SECTION 1
SOME KEY NOTIONS FOR NETWORKS

***

*_ Laurent Toutain_*

Lecturer at Télécom Bretagne

***

Lesson 2: Black box interactions

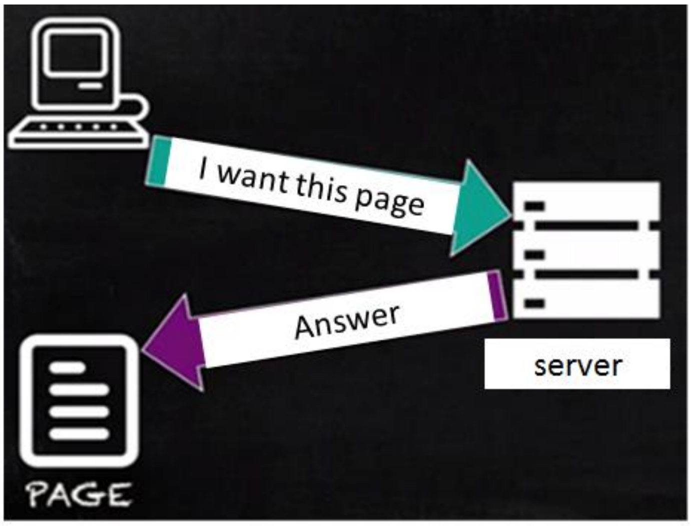
The previous video gave an insight into the complexity of networks. When we think about networks, we think of a tangle of different colored cables and blinking lights, and it may seem complicated to put all these elements together to provide services that enable us to surf the web, watch videos online or make telephone calls.

In fact, underneath all this complexity there is a hidden structure:

We are going to take a look at what is behind it all.

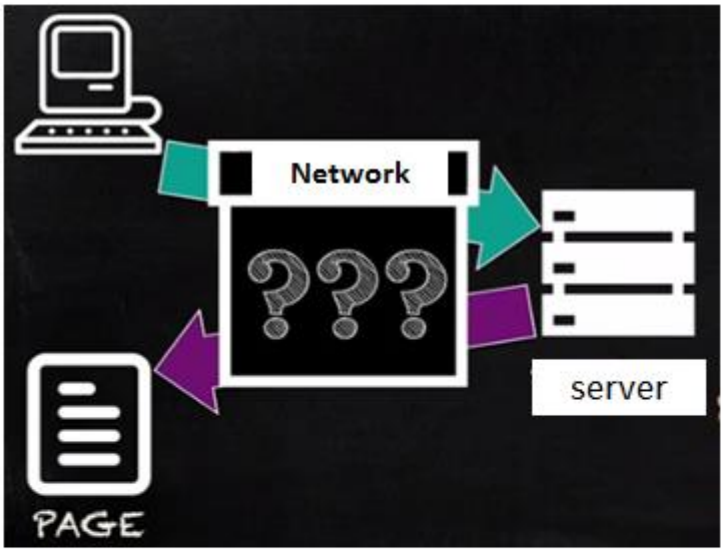
For a network to work on a large scale, its basic principles must be simple. Instead of seeing a network as a jumble of cables and equipment, we are going to try and present it in a simpler way.

When we surf the web, we type the name of a server and send it a message that says "I want this page". The server then processes this message and sends us a response.



We will imagine that the network is a “black box”.

We are interested in the interaction that takes place between us and the “black box”. How does data travel from one side of the box to the other? It is a mystery! All the user is interested in is that it should work! This sort of abstract view may also suit programmers developing applications using the network. But not us, of course! We are going to have a look at what is inside the “box”.



Inside the network

So, when I type the name France Université Numérique MOOC:

–> (www.france-universite-numerique-mooc.fr),

The first stage is that we have to go from a name that is relatively easy to remember to an address that can be understood by the network:

--> (193.48.168.199),

like when you go from the name of a person to their telephone number in a phone directory.

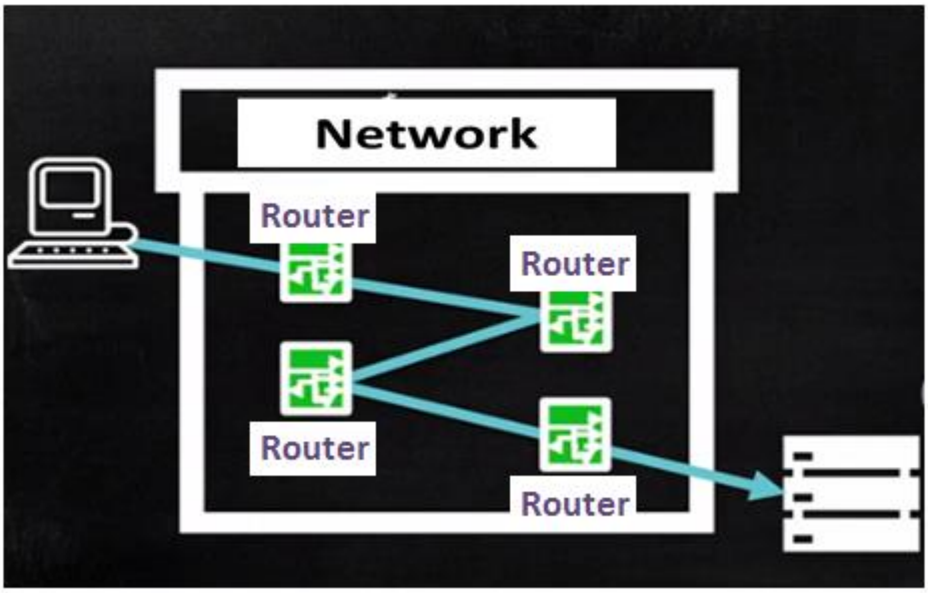
The transition process from name to address can be likened to a “black box”. So we are only going to look at the interactions with this “black box”:

- I want the address

- I have the address.

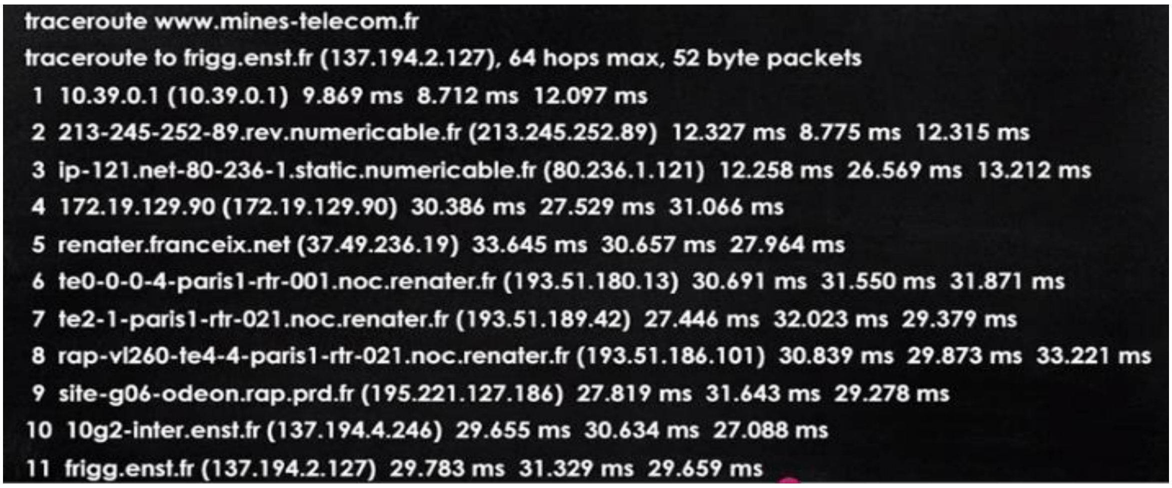
Once the address has been found we can send data over the network to obtain the information we want.

There is no direct connection between our computer and the servers at France Université Numérique. The information is copied from one intermediary device to another until it reaches the server. We will call these intermediary devices **routers**.



Routers analyze the data they receive and select the next router according to the address of the destination. The connection between routers varies depending on whether it uses fiber optics, a cable or radio wave link.

Caroline found the list of routers below when she was looking for the cause of the network problem using the **traceroute** program.



```
traceroute www.mines-telecom.fr
traceroute to frigg.enst.fr (137.194.2.127), 64 hops max, 52 byte packets
 1  10.39.0.1 (10.39.0.1)  9.869 ms  8.712 ms  12.097 ms
 2  213-245-252-89.rev.numericable.fr (213.245.252.89)  12.327 ms  8.775 ms  12.315 ms
 3  ip-121.net-80-236-1.static.numericable.fr (80.236.1.121)  12.258 ms  26.569 ms  13.212 ms
 4  172.19.129.90 (172.19.129.90)  30.386 ms  27.529 ms  31.066 ms
 5  renater.franceix.net (37.49.236.19)  33.645 ms  30.657 ms  27.964 ms
 6  te0-0-0-4-paris1-rtr-001.noc.renater.fr (193.51.180.13)  30.691 ms  31.550 ms  31.871 ms
 7  te2-1-paris1-rtr-021.noc.renater.fr (193.51.189.42)  27.446 ms  32.023 ms  29.379 ms
 8  rap-vl260-te4-4-paris1-rtr-021.noc.renater.fr (193.51.186.101)  30.839 ms  29.873 ms  33.221 ms
 9  site-g06-odeon.rap.prd.fr (195.221.127.186)  27.819 ms  31.643 ms  29.278 ms
10  10g2-inter.enst.fr (137.194.4.246)  29.655 ms  30.634 ms  27.088 ms
11  frigg.enst.fr (137.194.2.127)  29.783 ms  31.329 ms  29.659 ms
```

A recursive model

We see what can be called a recursive model, meaning that we find the same thing repeated several times. So when we connect to the servers of France Université Numérique, this can be presented as a “black box”.

This black box contains another black box whose purpose is to make communication reliable. If there is a loss of data or a transmission error, we will be able to recover the data.

This “black box” relies on other “black boxes” which are the routers that relayed the data to its destination.

The routers themselves rely on other black boxes that enable the transformation of a digital signal into a modulated signal that can be transmitted by a physical connection.

Summary

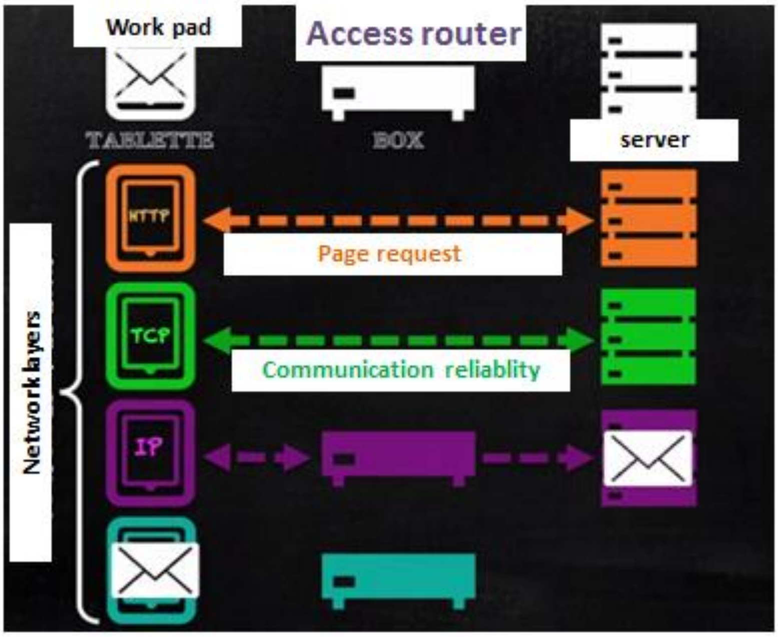
This model based on interaction is very powerful and flexible. We can easily move from one technology to another or keep up with any upgrades. What is important is to define what functionality we expect at each level.

Lesson 3: Network components

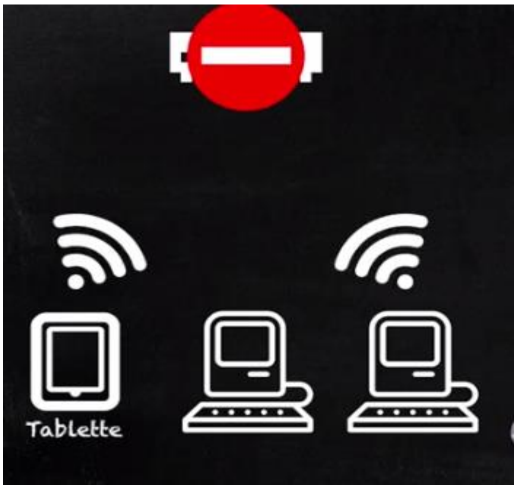
In the previous video we saw that although networks are complicated systems, they can be represented as a series of "black boxes" and interactions. In this video we are going to take a look inside some of these "black boxes". Caroline's tablet contains software enabling her to view web pages and videos. The program relies on an operating system that manages the device's resources. This operating system also manages the network.

- One layer within this operating system serves to make communication between the computer and the server more reliable: we will call it TCP from now on.

- Below that, there is the IP protocol that enables the routers to send packets between each other until they reach their destination. In Caroline's tablet, the IP protocol is in communication with the Wi-Fi. The Wi-Fi modulates the signal between the tablet and the box in order to transmit computer data.



As you know, Wi-Fi allows several devices to be connected together. A protocol means that only one device can communicate at a time, because if two signals overlap, neither will be comprehensible and the information will be lost.



Wi-Fi also allows modulation to be adapted to the quality of the network.

E.g. In the case of Caroline and Adrian's breakdown:

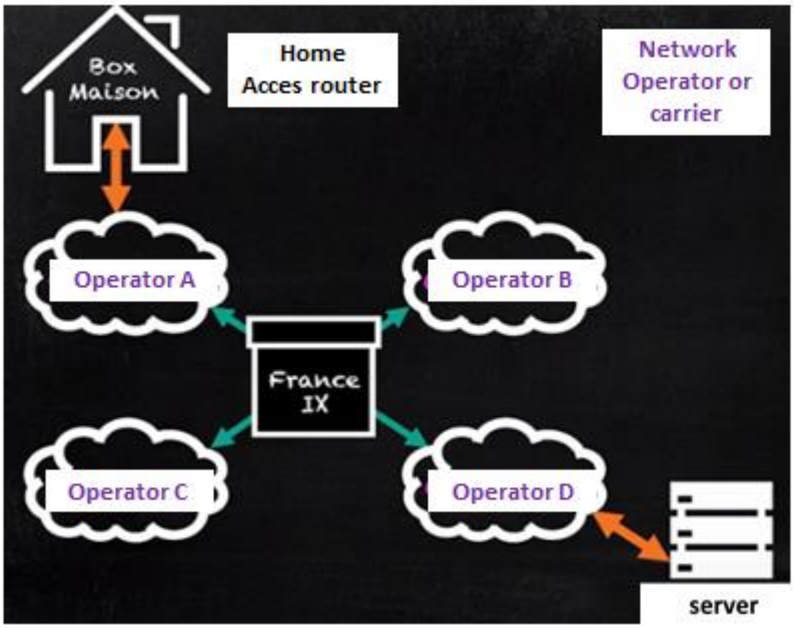
- one hypothesis could have been a modulation problem. Insufficient modulation could have prevented the video being sent.

- another hypothesis is that there may have been too much traffic generated by the neighbors, for example, which would have reduced Caroline's bandwidth.

Either way, checking the light on the Wi-Fi box is no good because all this shows is that the Wi-Fi is turned on.

The data sent by Wi-Fi will be received by the box, which retrieves the IP packets and analyses the recipient. Since that recipient is not in the house, the packet will be sent via ADSL and received by a router belonging to the operator. The operator's aim is to group its traffic together on higher-speed connections with the aim of reducing the cost of interconnection.

But the recipient is usually not on the operator's network, and the information therefore has to be sent to another operator. There are several ways of carrying out these exchanges of information: operators either exchange their data traffic directly or use neutral locations such as "France IX" (**France IX** is a French internet exchange point created in June 2010). In English this is referred to as a GIX (Global Internet eXchange).



We met Franck Simon, the Director.

What is a neutral internet exchange point or GIX (Global Internet eXchange)?

“A GIX is an internet exchange point. It is a sort of crossroads for data exchange where there are a lot of telecommunications players. Thanks to this crossroads, data can be exchanged via short routes that offer excellent quality. By connecting to this point you can reach a large number of internet destinations, with a high-quality service.

The bit rates here are much greater than what you get at home, because here the flows are aggregated. The clients of an exchange point are service providers such as operators. They aggregate or collect together a large number of customers themselves so there is a large quantity of data. This is why the units used are gigabytes, not megabytes like you may have at home.”

In the video we saw that the cause of the problem was at the Institut Mines-Télécom. We can therefore understand the advantages of sites like dailymotion which have interconnection points in different locations with operators and which can offer a better quality service. At the Institut Mines-Télécom there was a break in connection meaning that the traffic was sent via another, slower connection, and because the internet cannot recognize the difference between flows, some of the packets were lost. Although this loss concerned both the data and video outputs, it was much more noticeable on the video, because on the data flow it would simply have led to a much longer transfer time.

You can see that there are multiple challenges in managing a large-scale network. We must be capable of processing errors that occur, such as data loss, as well as building structures allowing the packets in the network to be sent to their destination. We must also be capable of managing periods of saturation that may occur at certain points in the network. On the internet, emitters have to reduce their output when they detect such saturation.

Lesson 4: The postal service: a network service

In the previous Lesson we saw how the life of a packet is organized in internet networks. We are now going to gradually develop these concepts. Historically speaking, one of the very first networks was the postal network. Postal services function thanks to their clients: anyone able to read and write. Each client must follow certain rules in order to use this network.

- The first rule is to have an address, meaning a place where they receive their mail.

- Another rule is to use a certain format on the envelope, notably to write the address of the destination.

Caroline Mooc
2 rue des Octets
101010 Résonum

We can also write our own address on the back, either so that the recipient can reply or so that the postal service can send it back to us if there is a problem.

If we go back to our “black box” model in order to represent the network, we see two points of interaction with the service:

- the client places the letter in a public “interface”: the letterbox

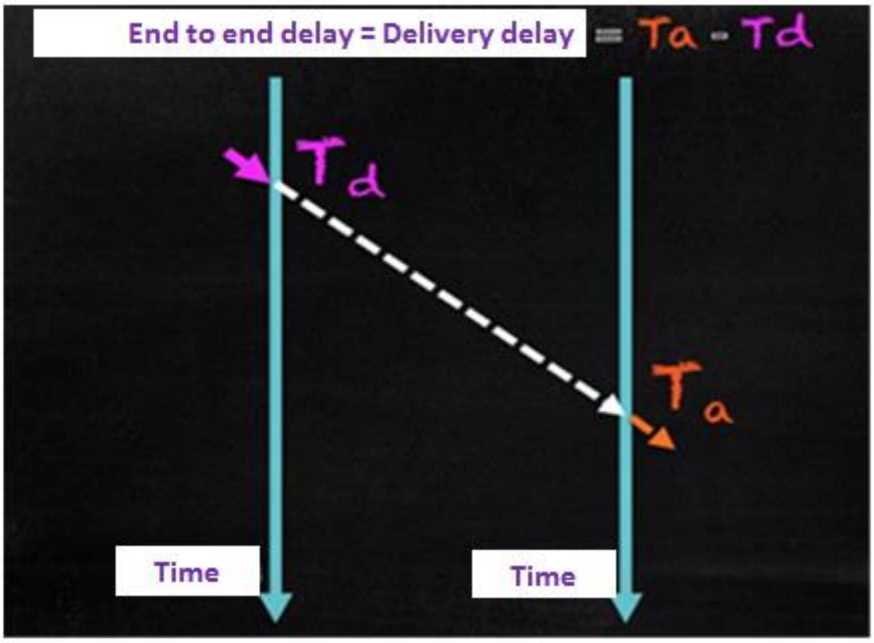
- the postal service puts the mail in a private “interface” at its destination.

The recipient will collect the mail when he wants. The service offered by the Post Office has its advantages, notably enabling communication with a far-off correspondent. However, it also has certain limitations. Postal services limit the weight of letters, so you can only send a certain number of pages per letter. So if you have a long document to send, you have to divide it into several letters and send each one separately.

The amount of time the letter remains in the network is also variable. The letter may remain in the post box where the sender posted it for a certain amount of time. Next, it is collected by a postman who sends it to sorting centers. Finally, there will be a lapse in time before the recipient can collect it from their letter box. Another interesting feature of the network is that the neither person receiving the letter nor the person sending it knows by what route it arrived at its destination. Last but not least, the postal service looks only at the envelope in order to send the letter to its destination and is not at all interested in the written content inside it.

This exchange can be represented chronologically in a diagram.

The letter was posted at a moment $t_{d,}$ travelled through the postal network and arrived in the recipient's letter box at a moment t_a. The journey time is of course t_a-t_d.



The postal service is not perfect; the amount of time spent in the network by the letter is not always the same, and it cannot be guaranteed that the letters will be received in the order they were sent. The sender and recipient must therefore develop strategies to resolve the problem. Let us take the simple example of a head chef and a kitchen hand. The head chef is teaching the kitchen hand his omelet recipe, and he sends him three letters containing the recipe.

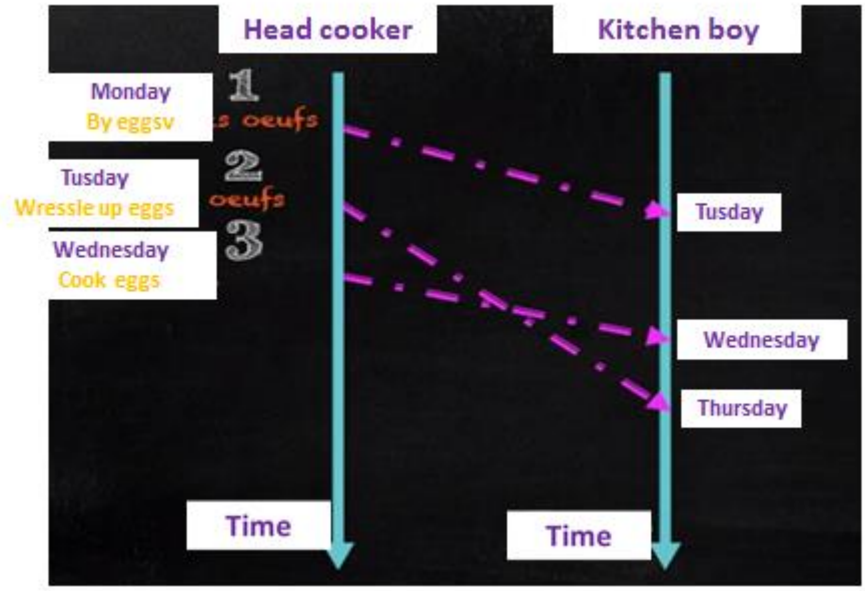
He sends the first on Monday, the second on Tuesday and the third on Wednesday.

The kitchen hand might receive them in a different order, such as,

- the first on Tuesday,
- the second on Thursday
- and the third on Wednesday.

If he follows them in this order he will not get his omelet!

One simple strategy is to number the letters. This way, the kitchen hand will know that the number corresponds to the order so that if he receives the first letter and then the third, he will wait for the second letter before attempting the omelet recipe.



Of course, this only solves part of the problem. I recommend taking a look at the document (Pierre Rolin's book) because in this document we outline different strategies to resolve all the problems caused by the postal network.

Summary

The strategies employed by the chef and the kitchen hand are like a **protocol**. A protocol is a set of rules accepted by both parties at either end (chef and kitchen hand) with the aim of improving the communication service or, at least, evolving towards the attributes desired for the service. A key point to remember with this chronogram is that we have an overall view of the system, whereas neither the sender nor recipient has such a view. They only have a partial vision of the events that occur locally:

-> I receive a letter

-> I send a letter.

However, neither can be aware of the eventuality "a letter has been lost". They would know only by other means that a letter has been lost, for example if there is a missing number.

Lesson 5: Notions for data networks

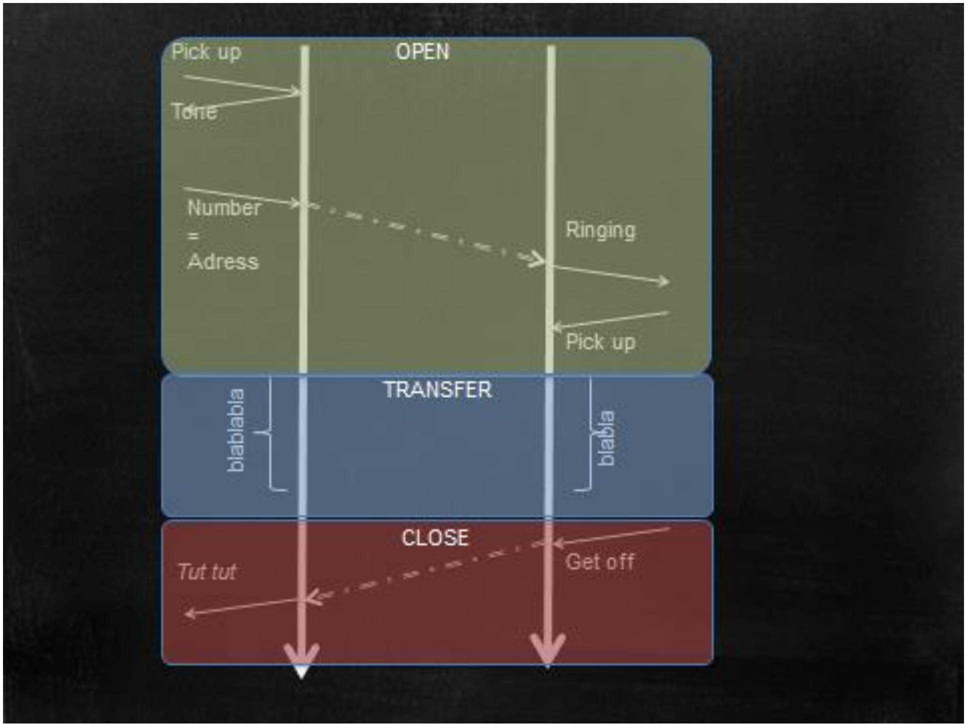
In the previous Lesson we used the postal service as an example and introduced certain concepts such as the packet, datagram, address, communication primitives and protocol. But the postal network is not the only network you are aware of. There is also the Ethernet, Wi-Fi, the internet and of course the phone network.

The phone network

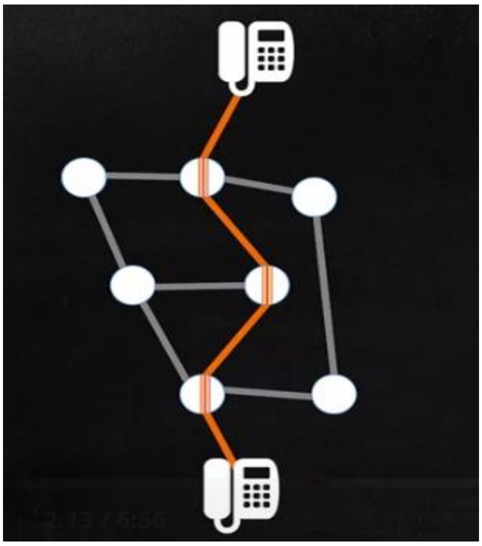
The phone network is another form of interaction with users. You must pick up your handset, type the desired number and the recipient of the call must pick up. Once this stage is complete, you can communicate freely. At the end of the conversation you put down the phone.

So, if we return to our chronograms we can see three types of interaction. The address is only transmitted during the first phase, during which the connection between the sender and recipient is established.

Next, there is a second phase when we transmit the information. During this phase the address is not repeated, in contrast to the postal network where the address has to be sent with every letter. At the end of the communication we hang up and the resources provided by the network are freed.



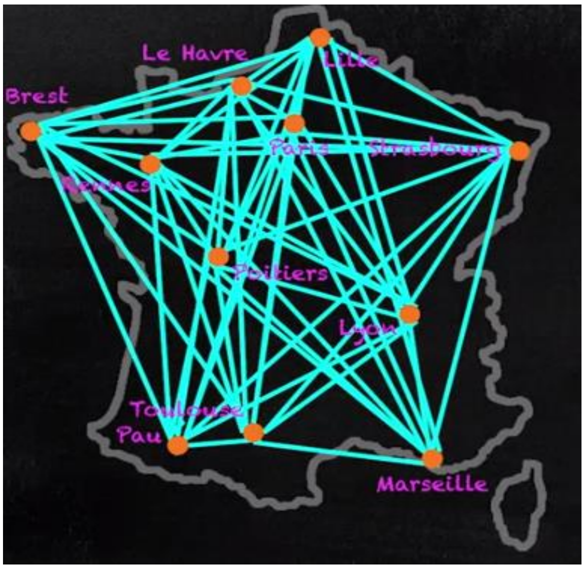
For a long time the telephone network was analog and communication occurred via copper wires placed end to end. The digitization of sound enabled this data to be transported in the form of little blocks similar to packets, but the principle of the circuit remained in the virtual sense.



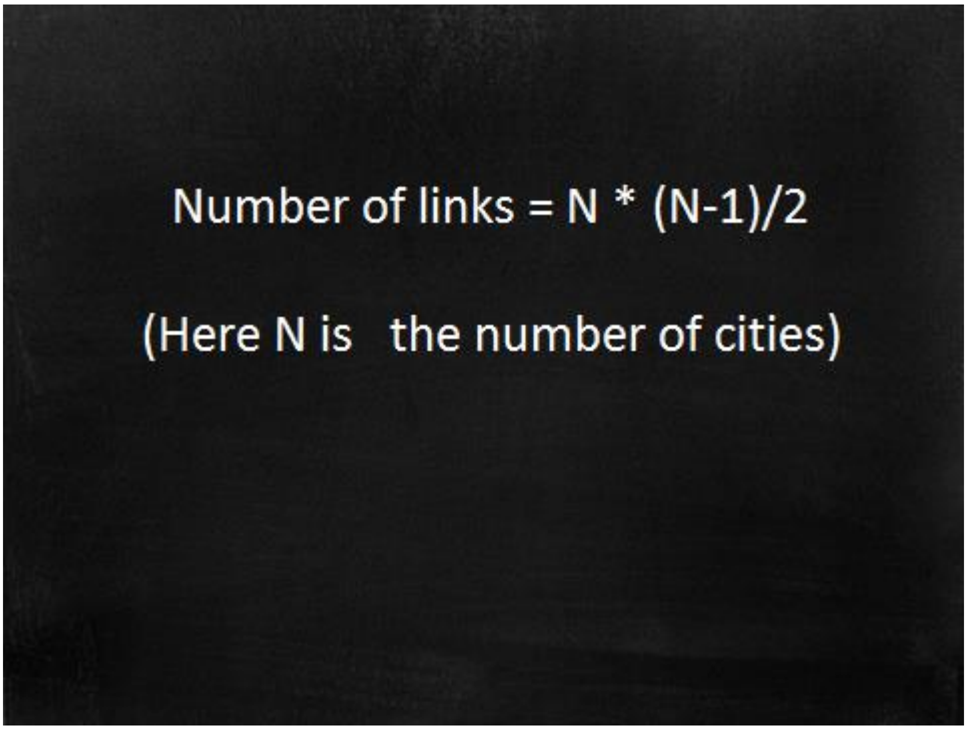
Packets (datagrams) of limited size

This has already been mentioned when we discussed the postal network, where letters have a weight limit. The same restriction exists in computer networks, and the data we emit has a limited size.

But why is there such a restriction? We will understand by looking at an example. Let us take the example of a bank in France that wants to connect its branches. An unrealistic approach would be to establish connections between each branch.



Why is this unrealistic? First of all because it is very expensive. Secondly, there would be relatively little traffic along each of the connections. Finally, it lacks flexibility because if we want to add a new branch into the network, we would have to establish another connection with each of the other branches.



Number of links = N * (N-1)/2

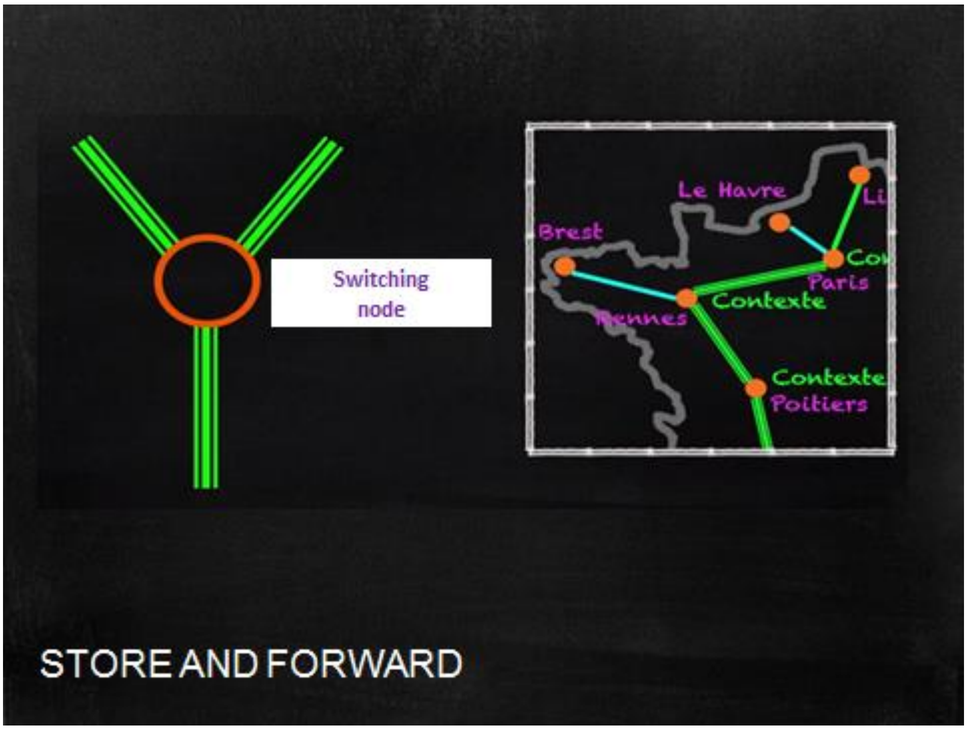
(Here N is the number of cities)

It is better to have an infrastructure or ***backbone*** to limit the number of connections and thus reduce the cost of communication. However, introducing a backbone will create other problems: branches are now only linked to the network via a single cable and so when they receive the data they do not know who emitted it. We could come up with rules such as establishing communication between certain branches during certain time periods, for example a branch in Lille can communicate with the branch in Marseille between 1am and 1:10am, but the problem is that if these branches have no data to transmit, we waste communication capacity.

To optimize communication, we can use one of the two methods that we have already seen. In a connection-based method, the branch communicates with the network to ask it to establish a connection with the recipient branch. The data is then transmitted and the branch, or both branches, cut off the connection. In a more datagram-based method, the recipient's address is transmitted ahead of each message.

Both cases avoid reserving resources. In fact, the messages are stored in each intermediary node. This method is known as store and forward.

The intermediary node receives the message, analyses it, checks it is complete and sends it to the following node, but only if the connection is free, of course. This method enables optimal use of the connections. On the other hand it causes delays, which is not a problem for IT transmissions because computers are generally patient.



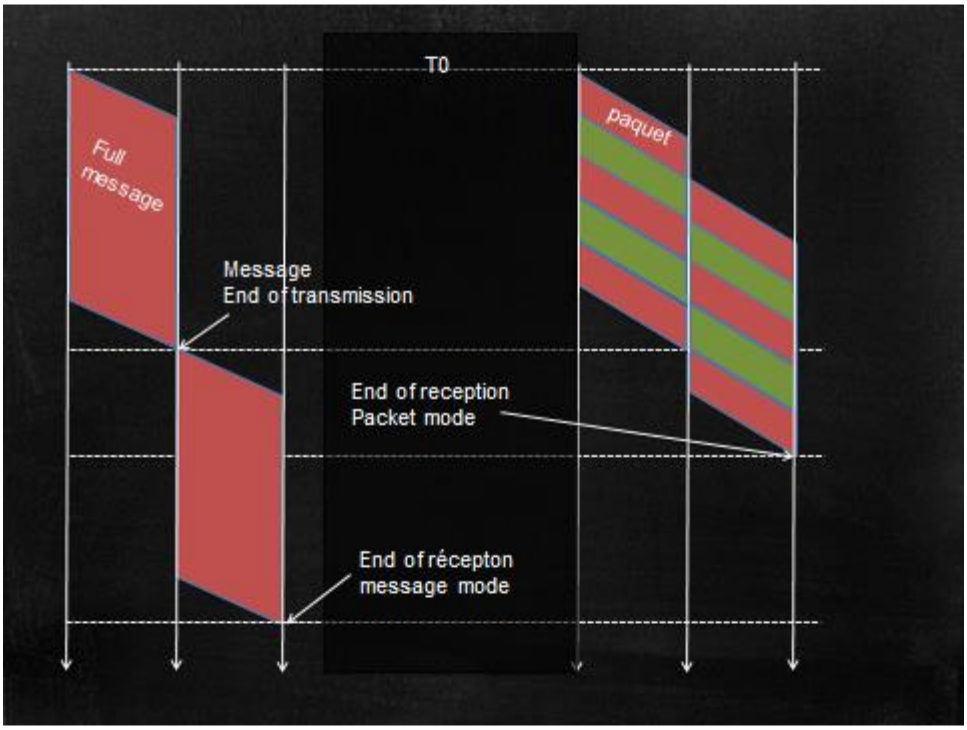
Switching node

STORE AND FORWARD

If we did not limit the size of the messages to be sent on the network, the intermediary nodes would need almost infinite resources in order to store all the information. The size restriction is therefore important because it allows smaller memories to be used.

In addition, the use of small blocks offers several advantages in terms of the network. Firstly, if there is a transmission error in one block, we do not have to re-transmit the whole sequence.

Secondly, transmission time across the network is significantly reduced because we can transmit one block and receive the following one at the same time.



Summary

As you have guessed, these small-sized data blocks are known as packets. The transmission time for a packet is important because it affects other communications. We must therefore limit this time as much as possible. One way of doing this is to have faster connections. In addition to transmitting more data, they will improve the quality of communication. This is very important for certain types of traffic, such as voice communication, which is sensitive to delays.

Lesson 6: Let's talk about addresses

Addresses play an important role in networks. Each type of network has a different address system but, in general terms, we can say that in order for an address to be effective:
- It has to be efficient in administrative terms: each device and object must be able to retrieve an address easily;
- It has to be efficient in technical terms: each device and object must be able to locate something easily in a network.

Addresses and identifiers
We must be careful not to confuse "addresses" and "identifiers".
- An identifier is unique but will simply identify a machine.
- An address enables it to be located within the network. We cannot expect the postman to use our passport numbers to deliver our mail. Here are a few examples of addresses.

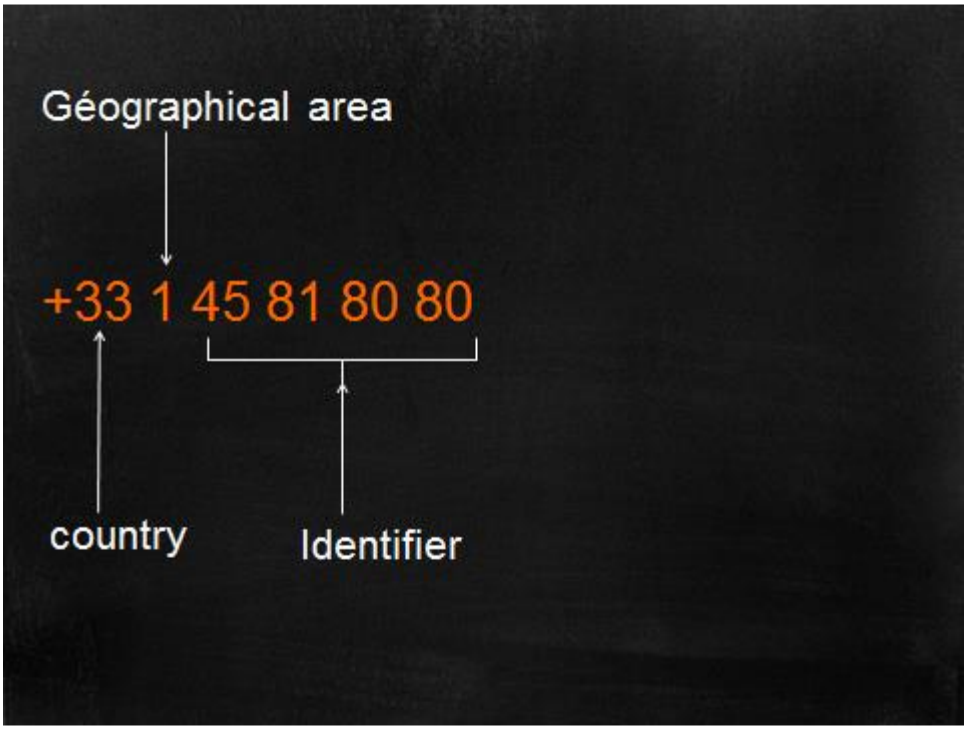
Examples of addresses

- **The postal network** that we have already looked at uses an address structure.

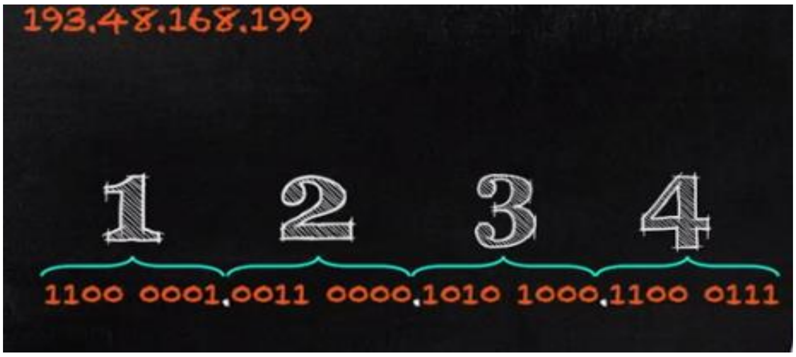
The address is hierarchical, and the hierarchy can be seen by reading the address from bottom to top. First of all there is the country, then a city and post code that can indicate the location within the city, then a street name, a street number and finally the name of a person.
When we read an address in this way we can locate the addressee.

- **Telephone numbers** also have this feature. Reading from left to right, we can see the country, a region within the country, an area and a recipient.

What is interesting, however, is that with the arrival of mobile phones and number portability, phone numbers are changing from an address into an identifier: its only role is that of being unique to a customer.

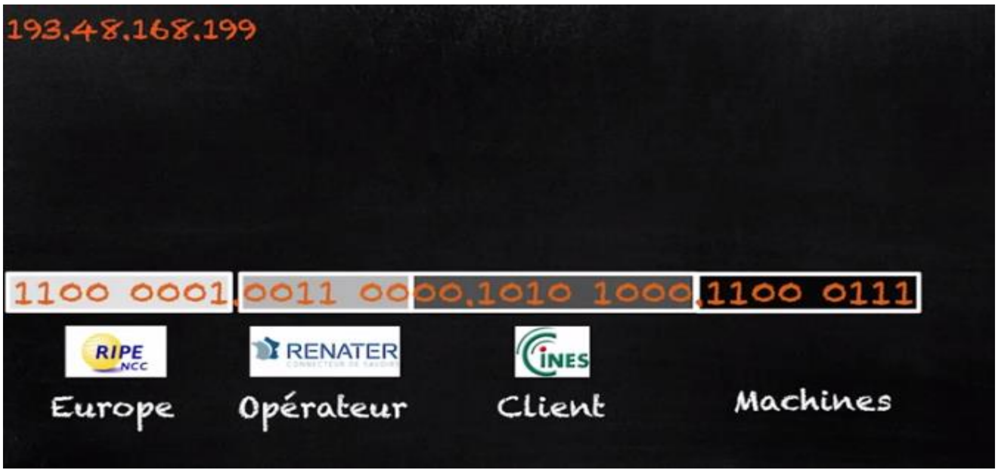


The internet network also uses an addressing system. As we saw in previous Lessons, an address can be broken down into four bytes separated by dots. An important characteristic of addresses within the internet network is that they have a fixed size. However, there is no apparent hierarchy as there is with the postal and telephone networks. In fact, the hierarchy is hidden.



Generally speaking,
- the first byte indicates an area
- the next two bytes indicate an operator within this area and a client of the operator in the area.
- the final byte is used by the end customer to number their machines.

Nevertheless, the limits are not as clearly defined as in the example that I have just given.



The internet also has another addressing system: domain names.
A domain name is a hierarchical structure.
E.g. www.mines-telecom.fr

On the right is the top-level domain, which is a name such as ".com" and ".net" (the number of these names is going to increase significantly over the next few years) or the name of a country such as ".fr".

Next (e.g. www.mines-telecom.fr) there is a subdomain that refers to a company, a service and then the name of the machine. A system known as **DNS** (**D**omain **N**ame **S**ystem) allows us to go from the domain name to an IP address.
PS: It is important to note that the IP hierarchy and the domain name hierarchy are completely separate.

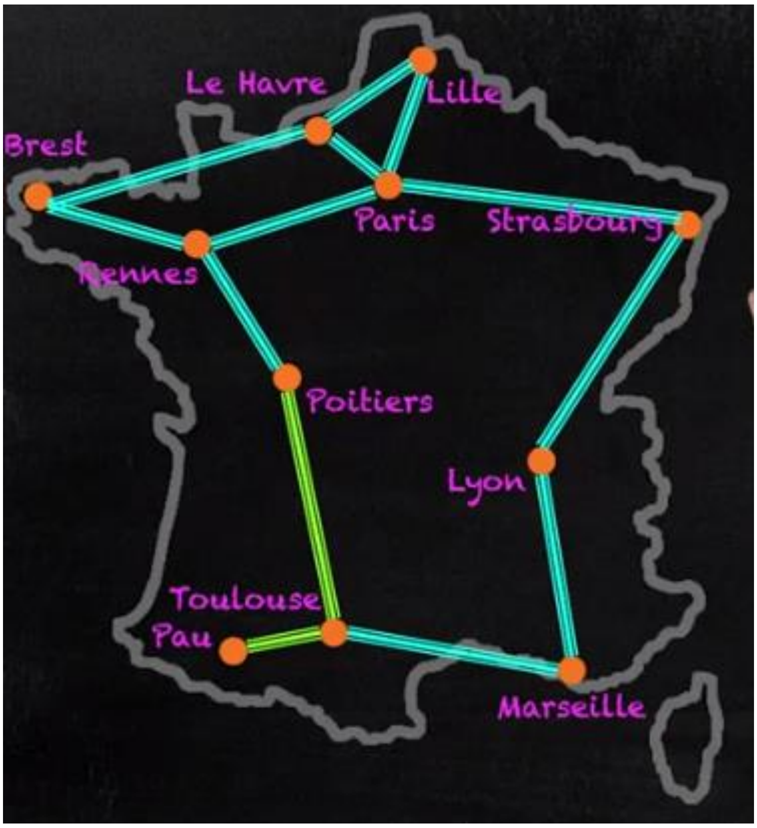
Some systems re-use the addressing scheme or the internet naming plan. The URLs that are used to find web pages online therefore contain either the address or the domain name.

Eg.: http://www.mines-telecom.fr/accueil

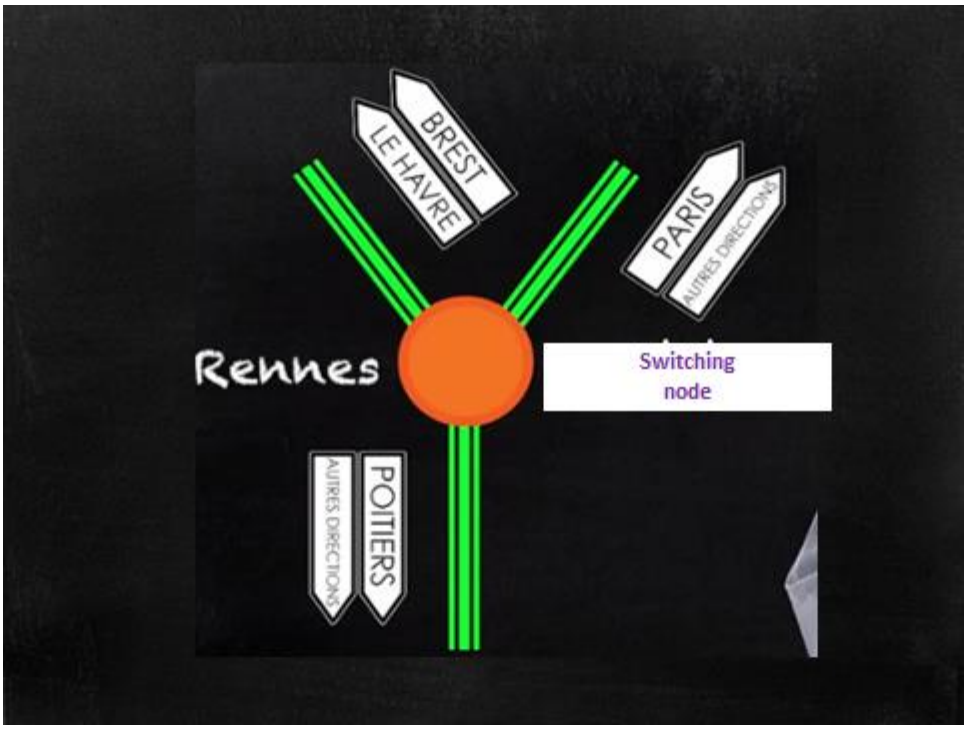
The first part of a URL indicates the protocol used to find the resource on the server. It is generally http or https. Next, there is either the name of the machine or its IP address, and finally an arborescence to locate the resource on the server.

Lesson 7: Architecture

In the previous video we saw that a fully interconnected network was unrealistic. Network architecture therefore relies on **partially interconnected** networks. However, it is important to have several possible routes between two network nodes so that any problem with a link or an intermediary node does not break the connection between the two devices.

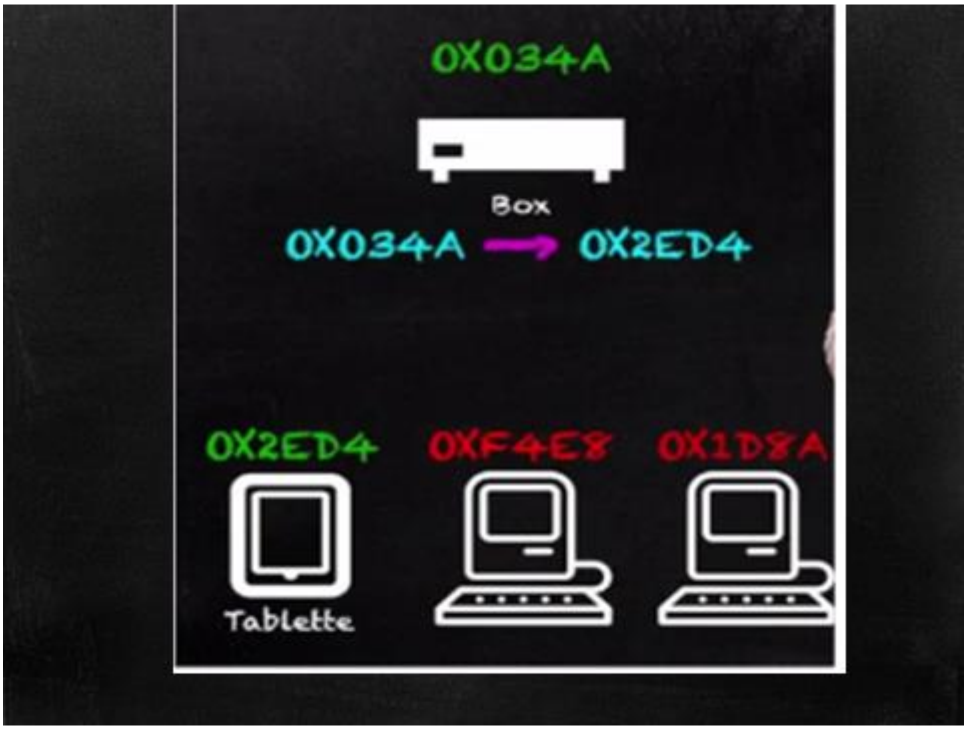


The nodes in the middle of the network direct information according to the address contained in the data. They can be compared to crossroads with road signs.

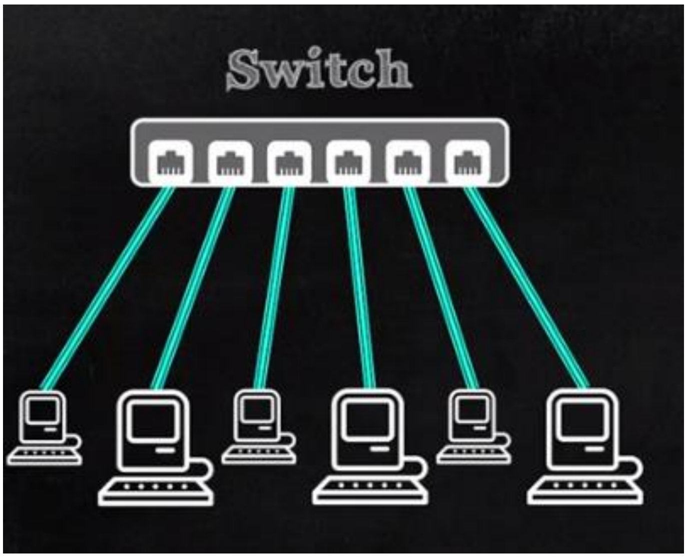


The signs offer very precise directions for close destinations, and much less precise directions for destinations that are further away, so there may be a sign saying “other directions” which indicates a very large number of cities or users.

Another network type is the **broadcast network**. We saw an example of this with Caroline and Adrien’s Wi-Fi network. All the connected machines could communicate directly. The address is therefore important in order to identify the sender and recipient. However, this address does not need to contain information on the location because everybody receives the data. In terms of our definition, the address here is more of a name than an address.



There is one final network type that is very commonly used: the star network. In this case, a device is connected by one link to a central machine.



This is how Ethernet networks operate.