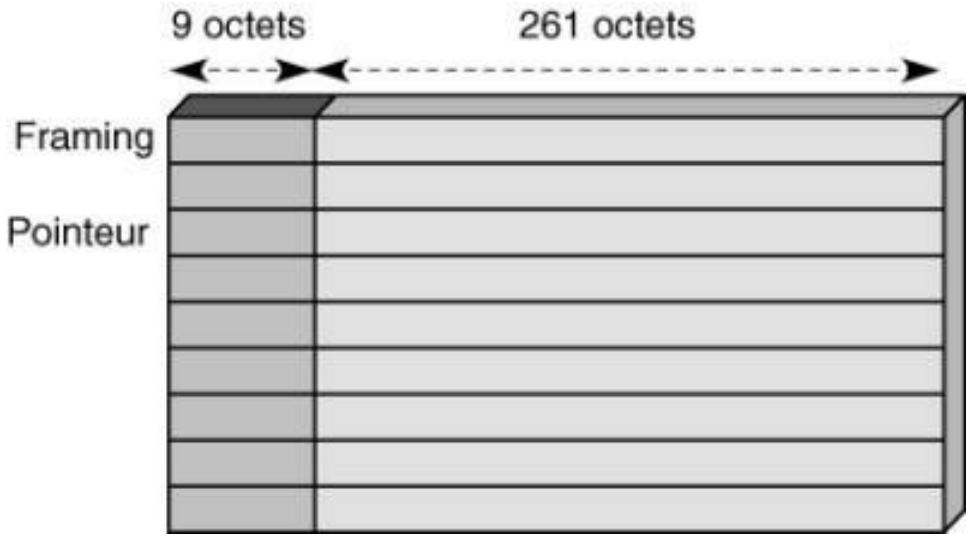


Figure 8.16
SDH frame

The information transported is indicated by a pointer that is located in the area of supervision of the frame. When the quantity of information to carry is greater than the area available in the frame SDH, it continues in the next frame, the end being indicated by a pointer to End.

Figure 8.17 shows how a stream to 140 Mbit/s can be transported on a link SDH in 155 Mbit/s (a flow rate of 140 Mbit/s represents 8 times 261 bytes plus 100 bytes, which corresponds to the shaded portion of the figure).

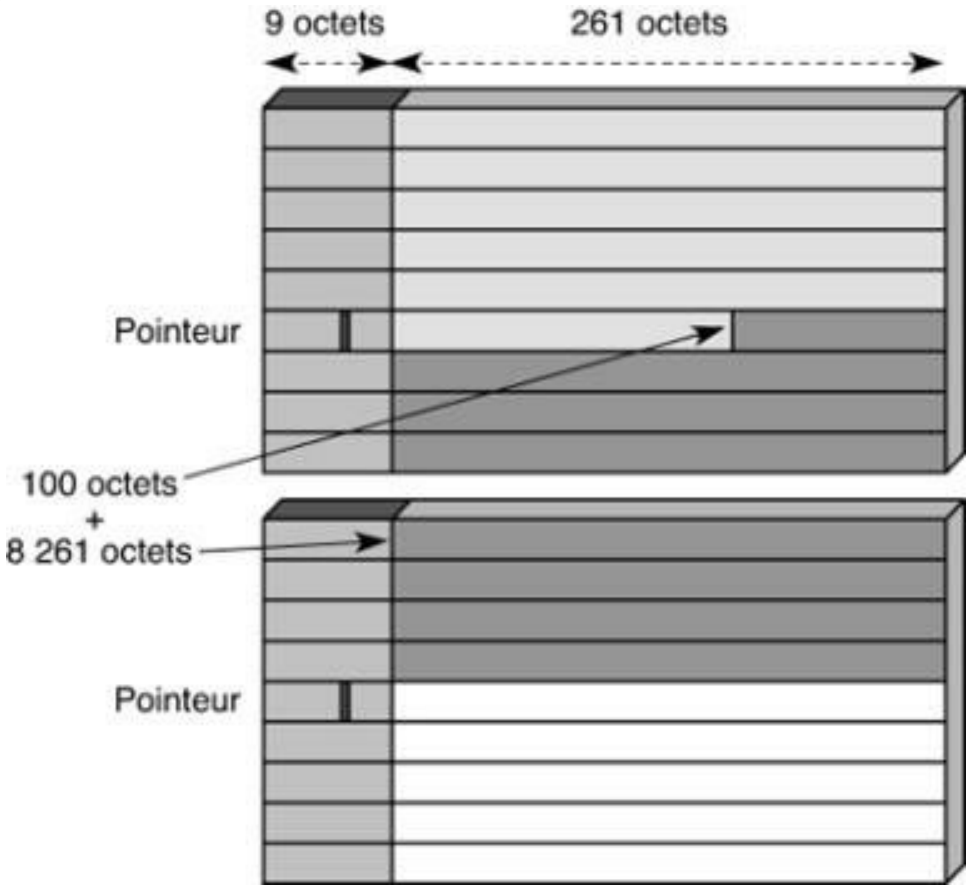


Figure 8.17
Transport of a stream SDH in 140 Mbit/s

ATM cells are transported in the SDH frame as soon as possible. It does not have to lose that a minimum of time during the transmission on each link to avoid that the variance of the response time of the cell increases too.

Figure 8.18 illustrates the transport of cells on a connection SDH basic: the cells are located anywhere, and they are reported in the area SOH or in the frame, because no time should be lost to

wait for a particular slot.

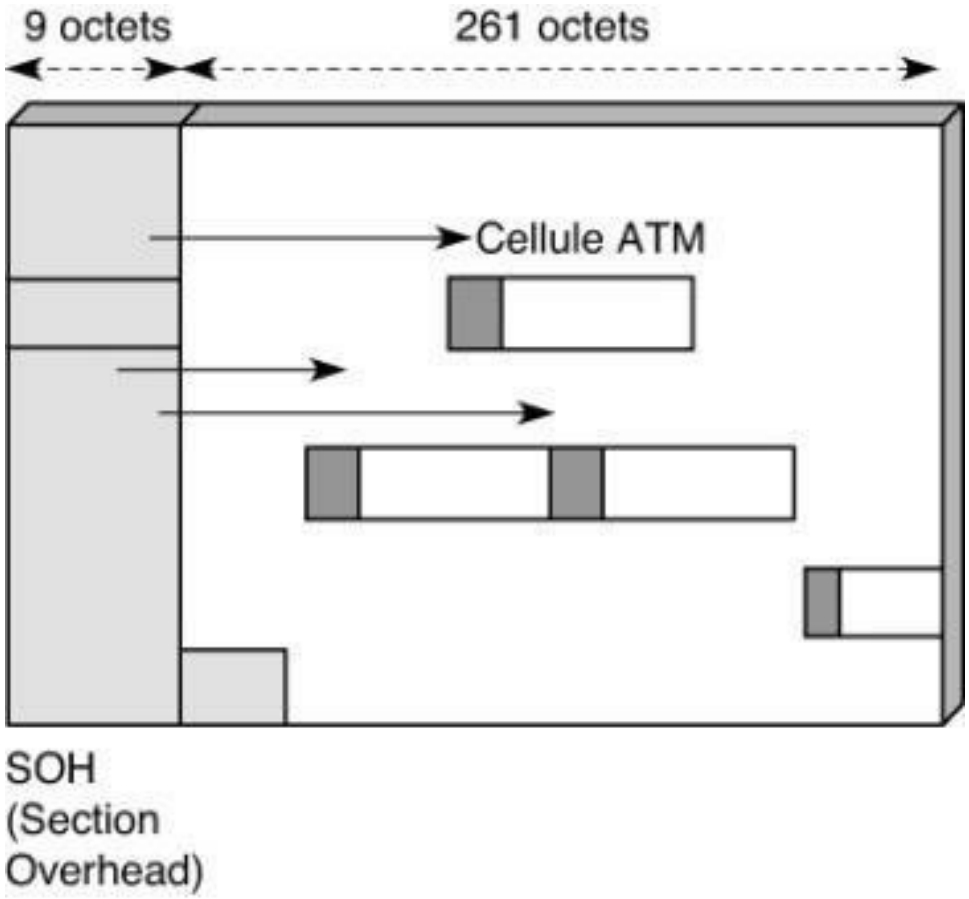


Figure 8.18
Transport of ATM cells on a link SDH

The base frame of SDH is called STM-1 (Synchronous Transport Module level 1). It is equivalent to STS-4 of the recommendation SONET. The SDH hierarchy of the ITU-T is summarized in Table 8.2.

STM-1	155.52 Mbit/s	STM-12	1 866,24 Mbit/s
STM-3	466,56 Mbit/s	STM-16	488,32 2 Mbit/s
STM-4	622,08 Mbit/s	STM-32	976,64 4 Mbit/s
STM-6	933,12 Mbit/s	STM-64	953,28 9 Mbit/s
STM-8	1 244,16 Mbit/s	STM-256	813,12 39 Mbit/s

Table 8.2 • SDH hierarchy of the ITU-T

The signals to carry come from the connections, which may be synchronous or asynchronous. For a transport more easy, are accumulates in a virtual container VC (Virtual Container), in the same way as for the recommendation SONET. As indicated previously, this packaging is called *adaptation. There are different virtual containers for each type of signal to transmit.*

The SDH connections used by the operators are five in number, corresponding to the STM-1, STM-4, STM-16, STM-64 and STM-256. The base frame is multiplied by 4 to go to the next level. This corresponds to speeds of 622 Mbit/s, 2,488 Gbit/s, 9,953 Gbit/s and 39 813 Gbit/s. The virtual containers for these levels are the VC-4, VC-16, VC-64 and VC-256. The transport of these containers on the frames STM-4 STM-16, STM-64 and STM-256 is carried out by a TDM, as shown in figure 8.19, in which 4 frames VC-4 are cut and intertwined byte by byte.

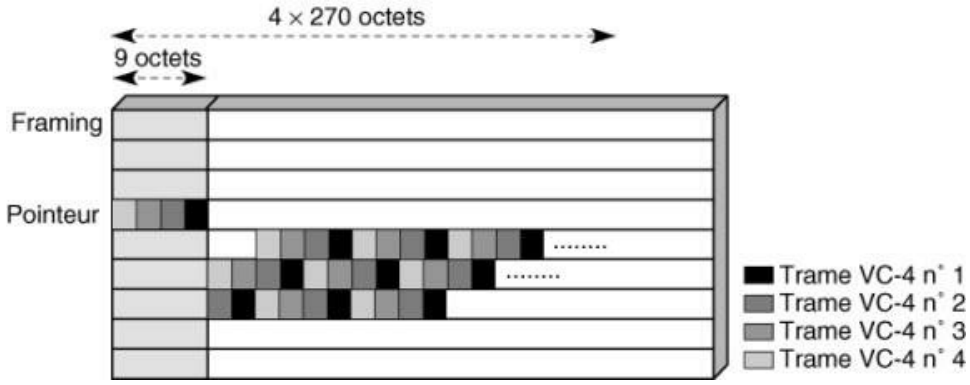


Figure 8.19

Multiplexing of containers VC-4 on a frame STM-4

As in SONET, the contents of the container with the pointers form an administrative unit, or (Administrative Unit). The administrative units are of several levels: AT-1, AT-4, at-16, at the-64 and the-256. The level STM-16 is formed from four STM-4, which are crossed on the physical media. The higher levels use the same intersecting.

In Europe, the ETSI has defined the European formats under the names of C-12, C-3 and C-4, which correspond to the values of containers. Of the intermediary formats, called TU (tributary Unit) and Tug (tributary Unit Groups), complement the hierarchy. This hierarchy somewhat complex is shown in Figure 8.20.

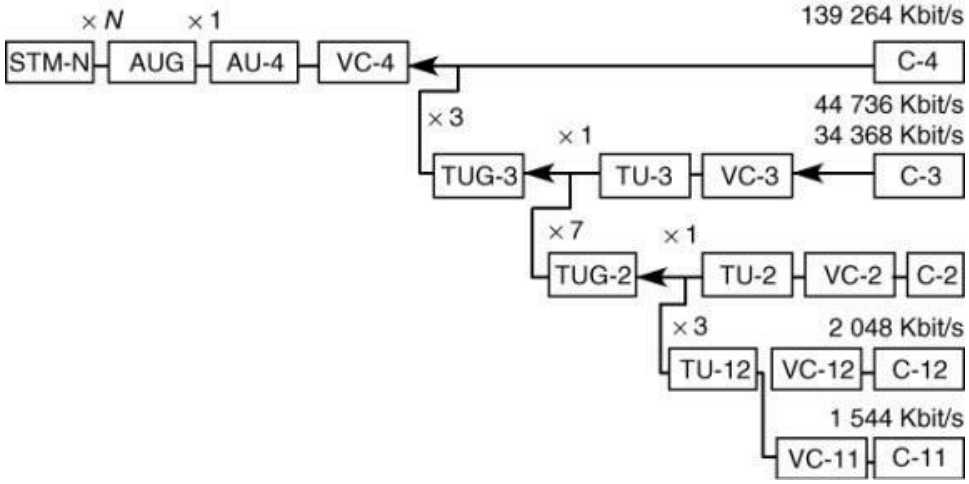


Figure 8.20

SDH hierarchy of the ETSI

Packet over SONET (PoS) and EOS (Ethernet over SONET)

As indicated previously, the SONET interface has been chosen at the outset by the ITU-T to interconnect the telephone networks. This interface determines a time of 125 microseconds between the issuance of two frames. The SONET frame is a kind of wagon, that can be completed by bytes or frames and more generally by containers. The general technique of transport of packets over SONET/SDH is called Pos. It is today widely used to route directly to high speed packets of any type on a support SONET/SDH.

The IP Interface over SONET is expected to call ip/PPP-HDLC over SONET. To format the flow of IP packets into frames, it uses PPP, which provides an encapsulation protocol, a control of error and a protocol for the opening of the connection. The PPP frames can be replaced by HDLC frames following the RFC 1662. PPP is described in RFC 1661.

Figure 8.21 illustrates the structure of the IP frame over SONET encapsulating the IP packet.

Figure 8.21

Format of the IP frame over SONET

Routers or switches must be capable of processing the flow rates permitted by SONET/SDH (155 Mbit/s, 622 Mbit/s, 2.5 Gbit/s, 10 Gbit/s, 40 Gbit/s and soon 160 Gbit/s). As IP packets are more and more often encapsulated in an Ethernet frame, SONET, is used for the transport of this type of frame to the inside of its area of data. It obtains the Ethernet technology over SONET (EOS). The advantage of this solution is to allow a synchronization of Ethernet frames. Convey the floor under Telephone IP and encapsulate it in SONET becomes very easy on long distances. The encapsulation of Ethernet frames is carried out by a method of encapsulation of blocks called Generic framing procedure which allows streams flow down to position themselves to the good places to the inside of the frame SONET.

SONET is also used for the transport of the 10 Gbit/s Ethernet (10GbE) by encapsulating the Ethernet frames to achieve long distances. This is the Ethernet switched.

The OTN interface (Optical transport network)

The standard SONET/SDH was introduced to carry the floor on the phone and it took many adaptations for the transport of frames and packets of type IP, ATM, Ethernet, or other. The Successor of SONET/SDH was standardized in early 2002 by the ITU-T under the name of OTN (Optical transport network). Its role is to pass packets on the connections at 2.5, 10, 40 and 160 Gbit/s. The corresponding Recommendation bears the number G.709.

Figure 8.22 illustrates the new format of the synchronous frame OTN.

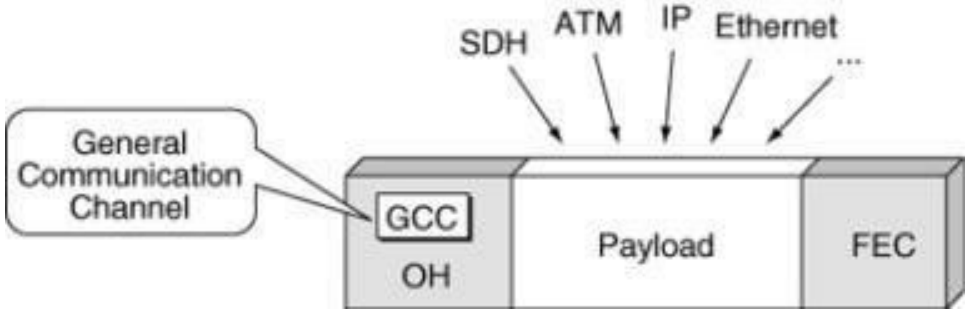


Figure 8.22

Format of the OTN frame

All types of frames must be able to be transported in a transparent manner in the frame OTN, without that they need to be changed. A field is intended to add a FEC (Forward Error Correction) in order to make the necessary corrections to achieve an error rate determined.

The OTN interface consists of several levels. Starting from the optical fiber, we find the following layers:

- OTS (Optical Transmission Section), which supports the transmission of the optical signal by checking its integrity.
- Who (Optical Multiplex section), which supports the capabilities to perform a wavelength division multiplexing.
- OCh (optical channel), which is the level of end-to-end of the optical signal. This level allows the modification of the connection and the rerouting, as well as the functions of maintenance of the connection.
- DW (Digital wrapper), which corresponds to the digital envelope.

The level Digital wrapper is itself decomposed into three sub-levels:

- OTUk (Optical Transport Unit), which gives the possibility to adopt a correction using a CEE.
- ODUk (Optical Data Unit), which manages connectivity independently of customers and provides a protection and a management of this connectivity.
- OPUk (optical payload Unit), which indicates a correspondence between the

signal and the type of customer.

The overall architecture of OTN is illustrated in Figure 8.23.

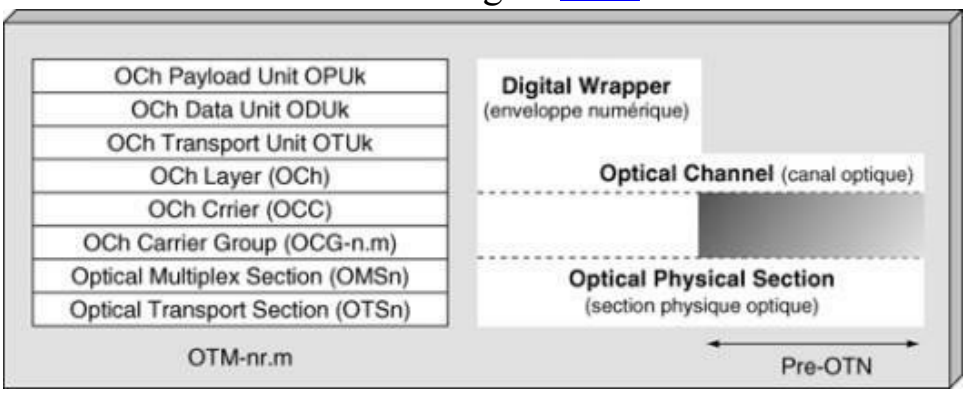


Figure 8.23

Layered structure of the OTN architecture

Figure 8.24 illustrates the different entities of transport on the interface OTH (Optical Transport Hierachy) and Figure 8.25 the structure of the frame and the flow rates of the interface OTN.

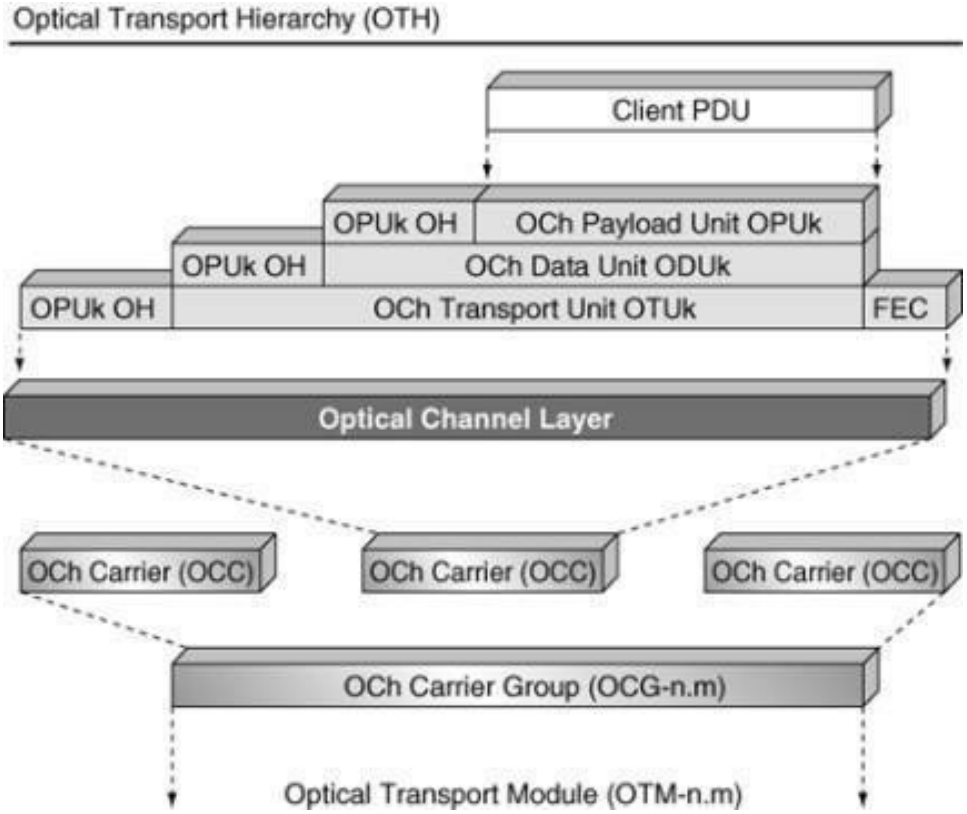
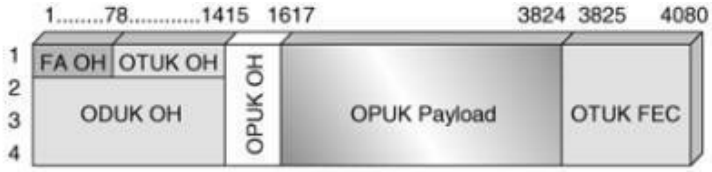


Figure 8.24

Entities of transport of the hierarchy OTH



OPU			ODU			OTU		
k = 1	k = 2	k = 3	k = 1	k = 2	k = 3	k = 1	k = 2	k = 3
48,971	12,191	3,035	48,971	12,191	3,035	48,971	48,971	3,035
20,420	82,028	329,489	20,420	82,028	329,489	20,420	20,420	329,489
2 488,320	9 995,277	40 150,519	2 498,775	10 037,274	40 319,219	2 666 057	10 709,225	43 018,413

OPU (Optical Packet Unit) ODU (Optical Data Unit) OTU (Optical Transport Unit)

Figure 8.25

Structure of the frame and flow rates of the interface OTN

MPLS-TP

This section is concerned with the part physical interface to MPLS-TP (Transport Profile). It is to be noted that the word "transport" refers not to the message level (layer 4, or transport) of the reference model, but the transport of binary elements on the physical level (layer 1). However, MPLS-TP brings together all of the specifications to achieve an MPLS network, from signaling to the switching in passing by the management system (see the Chapter 11, dedicated to MPLS).
The interface to which is interested in this section replaces the interface SONET synchronized by an asynchronous interface of type package. The aim is to reduce costs by replacing a synchronous interface quite complex by an asynchronous interface much more simple. The difficulty is to recover the synchronization to the end of the network. MPLS-TP also adds the techniques of redundancy in order to offer a high availability and, as SONET, a reconfiguration in a time of not impeding the floor the phone.
Among the possible duplication include the techniques 1:1, 1+1, N:1, N,M,O:

- 1:1: A path must be rescued by a second path.
- 1+1: The flow has opened two paths, with the possibility, in the event of the failure of a path, to continue to provide an acceptable quality of service while using only a single path.
- N:1: *N* paths have in common a path of relief.
- N+1: *N* + 1 paths are opened by the flow, with the possibility that if one of these *N* + 1 waves fails the quality of service remains acceptable in only using the *N* roads;
- N,M,O: *N* paths are rescued by *M* paths, which, themselves, are rescued by *O* paths. Many opportunities are offered by allowing or not to circulate the waves on the paths of relief that are less priority.

Conclusion

Optical networks respond well to requests to increase the flows of users. The use of a large number of wavelengths, at speeds of up to 40, 100 and then 160 Gbit/s, allows you to easily meet the current demand. The growth potential of the flows should allow to easily follow the request. However, the overcapacity, which will fundamentally been used between the years 2000 and 2010, is no longer of release since the arrival in 2010 of control methods reactive to establish communications.
All cores of network use of the optical fiber. There are still significant progress to be made to arrive at a network any optical, in which the signals in the form of light would be transported of end-to-end in the form of packets.
The interfaces of access to networks for the transport of data allow to determine the performance of the network. The frame used by the interface has a direct implication on both the terminal equipment and on the network. The dominant interface for high throughput is for many years SONET/SDH, which brings to the time of the transmission speeds important and a securing the interface by the ability to reconfigure the SONET loops in less than 50 milliseconds. However, the cost of this interface is high. The fact that it is not directly associated with a frame structure of level 2, it must in fact be encapsulate this frame of Level 2 in the SONET frame.
The MPLS technology-TP allows to reduce the cost of the interface using conventional techniques, already widely used in local networks. Another solution, represented by OTN, aims to find a universal interface to access a network of optical fiber. The problem of the single interface is however far from resolved, and it chooses today rather the interface depending on the type of terminal equipment to connect to and the network to cross.

9

The Ethernet networks

The Ethernet frame has become the standard of the frame layer, which is almost used everywhere. Frame Relay and ATM switching have virtually disappeared. This is the reason for which the technologies of Frame Relay and ATM are compiled in Appendix G.

The shared modes and dial-up

Ethernet operates according to two very different modes, but totally compatible, the shared mode and the dial-up mode, which allow all two of the transport of Ethernet frames. Since 2012, we can add a mode "routed", developed for the New Technologies of level 2 in order to interconnect data centers between them or for the communications to the inside of the datacenters.
The shared mode indicates that the physical media is shared between the devices equipped with Ethernet cards. In this mode, two stations which emit at the same time would see their signals enter in collision. In the Dial-up mode, the devices are connected to a switch, and there can be no collision since the device is the only one on the link is connected to the switch. The switch emits toward the station on the same link, but in full-duplex, i.e. in parallel, but in the other direction. Finally, the routed mode affects the Ethernet frames for which the network uses the Medium Access Control (MAC) address as an Ethernet address and not as a reference. Figure 9.1 illustrates the two modes, shared and switched, with five terminal stations.

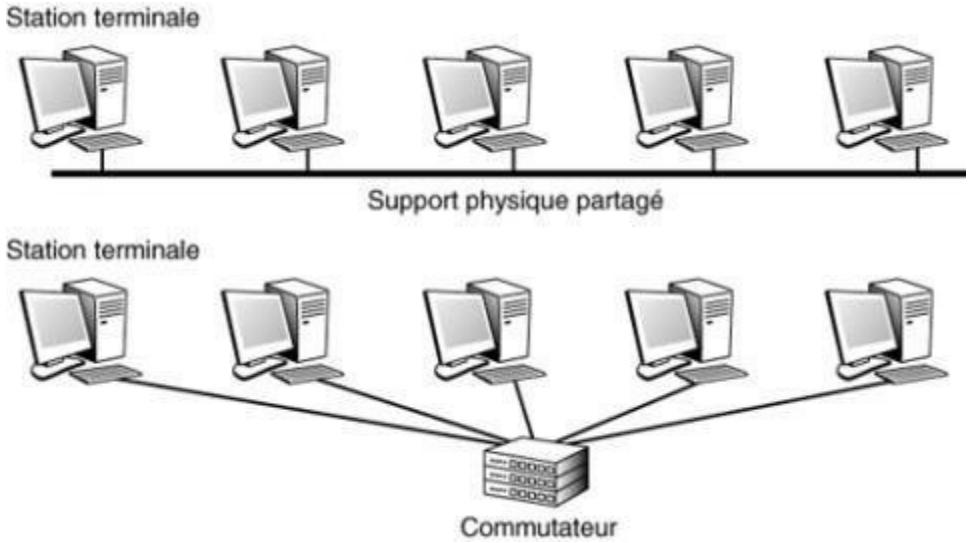


Figure 9.1
Comparison of the shared techniques and switched

The main advantages and disadvantages of the two modes are the following:

- There is no collision switched mode, but the frames must be stored in the switches, which requests a flow control.
- To connect a station in switching, there must be two couplers and a switch, while

to connect a station in shared mode, a single coupler is sufficient. The shared technical is therefore less expensive to implement.

- The technique allows switched of connections without constraint of distance, while the shared method is accompanied by a strong limitation of the distance to resolve the problem of the sharing of physical media.
- The routed mode is very similar to the Dial-up mode, but using a routing table in the place of a switching table.

The Ethernet networks shared

The Ethernet networks shared implement a technique for access to the physical media standardized by the Working Group IEEE 802.3 under the name of MAC access (Medium Access Control). MAC uses a general technique called random access.

Born of research carried out at the beginning of the 1970s on the techniques of random access, the IEEE 802.3 standard, which then gave birth to the standard ISO 8802.3, described the technique of access to a local Ethernet network shared. It is the society which Xerox has developed the first prototypes.

We can characterize the Ethernet networks shared by the technique of CSMA/CD access, whose flow varies from 1 to 10, 100 and 1 000 Mbit/s. Beyond, at a speed of 10 000 Mbit/s, only the switched solution is acceptable for reasons of distance (see later).

In the facts, the number of Ethernet networks standardized shared is impressive. The box below lists using the nomenclature IEEE.

The Ethernet networks standardized shared

The working group indicates the technique used: IEEE 802.3 corresponds to CSMA/CD; IEEE 802.3 Fast Ethernet to an extension of CSMA/CD, IEEE 802.9 to an interface CSMA/CD to which is added the B channels; IEEE 802.11 to an Ethernet by Hertzian waves, etc. Then come the speed then the modulation or not (Base = basic band and broad = broadband) and finally a complementary element, which, to the origin, was the length of a strand and is transformed into the type of physical media:

- IEEE 802.3 10BASE5 (shielded coaxial cable yellow);
- IEEE 802.3 10BASE2 (Cheapernet, coaxial cable Non Shielded Brown, thin Ethernet);
- IEEE 802.3 10Broad36 (Ethernet broadband, coaxial cable CATV);
- IEEE 802.3 1BASE5 (Starlan to 1 Mbit/s);
- IEEE 802.3 10BaseT, Twisted-Pair (twisted wire pairs);
- IEEE 802.3 10BaseF, Fiber Optic (optical fiber):
- 10BaseFL, Fiber Link;
- 10BASEFB, fiber backbone;
- 10BASEFP, Fiber passive;
- IEEE 802.3 100BaseT, Twisted-Pair or Fast Ethernet (100 Mbit/s In CSMA/CD):
- 100BaseTX;
- 100BaseT4;
- 100BaseFX;
- IEEE 802.3 1000BaseCX (two twisted pairs of 150 Ω);
- IEEE 802.3 1000BaseLX (pair of optical fiber with a wavelength high);
- IEEE 802.3 1000BaseSX (pair of optical fiber with a short wavelength);
- IEEE 802.3 1000BaseT (four pairs of Category 5 UTP);
- IEEE 802.9 10BaseM (multimedia);
- IEEE 802.11 10BaseX (terrestrial).
- The IEEE 802.12 standard defines the local network 100VG AnyLAN, which is compatible with Ethernet. The compatibility corresponds to the use of the same frame structure that in Ethernet. The technique of access, in contrast, is not compatible with the CSMA/CD, as indicated at the end of the chapter.

Characteristics

The characteristics of Ethernet networks shared are described in the ISO Standard 8802.3 10BASE5.

The topology of a network Ethernet includes strands of 500 meters to the maximum, interconnected to each other by repeaters. These repeaters are active elements which retrieve a signal and the retransmit after regeneration. The connections to the computer hardware can perform all the 2.5 meters, which allows up to 200 connections per strand. In many products, the specifications indicate that the signal should never cross more than two repeaters and only one may be far away. The regeneration of the signal is performed once crossed a line with a range of 1 000 meters. The maximum length is 2.5 kilometers, corresponding to the three strands of 500 meters and a remote repeater (see figure 9.2). This limitation of the distance 2.5 kilometers is however not a characteristic of the standard, and can be overcome these constraints of three repeaters and reach a total distance of the order of 5 kilometers.

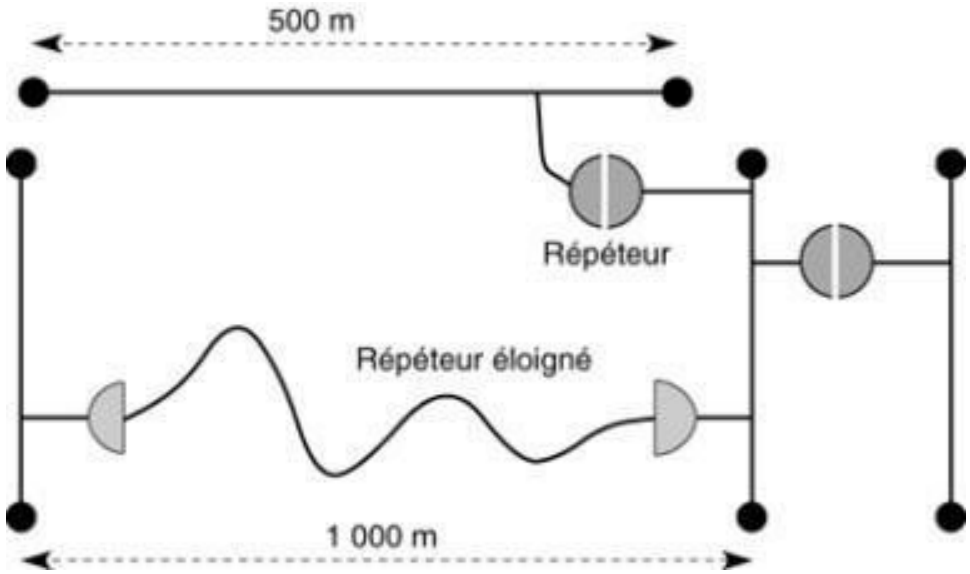


Figure 9.2
Topology of Ethernet

The only constraint to take into account is the maximum time that elapses between the emission and reception of the signal in the Coupler more distant. This time should not exceed a value of 25.6 microseconds. In effect, during a collision, the time before rebroadcasting is a multiple of 51.2 microseconds. To avoid a new collision between two frames reissued on slices of different time, it must elapse at maximum 51.2 microseconds between the time of the issuance and that of listening to the collision. The time to go is to the maximum of 25.6 microseconds, if the collision is performed just before the arrival of the remote signal. It should also be 25.6 microseconds to reassemble the collision up to the initial station (see figure 9.3). In addition, the length of a frame must be at least equal to the roundtrip time so that the issuer can save a collision. This minimum length is 64 bytes. We well 51.2 microseconds of minimum time of propagation in noting that 64 bytes equivalent to 512 bits, which, at the speed of 10 Mbit/s, require a time of issuance of 51.2 µs.

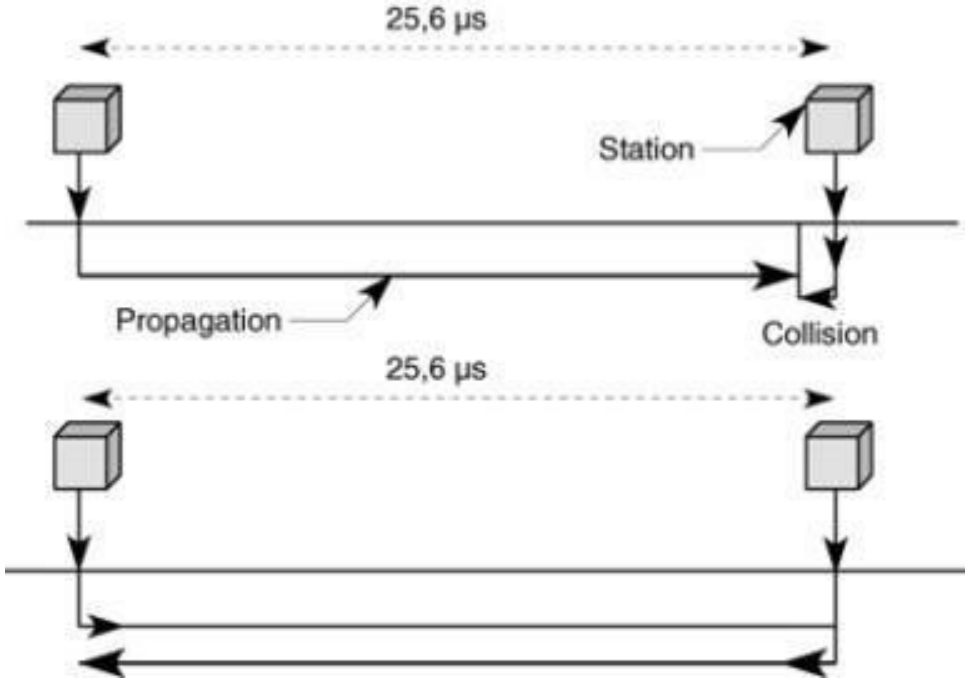


Figure 9.3
Maximum time between issuance and receipt of a collision

Any network for which the roundtrip time is less than 51.2 microseconds is likely to use the IEEE 802.3 standard. The speed of propagation on a coaxial cable being approximately 200,000 km/s, the maximum range on a same cable is roughly 5 kilometers. In the basic topology, a large part of the propagation time is lost in the repeaters. To achieve distances greater than 4 kilometers, some wiring use of stars passive optical, which allow to broadcast the signal to several strands Ethernet without loss of time. In this case, the loss of energy on the optical star that can reach several decibels, it is not possible to issue more than two or three in series. It then gets the topology shown in Figure 9.4.

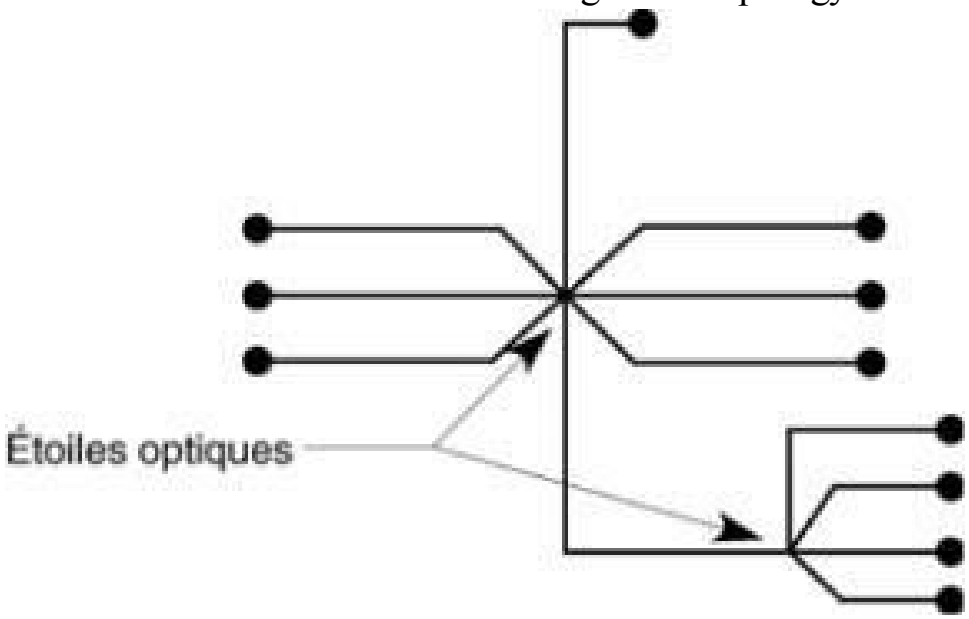


Figure 9.4
Ethernet topology with optical stars

The technique of access to the medium of transmission chosen by Ethernet is the random access with detection of the carrier of the persistent type. If the module of the emission-receipt detects the presence of other emissions on the cable, the Ethernet coupler may not issue of frame.

If a collision occurs, the module of the emission-receipt emits a signal to interrupt the collision and initialize the procedure of retransmission. The interruption of the collision occurs after the sending of a binary sequence, called sequence of Jam (jam), which verifies that the duration of the collision is sufficient to be noticed by all stations in transmission involved in the collision.

It is necessary to define several parameters to explain the recovery procedure on a collision. The time go-maximum return corresponds to the time that elapses between the two farthest points on the local network, from the issuance of a frame until the return of a signal of a collision. This value is 51.2 microseconds or 512 time to issuance of a bit, or even 512 elementary time. The sequence of jam lasts 48 elementary time. Ethernet specifies still a "slice of time," which is the minimum time before retransmission (51.2 microseconds). The time before retransmission also depends on the number *n* of collisions already carried out. The random delay of retransmission in Ethernet is a multiple of the time slice $r \times 51.2$ microseconds, where r is a random number such as $0 \leq r < 2k$, where $k = \min(n,10)$ and n is the number of collisions already carried out. If, at the end of 16 tests, the frame is still in the collision, the transmitter abandons its transmission. A recovery is performed from the protocols of the higher levels.
When the two frames will collide for the first time, they have a chance on two to enter again in collision: $r = 1$ or 0. Although the algorithm of retransmission, or back-off, does not seem optimal, c is the technique which gives the best results, because it is better to try and fill the transmission medium rather than wait time is too long and lose in flow.
A simple calculation shows that the retransmission time, after a dozen successive collisions, represent only a few milliseconds, i.e. a time still very short. CSMA/CD being a probabilistic technical, it is difficult to identify the time that elapses between the arrival of the frame in the coupler of the transmitter and the departure of the frame of the coupler receiver up to the recipient. This time depends on the number of collisions as well as, indirectly, of the number of stations, the network load and the average distance between two stations. The more the propagation time is important, the more the risk of collision increases.
All calculations reported here refer to an Ethernet network to 10 Mbit/s. If we increase the speed of the network by multiplying by ten its flow (100 Mbit/s), the maximum distance between the two stations the more remote areas is also divided by 10, and so on, so that we get in keeping the same minimum length of the frame:

- 10 Mbit/s → 5 kilometers
- 100 Mbit/s → 500 meters
- 1 Gbit/s → 50 meters
- 10 Gbit/s → 5 meters

These distances shall be understood without the existence of repeaters or hubs, which are asking for a certain time of crossing and reduce the extent of the maximum distance. To counter this problem, two solutions can be implemented: Increase the size of the Ethernet frame or to go to the switching. The Ethernet network 1 Gbit/s uses a minimum frame of 512 bytes, which allows him to return to a maximum distance of 400 meters. The network to 10 Gbit/s only uses the switching.

Performance of a network Ethernet 10 Mbit/s

Many of the performance curves show the actual flow rate in function of the flow rate offered, that is to say the flow from the new frames plus the flow rate caused by the retransmissions. Figure 9.5 illustrates the actual flow rate in function of the flow rate offered.

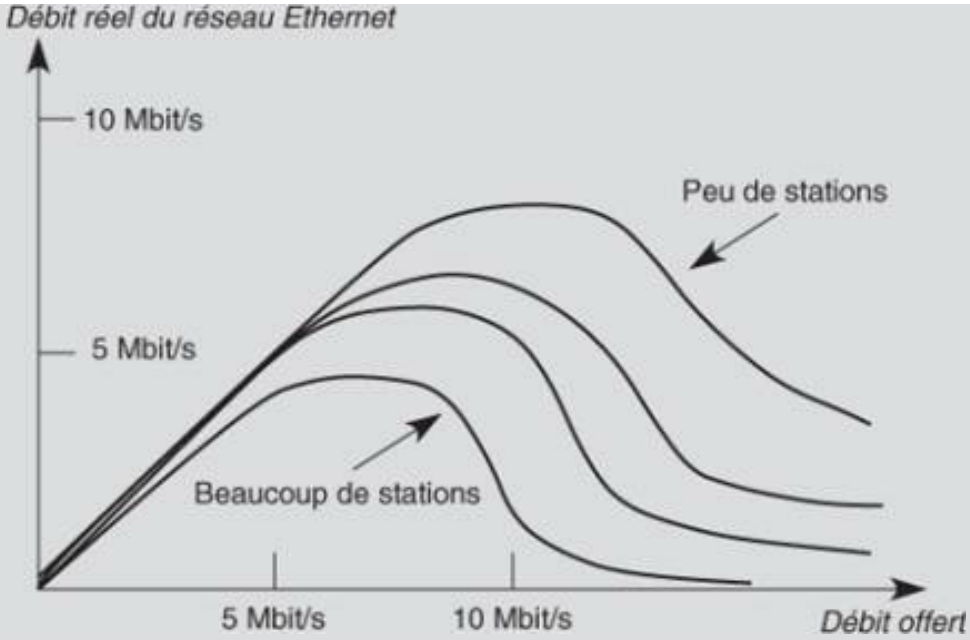


Figure 9.5

Performance of the Ethernet network

One can show that a malfunction occurs as soon as the flow offered exceeds a certain limit, due to collisions of more and more numerous. To avoid this type of problem on an Ethernet network, it is necessary that the instantaneous flow is less than 5 Mbit/s. To obtain this maximum value, we limit the number of stations to several hundreds of PC or to several tens of workstations. This are the figures cited most often.

To resolve the problems of flow, a solution is to achieve a set of Ethernet networks interconnected by gateways. The difference of the repeater, the Gateway is a body intelligent, able to select the frames that must be issued to the following networks. The gateway must manage a addressing. It limit of this fact the flow on each network. One obtains in this case of the topologies without constraint of distance. The gateway can be of different types. It is a bridge when it was only a problem of physical address to resolve. We then speak of filtering bridge to filter the frames that pass. The frames destined to a station located on the same network are stopped, while the other frames are reissued to the next network.

The example treaty on 10 Mbit/s is also valid for the whole of Ethernet technology to more high flow. The solution for do not fall on the saturation problems is of course to go to the upper range as soon as necessary. As, beyond 1 Gbit/s, control the flows becomes complex, Ethernet networks to 10 Gbit/s and 100 Gb/s no longer exist in shared: they only use the switching.

The Random Access

The random access, which is to issue to a moment completely random, relies on the method Aloha. This last takes its name of an experiment performed on a network linking the various islands of the Hawaiian archipelago at the beginning of the 1970s. In this method, when a coupler has information to transmit, it sends it, without worrying about the other users. If there is a collision, i.e. overlay of signals of two or several users, the signals become indecipherable and are lost. They are retransmitted later, as shown in figure 9.6, on which the couplers 1, 2 and 3 come into collision. The coupler 1 retransmits its frame in the first because it has drawn the most small timer. Then, the coupler 2 issues, and its signals collide with the coupler 1. All two withdraw a random time of retransmission. The coupler 3 has just listen to while the couplers 1 and 2 are silent, so that the frame of the coupler 3 passes with success. The Aloha technique is at the origin of all the methods of random access.

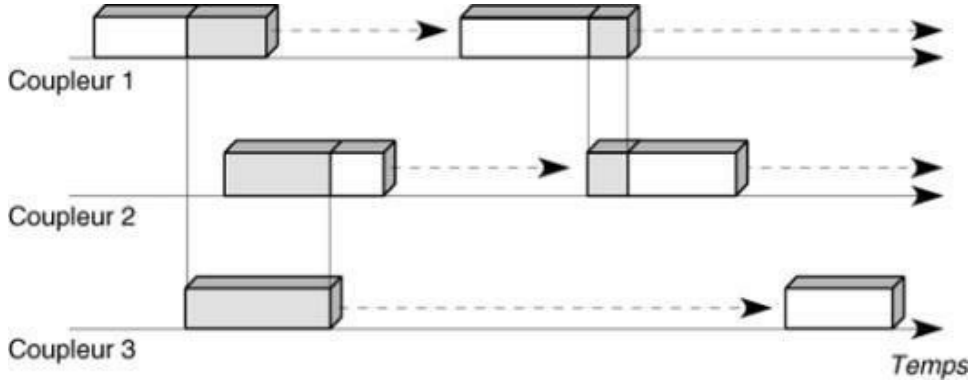


Figure 9.6
Principle of operation of the pure Aloha

In addition to its extreme simplicity, the aloha has the advantage of not requiring any synchronization and be completely decentralized. Its main disadvantage lies in the loss of information resulting from a collision and in its lack of effectiveness, since the transmission of frames in collision is not interrupted.

The flow of such a system becomes very low as soon as the number of couplers increases. It can be shown mathematically that if the number of stations tends toward the infinite, the flow becomes zero. From a certain point in time, the system is more stable. In order to decrease the likelihood of conflict between users, various improvements of this technique have been proposed (see the box below).

Slotted Aloha, ALOHA or in slices

An improvement of the Aloha technique has been to cut the time in slices of time, or slots, and not to authorize the issuance of frames that at the beginning of the installment, the transmission time of a frame asking exactly a slice of time. So, there is no collision if a single frame transmits at the beginning of the installment. In contrast, if several frames start to issue at the beginning of the installment, the emissions of frames are superimposed throughout the slot. In this latter case, there is retransmission after a random time.

This method improves the flow during the start-up period, but remains unstable. In addition, there is an additional cost from a complication of the devices, since all the emissions must be synchronized.

CSMA, or the random access with listening to the carrier

The technique Carrier Sense Multiple Access (CSMA) is to listen to the channel before issuing. If the coupler detects a signal on the line, it differs its emission at a later date. This considerably reduces the risk of collision, without, however, completely remove it. If, during the time of the spread between the torque of stations the more remote areas (period of vulnerability), a coupler does not detect the issuance of a frame, there may be superposition of signals. Of this fact, it must be later reissue the lost frames.

Many variants of this technique have been proposed, which differ by three characteristics:

- The strategy followed by the coupler after detection of the state of the channel.
- How collisions are detected.
- The policy of retransmission of messages after collision.

Its main variants are the following:

- **CSMA non-persistent.** The coupler listens to the channel when a frame is ready to be sent. If the channel is free, the coupler emits. In the contrary case, it starts the same process after a random delay.
- **CSMA persistent.** A coupler ready to issue previously listening the channel and transmits it is free. If it detects the occupation of the carrier, he continues to listen to up until the channel is free and emits at that time. This technique allows you to lose less time than in the previous case, but it has the disadvantage of increasing the probability of collision, since the frames that accumulate during the busy period are all transmitted at the same time.
- **CSMA p-persistent.** The algorithm is the same as previously, but, when the

channel becomes free, the coupler emits with the probability *P*. In other words, the coupler differs its emission with the probability 1 - *p*. This algorithm reduces the likelihood of a collision. Assuming that two terminals wish to issue, the collision is unavoidable in the standard case. With this new algorithm, there is a probability 1 - *p* that each terminal would not pass, which avoids the collision. On the other hand, it increases the time before transmission, since a terminal may choose not to issue, with a probability 1 - *p,* while the channel is free.

- **CSMA/CD (Carrier Sense Multiple Access/Collision Detection).** This technique for random access standardized by the Working Group IEEE 802.3 is currently the most used. To the prior listening on the network is added the listening during the transmission. A coupler ready to issue having detected the free channel transmits and continues to listen to the channel. The coupler persists to listen, which is sometimes indicated by the acronym CSMA/CD persistent. If there is a collision, it stops as soon as possible its transmission and sends signals Special, called bits of Jam, in order that all the couplers are warned of the collision. It tries to new sound emission following later an algorithm presented later.

Figure 9.7 illustrates the CSMA/CD. In this example, the couplers 2 and 3 attempt to issue during that the coupler 1 emits its own frame. The couplers 2 and 3 are in the listening and emit at the same time, in the propagation delay closely, as soon as the end of the Ethernet frame issued by the coupler 1. A collision is the result. As the couplers 2 and 3 continue to listen to the physical media, they realize the collision, stop their transmission and derive a random time to start the process of retransmission.

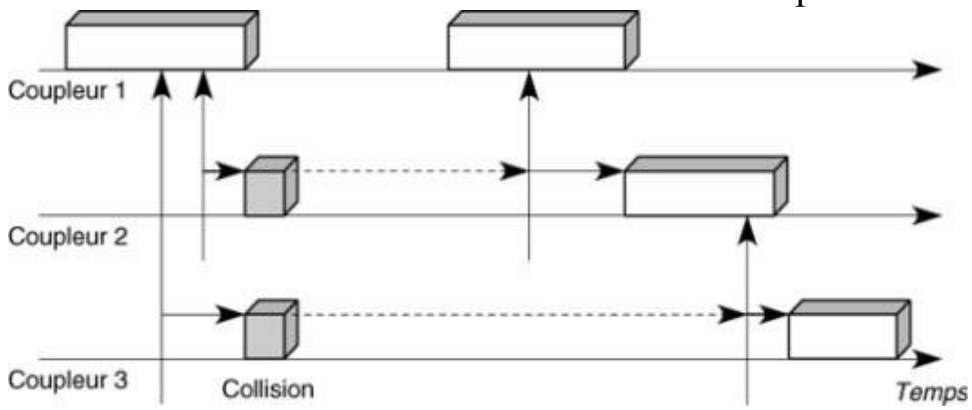


Figure 9.7
Principle of operation of the CSMA/CD

The CSMA/CD generates a gain in efficiency compared to other techniques of random access, because there is immediate detection of collisions and interruption of the transmission in progress. The couplers transmitters recognize a collision by comparing the signal issued with the one who passes on the line. The collisions are therefore no longer recognized by absence of acquittal, but by interference detection. This method of detection of conflicts is relatively simple, but it requires coding techniques powerful enough to easily recognize a superposition of signals. Typically used for this techniques of differential coding, such Manchester code differential.

- **CSMA/CA.** Less known than the CSMA/CD, the CSMA/CA (Carrier Sense Multiple Access/Collision Avoidance) begins to be heavily used in the Wi-Fi networks, i.e. the Ethernet networks wireless IEEE 802.11 (see Chapter 20). It is a variant of the CSMA/CD, which allows the method CSMA to operate when the collision detection is not possible, as in the Terrestrial. Its operating principle is to resolve the restraint before that the data are transmitted using

acknowledgments of receipt and timers.
The couplers wishing to issue test the channel on several occasions in order to ensure that no activity is detected. Any message received must immediately be acquitted by the receiver. The sending of new messages only took place after a certain period of time, so as to guarantee a transport without loss of information. The non-return of an acknowledgment of receipt, at the end of a predetermined time interval, allows to detect whether there was a collision. This strategy not only makes possible the implementation of a mechanism of acquittal at the level frame, but presents the advantage of being simple and economical, since it does not require a circuit of collision detection, unlike the CSMA/CD.
There are various techniques of CSMA with resolution of collisions, among which the CSMA/CR (Carrier Sense Multiple Access/Collision Resolution). Some variants of the CSMA also use of the mechanisms of priority that can enter under this word which avoid collisions by levels of separate priority, associated with different stations connected to the network.

The Ethernet networks switched

The previous section describes in detail the techniques used in Ethernet networks shared, in which a same cable is shared by several machines.
The other solution implementation in Ethernet networks is the switching. In this case, the physical media is not shared, two machines are exchanging Ethernet frames on a link. This solution completely changes the gives, because there is more of a collision.
The Ethernet switching, or Ethernet FDSE (full duplex switched Ethernet), was born at the beginning of the years 1990. Before the arrival of the switched Ethernet, Ethernet networks shared were cut in sub-shared networks stand alone, connected between them by bridges. Of this fact, we multiplied the traffic by the number of sub-networks.
The bridges are not in fact that Ethernet switches which store the frames and the emit back to other Ethernet networks. In pursuing this logic to the extreme, we can cut the network up to have only a single station by Ethernet network. It then gets the Ethernet switching.
The Ethernet network FDSE is a network particularly simple since there are only two stations: the one that you want to connect to the network and the switch of connection. It therefore has an Ethernet by terminal connected directly to the switch.
In switching, each Ethernet card is connected directly to an Ethernet switch, which is in charge of redirect the frames in the right direction. The switching request a reference which, *a priori,* does not exist in the Ethernet world, no package of supervision does not opening a path by asking of references. The Word of switch can therefore be considered as inaccurate since there is no reference. However, it is possible to speak of switching, if one considers the address of the recipient as a reference. The path is then determined by the result of references equal to the address of the recipient on 6 bytes. To achieve this switching of end-to-end, each switch must have the possibility to determine the connection of output in function of the reference, that is to say the value of the address of the receiver.
This technique of switching can present the following difficulties:

- Management of the addresses of all couplers connected to the network. The techniques of VLAN, which are discussed in detail in Chapter 22, allow to solve this problem.
- Management of possible congestion within a switch.

The fact of the second difficulty, it must put in place of control techniques likely to take in charge, on

the connections between switches, the frames simultaneously from all couplers Ethernet. Are the characteristics of the architectures of networks of switching. The distance limitation no longer exist, you can achieve networks in Ethernet switching to the size of the planet.
The Ethernet environment is currently imposes by its simplicity of implementation as long as the network remains of limited size. It is a network solution that presents the advantage to rely on the existing, namely the Ethernet cards present in all my Terminal machines and various Ethernet networks that many companies have put in place to create their local networks.
The disadvantage of the switching of the frame level lies in the addressing of level 2, which corresponds to the flat addressing of Ethernet. The flat addressing, or absolute, does not know the geographic location of an Ethernet card to after its value. As soon as the network has a large number of posts, which is the case if one accepts a mobility of terminals, for example, the updating of the switching tables (look-up table) becomes almost impossible because there is no standardization for the automation of this function to a large network. The Ethernet switching, introduced in Chapter 6, is only usable in small networks.
The limitation of the performance of the Ethernet environment is due to the sharing of the physical media by the whole of Ethernet cards. To remedy this disadvantage, can increase the base speed in passing to the 100 Mbit/s or 1 Gb/s. Another solution is to switch the Ethernet frames. The first step toward the switching consists, as previously indicated, to cut the Ethernet networks in small sections and to link between them by a bridge. The role of the bridge is to filter the frames in not leaving go than those intended for an Ethernet network other than the one where the frame comes from. Of this fact, we limit the number of Ethernet cards that share the same network. For this solution to be viable, the traffic must be relatively local.
In the switching, following the address, the switch forwards the frame to another switch or to an Ethernet card. The available capacity by terminal is 10 Mbit/s, 100 Mbit/s or 1 Gb/s, or 10 Gbit/s and even 40 and 100 Gb/s. The whole difficulty lies in the complexity of the networks to switching of Ethernet frames, with the problems of the opening of the paths and flow control that they pose.
A second solution of Ethernet switching is spreading rapidly through the MPLS techniques and Ethernet Carrier Grade. It is Ethernet networks for telecommunications operators. In the first case, it adds a new area in the Ethernet frame to bring the reference that is no longer the Ethernet address of destination. In the second case, it uses the VLAN field to introduce a path. These two solutions are explained in the following.

Ethernet for businesses

This section describes the Ethernet networks of business, and the following Ethernet networks of operators.
The first example of Ethernet networks to 10 Mbit/s is important because it determines the Ethernet networks following. As soon as all the cards of a network to ensure the 100 Mbit/s, the network automatically has this speed. On the other hand, as long as it remains in the network a card to 10 Mbit/s, even if this 10 Mbit/s is almost more used, the network is running to 10 Mbit/s.

The Ethernet networks 10/100 Mbit/s
The Ethernet networks to 10 Mbit/s have been the first to be introduced on the market. Today they are replaced by the networks to 10/100 Mbit/s, which adapt automatically to the 100 Mbit/s as soon as it no longer exists of Ethernet card to 10 Mbit/s on the network. This box reviews the different products of the shared Ethernet 10/100-Mbit/s.

Cheapernet

Cheapernet is a local Ethernet network shared using a coaxial cable particular, standardized under the word 10Base2/100base2. The

coaxial cable used is no longer the yellow cable Shielded, but a cable end of brown color Non Shielded, also called *thin cable* or cable RG-58. This cable has a lesser resistance to electromagnetic noise and induces a weakening the more important of the signal.
In the following, only the Ethernet 10 Mbit/s is detailed. The case 100 Mbit/s in deducted by dividing by 10 the values obtained. The strands of the Ethernet 10 Mbit/s are limited to 185 meters. The repeaters are of Ethernet type and work to 10 Mbit/s. The total length can reach 925 or 1 540 meters according to the versions. The constraints are the same as for the Ethernet network with regard to the round trip time. On the other hand, to obtain a comparable quality, it must limit the distance without repeater. The maximum length here has less importance, because the network Cheapernet is a capillary network enabling it to go up to the end user at a lower cost.

Starlan

Born of a study of AT&T on the quality of the telephone wiring from the dispatcher to floor, the Starlan network responds to a Any other need that the network Cheapernet. On the capillary networks of the company, speeds of 1 Mbit/s were acceptable in the 1980s. The arrival of Starlan corresponded to the willingness to use this infrastructure capillary, i.e. the telephone wiring, starting from the distributor of floor (see figure 9.8).

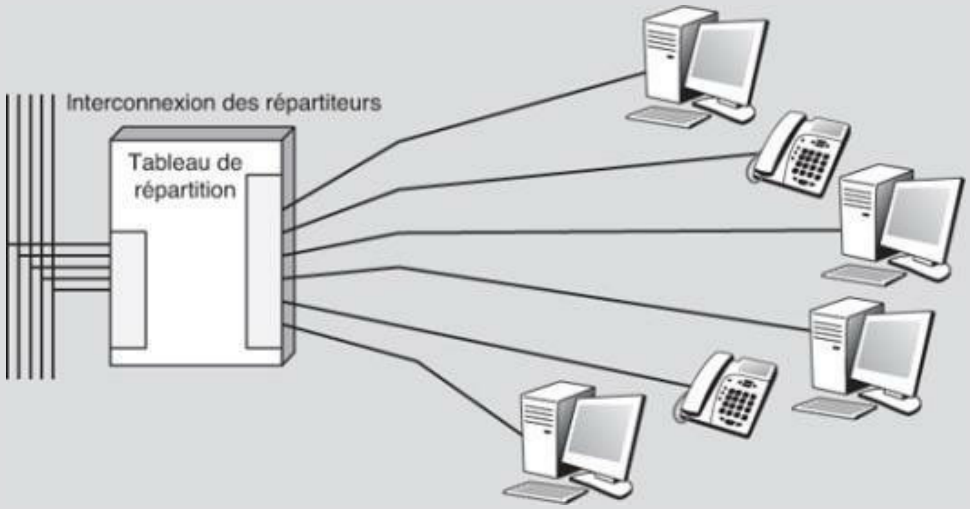


Figure 9.8
Wiring from the dispatcher to floor

As the cables from the dispatcher to floor did little but a flow rate of 1 Mbit/s, it has resumed the Ethernet technology in adapting to a wiring in star at a speed of 1 Mbit/s. It is always the CSMA/CD access method that is used on a star network active, such as the one shown in Figure 9.9. The speed of 1 Mbit/s was quickly replaced by a solution to 10 Mbit/s.

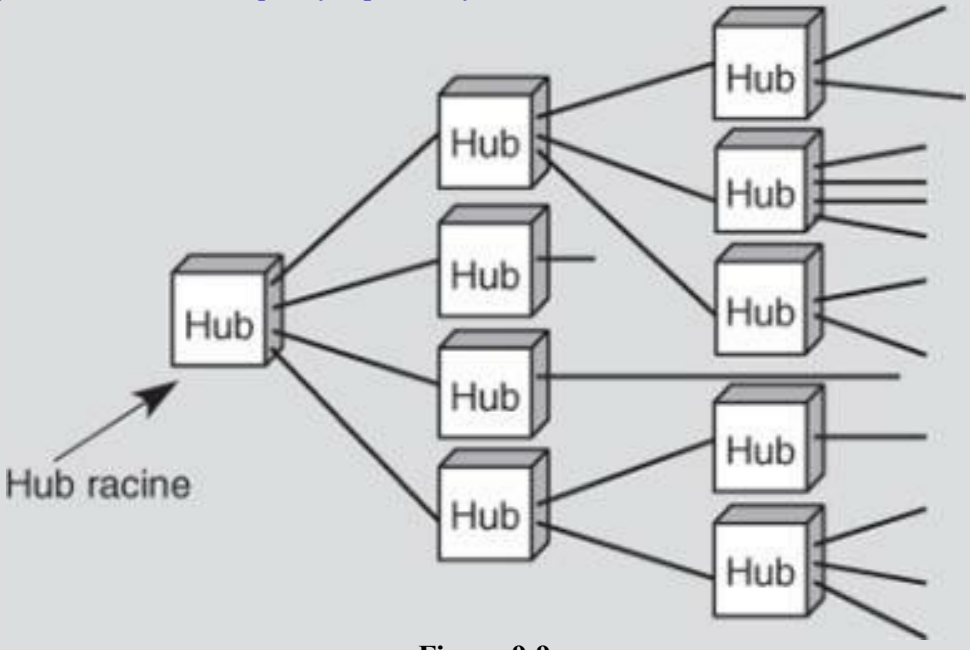


Figure 9.9
Architecture of a Starlan Network

Given the large number of companies who have renewed their wiring with pairs of wires in quality, the Starlan to 10 Mbit/s has met with a massive success. The Starlan networks to 10 Mbit/s also bear the name of Ethernet 10 Mbit/s on twisted wire pairs for well indicate that the coupler is the same as that of the Ethernet networks on coaxial cable to 10 Mbit/s. The dual 10/100 speed is imposed in the years 1990. In the following, only the 1 Mbit/s, but the case of 10 Mbit/s in deducted easily.
The Starlan standard IEEE 802.3 1BASE5 allows to have a maximum of five nodes, or hubs, successive, from the base node included. Between two nodes, a maximum distance of 250 meters is permitted. In reality, we find exactly the same constraints as in the Ethernet

network, that is to say, a roundtrip time maximum of 512 microseconds between the two points the most distant. For the local network Starlan to 10 Mbit/s, found the value of 51.2 microseconds.
The hub is an active node Able to regenerate the signals received toward the whole of the output lines, so that there is broadcast. The hub allows you to connect the equipment terminals located at the ends of the branches Starlan.
Classically, in each equipment corresponds a plug connection Starlan. However, to add an additional terminal in an office, it should pull a cable since the dispatcher to floor or the sub-dispatcher the closest, on condition that there is still a possible output. An alternative solution is to place it on the same socket several machines connected in series, as shown in Figure 9.10.

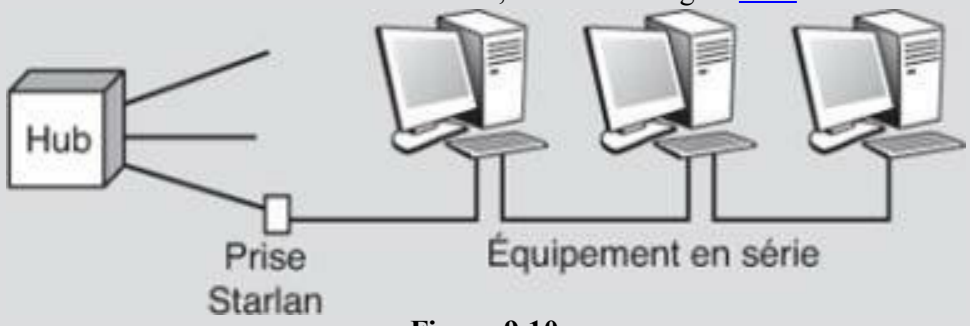


Figure 9.10
Equipment in series connected on a Socket Single Starlan

The 100 Mbit/s Fast Ethernet

Fast Ethernet is the name of the extension to 100 Mbit/s of the Ethernet network to 10 Mbit/s. It is the working group IEEE 802.3u which is at the origin. The technique of access is the same as in the Ethernet Version 10 Mbit/s, but at a speed multiplied by 10. The frames being transported are identical. This increase in speed can encounter the cabling system and to the possibility or not of the transit flows also important.
This is the reason for which three sub-standards have been proposed for the 100 Mbit/s:

- IEEE 802.3 100BaseTX, which requires two pairs unshielded (UTP) Category 5 or two pairs shielded (STP) of type 1.
- IEEE 802.3 100BaseT4, which requires four pairs of unshielded (UTP) of categories 3, 4 and 5.
- IEEE 802.3 100BaseFX, which requires two optical fibers.

The maximum distance between the two points the most distant is strongly reduced compared to the version to 10 Mbit/s. The minimum length of the frame is always 64 bytes, the transmission time is of 5.12 microseconds. It is deduced that the maximum distance that can be traveled in this period of time is of the order of 1 000 meters, which represents for the Fast Ethernet network a maximum length of approximately 500 meters. As the time of crossing of hubs is relatively important, most of the manufacturers limit the maximum distance to 210 meters for the Fast Ethernet. The time between two frames, or intergap, is reduced to 0.96 microseconds.
This solution has the advantage of offering a good compatibility with the version to 10 Mbit/s, which allows you to connect on a same hub stations to 10 Mbit/s and 100 Mbit/s. The cost of the connection to the 100 Mbit/s is today the same as that of the classic Ethernet, ten times less rapid.
The Fast Ethernet networks often serve networks of interconnection of networks Ethernet 10 Mbit/s. The distance relatively limited covered by the Fast Ethernet does allows him not however'" watering" a company a bit large. The Gigabit Ethernet, detailed below, does not resolve this issue further in its shared version. On the other hand, the switched version not having more constraint of distance, the Switched Gigabit Ethernet is one of the solutions for the interconnection of Fast Ethernet networks.
Another solution to extend the coverage of the Ethernet network is to connect Fast Ethernet by bridges designed to filter out the frames using the MAC address. These bridges having the same features as the switches, it is found today in large enterprises of networks to transfer of Ethernet frames that use

of Ethernet switches. These architectures are discussed later in this chapter.

The Gigabit Ethernet (GbE)

The Gigabit Ethernet, or GbE, is an evolution of the Ethernet standard. Several improvements have been made for this in Fast Ethernet at 100 Mbit/s.

The interface to new amended is called GMII (Gigabit Media Independent interface). It has a data path on 8 bit, instead of 4 in the version less powerful. The transceivers are working with a clock clocked at 125 MHz. The coding adopted comes from the Fiber Channel products to reach the gigabit per second. A single type of repeater is now accepted in this new version.

The different standardized solutions are the following:

- 1000BaseCX, two twisted pairs of 150 Ω;
- 1000BaseLX, to a pair of fiber optical wavelength high;
- 1000BaseSX, to a pair of optical fiber from Short Wavelength;
- 1000BaseT, to four pairs of Category 5 UTP.

The technique of access to the physical media, the CSMA/CD, is also amended. To be compatible with other versions of Ethernet, which is a basic principle, the size of the transmitted frame must be between 64 and 1 500 bytes. The 64 bytes, i.e. 512 bits, correspond to a time of issuance of 512 ns. This time of 512 ns represents the maximum distance of the media that a station in emission does not disconnect before having received a possible signal of collision. This represents 100 meters for a return trip. If no hub is installed on the network, the maximum length of the physical media is 50 meters. In the facts, with a hub of attachment and portions of the cable up to the couplers, the maximum distance is reduced to a few meters. To avoid this distance too short, standard setters have artificially increased the length of the frame for the bring to 512 bytes. If the length of the frame to be transmitted is less than 512 bytes, the coupler adds bytes of jam that are then removed by the coupler receiver.

If it is a good solution to enlarge the gigabit network, the useful rate is however very low if all frames to transmit have a length of 64 bytes, an eighth of the bandwidth being used in this case. It is in particular the case for transport of telephony over IP (ToIP), where the useful bytes for telephony are the number of 16. In this case, it is necessary to perform a jam of the frames of 64 bytes.

The Gigabit Ethernet accepts the repeaters or hubs when there are several possible directions. In this last case, an incoming message is copied on all lines of output. Figure 9.11 illustrates a Gigabit repeater corresponding to the IEEE 802.3z standard. The different solutions of Gigabit Ethernet can be interconnected by the intermediary of a repeater or a hub.

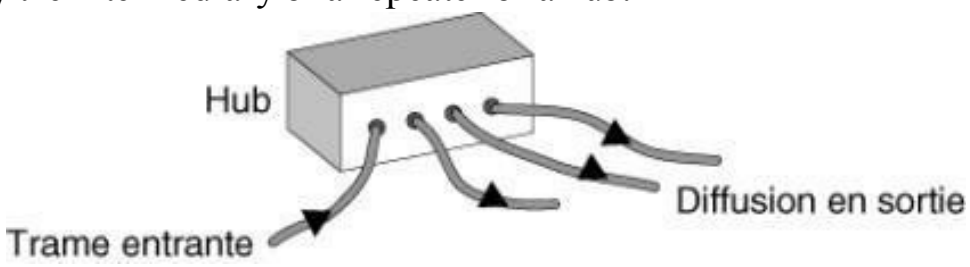


Figure 9.11
Hub Ethernet Gigabit

The Gigabit Ethernet also works in switched mode, in a configuration full-duplex. It can, by this means, interconnect Gigabit Ethernet between them or of the Fast Ethernet and conventional Ethernet.

Routers Gigabit are also available when it goes back up to the Network layer IP. In this case, you must retrieve the IP packet to perform a routing using the IP address that is located in the package to be transported to the inside of the Ethernet frame. Figure 9.12 illustrates an interconnection of two

switched networks by a Gigabit Router.
The management of the network Gigabit, like that of the older Ethernet networks, is ensured by conventional techniques, essentially Simple Network Management Protocol (SNMP). The Management Information Base (MIB) of Gigabit Ethernet is detailed in the standard IEEE 802.3z.

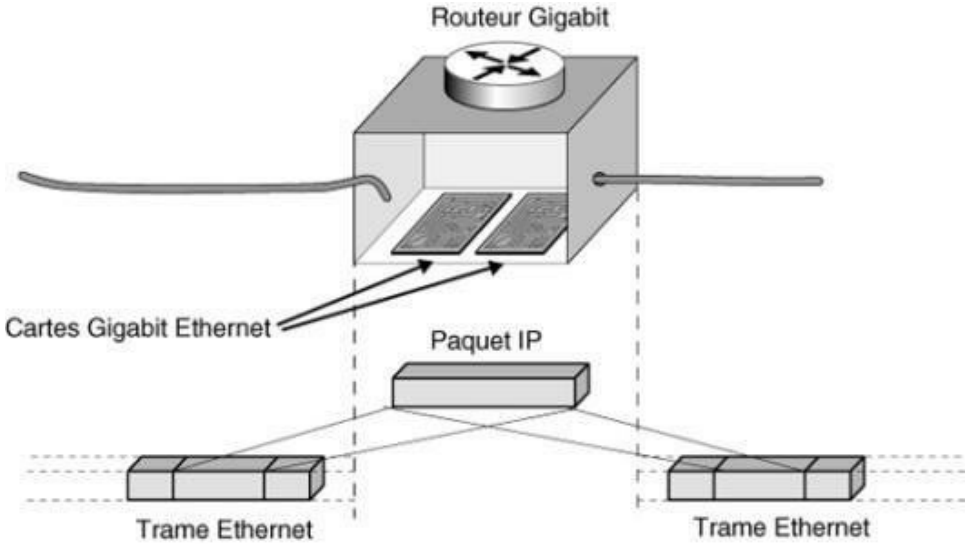


Figure 9.12
Interconnection of two Ethernet networks switched by a router

The 10 Gigabit Ethernet (10GbE)

The 10 Gigabit Ethernet, or 10GbE, is an evolution of the Ethernet standard that is surpassed only by the version 100GbE (100 Gbit/s). This technique is heavily used in the metropolitan networks and operators. It is a simple solution, because it is sufficient to multiplex ten networks GbE to multiply the speed by ten.
The 10 Gigabit Ethernet, or 10GbE has been standardized by the Working Group to the IEEE 802.3ae, in the objective to propose two types of solutions, all two in full-duplex and switching. The distance ranges from 65 meters with multimode fibers up to 40 kilometers with singlemode optical fiber. The two types of interfaces proposed are LAN-phy and WAN-phy.
The group IEEE 802.3ae has standardized in the PHY LAN a stream at the speed of 10,312 5 Gbit/s with a Coding 64B/66B. The WAN interface-PHY uses the same coding, but with a compatibility with the SONET interfaces OC-192 and SDH STM-64.
The proposed architecture by this working group is shown in Figure 9.13.
The Working Group of the IEEE incorporates a compatible interface SONET but which remains Ethernet. As previously explained, this interface implies the existence of a physical media 10GbE, called WAN PHY, which is equivalent to the support SONET/SDH Type of OC-192 or STM-64. The advantage of this compatibility is to allow to resume all the management environment and maintenance as well as the reliability of SONET/SDH. This solution has been defended by the 10 GEA (10 Gigabit Ethernet Alliance).

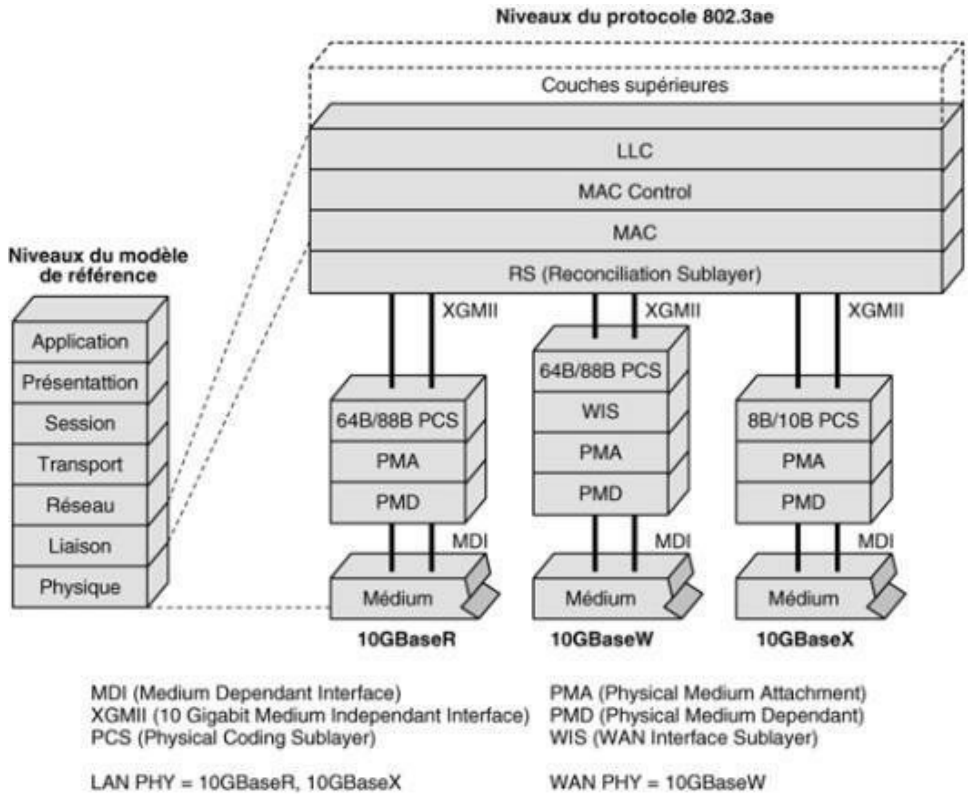


Figure 9.13
Architecture of networks 10GbE

The 100 Gigabit Ethernet (100GbE)

The 100 Gigabit Ethernet, or 100GbE, is the latest evolution of the Ethernet standard. This solution has been proposed by the Ethernet NG Forum, whose objective is to define the Ethernet environment of the new generation (Http://www.ng-ethernet.com/). The standard has been finalized and voted in June 2010 by the group IEEE 802.3ba, elsewhere with a version to 40 Gbit/s (40GbE).
The 40GbE and 100GbE are natural extensions switched mode of the 10GbE. It uses a technology CWDM (Coarse WDM) with 10 and 25 wavelengths. Several physical levels have been standardized which must be chosen in function of the distance to reach. The main levels are:

- On Metal Cable: 10 meters to 40 Gbit/s;
- On multimode fiber: 100 meters at 100 Gbit/s;
- On singlemode fiber: 40 kilometers to 100 Gbit/s.

Ethernet for the operators

Ethernet becoming the standard for the transport of IP packets, the fact of the strong cooperation between IP and Ethernet, all operators implement Ethernet to the place of ATM. The Ethernet solution is quite compatible with MPLS since the switched Ethernet uses a shim-label, or reference, operation classic, entering perfectly in this framework. The other solution for the operators comes from the Ethernet Carrier Grade.
The Ethernet solutions allow current to put in place this technology in many contexts, ranging from the metropolitan network to the WAN. The flow rates are in the range of 1 to 100 Gbit/s. The most important options concern the following solutions:

- The virtual networks Ethernet, for that the user has the impression that the remote network is connected to the network of the company by a local Ethernet network.
- Metropolitan-area networks, with the thrust of the MEF (Ethernet Metropolitan Forum).

- The ability to manage loops high flow at high reliability, as on SONET, with the standard RPR (Resilient Packet Ring), standardized by the Working Group IEEE 802.17. This solution is presented in Annex F.
- Ethernet MPLS forwarding, which is discussed in Chapter 11.
- Ethernet Carrier Grade.

The connections Ethernet Virtual

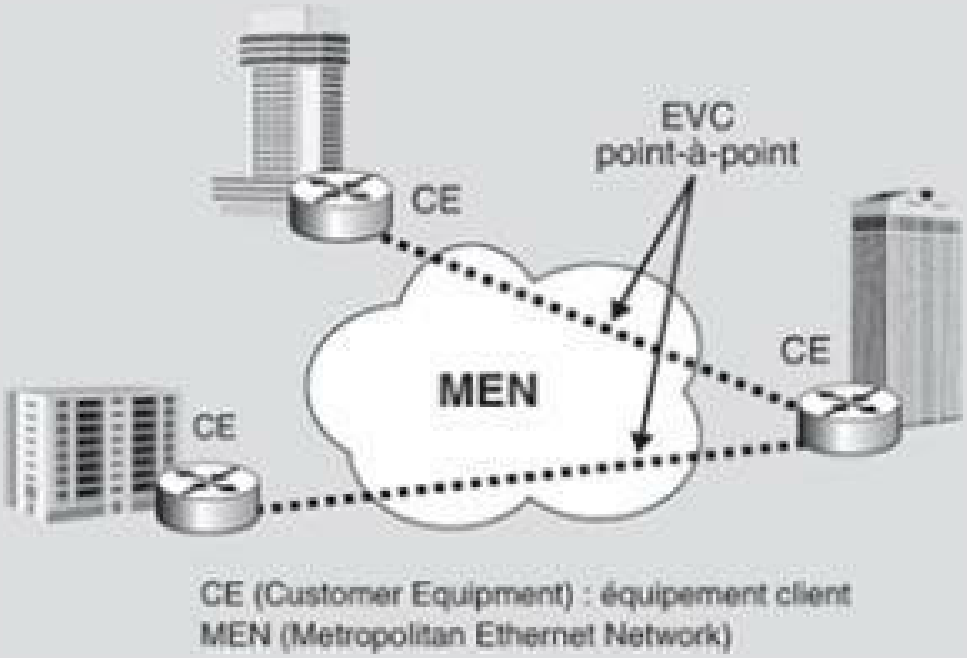


Figure 9.14

Virtual Private Network Ethernet to basis of EVC

Figure 9.14 illustrates a network of Ethernet operator using a virtual private network Ethernet built from EVC (Ethernet virtual connection). Thanks to this solution, the VPN-IP or MPLS are perfectly compatible with the Ethernet world.

A EVC can be either point-to-point or point-to-multipoint, is still multipoint-to-multipoint. A virtual private network Ethernet can therefore have all the properties of conventional VPN, but with a higher efficiency, since the system is located at level 2, and to a much lower cost through the Ethernet hardware used.

The the underlying solutions of VLAN (Virtual LAN) are detailed in Chapter 22, taking into account their importance for enterprise networks and of operators.

The Ethernet networks MEF

The MEF networks (Ethernet Metropolitan Forum) are known since a long time. They aim to interconnect networks of business in a same metropolis at very high flow rate. The techniques used are described in Chapter 6. They use the Ethernet networks switched to 1, 10, 40 and 100 Gb/s.

To achieve of IP telephony and services requiring strong temporal constraints, networks of operators Ethernet (Carrier Ethernet) allow the introduction of priorities, to this difference is that the field IEEE 802.1p, which serves to the introduction of these priorities, has only 3 bits. These 3 bits allow only 8 levels of priorities, to compare to the 14 levels defined by the IETF for services DiffServ.

The levels of priorities proposed by the MEF are the following:

- 802.1p-6 DiffServ Expedited Forwarding.
- 802.1P-5/4/3 DiffServ Assured Forwarding.
- 802.1p-5, which presents the loss the more low.
- 802.1p-3, which presents the loss the stronger.
- 802.1p-2 DiffServ Best Effort.

In the Ethernet environment, flow control is usually a delicate problem. Various proposals have been made to improve it. In particular, the methods of backpressure offer to send control messages on the part of the overloaded switches, which allow related switches to stop their emission to the congested node during a specified time in the primitive of control.

The choice made by the MEF is a control of type Frame Relay, where is found exactly the same parameters:

- CIR (Committed Information Rate)
- CBS (Committed Burst Size)
- PIR (Peak Information Rate)
- MBS (Maximum Burst Size)

These different proposals show that the world Ethernet is by train to nibble in small market shares and should become the technology a number of the network operators in a future which is approaching.

Ethernet Carrier Grade

Ethernet was designed for computer applications, not for multimedia applications. To upgrade and enter in the field of multimedia, the Ethernet environment has therefore had to transform. There is talk of Ethernet Carrier grade, i.e. acceptable to the operators of telecommunications with the control and management tools necessary in this case. This mutation concerns essentially the Ethernet switched.

The Ethernet Carrier grade must have the features that are found in the telecommunications networks, including the following:

- The reliability, which allows you to have only very few failures. The average time between failures, or MTBF (Mean Time Between Failure), must be at least 50 000 hours.
- The availability, which must reach the traditional values for telephony, i.e. be in running condition 99.999% of the time. This value is far from being reached by the Ethernet networks classic, which are rather to 99.9% of the time.
- The protection and restoration. When a failure occurs, the system must be able to be switched back on at the end of a maximum time of 50 milliseconds. This time comes from the telephony, which does not accept cuts than on time intervals less than this value. The SONET networks, for example, reach this time value of reconfiguration. The solutions are usually the redundancy, total or partial, which allows you to put on the road another path, planned in advance, in case of a power outage.
- The optimization of the performance by a monitoring active or passive. The performance are not all homogeneous when the waves of packets vary. It is therefore necessary to adapt the waves so that they can pass through without problem.
- The network must be able to accept the SLA (Service Level Agreement). The SLA is a notion of a typical network of operator when a customer wants to negotiate a guarantee of service. The SLA is determined by a technical part, the SLS (Service Level Specification), and an administrative part in which are negotiated the penalties if the system does not give satisfaction.
- The management is also an important feature of the networks of operators. In particular, systems of failure detection and warning signs must be available for the network to be in a state of walking.

From a technical point of view, the Ethernet Carrier Grade is an extension of the VLAN technology. A VLAN is a local network in which machines can be found in very distant points. The objective is to operate this network as if all points were geographically close to each other to form a local network. A VLAN can have several users. Each Ethernet frame is issued in broadcast to all of the machines in the VLAN. The tables that perform the routing of frames are fixed and may be seen as switching tables in which the addresses of the recipients are of references.

When the VLAN has only two points, the issuance of an Ethernet frame to a point to the other is similar to a switch to a path. It is this vision that has been retained in the Ethernet Carrier Grade. We form paths in determinant of the VLAN. The path is unique and simple if the VLAN has only two points and multipoint in if it has more than two points.

The problem with this solution comes from the limited size of the zone VLAN, which allows only twelve elements binaries. This was appropriate in the context of a business network with a Ethernet switching standard, but became quite inadequate in the framework of the Ethernet Carrier Grade,

which aims the networks of operators. It has therefore been necessary to increase the size of the VLAN field.
It may subdivide the Ethernet Carrier Grade in several solutions for the extension of the zone VLAN, all described in Figure 9.15. The most common solution is to use the standard IEEE 802.1ad, which is known by several names: Ethernet PROB (Provider Bridge), QiQ (Q in Q) or VLAN in cascade. The standard IEEE 802.1ah is also known under the names of MIM (MAC-in-Mac) or PBBS (provider backbone Bridge). The most advanced solution is called PBT (provider backbone transport), or PW over PBT (pseudowire). It allows you to return to the transport to a classical solution in which the Ethernet frames are switched following a succession of references corresponding to MPLS-SL (Label Service).
On the figure 9.15, these solutions are compared to the MPLS solution PW PBT, which uses both the PBT solution and the solution MPLS.

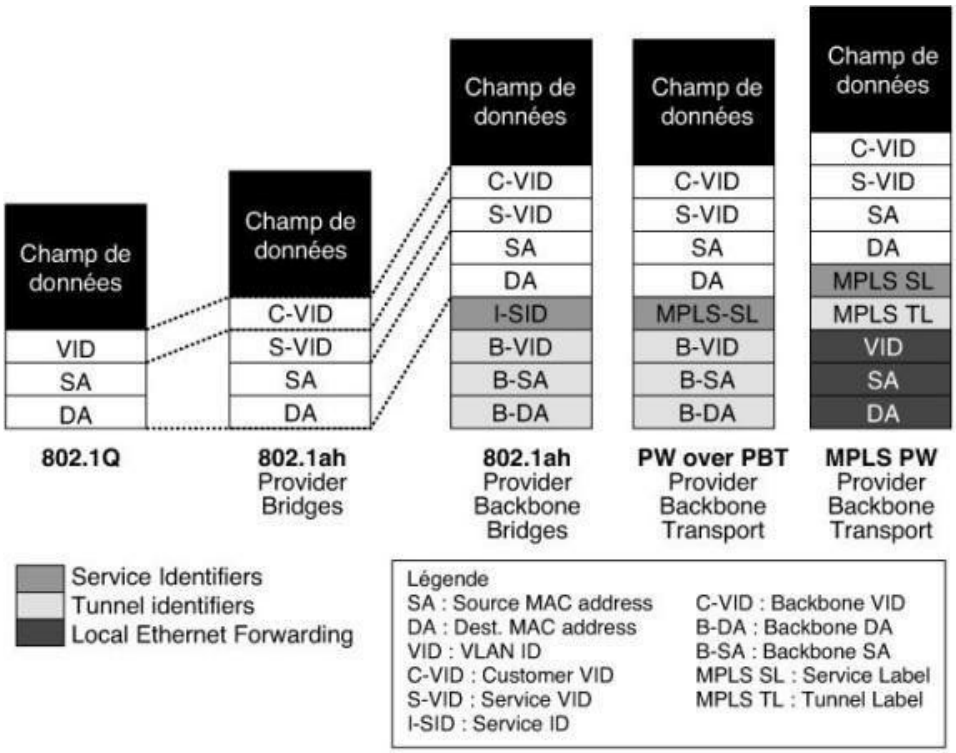


Figure 9.15
Technologies of the Ethernet Carrier Grade

Let us consider in a first time the Ethernet technology PROB (Provider flange). The Internet access provider adds a VLAN number to that of the client. There are therefore two VLAN numbers: the C-Vid (Customer-VLAN ID) and the S-VID (Service-VLAN ID). The bridge of the provider of the service allows you to extend the concept of VLANS in the network of the operator without destroying the VLAN of the user. This solution allows you to define the broadcasts to perform in the network of the operator. As the field of the Ethernet frame, in which the VLAN number is indicated, has a length of 12 bits, this allows you to define up to 4 094 entities of the network of the operator. These entities can be services, tunnels or in the areas of dissemination. However, if 4 094 is a sufficient value in the enterprise, it remains very inferior to the needs of an operator. Of implementations play on a translation of reference to enlarge the field, but this increases the complexity of management of the Whole. This solution is therefore not appropriate to large networks.
The solution proposed by the group IEEE 802.1ah PBB (provider backbone Bridge) improves the previous by switching the traffic of frames on the MAC address. This solution, called the MIM (MAC-in-MAC), encapsulates the MAC address of the client in a MAC address of the operator. This allows the operator the heart of not knowing that its MAC addresses. In the BBP network, the

correspondence of MAC addresses user and network MAC is known only by the nodes panel, avoiding the explosion of MAC addresses.
The third solution, called PBT (provider backbone transport), is fairly close to the MPLS technical, while providing the necessary properties in carrier grade, as a rate of unavailability of less than 50 milliseconds. It is somewhat of a tunnel MPLS rescued. The PBT tunnel is created as a tunnel MPLS, with corresponding references to the ends of the network. The VLAN numbers client and server are encapsulated in the tunnel MPLS, which may itself have a differentiation in VLAN operator. The actual reference is therefore of 24 bits + 48-bit, or 72-bit.
The last solution is that of service PS (pseudowire) of MPLS. In this case, the user VLANS and operator are encapsulated in a service tunnel MPLS, which itself can be encapsulated in a tunnel of MPLS transport. This solution comes from the encapsulation of tunnels in MPLS.
The Ethernet technology Carrier Grade is of interest to many operators. The future will tell what will be the winning solution. But it is already certain that the encapsulation of VLANS in the VLAN will be present in each of them.

The extensions of Ethernet

Ethernet is both a standard old and a standard of the future: old, by the techniques of local network, and the future, thanks to the Switching and its application to metropolitan-area networks and extended, but also to the local loop, to electrical networks, wireless, etc.

Ethernet in the local loop

The local loop is to connect users to the first node, router or switch, the operator in which the customer has a subscription. The solutions to high flow is shared between ATM and Ethernet, but since 2010 all new connections are Ethernet.
It includes everything from following the interest of this solution, which offers a continuity with the Machine Terminal, which generally has a Ethernet card. The terminal equipment is most of the time connected by Ethernet to the xDSL *or cable modem broadband. Why change of technology, that is to say decapsulating the IP packet that has been introduced in an Ethernet frame to put it in a frame ATM? The simplest solutions would have been either to put directly an ATM card in the PC, either the use of Ethernet modems in the place of ATM modems.*
The Working Group FSM (Ethernet in the First Mile) has proposed for this the standard IEEE 802.3ah, which includes three types of topologies, and physical media:

- Point-to-point in twisted pairs at a speed of 10 Mbit/s on a distance of 750 meters;
- Point-to-point in optical fiber to a speed of 1 Gbit/s on a distance of 10 kilometers;
- Point-to-multipoint in optical fiber to a speed of 1 Gbit/s on a distance of 10 kilometers.

The standard specifies the procedures for the administration and maintenance for the ends and the line itself. These procedures allow them to escalate the faults and to monitor the parameters of the connection.
The VDSL (Very high bit rate DSL), equivalent to the *xDSL modems for high speeds, is also compatible with Ethernet and offers complete continuity at high speed of the workstation transmitter until the workstation receiver by means of EFM modems.*

Power over Ethernet (PoE)

A major disadvantage of network equipment, and Ethernet devices in particular, comes from the need to feed into electrically. In the event of a power outage, the network stops and can put in difficulty the business. A solution that is needed more and more is the self-supply of the network equipment by the intermediary of the wiring itself.
The working group IEEE 802.3af proposes to electrically energise the Ethernet equipment by the Ethernet cable itself. The equipment can be as well of the switches in the network that Wi-Fi access points or other network equipment that connect on a Ethernet jack. The advantage of this solution is to secure the electrical power supplies of servers connected to the electrical network and therefore all other equipment connected to the server.
The electrical sources can be of two types: End-span, in which a particular server serves the whole of the network, and mid-span, where several equipment share the supply to the other network equipment. Metallic wires can be of category 3 with 4 wires or 5/6 with 8 wires, i.e. four pairs of wires. Two solutions are used to carry the 48 volts toward equipment who need: either the power supply uses two pairs while the other two pairs are dedicated to data, either the pairs are mixed when there are only two pairs, electricity and the current being worn on different frequencies.
The PoE solution is illustrated in Figure 9.16, where multiple network devices (IP phone, access point, access Bluetooth, webcam) are directly supplied through Ethernet. The equipment UPS (Uninterruptable Power Supply) can be added to avoid power outages.
Another Vision of Ethernet is the transport of frames directly on the electrical cables. This solution is strongly developing in the framework of the domotics, where individuals can connect directly to their card Ethernet coupler on their many electrical outlets.

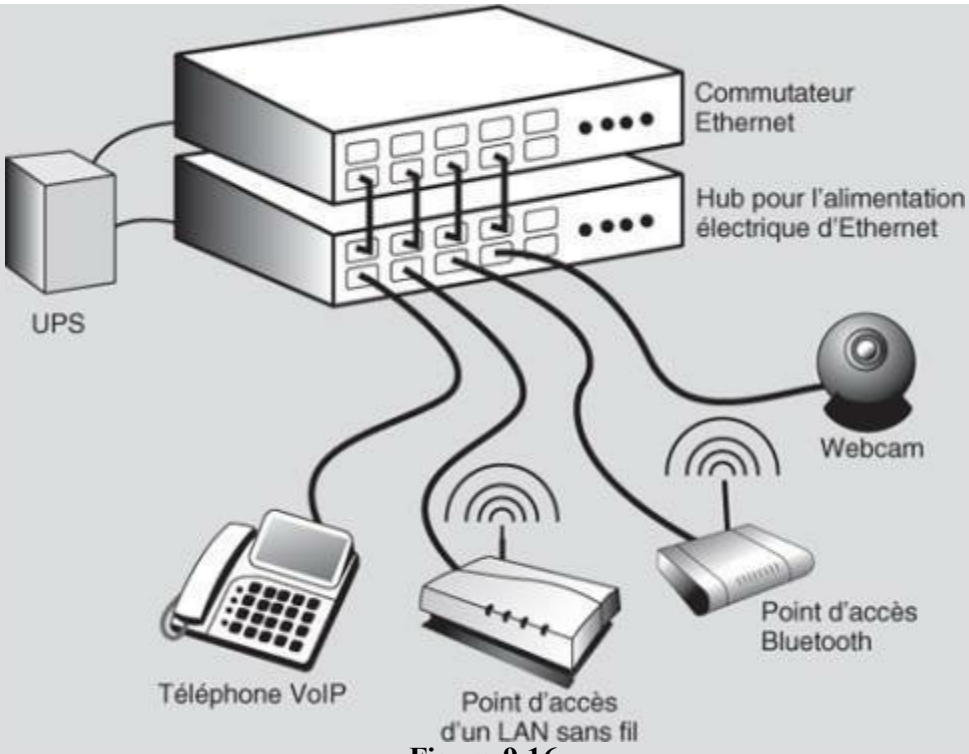


Figure 9.16
Architecture of the electrical supply PoE

Ethernet for data centers

Since 2010, the arrival of large physical sites where are located the computer equipment of businesses has changed in depth the traffic of the networks, which is done since then to more than 80% with these datacenters. This evolution has need to consolidate the waves to reach these datacenters. The transport networks of packages had to therefore go beyond the capabilities of everything that had been built in the years before 2010.

Two types of transport are required to support these new centers for processing of data: the transport between machines of a same datacenter and the transport between two data centers. The protocols promoted by the IETF to support the corresponding networks are respectively trill and LISP, detailed in the sections that follow, but there are other proposals, including the following:

- VXLAN (Virtual eXtensible LAN), which are discussed later in this chapter.
- SDN (Software-Defined Networking), which has taken such an importance that it is specifically studied in Chapter 12.
- The techniques associated with the Ethernet Carrier Grade, introduced earlier.

TRILL

TRILL (Transparent Interconnection of Lots of links) is an IETF standard implemented by nodes called RBridges (routers-bridges) or trill switch. TRILL combines the advantages of bridges and routers. In effect, Trill follows a routing of Level 2 using the state of the links. The RBridges are compatible with the bridges of Level 2 Defined in the IEEE standard 802.1 and they could gradually replace them. The RBridges are also compatible with the routers IPv4 and IPv6 and therefore perfectly compatible with IP routers present. The routing performed with the protocol IS-IS at level 2 between the RBridges replaces trill in the Spanning Tree Protocol (STP).

The fact to use a routing of level 2 between the Ethernet addresses allows you no longer have to configure the level 3 and the associated IP addresses. The protocol for the state of the links used for routing may contain additional information of type Type Length Values (TLVS). To avoid the potential

loops during the reroutages, the RBridges manage a *hop count, i.e. the number of nodes traversed. When this number reaches a maximum value, determined by the operator, the frame is destroyed.*
When an Ethernet frame arrives in the first RBridge of the network, called the IRB (Ingress RBridge), an additional header is added, the Trill header. It will be décapsulée by the RBridge output, or ERB (Egress RBridge). The new frame door in part trill the address of the ERB, which gives this technology the status of routing since it uses the address of the recipient and not a reference. This address of door RBridges on 2 bytes, replacing the 6 bytes of the classic Ethernet frame. These 2 bytes are chosen by the user himself. The header added includes six bytes in total, ending with the input and output addresses, preceded by the *hop count* and a flag.
Ethernet frames are retrieved after decapsulation to the output. There is then a classic mode of transfer, either on the VLAN number, the Ethernet address used for reference, either in décapsulant the Ethernet frame to find the IP packet.
An interesting point is the possibility to transfer the frames of a same stream by multiple paths simultaneously: it is called the *multipath. The Protocol ECMP (Equal Cost Multipath) allows you to detect the different routes to the same cost and to refer the frames on these different routes.*

VXLAN

VXLAN (Virtual eXtensible LAN) is a solution to achieve of communications in large clouds between datacenters. The technology is quite similar to that of the VLANS and the Ethernet Extensions Carrier Grade. It was originally proposed by Cisco and VMware. The disadvantage of the basic solution of VLANS IEEE 802.1Q is the limitation to 4 096 VLAN. VXLAN allows to extend the basic technology in parallel of the Ethernet Carrier Grade.
In the Ethernet frame base, it adds a VXLAN area of identification of 24 bits to achieve more than sixteen million of VLAN and then a UDP field of 64-bit, which is then encapsulated in an Ethernet frame with fields source address, destination and VLANS associated with XVLAN. The Ethernet frame base is therefore the part "data" of the UDP message, itself transported in an Ethernet frame. This property allows you to achieve VLANS beyond a domain Ethernet. These encapsulations are represented in Figure 9.17.

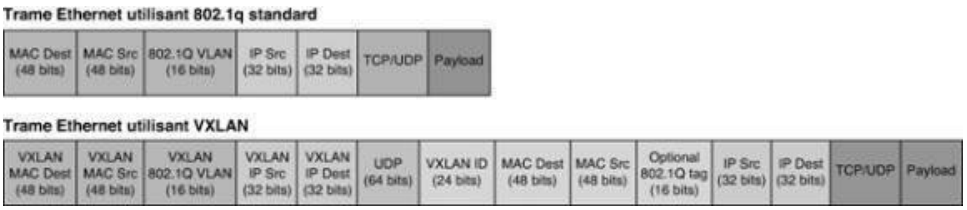


Figure 9.17
The frame VXLAN

It is to be noted that the overload is important because the frame VXLAN adds 36 bytes to the base frame.

Conclusion

The Ethernet networks dominate the market of enterprise networks for many years. These networks are only accentuate their advance, and soon, virtually 100% of enterprise networks will be Ethernet.
In the face of such a success, the Working Group 802 of the IEEE multiplies the offensives in other directions, as wireless networks, for which the wireless Ethernet became the main standard.
Of the metropolitan networks Ethernet are already marketed, and their success is no doubt. The networks of operators are also strongly moved to the Ethernet technology. Several solutions exist for this, including Ethernet MPLS forwarding and the Ethernet Carrier Grade. For the very high speeds in

the clouds, new protocols arrive on the market, always on concepts Ethernet.

10

IP networks

IP is an acronym very known in the field of networks. It corresponds to the architecture developed for the Internet network. In the strict sense, Internet Protocol (IP) is a packet level protocol. Above this Protocol, at the message level, two protocols are associated with it: TCP (Transmission Control Protocol) and UDP (User Datagram Protocol).
This chapter describes the general architecture of the IP networks and protocols that allow this environment to manage the problems of addressing and routing and more generally all the protocols associated with the IP protocol and located in the level package.

The IP architecture

The IP architecture is based on the mandatory use of the IP protocol, which has for basic functions the addressing and routing of IP packets. The IP level corresponds exactly at the level package of the architecture of the reference model.
Above of IP, two protocols have been chosen, TCP and UDP, which are introduced in Chapter 7. These protocols are consistent at the message level of the reference model. In fact, they incorporate a basic session, thanks to which TCP and UDP support the features of layers 4 (transport) and 5 (session).
The main difference between them lies in their mode, with connection for TCP and without connection for UDP. The TCP protocol is very comprehensive and ensures a good quality of service, in particular on the rate of error of packages transported. In contrast, UDP is a connectionless protocol, which supports applications less stringent in terms of quality of service.
The layer that is located at the top of TCP-UDP integrates the functionality of the layers 6 and 7 of the reference model and essentially represents the level application.
The TCP/IP architecture is shown in Figure 10.1. It contains three levels: the level package, the message level and a level containing the features of the upper layers.

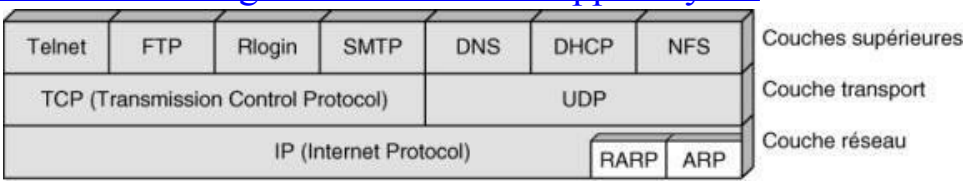


Figure 10.1
Architecture TCP/IP

The Internet

At the end of the 1960s, the U.S. Department of Defense decides to carry out a large network from a multitude of small networks, all different, which begin to cross-fertilize a little everywhere in North America. He had to find a way to make coexist these networks and give them a external visibility, the

same for all users. Where the appellation of InterNetwork (inter-network), abbreviated to Internet, given to this network of networks.
The Internet architecture is based on a simple idea: to ask all networks who want to be part to carry a single type of packet, a format determined by the IP protocol. In addition, the IP packet must carry an address defined with sufficient generality to be able to identify each of the computers and terminals scattered across the world. This architecture is illustrated in Figure 10.2.
The user who wishes to issue on this inter-network must store its data in IP packets, which are handed over to the first network to cross. The first network encapsulates the IP packet into its own structure of package, the package has, which circulates in this form up to an exit door, where he is decapsulated so as to retrieve the IP packet. The IP address is examined to locate, thanks to a routing algorithm, the next network to cross, and so on up to arrive at the destination device.
To complete the IP protocol, the U.S. defense has added the TCP protocol, which clarifies the nature of the interface with the user. This protocol also determines the way of transforming a stream of bytes in an IP packet, while ensuring a quality of the transport of this IP packet. The two protocols, assembled under the acronym TCP/IP, are in the form of a layered architecture. They correspond respectively to the packet level and at the message level of the reference model.
The Internet model is completed by a third layer, called the application level, which brings together the various protocols on which to build the Internet services. The email (SMTP), the transfer of files (FTP), the transfer of hypermedia pages, the transfer of distributed databases (World-Wide Web), etc., are a few of these services. Figure 10.3 illustrates the three layers of the Internet architecture.

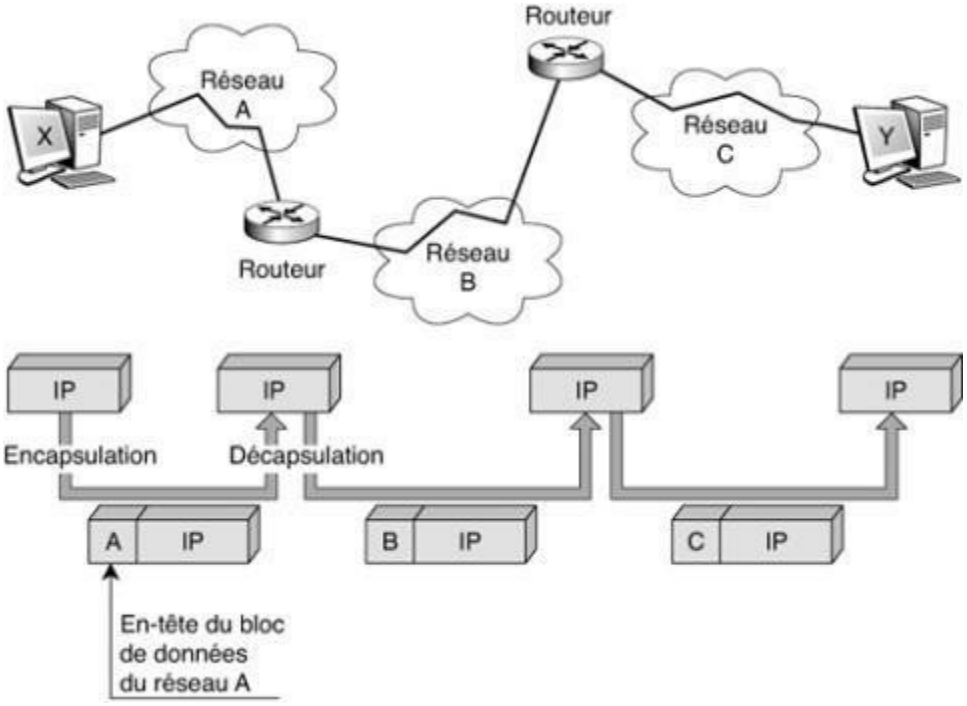


Figure 10.2
The architecture of the Internet

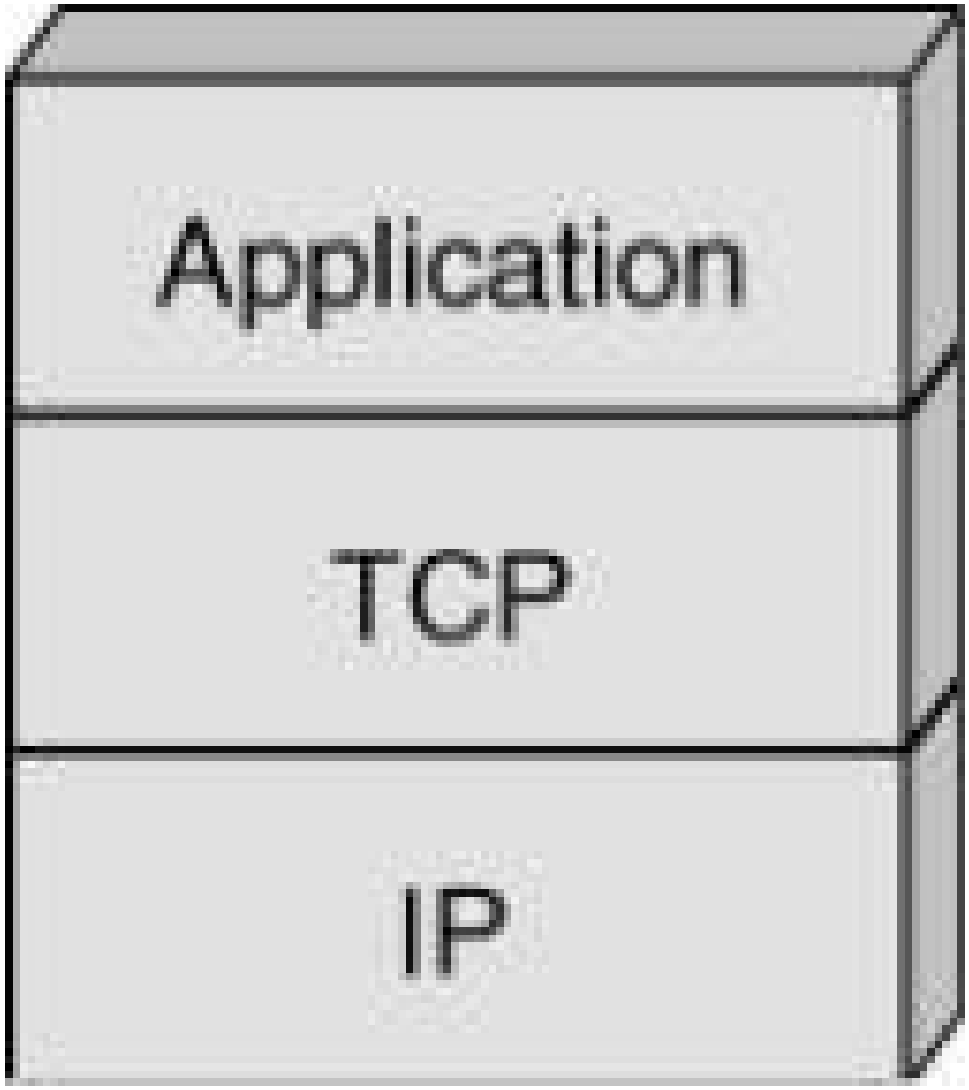


Figure 10.3
The three layers of the Internet Architecture

IP packets are independent of each other and are routed individually in the network by the interconnecting equipment sub-networks, the routers. The quality of service proposed by the IP protocol is very low and offers no detection of lost packets or a possibility of resumption on error.

The TCP protocol includes the features of the message level of the reference model. It is a protocol quite complex, which includes numerous options to solve all the problems of packet loss in the lower levels. In particular, a fragment lost can be recovered by retransmission on the stream of bytes. The TCP protocol uses a mode with connection.

The flexibility of the Internet architecture can sometimes be a default, in the measure where the global optimization of the network is carried out sub-network by sub-network, by a succession of local optimizations. This does not allow a homogeneity of the functions in the various sub-networks crossed. Another important characteristic of this architecture is to situate the entire control system, that is to say the intelligence and the control of the network, in the Machine Terminal, leaving virtually nothing in the network, at least in the current version, IPv4, the IP protocol. The intelligence of control can be found in the TCP software of the PC connected to the network.

This is the TCP protocol that is in charge to send more or fewer packets in function of the occupation of the network. A control window specifies the maximum number of fragments not paid can be issued. The control window of TCP increases or decreases the traffic depending on the time required to perform a back-to-back. The more this time increases, the more one sees the network congestion, and the more the flow of issuance must decrease to combat the saturation. In return, the cost of the infrastructure is extremely low, no intelligence are not found in the network. The service rendered by

the network of networks corresponds to a quality called best-effort, which means that the network does its best to sell the traffic. In other words, the quality of service is not assured.
The new generation of the IP protocol, the IPv6 protocol, introduces new features, which make the nodes of the network more intelligent. The next-generation routers are equipped of algorithms for the management of the quality of service, which allow them to ensure a transport capable of responding to the constraints of time or to packet loss. It is expected the arrival of IPv6 from a dozen of years, but it is always IPv4 Which Regent the world IP. The reason for this is that each new need achievable with IPv6, IPv4 has been able to find the algorithms necessary to do as well.
In IPv4, each new client is treated in the same way as those who are already connected, since resources are fairly distributed between all users. The policies for the allocation of resources of the networks of telecommunications operators are totally different, since, on these networks, a customer who already owns a certain quality of service suffers no penalty of the fact of the arrival of a new customer. The solution today recommended in the Internet environment is to foster customers with requirements of real time, in the middle of protocols adapted, using the levels of priority (see later).
The IP protocol exists since thirty years, but it is almost remained confidential for twenty years before takeoff, less by its properties that the fact of the failure of the protocols related directly to the reference model, too numerous and often incompatible. The growth of the world IP comes from the simplicity of its Protocol, with very few options, and its free.

Operation of TCP/IP networks

Most of the networks are independent entities, put in place to make service to a small population. Users choose the networks adapted to their specific needs, because it is impossible to find a satisfactory technology all types of needs. In this basic environment, users who are not connected to the same network cannot communicate. The Internet is the result of the interconnection of these different physical networks by routers. The interfaces of access must respect for this certain conventions. It is an example of open systems interconnection.
To obtain the interoperability of different networks, the presence of the IP protocol is mandatory in the nodes that provide routing between networks. Overall, the Internet is a network to transfer of packets. These packets go through one or several sub-networks to reach their destination, except of course if the transmitter is located in the same subnet as the receiver. The Packets are routed in gateways located in the nodes of interconnection. These gateways are the routers. More specifically, the Routers forward packets from input to output, in determining for each package on the best route to follow.
The Internet is a routed network, by opposition to networks X.25 or ATM, which are switched networks. In a routed network, each packet follows its own road, which is each moment optimized thanks to the dynamic of the routing tables, whereas in a switched network, the path is always the same.

The problems posed by the Synchronization

The IP architecture, using the routing and the service best-effort, presents a difficulty: the synchronism. In effect, the time of crossing of a package is a relatively random. It depends on the number of packets waiting in the output lines of the nodes and the number of retransmissions corresponding to errors in line. The fact of the transport of real-time applications with strong synchronizations, as the word real time, poses complex problems, which can only be resolved in some cases. If it is assumed that a telephone conversation interactive between two individuals accepts a latency of 600 milliseconds go-return, it is possible to resynchronize the bytes to the output, if the total time of paquétisation-dépaquétisation and crossing of the network is actually less than 300 milliseconds in each direction.
The resynchronization that it is possible to obtain is shown in Figure 10.4. It is necessary to determine a maximum time of crossing of the network and perform a resynchronization on this value. The software of the computer terminals, that can be termed intelligent, allows you to manage these timing problems if the time of crossing of the network is bounded. Then there is the question of whether the

maximum time of crossing is acceptable by the application. This maximum time of crossing of the network must be less than 28 milliseconds if the echoes exist at the ends and equal at most to 300 milliseconds if there is interactivity and to several seconds, or even tens of seconds if the application is, monodirectional as the video at the request or the dissemination of radio programs.

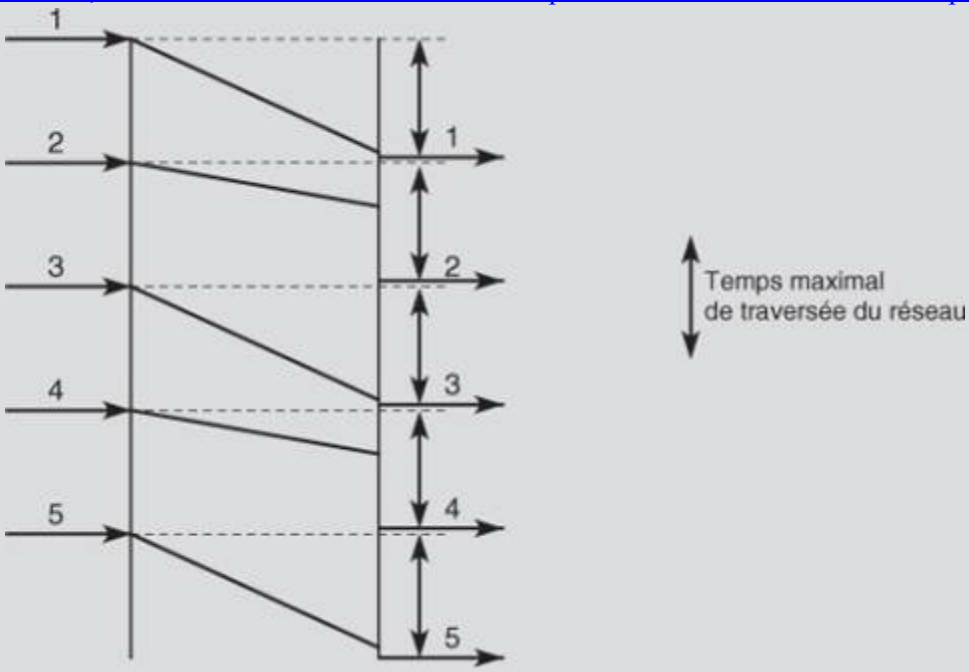


Figure 10.4

Resynchronize a isochronous application

It is obvious that if the device is not intelligent and analog, the reconstruction of the synchronous flow is impossible in a network to transfer of packets. In addition, it is necessary to put in place of the protocols allowing for a tighter control of the information isochronous in case of saturation. Of this fact, the Internet network has difficulties to carry data isochrons, at least until the arrival of the new generation, which has been designed in the spirit of the multimedia.

The IPv4 and IPv6 addressing

As the Internet is a network of networks, the addressing is particularly important. This section provides a first overview of the problems of addressing through the IP protocol of first generation IPv4 and the new generation IPv6.

The machines of the Internet have an IPv4 address represented on a 32-bit integer. The address consists of two parts: a network identifier and an identifier of the machine for this network. There are four classes of addresses, each for coding a different number of networks and machines:

- Class A: 128 networks and 16 777 216 hosts (7 bits for the networks and 24 for the hosts);
- Class B: 16,384 networks and 65,535 hosts (14 bits for the networks and 16 for the hosts);
- Class C : 2 097 152 networks and 256 hosts (21 bits for the networks and 8 for the hosts);
- Class D: group addresses (28 bits for the hosts belonging to a same group).

These addresses are detailed in Figure 10.5.

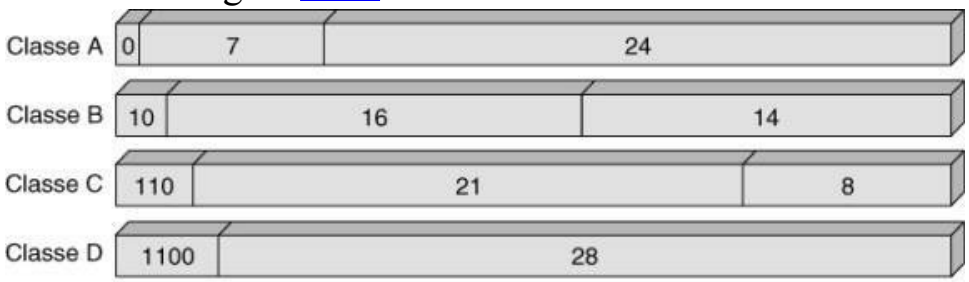


Figure 10.5

Classes of addresses of IPv4

The IP addresses have been defined to be processed quickly. The routers that perform the routing based on the number of network are dependent on this structure. A host connected to multiple networks has several IP addresses. In reality, an address does not identify just a machine, but a connection to a network.

To ensure the uniqueness of network numbers, Internet addresses are assigned by a central agency, the NIC (Network Information Center). We can also define its own addresses if it is not connected to the Internet. However, it is strongly recommended to obtain an official address to ensure interoperability in the future.

As the addressing of IPv4 is somewhat limited, it took propose an extension to cover the needs of years 2 000. This extension of address is often presented as the reason for the new version of IP, whereas it is only one reason among others.

The IPv6 address takes on 16 bytes. The number of potential addresses authorized by IPv6 exceeds 1023 for each square meter of the surface of the earth. The difficulty of use of this vast reserve of addresses lies in the representation and the rational use of these 128 bits. The representation is carried out by a group of 16-bit and takes the following form:

123:FCBA:1024:AB23:0:0:24:FEDC

Of the series of addresses equal to 0 can be abbreviated by the sign ::, which can only appear once in the address. In effect, this sign not indicating the number of successive 0, to deduct this number by examining the address, the other series can not be abbreviated.

IPv6 addressing is hierarchical. An allocation of addresses (i.e. a distribution between potential users) was proposed, of which the Table 10.1 provides the detail.

Address	The first few bits of the address	Characteristics
0:: /8	0000 0000	Reserved
100 :: /8	0000 0001	Not assigned
200 :: /7	0000 0001	ISO Address
400:: /7)	0000 010	Novell address (IPX)
600 :: /7	0000 011	Not assigned
800 :: /5	0000 1	Not assigned
1000 :: /4	0001	Not assigned
2000 :: /3	001	Not assigned
4000 :: /3	010	Aadresses of service providers
6000 :: /3	011	Not assigned
8000:: /3	100	Geographical Addresses of users
A000 :: /3	101	Not assigned
C000::/3	110	Not assigned
E000 :: /4	1110	Not assigned
F000:: /5	1111 0	Not assigned
F800:: /6	1111 10	Not assigned
FC00 :: /7	1111 110	Not assigned
FE00:: /9	1111 1110 0	Not assigned
FE80:: /10	1111 1110 10	Addresses of local links
FEC0:: /10	1111 1110 11	Addresses of local sites
FF00:: /8	1111 1111	Address of multipoint

Table 10.1 • allocation of IPv6 addresses

The Address Resolution Protocols ARP and RARP

The IP addresses are allocated independently of the hardware addresses of the machines. To send a datagram on the Internet, the network software must convert the IP address to a physical address, used to transmit the frame.

The Address Resolution refers to the determination of the address of a device from the address of this same equipment at another level of protocol. It resolves, for example, an IP address to an Ethernet address or in an ATM address.

It is the Address Resolution Protocol (ARP) who performs this translation between the world of IP and Ethernet based on the physical network. ARP allows the machines to resolve addresses without the use of a static table listing all the addresses of the two worlds. A machine uses ARP to determine the physical address of the recipient in broadcasting in the Subnet an ARP request containing the IP address to translate. The machine with the IP address concerned responds by returning its physical address. To make ARP more efficient, each machine maintains a table in its memory of the resolved addresses and thus reduces the number of programming in the Broadcast mode.

At the time of its initialization (bootstrap), a machine without a mass memory (diskless) must contact its server to determine its IP address and be able to use the TCP/IP services. The Protocol RARP (Reverse ARP) allows a machine to use its physical address to determine its logical address on the Internet. The RARP mechanism allows a computer to identify itself as a target by broadcasting on the network a RARP request. The servers receiving the message examine their table and respond. Once the IP address is obtained, the machine stores in RAM and no longer uses RARP until it is reset.

The ARP protocol is based on the physical network in order to perform address translation. To determine the physical address of the recipient, a machine broadcasts on the subnet an ARP request that contains the IP address to translate. The machine with the IP address concerned responds by returning its physical address. This process is illustrated in Figure 10.6.

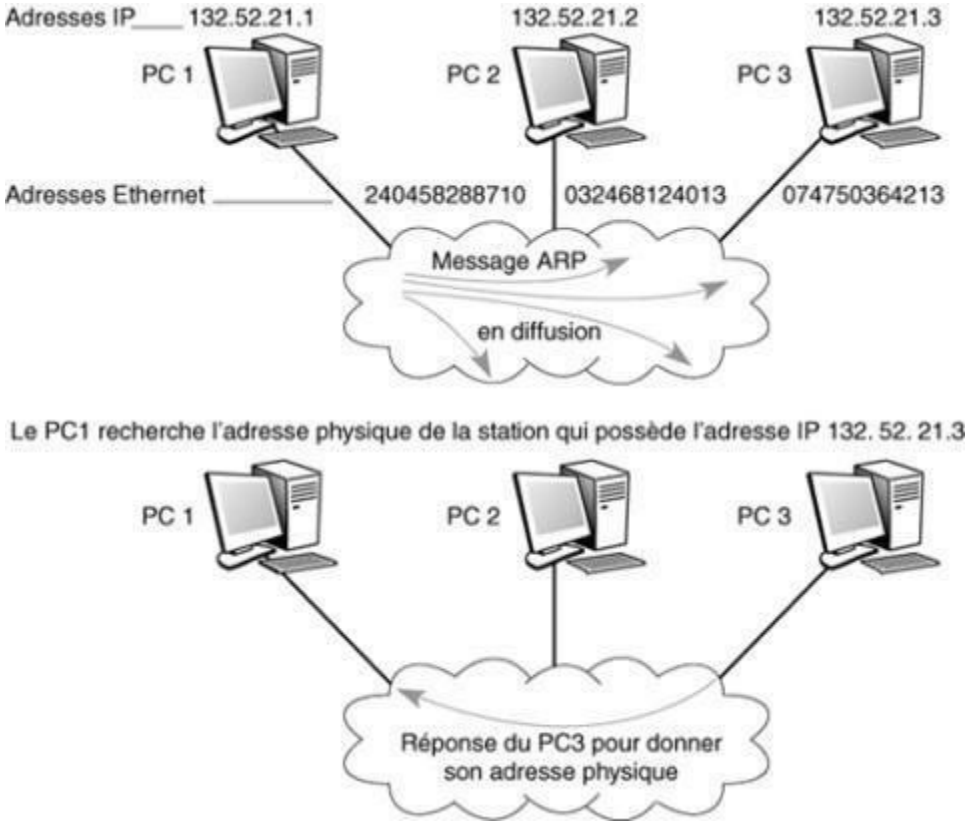


Figure 10.6

Operation of the ARP protocol

In an inverse manner, a station that connects to the network may know its own physical address without having an IP address. At the time of its initialization, the machine will contact his server to

determine its IP address and be able to use the TCP/IP services. The RARP protocol allows it to use its physical address to obtain its logical address on the Internet. Through the mechanism RARP, a station can be identify as the target by broadcasting on the network a RARP request. The servers receiving the message examine their table and respond. Once the IP address is obtained, the machine stores in RAM and no longer uses RARP until it is reset.
In IPv6, ARP and RARP are replaced by a neighbor discovery protocol, called ND (Neighbor Discovery), which is a subset of the protocol to control Internet Control Message Protocol (ICMP), presented in detail in Chapter 10.

Domain Name System (DNS)

As indicated previously, the structures of addresses are complex to handle, because they are in the form of groups of decimal digits of type *ABC* : *def:GHI:JKL, with a maximum value of 255 for each of the four groups. The IPv6 addresses held on 8 groups of 4 decimal digits. The input of such addresses in the body of a message would soon become unbearable. This is the reason for which the addressing uses a hierarchical structure completely different, much more simple to handle and store.*
The role of the DNS is to allow the mapping of physical addresses in the network and the logical addresses. The logical structure is hierarchical and uses at the highest level of the areas characterizing mainly the countries, which are indicated by two letters, as *fr* for France, and functional areas such as:

- *Com* : commercial organizations;
- *Edu* : academic institutions;
- *Org* : Organizations, institutional or not;
- *Gov* : U.S. Government;
- *Mil* : U.S. military organizations;
- *Net* : The network operators;
- *Int* : international entities.

On the inside of these major areas, there is sub-areas, which correspond to large companies or to important institutions. For example, *RP* represents the name of the team working in the field of networks and of the performance of the LIP6 laboratory of the University of Paris VI, which gives the address *Rp.lip6.fr* for the staff of this team within the laboratory.
To carry out this operation of translation, the World IP uses of the name servers, that is to say the servers that can respond to requests for name resolution or yet to be able to perform the translation of a name to an address. The servers of the names of the Internet are the DNS servers. These servers are hierarchical. When it is necessary to find the physical address IP to a user, the servers that manage the DNS is send requests to refit sufficiently in the hierarchy to find the physical address of the corresponding. These requests are made by the intermediary of small messages, which bear the question and the response in return.
Figure 10.7 illustrates the operation of the DNS. The customer *Guy.pujolle@reseau.lip6.fr* wants to send a message to *Xyz.xyz@systeme.lip6.fr.* To determine the IP address of *Xyz.xyz@systeme.lip6.fr,* a request is issued by the PC of Guy Pujolle, which queries the name server of the *network domain.lip6.fr.* If it has in memory the correspondence, it responds to the PC. In the contrary case, the request goes back in the hierarchy and reaches the server names of *LIP6.fr,* which again can respond positively if it knows the correspondence. In the contrary case, the request is routed to the server of *system names.lip6.fr,* who knows the correspondence. It is

therefore he who responds to the PC of departure.

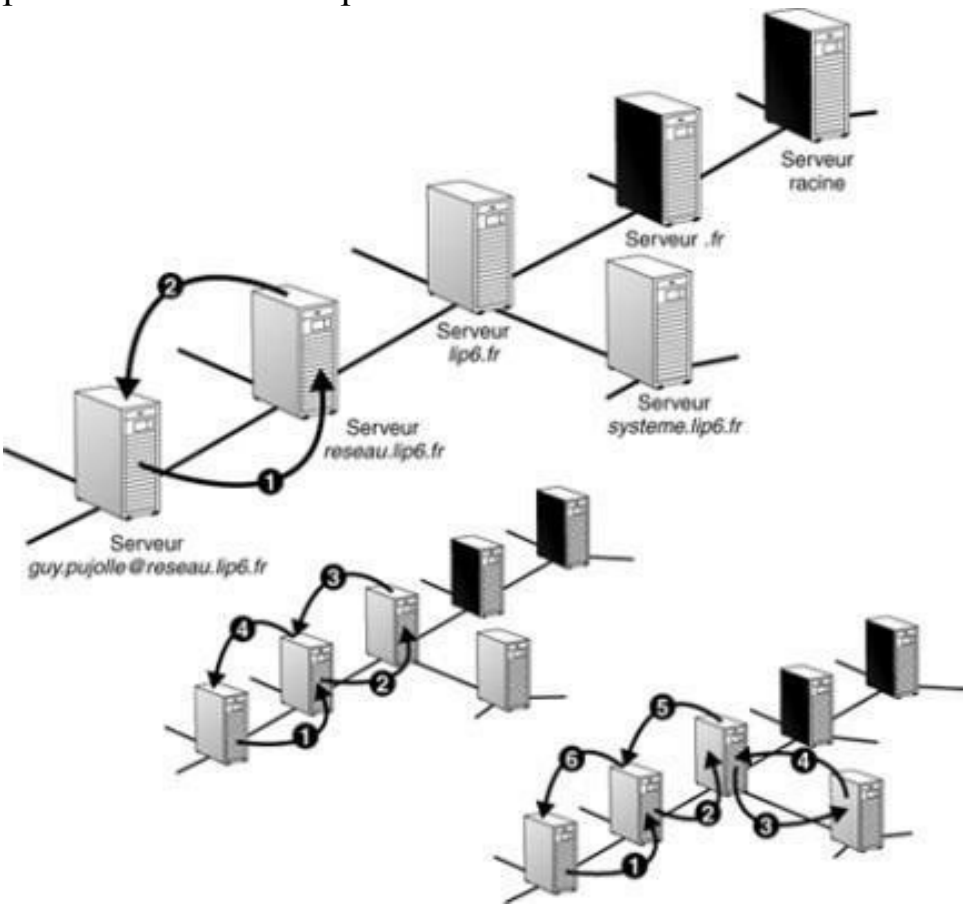


Figure 10.7
Operation of the DNS

The format of a DNS query is shown in Figure 10.8.

Identifier *(identificateur)*	Control *(contrôle)*
Number of Questions *(nombre de questions)*	Number of Answers *(nombre de réponses)*
Number of Authorities *(nombre d'autorités)*	Nombre de champs supplémentaires
Question *(question)*	
Answer *(réponse)*	
Authority *(autorité)*	
Additional *(champ supplémentaire)*	

Figure 10.8
Format of a DNS query

The first two bytes contain a reference. The customer chooses a value to place in this field, and the server responds using the same value, so that the client recognizes its request. The two following bytes contain the bits of control. These indicate if the message is a request of the client or a response from the server, if a request to another site must be performed, if the message has been truncated by lack of place, if the response message comes from the name server responsible or not of the requested address, etc. for the receiver who responds, a response code is also included in this field.
The six following possibilities have been defined:

- 0: No error.
- 1: The question is formatted in an illegal manner.
- 2: The server does not know answer.

- 3: the requested name does not exist.
- 4: The server does not accept the request.
- 5: The server refuses to answer.

Most of the requests make only one request at a time. The form of this type of query is shown in Figure 10.9. In the Question box, the content must be interpreted in the following way: 6 indicates that 6 characters follow; after the 6 characters of Network, 4 refers to the 4 characters of LIP6, 2 The two characters of *FR* and finally 0 the end of the field.

The field allows authority to servers which have authority on the name requested to make themselves known. The area additional fields allows you to carry information on the time during which the answer to the question is valid.

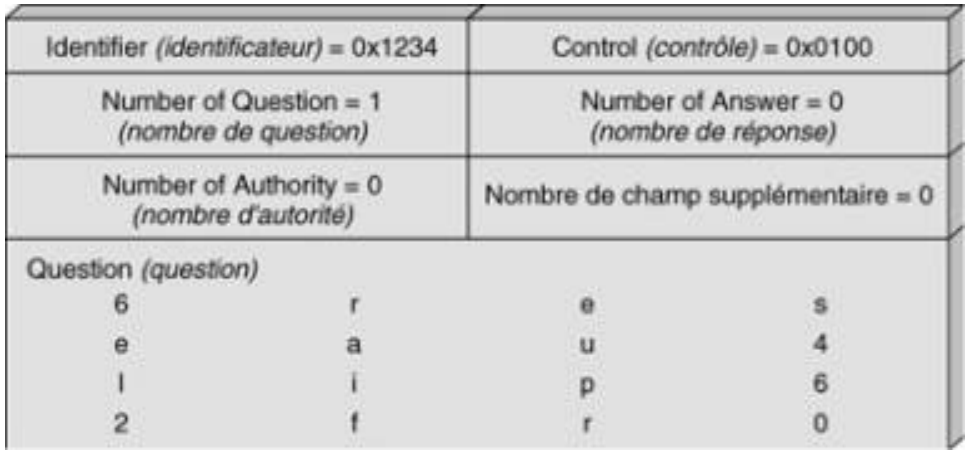


Figure 10.9
DNS query with a single request

IP routing

An internet environment is a result of the interconnection of physical networks by routers. Each router is connected directly to two or more networks, hosts usually being connected to a single network, but this is not mandatory.

There are several types of routes:

- **Direct Routing.** This is the case if the two machines who want to communicate are attached to the same network and therefore have the same IP network number. It may be of two hosts or of a router and a host. It is sufficient, to perform the transport of IP packet, to determine the physical address of the recipient and to encapsulate the datagram in a frame before sending it over the network.
- **Indirect Routing.** In this case, routing is more complex, because it is necessary to determine the router to which the datagrams must be sent. These can thus be passed from router to router until they reach the destination host. The routing function is based primarily on the routing tables. The routing is carried out from the network number of the IP address of the destination host. The table contains, for each number of network to reach, the IP address of the router to which to send the datagram. It may also include a default router address and the indication of direct routing. The difficulty of routing comes from the initialization and updating of routing tables.
- **The subnetting.** This technique of addressing and routing allows standardized to manage multiple physical networks from a same IP address of the Internet. The principle of subnetting is to divide the party host number of an IP address in the

subnet number and the host number. Outside of the site, the addresses are interpreted without that account is taken of the subnetting, cutting is known and treaty that of the Interior. The redistribution of the host number allows you to freely choose the number of machines in function of the number of networks on the site. At the conceptual level, the techniques of addressing and routing are the same. At the physical level, it uses an address mask.

The Internet network is so broad that it had to be split into autonomous systems to facilitate the management. Called autonomous system (AS) A set of routers and networks connected to each other, administered by the same organization and is exchanging packets through a same routing protocol.

The routing protocol shared by all routers in an autonomous system is called interior routing protocol, or IRP (Interior Routing Protocol). An internal protocol does not need to be implemented to the outside of the autonomous system. Of this fact, it can choose its routing algorithm to optimize the routing inside. The interior protocols are also called Interior Gateway Protocol (IGP).

When a network Internet involves several autonomous systems connected between them, it is necessary to appeal to an exterior routing protocol, or ERP (Exterior Routing Protocol). The ERP protocols must have a knowledge of the various as to accomplish their task. The protocols LES are also called an Exterior Gateway Protocol (EGP).

The figure 10.10 gives an example of autonomous systems with PIR protocols interconnected by an LES.

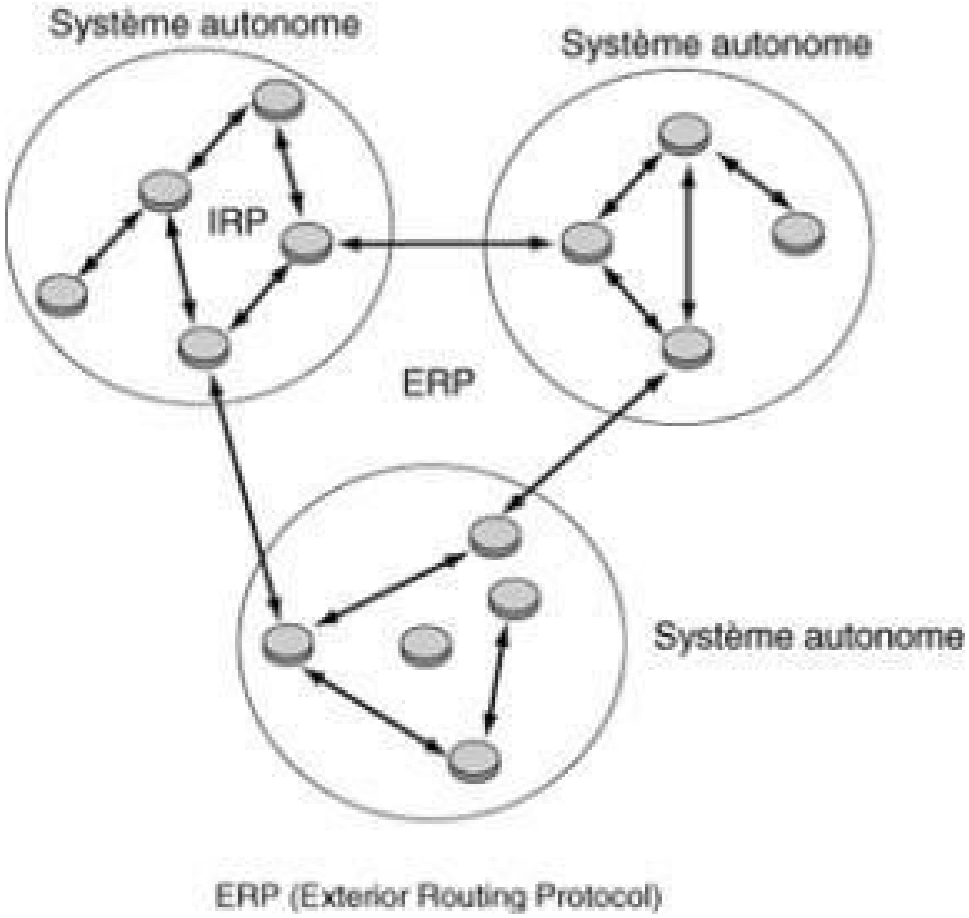


Figure 10.10

Routing protocols inside and outside

The routing algorithms

A routing algorithm is a process to determine the routing of packets in a node. For each node of a network, the algorithm determines a routing table, which, at each destination, match a line of output.

The algorithm must lead to a coherent routing, that is to say without the loop. This means that it should not be that a node route a packet to another node that could return the packet.
There are three broad categories of routing algorithms:

- Distance-vector (distance-vector routing);
- To state of the links (link state routing);
- Path Vector (path-vector routing).

Distance vector routing algorithms require that each Node s exchange of information between neighbors, that is to say between nodes directly connected. Of this fact, each node can maintain a table by adding information on all its neighbors. This table gives the distance to which are located each node and each network to reach. First to have been implemented in the ARPANET, this technique quickly becomes onerous when the number of nodes increases since it must carry a lot of information of node in the node. Routing Information Protocol (RIP) is the best example of a protocol that uses a distance vector.
In this type of algorithm, each router broadcasts to its neighbors a vector listing each network that it can achieve with the associated metric, that is to say the number of hops. Each router can therefore build a routing table with the information received from its neighbors, but has no idea of the identity of the routers that are located on the selected route. Of this fact, the use of this solution poses many problems for the exterior routing protocols. It is in effect assumed that all routers use the same metric, which may not be the case between autonomous systems. In addition, an autonomous system may have particular reasons to behave differently from another autonomous system. In particular, if a stand-alone system has need to determine by what other Autonomous System will spend its messages, for reasons of safety for example, it may not know.
The algorithms to state of the links had initially aim to alleviate the faults of the routing by distance vector. When a router is initialized, it must define the cost of each of its links connected to another node. The node then broadcasts the information to all the nodes in the autonomous system, and therefore not only to its neighbors. From all of this information, the nodes can perform a calculation allowing them to obtain a routing table indicating the cost necessary to achieve each destination. When a router receives information that modify its routing table, it notifies all routers involved in its configuration. As each node has the topology of the network and the costs of each link, the routing can be seen as centralized in each node. The Open Shortest Path First (OSPF) protocol implements this technique, which corresponds to the second generation of Internet protocols.
The algorithms to state of links resolve the problems mentioned above for the Outdoor routing, but raise other. The various autonomous systems may have different metrics as well as specific restrictions, so that it is not possible to obtain a coherent routing. The dissemination of all information necessary to the whole of the autonomous systems can also quickly become unmanageable.
The objective of the algorithms to path vector is to compensate for the shortcomings of the first two categories by dispensing metrics and in seeking to know what network can be reached by what node and what autonomous systems need to be traversed to this. This approach is very different from that by the distance vector since the vectors of path does not take into account the distances nor the costs. In addition, the fact that each routing information list all the autonomous systems that must be traversed to reach the receiving router, the approach by path vector is much more directed to the systems of external routing. The Border Gateway Protocol (BGP) belongs to this category.

RIP (Routing Information Protocol)

RIP is the most commonly used protocol in the TCP/IP environment to route packets between gateways in the Internet network. It is a Interior Gateway Protocol (IGP), which uses an algorithm to

find the shortest path.
By path, we mean the number of nodes crossed, which must be between 1 and 15. The value 16 indicates an impossibility. In other words, if the path to go from one point to another in the Internet network is greater than 15, the connection may not be put in place. The RIP messages to prepare the routing tables are sent approximately every 30 seconds. If a RIP message is not reached its neighbor at the end of three minutes, the latter considers that the link is no longer valid, the number of links being superior to 15. The RIP protocol is based on a periodic dissemination of the States of the network of a router to its neighbors. The version of RIP2 has a routing by sub-network, the authentication of messages, the multicast transmission, etc.

OSPF

OSPF is part of the second generation of routing protocols. Much more complex than RIP, but at the price of superior performance, it uses a distributed database, which keeps in memory the state of the links. These information form a description of the topology of the network and the status of the nodes, which allows you to define the routing algorithm by a calculation of the paths the shorter.
The algorithm OSPF allows, from a node, to calculate the shortest path, with the constraints specified in the associated content to each link. OSPF routers communicate between themselves through the OSPF protocol, placed at the top of IP. Let us now look at this protocol in a little more detail.
The hypothesis of departure for the protocols to state of the links is that each node is able to detect the state of the link with its neighbors (on or off) and the cost of this link. It must give each node enough information to enable him to find the road the less expensive to all destinations. Each node must therefore have the knowledge of its neighbors. If each node to the knowledge of the other nodes, a complete map of the network can be drawn up. An algorithm based on the state of the Neighbors requires two mechanisms: the dissemination of reliable information on the state of the links and the calculation of the roads by summation of the accumulated knowledge on the state of the links.
A first solution is to achieve a reliable flood of information, so as to ensure that each node receives its copy of the information on the part of all the other nodes. In fact, each node floods its neighbors, which, in their turn, flood their own neighbors. More specifically, each node creates its own update packets, called Link-State Packet (LSP), containing the following information:

- The identity of the node that creates the LSP.
- List of Nodes neighbors with the cost of the associated link.
- Sequence number.
- Timer (Time To Live) for this message.

The first two information are necessary for the calculation of the roads. The last two have for objective to make reliable flood. The sequence number allows you to put in order the information that would have been received in the disorder. The Protocol possesses elements of error detection and retransmission.
The calculation of the road is performed after receipt of all the information on the links. From the complete map of the network and of the costs of the links, it is possible to calculate the best route. The calculation is performed using the Dijkstra algorithm on the shortest path.
In the acronym Open Shortest Path First (OSPF), the word Open indicates that the algorithm is open and supported by the IETF. Using the mechanisms listed above, the OSPF protocol adds the additional properties:

- **Authentication of messages of routing.** malfunctions can lead to disasters. For example, a node which, following the receipt of the wrong messages, voluntarily or not, or messages of an attacker amending its routing table, calculates a routing

table in which all nodes can be achieved at a cost of nil automatically receives all the packets on the network. These malfunctions may be avoided by authenticating the transmitters of the messages. The first versions of OSPF possessed a password for authentication on 8 bytes. The latest versions possess authentication much more strong.

- **New hierarchy.** This hierarchy must allow a better transition to the scale. OSPF introduces a level of additional hierarchy by partitioning the areas in the eras (areas). This means that a router to the interior of a domain does not need to know how to reach all networks in the domain. It is sufficient that it knows how to reach the good era. This leads to a reduction of information which must be transmitted to and stored.

There are several types of OSPF messages, but they all use the same header, which is shown in Figure 10.11.

The current version is the 2. Five types have been defined with values from 1 to 5. The source address indicates the sender of the message. The identification of the era indicates the era in which is located the node issuer. The type of authentication is the value 0 if there is no authentication, 1 in the case of password authentication and 2 if an authentication technique is implemented and described in the 4 following bytes.

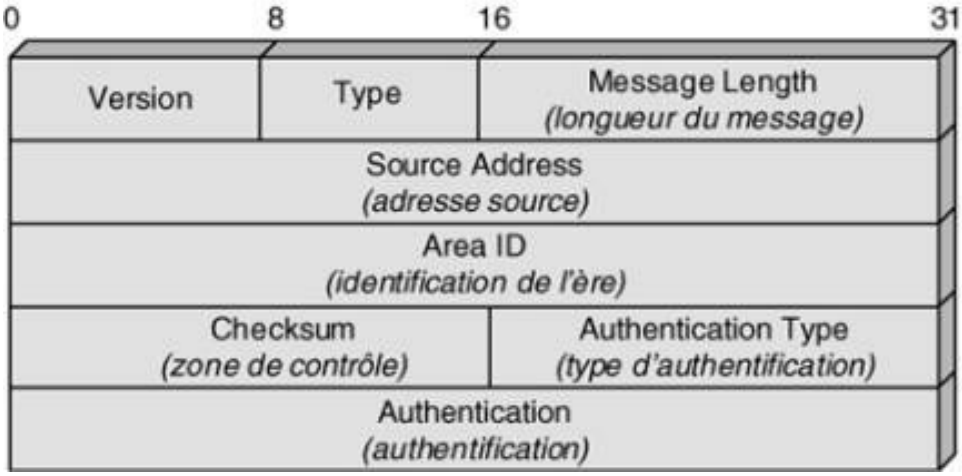


Figure 10.11

Format of the OSPF header

The five types of messages consist of the Hello message as type 1. This message is sent by a node to its neighbors for their indicate that it is always present and not in failure. The other four types are used to send information such as requests, shipments or the acquittals of the messages LSP. These messages transport primarily of Link-State Advertisement (LSA), i.e. information on the state of the links. An OSPF message can contain multiple LSA.

The Figure 10.12 shows an OSPF message of type 1 bearing a LSA.

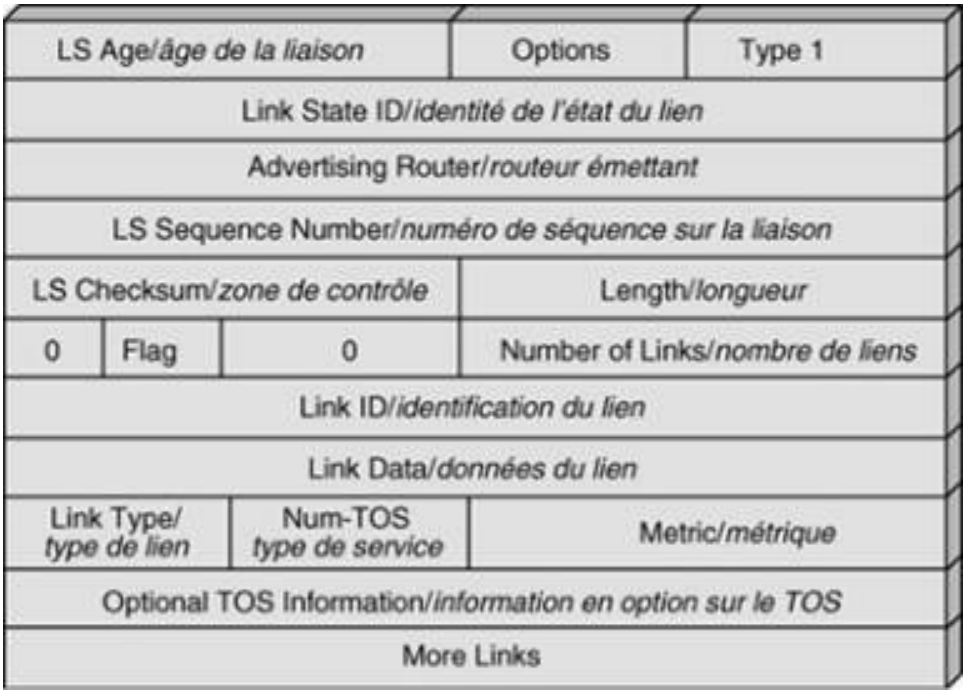


Figure 10.12

OSPF message bearing a LSA

This type indicates the cost of the links between the routers. The Type 2 is used to indicate the networks to which the issuer is connected. The types 3 and 4 are concerned with the indication of the eras. In Figure 10.12, the LS field Age is the equivalent of the TIMER TTL (Time To Live), if this is that the counter increases up to a certain preset value, whereas the TTL descends to 0.

The type here is the type 1. The two fields Link State ID and LS sequence number are identical and carry the identifier of the router which is issuing the message. The reason for this double field is to verify the identity of the router by two different means. The sequence number allows you to reséquencer received messages. The LS checksum allows to check the correction of the message. It takes into account the information from the field option. The Length field indicates the total length of the message. This are then information on the LSA which are transported: Link ID the ID each link, information about the link (link data) and Metric.

The field TOS (Type Of Service) allows the algorithm OSPF to choose the best route possible in function of the type of service. So there may be several metrics which depend on the type of service sought. The cost of the lines can also vary depending on the selected metric.

If the RIP protocol is adapted to the management of the routing in small networks, OSPF, applies to networks much more complex.

IS-IS

The algorithm IS-IS has been primarily developed by the ISO (ISO 10589). It describes a hierarchical routing based on the decomposition of communication networks in areas. In a domain, the different nodes indicate their state to routers is-is associated. The interdomain communications are performed by a routing to a point of access to the area determined by the routers responsible for external communications to the domain.

Interior Gateway Routing Protocol (IGRP)

Improved version of RIP, IGRP was designed by Cisco Systems for its own routers. It integrates the multipath routing, the management of Roads By default, the dissemination of information all the 90 seconds instead of all the 30 seconds, the detection of closures, that is to say of returns to a point by which the packet is already past, etc. This Protocol itself has been extended for a better protection against loops by the EIGRP protocol (Extended IGRP).

An Exterior Gateway Protocol (EGP)

EGP is the first routing algorithm to have been developed in the early 1980s, to route a packet from a standalone system to another.

It has three essential procedures, which allow the exchange of information. The first procedure applies to the definition of a neighboring gateway. This last being known, a second procedure determines the link that allows two neighbors to communicate. The third procedure concerns the exchange of packets between two neighbors connected by a link. The weaknesses of EGP have emerged with the exponential development of the Internet and the need to avoid certain areas politically sensitive.

Border Gateway Protocol (BGP)

To respond to the weaknesses of the EGP, a new algorithm has been launched by the IETF under the name of BGP. A first version, BGP-1, has been implemented in 1990, followed closely by BGP-2 and then BGP-3. At the end of a few years has been deployed BGP-4, which allows you to manage much more effectively the routing tables of large dimension by bringing together in a single line of several sub-networks.

BGP brings new properties compared to EGP, in particular that of managing the loops, which became common in EGP since this protocol is concerned only couples of neighbors, without taking into account the possible rebouclages by a third autonomous network.

The messages exchanged through BGP-4 are the following:

- OPEN : to open a relationship with a neighboring node.
- UPDATE : to convey information to the subject of a single road or ask for the destruction of roads that are no longer available.
- KEEPALIVE : to acknowledge the messages open and confirm that the neighbor relationship is still alive.
- NOTIFICATION : To send a message of error.

The three major procedures The following functional are defined in BGP:

- Acquisition of neighboring nodes;
- Possibility to reach the neighbor;
- Possibility to reach of the networks.

Two routers are considered as neighbors if they are in the same network. If the two routers are in areas separate autonomous, they may need to exchange routing information. For this, it must first achieve a acquisition of neighbors, that is to say, to allow two nodes that are not in the same autonomous system to exchange routing information. The acquisition must be made by a formal procedure because one of the two nodes may not want to exchange routing information. To perform the acquisition, a node sends the message open. If the remote router accepts the relationship, it returns a keepalive. Once a relationship is established, to maintain the active relationship the nodes exchange keepalive. Each node maintains a database of networks that it can reach and the route to arrive at these different networks. When a change occurs, the router broadcasts a message update to other routers, which allows the latter to update.

The Figure 10.13 illustrates the format of update package BGP.

Figure 10.13

Format of the package of BGP update

The Marker field is reserved for the authentication. The issuer may put a encrypted text which can be decrypted only by the receiver with the encryption key. Length gives the length of the message in byte. The message types are open, update, keepalive and notification.

To put in place a relationship of neighborhood, the router of departure initiates a TCP connection and then sends a message open. This message indicates the autonomous system in which the transmitter is located as well as the IP address of the router. It also includes a Hold Parameter TIME parameter, which specifies the number of seconds proposed for the Hold Timer Timer to determine the maximum time between two messages from the transmitter (Keepalive, update). If Remote the accepts the relationship, it calculates the minimum of its own Hold Timer Timer and that of the transmitter and sends it to the transmitter.

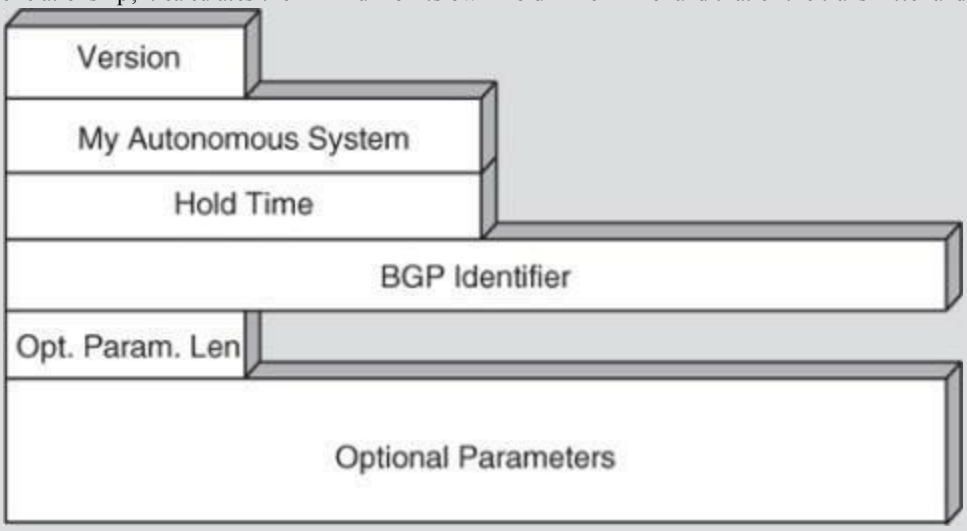


Figure 10.14

Fields of the open packet of BGP

The fields of the package open are shown in Figure 10.14. These fields are the following:

- Version: value on a byte indicating the version of BGP used (4 for BGP-4).
- My Autonomous System (my standalone system): VALUE ON 2 bytes indicating the autonomous system number of the transmitter.

- Hold Time (time restraint): field 2 bytes indicating the number of seconds that the issuer proposes for the counter of the retainer. The latter allows you to avoid the infinite closures in the autonomous systems. Once a device BGP receives a message open, it must calculate the value of the counter of restraint that is going to be used. For this, it chooses the smallest value between the holddown timer that it comes to receive in his message open and the value that has been configured for itself. The selected value is in fact the number of seconds between the reception of messages keepalive and update sent by the transmitter.
- BGP Identify (BGP ID): field of 4 bytes indicating the id BGP. This ID is based on the IP address assigned to the device BGP.
- Optional Parameters Length: field of a byte indicating the total size of the Optional Parameters field in byte. If the value is 0, it is that there is no optional parameters.
- Optional Parameters: field containing the list of optional parameters which are represented by triplets parameter type, Parameter Length and parameter value. The parameter field type uniquely identifies each optional parameter. The parameter field length specifies the size in bytes of the Parameter field value. The parameter field value is a variable size field (that is why its size is specified in the Parameter field length) containing the optional parameter itself.
- The keepalive message only takes into account the header of messages BGP. It must be issued often enough for that the Timer Hold Timer is not triggered.
- The update messages are used to route two types of messages:
- The information about a single route, which are recorded in the databases of information of the routers.
- The information about a list of roads that will be destroyed.

The notification messages are sent when an error has been detected. The following errors can be issued:

- MESSAGE HEADER ERROR : An error in the header of the message has been detected as a default of authentication or a syntax error.
- OPEN ERROR MESSAGE : An error in the syntax of the message open or a refusal of the value of the hold timer has been detected.
- UPDATE ERROR MESSAGE : An error in the message Update has been detected as a syntax error.
- HOLD timer expired : the Hold Timer Timer has expired, and the session of neighborhood is closed.
- FINITE State Machine Error : a procedural error has occurred.
- CEASE : To close a connection to another router in a circumstance that is not supported by the previous messages.

IDRP (Interdomain routing protocol)

The Estimates of departure provided that the Internet would be made up of tens of networks and of hundreds of machines. These figures have been multiplied by 10, 100, then 1 000 for the networks and by 1 000, 10 000 and 100 000 for machines. These gear reductions are not the only indicative of the success of the Internet. The measurements show that the flow which passes on the network exceeds very widely that represented by the set of the lyrics exchanged telephone in the world.

Such an explosion raises the question of the capacity of the mechanisms of routing put in place to bear the load. To reduce the risk of saturation and extend the current mechanisms, the immediate solution is to generalize the subnetting. The subnetting is to give a common address special, the mask, to the whole of the stations participating in the same logical network, even if the IP addresses of the stations of this logical network come from separate subnets. This allows the routing tables to grow more slowly.

In the environment IPv6, a new protocol, IDRP, the fruit of studies devoted to the routing between the areas of routing (Routing Domain) by the ISO, has been adapted to the world the Internet to perform routing between autonomous systems. The role of IDRP is slightly different from that of the protocols running inside a domain, because it defines a policy of routing between autonomous systems and not only a routing algorithm. The policy defined in this proposal led the routers in an autonomous system to agree, for example, do not go through a defined area or do not allow other autonomous systems to send IP packets to a standalone system determined. In other words, there must be a dialog between routers to provide only the indications corresponding to the policy defined.

The routing algorithms of type OSPF or RIP are applied by routers that all have the same goal: to find the best route possible, by minimizing the number of hops, either the time of crossing of the network, or by optimizing the transport capacity. These algorithms are based on the notions of weight: if the links have weight *or, the path is one in which the sum of the weights of the links crossed is the most low. The routing IDRP also has as an objective to find the good paths, but with restrictions for each autonomous system. The algorithm is based on distance vectors (Path Vector routing), which take into account the end-to-end path in addition to the weight to go to the neighboring nodes.*

As the number of autonomous systems can grow rapidly with the increase of the processing capabilities of the routers, it was decided to consolidate the autonomous systems in confederations. The Protocol IDRP works on the routing between these confederations. To convey the routing information, IDRP uses specific packages, brought in IP packets. In the IP area, the Next Header field contains the number 45 and indicates the protocol IDRP.

Network Address Translation (NAT)

The IP protocol version 4, that we use currently massively, offers a field of addressing limited and insufficient to allow any computer terminal to have an IP address. An IP address is indeed coded on a field of 32 bits, which offers a maximum of 232 possible addresses, either in theory 4 294 967 296 connectable devices to the same network.

To cope with this shortage of addresses, and pending the version 6 of the IP protocol, which will offer a number of addresses much more important on 128-bit, it is necessary to use a sharing of connection by using network address translation, or NAT (Network Address Translation).

This mechanism is frequently encountered both in business and among individuals. It distinguishes between two categories of addresses: the addresses say public, that is to say visible and accessible from anywhere (it is said also routable on the Internet), and addresses say private, that is to say non-routable on the Internet and only referable in a local network, to the exclusion of the Internet network.

The NAT is to establish relationships between the private addressing in a network and the public address to connect to the Internet.

Private addresses and public addresses

In the case of a network purely private and never led to connect to the Internet network, any IP address can be used. As soon as a private network can be led to connect on the Internet network, it is necessary to distinguish the private addresses public addresses. For this, each class of IP addresses has a range of addresses reserved, defined as private IP addresses and therefore not routable on the Internet. RFC 1918 summarizes these ranges of IP addresses, as indicated in the table 10.2.

Class of Addresses	Private Address Ranges	Network Mask	Addressable space
Has	10.0.0.0 to 10.255.255.255	255.0.0.0	On 24-bit, or 16 777 216 terminals
B	172.16.0.0 to 172.31.255.255	255.240.0.0	On 20 Bit, either 1 048 576 terminals
C	192.168.0.0. to 192.168.255.255	255.255.0.0	On 16-bit, or 65 536 terminals

Table 10.2 • Beaches of private addresses

In this framework, and before the introduction of the concept of NAT, the users who have a private IP address can communicate only on their local network, and not on the Internet, while with a public IP address, they can communicate on any IP network.

The private addressing can be used freely by any Administrator or user within its local network. On the contrary, addressing public is subject to restrictions of reporting and recording of the IP address from a specialised body, the Internet Assigned Numbers Authority (IANA), that the ISP perform overall in acquiring a range of IP addresses for their subscribers.

Figure 10.15 shows an example of a mixed addressing, in which one distinguishes the different possible communications, according to a mapping of type private or public.

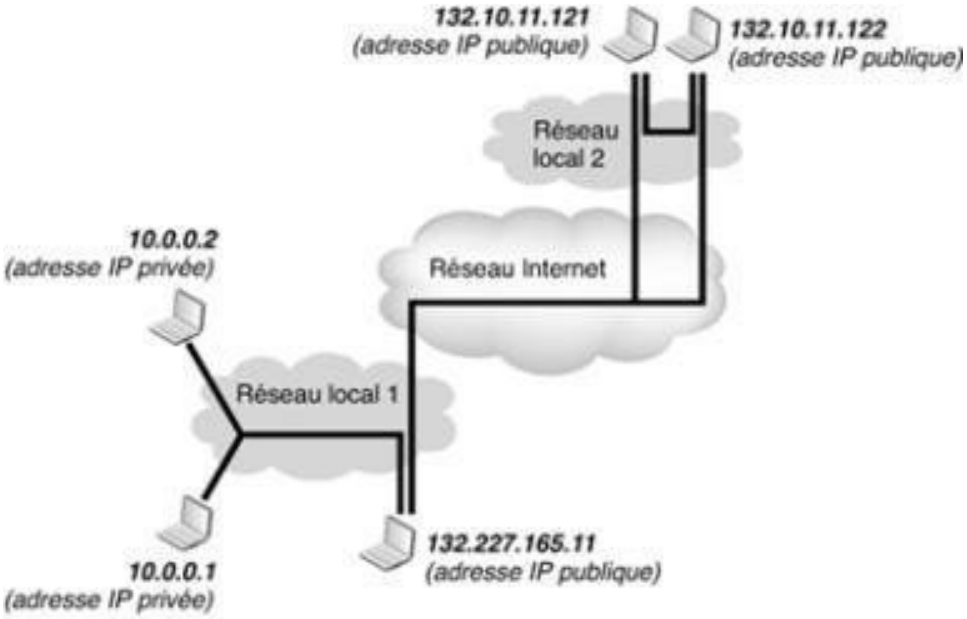


Figure 10.15
Private addresses Public and

Share a private IP address

Subject to the subscription of a internet access with an ISP, the latter provides its users with a private IP address. In a single household or company, two users may not communicate at the same time on the Internet With this single IP address provided. The private IP addresses generally agree to cover a private network, individual or company, but not to communicate directly with the public networks.

To resolve this problem and allow a device with a private IP address to communicate with the public network, the process of NAT fact intervene a third-party entity between a terminal, having a private IP address, and any other device with a public IP address. This mechanism is to insert a box between the Internet and the local network in order to perform the translation of the private IP address to a public IP address. Today, most of the enclosures, or home gateway, ISPS offer to their subscribers this feature. All of the machines that connect to it receive through the Service DHCP (Dynamic Host Configuration Protocol) A private IP address, that the unit is charging to translate to a public IP address.

Figure 10.16 shows an example in which a NAT gateway performs a translation of addresses for four terminals. This gateway has two network interfaces. The first is characterized by a public IP address (132.227.165.221). Connected to the Internet network, she is recognized and addressable normally in the network. The second interface is characterized by an IP address of a non-public (10.0.0.254). Connected to the local network, it can only communicate with devices that have an IP address of a non-public of the same class.

When a device with a private IP address attempts to connect to the Internet network, it sends its packets to the gateway Nat. It replaces the private IP address of origin by its own public IP address (132.227.165.221). This is called a translation of address. In this way, the terminals with a private IP address are recognized and addressable in the Internet network by a public IP address.

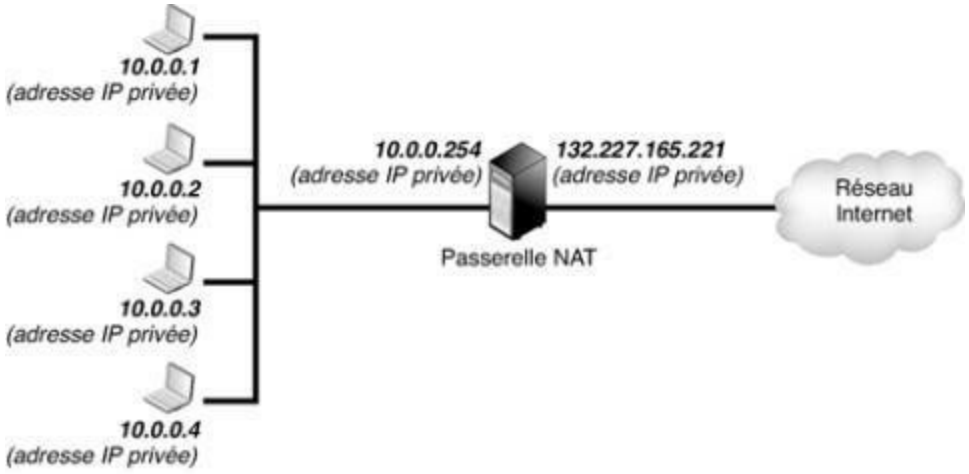


Figure 10.16
Translation of Addresses

The translation of address is of course carried out in both directions of a communication, in order to allow the issuance of requests as well as the receipt of the corresponding answers. For this, the unit NAT maintains a table of correspondence of the packets in a way to know which distribute the received packets.

For example, if an issuer whose IP address is 10.0.0.3 sends to the gateway Nat a packet from its port 12345, the NAT gateway modifies the packet by replacing the source IP address by the Siena and the source port by any port that it does not use, say the port 23456. It notes this correspondence in its table of NAT. In this way, when it will receive a packet to the destination of the port 23456, it will seek this port assignment in its table and return to the original source.

This case is shown in Figure 10.17.

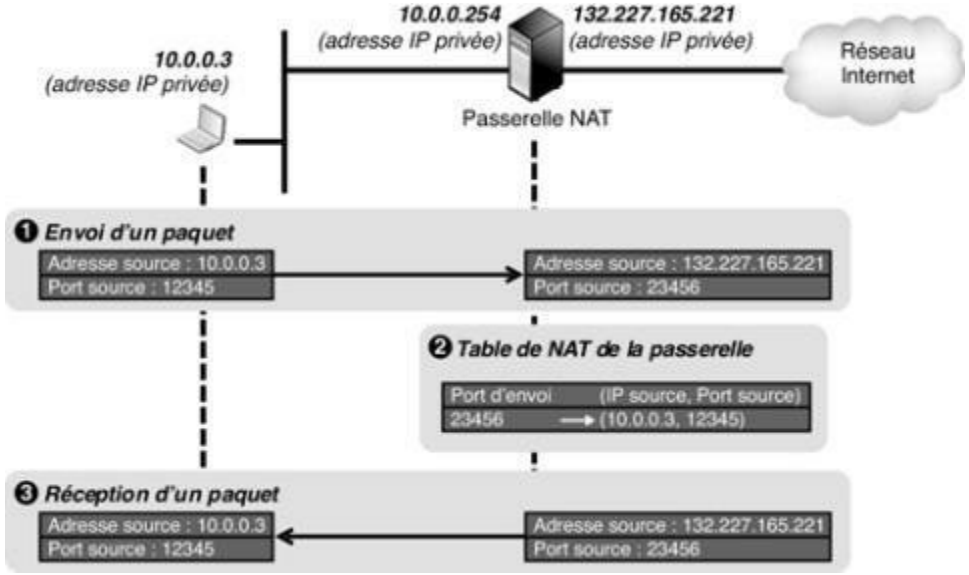


Figure 10.17
Modification of packets during the NAT

Benefits of NAT

The prime asset of the NAT is to simplify the management of the network, leaving the administrator free to adopt the plan of internal addressing that it wishes. Being private, the plan of internal addressing does not depend on external constraints, that administrators did not always master. For example, if a business uses an addressing plan public and for it to change ISP, it must change the address of all devices that make up its network. On the contrary, with the NAT and a plan of private addressing, the choice of a new Internet service provider has no impact on the terminals. In this case, the administrator does not need to reconfigure the IP addresses of all the terminals of its network. It is enough for him to modify, at the level of the NAT gateway, the pool of public IP addresses, which is

assigned dynamically to the private IP addresses of the devices on the local network.
The second asset of the NAT is to save the number of public IP addresses. The IP network protocol, which is used in the current Internet in its version 4, presents a significant limitation, because the number of IP addresses available is low compared to the number of devices that can be connected to the Internet network. As this resource is rare, its provision at a cost for the administrators who wish to benefit from it.
The NAT fills this shortage of own addresses to IP version 4 by offering the possibility to save the IP addresses at two distinct levels. All devices in a local network did not necessarily need to be reached from the outside, but may be limited to an internal connection to the network. For example, servers, intranet, corporate directories, servers dedicated to human resources with confidential information of personnel monitoring or even the servers of tests do not have to be contactable from the Internet network, but only internally within the company. Accordingly, these servers can be enough to a private IP address, which will never be "nattée" by the unit NAT since these servers receive requests, but do not emit ever.
A second level of economy of IP addresses service is operated using the mechanism mentioned in the previous section, which allows you to hide multiple terminals, each with a private IP address with a single public IP address, playing on the ports used. This method is very commonly used, because it imposes no requirement as to the number of devices that can potentially access to the Internet in the local network.
Another important advantage of the NAT is the safety and security. Terminals have in effect a supplementary protection, since they are not directly addressable from the outside. In addition, the Housing NAT offers the guarantee that all flows transiting between the internal network and the outside always go through him. If a device is poorly protected and does not have an efficient firewall, the network in which it connects can add the mechanisms of additional protection within the gateway Nat, since it represents a passage obliged for all flows. Overall, the Administrator concentrates the mechanisms for securing to a single point of control and centralized. This explains that, very often, the Nat enclosures are coupled with filtering firewall the flows.

LISP

The LISP protocol (Locator/Identify separation Protocol) has been developed to allow for the transport of virtual machines from one datacenter to another without that the VM (Virtual Machine) Exchange of IP address. For this, it is necessary to separate the two interpretations of the IP address: The identifier of the machine User (Identify) and the address used for the routing (locator). If we want to keep the same address of VM, it is necessary to differentiate between these two values. That is what the Protocol LISP, but this is not the only protocol to perform this separation: protocols HIP (Host Identity Protocol) and shim6 (level 3 Multihoming Shim Protocol for IPv6) The font also, but with mechanisms based on the Terminal Machines. These two protocols are detailed in Chapter 16.
In the LISP approach, the routers support the association between the address of the Machine Terminal, the Eid (Endpoint ID), and the address to the routing, the locator, or RLOC (Routing-Locator). At the start of the communication between a machine terminal and a server machine, the traffic is intercepted by a router to entry, RTID (ingress tunnel Router) which must determine the address to use for routing to access the server machine whose address is the one of the Eid. The output of the network to reach the EID is performed by the router eTR (Tunnel egress router). To perform this operation, the router s address to a directory service, the Eid-RLOC. Once obtained the address of Location of the recipient RLOC, the packet can be routed to this recipient. This process is

not visible from the machine and the server terminals.
It is to be noted that LISP allows the use of other addresses that those at level 3 IPv4 or IPv6, such as a GPS location or a MAC address.

The control protocols

Several protocols have been standardized for carrying packages of control. The sections that follow provide details Internet Control Message Protocol (ICMP) and IGMP (Internet Group Management Protocol), whose roles are to carry information on the anomalies of operation and to oversee the functioning of groups of users on the Internet.

Internet Control Message Protocol (ICMP)

In the transfer in mode without connection, described at the beginning of Chapter 2, each gateway and each machine operate independently. Similarly, the routing and the sending of datagrams are without coordination with the transmitter. This system works well as long as all the machines do not have a problem and that the routing is correct, but this is not always the case.
In addition to the network outages and machines, problems occur when a machine is disconnected from the network of a temporary or permanent basis, when the life of the datagram expires or when the congestion of a gateway is too important. To allow machines to account for these anomalies of operation, it was added to the Internet the protocol for sending ICMP control messages.
The recipient of an ICMP message is not a process application, but the Internet software of the machine. When a message is received, the IP addresses the problem. ICMP messages are not only sent by the gateways. Any machine on the network can send messages to any other machine. So, it was a single protocol for all control messages and information. The Table 10.3 summarizes the main ICMP control messages.

Code Message	Type of ICMP message
0	Echo reply
3	Destination Unreachable
4	Source quench
5	Redirect
8	Echo Request
11	Time exceeded for a datagram

Code Message	Type of ICMP message
12	Parameter problem it was Datagram
13	Timestamp Request
14	Timestamp Reply
17	Address Mask Request
18	Address Mask Reply

Table 10.3 • ICMP control messages

Each message has its own format, that is to say of the composition of the information fields of supervision. This format allows you to carry the appropriate information to account for the error. ICMP messages are carried in the Data field of the IP datagrams. The latter can be lost. In the event of error of a datagram containing a control message, no message of the report of the error is transmitted, in order to avoid an explosion of the number of packages transported in the network, called the avalanche.
ICMP takes much more importance with IPv6, the latest version of the IP protocol, where ARP (Address Resolution Protocol) is replaced by the function ND (Neighbor Discovery) of ICMP. This function allows a workstation to discover the router on which it depends and the hosts that it can

reach locally. The station built for this a knowledge base by examining the packets that pass through its intermediary and then takes of routing decisions and control. The correspondence of the IP address of a station with the local addresses is called address resolution. It is carried out by ND (Neighbor Discovery).
The station that uses ND emits a query neighbor solicitation on its line. The address of the recipient is a multicast address predetermined Type FF02::1:pruv:wxyz. This multicast address corresponds to a connection that part of a given point and is directed toward several points of destination. It is supplemented by the value pruv:wxyz of last 32 bits of the address of the station. In IPv6, the value of the Next-Header is 58 to indicate an ICMP message. The code of the ICMP message is 135 to indicate a query neighbor solicitation. If the station has no answer, it performs a new application. The stations which are recognized on the same line transmit to the sending station a Neighbor Advertisement. To discuss with a user on another network, the station has need to apply to a router. The Query Router Solicitation is used to this effect. The ND function allows the router Managing the station to know. The response message contains many options, such as the time of life of the Router - if the router does not give its new in this time, it is considered unavailable.
The messages Router Solicitation and Router Advertisement do not guarantee that the router that is known is the best. A router can be seen and send the packets of the station to another router through a redirect, in warning the workstation transmitter.
One last important function of ICMP Allows you to warn of the loss of communication with a neighbor. This function is provided by a query Neighbor Unreachability Detection.

IGMP (Internet Group Management Protocol)

The Internet defines groups of dissemination, formed of sets of machines participating in one and the same work, so that a message issued by a participant can reach the whole of the other participants. These groups of dissemination must be controlled for, for example, accommodate new customers or leave. The role of the IGMP protocol is to perform this test on the communications between the members of the group.
Distribution groups are dynamic. A machine can attach to a group or to exit at any time, the host should only be able to send and receive datagrams in multicast. This function ip for dissemination to the members of a group is not limited to a group located in a same sub-physical network. The Gateways also spreading the information of belonging to a group and manage the routing so that each machine receives a copy of each datagram sent to the group.
The machines shall communicate to the Gateways their belonging to a group using the IGMP protocol. The latter is designed to optimize the use of resources of the network. In most cases, the IGMP traffic consists of a periodic message sent by the gateway managing the multipoint and only one answer for each group of machines of a sub-network.
To achieve the communication to the inside of the group, the IP protocol Multicast is used. The latter manages the multipoint emissions toward the whole of the participants in a group and allows the sending of datagrams to multiple destinations at the same time effectively. IP uses the Class D address to indicate that it is sending to a multipoint.

Signaling protocols

The signage is an aspect that is particularly debated in the IP environment, since it is *a priori* contradictory with the philosophy of the world IP, which place a complete address in the packet in order to be able to router at any time this packet to its destination. The advantage of a standardized signaling and adopted by the entire IP community is to put in place the equivalent of

virtual circuits. On such circuits, the stream of packets can have a quality of service enabling him to accommodate demanding applications, such as the floor the phone.
The sections that follow present signaling protocols of the world IP, in the first rank of which RSVP (Resource ReSerVation Protocol). A specific group of the IETF, nsis (Next Steps In signaling), was created in 2002 to decide on the signaling to follow in the future World IP. This group has happened to first conclusions, which are detailed in Chapter 23, vested in the signage.

RSVP

RSVP seems to be the most interesting of the signaling protocols of new generation. Its role is to warn the intermediate nodes of the arrival of a stream corresponding to the qualities of service determined. By itself, RSVP does not explicitly launch the reservation of resources at the request of an application and then release these resources at the end.
The signaling is performed on a stream (flow), which is sent to one or more receivers. This stream is identified by an IP address or a destination port, or a reference of flow, or flow-label, in IPv6.
From the point of view of the network operator, the RSVP is linked to a reservation, which must be performed in the nodes of the network on a particular route or on the specified routes by a multipoint. The difficulties encountered in implementing this mechanism are of two orders: to determine the quantity of resources to book at any moment and reserve the resources on a single road, given that the routing of IP packets fact vary the path to follow.
RSVP makes the reservation from the receiver, or of receptors in the case of a multipoint. This may seem surprising at first sight, but this solution adapts to many cases of figure, in particular the multipoint. When a new item is added to the multipoint, this last can achieve the addition of booking a simpler way that could do the transmitter.
The RSVP packets are transported in the data area of the IP packets. The upper part of Figures 10.18 to 10.20 illustrates the headers of IPv6. The value 46 in the Next Header field indicates that a RSVP packet is transported in the data area.
In addition to two fields reserved, the RSVP packet contains the eight following fields:

- **Version.** indicates the version number of the RSVP.
- **Flags.** Four bits are reserved for later use.
- **RSVP type.** The type characterizes the RSVP message. Currently, two types are more specifically used: the message of path and the message of the reservation. The values that have been selected for this field are the following:

1 Path message
2 reservation message
3 error indication in response to PATH message
4 error indication in response to reservation message
5 Path teardown message
6 reservation teardown message

- **Checksum.** Allows you to detect errors on the RSVP packet.
- **The length of the message.** Indicated On 2 bytes.
- **Reserved.** First field reserved for subsequent extensions.
- **The identifier of the message.** contains a value common to the whole of the fragments of the same message.
- **Reserved.** Other field reserved for subsequent extensions.

- **More fragment.** bit indicating that the fragment is not the last. A zero is placed in this field for the last fragment.
- **Position of the fragment.** Indicates the location of the fragment in the message.

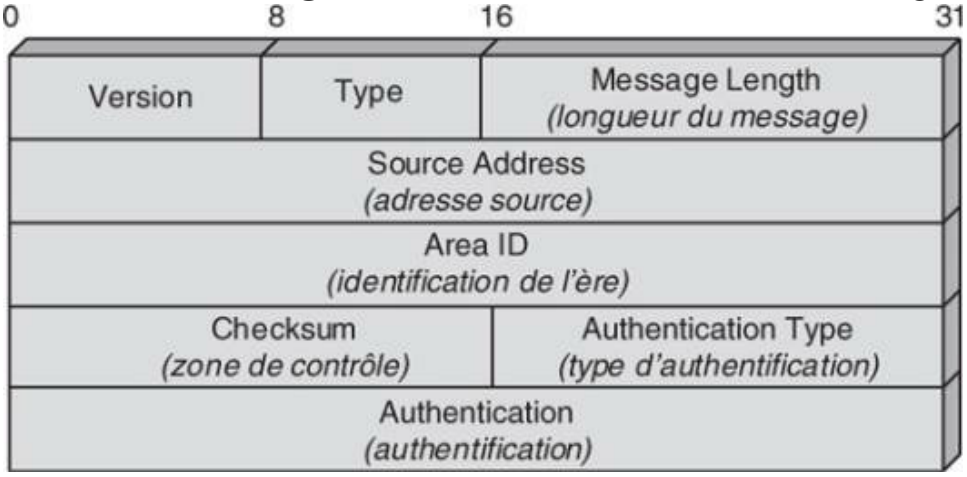


Figure 10.18
Format of the RSVP message

The message portion RSVP brings together a series of objects. Each object is present in the same way, with a field length of the object on 2 bytes and then the number of the object on a byte, which determines the object, and finally a byte to indicate the type of object.
Table 10.4 summarizes the Fifteen defined objects.

Number Subject	Type	Description
0	NULL	Ignored by the receiver
1 Session	1	Session IPv4 (destination of flow)
	2	Session IPv6
3 RSVP_Hop	1	Address of the next node (IPv4)
	2	Address of the next node (IPv6)
4 Integrity		Authentication Data
5 TIME_VALUES	1	Refresh frequency
6 error_spec	1	Error Information for IPv4
1 Session	2	Error Information for IPv6
7 SCOPE	1	List of IPv4 hosts on which the reservation is exercised
	2	List of IPv6 hosts on which the reservation is exercised
STYLE 8	1	Style of reservation
9 FLOWSPEC	1	Specification of a stream Requesting a control the time delay
	2	Specification of a stream requesting a quality of service
	3	Specification of a stream requesting a guaranteed quality of service
	254	Specification of a stream containing several sub-Flots
10 filter_spec	1	Filter for IPv4 to apply the flow
	2	Filter for IPv6 using the values of the source port

	3	Filter for IPv6 using the values of the labels of the waves
11 sender_TEMPLATE	1	Description of the flood IPv4 that the transmitter generates
	2	Description of the flood IPv6 that the transmitter generates
12 sender_TSPEC	1	Top terminal on the traffic generated by the transmitter
13 ADSPEC		Information warning of a stream emitted by the transmitter
14 policy_Data	1	Information on the policy followed by a stream
	254	Information on the policies followed by several waves
20 TAG	1	Collection of objects associated with a given name

Table 10.4 • objects of RSVP

The specifications of RSVP contain the specific descriptions of the paths followed by the messages, including the necessary objects and the order in which these objects appear in the message.

The Figures 10.19 and 10.20 Give examples of RSVP messages. The figure 10.21 describes as a complement to the format of the indication of the errors in RSVP.

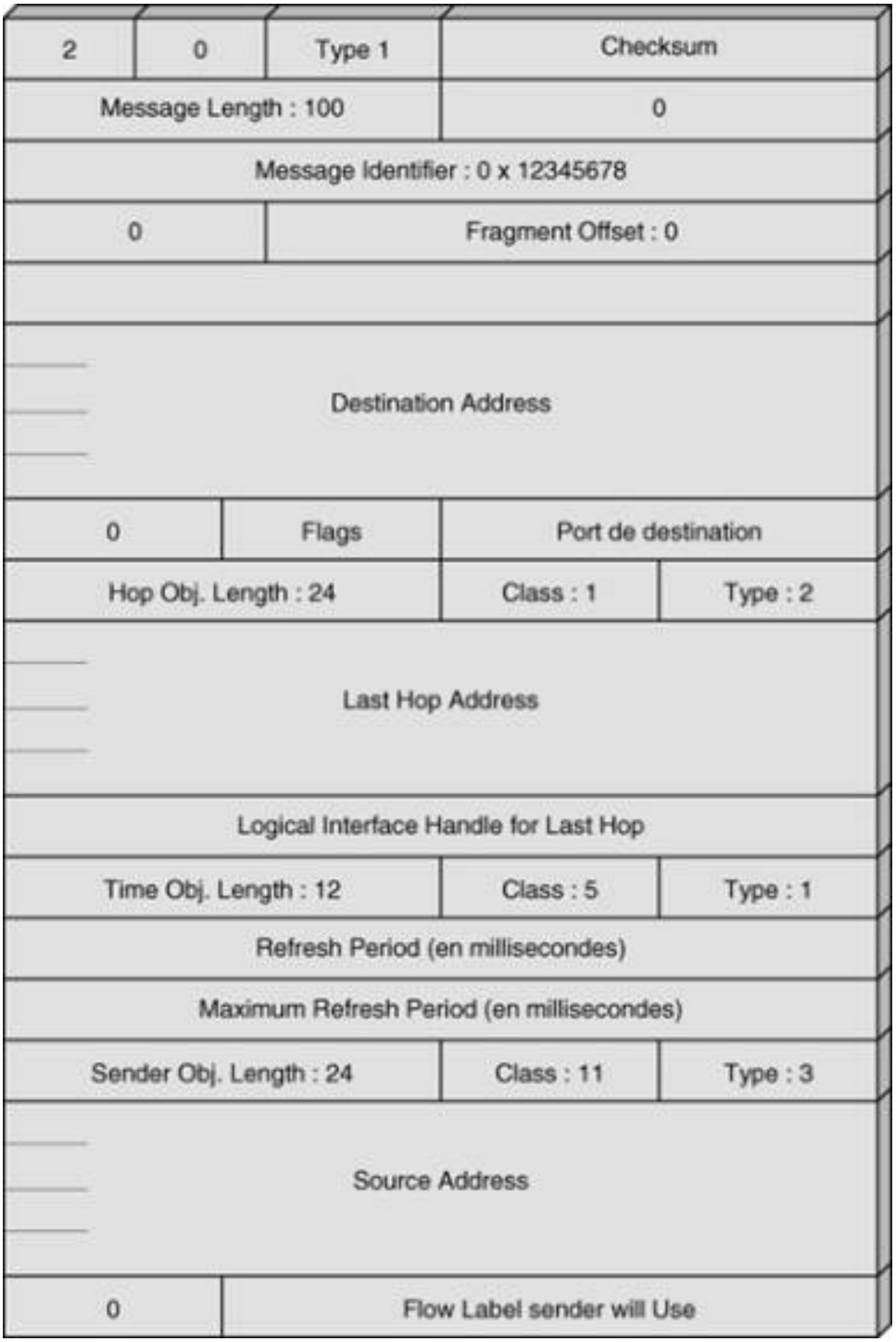


Figure 10.19
Message to determine the path RSVP

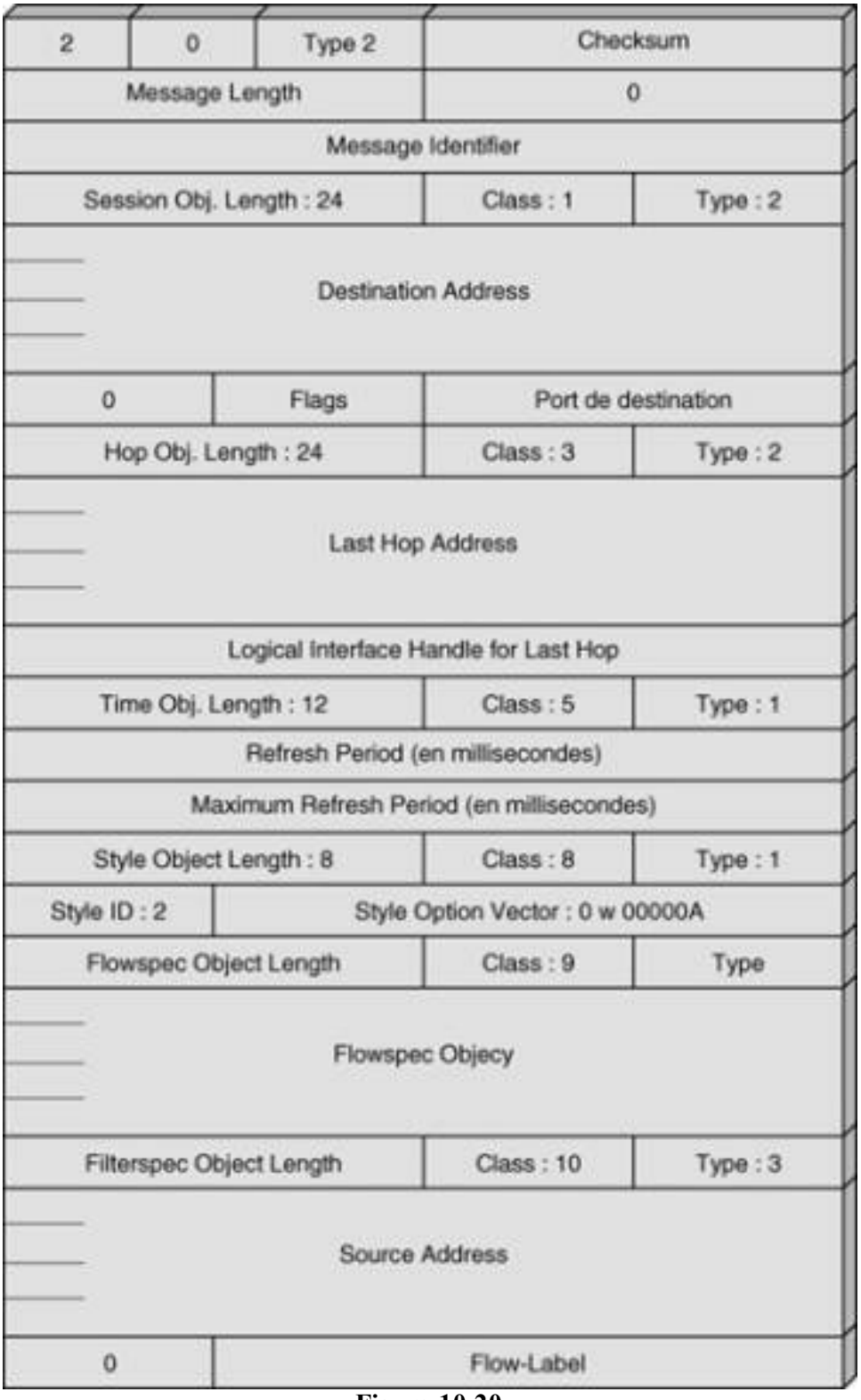


Figure 10.20

Package of RSVP reservation

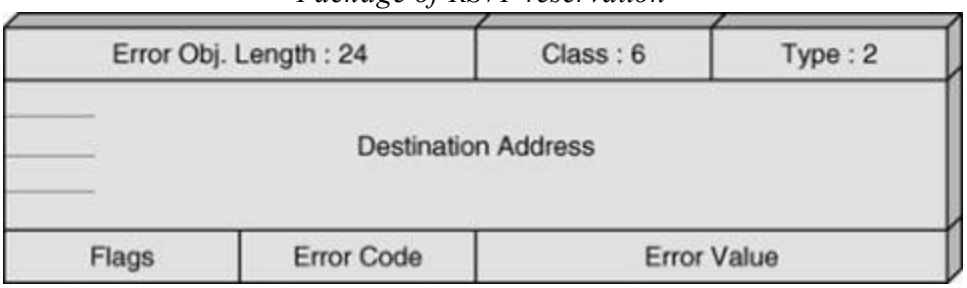


Figure 10.21

Format of the indication of the errors in RSVP

As explained previously, RSVP is not a protocol of reservation, but a signaling, which allows nodes to do their best in relation to the knowledge they have of the main waves that will cross. It must then

apply a policy of scheduling of packets, of type fair queuing, to optimize the different qualities of service required on the waves.

RTP (Real-time Transport Protocol)

The support for real-time applications, such as the floor the phone or video conferencing, is the challenge to the Internet. These applications request of the qualities of service that classical protocols of the Internet can offer. RTP was designed to try to resolve this problem, which is more directly in a multipoint environment, taking at its charge as well the management of the real time that the administration of the session multipoint.

To achieve this, two intermediaries are necessary, of translators and of mixers. A translator has the function to translate an application coded in a certain format to another format, better adapted to the passage by a sub-network. For example, a videoconferencing application coded in MPEG could be decoded and recoded in H.261 to reduce the quantity of information transmitted. A blender has a role to consolidate multiple applications corresponding to several separate streams in one stream keeping the same format. This approach is particularly interesting for the flow of digital words.

To achieve the transport in real time, a second protocol, RTCP (Real-Time Control Protocol), has been added to RTP since the RTP packets carry only the data of users and non-the information of supervision.

The RTCP protocol accepts the five types of packages:

- 200: sender Report (Report of the transmitter);
- 201: receiver report (report of the receiver);
- 202: Description Source (description of the source);
- 203: Bye (goodbye);
- 204: Application Specific (application specific).

These different packages give to the nodes of the network the necessary instructions to a better control of real-time applications.

The quality of service

The quality of service is a necessary condition to the passage of the multimedia in IP networks. The reflections carried out on the TCP/IP architecture to go in this direction are numerous.

The IETF has first proposed the ISA architecture (Integrated System Architecture), which includes protocols such as IPv6 and RSVP. For this architecture to be effective, it must be placed in a network intranet, where an assignment of costs in relation to the requested service is possible. In the contrary case, the set of users quickly acquires the highest priority.

This architecture is based mainly on the knowledge of what is being transported in the IP packet. In IPv4, the information is located in the field TOS (Type of Service). In IPv6, it uses the priority zone and possibly the flow-label. Routers can take into account this information and treat the following packages a scheduling predetermined. Several solutions to the authorisation can be selected, one of the most classic is the fair queuing.

The Fair queuing is to place incoming clients in multiple queues according to their priority and to treat them in an order that satisfies all the better. If one takes the clients strictly in their order of priority, the last to be served are likely to have a quality of service disastrous, even though a priority packet could sometimes wait without that the disadvantage.

The IETF has made many other proposals in recent years to introduce the quality of service in IP networks. Are not described here that the two most important, IntServ and DiffServ.

The IETF proposes the use of two major categories of services, which are divided into sub-services with different qualities of service: integrated services IntServ (Integrated Services) and the differentiated services Differentiated Services (DiffServ). The integrated services are managed independently of each other, while the differentiated services bring together several applications simultaneously. The first solution is often chosen for the access network and the second for the interior of the network when there is a lot of waves to manage.
The integrated services IntServ have the three following classes:

- The guaranteed service (Guaranteed Service), which is the equivalent of the CBR and VBR-rt of the ATM.
- The service controlled (Controlled Load), which is the equivalent of the ABR Service with a guaranteed minimum (guaranteed minimum cell rate).
- The best-effort, which is the equivalent of the UBR or of the GFR.

The differentiated services, or DiffServ, have the three following classes:

- The guaranteed service (expedited forwarding), or Premium Service, which is the equivalent of the CBR and VBR-rt of the ATM.
- The service controlled (Assured Forwarding), which is the equivalent of the ABR service with guaranteed minimum (guaranteed minimum cell rate).
- The best-effort, equivalent to the UBR or to the GFR.

IntServ (Integrated Services)

The IntServ service integrates two different service levels with guarantees of performance. It is a service-oriented flow, that is to say that each flow can make its specific request of quality of service. To obtain a guarantee precise, the Working Group IntServ has considered that only a reservation of resources was able to make for sure the means to ensure the request.
As explained previously, three sub-types of services are defined in IntServ: a service with a total guarantee, a service with a partial guarantee and service best-effort. The first corresponds to the Rigid services with strong constraints to comply, and the second and third to the so-called services elastic, who do not have strong constraints.
When they receive a request *via* the RSVP, the routers can accept or refuse. This request is carried out as for the RSVP protocol from the receiver to the Transmitter after a phase to go. Once the application is accepted, the routers place the corresponding packets in a queue of the class of service requested.
The IntServ service must have the following components:

- A procedure of signage to warn the nodes traversed. The RSVP protocol is supposed to perform this task.
- A method to indicate the demand for quality of service for the user in the IP packet in order that the nodes can be taken into account.
- A traffic control to maintain the quality of service.
- A mechanism to move the level of quality to the underlying network, if it exists.

Guaranteed service GS affects a top terminal to the time limit for routing. For this, a protocol of reservation as RSVP is necessary. The booking request consists of two parts: a specification of the quality of service determined by a FlowSpec and a specification of packages that must be taken into account by a filter, the FilterSpec. In other words, some packages from the stream can have a quality of service, but not necessarily the other. Each flow has its quality of service and its filter, which can be fixed (fixed filter), shared with other sources (shared-explicit) or specific yet (wildcard filter).

The service partially guaranteed CL (controlled load) must guarantee a quality of service to almost equal to that offered by a network little loaded. This class is used primarily for the transportation of elastic services. The transit time of the in the network of the waves CL must be similar to those of clients of a class best-effort in a network very little loaded. To arrive at this fluidity of the network, it must integrate a control technique.
The two services must be able to be claimed by the application *via* the interface. Two possibilities are mentioned in the proposal IntServ: the use of the GQoS specification Winsock2, which allows the transport of applications Point-to-point and multipoint, and RAPI RSVP (API), which is an Application Interface on UNIX.
The scheduling of packets in the routers is a second mechanism necessary. One of the more classically proposed is the WFQ (Weighted Fair Queuing). This algorithm placed in each router requests a queued packets depending on their priority. The queues are served in an order determined by a scheduling dependent on the operator. Generally, the number of packets served at each passage of the server depends on the setting of the weight of the queue.
There are many solutions to manage the way in which the service is assigned to queues, generally based on levels of priority. These include the algorithm Virtual clock, which uses a virtual clock to determine the time of emission, and QRC (Self-Clocked Fair Queuing), who works on intervals of time between minimum two emissions of packages of the same class, interval depending on the priority.
The service IntServ poses the problem of the transition to the scale, or scalability, which refers to the possibility to continue to behave well when the number of waves to manage becomes very large, as is the case on the Internet. The control IntServ is doing on the basis of individual streams, the routers in the network IntServ must indeed keep in memory the characteristics of each stream. Another difficulty concerned the treatment of the different waves in the Nodes IntServ: Which stream treat before such other when thousands of waves arrive simultaneously with classes and associated parameters separated?
In the absence of recognized solution to all these problems, the second large control technique, DiffServ, tries to sort the waves in a small number of well defined classes, Multiplexing the waves of the same nature in the waves more important, but always in a limited number. IntServ may however apply to small networks such as the networks of access. Other research toward management processors specialized in the quality of service have recently resulted on equipment capable of processing several tens or even hundreds of thousands of waves.
The Working Group ISSLL (Integrated Services over specific link layers) of the IETF seeks to define a model IntServ acting on a level type frame ATM, Ethernet, Frame Relay, PPP, etc. In other words, the objective is to propose mechanisms to move the level of priority of the class to the classes sometimes not equivalent and to choose in the underlying network of algorithms likely to give a result equivalent to that which would be obtained in the IP world.

Differentiated Services (DiffServ)

The main objective of DiffServ is to propose a general diagram to deploy the quality of service on a large IP network and achieve this deployment quite quickly.
DiffServ separates the architecture in two major components: the technique of transfer and the configuration of the parameters used in the transfer. This concerns both the treatment received by the packets during their transfer to a node that the management of queues and the discipline of Service. The configuration of all the nodes of the path is carried out according to a manner called PHB (Per-

Hop Behavior). These PHB determine the different treatments corresponding to the waves which have been differentiated in the network.
DiffServ defines the general semantics of PHB and not the specific mechanisms that allow them to implement. The PHB are defined once and for all, while the mechanisms can be modified and improved, or even be different depending on the type of the underlying network.
The PHB and the associated mechanisms need to be able to easily be deployed in IP networks, which requests that each node can manage the waves through a number of mechanisms, such as the scheduling, the implementation form or the loss of packets traveling through a node.
DiffServ aggregates the waves in classes, called aggregates, which offer qualities of specific service. The quality of service is ensured by treatments carried out in the specified routers by an indicator located in the IP packet. The points of aggregation of incoming traffic are generally placed at the entrance of the network. The routers are configured through the DSCP field (DiffServ Code points) of the IP packet, which forms the first part of a field more general called DS (Differentiated Service) and containing also a field Cu (currently unused). In IPv4, this field DS is taken on the area TOS (Type of Service), which is to this fact redefined by report at its first use. In IPv6, this field is located in the zone tc (Traffic Class) of the class of service.
The Figure 10.22 shows the DS field in the IPv4 and IPv6. The DSCP field takes place in the field TOS (Type of Service) of IPv4 and in the field Traffic Class of IPv6. The DSCP field takes on 6 of the 8 bits and is complemented by two bits CU. The DSCP determines the class of service PHB.

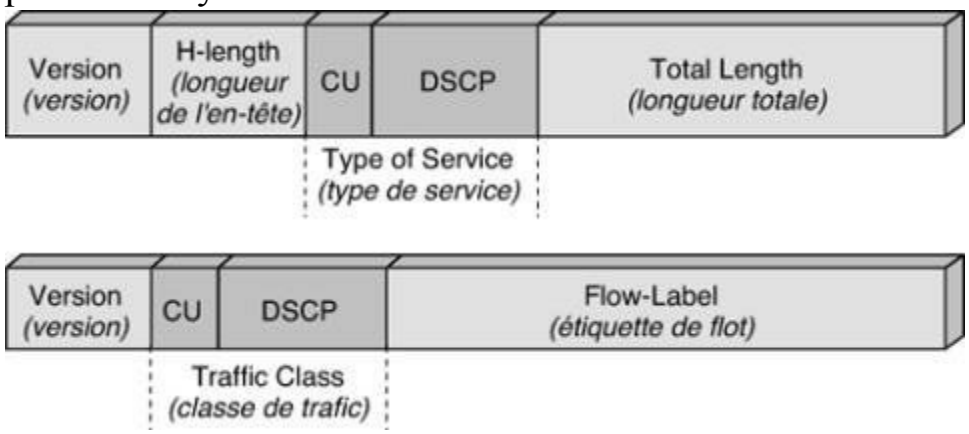


Figure 10.22
DSCP field IPv4 and IPv6 packets

The field of 6 bits of the DSCP must be interpreted by the node to assign to the packet the treatment corresponding to the class PHB indicated. The two bits CU must be ignored during the treatment in a DiffServ node standardized. By the intermediary of a table, the values of the DSCP determine the PHB acceptable by the node. A default value must always be indicated when the DSCP field corresponds to no PHB.
The operators of telecommunications can define their own values of DSCP to a PHB given, in place of the value recommended by the standardization of the IETF. These operators must however provide in the gateways to output the standard value of the DSCP so that this field to be interpreted properly by the next operator. In particular, a DSCP not recognized must always be interpreted by a default value.
The definition of the structure of the field DS is incompatible with that of the ToS field of the RFC 791 which defines IPv4. This TOS field had been designed to indicate the criteria to focus in the routing. Among the criteria specified are delay, reliability, cost and safety.
In addition to the service be (best-effort), two PHB quite similar to those of IntServ are defined in DiffServ:

- EF (expedited forwarding), or guaranteed service, also called Premium Service or first.
- AF (Assured Forwarding), or insured service, also sometimes referred to as the Assured service or Olympic service or the Olympic Games.

There are four sub-classes of services in AF determinant of the rate of loss acceptable for the waves of packets considered. We can classify them in platinum (platinum), gold (gold), Silver (money) and Bronze (bronze). As this terminology is not standardized, it is possible to meet other. Within each of these classes, three sub-classes sorted according to their degree of priority in relation to each other are defined. The class AF1x is the highest priority, then comes the class AF2x, etc. so there is total of twelve standardized classes, but little of operators implement them. As a general rule, the operators meet three basic classes of service AF, and therefore of five classes total by adding the services EF and BE.

The values brought by the DSCP field associated with these different classes are illustrated in Figure 10.23. For example, the value 101110 of the DSCP field indicates that the packet is of type EF (expedited forwarding). The class Best Effort focuses the value 000000.

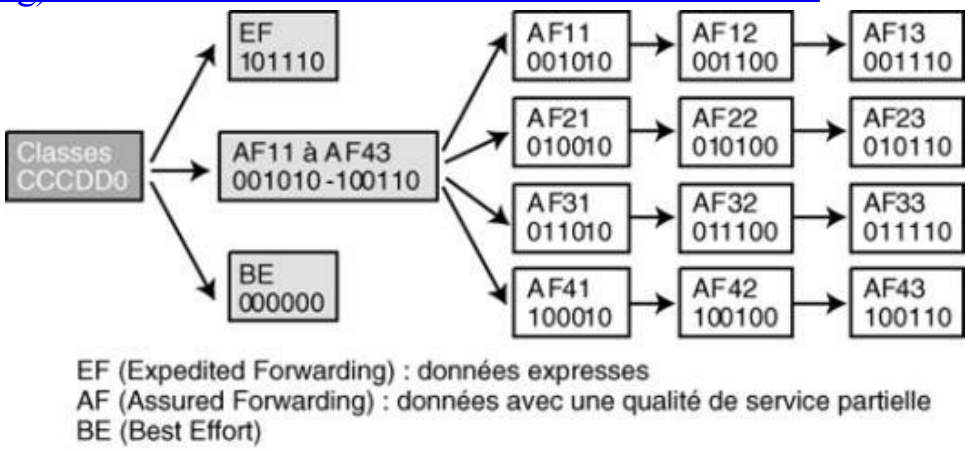


Figure 10.23

Classes of service of DiffServ and values of the DSCP fields associated

The DSCP 11x000 is reserved to classes of customers even more priority than the class EF. For example, it can be used for signaling packets.

EF (expedited forwarding)

The PHB EF (expedited forwarding) is defined as a transfer of packets for an aggregation of waves from DiffServ nodes such as the service rate of packets of this aggregate is higher at a rate determined by the operator. The EF traffic must be able to receive a rate of service independently of the other traffic circulating in the network. In terms even more specific, the rate of traffic EF must be greater than or equal to the rate determined by the operator measured on any interval of time at least equal to the size of an MTU (Maximum Transmission Unit). If the PHB EF is implemented through a mechanism on the priority of the other traffic, it is necessary that the traffic rate of the aggregate EF does not exceed a limit which would be unacceptable for the PHB of other classes of traffic.

Several types of mechanisms of scheduling can be used to respond to these constraints. A priority queue is the simplest mechanism to achieve the service (PHB) EF as long as there is no other queues more priority to preempt the packets EF of more than a packet for a proportion of time determined by the rate of service packages of the aggregate EF. It is possible to use a queue in a normal group of queues managed by a mechanism of tower of role with weight (Weighted Round Robin) or to use a sharing of the bandwidth of the output queue of the node, allowing the queue EF to achieve the rate of service guaranteed by the operator. Another potential mechanism, called SHARE CBQ (Class Based

Queuing), gives to the queue EF a sufficient priority to obtain at least the rate of service guaranteed by the operator.
The traffic Expedited Forwarding corresponds to the traffic sensitive to delay and jitter. It is equipped with a high priority in the nodes, but must be controlled to ensure that the sum of trafficking from different sources and passing on a same link does not exceed the rated capacity determined by the operator.
Several solutions allow you to reserve bandwidth proposed to the waves of packets EF. A type of protocol RSVP, for example, can carry out the reservations of bandwidth required. Another solution is to use a dedicated server in the distribution of the bandwidth, or bandwidth Broker. This server of bandwidth performs the admission control in proposing a centralized booking.

AF (Assured Forwarding)

The PHB AF (Assured Forwarding) ensure the transfer of IP packets for which a certain quality of service can be guaranteed. Trafficking AF are subdivided into *n* AF classes separate. In each Class A IP packet is assigned a rate of loss maximum and a priority to the loss, corresponding to classes of priority. An IP packet which belongs to the class AFi and which has a rate of loss corresponding to the priority I is marked by a DSCP AFij (see Figure 10.24). As explained previously, twelve classes are defined for DiffServ, corresponding to the four AF classes with guarantees on the loss of packets. The four classes corresponding to the rate of loss guarantee are called Platinum, Gold, Silver and Bronze, each class with three priority levels.
The packets of a class AF are transferred independently from those of other AF classes. In other words, a node can not aggregate streams with different DSCP In a common class.
A DiffServ node must allocate a set of minimum resources to each PHB AF so that they can fill the service for which they have been put in place. A class AF must possess the minimum resources in memory and bandwidth for a rate of minimum service, determined by the operator, can be achieved on a time scale potentially quite long. In other words, on a relatively long time interval, which can be Count in second, a guarantee of flow must be provided to the services AF.
A node AF must be able to be configured to allow a class AF to receive more resources to transfer that the minimum when additional resources are available in the network. This additional allocation is not necessarily proportional to the level of the class, but the operator must be able to reallocate the resources released by the class EF on the PHB AF. The priorities must nevertheless be respected, a class of priority best not to lose more of packages that a class with a lower priority, even if the loss remains below the permissible level.
A domain implementing the services AF must, by the intermediary of the border routers, be able to check the entries of trafficking AF for that the qualities of service determined for each class AF are met. The Routers border must for this put in place mechanisms for the format of the Traffic (Shaper), destruction of packets (dropper), to increase or decrease the loss of packets by class AF and rewiring of trafficking AF in other AF classes. The actions of scheduling must not cause of discount in order of packages of a same microflot, a microflot being a particular stream to the inside of an aggregate of a class of PHB.
The implementation of a strategy AF must minimize the rate of congestion in the interior of each class, even if congestion of short duration are eligible as a result of overlays of continuous streams of packets (bursts). This request an algorithm dynamic management in each node AF. An example of such an algorithm is red (Random Early discard). The congestion in the long term must also be avoided thanks to packet loss corresponding to the levels of priority, and that in the short term through the queues to hold some packages. The Algorithms for the format of the traffic must also be able to

detect the packets likely to cause congestion in the long term.
The basic algorithm to perform the control of traffic AF is WRED, or Weighted RED. It is to try to maintain the network in a fluid state. The loss of packets must be proportional to the length of queues. In other words, the packets in surplus and then the normal packets are eliminated as soon as the traffic is more fluid. The elapsed time since the last loss on a same aggregate is taken into account in the algorithm. The procedure tries to distribute the control to all of the nodes and no longer at the single node congested. The algorithms of destruction of packets must be independent of the short term and microflots, as well as of the microflots inside the aggregates.
The interconnection of AF services can be quite difficult to make because of the relative vagueness of the service levels of the different operators interconnecting.
A solution to allow the crossing of an aggregate in an IP network non-compliant with DiffServ is to achieve a tunnel with a quality of service greater than that of the PHB. When an aggregate of packets AF uses the tunnel, the quality of service provided by the latter must allow the PHB in the basis of be respected at the exit of the tunnel.
A client who requests a traffic Assured Forwarding must negotiate an approval of service, or ALS, corresponding to a profile determined by a set of parameters of quality of service, or SLS. The SLS indicates a rate of loss, and for the services ef, a average response time and a jitter of response time. The traffic not falling in the profile is destroyed in priority if a risk of congestion exists which would not allow The conforming traffic to reach its quality of service.

Architecture of a DiffServ node

The architecture of a DiffServ node is shown in Figure 10.24. It includes an entry containing a classifier (classify), whose role is to determine the right path to the inside of the node. The branch line chosen depends on the class detected by the classifier.
Next come the organs called meter, or quantity surveyor. A quantity surveyor determines if the package has the performance required by its class and decides the result of the treatment. The Quantity Surveyor knows all the queues of the node as well as the parameters of quality of service requested by the aggregate to which belongs the packet. It may decide for the eventual destruction of a package, if its class allows, or sending it to an output queue. The DiffServ node can also decide to change this package of class or well to the multiplexer with other waves (see later). The dropper body, or suppressor, may decide to lose or not, that is to say to destroy or not the packet, whereas the absolute suppressor (absolute dropper) automatically eliminates the packet.
In other words the quantity surveyor (meter) may take a decision of destruction and send the packet in a absolute suppressor (absolute dropper), whereas the quantity surveyor (meter) does that determine the performance parameters and leaves the suppressive (dropper) care to destroy or not the next packet other criteria that the gross measure of performance.
For some packages, such as the Packages BE (Best Effort), it is not necessary to ask the question of the performance since there is no warranty on the aggregate. It is enough to know if the packet should be lost or not. This corresponds to the branch of on the Figure 10.7. On this same figure, the first branch (A) corresponds to customers EF or premium, the following two (B and C) to customers af, with Gold customers in the path top and silver or bronze in the other path, and the last branch (D) to customers be.
The architecture of a DiffServ node ends by queues intended to put on hold the packets before their issuance on the output line determined by the routing. An algorithm of priority is used to treat the order of emission of the packets. The scheduler (Scheduler) is responsible for this function. The algorithm the most simple returns to treat the files according to their order of priority and not to let

pass the clients to another queue as long as there is still of customers in a Priority Queue.

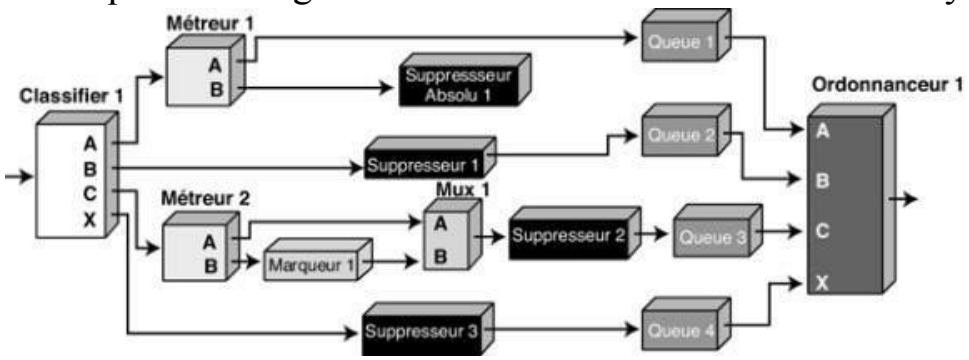


Figure 10.24

Architecture of a DiffServ node

Many other algorithms allow to give a specific weight to queues in such a way that a customer not a priority can be served before a customer priority. Among these algorithms include WFQ (Weighted Fair Queuing), in which each queue has a weight, for example 70 for the queue EF, 20 for the queue AF Gold and 10 for the other queue AF. The scheduler (scheduler) leaves go during 70% of the time customers EF. If these clients exceed the use of 70%, the scheduler agrees to let go of clients AF Gold during the 20 per cent of the remaining time and during 10% of clients AF Silver or Bronze.

The whole of the actions suffered by a package in a DiffServ node is carried out by a body called General Conditioner (conditioner). A conditioner of traffic can contain the following elements: Quantity Surveyor, marker (marker), Director, form and suppressor of packets. A stream is selected by the classifier. A quantity surveyor is used to measure the traffic in comparison to the profile. The measurement is performed by the quantity surveyor for a packet determined can be used to determine whether it must send the packet to a marker or a suppressor of package.

When the packet exits the conditioner, it must possess the appropriate value of the DSCP. The Quantity Surveyor obtains the temporal properties of the stream of packages selected by the classifier in function of a profile determined by a TCA (Traffic Conditioning Agreement). The Quantity Surveyor sends the information to the other organs of the conditioner, which implement specific functions adapted to the packages for that they receive appropriate treatment, that they are located in the profile or outside profile.

The markers of packets positioned the DSCP field to a particular value and add the package to the flow corresponding aggregate. The marker can be configured to mark all the packets to the correct value of the DSCP or to choose a particular DSCP for a set of predetermined PHB.

The directors of form have for objective of delaying packets from a same stream for the put in conformity with a specific profile. A director in the form usually has a memory of finite size for delaying the packets by putting them on hold. These can be destroyed if there is no space available in the memory for the put in compliance.

The suppressors destroy the packets of the same stream which do not conform to the profile of traffic. This process is sometimes called "policing" of traffic. A suppressor is sometimes implemented in the Director in the form when a packet must be dismissed, if it is impossible to install in the profile.

The processors of traffic are most often placed in the nodes of entry and exit areas DS. Since the marking of packages is carried out by the input nodes of the domain, an aggregate from another operator is assumed to be consistent with the appropriate TCA.

The PHB (Per Hop Behavior)

A PHB is a description of the way to transfer packets of the same aggregate to the interior of a DiffServ node. The way to transfer packets is a very general concept in this context. For example, if there is only one PHB in a node, the way to transfer the packets is simple and allows easily to deduct

the response time, the losses and the jitter, that only depend on the load of the node. Distinctions on the way to transfer packets can be observed when several aggregates with different PHB come into competition in a node to acquire the memory or bandwidth. The PHB must therefore define the means by which resources are allocated to the different aggregates.
The PHB the most simple is the one that guarantees a bandwidth of at least *X* % of a physical link on a time interval that is significant in relation to the life of the aggregate. This PHB is easy to measure, even in the presence of many other aggregates with PHB different. A PHB a little more complex could guarantee at least *X* % of a physical connection with a fair sharing of the Bandwidth not distributed between the aggregates across a node. As a general rule, the observation of a PHB depends on the constraints posed by the different aggregates which share the transfer nodes.
The PHB can be specified for the priority of allocation of resources (memory, bandwidth, etc.) of transfer nodes on the basis of the guarantees granted to the different packages of aggregates (delay, loss, etc.). These PHB can be used as basic bricks which allow to build groups of PHB consistent. The groups of PHB share common constraints applying to each PHB to the inside of a group. The relations between the PHB in the interior of a group can be expressed in terms of priority of packet loss in relative or absolute, in other words with thresholds deterministic or stochastic, or balanced sharing of resources.
The PHB are implemented in the nodes in the middle of managers of briefs and mechanisms of scheduling of packets. They are defined by the way in which the transfers are made and strongly depend on the policies for the allocation of resources, rather than the implementation of specific mechanisms. Many mechanisms can be used to implement a group of PHB particular. Several groups of PHB can be implemented in a node and in a field, for example. The groups of PHB must be defined in such a way that the allocation of resources between the groups can be easily inferred and that mechanisms can be shared by different groups of PHB. The definition of a group of PHB indicates the possible conflicts with the groups of PHB already defined so as to avoid inconsistent operations.
A PHB is selected in a node by the value of the DSCP scope by a packet. The PHB standards have a DSCP associated. The whole of the DSCP being greater than the space available for the DSCP recommended in DiffServ, the field DS leaves the possibility in local to match a PHB to several DSCP. All the DSCP must however have a corresponding PHB. If this is not the case, a PHB by default is assigned to a DSCP is not known.

The architecture model of DiffServ

In the Architecture model of a DiffServ network traffic entering the network is classified and possibly conditioning in the border routers in the network. Once the classification done, trafficking are aggregated by the PHB in aggregates. To the inside of the network, packets are transferred according to the value of their DSCP.
A domain DiffServ, commonly called Field DS, corresponds to a set of nodes DiffServ, or DS Nodes, contiguous which operate according to a policy of common control and a set of PHB in common. This area includes DS a border composed of DS nodes capable of performing the classification in the PHB in the field and to condition the trafficking. The border routers can use either the DSCP standards, or transform the DSCP by default in the DSCP owners. The PHB determined by the classifiers must allow to achieve the SLAS negotiated by the customers of these aggregates.
A domain DS consists of one or multiple IP networks with the same management system. The manager of the network determines the characteristics of the PHB that it may propose in function of the SLA of its clients and resources available to it. The field contains the routers border and the interior routers. The border routers are able to perform the classification of incoming streams in the network. They

must eventually have the functions of conditioning defined by an approval of conditioning of traffic, or TCA (Traffic Conditioning Agreement). The interior nodes are able to perform some basic functions for conditioning of the traffic, such as the transformation of the marking of the DSCP.
A host attached directly to a internal router must have the router functions border or be connected to a internal router with these functions. A border router has at the time the properties of a node of entry and an exit node. A node of entry is features the treatment of TCA, and an exit node from those of the conditioning of the waves in order that they conform to the normalization.
A region implementing differentiated services, or region DiffServ or DS, is an environment comprising one or several areas DS. A region DS allows the implementation of differentiated services on the whole of the region. The areas DS to the inside of a same region can achieve the immediate packaging necessary for the implementation of a differentiated service end to end in the region.
The access of a stream in a domain DS is carried out thanks to the treatment of the SLA from the outside of the network and linked the flow examined. An SLA can specify a classification or a profile of traffic and the actions to perform on the incoming traffic or not in the SLA. The TCA between areas is derived from this SLA.
The process of classification identifies the or the PHB in which traffic should be inserted. A packaging can be conducted to enable this adequacy. The packaging is done by a director in the form, a controller (policing entity) or a remarking, so that the characteristics of the flow correspond well to the definition of TCA in the domain.
The packet classifier selects the waves through a study of their characteristics. This study may focus on the analysis of the application by its port number or through a filter. Two types of classifiers are proposed in the DiffServ networks: the classifiers BC (Behavior Classify), which work only on the DSCP of the field DS, and classifiers MF (Multi-Field), that select the packets on combinations of information from the analysis of the packages or waves. The classifiers should be configured by a management system respecting the TCA. The Classifier must be able to authenticate the flow for security reasons.
The traffic profile specifies the temporal properties of the selected traffic by the classifier. The classifier determines the rules to follow if a particular package is not consistent with the profile determined by the CAW. For example, a profile based on a token-bucket implies that:

If DSCP = *x,* use the token-bucket *R, B*

This profile indicates that all packets marked with the DSCP *x* must see their performance compared to a metric. The state associated with a packet is obtained by the quantity surveyor and must match the token-bucket of rate *r* and a size of burst equal to *b.* In this example, the packets out profile are those who arrive when an insufficient number of tokens (tokens) is available. The concept "in and out" profile can be extended to more than two levels. Several levels of compliance for a given profile can therefore be defined.
The actions of packaging may be decided on for incoming and outgoing packets. Of this fact, different solutions of accounting may be implemented to determine the cost of the service. The packets in the profile may be admitted to enter in the field DS without additional conditioning or, conversely, their DSCP can be changed. This occurs when the DSCP is positioned at a value which is not the one defined by default or that the packet between In a domain DS using a PHB which is not consistent with the standard. A packet out profile can either be put on hold until the packet between in the profile, either destroyed or marked by a new DSCP. The packets out profile may possibly be aggregated to a class lower PHB and, for example, transported in a class be. A profile of traffic is a

optional component of the TCA.

Allocation of Resources

The implementation, configuration and administration of the groups of PHB accepted for the nodes of a domain DS must allow the partition of resources between the different aggregates, in agreement with the policies for the allocation of resources. The processors of traffic must check that the resources are properly allocated in function of the TCA. Although a set of services can be deployed in the absence of functions of the complex conditioning (using policies of static marking), functions as the control of the waves (policing), formatting, and the dynamic remarking allow deployment of services next of performance metrics predetermined.

An entity of control must be able to arbitrate between the decisions of allocation of resources contradictory. There is a large variety of models to perform these checks. However, the passage to the scale of the DiffServ technique requires that the control of areas does not require a microscopic management of network resources. The control passing the more easily the scale is done in open loop at the scale of the processing time of the packets.

The standardization of the PHB must specify a DSCP recommended among the values of the DSCP available. The corresponding functions must include the management of the queue, the allocation of the briefs, the destruction of the packets and the selection of the output line. Finally, a specification of the methods to resolve incompatibilities between the groups of PHB must be added to the standardization.

The specification of a group of PHB must indicate the number of individual PHB present in the group. If several PHB working in parallel in a same group, interactions and the constraints to be respected between the PHB must be clearly indicated. For example, the specification of the Group should indicate if the probability of reordering of packets in a same microflot increases when different packages of this microflot borrow different PHB to the inside of the group and therefore are marked by different DSCP.

When the operation of a group may depend on constraints between PHB, the definition of PHB must describe the behavior of the processors when these constraints are violated. In addition, if actions such as a packet loss or a remarking are required when the constraints are raped, these actions must be perfectly stipulated.

A group of PHB can be specified for a local use to the interior of a domain to allow to define the specific functions of a domain. In this case, the specifications of a PHB are useful to allow for the interoperability of these PHB to the inside of a group. However, all groups of PHB that are defined for a local use must not be taken into account in standardisation. Only groups of PHB standards must be absolutely specified to allow interoperability of transfer nodes of the OEMS.

It is possible that a package marked for a PHB to the inside of a group of PHB be noticed if it is selected to be transferred by another PHB of this group. Three reasons may require this remarking:

- The DSCP associated with a group of PHB correspond to the States of the network.
- The conditions involve a change in the level of the PHB used for a stream.
- The border between the two areas is not covered by an SLA. In this case, the DSCP to select at the crossing of the border are determined by local policies.

The specification of a PHB must clearly indicate the circumstances in which marked packets for a PHB given to the inside of a group of PHB must be directed toward another PHB in the group. If it is forbidden that PHB a package to be amended, the specification must clearly indicate the risks incurred when the PHB is amended despite everything. The risk of change in a PHB to the inside or

outside of a group of PHB strongly increases the likelihood of having to reorder a microflot. The PHB in the interior of a group can carry different semantic that it may be difficult to duplicate if the packets are noticed in another PHB.
For some groups of PHB, it may be appropriate to indicate a change of state in noting the packets to another PHB inside the group. If a group of PHB is determined to reflect the state of the network, the definition of a PHB must be able to describe the relationship between the PHB and member of the network that they reflect. In addition, if these PHB limit the possibilities of transfer that a node can achieve, these constraints may be specified in the actions that a node can take.
The process of specifying a group of PHB is by nature incremental. When a new group of PHB is proposed, the interactions with the PHB already implemented must be documented. When a new group of PHB is created, it can be totally new with respect to its subject matter, but also be an extension more or less complex of a group already defined. In both cases, the interactions to the interior of the new group must be specified as a function of the other groups of PHB. In particular, the reordering of packets of microflots must be discussed. If the concurrent operations must be carried out by PHB belonging to different groups, it must specify the interactions. If the group of PHB is an extension of a group of PHB already exists, it is necessary that the interactions are specified with care.
The conformity of a PHB with its definition must be able to be verified by various rules, such as the use of tables of compliance, the taking into account of tests or pseudo-codes, etc. In addition, the specifications of a PHB must include elements of security. In particular, groups of PHB must indicate the process to activate in response to attacks by denial of service or to attacks aimed at reducing or violate the contract of service.
A specification of PHB must finally include a section detailing the means employed for the configuration and management of PHB.

Element of standardization of PHB

As explained earlier, these are the characteristics of a PHB which must be standardized and non-specific algorithms or the mechanisms implemented to achieve it. The behavior of a node is defined by a set of parameters that allow you to determine how the control of packets must be exercised on the interfaces (number of queues, priorities attached, length of queues, service disciplines, algorithms of packet loss, weight associated with preferences, thresholds, etc.).
To illustrate the distinction between a PHB and a mechanism, indicate that a conditioner of traffic adapted to the PHB is a system of queues with algorithms type WFQ (Weighted Fair Queuing), Weighted Round Robin (WRR) or their variants, or CBQ (Class Based Queuing), separately or in combination. The PHB can be defined individually or in a group. The specification of the Group of PHB must describe the conditions according to which a packet must be noticed to select another PHB inside the group.
Each standardized PHB must have an associated DSCP, allocated among the 32 potential DSCP. This specification leaves the place to make evolve the DSCP In using all the possibilities offered by the field DS. Equipment suppliers are free to offer specific parameters for their PHB. When a standardized PHB is accepted in a node, a supplier must be able to use any algorithm that meets the definition of PHB to achieve its implementation. The possibilities offered by the resources implemented in the node and the means to configure the node determine how packets are treated to achieve the PHB.
The operators are not required to use the same mechanisms of configuration to perform the differentiated services to the inside of their network and are free to configure their way the

parameters of their transfer nodes so as to satisfy the PHB taken into account.

Conclusion

Fifteen years ago, we would have been able to count a 20 architectures for packet level, each major computer company deploying its own architecture.

Today, we can say that there are virtually no more that a single architecture of packet level in the world, that from the Internet. If it is still challenged, it is no longer by other solutions for level 3, but by the architectures of frame level, which include placing the IP packet in a frame in the Machine Terminal and to transport it to the inside of the frame all along the path. Arrival in the machine terminal end of the recipient, the frame is décapsulée to restore its IP packet. The architecture of this solution is level frame (layer 2, or connection).

The networks of the future will no doubt be the witnesses of a confrontation of the architectures IP between them since this standard is imposed. The discussions focus on the way to carry the IP packet from one end to the other end of the network…

11

The Label Switching: MPLS and GMPLS

The switching has started with the technology X.25. The reference was then in the layer packet. To provide higher flows, the reference was changed from place to be put in the frame and avoid this fact the décapsulations/necessary encapsulations to the passage by the IP level. From there was born the Frame Relay, presented in detail with X.25 in Annexs E and G. The next step was to choose a frame much smaller and better adapted to the multimedia, the frame ATM, treated equally at the Annex G, because it has virtually disappeared.
The frames associated with SPLM have first been of frames ATM, but little by little Ethernet has taken the above. Of this fact, this technology could have been classified in Ethernet networks, but all its originality comes from the signage required to put in place the paths. X.25, then the Frame Relay, and ATM finally have need of signaling protocols specific and generally complex. Conversely, MPLS has chosen the IP packet to route its signage. This native compatibility with IP has ensured its success.

MultiProtocol Label Switching (MPLS)

MPLS is a proposed standard by the IETF, the standardisation organization of the Internet, for the whole of the architectures and protocols of high level, which it does today remains that the IP protocol. The nodes of specific transfer used in MPLS are called LSR (Label Switch Router). The LSR behave as switches for the waves of user data and such as routers for signaling. To route user frames, it uses references, or *labels*. to a reference of entry is a reference of output. The succession of references defines the path followed by the set of frames containing the packages for the IP stream.
Any frame used in switching, or Label Switching, may be used in an MPLS network. The reference is placed in a specific field of the frame or in a field added to this purpose.
The LSR supersede the routers in working either in router mode, to trace the path by the signalling, either in mode switching, for all frames that follow the path. The path is determined by the IP mode and thus by a routing algorithm of the Internet.
The routing solution-switching of this technique is illustrated in Figures 11.1 and 11.2. The first assumes that the frame of the level frame is ATM and the following that the frame of the level packet is Ethernet. It should be noted that, in the case of the solution ATM, the IP packet to carry is cut in the AAL layer (ATM Adaptation Layer) in 48 bytes to be encapsulated in the frame ATM. This step does not exist in the Ethernet world, and it is one of the reasons for which we prefer very widely switch Ethernet frames that ATM.

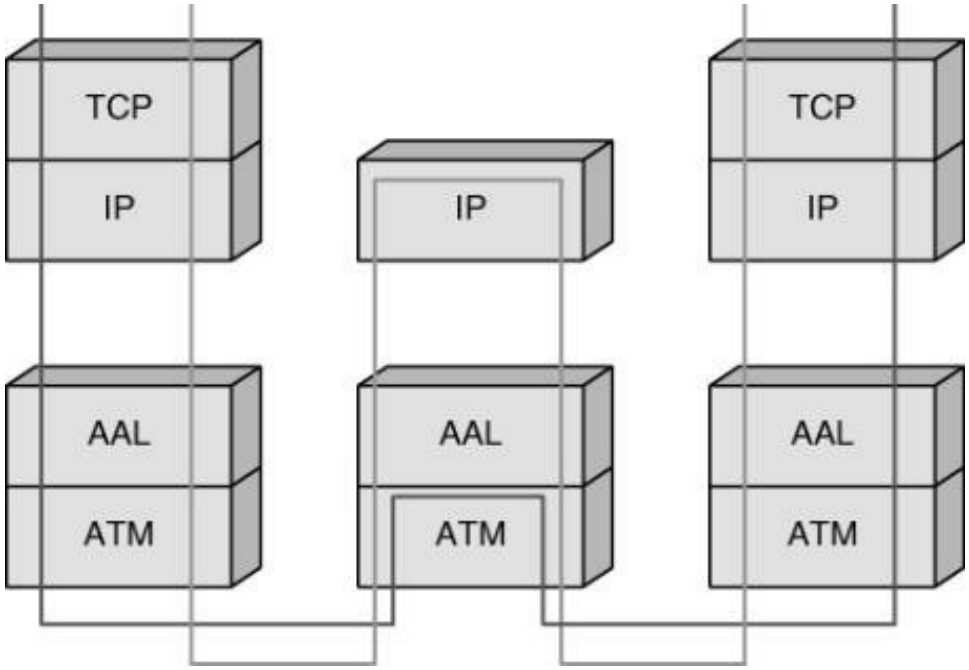


Figure 11.1
TCP/IP over ATM MPLS in

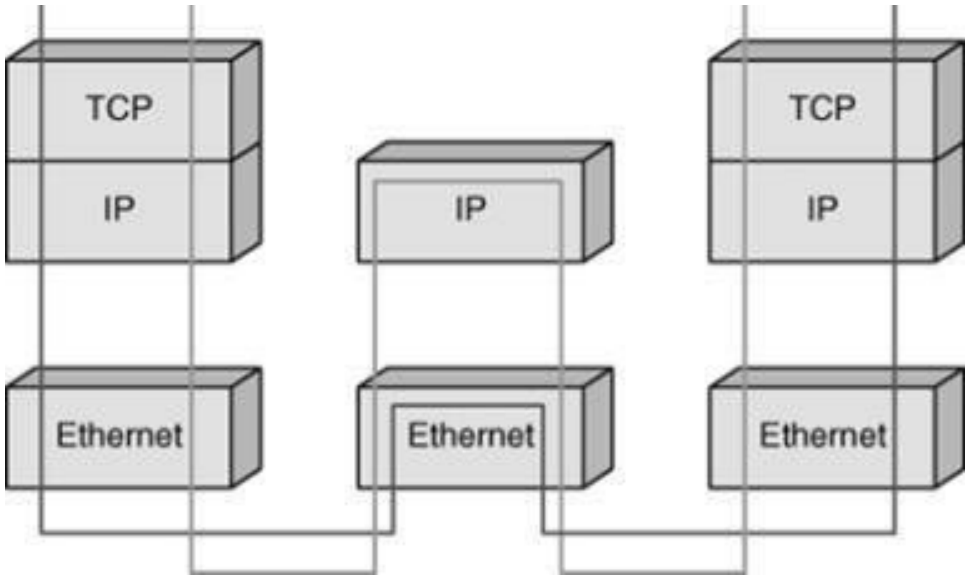


Figure 11.2
Architecture of an IP environment on Ethernet

This architecture is dictated by the Ethernet environment, which shines by its simplicity of implementation. It has the advantage of relying on the existing, the couplers and the various Ethernet networks, that many companies have put in place to create their local networks.

Since the data produced in the format IP, IPX or another are placed in Ethernet frames in order to be transported in the local environment, it is tempting to directly switch the Ethernet frames from one local network to another. As all of the networks in the Ethernet environment are compatible and speak the same language, the machines emitting Ethernet frames can interconnect easily. We can thus realize the networks extremely complex with shared segments on the local parties, the switched links on the long distances or between the Ethernet switches and crossings by routers when an escalation up at the IP level is required.

Characteristics

MPLS is the logical culmination of all the proposals that have been made in the years 1990. The idea of the IETF has been to gather the proposals in a common standard for the transport of IP packets on sub-networks working in switched mode. The nodes are routers-switches, or LSR (Label Switch

Router), able to reassemble either at the IP level to perform a routing, either at the level frame to perform a switching.
The most important characteristics of the standard SPLM are the following:

- Specification of the mechanisms for transport of the waves of IP packets with various granularités flows between two points, two machines or two applications. The Granularity refers to the size of the flow, which can integrate more or less WAVES user.
- Independence of the frame level and the level package, although only the transport of IP packets is actually taken into account.
- Relation of the IP address of the recipient with a reference to entry in the network.
- Recognition by edge routers of routing protocols of OSPF type and signage As RSVP.
- Use of different types of frames.

A few additional properties deserve to be highlighted:

- Opening of the path based on the topology, although other possibilities are also defined in the standard.
- Assignment of references made by the downstream, that is to say, at the request of a node which emits a message in the direction of the transmitter.
- Variable granularity of references.
- Stock of references managed according to the method "last come first served".
- Ability to prioritize the applications.
- Use of a timer TTL.
- Encapsulation of a reference in the frame including a TTL and a high quality of service.

A benefit provided by the MPLS protocol is the possibility, shown in Figure 11.3, to carry IP packets on several types of switched networks. It is thus possible to move from an ATM network to an Ethernet network. In other words, it can be any type of frame, from the moment where a reference may be included. However it is possible to add a reference when the frame does not provide for this (see below). The advantage of this solution is the simple migration of old technologies toward new, as ATM to Ethernet, simply enough.

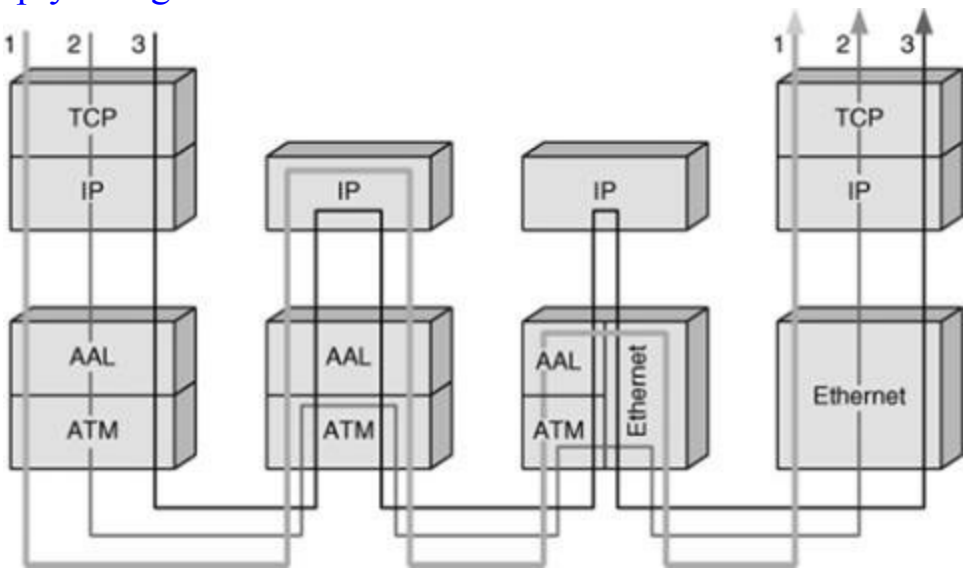


Figure 11.3
MPLS network with sub-separate networks

Operation

The transmission of data is carried out on paths, appointed LSP (Label switched path). An LSP is a result of references from the source and ranging up to the destination. The LSPS are established prior to the transmission of data (control-driven) or to the detection of a stream which wishes to traverse the network (data-driven).

The references included in the frames are distributed using a signaling protocol. The most important of these protocols is LDP (Label Distribution Protocol), but it also uses RSVP (Resource ReSerVation Protocol), possibly associated with a routing protocol, such as Border Gateway Protocol (BGP), or OSPF. The FRAMES routing IP packets carry the references of node in the node.

LSR and 1

The nodes that participate in SPLM are classified into 1 (label edge router) and LSR. A LSR is a router in the heart of the network that participates in the establishment of the path by which the frames are forwarded. A 1 is a node for access to the network MPLS. A 1 may have multiple ports allowing access to several separate networks, each may have its own technique of switching. The 1 play an important role in the establishment of references.

LSR (Label Switch Router)

A piece of equipment that performs a switching to a reference is called a LSR. The switching tables (LSFT Label Switching forwarding table) consist of a set of references of entry which correspond the output ports. to a reference of entry can match several output ports to take account of multipoint addresses.

The switching tables can be more complex. To a reference of entry can match the output port of the node in a first sub-entry, but also, in a second sub-entry, a second output port to corresponding to the output queue of the next node that will be crossed, and so on. So, to a reference can match a set of output ports that will be borrowed when routing the packet.

The switching tables can be specific to each port of entry to a LSR and consolidate additional information, such as a quality of service or a specific request for resources.

FEC (Forwarding equivalence class)

In MPLS, the routing is carried out by the intermediary of the equivalence classes, called FEC. A class represents a stream or a set of waves having the same properties, including the same prefix in the IP address. All frames of a CEE are treated in the same way in the nodes of the network MPLS. The frames are introduced in a CEE at the entry node and can no longer be distinguished within the class of the other streams.

A FEC can be built in different ways. It may have a destination address well determined, a same address prefix, a same class of service, etc. Each LSR has a switching table which indicates the references associated with the FEC. All frames of a same CEE are transmitted on the same output interface. This switching table is called lib (Label Information Base).

The references used by the FEC can be grouped in two ways:

- By platform: The values of references are unique on the whole of the LSR of a domain, and references are distributed on a common set managed by a particular node.
- By interface: The references are managed by interface, and a same value of Reference can be found on two different interfaces.

MPLS and references

A reference in the entry allows therefore to determine the CEE by which conveys the flow. This solution is similar to the concept of virtual duct in the world ATM, where virtual circuits are multiplexed. Here we have a multiplexing of all virtual circuits to the inside of a FEC, so that in this duct, we can no longer distinguish the virtual circuits.

The LSR examines the reference and sends the frame in the direction indicated. One sees well as well the crucial role played by the LER, which assign the streams of packets of references that allow you

to switch the frames on the good virtual circuit. The reference has no meaning that locally, since there is a modification of its value on the following connection.
Once the packet is classified in a FEC, a reference is assigned to the frame that will carry. This reference determines the point of exit by the chaining of references. In the case of conventional frames, as LAP-F of Frame Relay or ATM, the reference is positioned in the DLCI (Data Link Connection Identifier) or in the VPI/VCI.
The signage required to remove the value of references along the path is determined for a FEC can be managed either to each flow (data-driven), either by an environment of independent control of the Waves user. This last solution is preferable in the case of large networks of the fact of its capacity for transition to the scale.
The references can be distributed for:

- A unicast routing to a particular destination;
- A traffic management, or TE (traffic engineering);
- A multicast ;
- A virtual private network;
- A quality of service.

The format of the reference MPLS is shown in Figure 11.4. The reference is encapsulated in the header of the frame level of the standardized field to carry the reference or just between the header of level frame and the header of packet level.

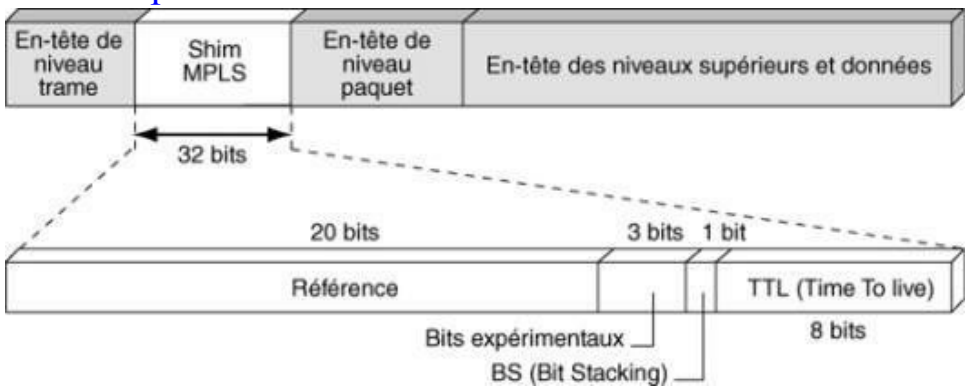


Figure 11.4

Generic format of a reference in MPLS

Figures 11.5 and 11.6 illustrate the implementation of the reference in the case of ATM and the figure 11.7 deals with the case where the frame is not designed at the outset for a Label Switching, as the PPP frame.

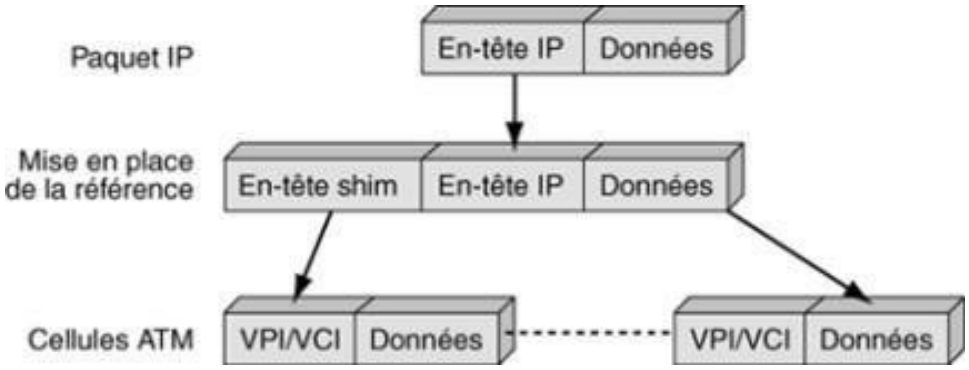


Figure 11.5

Establishment of references in the ATM

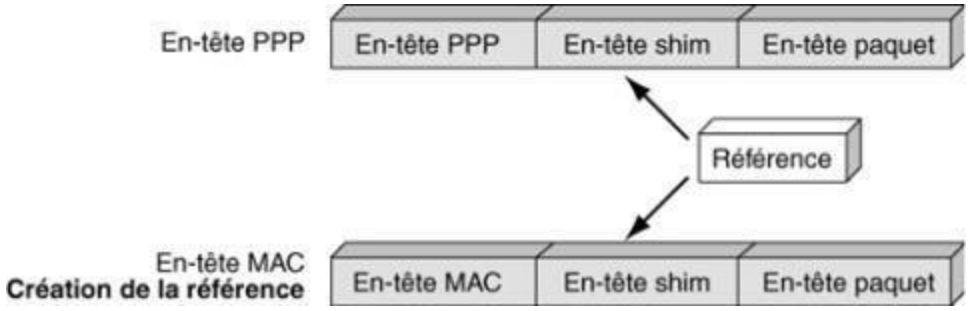


Figure 11.6

Establishment of references in the PPP frame

Distribution of References

MPLS normalizes several methods to achieve the distribution of references. The distribution indicates that each node has its own references and that he must put in correspondence with the references of its neighbors.

The methods of distribution of references are the following:

- Topology-based, or based on the topology, which uses the messages intended for the management of the routing, such as OSPF and BGP.
- Request-based, or based on the stream, which uses a query of the request to open a path to a stream IP. It is the case of RSVP.
- Traffic-based, or based on the traffic: to the receipt of a packet, a reference is assigned to the frame that the transports.

Methods based on the topology and on the flow correspond to a control (control-based), while that based on the traffic corresponds to the data.

The routing protocols, including Interior Gateway Protocol (IGP), have been improved to carry an additional reference. The OSPF protocol has been supplemented by the taking into account of the flows on the connections: OSPF-TE (Traffic Engineering). Similarly, the RSVP protocol includes a version associated with SPLM which allows him to carry a reference. The version of the most successful is RSVP-TE (Traffic Engineering), which allows the opening of roads, taking into account the resources of the network.

The IETF has also standardized a new signalling protocol, LDP (Label Distribution Protocol), to manage the distribution of references. Extensions of this Protocol, as CR-LDP (constraint-based routing-LDP), allow you to choose the routes followed by the clients of the FEC with a quality of service predefined.

The main signaling protocols are the following:

- LDP, which is the matching of the unicast IP addresses and references.
- RSVP-TE and CR-LDP, which open up roads with a high quality of service.
- Protocol Independent Multicast (PIM), which is the matching of the multicast IP addresses and associated references.
- BGP, which is used to determine of the references in the framework of virtual private networks.

LSP (Label switched path)

A domain MPLS is determined by a set of nodes MPLS on which are determined of the FEC. The LSP paths are determined by the references positioned by the signage. The LSPS are determined on a domain before the arrival of the data in the classical case. Two options are used to this end:

- The routing hop-by-hop (hop-by-hop). In this case, the LSR select the next hops independently of each other. The LSR uses for this a routing protocol such as

OSPF.

- The explicit routing, identical to the routing by the source. The 1 of entry of the domain MPLS specifies the list of nodes by which the signage has been routed, the choice of this road that may have been forced by requests for Quality of service.

The path followed by the frames in a direction of the communication may be different in the other direction.

Aggregation of the waves

The waves from different interfaces can be gathered and switched on a same reference if they go to the same direction of output. This corresponds to an aggregation of waves. This technique is already exploited on the ATM networks, in which a conduit may aggregate multiple waves from different input nodes to a common point, where streams are disaggregated.

The aggregation of the waves has for objective to avoid the explosion of the number of references to use or, what is equivalent, to prevent the switching tables to become too important.

Signage

As explained previously, several mechanisms of distribution of references, called signaling, can be implemented in the nodes of an MPLS network, including the following:

- Request for reference: a LSR emits a request for a reference to its neighbors to downstream (downstream), that it can link to the value of a CEE. This mechanism can be used in node in node up to the exit node of the network MPLS.
- Correspondence of reference: in response to a request for a reference to a node upstream, a LSR sends a reference from a mechanism of correspondence known already in place to go until the exit node.

The figure 11.7 gives an illustration of these two mechanisms.

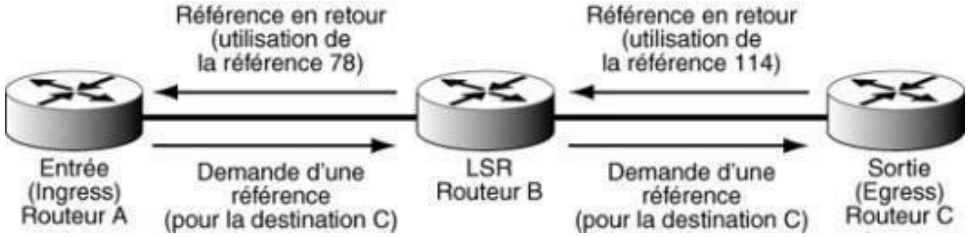


Figure 11.7

Signaling mechanisms of MPLS

LDP (Label Distribution Protocol)

LDP is the protocol for the distribution of references which tends to become the most widely used standard in MPLS. This Protocol takes into account the unicast and multicast. The routing is explicit and is managed by the nodes in the output. The exchanges are under the protocol TCP to ensure an acceptable quality.

Two classes of messages are accepted, that of adjacent messages and the messages indicating the references. The first allows you to query the nodes that can be achieved directly from the node origin. The second class of messages shall transmit the values of the reference when there is agreement between the adjacent nodes. These messages are encoded in the classic form, which allows you to describe an object: it indicates in a first field the type of object, in a second The total length of the message describing the object and in a third the value of the object. This encoding is called Type Length Values (TLVS).

The routing is performed, as shown previously, by classes of equivalence, or FEC. A class represents a destination or a set of destinations with the same prefix in the IP address. Of this fact, a packet that

has a given destination belongs to a class and follows a common road with the other packets in this class. This defines a tree, whose root is the recipient and whose leaves are the transmitters. The packets that do not have more than to follow the shaft up to the root, the Waves overlaying little by little in going to the root. This solution allows you to do not use too much of different references.
The granularity of references, i.e. the size of the waves that use the same reference, is the result of the size of the equivalence classes: if there is little, the waves are important, and the granularity is strong; if there has been a lot, the waves are low, and the granularity is fine. For example, a destination may correspond to an important network, in which all the addresses have a common prefix. The destination can also correspond to a particular application on a given machine, which gives a strong granularity. This latter case is shown in Figure 11.8, in which the receiver is the machine 1 and the FEC is determined by the shaft whose leaves are the terminal machines 1, 2 and 3. The class of equivalence, descending the shaft from 1, starts by the references 28 and then 47 and continues by the branches 77 and then 13 and then 36. From 2, references 53 and then 156 are used to go to the root. From 3, this are the references 134 and 197 which are used. All the above references belong to the same class of equivalence.

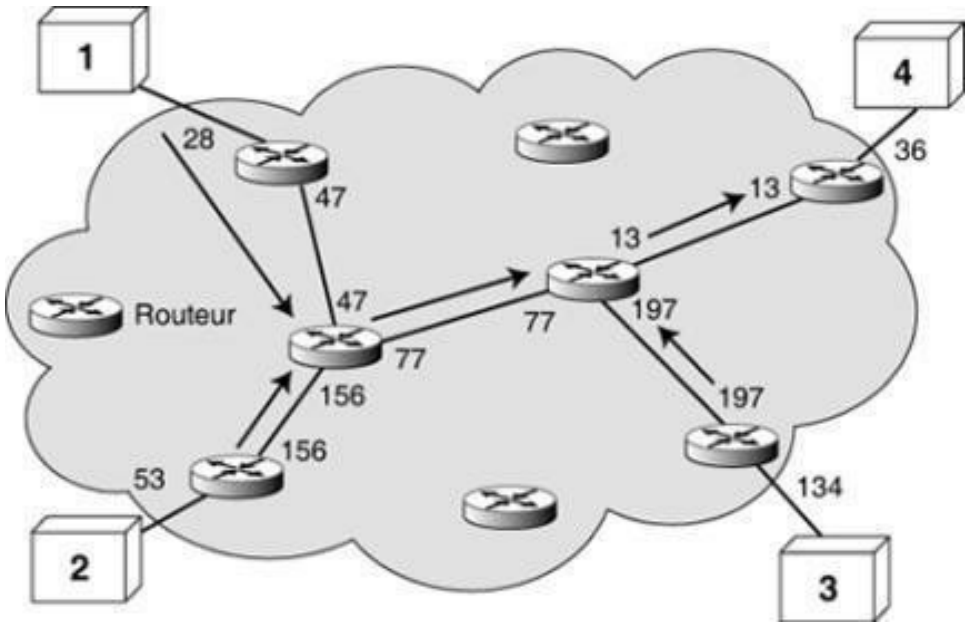


Figure 11.8
The equivalence classes (FEC) in an MPLS network

In this example, the terminals 1, 2 and 3 wish to issue a flow of IP packets to the terminal station 4. For this Station 1 emits its frames (encapsulating IP packets) with the reference 28, which is switched to the reference 47 and then switched to the references 77 and then 13 and then 36. The flow from the station 2 is switched from 53 in 156 and then 77, 13 and 36. Finally, the third flow, starting from the station 3, is switched from the values 134 and then 197, 13 and 36. It can be seen that the aggregation is performed on the first two waves with the only value 77 and that the three streams are aggregated on the values 13 and 36. The station 4 could be replaced by a sub-network, which would certainly have allowed to aggregate much more of flows and have a granularity least fine.
A problem posed by the routing tables involving the FEC is that of potential loops, i.e. a possible return to a station that has already seen the frame. If the routing uses a protocol such as OSPF, it prevents loops using an information message.
The LPD protocol includes the following messages:

- Message of discovery (discovery message), which announcement and maintains the presence of a LSR in the network.
- Message of session (Session Message), which establishes, maintains and

terminates sessions between the ports LDP.

- Warning Message (advertisement message), which creates, maintains and destroyed the correspondence between the references and the FEC.
- Notification message (notification message), which provides information of error or problem.

The switching tables can be built and controlled in various ways. The routing protocols of the Internet, such as OSPF, BGP, PIM, etc., are generally used to this effect. It must be their add procedures to match the references and the classes of equivalence FEC.

As explained earlier, the distribution of references is performed by the downstream, in dating back toward the station of emission. In reality, it is indicated in the standard MPLS that the distribution of references can be performed by the downstream (downstream) or by the upstream (upstream). In the first case, the recipient indicates to the Nodes upstream of the value of the reference to put in the switching table. In the second case, the packet arrives with a reference, and the node updates its switching table.

In the upstream distribution (upstream), a downstream node sends the value of the reference that it wishes to receive to switch a packet on a CEE. This are the nodes located the further downstream that trigger the process and indicate the recipients and their granularity. The changes are carried out during the reception of a frame or by the intermediary of information of supervision.

The distribution of identifiers may be carried out by the intermediary of the Protocols RSVP-TE or PIM.

The batteries of References

The mechanism of batteries of references to MPLS allows an LSP to transit by nodes non-MPLS or by areas hierarchical. For this, the area bearing the reference can store not more a value, but a stack of values, i.e. a stack of references. Depending on the level of the hierarchy of references there uses the reference of the corresponding hierarchy in the stack.

The batteries of references can achieve the tunnels, in which are grouped the references of a same level of the hierarchy. At the exit of the tunnel, it returns to the hierarchy just below, as shown in Figure 11.9. On this figure, the flow from the station 1 is switched on the values 28 and then 53. The node has, a stack of references is created with the addition of the reference 156, which is used in the next node to switch on the values 77 and then 197. The Node B allows the output of the tunnel using the reference again 53 After popping the references. We see that between the node A and Node B a tunnel is constituted, which, to a reference of entry 53, match a output reference 13.

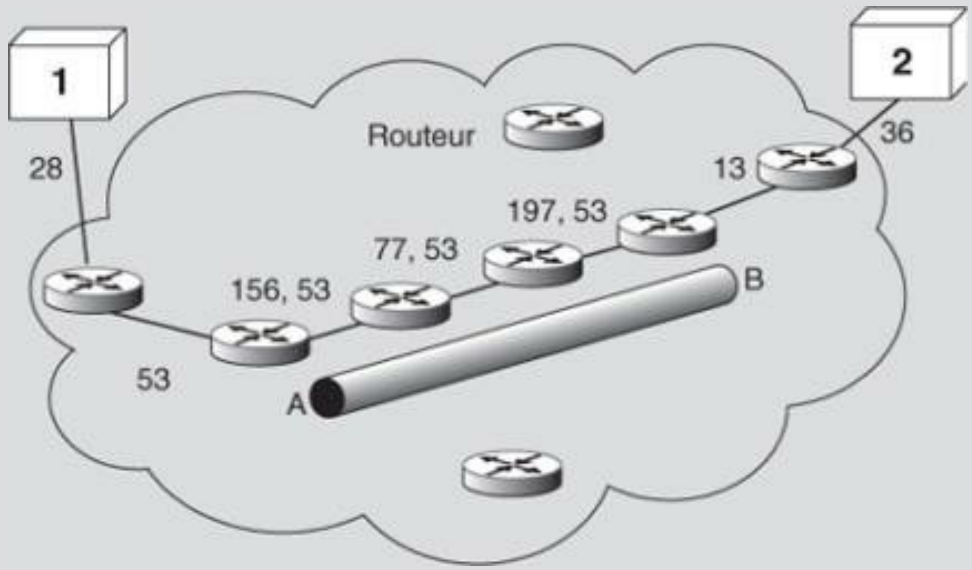


Figure 11.9

MPLS tunnel carried out thanks to a stack of References

The traffic engineering

It is difficult to achieve a traffic engineering in Internet of the fact that the BGP protocol uses only the

information about the topology of the network. The IETF has introduced in the architecture MPLS routing at basis of constraint and a routing protocol internal to state of extended links in order to achieve a traffic engineering effective.
Each frame encapsulating an IP packet that between in the MPLS network sees add by the LSR of entry, or Ingress LSR, a reference to the level of the header to route the frame in the network. The paths are previously opened by a Protocol resource reservation, RSVP or LDP. At the exit of the network, the reference is added to the header of the frame is deleted by the LSR output, or Egress LSR.
Attributes allowing control of the resources allocated to these paths are associated with the LSP, which is the road built between the LSR of entry and the LSR output. These attributes are summarized in table 11.1. They essentially concern the necessary bandwidth to the path, its level of priority, its dynamic aspect, through the protocol used for its openness and its flexibility in the case of failure.

Attribute	Description
Bandwidth	Minimum requirements of bandwidth to book on the path of the LSP
Path attribute	Indicates if the path of the LSP must be specified manually or dynamically by the algorithm CBR (Constraint-Based routing).
Priority of preemption	The LSP with the highest priority is allocated a resource requested by several LSP.
Priority of preemption	Indicates if a resource of a LSP can be withdrawn to be assigned to another LSP more priority.
Affinity or color	Expresses the administrative specification.
Adaptability	Indicates whether the path of an LSP must be modified to have an optimal path.
Flexibility	Indicates if the LSP must be rerouted in the event of a fault on the path of the LSP.

Table 11.1 • Attributes of LSP paths in an MPLS network

The algorithm CR (constraint-based routing)

The CR algorithm is applied at the time of the opening of the path or of its Reopening If the path is dynamic.
In addition to the constraints of topology used by routing algorithms classics, the CR algorithm calculates routes based bandwidth constraints or administrative. The paths calculated by the CR protocol are not necessarily the most short. In effect, the shortest path may not meet the bandwidth capacity requested by the LSP. The LSP can therefore borrow another path, slower, but with the bandwidth capacity requested. So, the traffic is distributed more evenly on the network.
The algorithm CR can be carried out in real time or not. In the first case, the number of LSPS to cross is calculated to any moments by routers on the basis of local information. In the second case, a server loads, from information collected on the entire network, to calculate the paths periodically and automatically reconfigure the routers with the New Paths calculated.
The routing protocol is necessary for the transport of the routing information. In the case of the algorithm CR, the routing protocol must carry, in addition to the information in the topology, of constraints such as bandwidth needs. The spread of these information was done more frequently than in the case of a PGI standard, since there is more of the factors likely to change. Not to overload the network, it must however ensure that the frequency of the spread of information is not too important. A compromise must be found between the need to update the information and to avoid the excessive

propagations.
The design of a system MPLS for the engineering of trafficking requires to browse through the following steps:
1.Definition of the geographical extent of the system MPLS. Depends on the administrative policy and of the architecture of the network.
2.Definition Of routers Members of the MPLS system. It is to define the LSR of entry, transit and exit the system MPLS. For various reasons, the latter does not necessarily contain all of the routers in the network, including if a router is not powerful enough or if it is not secure.
3.Definition of the system hierarchy MPLS. Two cases are possible: Connect all the LSR of MPLS system and create a single hierarchy level forming a large MPLS system or divide the network into several levels of hierarchy. In this last case, the LSR of first and second level of the hierarchy, which form the heart of the network, are highly meshed.
4.Definition Of bandwidth needs of LSP. The bandwidth requirements can be defined by the matrix of end-to-end traffic, which is not always available, or by a statistical calculation based on the exploitation of the LSP and the regular update of this information by constantly watching their traffic.
5.Definition of paths to the LSP. The paths are usually calculated dynamically by a CR real time. When it proves difficult to perform this calculation in real time, we can use an algorithm CR Non-real time.
6.definition of priorities of LSP. We can assign the highest priority to the LSP before drain a large traffic. This allows you to borrow the shortest paths and to avoid overloading a large number of links in the network, while offering a routing stability and a better use of resources.
7.Definition of the number of parallel paths between the two ends of any kind. You can configure multiple paths in parallel with highways physically different. This ensures a distribution of the traffic load more uniform. The underlying idea is to define the LSP small in size with a view to a better flexibility in routing. This flexibility is the first motivation of LSP parallel.
8.Definition of the affinity of the LSP and links. Colors can be attributed to the PHLS and links in function of administrative constraints. These colors are used to determine the paths to choose for the LSP.
9.Definition of Attributes of adaptation and flexibility. According to the evolution of the behavior of the network, it is possible to find the optimal paths for the LSP already calculated. The network administrator can accept or refuse a new optimization of the LSP. It must not be that the latter is too frequent, because it could introduce a instability of the routing. It must also provide mechanisms for the rerouting of the LSP in case of failure of a LSR.
The operation of an MPLS network follows the steps listed below:
1.Data collection for statistics using the LSP at startup of the system. The objective of this step is to calculate the rate of traffic between each pair of routers. The methods existing statistics allow to calculate the rate of traffic to the entry and exit of an interface, but not the one going to a particular destination. The construction of the matrix of end-to-end traffic is carried out by estimation, which makes the traffic engineering difficult and little efficient. The use of the LSP to the start of a system MPLS precisely gives the rate of traffic between two ends any function of destinations.
2.Operation of the LSP with bandwidth constraints defined in the previous step. The Step 1 above having allowed to know the bandwidth needs of each LSP, this information is used by the algorithm CR to recalculate the LSP with their real need for bandwidth.
3.periodic updating of bandwidths of the LSP. A periodic update of bandwidths of the LSP is

necessary to ensure the evolution and adaptation of the network to the change of the traffic in the network.

4.execution of the algorithm CR in real time. For an efficient use of links, the algorithm CR must be run on a dedicated server. Calculated on a server with all of the topology information and attributes of all the LSP, this algorithm may allow to reach the real time. The algorithm offers LSP with better performance compared to those of the LSP already open. The algorithm CR must be able to run in real time to take account of a failure of an LSP. The algorithm can then quickly determine a new LSP capable of flow traffic in the hold.

The quality of service

MPLS allows therefore to make engineering and perform calculations to determine the resources to assign to a path when the system is relatively static. If the system is dynamic, the paths must open and close to meet the constraints which are expressed on shorter time periods. The basic idea is to open the paths through an algorithm taking into account the resources. The proposal CR-LDP having been partially abandoned, another algorithm, RSVP-TE, has taken a place of choice among the OEMS.

In Cr-LDP, the two ports that must communicate to exchange their set of references to establish the connection. In RSVP-TE, there is no negotiation of references. It is the management plan which takes charge of this negotiation. For very large networks, the establishment of the path with LDP may require considerable resources, which explains its failure for the moment.

CR-LDP can specify the route from the source by a field of type TLV and RSVP-TE through the subject "explicit road". The two protocols send a response to the node of entry to indicate the success or failure of the opening of the path.

Tables 11.2 and 11.3 respectively summarize the similarities and differences between the two techniques.

Characteristic	CR-LDP	RSVP-TE	Comment
Initialization of the opening	Message LABEL_REQUEST	RSVP-TE Message path containing the object label_REQUEST	
Opening	DIFF-SERV_psc TLV	The DIFFSERV object_psc	Both contain the information corresponding to the DiffServ Code Point (DSCP) included in the request message of openness.
Accepts the LSP point-to-multipoint	Non	Non	Waiting for an RFC
Possibility of a routing by the source	Transported PA the list TLV of explicit_ROUTE	Transported by the object explicit_ROUTE	Specifies the path to follow.

Table 11.2 • similarities between RSVP-TE and CR-LDP

Characteristic	CR-LDP	RSVP-TE	Comment
Stage of Development	The youngest but not used today	The oldest, with additions to take account of the various networks available in MPLS	Some objects of RSVP have been modified to be used in MPLS.
Signage	UDP for the discovery and TCP for the session	IP packets or encapsulation in UDP for the exchange of messages	

			No detection of deterministic failure with RSVP-TE. A problem on TCP can have a catastrophic impact on the paths in CR-LDP.
Status of the connection	Hard State	Soft State	The soft State is generally not the scale. RSVP supports the aggregation of messages of refresh.
Reliability	Defined to support most of the techniques frame, such as ATM, Frame Relay or Ethernet.	Tunneling through the ATM network which must be configured manually.	

Table 11.3 • Differences between RSVP-TE and CR-LDP

MPLS-TP

The MPLS protocol is today unanimously chosen for the heart of the networks of operators. However, MPLS is complex, and the compatibility between OEMS is in general not ensured. The lack of a system of management is also glaring. In addition, the IETF wishes to extend to the periphery and propose to the operators a solution more simple to configure and less expensive. To get there, the Working Group MPLS-TP (Transport Profile) has been initiated and becomes available in 2011.

The difference between the basic standard MPLS and the new generation MPLS-TP is shown in Figure 11.10. MPLS-TP is a sub-set of MPLS by removing options, in particular on the signage. In contrast, MPLS-TP contains a new environment of management says OAM (Operation and Maintenance) and a protection of the transportation part on the physical media.

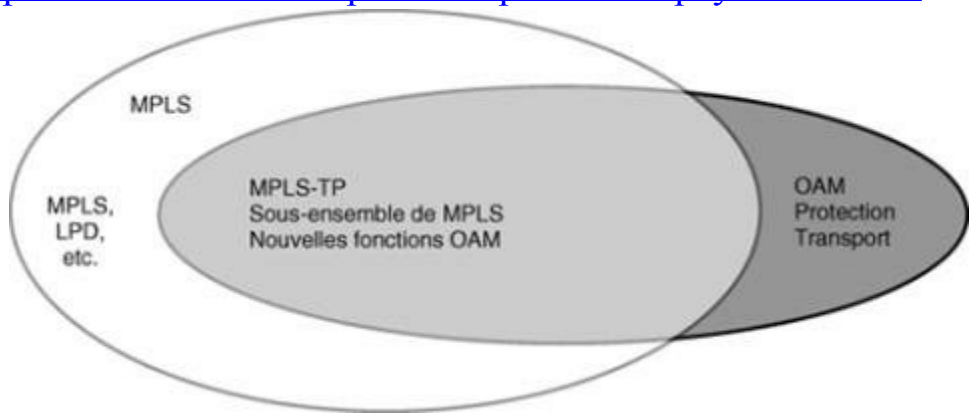


Figure 11.10
The differences between MPLS and MPLS-TP

The features of MPLS-TP may be summarized in the following manner:

- MPLS is resumed for the part switching with a simplification at the level of the implementation of the LSP.
- An architecture of pseudowire (emulation of a Level 2 in mode connection, i.e. emulation of a cable) said PWE3 (pseudowire emulation edge-to-edge).
- Use of the Control Plan, static or dynamic, from GMPLS-which is described later in this chapter.
- The features of the management from an environment OAM.
- Procedures to increase the reliability and availability to obtain performance of the same order as those of SONET.
- Use of a set of management functions from a protocol said generic associated channel (G-ACh).
- Introduction of the Multipoint.

A very important point concerns standardization. At the IETF is added the ITU-T in order to reach a standard bringing together both the telecommunications and informatics. This convergence is shown in Figure 11.11. The high part of the architecture is retained, but in a simplified way. The lower part, concerning the transport of binary elements, is integrated to MPLS-TP by the solution OTN pushed by the ITU-T. Finally, the synchronization that are found in SONET as well as the reliability and availability are occasions in MPLS-TP.

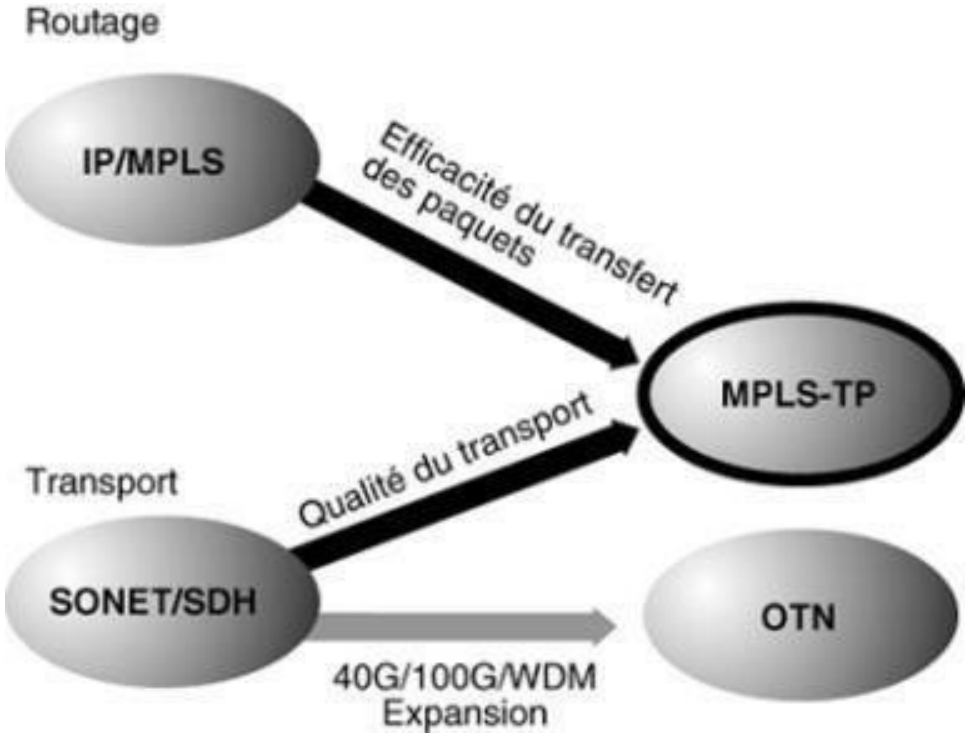


Figure 11.11

The convergence between the IETF and ITU-T

Figure 11.12 details more this convergence. The standardization of base has been carried out by the IETF and then resumption by the ITU-T under the name T-MPLS in order to add the transport function of binary elements. This intermediate standard has been resumed by the IETF to reinforce in the MPLS version-TP in parallel to the work of improvement of T-MPLS by the ITU-T.

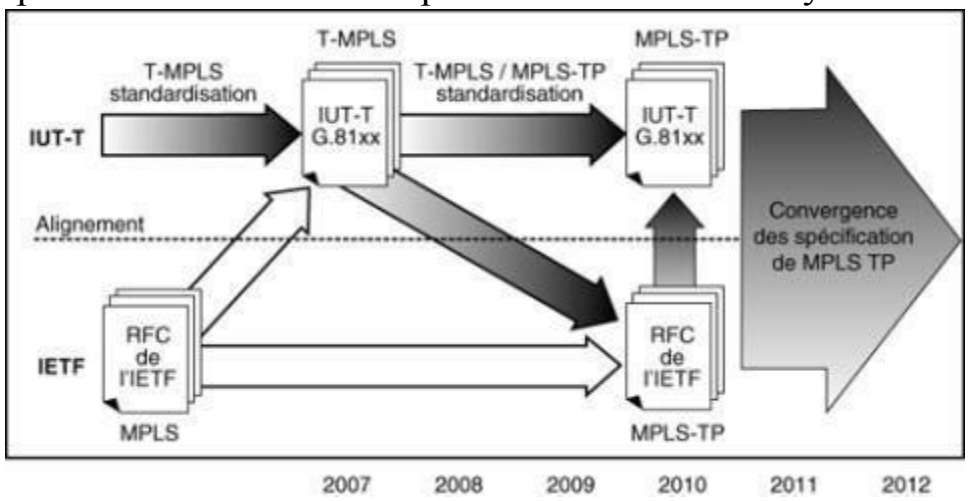


Figure 11.12

The process of normalization of MPLS-TP

MPLS-TP is the large standard that all operators were waiting. It applies as well to the inside of the network to the periphery, for example to connect networks of user access or access networks of antennas 3G and 4G.

GMPLS (Generalized MPLS)

As its name indicates, GMPLS is a generalization of the Protocol MPLS. This generalization is fairly simple to explain, since everything that can play the role of a reference - Number of a wave length, number of a slot, etc. - can enter in GMPLS. The structure of GMPLS is however more complex than it seems and a global management is necessary to arrive at well check this environment. GMPLS also brings a supervision plan which becomes a standard even for the simplified version of MPLS: MPLS-TP.

The extensions to MPLS

At the level frame (layer 2, or connection), MPLS works only on frame structures of Level 2: This is what is called the L2S (Level 2 switching). The extensions allow however to introduce references on other media, such as the number of a slice of time in a temporal sharing or a number of wavelength on an optical fiber.

The main expansion possibilities to MPLS are the following:

- PSC (Packet Switching capable), for packets able to receive a reference. One could imagine a IPv6 packet with the flow-label as reference, but this solution is not acceptable in the State, because a package cannot be transmitted directly on a physical media: it must encapsulate it in a frame. It is generally the PPP frame which serves as a carrier.
- The2SC (Level 2 Switching capable), which corresponds to the Label Switching used in the standard MPLS.
- TDMC (Time Division Multiplexing capable), which introduces the reference as a slot in a TDM. All the techniques which include a structure in the form of a frame with slots on the inside are part of this class. In particular, all of the techniques with hertzian temporal division integrate in GMPLS.
- LSC (Lambda Switching capable), which takes the number of the wavelength at the interior of an optical fiber as a reference for switching. This technique has been the first extension to MPLS under the name of MPλS.
- FSC (fiber switching capable), which takes the number of an optical fiber among a beam of optical fibers as a reference for switching. In a wiring harness, the fibers are numbered 1 to *n,n* corresponding to the number of optical fibers.

Table 11.4 summarizes the transferring techniques offered by a GMPLS network.

Field of transfer	Type of traffic	Type of transfer	Example of station	Parts List
Frame	ATM, Ethernet	Use of References	ATM switch or Ethernet	The2SC (Layer 2 Switching capable)
Packet	IP	Routing	IP routing	PSC (Packet Switching capable)
Time	TDM/SONET	Slot of time is repeating by cycle	Brewer and switch	TDMC (Time Division Multiplexing capable)
Wavelength	Transparent	Lambda	DWDM	LSC (Lambda Switching capable)
Physical space	Transparent	Optical Fiber	OXC (Optical Cross Connect)	FSC (fiber switching capable)

Table 11.4 • techniques of transfer of GMPLS-

Other extensions are imaginable, such as the association of a code in a communication, either in a CDMA or in a transmission of any kind. By these extensions, it is possible to match in input and in

output of references that do not come from the same technology. On the other hand, the different solutions are not necessarily of flows identical. For example, if one chooses as reference a slice with a number well determined a temporal multiplex terrestrial, who risk to give the better a few megabits per second, it is difficult to match an output wavelength of an optical fiber which may have a capacity of 10 Gbit/s. A Prioritization of media is therefore necessary.

Hierarchy of media

The Figure 11.13 shows a possible hierarchy between the media that can be used in GMPLS. In this figure, a stream of IP packets gives birth to a PSC, itself embedded in a the2SC FEC type, i.e. bringing together several IP streams with a common property, such as a same LSR output.

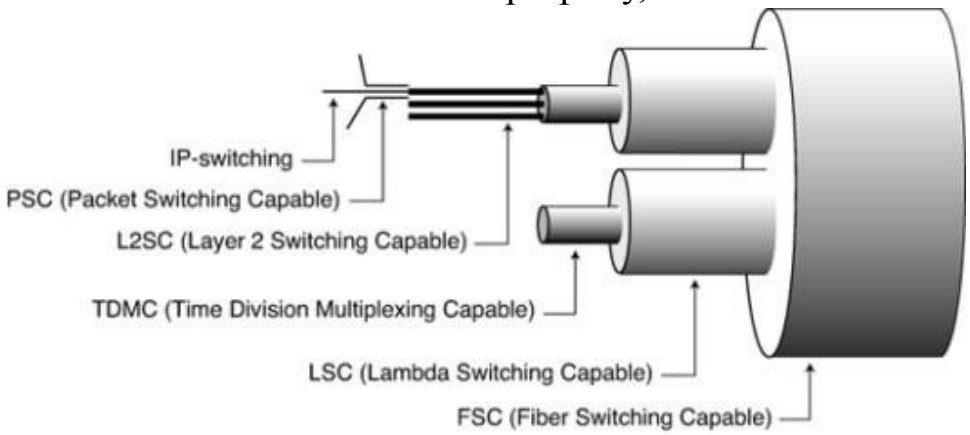


Figure 11.13

Hierarchy of techniques of transfer in GMPLS

The Waves level the2CS themselves can be encapsulated in a slot of a technique of type SONET/SDH. Continuing in the hierarchy, the TDMC streams in turn may be multiplexed in a same wavelength, that is to say in a LSC. By continuing the hierarchy to arrive at the highest level, the wavelengths may themselves be integrated in a particular fiber of a beam of optical fiber.

Overlay Network

Another important characteristic of the MPLS networks and GMPLS-is to work in overlay network, that is to say in a hierarchy of networks, as shown in Figure 11.14, where three levels are represented.

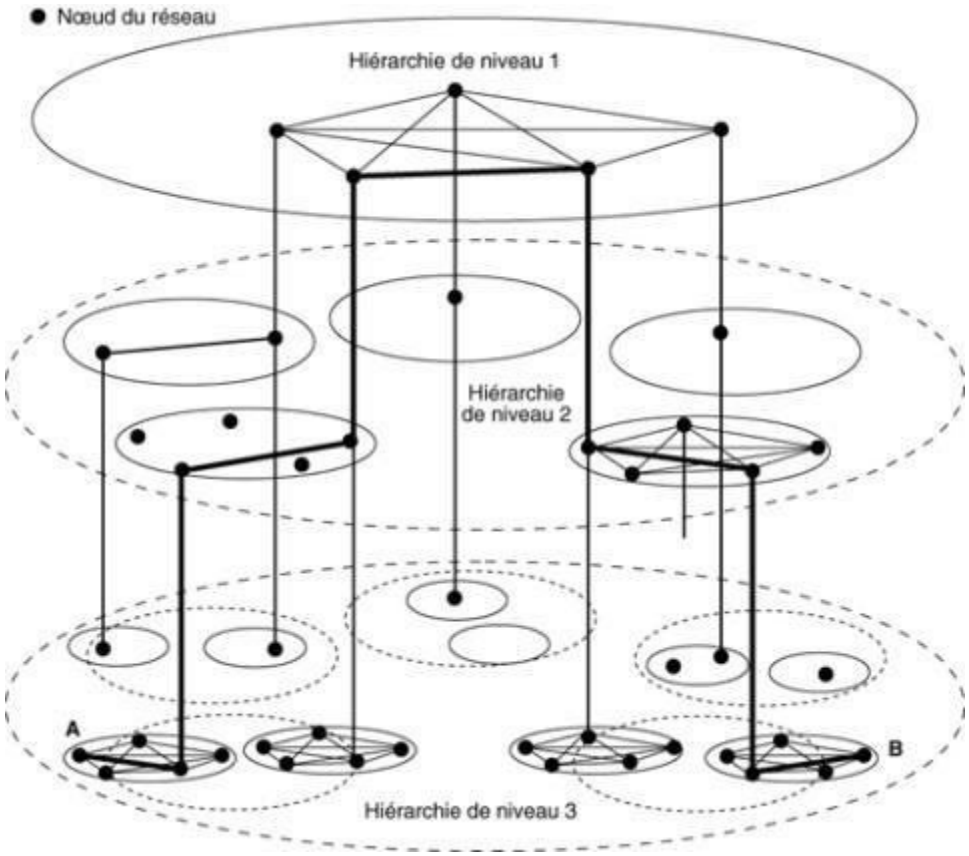


Figure 11.14

Hierarchy of network to three levels

If it is assumed, for simplicity, that the global network includes only two levels of hierarchy, as shown in Figure 11.15, each node of the overlay network serves a network of the underlying level. To go from one point to another, from A to D for example, the packet must be sent by the local network to the node of network entry overlay, that is to say from A to B on the figure, and then transmitted in the overlay network from B to C and finally in the local network of arrival from C to D.

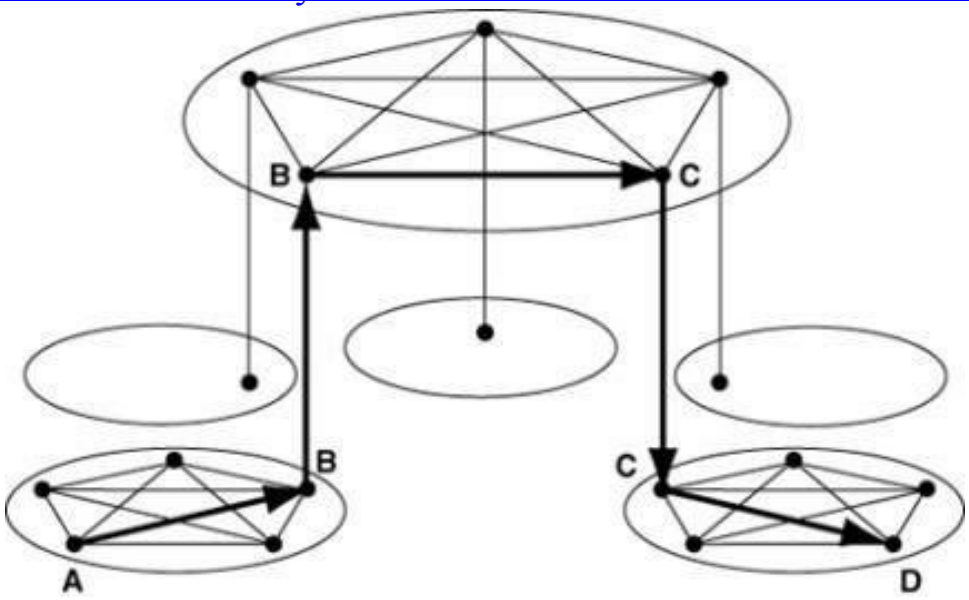


Figure 11.15

Operation of a network overlay

If the different levels of the hierarchy contain mesh networks, which allow you to go directly from one point to another in the network, it can be seen that this solution of network allows to limit the number of nodes to cross. In the case of the Figure 11.15, to go from A to D, it only passes by two intermediate nodes, then that if all the nodes of the network had been at the same level, it would have had perhaps a dozen of hops.

The hierarchical structure of the transmission media of GMPLS-allows you to put in place this type of

network. It may, for example, in a simple case, have areas SPLM to level 2 interconnected by a network overlay Using a wavelength on an optical fiber. This overlay network connects the points of the basic areas selected to be part of the network overlay.
To open paths on different networks to each other, a set of control protocols and monitoring is necessary. A first problem posed by the routing in the overlay networks concerns control of the connectivity, which is supported by messages of type hello, sent regularly on all interfaces. Each Hello must be acquitted explicitly. When no ACK is received, the line is considered as failed. In the case of GMPLS-on optical fiber, it is not possible to send hello messages. Control of the connectivity must therefore be done by a new protocol.
A second problem posed by the overlay networks comes from the impossibility for nodes of the same level, but not belonging to the same domain to transmit directly of control messages. It is necessary to go through a network of higher level, which may not be able to interpret the messages of the lower levels. There is therefore no overall vision of the network.

Control and management of MPLS

To improve the control and management, it is necessary to separate the user plans, management and control, particularly if the network is complex. This applies even more in the networks using the optical fiber.
As for the ATM, it distinguishes three plans in GMPLS:

- The plan user, who is responsible to carry user data from one end to the other.
- The Control Plan, intended to put in place the virtual circuits and then to destroy at the end of the transmission or to maintain them if necessary.
- The management plan, which carries the necessary messages to the management of the network.

The working groups of GMPLS-have developed such an architecture to allow for control by a specific plan all of the components of the network.
To adapt to the GMPLS protocol, signaling protocols (RSVP-TE, CR-LDP) and routing protocols (OSPF-TE, IS-IS-Te) have been extended. A new protocol management, called PML (Link Management Protocol), has been introduced to manage the user plans and control. LMP is an IP protocol that contains extensions for RSVP-TE and CR-LDP.
The Table 11.5 summarizes the properties of these protocols and their extensions in the framework of GMPLS-.

Protocol	Description
Routing (OSPF-TE, IS-IS-Te)	Intended to the automatic discovery of the topology of the network and to the extent of the availability of resources (bandwidth, type of protection). The main improvements are the following: - - - - - - - Indication of the type of protection (1+1, 1:1, not protected, traffic in more). 1+1 indicates that a backup path that is permanently open, 1:1 that in the event of a fault a failover path is planned but without the reservation of resource. - Implementation of lines of bypass to improve the passage to the scale. - Acceptance and indication of connections which do not have IP address; use of an identification Link ID. - Identity of the interfaces of entry and exit (interface ID). - Discovery of a path for a back-up using a different path of the primary path (shared-risk link group).

Signage (RSVP-TE, CR-LDP)	Intended for the establishment of the paths by a traffic engineering. The main improvements are the following: - Exchange of references with networks not package (generalized reference). - Establishment of LSP path bidirectional. - Signals for the opening of a path of back-up. - Proposal of suggested references. - Accepts the switching of wavelength.
LMP (Link Management Protocol)	Includes the following extensions: - Control Channel Management: establishes, during the negotiation, the parameters of the connection, such as the frequency of issuance of messages keep_alive and hello. - Link Connectivity Verification: allows you to ensure the physical connectivity between the nodes neighbors thanks to messages of type ping. - Link property Correlation: determines the mechanisms of protection. - Fault Isolation: isolates the simple mistakes or multiples of the optical domain.

Table 11.5 Properties and Extensions of the protocols of GMPLS-

The different layers examined form above the architecture called multilayer GMPLS: frame, temporal slot, wavelength, set of wavelengths, optical fiber, group of optical fiber.

Control Plan GMPLS

One of the difficulties encountered in establishing the LSP is to find the best path, taking into account the multiple layers of the architecture. For example, it is possible to open an optical link connecting two optical switches and through several other switches in a fully transparent way. Of this fact, this link, often called Te-Link, is seen as a connection to a jump. The optimization of the path to open thus has every interest to pass by Te-link of the lowest level possible.

The architecture of the control plan to achieve the opening of the LSP is shown in Figure 11.16. This architecture contains the lower layers of the architecture GMPLS, with the different possibilities to carry IP packets to check on the various commutations accepted by GMPLS. The IP packets are routed by routing protocols of OSPF type-TE, that is to say taking into account the traffic engineering. Once the path is determined, a reservation is done, essentially by the RSVP-TE. Other possibilities, such as CR-LDP or BGP, may be employed, but they have not yet met with the same success as RSVP-TE.

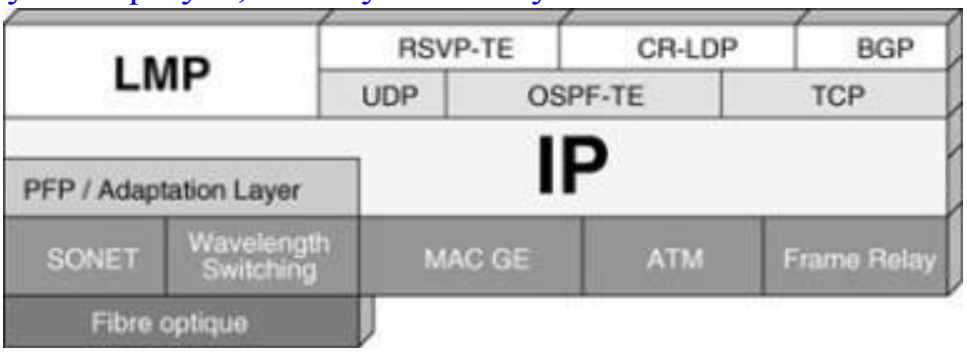


Figure 11.16
Architecture of GMPLS-

The plan for the control of GMPLS-will require yet many developments and tests before to be really optimized. Today it is mainly used for the optical part, but as there is no buffer in the nodes, the openings and closures must be done on the fly.

To put in place of the LSP to be done several hours in advance and often in a manual way should succeed an automatic process to open and close the LSP almost instantly as the requests.

Conclusion

The MPLS networks and GMPLS are promised to a bright future. All the major operators have invested in this direction, not without a certain apprehension regarding the overall complexity of this new architecture, which can be viewed as a compromise between a large number of different architectures. The proposal MPLS-TP is the solution to these problems and should become the very major standard networks of operators.

The plan user seems well designed and allows you to optimize relatively easily for the establishment of the network and its engineering, in particular with regard to the quality of service, the maintenance and management.