INTERNET : LA TOILE DE FOND DU WEB

On parle beaucoup du Web, la toile internationale de services et d'information et on en oublie qu'elle repose en fait sur une toile matérielle bien réelle, celle des réseaux interconnectés qui forment l'Internet. Internet n'est pas le Web. Internet est l'interconnexion globale de réseaux et d'ordinateurs indépendants alors que le Web n'est qu'une application qui utilise l'infrastructure d'Internet. Nous allons survoler ici cette infrastructure pour essayer de comprendre les grands principes qui régissent ces réseaux et leur interconnexion... interconnected networks... internetworking... internetting ... internet, la technologie qui donna vie à Internet.

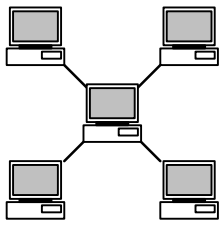
I. RÉSEAUX LOCAUX

Les réseaux se divisent couramment en deux grandes familles:

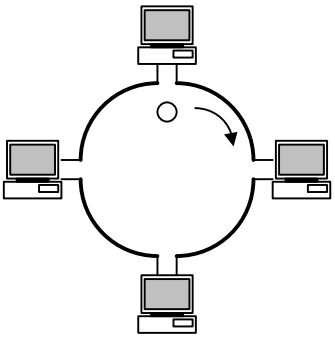
- Les WAN (Wide Area Networks) : réseaux de grande superficie qui connectent des réseaux sur une zone géographique assez large : entre plusieurs immeubles, villes ou pays.
- Les LAN (Local Area Networks) : réseaux de petite superficie qui connectent des ordinateurs à l'intérieur d'un bureau, d'un étage ou d'un bâtiment. Ils se connectent en général à un même médium que l'on appelle le backbone (colonne vertébrale ou épine dorsale en français). Ce backbone relie en général tous les réseaux locaux d'un même organisme et permet aussi de les relier à d'autres réseaux locaux, à d'autres backbones et à des WAN.

La plus part des organisations relient donc leurs ordinateurs en les organisant en un ou plusieurs réseaux locaux. Les postes sont repérés par des adresses physiques uniques assignées par le constructeur des adaptateurs/cartes réseau et appelées adresses MAC (Media Access Control).

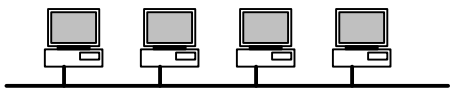
Il existe trois grands types de topologies de réseaux locaux (LAN) qui sont les topologies en étoile, en anneau ou de bus. Des topologies plus complexes peuvent être obtenues en combinant ou en dérivant ces topologies de base.



Dans la topologie en étoile un serveur central reçoit et distribue les données du réseau. Le trafic entre chaque nœud est donc minimal puisque seulement l'information destinée ou en provenance d'un poste transite sur le câble entre lui et le serveur. Si l'un des postes tombe en panne cela n'a aucune incidence sur la communication entre les autres postes. Par contre si le serveur central est ralenti ou arrêté pour une raison ou pour une autre, tout le réseau s'effondre. Cette topologie à donc un talon d'Achille : le serveur central. C'est la topologie typiquement utilisée par les mainframe : de gros ordinateurs centraux avec un ensemble de terminaux autour.



Dans la topologie en anneau, les postes sont reliés entre eux pour former une boucle. L'ordre d'accès au réseau se fait grâce à un unique 'jeton électronique' qui est passé d'un poste à l'autre tout autour de l'anneau en bouclant sans cesse. Un poste ne peut transmettre une information sur le réseau que lorsqu'il a le jeton. Chaque nœud du réseau a ainsi la même possibilité de communiquer, ou si vous préférez le même temps de parole que les autres. Les données sont alors émises avec en en-tête l'adresse de l'émetteur et du destinataire, elles font le tour de l'anneau, tous les postes les reçoivent mais seul le destinataire les garde. Le plus gros problème est que si l'un des nœuds connaît une avarie tout le réseau en pâtit. La technologie de réseau en anneau la plus connue est la technologie IBM Token Ring (en français Anneau à Coupon/Bon/Jeton).



La troisième topologie est probablement la plus utilisée, c'est la topologie de bus. Elle utilise un seul médium de transmission (par exemple un seul câble). Tous les postes partagent le même bus et les mêmes communications. Comme avec la topologie en anneau, les données sont émises avec une en-tête identifiant l'émetteur et le récepteur, chaque poste écoute le bus et ne relève parmi les paquets qui passent que ceux qui l'intéressent. Ici si l'un des nœuds a un problème le réseau marche toujours très bien. En contre partie il faut mettre en place un algorithme distribué de communication pour gérer la prise de la parole et les collisions de paquets. Une technologie typique pour les réseau en bus est la technologie Ethernet 2.0.

Une topologie complexe typique est la topologie en arbre obtenue en combinant la topologie en étoile et la topologie en bus. On peut aussi relier plusieurs réseaux d'un même type en utilisant un boîtier de connexion que l'on appelle un pont. Cependant ces technologies de LAN ne permettent pas de relier directement un grand nombre de machines et encore moins de relier différents types de réseaux.

II. RÉSEAUX INTERCONNECTÉS

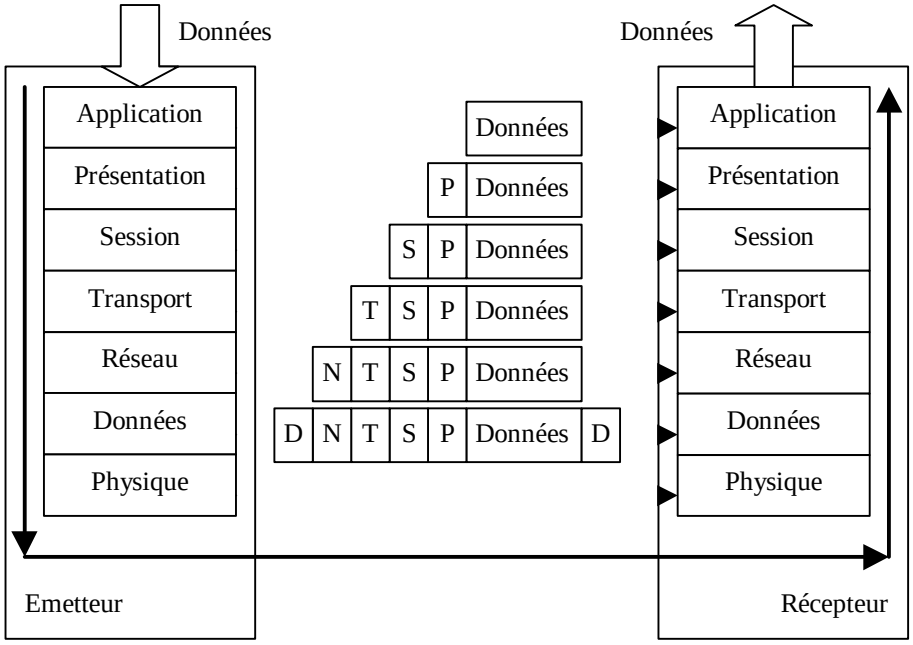
Les technologies des réseaux telles que nous venons de les voir permettent une connexion fiable entre deux ordinateurs reliés au même réseau. Cependant, elles ne permettent pas la connexion inter-réseaux, où des données pourraient être transmises entre deux nœuds appartenant à deux réseaux différents. Cette interconnexion de réseaux est connue sous le nom de 'internetworking' ou internet pour les intimes et chaque sous-partie est appelée un sous-réseau (subnetwork ou subnet)

II.1 MODELE OSI

La transmission obéit au modèle en couche OSI (Open Systems Interconnection) qui est une norme ISO (International Standards Organisation). Les données passent de la couche la plus haute de l'émetteur vers la couche la plus basse et de la couche la plus basse du récepteur vers la couche la plus haute. Ainsi chaque couche étant indépendante tout se comporte comme si une couche de l'émetteur s'adressait directement à la couche correspondante du récepteur créant ainsi un flot virtuel de données entre les couches.

La couche la plus haute (application) reçoit des données d'une application et leur ajoute des informations destinées à l'application du récepteur, puis passe le tout à la couche du dessous. La couche de présentation ajoute des informations sur le format des données, et ainsi de suite jusqu'au niveau physique. La couche physique est responsable de la transmission des données au récepteur. Les données envoyées peuvent être appelées des paquets.

La couche physique représente le niveau technologique où l'on définit les caractéristiques physiques du médium de communication et les signaux transmis (voltage, connecteurs, câbles...). La couche des données s'assure que la transmission est fiable et cherche à détecter les erreurs, les pertes, les collisions. Ces deux premières couches correspondent aux technologies des réseaux locaux dont nous avons parlé en première partie. Elles s'occupent de la communication entre un nœud et le suivant.
L'interconnexion entre les réseaux se joue au niveau des deux couches suivantes : Réseau et Transport. La couche réseau route les paquets au travers du réseau. Si un paquet doit sortir du réseau, la couche de réseau le route à travers les réseaux interconnectés. Cela peut impliquer de découper les paquets et les ré-assembler. La couche de transport permet de mettre en place une connexion fiable et maintenue entre deux nœuds.



Les trois autres couches sont, en général, incluses dans les applications basées sur l'abstraction créée par les couches inférieures et s'occupent respectivement de gérer l'ouverture et la fermeture d'une session, d'assurer de la compatibilité la traduction, le cryptage des données échangées entre deux systèmes et enfin de l'application en elle même (web, e-mail, ftp, telnet,...). Comme le montre le schéma ci-contre, typiquement, les couches physique et de données sont couvertes par les technologies de réseau local ethernet, token ring,... la couche réseau est couverte par le protocole IP (Internet Protocol) qui s'occupe de router les paquets à travers les réseaux interconnectés et la couche transport relève du protocole TCP (Transport Control Protocol) qui s'occupe de la fiabilité des transferts entre deux nœuds. On appelle protocole de communication, un ensemble de règles respectées par les deux partis communicant qui permettent le bon déroulement de l'échange d'information. TCP/IP est un couple de protocoles permettant à deux sous réseaux de communiquer nous allons les voir plus en détail dans la section suivante.

Applications utilisateur, processus …	Application	Applications DNS FTP Telnet Mail…
Interprétation des données et formats	Présentation	
Administration et contrôle des sessions entre 2 nœuds	Session	
Transfert des données indépendamment du réseau	Transport	TCP, UDP
Routage, aiguillage et contrôle des débits	Réseau	IP, ICMP
Contrôle des liens des erreurs et des débits	Données	Ethernet Token Ring Modem…
Caractéristiques et mécaniques	Physique	

II.2 RESEAU ET TRANSPORT

Pour que les données soient transmises au travers des réseaux il faut mettre en place une structure d'adressage qui soit lu par les passerelles (gateway) ou les routeurs (router) qui font le lien entre les réseaux et routent les données jusqu'à leur destinataire. Une gateway est utilisée lorsque les deux réseaux à relier sont de différents types et qu'une traduction doit être effectuée pour passer de l'un à l'autre.

Pour cela, les principales fonctions du protocole IP sont :

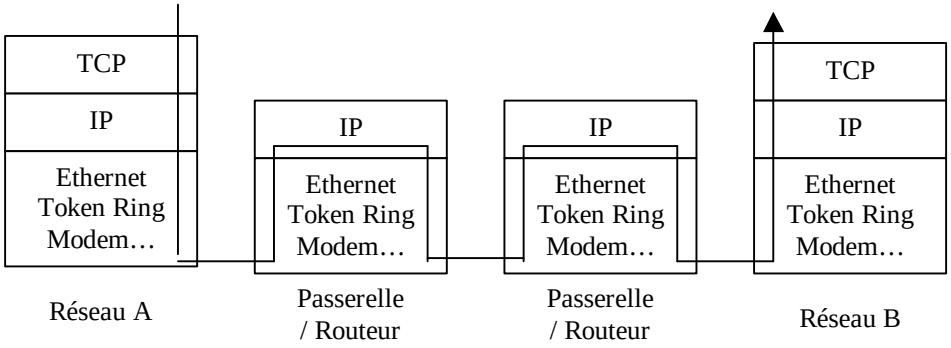
- Router les paquets d'information: Les logiciels du protocole IP installés sur chaque nœud doivent connaître la gateway (passerelle) du réseau. Les gateways connaissent les réseaux qui leur sont connectés. Et c'est ainsi que les données traversent Internet de nœuds en passerelles, de passerelles en passerelles, et finalement de passerelles en nœuds.
- Fragmenter les données si la quantité de données est trop grande pour passer d'un trait.

A IP s'ajoute un autre protocole moins connu (ICMP) servant à reporter les erreurs de transfert (ex: données endommagées, destinataire inconnu) et empêcher les boucles en effaçant les paquets trop vieux.

L'adresse définie pour le protocole IP (appelée 'adresse IP') est de la forme A.B.C.D où A, B,C et D sont entre 0 et 255, exemple : 192.52.154.87. La répartition des adresses est faite par le NIC (Network Information Center) . Les adresses ne sont souvent pas attribuées une par une par le NIC, mais par classes. Une classe d'adresses est un ensemble de numéros que l'organisme pourra utiliser à sa guise. On parle de 'classe A' si la classe attribuée est de la forme A.X.X.X où A est fixé par le NIC et le reste est libre. On parle de 'classe B' si la classe attribuée est de la forme A.B.X.X où A et B sont fixés par le NIC et le reste est libre. On parle de 'classe C' si la classe attribuée est de la forme A.B.C.X où A, B et C sont fixés par le NIC et le reste est libre. De ce fait, les adresses se divisent en deux parties. La première partie identifie le réseau, et la deuxième est entièrement libre pour l'administrateur du réseau qu'il veuille la subdiviser en sous réseaux ou la répartir en numéros de machines. Moins la première partie inclut de nombres, plus le nombre d'adresses pour l'organisme est grand mais bien sûr moins il y a de classes de ce type disponibles et donc plus il est difficile et cher de s'en faire attribuer une. Une classe C permet déjà de numéroter jusqu'à 254 machines.

Les administrateurs réseau ont donc à leur disposition un certain nombre d'adresses voire de classes. Ils attribuent ces numéros, et au besoin créent des sous réseaux en divisant les classes à l'aide de masques (filtres permettant de savoir à quel sous réseau appartient une adresse IP). Ils relient alors leurs réseaux locaux entre eux par des passerelles ou des routeurs appelés gateway IP.

Les gateway IP utilisent l'adresse IP pour router les paquets d'information. Si l'adresse est connue de la gateway alors le paquet est envoyé à son destinataire. Sinon si le réseau ou le sous réseau correspondant à l'adresse est connu de la gateway alors le paquet est envoyé à la gateway de ce sous réseau. Sinon le paquet est envoyé à une gateway par défaut et ainsi de suite jusqu'à ce qu'il soit reconnu ou qu'il soit prouvé que l'adresse n'existe pas.



Les gateways ont donc des tables de routage indiquant les machines de leur réseau local ainsi que les gateway et les masques des réseaux voisins. Les administrateurs réseaux maintiennent ces tables à jour de façon à ce que tous les réseaux puissent communiquer.

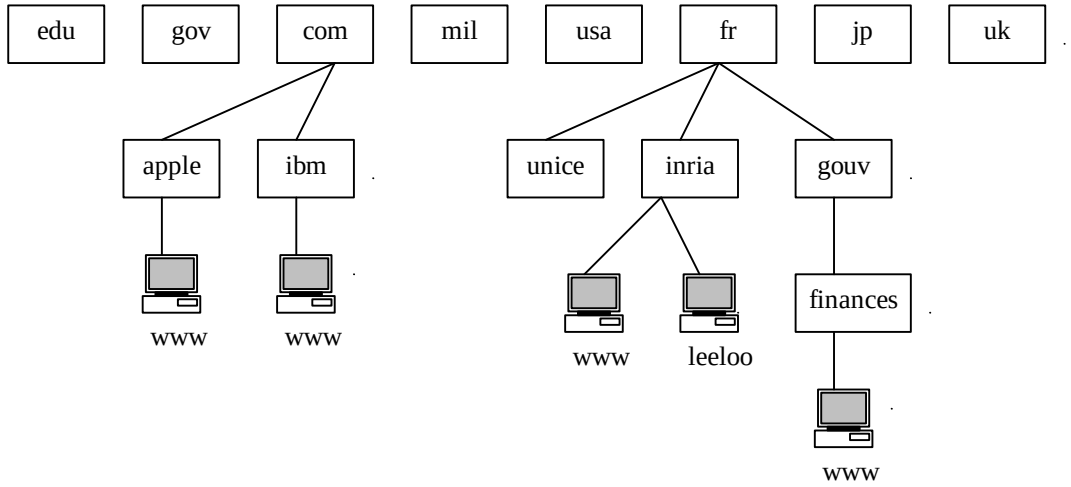
Au-dessus de IP/ICMP sont définis deux protocoles TCP et UDP. Le protocole TCP (Transmission Contol Protocol) permet de mettre en place une connexion fiable et maintenue entre deux nœuds pendant toute la durée nécessaire à leur communication. Une fois cette connexion mise en place tous les messages échangés sont vérifiés, leur réception est accusée et si nécessaires les paquets perdus ou abîmés sont réémis. Les applications obtiennent avec TCP une connexion bidirectionnelle fiable avec une autre application de façon complètement transparente sans avoir à se soucier des architectures et aléas des réseaux. UDP (User Datagram Protocol) est une connexion non fiable et non connectée utilisée lorsque l'application n'a pas d'intérêt à réémettre ses paquets en cas de problème pour l'un d'eux. (par exemple une application qui émet l'heure GMT ne va pas réémettre un paquet qui a échoué puisque le temps qu'elle le fasse, son contenu est déjà périmé).

II.3 LES NOMS SYMBOLIQUES

L'adresse IP est bien pratique pour l'interconnexion des réseaux, mais pour nous humains, il n'est pas facile de se rappeler les adresses IP des machines auxquelles on souhaite se connecter. Il nous faudrait avoir un répertoire IP au même titre que l'on a un répertoire téléphonique pour se rappeler le numéro de téléphone de nos amis. C'est pour cela qu'ont été inventés les noms symboliques.

Lorsque qu'un utilisateur ou un programme utilise un nom symbolique, un service d'annuaire sur le réseau Internet détermine l'adresse IP correspondant au destinataire. Cela permet aussi à un utilisateur ou un programme d'être déplacé sans changer son nom symbolique, seule la correspondance entre son adresse IP et le nom est modifiée dans l'annuaire. C'est exactement le même principe que l'annuaire téléphonique sauf qu'ici la mise en correspondance est automatique.

Les noms sur Internet suivent une certaine structure composée de termes séparés par des points comme par exemple www.inria.fr. La génération de ces noms suit une structure hiérarchique :



A la base furent définis de grands domaines tels que
- edu: organismes educatifs
- gov: organismes gouvernementaux
- com: organismes commerciaux
- mil: organismes militaires

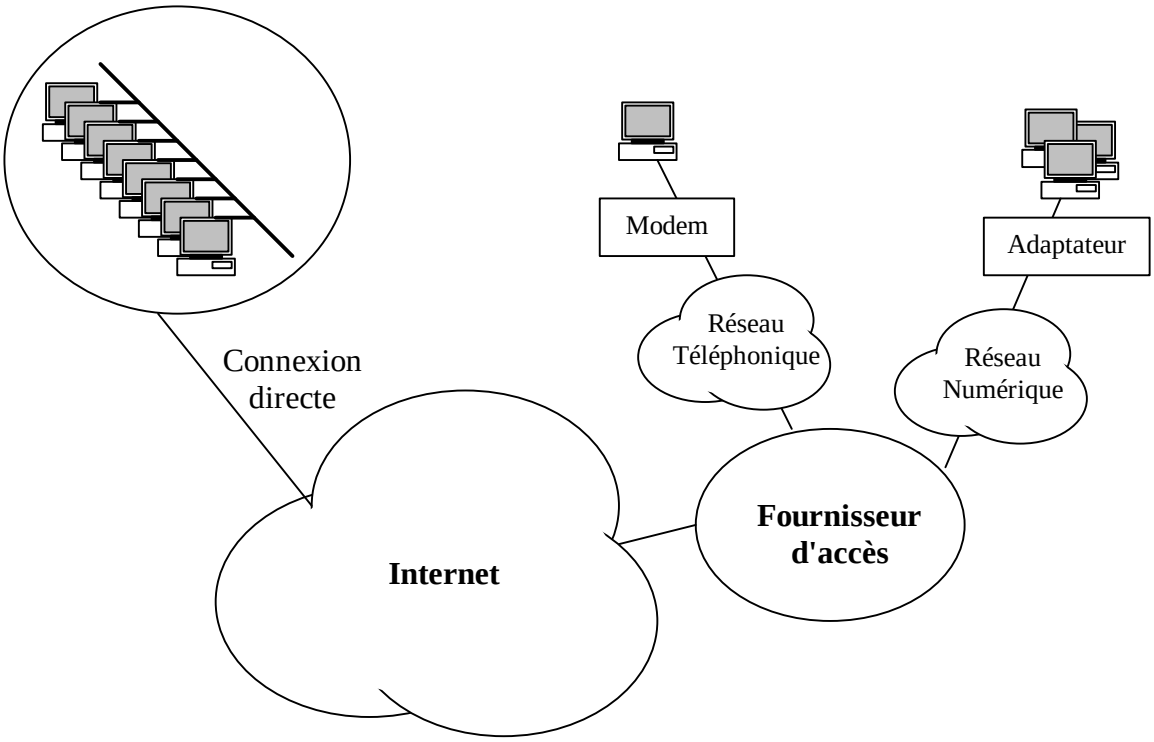
Puis on a introduit des domaines par pays. Chacun de ces domaines est ensuite subdivisé en sous domaines en essayant de garder les noms le plus court et le plus facile à retenir possible. Ici par exemple dans le domaine .fr (France) on trouve le sous domaine .gouv (gouvernement) qui lui même contient d'autres sous domaines parmi lesquels .finances, et dans ce sous domaine il y a une machine sur laquelle tourne un service web (pour publier des pages sur le web) designée par www. Le nom de ce service web et de la machine à contacter pour y accéder est donc www.finances.gouv.fr

Chaque institution sur Internet a une machine sur laquelle tourne le DNS (Domain Name Server), c'est à dire le service qui garde à jour l'annuaire des machines de l'institution qui sont sur Internet. Une machine peut avoir plusieurs noms en particulier si elle remplit plusieurs services. Une machine souhaitant communiquer avec une autre machine s'adresse donc au DNS pour connaître son adresse IP.

III. LA CONNEXION DISTANTE.

Les méthodes que nous avons vues jusqu'à maintenant permettent de connecter un ordinateur à un réseau puis de connecter les réseaux locaux entre eux mais cela suppose qu'il y ait un réseau pas trop loin. Les grands réseaux sont reliés entre eux par des fibres optiques à haut débit qui leur sont dédiées et qui ont parfois été mises en place spécialement pour ça. Mais comment cela se passe-t-il pour une entreprise et à plus forte raison pour un particulier qui veut se relier au réseau ?

Pour les entreprises il y a la possibilité de demander directement une liaison spécialisée vers un opérateur Internet ou de passer par exemple par une liaison Numéris (ISDN) de France Télécom vers un fournisseur d'accès lui-même relié à Internet. Les différentes offres et possibilités sont ici hors sujet, mais le budget et les restrictions attachées à chacune d'entre elles déterminent le choix du client.



Pour les particuliers il faut exploiter les câbles qui arrivent déjà chez lui (téléphone, chaînes câblées, électricité...). La ligne la plus répandue dans les foyers c'est la ligne téléphonique. Normalement le réseau téléphonique n'est pas adapté au transfert de données digitales, les lignes téléphoniques étant des lignes qui laissent passer la voix, (ce qu'on appelle un signal analogique, par opposition au signal numérique). Pour transformer le signal numérique en un signal analogique compatible avec une ligne téléphonique, on utilise un Modulateur Démodulateur, appelé modem. Grâce à son modem, votre ordinateur compose donc le numéro de téléphone du central de votre fournisseur d'accès, à l'autre extrémité de la ligne un modem du fournisseur d'accès lui répond et la communication s'installe. Pour information, les protocoles utilisés pour cette connexion par modem sont souvent PPP ou SLIP; ils sont au niveau physique et données et au-dessus d'eux on retrouve bien évidement TCP/IP ce qui permet à vos applications réseau de fonctionner chez vous comme ailleurs. Le réseau du fournisseur d'accès est aussi connecté à d'autres réseaux, à un WAN et souvent directement à l'Internet. Le plus gros problème c'est que les liaisons téléphoniques ne vont, dans le meilleur des cas, que jusqu'à 56000 bits par seconde. Avec de telles liaisons, on peut transporter un fichier de 1 Mo en 4 à 5 minutes théoriques, c'est-à-dire en fait en 8 à 10 minutes. Ce qui veut dire que pour transférer un CD audio (640 Mo) il faudra 4 jours... Bonjour la facture de téléphone !!!!

C'est pour cela que tout un tas de nouvelles technologies émergent pour améliorer l'accès à Internet, les plus en vogue actuellement sont l'utilisation du câble (des chaînes câblées) et l'ADSL (nouvelle technologie utilisant les câbles du téléphone sans en déranger la ligne). On utilise alors plus un modem classique mais une carte/adaptateur dédiée, les taux de transfert sont bien meilleurs et de plus la machine peut rester connectée sur le réseau en permanence. Les coûts d'installation et d'abonnement sont cependant encore assez élevés.