

Introduction

Les trois derniers siècles ont chacun été marqués par des progrès technologiques spectaculaires. Le XVIII^e siècle a été celui des grands systèmes mécaniques issus de la révolution industrielle, le XIX^e nous a donné la première locomotive à vapeur et le XX^e a été l'ère de la collecte, du traitement et de la distribution des informations. Cette dernière période a aussi connu d'autres développements majeurs, notamment le déploiement de réseaux téléphoniques à l'échelle mondiale, l'invention de la radio et de la télévision, l'explosion de l'industrie informatique, le lancement de satellites de communication, et, bien entendu, l'Internet.

En raison des rapides progrès technologiques que nous connaissons, ces domaines convergent rapidement au XXI^e siècle, et certaines différences qui existaient entre la collecte, le transport, le stockage et le traitement des informations disparaissent progressivement. Une entreprise, si gigantesque soit-elle, peut désormais, le plus simplement du monde, connaître la situation précise de n'importe lequel de ses bureaux quel qu'en soit l'éloignement géographique. Mais à mesure que nos capacités à gérer l'information augmentent, nos besoins en traitements toujours plus sophistiqués croissent aussi.

Bien que plus récente que d'autres industries comme l'automobile ou les transports aériens, l'industrie informatique a accompli des progrès fantastiques en peu de temps. Durant les deux premières décennies, les ordinateurs jouissaient du statut de joyaux technologiques. Ils étaient centralisés, fréquemment placés dans une seule pièce aux cloisons souvent vitrées au travers desquelles les visiteurs pouvaient les admirer. Une société ou une université de taille moyenne n'en possédait peut-être qu'un ou deux, tandis que les établissements les plus grands en comptaient au plus quelques dizaines. L'idée qu'une quarantaine d'années plus tard des systèmes de la taille d'un timbre-poste mais beaucoup plus puissants allaient être produits par milliards relevait de la science-fiction pure et simple.

Le rapprochement du monde informatique et de celui des télécommunications a profondément influencé la façon dont les systèmes informatiques sont organisés. Le concept de « salle informatique » autrefois dominant – une pièce abritant un gros ordinateur dans laquelle les utilisateurs apportaient leurs travaux à traiter – est

aujourd’hui complètement obsolète (même si les centres d’hébergement contenant des centaines de serveurs Internet deviennent courants). Le vieux modèle de l’ordinateur unique répondant à tous les besoins de l’entreprise a été supplanté par un autre dans lequel les traitements sont effectués par plusieurs machines distinctes mais interconnectées. Ces systèmes sont les **réseaux d’ordinateurs**, et ce livre traite de leur conception et de leur mise en œuvre.

Tout au long de ce livre, nous emploierons le terme de « réseau d’ordinateurs » pour désigner un ensemble d’ordinateurs autonomes interconnectés au moyen d’une seule technologie. Deux ordinateurs sont dits interconnectés s’ils peuvent échanger des informations. La connexion physique n’est pas nécessairement réalisée à l’aide d’un câble en cuivre : il est possible d’employer de la fibre optique, des micro-ondes, des ondes infrarouges ou encore des satellites de communication. Comme nous le verrons plus loin, les réseaux peuvent être de tailles, de formes et de types différents. Ils sont généralement interconnectés pour constituer de plus grands réseaux, l’**Internet** étant l’exemple le plus célèbre d’un réseau de réseaux.

Il règne une grande confusion dans la littérature spécialisée entre les notions de réseau d’ordinateurs et de **système réparti** ou **distribué**. La distinction fondamentale est le niveau d’abstraction obtenu dans ce dernier. Un système réparti est un ensemble d’ordinateurs indépendants, présenté à l’utilisateur comme un système unique cohérent (avec généralement un seul modèle ou paradigme). Souvent, une couche logicielle intermédiaire appelée **middleware**, située au-dessus du système d’exploitation, est responsable de l’implémentation de ce modèle. Un exemple connu de système réparti est le **Web**. Il s’exécute au-dessus de l’Internet et présente un modèle dans lequel tout apparaît sous la forme d’un document (une page web).

Avec un réseau d’ordinateurs, les notions de cohérence, de modèle et de middleware, disparaissent. L’utilisateur se retrouve face à la réalité des machines et à leurs caractéristiques diverses. Le système ne tente pas de présenter ou de faire agir les machines de façon cohérente. Si le réseau se compose d’équipements et de systèmes d’exploitation variés, les différences sont pleinement visibles. Pour exécuter un programme sur une machine distante, l’utilisateur doit y ouvrir une session.

Un système réparti est donc un logiciel élaboré au-dessus d’un réseau pour apporter un haut degré de cohésion et de transparence. La différence entre un réseau et un système réparti se situe donc davantage au niveau du logiciel (surtout du système d’exploitation) que du matériel.

Ces deux systèmes possèdent néanmoins de profondes similitudes fonctionnelles. Par exemple, tous deux requièrent que des fichiers puissent être déplacés. La différence réside dans l’élément initiateur du déplacement, qui est le système dans un cas, l’utilisateur dans l’autre.

Bien que ce livre traite principalement des réseaux, de nombreux sujets abordés ont également leur importance dans le cadre des systèmes répartis. La bibliographie donnée en fin d’ouvrage permettra au lecteur de compléter son information sur ces différents sujets.

1.1 Usage des réseaux d'ordinateurs

Avant de nous attaquer aux aspects techniques propres aux réseaux, voyons l'intérêt que présentent ces derniers ainsi que leurs applications. Après tout, si personne n'était intéressé par leur emploi, on en construirait peu. Nous débuterons par les utilisations traditionnelles, telles que celles faites par les entreprises et les particuliers, avant de poursuivre avec les réseaux domestiques et les récents développements concernant les utilisateurs nomades, pour terminer par les aspects sociaux.

1.1.1 Applications professionnelles

La plupart des entreprises possèdent un nombre considérable d'ordinateurs. Elles peuvent par exemple en affecter un à chaque employé et les utiliser pour concevoir des produits, rédiger des documents et élaborer la paie. Si chacun d'eux à l'origine fonctionnait seul, indépendamment des autres, on a pu décider à un moment donné de les interconnecter pour pouvoir distribuer les informations dans toute l'entreprise.

Plus généralement, la question abordée ici est celle du **partage des ressources**. L'objectif est de rendre les programmes, les équipements et surtout les données accessibles à tout utilisateur du réseau, indépendamment de leur emplacement physique ou de celui de la ressource. Un exemple courant est le partage d'une imprimante par un groupe d'employés. Aucun d'eux n'a besoin d'une imprimante pour lui seul, et une imprimante en réseau capable d'imprimer des volumes importants est souvent moins coûteuse, plus rapide et plus facile à maintenir que de nombreuses imprimantes individuelles.

Toutefois, un autre type de partage est probablement encore plus important que celui des ressources physiques comme les imprimantes et les systèmes de sauvegarde sur bande : celui des informations. Quelle que soit leur taille, toutes les organisations dépendent aujourd'hui de façon vitale des informations que renferment leurs systèmes informatiques. La plupart ont en ligne des fichiers clients, des fiches produits, des états des stocks, des rapports financiers, des données fiscales et bien d'autres informations encore. En cas de panne informatique soudaine, une banque ne pourrait pas poursuivre son activité au-delà de cinq minutes, et une usine de production moderne s'appuyant sur une chaîne de montage assistée par ordinateur ne durerait même pas cinq secondes. Même la survie d'une petite agence de voyages ou d'un modeste cabinet juridique dépend désormais totalement du réseau d'ordinateurs qui doit permettre aux employés d'accéder instantanément aux informations et aux documents.

Au sein d'une petite société, tous les ordinateurs se trouvent souvent dans un même bureau ou un même immeuble. Dans les plus grandes, les ordinateurs et les employés peuvent être disséminés à travers des dizaines de bureaux ou de bâtiments situés dans des pays différents. Un commercial doit pouvoir consulter une base de données de produits située à l'autre bout du globe s'il le souhaite. Des **réseaux privés virtuels** ou **VPN** (*Virtual Private Networks*) peuvent servir à relier les réseaux individuels des différents sites en un seul réseau étendu. Autrement dit, le simple fait qu'un utilisateur se trouve à 15 000 km de ses données ne doit pas l'empêcher de pouvoir les

exploiter comme si elles étaient stockées localement. En atteignant cet objectif, on se libère pour ainsi dire de la « tyrannie géographique ».

Pour simplifier, on peut dire qu'un système d'information d'entreprise est un ensemble auquel participent, d'une part, une ou plusieurs bases de données et, d'autre part, un certain nombre d'employés devant accéder à distance à celles-ci. Dans ce modèle, les données sont stockées sur des ordinateurs puissants appelés **serveurs**, qui sont souvent groupés physiquement et gérés par un administrateur système. Les employés travaillent sur des ordinateurs plus simples, appelés **clients**, à l'aide desquels ils accèdent aux données pour, par exemple, les traiter dans un tableur. *Nous emploierons parfois le terme « client » pour désigner l'utilisateur de la machine, mais le contexte devrait vous permettre de déterminer chaque fois à quoi il renvoie.* Les machines client et serveur sont interconnectées par l'intermédiaire d'un réseau, comme dans l'exemple de la figure 1.1. Notez que le réseau est représenté par un simple ovale, sans autre détail. Nous utiliserons cette forme d'illustration pour désigner un réseau au sens le plus abstrait. Lorsque davantage de détails seront nécessaires, ils seront inclus.

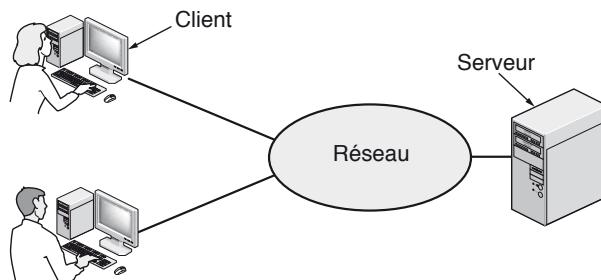


Figure 1.1 • Un réseau avec deux clients et un serveur.

Ce type d'organisation porte le nom de **modèle client-serveur**. Il est largement répandu et sert de base à de nombreuses applications. La réalisation la mieux connue est celle de l'**application web**, dans laquelle le serveur génère des pages web à partir de sa base de données, en réponse aux requêtes des clients qui peuvent mettre à jour ladite base. Ce modèle est applicable lorsque le client et le serveur se situent dans le même immeuble (et appartiennent à une même société), mais également lorsqu'ils sont éloignés géographiquement, comme dans le cas d'une personne accédant à une page sur le Web. Dans ce cas, le serveur est le serveur web distant et le client est l'ordinateur personnel de l'utilisateur. La plupart du temps, le serveur peut prendre en charge simultanément des centaines ou des milliers de clients.

Si nous examinons le modèle client-serveur en détail, nous constatons que deux processus (deux programmes qui s'exécutent) sont impliqués : l'un sur la machine client et l'autre sur la machine serveur. La communication prend alors la forme d'un processus client qui envoie un message au processus serveur, puis attend un message en réponse. Quand le processus serveur reçoit la requête du client, il exécute la tâche ou

recherche les données demandées, puis il envoie une réponse. La figure 1.2 illustre cet échange de messages.

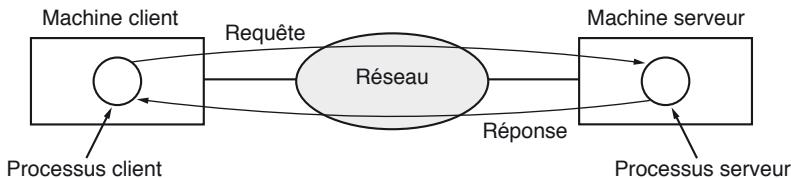


Figure 1.2 • Le modèle client-serveur met en œuvre l'échange de requêtes et de réponses.

Un deuxième objectif de l'implémentation d'un réseau d'ordinateurs concerne les personnes, et non plus les informations ni même les ordinateurs. En effet, un tel système peut représenter un formidable **moyen de communication** entre employés. Pratiquement n'importe quelle société disposant d'au moins deux ordinateurs possède aujourd'hui un système de messagerie ou de **courrier électronique** que le personnel emploie généralement pour une grande partie de ses échanges quotidiens. Preuve de l'engouement pour cet outil : les plaintes qui reviennent souvent concernant la quantité de messages que tout un chacun doit traiter.

Les appels téléphoniques entre employés peuvent transiter par le réseau informatique au lieu de celui de l'opérateur téléphonique. En cas d'emploi d'une technologie internet, on parle de **téléphonie IP** ou de **voix sur IP** (*VoIP, Voice over IP*). À chaque extrémité, le microphone et l'écouteur peuvent appartenir à un combiné équipé pour VoIP ou à l'ordinateur des employés. Les entreprises y voient un moyen extraordinaire de réaliser des économies sur les notes de téléphone.

Les réseaux rendent également possibles des formes de communication enrichies, comme la vidéoconférence. On peut associer la vidéo au son, afin que des employés puissent se voir et s'entendre en tenant des réunions à distance. Cette technique est un outil puissant qui permet de gagner du temps et d'éliminer les coûts autrefois consacrés aux déplacements. Le **bureau partagé** permet de voir un écran graphique interactif. Plusieurs personnes éloignées physiquement peuvent ainsi facilement lire et écrire sur un tableau virtuel partagé ou rédiger un rapport ensemble. Lorsque l'une d'elles modifie un document en ligne, les autres peuvent en prendre connaissance immédiatement, au lieu de devoir attendre plusieurs jours la réception d'une lettre. Un tel gain de temps facilite la coopération là où elle était impossible auparavant. On commence maintenant à utiliser des formes de coordination plus ambitieuses, comme la télémédecine (par exemple pour la surveillance de patients à distance), dont l'emploi pourrait se généraliser. On dit parfois que la communication et le transport sont deux rivaux engagés dans une course, et que la victoire de l'un signifiera l'obsolescence de l'autre.

Un troisième objectif de plus en plus d'entreprises est l'exercice de leur activité économique en ligne, en particulier avec leurs fournisseurs et leurs clients. Ce nouveau modèle, qualifié de **commerce électronique**, ou **e-commerce**, s'est rapidement

développé ces dernières années. Les compagnies aériennes, les librairies, etc., ont constaté que de nombreux clients apprécient de pouvoir faire leurs achats sans avoir à se déplacer. De nombreuses sociétés fournissent donc des catalogues de produits ou de services et prennent des commandes en ligne. Des fabricants d'automobiles, d'avions ou d'ordinateurs, parmi d'autres, achètent des sous-systèmes auprès de différents fournisseurs et les assemblent ensuite. Les réseaux d'ordinateurs leur permettent de passer leurs commandes électroniquement au fur et à mesure de leurs besoins, ce qui évite de devoir maintenir des stocks volumineux et améliore l'efficacité.

1.1.2 Applications domestiques

En 1977, Ken Olsen était le président de Digital Equipment Corporation, le deuxième constructeur d'ordinateurs au monde à l'époque (après IBM). Lorsqu'on lui a demandé pourquoi Digital n'investissait pas franchement sur le marché de l'ordinateur individuel, il a répondu qu'il ne voyait aucun intérêt pour le particulier à disposer d'un ordinateur chez lui. L'histoire lui a donné tort, et Digital n'existe plus. À l'origine, on achetait un ordinateur pour le traitement de texte et les jeux. Plus récemment, la principale raison d'acquérir un ordinateur domestique a sans doute été l'accès à l'Internet. De nos jours, de nombreux produits électroniques destinés au grand public, comme les décodeurs TV, les consoles de jeu et les radioréveils, contiennent des ordinateurs embarqués, et les réseaux informatiques, surtout les réseaux sans fil et les réseaux domestiques, sont largement utilisés pour le divertissement, notamment pour la création et la diffusion de musique, de photos et de vidéos.

L'accès à l'Internet fournit aux utilisateurs domestiques la **connectivité** à des ordinateurs distants. Comme les utilisateurs professionnels, ils peuvent consulter des informations, communiquer avec d'autres personnes et acheter des biens et des services en ligne. Le principal avantage provient maintenant de la possibilité de se connecter sans sortir de chez soi. Bob Metcalfe, l'inventeur d'Ethernet, estimait que la valeur d'un réseau est proportionnelle au carré du nombre de ses utilisateurs, parce que c'est approximativement le nombre de connexions différentes possibles. Connue sous le nom de « loi de Metcalfe », cette hypothèse aide à expliquer en quoi la fantastique réussite de l'Internet est due à sa taille.

L'accès aux informations à distance prend de nombreuses formes, que l'internaute recherche des données spécifiques ou qu'il surfe sur le Web pour le plaisir. Les informations disponibles touchent tous les sujets : arts, affaires, cuisine, gouvernement, santé, histoire, loisirs, détente, science, sports, voyages, et bien d'autres encore. Les sources de divertissement sont trop nombreuses pour qu'on les cite toutes, sans compter celles qu'il est préférable de ne pas évoquer.

De nombreux journaux sont aujourd'hui accessibles en ligne et personnalisables. Vous pouvez par exemple indiquer que vous souhaitez tout savoir sur les politiciens corrompus, les grands incendies, les scandales impliquant des célébrités et les épidémies, mais que le football ne vous intéresse pas. Il est parfois possible de recevoir des articles sélectionnés, qui sont téléchargés sur votre ordinateur pendant que vous dormez. Si cette tendance se poursuit, elle fera de nombreux chômeurs parmi les livreurs

et les vendeurs de journaux, mais la presse l'apprécie, parce que la distribution a toujours constitué le maillon le plus faible de la chaîne de production. Bien entendu, pour que ce modèle fonctionne, il faudra d'abord déterminer comment dégager du profit dans ce nouveau contexte, ce qui n'est pas forcément facile quand les utilisateurs de l'Internet s'attendent à la gratuité totale.

L'étape suivante après les journaux (ainsi que les magazines et les revues scientifiques) est la bibliothèque numérique en ligne. Les publications et les actes des conférences de nombreuses organisations professionnelles, comme l'**ACM** (*Association for Computing Machinery*, www.acm.org) et l'**IEEE Computer Society** (www.computer.org), sont déjà accessibles sur le Web. Les lecteurs de livres électroniques et les bibliothèques en ligne pourraient bien finir par rendre le livre imprimé obsolète. Si vous en doutez, pensez aux effets que l'imprimerie a eus sur les manuscrits enluminés du Moyen Âge.

Le modèle client-serveur permet d'accéder à nombre de ces informations, mais il en existe un autre, appelé **poste à poste** ou **pair à pair** (*peer-to-peer*) souvent abrégé en P2P. Dans ce modèle, les utilisateurs forment un groupe informel au sein duquel chacun peut communiquer avec tous les autres, comme le montre la figure 1.3. L'échange se fait en principe sur un pied d'égalité, et il n'existe pas de division en clients et en serveurs.

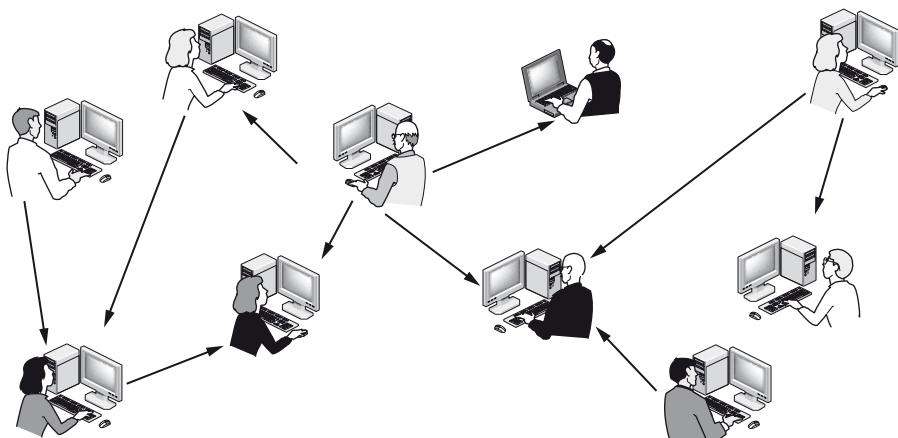


Figure 1.3 • Dans un système pair à pair, il n'existe pas de clients ni de serveurs fixes.

Dans la plupart des systèmes pair à pair, comme BitTorrent, la base de données centrale disparaît. À la place, chaque utilisateur maintient sa propre base de données localement et fournit une liste des autres personnes du voisinage qui sont membres du système. Un nouvel utilisateur peut ainsi rendre visite à un membre existant pour voir ce qu'il propose et obtenir les noms d'autres membres pour rechercher d'autres contenus et obtenir d'autres noms. Ce processus peut se répéter indéfiniment et permettre d'établir une volumineuse base de données locale. Cette tâche serait fastidieuse pour des humains, mais les ordinateurs y excellent.

La communication pair à pair sert souvent à partager de la musique et des vidéos. Elle a connu son apogée aux alentours de l'an 2000, avec un service de partage de musique nommé Napster, qui fut fermé après ce qui fut sans doute le cas le plus important de non-respect des droits d'auteur de toute l'Histoire. Il existe également des applications parfaitement légales de ce modèle. Les fans qui partagent de la musique appartenant au domaine public, les familles qui échangent des photos ou les utilisateurs qui téléchargent des logiciels gratuits sont autant d'exemples d'activités autorisées. En fait, l'une des applications les plus populaires de l'Internet, le courrier électronique, est intrinsèquement peer-to-peer. Cette forme de communication est appelée à se développer considérablement dans le futur.

Toutes ces applications impliquent des interactions à distance entre une personne et une base de données renfermant quantité d'informations. Mais l'Internet est aussi principalement employé à des fins de communication interpersonnelle et représente en quelque sorte au xxie siècle ce que le téléphone représentait au xixe. Des millions d'individus dans le monde emploient quotidiennement le courrier électronique, et son utilisation augmente rapidement. Il sert à transporter aussi bien du texte et des images que des données audio et vidéo. Quant aux odeurs, il faudra patienter encore quelque temps.

La plupart des adolescents ne peuvent plus se passer aujourd'hui de la **messagerie instantanée**. Ce service, qui s'inspire du programme UNIX *talk* utilisé dans les années 1970, permet à deux personnes d'échanger des messages en temps réel. Il existe également des services de messagerie collective, comme **Twitter**, qui permet d'envoyer de courts messages texte appelés « tweets » (gazouillis) à son cercle d'amis ou à qui souhaite les recevoir.

Des applications peuvent également utiliser l'Internet pour transporter des données audio (comme les stations de radio en ligne) et vidéo (comme YouTube). Outre qu'elles peuvent représenter un moyen bon marché d'appeler des amis éloignés, ces applications peuvent procurer des expériences intéressantes, comme le télenseignement, qui présente l'avantage de pouvoir suivre un cours à 8 heures du matin sans devoir quitter son lit. À long terme, l'utilisation des réseaux pour améliorer la communication interhumaine sera peut-être plus importante que toutes les autres. Elle peut même devenir capitale pour les personnes géographiquement isolées, en leur offrant les mêmes services qu'à celles qui vivent au cœur d'une grande ville.

Entre les communications interpersonnelles et l'accès à l'information, on trouve les applications de réseautage social, ou **réseaux sociaux**. Dans ces derniers, le flux d'information est régi par les relations que les personnes déclarent exister entre elles. L'un des sites les plus connus en la matière est **Facebook**, Il permet à ses membres d'actualiser leur profil personnel et de partager les mises à jour avec ceux qu'ils affirment être leurs amis. Dans d'autres applications de réseautage social, il est possible d'être introduit par des amis d'amis, d'envoyer des messages à des amis, comme sur Twitter mentionné précédemment, et bien plus encore.

De façon encore plus informelle, des groupes d'individus peuvent travailler ensemble pour créer des contenus. Un **wiki**, par exemple, est un site web collaboratif qu'une communauté crée et modifie. Le plus célèbre d'entre eux est **Wikipédia**,

une encyclopédie à laquelle chacun peut contribuer, mais il en existe des centaines d'autres.

Notre troisième catégorie d'applications concerne le commerce électronique au sens le plus large du terme. Nombre d'utilisateurs apprécient de faire leurs achats depuis leur domicile en consultant les catalogues en ligne de milliers d'entreprises. Certains d'entre eux sont interactifs et montrent les produits sous différents angles et dans des configurations personnalisables. Et si le client rencontre des difficultés lors de l'emploi d'un produit après son achat électronique, il peut faire appel à un service d'assistance en ligne.

Le commerce électronique a aussi largement pénétré le secteur des institutions financières. De nombreux clients paient leurs factures, gèrent leurs comptes bancaires et réalisent des investissements par voie électronique. Cette tendance se confirmera certainement, à mesure que les réseaux seront davantage sécurisés.

Un domaine que personne n'avait prévu est celui de la braderie en ligne. Les ventes aux enchères de marchandises d'occasion représentent une véritable industrie. À la différence du commerce électronique traditionnel qui s'appuie sur le modèle client-serveur, celle-ci est du type peer-to-peer, au sens où les consommateurs peuvent être aussi bien acheteurs que vendeurs.

Certaines de ces formes de commerce en ligne ont reçu une appellation tirée de l'anglais (où le mot *to* est remplacé par 2, dont la prononciation est identique). La figure 1.4 présente les plus connues.

Figure 1.4 • Quelques formes de commerce en ligne.

Appellation	Modèle métier	Exemple
B2C (<i>Business-to-Consumer</i>)	Entreprise à consommateur	Achat de livres en ligne
B2B (<i>Business-to-Business</i>)	Entreprise à entreprise	Fabricant d'automobiles commandant des pneus à un fournisseur
G2C (<i>Government-to-Consumer</i>)	Gouvernement à consommateur	Distribution de formulaires fiscaux
C2C (<i>Consumer-to-Consumer</i>)	Consommateur à consommateur	Vente aux enchères de produits d'occasion
P2P (<i>Peer-to-Peer</i>)	Égalitaire	Partage de musique

Une quatrième catégorie d'applications concerne le divertissement. Elle a fait une énorme percée ces dernières années, avec la distribution en ligne de musique, de programmes de radio et de télévision et de films, qui commence à concurrencer les mécanismes traditionnels. Les usagers peuvent trouver, acheter et télécharger des morceaux au format MP3 et des films de qualité DVD, et les ajouter à leur collection personnelle. Les émissions télévisées atteignent de nombreux foyers *via* des systèmes IPTV (*IP Television*), qui s'appuient sur une technologie IP au lieu du

câble ou de la transmission hertzienne. Des applications de *streaming* (diffusion en flux) permettent aux utilisateurs d'écouter des radios internet (ou Web radios) ou de voir de récents épisodes de leurs séries favorites. Naturellement, tous ces contenus peuvent être distribués dans toute la maison, entre différents équipements audio et vidéo, généralement grâce à un réseau sans fil.

Bientôt, il sera peut-être possible de rechercher n'importe quel film ou programme télévisé jamais produit dans n'importe quel pays et de le voir diffusé instantanément sur son écran. Les nouveaux films pourront devenir interactifs, l'utilisateur étant invité de temps à autre à choisir le cours de l'histoire parmi différents scénarios possibles (par exemple, Macbeth doit-il assassiner Duncan ou attendre son heure ?). Les jeux télévisés en direct pourraient également devenir interactifs et proposer au public de participer, de déporter les concurrents, et ainsi de suite.

Le jeu est une autre forme de divertissement. Il existe déjà des jeux de simulation multijoueurs, comme les jeux de rôle dans des donjons virtuels, ou les simulateurs de vol où les joueurs d'une équipe tentent d'abattre les avions de l'équipe adverse. Les mondes virtuels constituent un environnement persistant, dans lequel des milliers d'utilisateurs peuvent vivre une réalité partagée grâce à des images tridimensionnelles.

Notre dernière catégorie est l'**informatique ubiquitaire**, dans laquelle les ordinateurs sont omniprésents dans la vie quotidienne. Nombre de foyers sont déjà équipés de systèmes de sécurité, avec des capteurs placés sur les portes et les fenêtres. Bien d'autres capteurs pourraient être embarqués dans un petit système de surveillance domotique, pour contrôler par exemple la consommation d'énergie. Les compteurs d'électricité, de gaz et d'eau pourraient aussi transmettre directement les données sur le réseau, ce qui économiserait de l'argent en évitant les relevés. De même, les détecteurs de fumée pourraient appeler les pompiers au lieu d'émettre un bruit strident (qui ne sert pas à grand-chose quand il n'y a personne à la maison). Comme les prix des **capteurs** et des communications chutent, un nombre accru de mesures et de rapports passera par les réseaux.

Les équipements électroniques grand public sont de plus en plus connectés à des réseaux. Par exemple, certains appareils photo haut de gamme disposent déjà de fonctionnalités sans fil et les utilisent pour envoyer des images et les afficher sur un écran proche. Les photographes sportifs professionnels peuvent également envoyer à leur agence leurs photos en temps réel, d'abord en sans-fil vers un point d'accès, puis sur l'Internet. Des appareils comme les téléviseurs, qui se branchent sur le secteur, peuvent passer par le réseau électrique pour envoyer des informations dans toute la maison par courant porteur en ligne. S'il n'est pas très surprenant de voir ce type d'objets sur le réseau, il en existe d'autres qui peuvent également capter et communiquer des données. Par exemple, votre douche peut enregistrer votre consommation d'eau, vous donner un feedback visuel pendant que vous vous savonnez et envoyer un rapport à une application domotique écologique quand vous avez fini pour vous permettre d'économiser sur votre facture d'eau.

Une technologie du nom de **radio-identification**, ou **RFID** (*Radio Frequency Identification*), ira encore plus loin dans le futur. Les étiquettes RFID, ou

radio-étiquettes, sont des puces passives (autrement dit sans batterie) de la taille d'un timbre-poste, que l'on peut déjà fixer sur un livre, un passeport, un collier de chien, une carte de crédit et toutes sortes d'autres articles à la maison ou à l'extérieur. Cela permet à des lecteurs RFID de les localiser et de communiquer avec eux sur une distance pouvant atteindre plusieurs mètres, selon le type de RFID. Commercialisée à l'origine pour remplacer les codes-barres, cette technique a rencontré peu de succès, parce que les codes-barres sont gratuits alors que les radio-étiquettes coûtent quelques centimes. Naturellement, elles offrent beaucoup plus de possibilités, et leur prix diminue rapidement. Elles pourraient transformer le monde réel en un Internet des objets.

1.1.3 Utilisateurs nomades

Les équipements mobiles comme les ordinateurs ultraportables, les assistants personnels ou PDA (*Personal Digital Assistant*) et les téléphones intelligents (*smartphones*) représentent l'un des segments à plus forte croissance de l'industrie informatique. Leurs ventes ont déjà dépassé celles des ordinateurs de bureau. Quel est leur intérêt ? En déplacement, les utilisateurs veulent pouvoir lire et envoyer des courriers électroniques et des messages instantanés, regarder des films, télécharger de la musique, jouer ou tout simplement surfer sur le Web à la recherche d'informations, autrement dit faire tout ce qu'ils font chez eux et au bureau. Bien entendu, ils veulent le faire de n'importe où sur terre, en mer ou dans les airs.

La **connectivité** à l'Internet permet nombre de ces usages nomades. Puisqu'une connexion filaire est impossible à bord d'un véhicule, d'un bateau ou d'un avion, les réseaux sans fil suscitent un très vif intérêt. Les réseaux cellulaires offerts par les compagnies téléphoniques sont un exemple familier de ce type de réseau qui nous fournit une couverture pour les téléphones mobiles. Les **hotspots** sans fil, basés sur la norme 802.11 sont un autre type de réseau sans fil pour les ordinateurs portables. Ils ont survécu un peu partout et forment désormais un patchwork qui couvre les cafés, les hôtels, les aéroports, les écoles, les trains et les avions. Toute personne équipée d'un ordinateur portable et d'un modem sans fil peut se connecter à l'Internet *via* le hotspot comme s'il était connecté à un réseau filaire.

Les réseaux sans fil sont aussi très précieux pour les flottes de camions ou de taxis, et pour les livreurs ou les réparateurs qui gardent ainsi le contact avec leur entreprise. Dans de nombreuses villes, les chauffeurs de taxi sont des travailleurs indépendants et non les employés d'une entreprise. Dans certaines d'entre elles, les voitures sont équipées d'un dispositif comprenant un écran sur lequel s'affichent les lieux de prise en charge et de destination lorsqu'une nouvelle demande est transmise par un centre d'appels. Un signal sonore avertit le chauffeur de l'arrivée de la demande et le premier qui appuie sur un bouton de l'écran obtient la course.

Les réseaux sans fil jouent aussi un rôle important pour l'armée. Pour pouvoir partir en « mission » n'importe où dans un bref délai, il serait déraisonnable de compter sur l'infrastructure de réseau locale, et il est préférable d'employer la sienne.

Bien que le réseau sans fil et l'informatique mobile soient souvent associés, il s'agit de deux choses différentes, comme le montre le tableau de la figure 1.5. Nous voyons qu'il existe une différence entre les **accès sans fil fixes** et les **accès sans fil mobiles**. Même des ordinateurs portables sont parfois interconnectés au moyen d'un câble. Par exemple, un voyageur reliant son notebook à la prise du réseau filaire de sa chambre d'hôtel bénéficie de la mobilité sans recourir à un réseau sans fil.

Figure 1.5 • Relations entre réseaux sans fil et informatique mobile.

Sans fil	Mobile	Applications
Non	Non	Ordinateur personnel de bureau
Non	Oui	Ordinateur portable utilisé dans une chambre d'hôtel
Oui	Non	Réseau dans des immeubles anciens sans câblage
Oui	Oui	Ordinateur de poche pour l'inventaire de magasin

Inversement, certains ordinateurs sans fil ne sont pas mobiles. À la maison, et dans les bureaux ou les hôtels qui ne sont pas câblés de façon appropriée, il peut être plus pratique de connecter des ordinateurs de bureau ou des lecteurs multimédias par une liaison sans fil que de tirer des câbles. L'installation d'un réseau sans fil ne demande guère plus que d'acquérir un petit boîtier avec l'électronique associée, de déballer le tout et de le mettre en place. Cette solution est de loin plus économique que le recours à des ouvriers pour passer des câbles.

Enfin, il existe également des applications sans fil réellement mobiles, comme celles qu'utilisent les employés pour faire l'inventaire grâce à un ordinateur de poche dans un magasin. Dans les grands aéroports, les sociétés de location de véhicules équipent les employés affectés à la réception des véhicules d'ordinateurs portables communiquant sans fil. Sur le parking, ils scannent les codes-barres ou les puces RFID des véhicules rendus, et leur équipement mobile, qui dispose d'une imprimante intégrée, appelle l'ordinateur central, récupère les données du contrat et imprime la facture sur-le-champ.

Mais peut-être le principal facteur de développement des applications sans fil nomade est-il le téléphone mobile. La **messagerie textuelle** est extrêmement répandue. Elle permet à un utilisateur de taper un court message, qui est ensuite remis à un autre abonné par l'intermédiaire du réseau cellulaire. Qui aurait prédit il y a dix ans que les adolescents adoreraient taper péniblement des SMS et que cela rapporterait des fortunes aux opérateurs téléphoniques ? Car le SMS (*Short Message Service*) est très lucratif, puisqu'il ne leur coûte qu'une minuscule fraction de centime alors qu'ils le facturent beaucoup plus cher.

La convergence si attendue de la téléphonie et de l'Internet est enfin arrivée, et elle va accélérer la croissance des applications mobiles. Les **téléphones intelligents**, ou **smartphones**, comme l'iPhone, combinent des aspects des ordinateurs et des téléphones mobiles. Les réseaux cellulaires (3G et 4G) auxquels ils se connectent peuvent fournir des services de données rapides passant par l'Internet en plus des

communications téléphoniques. De nombreux téléphones évolués se connectent également aux hotspots et passent automatiquement d'un réseau à l'autre afin de choisir la meilleure option pour l'utilisateur.

D'autres équipements électroniques grand public peuvent aussi utiliser les réseaux cellulaires et les hotspots pour rester connectés à des ordinateurs distants. Les lecteurs peuvent télécharger sur une liseuse le dernier livre qu'ils ont acheté, la nouvelle édition d'un magazine ou le journal du jour partout où ils vont, et les panneaux numériques peuvent actualiser leur affichage à tout moment.

Comme les téléphones mobiles connaissent leur position, souvent parce qu'ils sont équipés de récepteurs **GPS** (*Global Positioning System*), certains services sont intentionnellement dépendants de la position géographique. Les cartes et les itinéraires sont des exemples évidents, et le GPS de votre téléphone ou de votre voiture a probablement une meilleure idée de l'endroit où vous êtes que vous. Il en va de même si vous recherchez une librairie ou un restaurant chinois à proximité, ou encore si vous voulez la météo locale. D'autres services peuvent enregistrer la position, par exemple pour annoter des photos et des vidéos avec le lieu où elles ont été prises. Ces annotations sont connues sous le nom de balises de géolocalisation (*geotags*).

Un domaine dans lequel on commence à utiliser le téléphone mobile est le **commerce mobile** (**m-commerce**). Au lieu d'insérer de la monnaie ou une carte bancaire, l'usager envoie un SMS au distributeur automatique pour payer un café, un ticket de cinéma ou un autre petit article. Le paiement est ensuite imputé sur la facture du téléphone mobile. S'il est équipé de la technologie de **communication en champ proche** ou **NFC** (*Near Field Communication*), le mobile peut faire office de carte RFID et communiquer avec un lecteur proche pour régler un achat. Les forces motrices derrière ce phénomène sont les constructeurs d'équipements nomades et les opérateurs de télécommunications, qui déploient des efforts considérables pour obtenir une part du gâteau. Du point de vue du commerçant, ce système lui évite la majeure partie des frais liés à l'emploi des cartes de crédit, qui peuvent représenter un pourcentage important. Bien entendu, la médaille a un revers : les clients d'un magasin peuvent utiliser leur lecteur de codes-barres ou de radio-étiquettes pour connaître les tarifs des concurrents avant d'acheter ou obtenir une liste détaillée des autres points de vente proposant le même produit avec son prix.

Un gros avantage dont peut profiter le commerce mobile est que les utilisateurs de portables sont habitués à payer pour tout (à la différence des utilisateurs de l'Internet qui s'attendent au « tout-gratuit »). Un site web soulèverait un tollé général s'il facturait un coût supplémentaire pour autoriser le client à payer par carte de crédit, alors que l'application d'un coût supplémentaire par un opérateur pour permettre le paiement par téléphone portable dans un magasin sera probablement considérée comme naturelle. L'avenir nous le dira.

Il ne fait aucun doute que les usages des ordinateurs mobiles et sans fil se développeront rapidement à mesure que les ordinateurs se miniaturiseront, et probablement en des façons que personne ne peut encore prévoir. Voyons rapidement quelques possibilités. Les **réseaux de capteurs** sans fil sont constitués de nœuds qui recueillent et transmettent des informations sur l'état du monde physique. Ces nœuds peuvent

faire partie d'objets familiers, comme un téléphone ou une voiture, ou bien être de petits dispositifs autonomes. Par exemple, votre voiture pourrait collecter des données sur sa position, sa vitesse, les vibrations qu'elle émet et son efficience énergétique, puis les envoyer à une base de données. Celle-ci pourrait aider à détecter des nids-de-poule, trouver des itinéraires pour contourner les embouteillages et vous dire si vous êtes un « goinfre en essence » par rapport aux autres conducteurs empruntant la même portion de route.

Les réseaux de capteurs sont en train de révolutionner la science en fournissant une profusion de données sur des comportements qu'il était auparavant impossible d'observer. On a pu par exemple suivre des migrations de zèbres en plaçant un petit capteur sur chaque animal. Des chercheurs ont réussi à embarquer un ordinateur sans fil dans un cube de 1 mm d'arête. Il est ainsi possible de suivre à la trace des oiseaux, des rongeurs et des insectes, si petits soient-ils.

Même des utilisations banales peuvent être significatives, parce qu'elles emploient des données auparavant indisponibles. Par exemple, les parcmètres peuvent accepter des paiements par carte de débit ou de crédit sur une liaison sans fil, et indiquer s'ils sont utilisés ou non. Cela permettrait aux conducteurs de télécharger un plan de quartier actualisé en temps réel pour trouver une place plus facilement. Bien entendu, une fois le temps de stationnement écoulé, le parcmètre pourrait également vérifier la présence d'une voiture (en captant le signal qui en provient) et avertir les autorités de l'expiration du délai.

Les **ordinateurs « prêts à porter »** (*wearable computers*) représentent une autre application prometteuse. Si les montres intelligentes équipées de radios font partie de notre imaginaire depuis leur apparition dès les années 1940 dans les films et les bandes dessinées, nous pouvons maintenant les acheter. D'autres dispositifs, comme les pacemakers ou les pompes à insuline, sont implantables, et certains peuvent être contrôlés par le biais d'un réseau sans fil, ce qui permet aux médecins de les tester et de les reconfigurer plus aisément. En revanche, s'ils sont aussi peu sûrs et aussi faciles à pirater que le PC moyen, les risques encourus ne sont pas négligeables.

1.1.4 Aspects sociaux

À l'instar de l'imprimerie il y a plus de cinq siècles, les réseaux d'ordinateurs permettent aux citoyens ordinaires de diffuser et de voir des contenus de façons imaginables auparavant. Mais toute médaille a son revers, et cette nouvelle liberté s'accompagne de nombreux problèmes sociaux, éthiques et politiques non résolus. Nous n'en aborderons que quelques-uns, car une étude approfondie nécessiterait au moins un livre entier.

Les réseaux sociaux, les forums, les sites de partage de contenus et une foule d'autres applications permettent aux personnes de partager leurs vues avec des individus de même sensibilité. Tant que les sujets abordés restent techniques, ou se limitent à des violons d'Ingres comme le jardinage, peu de problèmes se présentent.

La difficulté surgit lorsqu'ils touchent à des domaines plus sensibles, comme la religion, la politique ou la sexualité, et les opinions exprimées publiquement peuvent alors être choquantes ou, pire encore, insultantes. De plus, elles ne se limitent pas nécessairement au texte : les photographies en couleurs haute résolution et les clips vidéo sont faciles à partager sur les réseaux. Certains prônent la tolérance, d'autres pensent au contraire que certains messages sont purement et simplement inacceptables et doivent être censurés, par exemple les attaques verbales contre des pays ou des religions, les contenus pornographiques, etc. Les lois diffèrent ou s'opposent selon les pays, et le débat fait rage.

Par le passé, des personnes ont poursuivi en justice des opérateurs, les tenant pour responsables du contenu qui circule sur leurs réseaux, comme c'est le cas des magazines ou des journaux. La réponse, inévitable, est que leur rôle s'apparente plutôt à celui d'une compagnie téléphonique ou du service postal, et que l'on ne peut exiger d'eux de contrôler ce que les utilisateurs disent ou envoient.

Maintenant, ne soyons pas surpris d'apprendre que certains opérateurs bloquent des contenus pour des raisons qui leur sont propres. Certains utilisateurs de réseaux P2P se sont vu refuser les services réseau, parce que les fournisseurs ne jugeaient pas rentable de transporter les gros volumes de trafic envoyés par ces applications. Ces mêmes opérateurs aimeraient probablement réservier des traitements distincts à différentes entreprises, selon qu'elles sont puissantes ou misérables. Les opposants à cette pratique avancent que le P2P doit être traité de la même manière que les autres contenus, parce qu'il ne constitue qu'une infime partie du trafic. Cet argument voulant que les communications ne soient pas différencieres en fonction de leur source ou de leur contenu est connu sous le nom de **principe de neutralité du réseau**. Sans aucun doute, ce débat risque de se poursuivre encore longtemps.

Nombre d'autres parties sont engagées dans la bagarre sur les contenus. Par exemple, la musique et les films piratés ont alimenté la croissance massive des réseaux P2P, ce qui n'a pas plu aux détenteurs de copyright qui ont menacé de porter plainte, voire parfois intenté des procès. Il existe maintenant des systèmes automatiques qui surveillent ces réseaux et envoient des avertissements aux opérateurs et aux utilisateurs suspects d'infraction au droit d'auteur. Aux États-Unis, ces avertissements sont connus sous le nom de **notices DMCA**, nommées ainsi d'après le *Digital Millennium Copyright Act*. En France, c'est l'**Hadopi** (Haute autorité pour la diffusion des œuvres et la protection des droits sur Internet) qui veille. Cette surveillance s'apparente à une course aux armements, parce qu'il est très difficile de détecter de manière fiable les infractions à la loi.

Les réseaux d'ordinateurs facilitent beaucoup la communication. Ils permettent également aux opérateurs d'espionner aisément le trafic. Cela entraîne des conflits sur d'autres sujets, comme celui des droits respectifs des employés et des employeurs. Beaucoup de personnes rédigent et lisent des messages électroniques sur leur lieu de travail. Nombre d'employeurs revendiquent le droit de lire et au besoin de censurer ces messages, y compris ceux expédiés depuis le domicile en dehors des heures de travail. Tous les employés ne sont pas d'accord, surtout avec la dernière proposition.

Un autre conflit tourne autour des droits du gouvernement et de ceux des citoyens. Le FBI a mis en place chez plusieurs fournisseurs d'accès à Internet un système qui lui permettait d'espionner les messages entrants et sortants afin de collecter des éléments intéressants. L'un des premiers systèmes s'appelait Carnivore, mais fut rebaptisé d'un nom plus discret, DCS1000, en raison de la mauvaise publicité dont il a fait l'objet. Ces dispositifs ont pour but de surveiller des millions de personnes afin de détecter les activités illégales. Malheureusement pour les espions, le quatrième amendement de la Constitution des États-Unis interdit les investigations sans mandat, mais le gouvernement l'ignore souvent.

Bien entendu, les gouvernements ne sont pas les seuls à menacer la vie privée des particuliers. Le secteur privé s'en mêle aussi en **profilant** les utilisateurs. Par exemple, les navigateurs web enregistrent sur l'ordinateur des visiteurs des petits fichiers appelés **cookies**, qui permettent aux entreprises de suivre leur activité dans le cyberspace, mais qui peuvent aussi occasionner la fuite d'informations confidentielles, comme des numéros de carte de crédit ou de Sécurité sociale. Les sociétés qui fournissent des services fondés sur le Web peuvent conserver d'importants volumes de données personnelles sur les utilisateurs, données qui leur permettent d'étudier leurs activités directement. Par exemple, si vous utilisez **Gmail**, Google peut lire vos courriels et afficher sur votre écran des publicités correspondant à vos intérêts.

Un nouveau problème lié aux équipements mobiles est celui de la géoconfidentialité. Dans le cadre de leur rôle de fournisseurs de services, les opérateurs peuvent apprendre où vous vous trouvez aux différents moments de la journée, suivre vos déplacements à la trace et savoir quelle boîte de nuit ou quel cabinet médical vous fréquentez.

Les réseaux informatiques offrent également la possibilité d'augmenter le degré de confidentialité grâce à l'envoi de messages anonymes. Cela peut être souhaitable dans certaines situations, par exemple, lorsque des étudiants, des soldats, des employés ou des citoyens veulent dénoncer le comportement répréhensible de professeurs, d'officiers, de supérieurs ou d'hommes politiques sans risquer des représailles. D'un autre côté, dans une société démocratique, la loi prévoit explicitement qu'un accusé puisse être confronté à son accusateur, et une accusation anonyme ne saurait donc constituer une preuve.

L'Internet permet de trouver rapidement des informations, mais nombre d'entre elles sont considérées comme trompeuses, voire carrément fausses. Le conseil médical que vous venez de trouver sur le Web concernant une douleur dans la poitrine peut provenir aussi bien d'un lauréat du prix Nobel que d'un lycéen en échec scolaire.

D'autres informations sont fréquemment indésirables. Le courrier électronique non sollicité (*spam*) est devenu aujourd'hui chose courante, parce que les spameurs ont collecté des millions d'adresses électroniques et que de prétendus commerciaux peuvent leur vendre à bas prix des messages générés par ordinateur. Le flot de spams résultant interfère avec celui des messages émanant de personnes réelles. Heureusement, les programmes de filtrage sont capables de lire et de rejeter les messages produits par des robots, bien qu'avec plus ou moins de succès.

D'autres contenus encore sont associés à des comportements délictueux. Les pages web et les messages dont le contenu est actif (essentiellement des programmes ou des macros qui s'exécutent sur la machine du destinataire) peuvent aussi renfermer des virus capables d'infester votre ordinateur. Ils peuvent alors servir à pirater le mot de passe de votre compte bancaire, ou faire en sorte que votre machine envoie des spams en se joignant à un **botnet**, un ensemble de machines compromises.

Les messages de filotage (*phishing*) font semblant de provenir d'un interlocuteur de confiance, par exemple votre banque, pour tenter de vous extorquer des informations sensibles comme des numéros de cartes bancaires. Les usurpations d'identité sont devenues un problème grave car leurs auteurs s'emparent de suffisamment d'informations pour obtenir des cartes de crédit ou d'autres documents au nom de leurs victimes.

Il peut être difficile d'empêcher des ordinateurs de se faire passer pour des personnes sur l'Internet. Ce problème a conduit au développement des **captchas**, qui demandent à un utilisateur d'exécuter une courte tâche de reconnaissance, par exemple de taper les caractères contenus dans une image déformée, pour prouver qu'ils ne sont pas des machines. Ce processus est une variante du célèbre test de Turing, dans lequel une personne pose des questions pour juger si l'entité qui répond est humaine ou non.

L'industrie informatique pourrait résoudre bon nombre de ces problèmes si elle prenait la sécurité au sérieux. Si tous les messages étaient chiffrés et authentifiés, certains méfaits seraient plus difficiles à commettre. De telles technologies sont éprouvées, et nous les étudierons en détail au chapitre 8. L'ennui est que les fournisseurs d'équipements et de logiciels savent que l'intégration de fonctions de sécurité est coûteuse et que leurs clients n'en demandent pas tant. De plus, un nombre considérable de problèmes trouvent leurs causes dans les nombreuses malfaçons ou imperfections que provoquent les ajouts continuels de nouvelles fonctionnalités, la multiplication des lignes de code entraînant celle des bogues. Une taxe sur les mises à jour pourrait aider à en réduire le nombre mais l'idée serait probablement difficile à vendre au niveau de certaines instances. Un remboursement des programmes défectueux serait envisageable, mais toute l'industrie du logiciel ferait faillite au bout d'un an.

Les réseaux informatiques soulèvent de nouveaux problèmes d'ordre juridique lorsqu'ils sont confrontés à des lois plus anciennes. Les paris en ligne en sont un exemple. Puisque les ordinateurs simulent toutes sortes de choses depuis des décennies, pourquoi ne pas simuler des machines à sous et des jeux de roulette ou de black-jack ? Parce que c'est une pratique illégale dans de nombreux pays. Le problème est qu'elle est légale dans d'autres (au Royaume-Uni par exemple), et que les propriétaires de casinos ont bien compris le potentiel financier des paris sur l'Internet. Mais que se passe-t-il si le parieur, le casino et le serveur se trouvent dans des pays différents, et que leurs lois ne s'accordent pas ? Bonne question.

1.2 Caractéristiques physiques des réseaux

Maintenant que nous avons abordé les applications et les aspects sociaux des réseaux, il est temps d'examiner les aspects techniques liés à leur conception. Il n'existe pas de taxonomie générale des réseaux, mais deux critères importants permettent de les caractériser : la technologie de transmission et la taille. Nous les examinerons l'un après l'autre, dans cet ordre.

D'un point de vue général, on distingue deux types de technologies de transmission largement répandues : la **diffusion** et le **point-à-point**.

Les liens point-à-point connectent des paires de machines individuelles. Pour aller de sa source à sa destination sur un réseau formé de tels liens, un court message, appelé **paquet** dans certains contextes, peut devoir transiter par une ou plusieurs machines intermédiaires. Comme plusieurs routes de longueur différente sont souvent possibles, il est important de pouvoir trouver les meilleures. La transmission point-à-point entre exactement un émetteur et un destinataire est appelée **envoi individuel** ou parfois **diffusion individuelle** (ou *unicast*).

En revanche, dans un réseau à diffusion (*broadcast*) un seul canal de transmission est partagé par tous les équipements : les paquets sont reçus par toutes les machines. Dans chaque paquet, un champ d'adresse permet d'identifier le destinataire réel. À réception d'un paquet, une machine lit ce champ et procède au traitement du paquet si elle reconnaît son adresse, ou l'ignore dans le cas contraire.

Un réseau sans fil est un exemple courant de réseau à diffusion, avec des communications partagées dans une zone de couverture qui dépend du canal sans fil et de la machine émettrice. Par analogie, imaginez qu'un responsable de service se tienne dans un couloir donnant sur de nombreux bureaux ouverts et appelle quelqu'un par son nom. Bien que toutes les personnes présentes l'entendent, une seule répondra à l'appel et les autres l'ignoreront.

Les systèmes à diffusion offrent généralement la possibilité d'adresser un paquet à *toutes* les destinations en utilisant une valeur spéciale dans le champ d'adresse. Dans ce cas, le paquet est non seulement reçu mais aussi traité par toutes les machines. Ce mode de transmission est appelé **diffusion générale** (ou *broadcast*). Certains systèmes permettent aussi d'adresser un paquet à un sous-ensemble des machines du réseau. On parle alors de **diffusion restreinte** (ou *multicast*).

L'autre critère de différenciation des réseaux est leur taille. La distance est une métrique de classification importante, car elle dicte l'emploi de technologies différentes.

La figure 1.6 présente plusieurs systèmes classés en fonction de leur taille approximative. On trouve en premier le **réseau personnel**, ou **PAN** (*Personal Area Network*), destiné à une seule personne. Viennent ensuite les réseaux opérant sur de plus longues distances, qui se répartissent en trois catégories : les réseaux locaux ou **LAN** (*Local Area Network*), les réseaux métropolitains ou **MAN** (*Metropolitan Area Network*) et les réseaux étendus ou **WAN** (*Wide Area Network*), leur taille augmentant à chaque fois. Enfin, l'interconnexion de plusieurs réseaux s'appelle un **interréseau**. L'Internet, qui fonctionne à l'échelle mondiale, est l'exemple le plus connu, (mais non le seul)

d'interréseau. Nous en connaîtrons bientôt de plus grands encore, avec l'**Internet interplanétaire**, pour connecter des réseaux à travers l'espace.

Distance entre processeurs	Emplacement des processeurs	Exemple
1 m	Un mètre carré	Réseau personnel
10 m	Une salle	
100 m	Un immeuble	Réseau local
1 km	Un campus	
10 km	Une ville	Réseau métropolitain
100 km	Un pays	
1 000 km	Un continent	Réseau longue distance
10 000 km	Une planète	

Figure 1.6 • Classification des réseaux d'après leur taille.

1.2.1 Réseaux personnels (PAN)

Les **réseaux personnels**, ou PAN (*Personal Area Networks*), permettent aux équipements de communiquer à l'échelle individuelle. Un exemple courant est celui du **réseau sans fil**, qui relie un ordinateur à ses périphériques. Pratiquement tous les ordinateurs s'accompagnent d'un moniteur, d'un clavier, d'une souris et d'une imprimante. En l'absence de liaisons sans fil, les connexions doivent être câblées. C'est pourquoi nombre de nouveaux utilisateurs ont du mal à trouver les bons câbles et à les brancher au bon endroit (même s'ils sont de différentes couleurs), et que la plupart des fournisseurs offrent, en option, l'installation par un technicien. Pour les aider, plusieurs sociétés se sont réunies pour mettre au point un réseau à courte portée nommé **Bluetooth** pour connecter ces équipements sans fil : s'ils sont équipés de Bluetooth, le câblage devient inutile. Il suffit de les poser sur son bureau et de les mettre sous tension pour que les connexions fonctionnent. Pour beaucoup, cette facilité est un plus important.

Sous leur forme la plus simple, les réseaux Bluetooth s'appuient sur le paradigme maître-esclave de la figure 1.7, où le PC est normalement le maître, et la souris, le clavier, etc., sont les esclaves. Le maître indique aux esclaves les adresses à utiliser, le moment où ils peuvent diffuser, pendant combien de temps ils peuvent émettre, les fréquences qu'ils peuvent employer, et ainsi de suite.

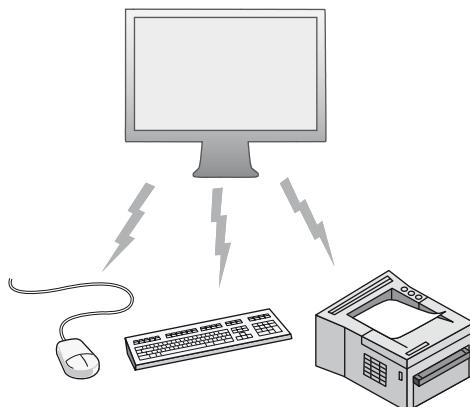


Figure 1.7 • Configuration d'un réseau personnel Bluetooth.

Bluetooth peut également servir dans d'autres contextes, notamment pour connecter une oreillette à un téléphone mobile, ou un baladeur numérique à votre voiture en vous plaçant simplement à portée. Un type de réseau personnel complètement différent est formé lorsqu'un appareil médical implanté, comme un stimulateur cardiaque, une pompe à insuline ou une aide auditive, communique à distance avec l'utilisateur. Nous verrons Bluetooth plus en détail au chapitre 4.

D'autres technologies de communication à courte portée, comme RFID pour les cartes sans contact et les livres de bibliothèque, permettent aussi de construire des PAN. Nous étudierons RFID au chapitre 4.

1.2.2 Réseaux locaux (LAN)

L'étape suivante est le **réseau local**, ou LAN (*Local Area Network*). Les LAN sont des réseaux privés, qui fonctionnent dans un seul bâtiment (ou à proximité), comme une maison, un immeuble de bureaux ou une usine. Ils sont fréquemment utilisés pour relier des ordinateurs personnels et des équipements électroniques grand public (par exemple des imprimantes) pour leur permettre de partager des ressources et d'échanger des informations. Quand ils sont employés par des organisations, on parle de **réseaux d'entreprise**.

Les LAN sans fil sont très répandus de nos jours, surtout dans les habitations, les immeubles de bureaux anciens, les cafétérias et autres lieux où l'installation de câbles poserait trop de problèmes. Dans ces systèmes, chaque ordinateur dispose d'un modem radio et d'une antenne, au moyen desquels il communique avec les autres ordinateurs. Dans la plupart des cas, chaque machine communique avec un équipement installé dans le plafond, comme le montre la figure 1.8(a). Cet équipement, appelé **point d'accès** (AP, Access Point), **routeur sans fil** ou **station de base**, relaie les paquets entre les ordinateurs sans fil, et entre ceux-ci et l'Internet. Être un point d'accès est un peu comme être le chouchou de la classe : tout le monde veut vous parler. Toutefois, si les machines sont suffisamment proches, elles peuvent communiquer directement dans une configuration poste-à-poste.

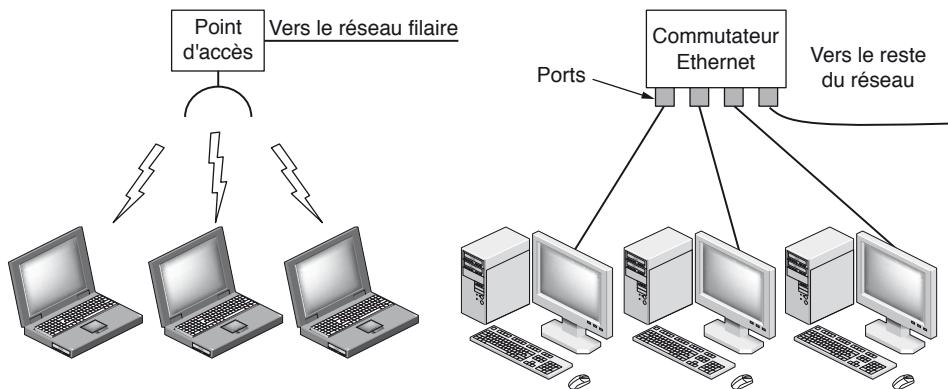


Figure 1.8 • LAN sans fil et filaires. (a) 802.11. (b) Ethernet commuté

Il existe pour les LAN sans fil une norme appelée **IEEE 802.11**, plus connue sous le nom de **Wi-Fi**, qui est maintenant très répandue. Elle permet des débits de un à plusieurs centaines de mégabits par seconde. (Dans ce livre, nous respecterons la tradition qui veut que l'on mesure le débit d'une ligne en mégabits par seconde, où 1 Mbit vaut un million de bits, et en gigabits par seconde, où 1 Gbit vaut un milliard de bits.) Nous étudierons la norme 802.11 au chapitre 4.

Les LAN filaires font appel à différentes technologies de transmission. La plupart d'entre elles utilisent du fil de cuivre, mais certaines sont à base de fibre optique. Les LAN sont limités en taille, ce qui veut dire que le temps de transmission le plus long est également limité et connu d'avance. Connaître ces restrictions est utile pour la conception des protocoles réseau.

Généralement, les LAN filaires offrent des débits de 100 Mbit/s à 1 Gbit/s, un faible délai (de l'ordre de quelques microsecondes ou nanosecondes) et connaissent très peu d'erreurs. Les plus récents peuvent atteindre 10 Gbit/s. Leurs performances sont supérieures en tout point à celles des réseaux sans fil : il est tout simplement plus facile de faire voyager des signaux sur du cuivre ou de la fibre que par voie aérienne.

La topologie de nombreux LAN filaires est construite à partir de liens point-à-point. La norme IEEE 802.3, plus connue sous le nom d'**Ethernet**, est de loin la plus courante pour les LAN filaires. La figure 1.8(b) représente un exemple de topologie pour un réseau **Ethernet commuté**. Chaque ordinateur « parle » le protocole Ethernet et est connecté à un **commutateur** (d'où son nom) par un lien point-à-point. Un commutateur possède plusieurs **ports**, chacun d'eux pouvant être connecté à un ordinateur. Sa tâche consiste à relayer les paquets entre les ordinateurs auxquels il est relié, en se basant sur l'adresse présente dans chaque paquet pour déterminer auquel l'envoyer.

Pour construire des LAN de plus grande envergure, il est possible de connecter des commutateurs en utilisant leurs ports. Que se passe-t-il si vous les connectez en boucle ? Le réseau continuera-t-il à fonctionner ? Heureusement, les concepteurs y ont pensé : c'est au protocole qu'il appartient de déterminer quel chemin les paquets

doivent emprunter pour atteindre sans entraves l'ordinateur ciblé. Nous verrons comment ce processus fonctionne au chapitre 4.

On peut également subdiviser un grand LAN physique en deux LAN logiques plus petits. Vous vous demandez peut-être quel est l'intérêt de l'opération. Il arrive que l'agencement de l'équipement réseau ne corresponde pas à la structure de l'organisation. Par exemple, il se peut que les ordinateurs du bureau d'études et de la comptabilité se trouvent sur le même LAN physique parce qu'ils résident dans la même aile du bâtiment, mais que le système soit plus facile à gérer si chaque département dispose de son propre réseau logique, autrement dit un **réseau virtuel** ou **VLAN** (*Virtual LAN*). Dans ce cas, chaque port est marqué d'une « couleur » différente, par exemple vert pour le bureau d'études et rouge pour la comptabilité. Le commutateur transmet alors les paquets de telle sorte que les ordinateurs reliés aux ports verts soient séparés des ordinateurs reliés aux ports rouges. Les paquets de diffusion envoyés sur un port rouge ne seront pas reçus sur un port vert, exactement comme s'il s'agissait de deux réseaux différents. Nous aborderons les VLAN à la fin du chapitre 4.

Il existe également d'autres topologies de LAN filaires. En fait, l'Ethernet commuté est une version moderne de l'Ethernet d'origine, qui diffusait tous les paquets sur un seul câble linéaire : une seule machine pouvait émettre à la fois, et un mécanisme d'arbitrage servait à résoudre les conflits. Il s'appuyait sur un algorithme simple : les ordinateurs pouvaient émettre chaque fois que le câble était inoccupé. Si deux paquets ou plus entraient en collision, chaque ordinateur se contentait d'attendre un laps de temps aléatoire et réessayait plus tard.

Pour plus de clarté, nous qualifierons cette version d'**Ethernet classique** et, comme vous vous en doutiez, nous la verrons plus en détail au chapitre 4.

Les réseaux à diffusion (tant filaires que sans fil) se répartissent aussi en systèmes statiques et systèmes dynamiques, selon la façon dont le canal est alloué. Une méthode d'allocation statique type consiste à diviser le temps en intervalles discrets et à utiliser un algorithme de tourniquet (*round robin*) : chaque machine émet à tour de rôle lorsque la tranche de temps qui lui a été accordée se présente. Ce fonctionnement gaspille toutefois la capacité du canal lorsqu'une machine n'a rien à transmettre, raison pour laquelle la plupart des systèmes tentent d'allouer le canal dynamiquement (c'est-à-dire à la demande).

L'allocation dynamique d'un canal partagé se fait de façon centralisée ou décentralisée. Dans le premier cas, il n'existe qu'une seule entité, par exemple la station de base d'un réseau cellulaire, qui détermine la prochaine machine autorisée à émettre. Pour cela, elle peut accepter des requêtes et un algorithme interne guide son choix. Dans le second cas, aucune entité n'assure l'arbitrage et chaque machine doit décider elle-même du moment à émettre. Cette technique ne mène pas au chaos comme on pourrait le penser de prime abord. De nombreux algorithmes, que nous étudierons plus loin, permettent de l'éviter.

Prenons un moment pour évoquer les réseaux domestiques. Dans le futur, il est probable que chaque appareil d'une habitation pourra communiquer avec tous les autres et que tous seront accessibles par l'Internet. C'est sans doute le genre de concept

visionnaire que personne n'a demandé, à l'instar de la télécommande de téléviseur ou du téléphone mobile, mais dont on ne peut plus se passer une fois qu'il s'est concrétisé.

Beaucoup d'appareils sont déjà aptes à être interconnectés : ordinateurs, appareils de divertissement, comme les téléviseurs et les lecteurs de DVD, téléphones et autres produits électroniques grand public comme les appareils photo, radioréveils et équipements d'infrastructure comme les compteurs d'eau et les thermostats. Cette tendance ne peut que se confirmer. Par exemple, le foyer moyen disposera probablement d'une dizaine d'horloges (dans les appareils), qui pourraient toutes être remises à l'heure automatiquement lors des changements d'horaire liés aux économies d'énergie si elles étaient reliées à l'Internet. Une application gagnante pourrait être la télésurveillance du domicile. Beaucoup de gens sont prêts à dépenser une certaine somme pour permettre à leurs parents âgés de vivre en sécurité dans leur propre maison.

S'il est possible de voir dans un réseau domestique un réseau comme un autre, il faut reconnaître qu'il présente des propriétés bien particulières. Premièrement, les équipements doivent être faciles à installer. Les routeurs sans fil représentent l'article le plus fréquemment retourné. On l'achète parce que l'on veut un réseau sans fil, on se rend compte qu'il ne fonctionne pas sans procédure d'installation et on le rapporte au magasin au lieu d'écouter la musique d'attente téléphonique de l'assistance technique.

Deuxièmement, le réseau et les appareils devront avoir un fonctionnement infaillible. Les utilisateurs de climatiseurs avaient l'habitude de ne voir qu'un bouton avec quatre positions : arrêt, bas, moyen et élevé. Ils sont maintenant confrontés à un manuel d'utilisation de 30 pages. Une fois que les appareils seront en réseau, attendez-vous à ce que le chapitre sur la sécurité fasse à lui seul 30 pages. C'est un problème, parce que seuls les utilisateurs d'ordinateurs sont habitués à se débrouiller avec des produits qui ne fonctionnent pas. Ceux qui achètent des téléviseurs, des voitures et des réfrigérateurs sont beaucoup moins tolérants : ils veulent que ces produits fonctionnent à 100 % sans devoir recourir aux services d'un spécialiste.

Troisièmement, un prix modique est la condition *sine qua non* pour un déploiement réussi. Personne ne paiera un abonnement de 50 € pour avoir un thermostat internet, car peu de gens considèrent qu'il est important de surveiller la température de leur habitation depuis leur lieu de travail. Mais à 5 €, l'idée pourrait faire des adeptes.

Quatrièmement, le particulier doit avoir la possibilité de commencer avec un ou deux appareils, puis d'étendre progressivement la portée de son réseau. Cela impose de bannir les guerres de formats. Encourager les consommateurs à acheter des périphériques munis d'une interface IEEE 1394 (*FireWire*), puis les orienter quelques années après vers l'USB 2.0 sous prétexte que c'est l'interface à la mode, pour passer ensuite à 802.11g, à moins que ce ne soit 802.11n ou 802.16 (des normes de réseau sans fil différentes), provoque de sérieuses réticences. L'interface réseau devra rester stable pendant des décennies, comme c'est le cas des normes de télédiffusion.

Cinquièmement, la sécurité et la fiabilité seront très importantes. Perdre quelques fichiers à cause d'un virus logé dans un courriel est une chose, se faire piller son domicile après qu'un voleur a désactivé le système de sécurité au moyen de son ordinateur portable en est une autre.

Une question intéressante est de savoir si les réseaux domestiques seront filaires ou sans fil. Les aspects financiers font pencher la balance du côté de la technologie sans fil, car il n'y a pas de câbles à installer ou, pire, à maintenir. L'aspect sécuritaire plaide en faveur du réseau filaire, parce que les ondes radio employées par le sans-fil traversent aisément les murs. Personne ne se réjouirait à l'idée de savoir qu'un voisin pirate sa connexion à l'Internet et lit son courriel. Au chapitre 8, nous étudierons comment le chiffrement peut servir à sécuriser les informations, ce qui est plus facile à dire qu'à faire pour des utilisateurs inexpérimentés.

Une troisième solution, qui peut être intéressante, consiste à exploiter les réseaux existants. Le **courant porteur en ligne**, ou CPL, permet aux appareils qui se connectent au secteur de diffuser des informations dans toute la maison. Certes, il faut brancher le téléviseur, mais l'on obtient la connectivité à l'Internet en même temps. Le problème réside dans la façon de transporter des signaux électriques et des données en même temps. Le fait qu'ils utilisent des bandes de fréquences différentes est une partie de la réponse.

En résumé, le réseau domestique offre des opportunités mais présente aussi des difficultés. La plupart de ces dernières sont dues au fait qu'il doit être simple à gérer, fiable et sûr, surtout entre des mains non techniciennes, tout en restant peu coûteux.

1.2.3 Réseaux métropolitains (MAN)

Un **réseau métropolitain**, ou MAN (*Metropolitan Area Network*), couvre une ville. L'exemple le plus connu de MAN est le réseau de télévision par câble que l'on trouve dans nombre d'agglomérations. Celui-ci a évolué à partir de l'ancien système d'antenne collective qui était utilisé dans les zones souffrant d'une mauvaise réception : on plaçait une grosse antenne en haut d'une colline voisine pour conduire ensuite le signal par câble jusqu'au domicile des abonnés.

Il s'agissait au départ de systèmes *ad hoc* dont la conception ne répondait qu'à des besoins locaux. Lorsque les entreprises se sont attaquées au marché, elles ont obtenu des administrations locales des contrats leur permettant de câbler des villes entières. L'étape suivante a été la programmation d'émissions de télévision, et même des chaînes entières expressément conçues pour le câble. Celles-ci étaient souvent très spécialisées, ne diffusant que des actualités, des émissions sportives, des recettes de cuisine, des conseils en jardinage, etc. Toutefois, lors de son apparition à la fin des années 1990, le système ne servait qu'à la réception des signaux de télévision.

Lorsque l'Internet a attiré une audience de masse, les câblo-opérateurs ont commencé à se rendre compte qu'avec peu de changements ils pourraient offrir un service d'accès à l'Internet bidirectionnel dans certaines parties inutilisées du spectre. Le système de télévision a alors commencé à se transformer en réseau métropolitain. En première approximation, on peut dire qu'un MAN ressemble à l'illustration de la figure 1.9. On y voit à la fois des signaux de télévision et le flux Internet qui transitent par la **tête de réseau** pour être distribués aux différentes habitations. Nous reviendrons en détail sur ce sujet au chapitre 2.

Toutefois, le réseau de télévision par câble n'est pas le seul réseau métropolitain. Les récents développements de l'accès à haut débit sans fil à l'Internet ont entraîné l'apparition d'un autre MAN, avec la norme IEEE 802.16, plus connue sous le nom de **WiMAX**. Nous l'étudierons au chapitre 4.

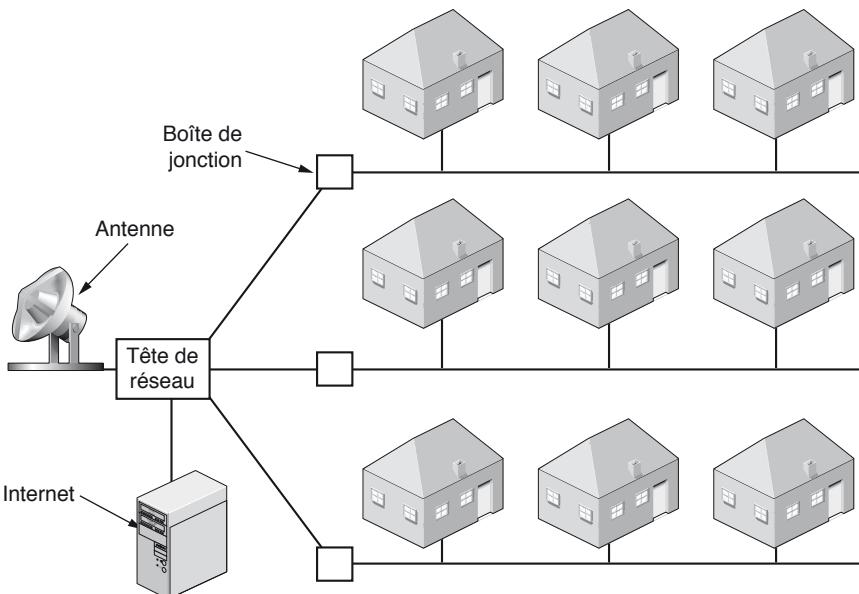


Figure 1.9 • Un réseau métropolitain fondé sur un réseau de télévision câblée.

1.2.4 Réseaux étendus (WAN)

Un **réseau étendu**, ou **WAN** (*Wide Area Network*), s'étend sur une vaste zone géographique (un pays, voire un continent). Nous commencerons par étudier les WAN filaires, en prenant l'exemple d'une entreprise possédant des filiales dans différentes villes.

Le WAN de la figure 1.10 est un réseau qui connecte des bureaux situés à Perth, Melbourne et Brisbane. Chacun de ces bureaux contient des ordinateurs destinés à l'exécution de programmes utilisateur (autrement dit d'applications). Nous respecterons la tradition et appellerons ces machines des **hôtes**. Le reste du réseau qui connecte ces hôtes s'appelle un **sous-réseau de communication**, souvent abrégé en **sous-réseau**. Celui-ci a pour tâche l'acheminement des messages d'un hôte à un autre, de la même manière que le système téléphonique transporte des paroles (des sons en réalité) d'un interlocuteur à un autre.

Dans la plupart des WAN, le sous-réseau se compose de deux types de composants : les lignes de transmission et les équipements de commutation. Les **lignes de transmission** transportent les bits d'une machine à une autre et peuvent être à base de fil

de cuivre, de fibre optique ou même prendre la forme de liaisons radio. La plupart des entreprises ne possédant pas de lignes de transmission, elles en louent à un opérateur téléphonique. Les équipements de **commutation**, ou plus simplement **commutateurs**, sont des ordinateurs spécialisés qui servent à interconnecter trois lignes de transmission ou plus. Lorsque des données arrivent sur une ligne entrante, l'équipement de commutation doit choisir une ligne sortante vers laquelle les aiguiller. Par le passé, plusieurs termes ont désigné ces machines, le plus usité aujourd'hui étant **routeur**.

Le terme « sous-réseau » mérite un petit commentaire supplémentaire. Il avait au départ une **seule** signification : l'ensemble des routeurs et des lignes de transmission chargés d'acheminer des paquets depuis un hôte source jusqu'à un hôte de destination. Sachez qu'il a acquis un second sens, plus récent, en rapport avec l'adressage. Nous traiterons ce sujet au chapitre 5, et nous en tiendrons jusque-là à l'acception traditionnelle (un ensemble de lignes et de routeurs).

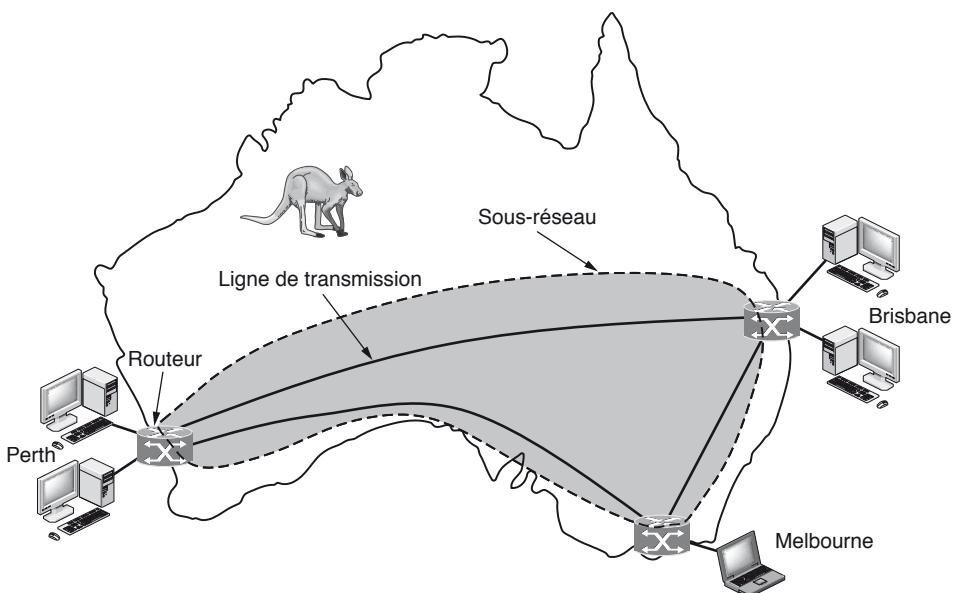


Figure 1.10 • WAN connectant des filiales en Australie.

Le WAN tel que nous l'avons décrit peut sembler analogue à un grand LAN filaire, mais il présente des différences importantes qui dépassent la question de la longueur des câbles. Généralement, dans un WAN, des personnes différentes possèdent et gèrent les hôtes et le sous-réseau. Dans notre exemple, les employés pourraient être responsables de leurs ordinateurs, tandis que le service informatique de l'entreprise aurait la charge du reste du réseau. Nous verrons des frontières plus claires dans les exemples à venir, où c'est le fournisseur d'accès ou l'opérateur téléphonique qui gère le réseau. La séparation entre l'aspect communication pure (le sous-réseau) et l'aspect applications (les hôtes) simplifie grandement la conception globale du réseau.

Deuxième différence, les routeurs connecteront généralement des réseaux utilisant différents types de technologies. Par exemple, les réseaux internes aux filiales peuvent être en Ethernet commuté, alors que les lignes de transmission longue distance peuvent être des liens SONET (voir le chapitre 2). Des équipements doivent leur permettre de communiquer. Le lecteur attentif aura remarqué que cela dépasse notre définition d'un réseau. Autrement dit, de nombreux WAN seront des **interréseaux**, des réseaux composites formés de plusieurs réseaux.

Enfin, une dernière différence tient à la nature des éléments connectés. Il peut s'agir d'ordinateurs individuels, comme dans le cas des LAN, ou bien de LAN entiers. C'est ainsi que l'on construit des grands réseaux à partir de plus petits. En ce qui concerne les sous-réseaux, leur rôle est identique.

Nous sommes maintenant en mesure de voir deux autres variétés de WAN. Premièrement, au lieu de louer des lignes de transmission dédiées, une entreprise peut connecter ses filiales à l'Internet, ce qui lui permet de mettre en œuvre les connexions entre elles sous forme de liens virtuels. Cet agencement, représenté à la figure 1.11, s'appelle un **réseau privé virtuel**, ou VPN (*Virtual Private Network*). Comparé à l'agencement dédié, un VPN offre l'avantage habituel de la virtualisation : la réutilisation facile d'une ressource (la connectivité à l'Internet). Pour le constater, voyez comme il serait facile d'ajouter une nouvelle filiale. En revanche, il en présente aussi l'inconvénient habituel : l'absence de contrôle sur les ressources sous-jacentes. Avec une ligne dédiée, la capacité est claire. Avec un VPN, elle peut varier en fonction de votre service Internet.

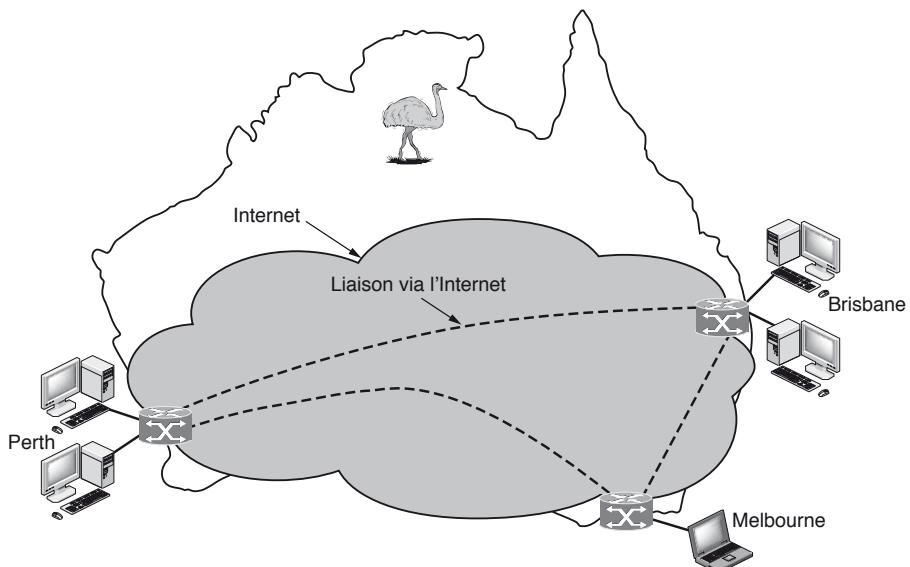


Figure 1.11 • WAN utilisant un réseau virtuel privé.

Dans la seconde variante, le sous-réseau peut être géré par une société différente – le fournisseur de services réseau – et les filiales sont ses clients. Cette structure est représentée à la figure 1.12. L'opérateur du sous-réseau sera également connecté à d'autres clients, tant qu'ils peuvent payer et qu'il peut fournir le service. Comme ce service serait bien décevant si les clients ne pouvaient que s'envoyer des paquets entre eux, cet opérateur les connectera aussi à d'autres réseaux qui font partie de l'Internet. Un tel opérateur se nomme un **FAI** (*fournisseur d'accès à l'Internet*), et le sous-réseau est un réseau de FAI. Les clients qui sont connectés au FAI reçoivent le service Internet.

Nous pouvons utiliser le réseau de FAI pour avoir un aperçu de sujets importants que nous étudierons aux chapitres suivants. La plupart des WAN contiennent plusieurs lignes de transmission, chacune reliant deux routeurs. Si deux routeurs qui ne partagent pas de ligne commune souhaitent communiquer, ils doivent le faire indirectement, en passant par d'autres routeurs. Or, plusieurs chemins peuvent connecter ces deux routeurs. La façon dont le réseau décide du chemin à emprunter s'appelle l'**algorithme de routage**. Il en existe de nombreux. La façon dont chaque routeur décide de l'endroit où envoyer un paquet s'appelle l'**algorithme de transfert (forwarding)**. Il en existe aussi de nombreux. Nous en étudierons plusieurs de chaque type au chapitre 5.

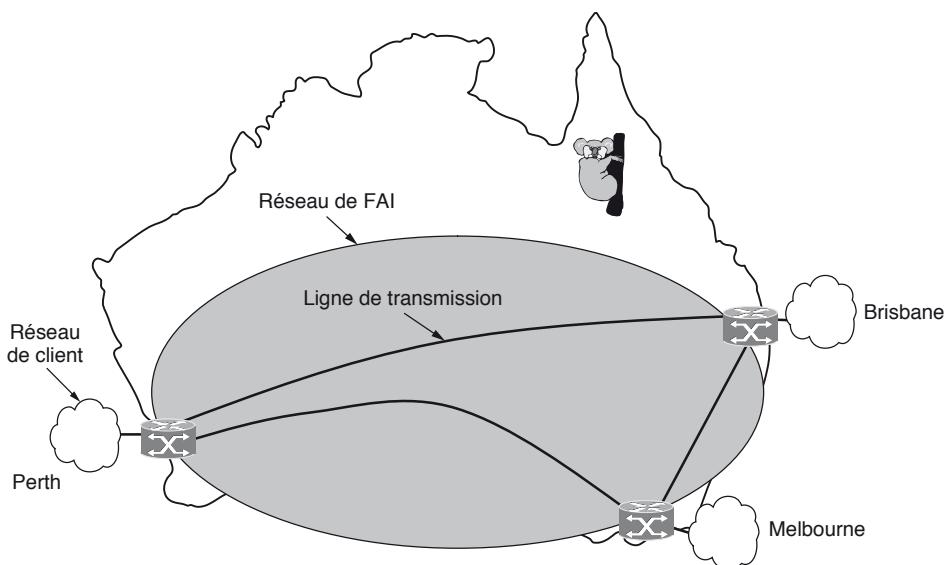


Figure 1.12 • WAN utilisant un réseau de FAI.

D'autres types de WAN utilisent intensivement les technologies sans fil. Dans les systèmes satellitaires, tous les ordinateurs au sol sont équipés d'une antenne grâce à laquelle ils peuvent émettre des données vers un satellite en orbite et en recevoir. Tous entendent les émissions *en provenance* du satellite, et, dans certains cas, celles des

ordinateurs voisins *vers* le satellite. Les réseaux satellitaires sont à diffusion générale par nature, et ce sont les plus adaptés lorsqu'il est important de pouvoir bénéficier de cette propriété.

Le réseau téléphonique cellulaire est un autre exemple de WAN qui s'appuie sur une technologie sans fil. Ce système a déjà connu trois générations, et une quatrième est à l'horizon. La première génération était analogique et ne transportait que la voix. La deuxième était numérique, mais toujours dédiée uniquement à la voix. La troisième génération, numérique aussi, a ajouté le transport de données. Chaque station de base couvre une distance beaucoup plus importante qu'un LAN sans fil, avec une portée mesurée en kilomètres et non en dizaines de mètres. Les stations sont connectées entre elles par un réseau fédérateur (*backbone*). Les réseaux cellulaires ont un débit de l'ordre de 1 Mbit/s, donc très inférieur à celui d'un LAN sans fil qui peut atteindre 100 Mbit/s. Nous en dirons beaucoup plus sur ces réseaux au chapitre 2.

1.2.5 Interréseaux

Il existe beaucoup de réseaux dans le monde, souvent composés de matériels et de logiciels différents. Les personnes connectées à un réseau souhaitent fréquemment communiquer avec d'autres personnes reliées à d'autres réseaux. Pour satisfaire à cette demande, les divers réseaux, souvent incompatibles entre eux, doivent être interconnectés. Un ensemble de réseaux ainsi reliés s'appelle un **interréseau** (*internetwork*), l'Internet en étant l'exemple le plus connu aujourd'hui. Il utilise des réseaux de FAI pour connecter des réseaux d'entreprise, des réseaux domestiques et bien d'autres réseaux. Nous étudierons l'Internet très en détail au long de ce livre.

Une certaine confusion règne quant à la signification des termes de sous-réseau, réseau et interréseau. Le terme de sous-réseau prend tout son sens dans le contexte d'un réseau étendu, où il renvoie à l'ensemble des routeurs et des lignes de transmission qui appartiennent à l'opérateur du réseau. En procédant par analogie, prenez le système téléphonique qui se compose de centres de commutation reliés entre eux par des lignes à haut débit et qui sont connectés aux habitations et aux entreprises par des lignes à faible débit. Ces lignes et ces équipements, appartenant à l'opérateur de télécommunications qui en assure la maintenance et le bon fonctionnement, forment le sous-réseau du système téléphonique. Les téléphones eux-mêmes (les hôtes dans notre analogie) n'en font pas partie.

Un réseau est formé par la combinaison d'un sous-réseau et de ses hôtes. Toutefois, le mot réseau est souvent employé dans un sens imprécis. On peut dire qu'un sous-réseau est un réseau, comme dans le cas du « réseau de FAI » de la figure 1.12. On peut également dire qu'un interréseau est un réseau, comme dans le cas du WAN de la figure 1.12. Nous adopterons une pratique similaire et, pour distinguer un réseau d'autres agencements, nous nous en tiendrons à notre définition d'origine : un ensemble d'ordinateurs interconnectés par une seule technologie.

Mais précisons ce qu'est un interréseau. Nous savons qu'un interréseau est formé lorsque des réseaux distincts sont interconnectés. De notre point de vue, la connexion d'un LAN et d'un WAN ou de deux LAN est la façon habituelle de constituer un

interréseau, mais il n'y a pas de consensus sur cette terminologie au sein de l'industrie. Deux règles empiriques sont utiles. Premièrement, si différentes parties du réseau ont été financées par diverses organisations et que celles-ci assurent la maintenance de leurs composants respectifs, on est en présence d'un interréseau et non d'un réseau unique. Deuxièmement, si les technologies de communication sous-jacentes propres aux diverses parties diffèrent (par exemple à diffusion *vs* point-à-point, ou sans fil *vs* filaire), on a probablement affaire à un interréseau.

Pour aller plus loin, nous devons parler de la façon dont deux réseaux différents peuvent être connectés. Le nom générique d'un équipement qui établit une connexion entre deux réseaux ou plus, tant en termes de matériels que de logiciels, est une **passerelle** (*gateway*). On distingue les passerelles selon la couche au niveau de laquelle elles opèrent dans la hiérarchie de protocoles. Nous verrons plus en détail les couches et les hiérarchies de protocoles à partir de la section suivante. Pour l'instant, dites-vous que les couches supérieures sont plutôt liées aux applications, comme le Web, et que les couches inférieures sont plutôt liées aux liens de transmission, comme Ethernet.

L'avantage d'un interréseau est qu'il permet de connecter des ordinateurs appartenant à des réseaux différents. En conséquence, on n'utilisera pas de passerelle de trop bas niveau, sous peine d'être incapable d'établir de connexions entre des réseaux de différents types. De même, avec une passerelle de trop haut niveau, la connexion ne fonctionnerait que pour des applications particulières. Le niveau intermédiaire, le « juste milieu », est souvent appelé la couche réseau, et un routeur est une passerelle qui commute des paquets à ce niveau. Nous pouvons maintenant repérer un interréseau au fait que c'est un réseau qui contient des routeurs.

1.3 Logiciels de réseaux

Lors de la conception des premiers réseaux d'ordinateurs, les ingénieurs se sont surtout concentrés sur l'aspect matériel, reléguant l'aspect logiciel au second plan. Cette stratégie ne tient plus. Les logiciels de réseaux sont aujourd'hui hautement structurés. Les sections suivantes examinent cette technique de structuration. L'approche que nous décrivons constitue la clé de voûte de tout l'ouvrage, et nous nous y référons souvent.

1.3.1 Hiérarchie de protocoles

Afin de réduire la complexité de conception, la plupart des réseaux sont organisés en strates, appelées **couches** ou **niveaux**, chacune étant placée au-dessus de la précédente. Le nombre de couches ainsi que le nom, le contenu et la fonction de chacune d'elles diffèrent selon les réseaux. Le rôle de chaque couche est de fournir des services à la couche immédiatement supérieure tout en lui dissimulant les détails d'implémentation. En un sens, on peut dire que chaque niveau représente une sorte de machine virtuelle qui offre certains services au niveau supérieur.

Il s'agit d'ailleurs d'un concept familier et très répandu en informatique, connu sous diverses appellations : masquage des informations, types de données abstraits, encapsulation de données ou encore programmation orientée objet. Quoi qu'il en soit, l'idée de base est de faire en sorte qu'un composant, logiciel ou matériel, puisse proposer un service à ses utilisateurs sans que ces derniers aient à connaître le détail de son état ni ses algorithmes.

Lorsque la couche n d'une machine dialogue avec la couche n d'une autre machine, les règles et les conventions qui gouvernent cette communication sont regroupées sous le nom de protocole de couche n . En substance, un **protocole** est une convention acceptée par les parties communicantes sur la façon dont leur dialogue doit avoir lieu. À titre de comparaison, imaginez une situation où une femme est présentée à un homme. Elle décide de lui tendre la main. L'homme peut opter pour une poignée de main ou pour un baisemain, selon qu'il s'agit d'une avocate dans une réunion professionnelle ou d'une princesse lors d'un bal officiel. La violation du protocole rendrait la communication difficile, voire impossible.

La figure 1.13 illustre un réseau en cinq couches. Les entités comprenant les couches correspondantes sur des machines différentes sont appelées **pairs**. Ces pairs peuvent être des processus logiciels, des équipements ou même des êtres humains. En d'autres termes, ce sont les pairs qui communiquent, en utilisant le protocole pour dialoguer.

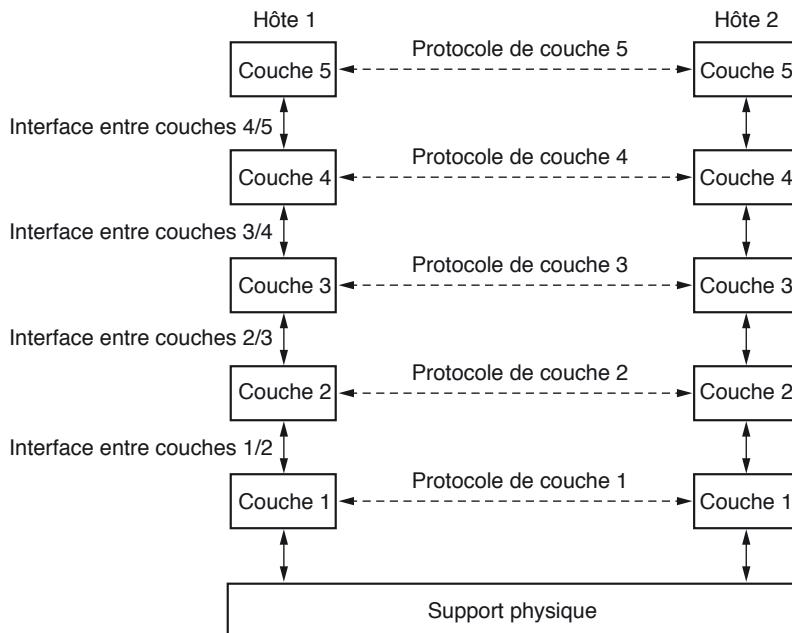


Figure 1.13 • Les couches, les protocoles et les interfaces.

En réalité, aucune donnée n'est directement transmise de la couche n d'une machine à la couche n d'une autre machine. Chaque couche transmet les données et les informations de contrôle à la couche immédiatement inférieure, jusqu'à ce que la couche la plus basse soit atteinte. Vient ensuite le **support physique** grâce auquel la communication a lieu. À la figure 1.13, la communication virtuelle est illustrée par des lignes pointillées et la communication physique par des lignes continues.

Entre deux paires de couches adjacentes, on trouve une **interface**. Celle-ci définit les opérations fondamentales et les services que la couche inférieure offre à la couche supérieure. Lors de la détermination du nombre de couches à intégrer dans un réseau et des opérations que chacune d'elles doit réaliser, l'une des tâches les plus importantes pour les concepteurs est de définir des interfaces claires entre les couches. Pour cela, chaque couche doit effectuer un ensemble de fonctions bien définies. Outre le fait de réduire la quantité d'informations passées entre les différentes couches, disposer d'interfaces claires permet aussi de facilement remplacer l'implémentation d'une couche par un autre protocole ou une autre implémentation totalement différente (par exemple le remplacement de toutes les lignes téléphoniques par des canaux satellite). En effet, il suffit au nouveau protocole ou à l'implémentation d'offrir le même ensemble de services à la couche supérieure que l'ancienne. Il n'est pas rare que des hôtes différents utilisent des implémentations différentes du même protocole (souvent écrites par des fournisseurs différents). En fait, le protocole lui-même peut changer dans une couche sans même que les couches supérieures et inférieures s'en aperçoivent.

Un ensemble de couches et de protocoles forme une **architecture de réseau**. La spécification d'une architecture doit contenir suffisamment d'informations pour qu'un développeur puisse écrire le programme (ou construire le matériel) pour chaque couche de façon que celui-ci obéisse correctement au protocole concerné. Ni les détails de l'implémentation, ni la spécification des interfaces ne font partie de l'architecture, car ils sont dissimulés dans la structure des machines, invisibles à l'environnement extérieur. Il n'est même pas utile que les interfaces soient les mêmes sur toutes les machines d'un réseau, du moment que chaque machine peut utiliser correctement tous les protocoles. L'ensemble des protocoles utilisés par un système donné, avec un protocole par couche, s'appelle une **pile de protocoles**. Les architectures de réseau, les piles de protocoles et les protocoles eux-mêmes sont les sujets principaux de ce livre.

Pour mieux comprendre ce que représente la communication en couches, imaginez la situation suivante. Deux philosophes (processus pairs de couche 3) souhaitent s'entretenir, mais l'un parle ourdou et anglais, et l'autre chinois et français. Faute de pouvoir partager une langue commune, ils engagent chacun un traducteur (processus pairs de couche 2), qui fait appel à son tour à une secrétaire (processus pairs de couche 1). Le premier philosophe souhaite communiquer à son homologue distant sa passion pour *Oryctolagus cuniculus*. Il transmet un message (en anglais) à son traducteur par l'intermédiaire de l'interface 2/3, en indiquant « I like rabbits », comme illustré à la figure 1.14. Comme les traducteurs se sont accordés sur une langue commune, en l'occurrence le néerlandais, le message est converti en « Ik vind konijnen leuk ».

Le choix de cette langue fait partie du protocole de couche 2 et appartient aux processus pairs de ce niveau.

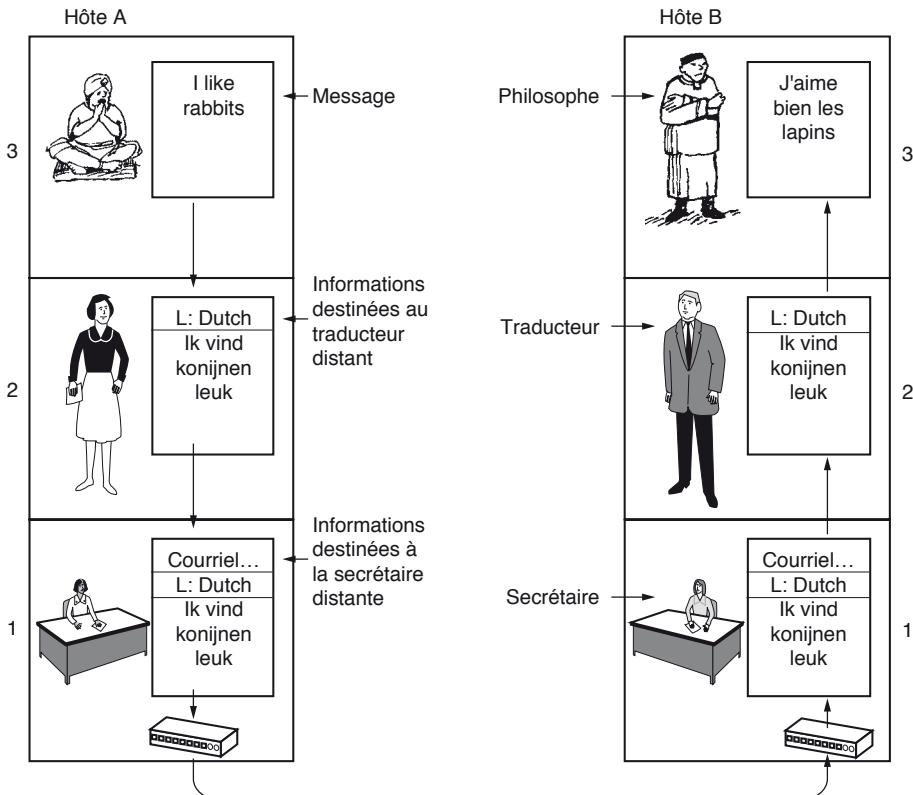


Figure 1.14 • L'architecture philosophe-traducteur-secrétaires.

Le traducteur passe le message à sa secrétaire pour transmission, par exemple, par courrier électronique (protocole de couche 1). Quand le message parvient à l'autre secrétaire, il est transmis au traducteur local, qui le traduit en français et le transmet par l'intermédiaire de l'interface 2/3 au deuxième philosophe. Notez que les protocoles sont complètement indépendants les uns des autres tant que les interfaces ne sont pas changées. Les traducteurs pourraient passer au finnois, pourvu qu'ils soient tous les deux d'accord et qu'aucun ne change ses interfaces avec les couches 1 et 3. De même, les secrétaires pourraient choisir le téléphone à la place du courrier électronique, sans déranger, ni même aviser, les autres couches. Chaque processus peut ajouter certaines informations à l'intention du processus pair uniquement, et celles-ci ne seront pas transmises à la couche supérieure.

Prenons maintenant un exemple plus technique pour comprendre comment se produit la communication entre deux processus pairs d'une couche supérieure.

Examinez le réseau à cinq couches illustré à la figure 1.15. Un message, M , est généré par un processus actif (une application quelconque) de la couche 5 et passé à la couche 4 pour transmission. La couche 4 place un **en-tête** au début du message et transmet l'unité obtenue à la couche 3. L'en-tête comprend des informations de contrôle, comme des adresses, pour permettre à la couche 4 sur l'ordinateur de destination d'assurer la livraison du message. Dans certaines couches, l'en-tête peut inclure d'autres informations de contrôle, comme des numéros de séquence (pour le cas où la couche inférieure ne respecterait pas l'ordre prévu) et des indications de taille et de temps.

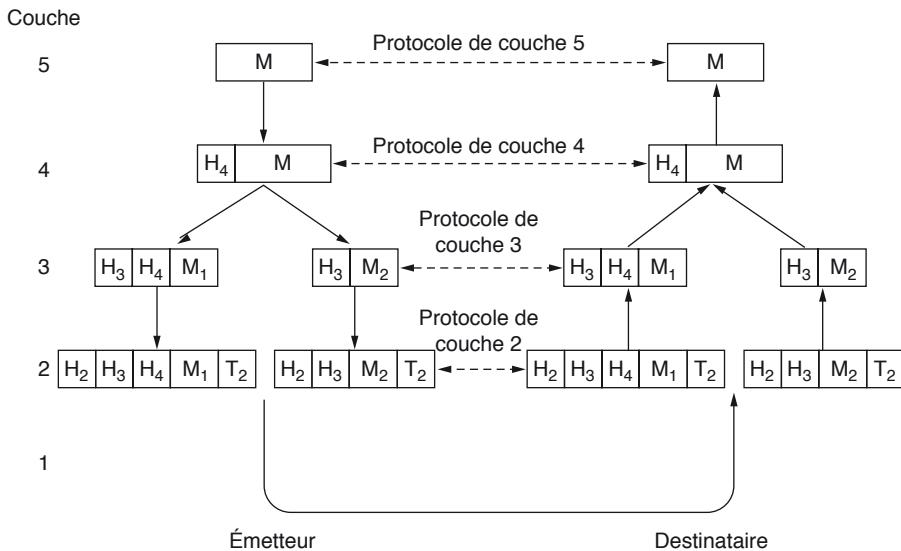


Figure 1.15 • Flots d'informations supportant une communication virtuelle de niveau 5.

Dans de nombreux réseaux, aucune limite de taille n'est imposée aux messages transmis dans le protocole de couche 4, alors que celui de la couche 3 en impose pratiquement toujours une. En conséquence, la couche 3 doit scinder les messages entrants en unités plus petites, des paquets, auxquelles elle ajoute un en-tête approprié. Dans notre exemple, M est divisé en deux parties, M_1 et M_2 , qui seront transmises séparément.

La couche 3 décide de la ligne sortante à emprunter et passe les paquets à la couche 2. Celle-ci ajoute à chacun non seulement un en-tête mais aussi un en-queue, et remet l'unité produite à la couche 1 qui se charge de l'envoi sur le support physique. Arrivé sur la machine de destination, le message remonte la pile, couche par couche, chaque niveau supprimant au passage l'en-tête le concernant. Aucune couche ne reçoit donc un en-tête d'une couche inférieure.

Ce qu'il importe de saisir dans l'exemple de la figure 1.15, c'est la relation entre communication virtuelle et communication réelle, et la différence entre protocoles et interfaces. D'un point de vue conceptuel, les processus pairs de la couche 4, par exemple, communiquent « horizontalement » grâce au protocole de la couche. Chaque processus dispose vraisemblablement de procédures du genre *envoie_à_l'autre_extrémité* et *reçoit_de_l'autre_extrémité*, même s'il communique en réalité avec la couche inférieure au travers de l'interface 3/4 et non avec l'autre extrémité.

L'abstraction sur laquelle repose la communication entre processus pairs est cruciale pour toute réalisation d'un réseau. La conception d'un réseau entier, qui serait ingérable autrement, peut ainsi être morcelée en problèmes plus petits et plus faciles à gérer, à savoir la conception des différentes couches.

Bien que cette section s'intitule « Logiciels de réseau », on va voir que, fréquemment, les couches inférieures d'une hiérarchie de protocoles sont implémentées au niveau matériel ou sous forme d'un microprogramme. Des algorithmes complexes sont néanmoins mis en œuvre, même s'ils sont, pour tout ou partie, intégrés dans le matériel.

1.3.2 Principes de conception des couches

Certains principes de conception des réseaux se retrouvent dans plusieurs couches. Nous décrivons ci-après brièvement les plus importants.

La fiabilité est le principe qui veut qu'un réseau fonctionne correctement, même s'il est constitué d'un ensemble de composants pas entièrement fiables. Pensez aux bits d'un paquet transitant sur le réseau. Il existe un risque que certains d'entre eux soit reçus endommagés (inversés) en raison d'un bruit électrique fortuit, de signaux radio aléatoires, de défauts matériels, de bogues logiciels et ainsi de suite. Comment est-il possible de s'apercevoir de ces erreurs et de les réparer ?

L'un des mécanismes permettant de repérer la présence d'erreurs dans les informations reçues utilise des codes de **détection d'erreur**. Les données qui ont été incorrectement reçues peuvent alors être retransmises jusqu'à ce qu'elles arrivent correctement. Des codes plus puissants permettent la **correction d'erreur**, dans laquelle le message correct est récupéré à partir des bits éventuellement incorrects reçus à l'origine. Ces deux mécanismes fonctionnent par ajout d'informations redondantes. Ils sont utilisés dans les couches inférieures, pour protéger les paquets envoyés sur les liens individuels, et dans les couches supérieures, pour vérifier que les bons contenus ont été reçus.

Une autre question liée à la fiabilité est celle du choix d'un chemin efficace dans un réseau. Il existe souvent plusieurs chemins menant d'une source à une destination, et, dans un grand réseau, certains liens ou certains routeurs peuvent être défaillants. Supposez une panne de réseau en Allemagne. Les paquets envoyés de Londres à Rome via l'Allemagne n'atteindront pas leur but, mais on pourrait à la place envoyer les paquets de Londres à Rome en passant par Paris. Le réseau doit prendre cette décision automatiquement : c'est ce que l'on appelle le **routage**.

Un deuxième principe de conception concerne l'évolution du réseau. Au fil du temps, les réseaux deviennent de plus en plus grands, et de nouvelles structures émergent, qui doivent être connectées à l'existant. Nous avons vu récemment un mécanisme de structuration capital utilisé pour prendre en charge les changements en décomposant le problème global et en masquant les détails de l'implémentation : l'organisation en **couches de protocoles**. Il existe beaucoup d'autres stratégies.

Comme de nombreux ordinateurs résident sur le réseau, chaque couche a besoin d'un mécanisme pour identifier les émetteurs et les récepteurs impliqués dans un message donné. Ce mécanisme s'appelle l'**adressage** dans les couches inférieures et le **nommage** dans les couches supérieures.

Autre aspect important : des technologies différentes ont souvent des contraintes différentes. Par exemple, tous les canaux de communication ne préservent pas l'ordre des messages qu'ils transportent, ce qui oblige à des solutions de numérotation des messages. Autre exemple, les différences dans la taille maximale d'un message que les réseaux peuvent transmettre. Il faut donc des mécanismes pour les désassembler, les transmettre puis les réassembler à l'arrivée. Ce sujet général s'appelle l'**interconnexion de réseaux**.

Quand les réseaux se développent, de nouveaux problèmes surgissent. Les grandes agglomérations peuvent connaître des embouteillages ou une pénurie de numéros de téléphone, et il est facile de s'y perdre. Si peu de personnes souffrent de ces problèmes dans leur propre quartier, ils peuvent devenir très ennuyeux à l'échelle d'une ville. Les conceptions qui continuent à bien fonctionner lorsque le réseau s'agrandit sont dites **évolutives**.

Un troisième principe de conception a trait à l'allocation des ressources. Les réseaux s'appuient sur leurs ressources sous-jacentes, comme la capacité des lignes de transmission, pour fournir un service aux hôtes. Pour le faire bien, il leur faut des mécanismes qui répartissent ces ressources de telle sorte qu'un hôte donné n'interfère pas trop avec un autre. Par exemple, on peut penser à partager la bande passante dynamiquement, en fonction des besoins à court terme des hôtes, au lieu d'en attribuer à chacun une fraction fixe qu'il utilisera ou non. Cette méthode s'appelle le **multiplexage statistique**, ce qui signifie que le partage est fondé sur les statistiques des demandes. On peut l'appliquer au niveau des couches inférieures pour un seul lien, ou au niveau des couches supérieures pour un réseau, voire aux applications qui utilisent le réseau.

Un problème d'allocation se pose à chaque niveau : comment empêcher un émetteur rapide de submerger de données un récepteur lent ? Ce sujet est celui du **contrôle de flux**. Parfois, le problème tient au fait que le réseau est en « surréservation », parce que trop d'ordinateurs veulent émettre trop de trafic, et que le réseau ne peut pas tout acheminer. On qualifie cette surcharge de **congestion**. Une stratégie consiste à faire en sorte que chacune réduise sa demande quand il constate une congestion. Elle aussi peut être appliquée au niveau de chaque couche.

Il est intéressant d'observer que les ressources que le réseau a à offrir ne se limitent pas à la bande passante. Pour des usages tels que le transport de la vidéo en direct, tout retard dans la transmission est catastrophique. La plupart des réseaux doivent

fournir un service aux applications qui ont besoin de cette livraison en **temps réel** en même temps qu'à celles qui ont besoin d'un débit important. **Qualité de service** est le nom donné aux mécanismes qui concilient ces deux demandes opposées.

Le dernier principe de conception majeur est celui qui consiste à sécuriser le réseau en le défendant contre différents types de menaces. L'une d'elles, déjà mentionnée, est l'espionnage des communications. Les mécanismes qui assurent la **confidentialité** protègent de ce danger et sont appliqués dans plusieurs couches. Les mécanismes d'**authentification** empêchent une personne d'usurper l'identité d'une autre. On peut les utiliser pour discerner les faux sites bancaires des vrais, ou faire en sorte que le réseau cellulaire vérifie qu'un appel provient bien de votre téléphone afin de vous facturer la communication. D'autres enfin, conçus pour l'**intégrité**, empêchent de modifier subrepticement des messages, comme changer « débitez mon compte de 10 € » en « débitez mon compte de 1 000 € ». Tous s'appuient sur la cryptographie, que nous étudierons au chapitre 8.

1.3.3 Services avec connexion et sans connexion

Les couches peuvent offrir deux types de services différents aux couches supérieures : avec connexion et sans connexion. Cette section examine ces services ainsi que leurs différences.

Un **service avec connexion** suit le modèle du système téléphonique. Pour converser avec quelqu'un, vous décrochez le téléphone, composez le numéro, parlez, puis raccrochez. De même, pour utiliser un service réseau avec connexion, l'utilisateur doit d'abord établir une connexion, l'utiliser, puis la libérer. L'aspect essentiel d'une connexion est qu'elle agit tel un tuyau : l'émetteur y injecte des objets (des bits) à une extrémité et le récepteur les récupère à l'autre bout. Dans la plupart des cas, l'ordre d'envoi des bits est préservé.

Dans certains cas, lorsqu'une connexion est établie, l'émetteur, le récepteur et le sous-réseau s'engagent dans une **négociation** à propos des paramètres à utiliser, tels que la taille maximale des messages, la qualité de service requise ou d'autres caractéristiques. Généralement, une proposition est faite par une extrémité et l'autre côté peut l'accepter, la rejeter ou faire une contre-proposition. Un **circuit** est un autre nom d'une connexion avec des ressources associées, comme une bande passante fixe. Ce terme date du réseau téléphonique, dans lequel un circuit était un chemin sur un fil de cuivre qui transportait une conversation.

À l'inverse, un **service sans connexion** suit le modèle du système postal. Chaque message (lettre) contient l'adresse de destination complète, et chacun est routé *via* les noeuds intermédiaires dans le système indépendamment de tous les messages suivants. Les messages portent des noms différents selon le contexte : un **paquet** est un message au niveau de la couche réseau. Lorsque les noeuds intermédiaires reçoivent un message en entier avant de le transmettre au prochain noeud, on parle de **commutation en mode différé** (*store-and-forward*). L'autre possibilité, dans laquelle un noeud commence à retransmettre un message avant de l'avoir entièrement reçu, est la **commutation en mode direct** (*cut-through*). Normalement, lorsque deux messages

sont envoyés à une même destination, le premier envoyé sera le premier à arriver. Toutefois, des retards peuvent se produire et faire en sorte que le second arrive avant.

Chaque service peut encore être caractérisé par sa **fiabilité**. Certains services sont qualifiés de fiables au sens où ils ne perdent jamais de données. En principe, un service fiable est implémenté au moyen d'un système de notification dans lequel le récepteur accuse réception de chaque message afin que l'émetteur soit certain qu'il est arrivé. Le processus d'acquittement provoque une surcharge de service et des délais, qui valent souvent la peine qu'on les supporte mais sont parfois indésirables.

Un cas typique d'emploi approprié d'un service fiable avec connexion est le transfert de fichiers. Le propriétaire d'un fichier veut être sûr que les bits arrivent correctement et dans le bon ordre. Peu de clients utilisant le transfert de fichiers préféreraient un service susceptible de mélanger ou de perdre de temps à autre quelques bits, même s'il était beaucoup plus rapide.

Le service fiable avec connexion possède deux variantes mineures : par séquences de messages et par flot d'octets. Dans la première, les limites des messages sont préservées. Lorsque deux messages de 1 024 octets chacun sont envoyés, ils arrivent sous forme de deux messages distincts, jamais sous forme d'un seul message de 2 048 octets. Dans la seconde, la connexion prend la forme d'un flot d'octets, sans limite de messages. Lorsque 2 048 octets sont reçus, il n'existe aucun moyen de savoir s'ils ont été envoyés sous forme d'un message de 2 048 octets, de deux messages de 1 024 octets ou de 2 048 messages d'un octet chacun. Si des pages d'un livre sont envoyées par le réseau chez un photocomposeur sous forme de messages séparés, il est important que leurs limites soient conservées. En revanche, pour télécharger un film, un flot d'octets du serveur vers l'ordinateur de l'utilisateur suffit. Dans un film, les limites de messages ne sont pas pertinentes.

Dans certaines applications, les délais introduits par les acquittements sont inacceptables. L'une d'elles est le transport de la voix numérisée, pour la **voix sur IP (VoIP)**. Il est moins gênant pour les usagers d'entendre du bruit sur la ligne de temps à autre que de supporter un retard occasionné par l'attente des acquittements. Dans le même ordre d'idées, quelques pixels défectueux ne posent pas de problèmes lors d'une conférence vidéo, alors qu'il serait irritant d'avoir des images saccadées en raison des arrêts de transmission liés à la correction d'erreurs.

Toutes les applications ne requièrent pas de connexion. Par exemple, les spameurs envoient des courriers indésirables (pourriels) à une foule de destinataires. Ils ne veulent certainement pas établir puis libérer une connexion vers chaque utilisateur pour envoyer un message à la fois. La livraison fiable à 100 % n'est pas essentielle non plus, surtout si elle coûte plus cher. Il suffit de pouvoir envoyer un seul message ayant une forte probabilité d'arriver, mais sans garantie. Un service non fiable sans connexion (c'est-à-dire non acquitté) est souvent appelé **service de datagramme**, par analogie avec le service télégraphique qui n'envoie pas non plus d'accusé de réception à l'émetteur. Malgré son caractère peu fiable, c'est le modèle dominant dans la plupart des réseaux, pour des raisons que nous expliciterons plus loin.

Dans d'autres situations, il est souhaitable pour des questions pratiques d'éviter d'établir une connexion pour envoyer un message, mais la fiabilité est essentielle.

Le **service de datagramme acquitté** peut alors être utilisé. Il s'apparente à l'envoi d'une lettre recommandée avec accusé de réception. Lorsque l'émetteur reçoit ce dernier, il a l'assurance que la lettre a été remise au destinataire et ne s'est pas perdue en route. La messagerie texte des téléphones mobiles en est un exemple.

Il existe un autre type de service appelé **demande-réponse** ou **requête-réponse**. Il permet d'envoyer un seul datagramme contenant une demande et de recevoir en retour la réponse. Par exemple, le possesseur d'un téléphone mobile peut envoyer une requête à un serveur pour obtenir une carte de l'endroit où il se trouve. La figure 1.16 résume ces types de services.

	Services	Exemples
Avec connexion	Flot de messages fiable	Suite de pages
	Flot d'octets fiable	Téléchargement d'un film
Sans connexion	Connexion non fiable	Voix sur IP
	Datagramme non fiable	Prospectus électronique
Sans connexion	Datagramme acquitté	Messagerie avec accusé de réception
	Demande-réponse	Interrogation d'une base de données

Figure 1.16 • Six types de services différents.

Le concept de communication non fiable peut être déroutant au départ. Après tout, pour quelle raison la préférerait-on à une communication fiable ? Tout d'abord, la communication fiable (au sens où nous l'entendons, c'est-à-dire avec acquittement), n'est pas toujours disponible dans une couche donnée. Par exemple, Ethernet ne propose pas ce type de service. Les paquets peuvent de temps à autre subir des dommages lors du transit. Il appartient alors aux protocoles des couches supérieures de résoudre ce problème. Par exemple, beaucoup de services fiables sont construits au-dessus de services de datagrammes non fiables. Ensuite, les délais inhérents à la fourniture d'un service fiable peuvent être inacceptables, surtout pour les applications en temps réel comme le multimédia. C'est pour ces raisons que ces deux services coexistent.

1.3.4 Primitives de service

Un service est formellement défini par un ensemble de **primitives** (opérations) qu'un processus utilisateur emploie pour accéder au service. Ces primitives indiquent au service d'exécuter une action ou de rendre compte d'une action entreprise par un processus pair. Si la pile de protocoles se trouve dans le système d'exploitation, comme c'est souvent le cas, les primitives prennent normalement la forme d'appels système. Ces appels provoquent un déroutement (*trap*) dans le mode noyau, ce qui permet au système d'exploitation d'envoyer les paquets requis.

L'ensemble des primitives disponibles dépend de la nature du service fourni. Un service avec connexion diffère d'un service sans connexion. La figure 1.17 énumère quelques primitives disponibles pour un flot d'octets fiable. Elles seront familières aux fans des sockets de Berkeley, car elles constituent une version simplifiée de cette interface.

Figure 1.17 • Six primitives de service implémentant un service simple avec connexion.

Primitives	Description
LISTEN	État bloquant dans l'attente d'une connexion entrante
CONNECT	Établissement d'une connexion avec un processus pair à l'écoute
ACCEPT	Acceptation d'une connexion entrante d'un processus pair
RECEIVE	État bloquant dans l'attente d'un message entrant
SEND	Envoi d'un message au processus pair
DISCONNECT	Libération d'une connexion

Ces primitives pourraient être utilisées pour une interaction requête-réponse dans un environnement client-serveur. Pour illustrer ce cas, nous esquisserons un protocole simple qui implémente le service en utilisant des datagrammes acquittés.

Tout d'abord, le serveur exécute la primitive LISTEN pour indiquer qu'il est prêt à accepter des connexions entrantes. Cette primitive est souvent implémentée sous forme d'un appel système bloquant. Après son exécution, le processus serveur reste dans un état bloquant jusqu'à ce qu'une requête de connexion se présente.

Ensuite, le processus client exécute CONNECT pour établir une connexion au serveur. L'appel doit spécifier l'entité avec laquelle la connexion doit être établie et peut donc contenir un paramètre fournissant l'adresse du serveur. En général, le système d'exploitation envoie alors un paquet au processus pair demandant l'établissement d'une connexion, comme illustré en (1) à la figure 1.18. Le processus client est suspendu jusqu'à ce qu'une réponse arrive.

Lorsque le paquet arrive sur le serveur, le système d'exploitation constate qu'il s'agit d'une requête de connexion. Il vérifie qu'un processus est bien à l'écoute et, si c'est le cas, le débloque. Le processus serveur peut alors établir la connexion au moyen de l'appel ACCEPT, ce qui envoie une réponse (2) au processus client pour accepter la connexion. L'arrivée de cette réponse provoque alors la libération du client. À ce stade, les processus client et serveur sont tous deux actifs et partagent une connexion.

Ce protocole est comparable à une situation fréquente de la vie courante : un client appelant le directeur d'un service clientèle. Au début de la journée, le directeur se place près du téléphone, pour le cas où il sonnerait. Plus tard, un client appelle. Au moment où le directeur décroche, la connexion est établie.

L'étape suivante pour le serveur consiste à exécuter RECEIVE qui le prépare à la réception d'une première requête. Normalement, le serveur exécute immédiatement cette primitive dès qu'il est libéré de l'état LISTEN, avant même que l'acquittement ne parvienne au client. L'appel RECEIVE bloque le serveur.

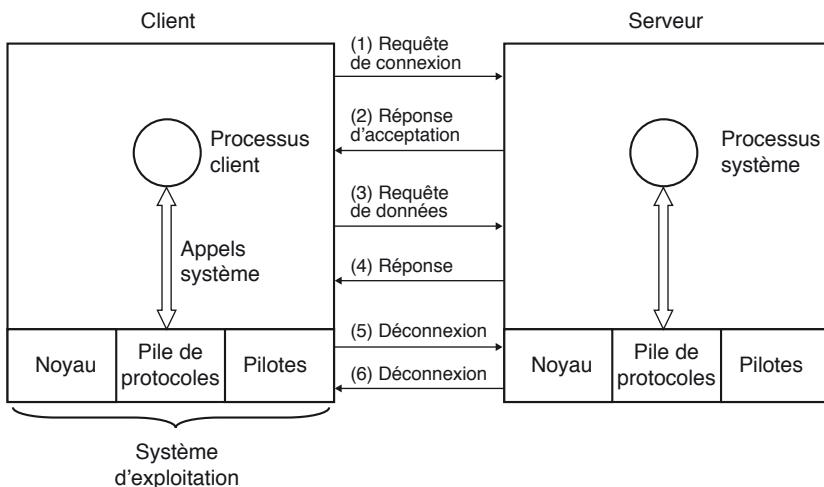


Figure 1.18 • Une interaction client-serveur simple utilisant des datagrammes acquittés.

Le client exécute ensuite successivement les primitives SEND, pour transmettre une requête (3), et RECEIVE, pour se préparer à recevoir une réponse. L'arrivée du paquet de requête sur la machine serveur débloque le processus serveur, pour qu'il puisse traiter la requête. Après avoir terminé la tâche requise, le serveur envoie la réponse au client au moyen de la primitive SEND (4). La réception de ce paquet sur la machine du processus client débloque ce dernier, qui peut alors inspecter la réponse. Si le client a des requêtes supplémentaires, il peut maintenant les émettre.

Lorsque le client a terminé, il exécute DISCONNECT pour mettre fin à la connexion (5). Le premier appel de cette primitive est généralement bloquant et provoque l'envoi d'un paquet au serveur lui indiquant que la connexion n'est plus nécessaire. Lorsque le serveur reçoit le paquet, il exécute sa propre primitive DISCONNECT, accusant ainsi réception au client et libérant la connexion (6). Lorsque le paquet du serveur parvient au client, le processus client est débloqué et la connexion est rompue. Globalement, une communication en mode connexion fonctionne de cette manière.

Bien sûr, la vie n'est pas aussi simple et beaucoup de choses peuvent venir perturber les échanges, notamment une mauvaise synchronisation (par exemple, un appel CONNECT précédant un appel LISTEN), des paquets perdus, etc. Nous traiterons cela en détail plus loin. Pour l'instant, la figure 1.18 résume la façon dont la communication client-serveur pourrait fonctionner avec des datagrammes acquittés, pour que l'on puisse ignorer les paquets perdus.

Étant donné que six paquets sont nécessaires pour accomplir ce protocole, d'aucuns peuvent se demander pourquoi ne pas utiliser à la place le mode sans connexion. La réponse est que ce serait possible dans un monde parfait, auquel cas deux paquets suffiraient : un pour la requête et un pour la réponse. Toutefois, en présence de messages volumineux (par exemple un fichier d'un mégaoctet) dans les deux sens, d'erreurs de transmission et de paquets perdus, la situation change. Si une réponse

se compose de centaines de paquets, certains peuvent se perdre durant la transmission. Dans ce cas, comment le client peut-il savoir que rien ne manque ou que le dernier paquet reçu est bien le dernier envoyé ? Supposez que le client souhaite recevoir un second fichier. Comment peut-il distinguer le paquet 1 du second fichier d'un paquet 1 perdu provenant du premier et qui aurait soudain retrouvé son chemin jusqu'au client ? En résumé, dans le monde réel, un simple protocole du genre demande-réponse dans un réseau non fiable est souvent insuffisant. Au chapitre 3, nous étudierons en détail des protocoles qui permettent de régler ces problèmes parmi d'autres. Pour le moment, il suffira de dire qu'il est parfois pratique de pouvoir disposer d'un flot d'octets fiable et ordonné.

1.3.5 Relations des services aux protocoles

Les services et les protocoles sont des concepts différents. Cette distinction est si importante que nous insistons une fois encore. Un *service* est un ensemble de primitives (d'opérations) qu'une couche fournit à la couche supérieure. Il définit les opérations que la couche est prête à exécuter pour le compte de ses utilisateurs, mais ne renseigne pas sur la façon dont ces opérations sont implémentées. Il se rapporte à une interface entre deux couches, la couche inférieure étant le fournisseur de service et la couche supérieure l'utilisateur du service.

Un *protocole*, quant à lui, est un ensemble de règles qui déterminent le format et la signification des paquets ou des messages qui sont échangés par des entités paires au sein d'une couche. Ces entités utilisent des protocoles pour implémenter les définitions de services. Elles sont libres de changer leurs protocoles comme bon leur semble, à condition de ne pas modifier le service tel que les utilisateurs le voient. On constate ainsi que les concepts de service et de protocole sont totalement dissociés. C'est là une notion capitale, que les concepteurs de réseaux doivent maîtriser.

Pour répéter ce point crucial, les services se rapportent aux interfaces entre les couches, comme illustré à la figure 1.19, alors que les protocoles se rapportent aux paquets qui sont échangés par deux entités paires sur deux machines différentes. On voit là encore qu'il est primordial de distinguer les deux concepts.

Nous pouvons aussi établir une analogie avec les langages de programmation. Un service s'apparente à un type de données abstrait ou à un objet dans un langage de programmation orienté objet ; il définit les opérations qui peuvent être exécutées sans spécifier comment elles sont implémentées. En revanche, un protocole correspond à l'*implémentation* du service et, en tant que tel, il est invisible pour l'utilisateur du service.

Beaucoup de protocoles anciens n'observaient pas cette séparation entre service et protocole. En effet, une couche typique pouvait offrir une primitive de service SEND PACKET et l'utilisateur fournissait un pointeur vers un paquet entièrement assemblé. Une telle organisation signifiait qu'un changement apporté à un protocole était immédiatement visible par l'utilisateur. Aujourd'hui, la plupart des concepteurs de réseau considèrent cela comme une grave erreur.

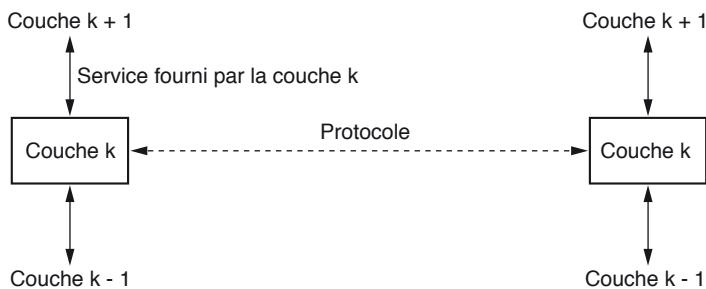


Figure 1.19 • La relation fonctionnelle entre service et protocole.

1.4 Modèles de référence

Maintenant que nous avons abordé l'aspect abstrait des réseaux en couches, nous pouvons examiner quelques exemples. Nous verrons deux architectures de réseau de première importance : les modèles de référence OSI et TCP/IP. Bien que les *protocoles* associés au modèle OSI soient rarement utilisés aujourd'hui, le *modèle* en lui-même est tout à fait général et reste valide : les fonctionnalités examinées au niveau de chaque couche sont toujours très importantes. Avec le modèle TCP/IP, c'est l'inverse qui se produit : le modèle n'est pas très employé alors que ses protocoles sont largement déployés. Nous étudierons donc en détail ces deux modèles. De plus, on apprend parfois plus des échecs que des succès.

1.4.1 Le modèle de référence OSI

Le modèle OSI (moins le support physique) est illustré à la figure 1.20. Il s'appuie sur une proposition qui a été développée par l'ISO (*Organisation internationale de normalisation*) comme une première étape vers la normalisation internationale des protocoles utilisés dans les diverses couches. Il a été révisé en 1995. On l'appelle **modèle de référence OSI** (*Open Systems Interconnection*) car il traite des systèmes ouverts, c'est-à-dire des systèmes ouverts à la communication avec d'autres systèmes. Nous l'appellerons simplement le **modèle OSI**.

Ce modèle se compose de sept couches. Les principes qui ont été appliqués pour parvenir à ce nombre peuvent être résumés brièvement de la façon suivante :

1. Une couche doit être créée lorsqu'un nouveau niveau d'abstraction est nécessaire.
2. Chaque couche doit assurer une fonction bien définie.
3. La fonction de chaque couche doit être choisie en visant la définition de protocoles normalisés au niveau international.
4. Les limites d'une couche doivent être fixées de manière à réduire la quantité d'informations devant passer au travers des interfaces.

5. Le nombre de couches doit être, d'une part, assez grand pour que des fonctions très distinctes ne soient pas regroupées dans une même couche et, d'autre part, suffisamment faible pour que l'architecture ne devienne pas trop complexe.

Les sections suivantes examinent chaque couche tour à tour en commençant par la plus basse. Notez que le modèle OSI ne constitue pas en lui-même une architecture de réseau car il ne spécifie pas les services et les protocoles précis à utiliser dans chaque couche. Il précise simplement ce que chaque couche doit faire. Toutefois, l'ISO a également produit des normes pour chaque couche, bien qu'elles ne fassent pas partie du modèle. Chacune d'elles a été publiée sous forme de norme internationale distincte. Le *modèle* (en partie) est largement utilisé, même si les protocoles associés sont oubliés depuis longtemps.

La couche physique

La **couche physique** se charge de la transmission de bits à l'état brut sur un canal de communication. L'un des objectifs de conception de ce niveau est de s'assurer qu'un bit à 1 envoyé sur une extrémité arrive aussi à 1 de l'autre côté et non à 0. Les questions que l'on se pose portent généralement sur les signaux électriques à utiliser pour représenter un 1 et un 0, le nombre de nanosecondes que doit durer un bit, la possibilité de transmission bidirectionnelle simultanée, la façon dont une connexion initiale est établie puis libérée lorsque les deux extrémités ont fini, ou encore le nombre de broches d'un connecteur et le rôle de chacune. Les problèmes de conception concernent principalement les interfaces mécaniques et électriques et la synchronisation, ainsi que le support physique de transmission sous-jacent à la couche physique.

La couche liaison de données

Le rôle principal de la **couche liaison de données** est de faire en sorte qu'un moyen de communication brut apparaisse à la couche réseau comme une liaison exempte d'erreurs de transmission non détectées. Pour ce faire, elle masque les erreurs réelles afin que la couche réseau ne les voie pas. Elle accomplit cette tâche en décomposant les données de l'entrée en **trames de données** (généralement de quelques centaines ou quelques milliers d'octets), qu'elle envoie en séquence. S'il s'agit d'un service fiable, le récepteur confirme la bonne réception de chaque trame en envoyant à l'émetteur une **trame d'acquittement**.

Un autre point important à résoudre dans cette couche, mais aussi dans la plupart des couches supérieures, est de savoir comment éviter qu'un récepteur lent soit submergé par les données d'un émetteur rapide. Recourir à des mécanismes de régulation du trafic peut être nécessaire pour permettre à l'émetteur de savoir quand le récepteur peut accepter plus de données.

Les réseaux à diffusion sont confrontés à une difficulté supplémentaire : comment contrôler l'accès au canal partagé. C'est une sous-couche spéciale de la couche liaison de données, la sous-couche de **contrôle d'accès au média**, ou **sous-couche MAC**, qui gère ce problème.

La couche réseau

La **couche réseau** contrôle le fonctionnement du sous-réseau. Un élément essentiel de sa conception vise à déterminer la façon dont les paquets sont routés de la source vers la destination. Les routes peuvent être choisies au moyen de tables statiques « câblées » dans le réseau et rarement modifiées ou, plus souvent, mises à jour automatiquement pour éviter les composants défaillants. Elles peuvent également être déterminées au début de chaque conversation, par exemple, lors d'une connexion à partir d'un terminal (comme une ouverture de session sur une machine distante). Enfin, elles peuvent aussi être très dynamiques et déterminées à nouveau pour chaque paquet, afin de prendre en compte la charge actuelle du réseau.

La présence de trop de paquets en même temps sur le sous-réseau provoque des goulets d'étranglement. La gestion de la congestion incombe également à la couche réseau, en conjonction avec les couches supérieures qui adaptent la charge qu'elles placent sur le réseau. Plus généralement, la qualité du service fourni (délai, temps de transmission, gigue, etc.) relève également de cette couche.

Lorsqu'un paquet doit passer d'un réseau à un autre pour atteindre sa destination, de nombreux problèmes peuvent surgir : la technique d'adressage du second réseau peut être différente de celle du premier, le second réseau peut refuser le paquet s'il est trop grand, les protocoles peuvent être différents, etc. La couche réseau doit gérer tous ces problèmes pour que des réseaux hétérogènes puissent être interconnectés.

Sur un réseau à diffusion, le problème du routage est simple et la couche réseau est donc souvent mince, voire inexistante.

La couche transport

La fonction de base de la **couche transport** est d'accepter des données de la couche supérieure, de les diviser en unités plus petites si nécessaire, de les transmettre à la couche réseau et de s'assurer qu'elles arrivent correctement à l'autre bout. De plus, tout cela doit être réalisé efficacement et d'une façon qui isole les couches supérieures des inévitables changements matériels dus aux progrès technologiques.

La couche transport détermine aussi le type de service à fournir à la couche session, et, en fin de compte, aux utilisateurs du réseau. Le type de connexion de transport le plus courant est le canal point-à-point exempt d'erreurs, qui délivre les messages ou les octets dans l'ordre où ils ont été envoyés. Il existe d'autres types de services de transport, tels que la remise de messages isolés sans garantie de l'ordre d'arrivée, ou la diffusion de messages à plusieurs destinations. Le type de service à assurer est déterminé lors de l'établissement de la connexion. (Par parenthèse, il est tout à fait impossible d'obtenir un canal totalement exempt d'erreurs : cette expression signifie en réalité que le taux d'erreur est suffisamment faible pour être ignoré en pratique).

La couche transport est une vraie couche de bout-en-bout : elle transporte les données tout du long, de la source à la destination. Autrement dit, un programme sur la machine source entretient une conversation avec un programme similaire sur la machine de destination, en utilisant les en-têtes des messages et des messages de

contrôle. Dans les couches plus basses, chaque protocole prend place entre une machine et ses voisins immédiats, non entre les machines source et de destination, qui peuvent être séparées par de nombreux routeurs. La différence entre les couches 1 à 3, qui sont chaînées, et les couches 4 à 7, qui sont de bout-en-bout, est illustrée à la figure 1.20.

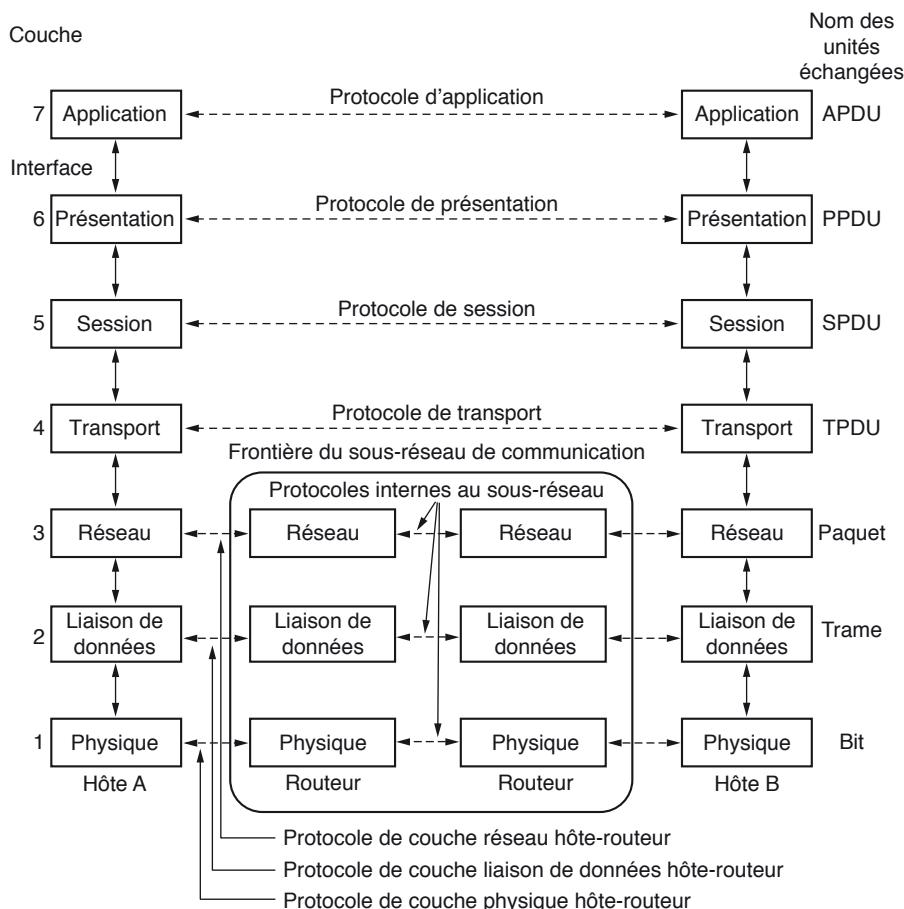


Figure 1.20 • Le modèle de référence OSI.

La couche session

La couche session permet aux utilisateurs de différentes machines d'établir des **sessions** entre eux. Une session offre divers services, parmi lesquels la **gestion du dialogue** (suivi du tour de transmission), la **gestion du jeton** (empêchant deux participants de tenter la même opération critique au même moment) et la **synchronisation**

(gestion de points de reprise permettant aux longues transmissions de reprendre là où elles en étaient, à la suite d'une interruption).

La couche présentation

À la différence des couches les plus basses, qui sont principalement concernées par le transport des bits, la **couche présentation** s'intéresse à la syntaxe et à la sémantique des informations transmises. Pour permettre à des ordinateurs utilisant des représentations de données internes différentes de communiquer, les structures de données à échanger peuvent être définies de façon abstraite et associées à un système de codage standard à utiliser « au fil de l'eau ». C'est la couche présentation qui gère ces structures de données abstraites et permet de définir et d'échanger des structures de données de plus haut niveau (par exemple des enregistrements bancaires).

La couche application

La **couche application** contient différents protocoles dont les utilisateurs ont couramment besoin. **HTTP** (*HyperText Transfer Protocol*), qui forme la base du World Wide Web, est un protocole d'application largement utilisé. Lorsqu'un navigateur veut afficher une page web, il transmet son nom au serveur qui l'héberge au moyen du protocole HTTP, et le serveur envoie la page en réponse. D'autres protocoles d'application sont utilisés pour le transfert de fichiers, le courrier électronique et les nouvelles (*news*).

1.4.2 Le modèle de référence TCP/IP

Venons-en maintenant au modèle de référence utilisé par l'ancêtre des réseaux étendus, l'ARPAnet, et par son célèbre descendant, l'Internet mondial. L'histoire de l'ARPAnet sera brièvement évoquée plus loin, mais il est utile d'en mentionner dès maintenant quelques aspects importants. À l'origine, il s'agissait d'un réseau destiné à la recherche, subventionné par le ministère de la Défense des États-Unis (DoD, *Department of Defense*). Il a fini par connecter des centaines de sites universitaires et gouvernementaux au moyen de liaisons téléphoniques louées. Lorsque les réseaux satellitaires et radio ont fait leur apparition, on a dû développer une nouvelle architecture de référence pour permettre aux protocoles existants d'interagir avec eux. Ainsi, dès le tout début, la possibilité d'interconnecter de nombreux réseaux de façon transparente a été l'un des objectifs de conception majeurs de l'ARPAnet. Cette architecture est devenue plus tard le **modèle de référence TCP/IP**, nommé d'après ses deux principaux protocoles.

Craignant que certains de ses précieux hôtes, routeurs et passerelles d'interconnexion ne soient soudainement neutralisés par une attaque de l'Union soviétique, le DoD avait un autre objectif majeur : faire en sorte que le réseau survive à la perte d'un équipement de sous-réseau, sans que les conversations existantes soient interrompues. Autrement dit, tant que les machines source et de destination étaient en état de fonctionner, les connexions existantes ne devaient pas être affectées par des équipements ou des lignes de transmission hors d'usage. Une architecture très souple

était également nécessaire, puisqu'il était prévu d'utiliser des applications aux exigences diverses, notamment pour le transfert de fichiers et la transmission de la parole en temps réel.

La couche liaison

Toutes ces exigences ont conduit au choix d'un réseau à commutation de paquets fondé sur une couche sans connexion transversale aux différents réseaux. La couche la plus basse de ce modèle, la **couche liaison**, décrit ce que les liens comme les lignes séries et les connexions Ethernet classiques doivent faire pour répondre aux besoins de cette interconnexion sans connexion. Il ne s'agit pas du tout d'une couche au sens normal du terme, mais plutôt d'une interface entre les hôtes et les liens de transmission. Les premiers documents sur le modèle TCP/IP n'étaient pas très bavards à son sujet.

La couche internet

La **couche internet** est l'axe central qui soutient toute l'architecture. La figure 1.21 la représente comme correspondant approximativement à la couche réseau du modèle OSI. Elle permet aux hôtes d'introduire des paquets sur n'importe quel réseau et fait en sorte qu'ils soient acheminés indépendamment les uns des autres vers leur destination (pouvant se trouver sur un réseau différent de celui d'origine). Il est même possible que les paquets arrivent dans un ordre complètement différent, auquel cas les couches supérieures se chargeront de les réordonner si cela fait partie des exigences de livraison. Notez que le terme « *internet* » est employé ici au sens générique d'interréseau, même si cette couche est présente dans l'Internet.

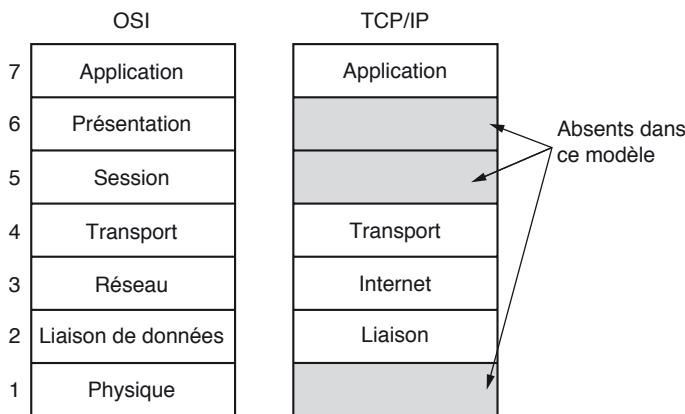


Figure 1.21 • Le modèle de référence TCP/IP.

On peut établir une analogie avec le courrier postal. Lorsqu'une personne poste plusieurs lettres pour l'étranger, il y a de fortes chances qu'elles arrivent à destination. Elles seront probablement traitées par un ou plusieurs centres de transit internationaux, mais cela reste transparent pour les utilisateurs, qui n'ont pas non plus à

se préoccuper du fait que chaque pays (c'est-à-dire chaque réseau) dispose de ses propres timbres, formats d'enveloppes et règles de distribution.

La couche internet définit un format de paquet officiel et un protocole nommé **IP** (*Internet Protocol*), plus un protocole compagnon, **ICMP** (*Internet Control Message Protocol*), qui l'aide à fonctionner. Son rôle est d'acheminer les paquets IP jusqu'à leur destination. Les deux préoccupations majeures à ce niveau sont donc le routage et les congestions (bien qu'IP ne se soit pas montré très efficace pour les éviter).

La couche transport

La couche directement supérieure à la couche internet dans le modèle TCP/IP se nomme aujourd'hui la **couche transport**. À l'instar de la couche transport du modèle OSI, son rôle est de permettre à des entités paires sur les hôtes source et de destination de mener une conversation. Deux protocoles de bout-en-bout ont été définis. Le premier, **TCP** (*Transmission Control Protocol*), est un protocole fiable avec connexion qui garantit la livraison sans erreur à n'importe quel hôte de l'interréseau d'un flot d'octets émis par une machine. Il segmente le flot d'octets entrant en messages discrets et transmet chacun d'eux à la couche internet. À l'arrivée, le processus TCP destinataire réassemble les messages reçus en un flot de sortie. TCP assure aussi un contrôle de flux pour éviter qu'un émetteur rapide ne submerge un récepteur lent de plus de messages qu'il ne peut en traiter.

Le second, **UDP** (*User Datagram Protocol*), est un protocole non fiable sans connexion qui permet aux applications d'assurer elles-mêmes le séquençage et le contrôle de flux au lieu de faire appel à TCP. Ce protocole est aussi largement utilisé par les interrogations ponctuelles de type demande-réponse dans les environnements client-serveur, et par les applications pour lesquelles la livraison des données à temps est plus importante que leur précision, comme dans la transmission de sons ou d'images. Les relations entre IP, TCP et UDP sont illustrées à la figure 1.22. Depuis que ce modèle a vu le jour, IP a été implémenté sur de nombreux autres réseaux.

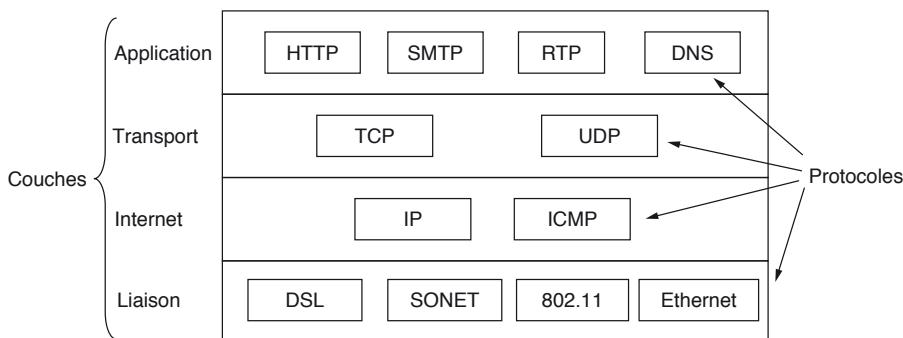


Figure 1.22 • Les protocoles du modèle TCP/IP.

La couche application

Le modèle TCP/IP ne comprend ni couche session, ni couche présentation, leur nécessité n'ayant pas été perçue. À la place, les applications incluent simplement les fonctions correspondantes si elles ont besoin. L'expérience du modèle OSI a confirmé la pertinence de ce choix : ces couches sont très peu utiles à la plupart des applications.

Directement au-dessus de la couche transport, on trouve la **couche application**, qui contient tous les protocoles de haut niveau. Les premiers à avoir été développés sont Telnet (protocole de terminal virtuel), FTP (protocole de transfert de fichiers) et SMTP (protocole d'échange de courrier électronique). Beaucoup d'autres protocoles, que nous étudierons plus tard, ont été ajoutés au fil des ans (voir figure 1.22). Citons notamment DNS (*Domain Name System*), pour associer des noms d'hôtes à leurs adresses réseau, HTTP (*HyperText Transfer Protocol*), pour transférer des pages web et RTP (*Real-Time Transfer Protocol*) pour diffuser en temps réel des médias comme la voix et la vidéo,

1.4.3 Le modèle utilisé dans ce livre

Comme nous l'avons vu, le point fort du modèle OSI est le *modèle* lui-même (moins les couches présentation et session), qui s'est montré exceptionnellement utile pour comprendre les réseaux d'ordinateurs. À l'inverse, la force du modèle TCP/IP réside dans les *protocoles*, qui sont utilisés depuis bien des années. Comme les informatiens aiment avoir le beurre et l'argent du beurre, nous adopterons pour cadre de référence de ce livre le modèle hybride de la figure 1.23.



Figure 1.23 • Le modèle de référence utilisé dans ce livre.

Ce modèle est organisé en cinq couches, allant de la couche physique à la couche application, en passant par les couches liaison, réseau et transport. La couche physique spécifie comment transmettre des bits sur différents types de support, sous forme de signaux électriques (ou d'autres formes de signaux analogiques). La couche liaison s'occupe de la façon d'envoyer des messages de longueur finie entre des ordinateurs directement connectés, avec des niveaux de fiabilité spécifiés. Ethernet et 802.11 sont des exemples de protocoles de cette couche.

La couche réseau permet de combiner plusieurs liens en réseaux, et les réseaux en interréseaux, afin de pouvoir transmettre des paquets entre ordinateurs

distsants. Cela nécessite de trouver des chemins sur lesquels transmettre les paquets. IP est le principal protocole que nous étudierons pour cette couche.

La couche transport renforce les garanties de livraison de la couche réseau, généralement avec une fiabilité accrue, et fournit certaines abstractions, comme un flot d'octets fiable, qui répondent aux besoins des différentes applications. TCP est un exemple important de protocole de la couche transport.

Enfin, la couche application contient des programmes qui utilisent le réseau. Beaucoup d'applications réseau, mais pas toutes, disposent d'interfaces utilisateur, par exemple les navigateurs web, mais nous nous intéresserons à la portion du programme qui utilise le réseau. Dans le cas du navigateur web, il s'agit du protocole HTTP. Elle contient également des programmes de support importants, comme le DNS, qui sont utilisés par de nombreuses applications.

L'ordonnancement des chapitres s'appuie sur ce modèle. De cette manière, nous conservons toute la valeur du modèle OSI pour comprendre les architectures de réseau, tout en nous concentrant avant tout sur les protocoles qui sont importants dans la pratique, de TCP/IP et des protocoles apparentés à d'autres plus récents, comme 802.11, SONET et Bluetooth.

1.4.4 Comparaison des modèles de référence OSI et TCP/IP

Les modèles de référence OSI et TCP/IP ont beaucoup de points communs. Tout d'abord, ils reposent tous deux sur le concept de pile de protocoles indépendants. Ensuite, leurs couches sont à peu près identiques sur le plan fonctionnel. Par exemple, les couches inférieures, couche transport incluse, ont pour rôle d'offrir un service de transport de bout-en-bout, indépendant du type de réseau, aux processus qui souhaitent communiquer. De même, les couches au-dessus de la couche transport sont des couches orientées applications utilisatrices du service de transport.

Malgré ces similitudes, les deux modèles présentent également de nombreuses différences, dont certaines fondamentales que nous examinerons dans cette section. Notez bien que nous comparons ici les *modèles de référence*, et non les *piles de protocoles* correspondantes. Il sera question des protocoles ultérieurement.

Trois concepts sont au cœur du modèle OSI :

1. Les services.
2. Les interfaces.
3. Les protocoles.

La plus grande contribution du modèle OSI a probablement été de rendre explicite la distinction entre ces trois concepts. Chaque couche rend des *services* à la couche située au-dessus d'elle. La définition du service indique ce que fait une couche, mais pas comment elle le fait ni comment les entités supérieures y accèdent. Le service spécifie la sémantique de la couche.

L'*interface* d'une couche indique aux processus situés au-dessus d'elle comment accéder à cette couche. Elle spécifie les paramètres à utiliser et les résultats à attendre. Elle non plus ne dit rien du fonctionnement interne de la couche.

Enfin, les *protocoles* pairs utilisés dans une couche ne regardent qu'elle. Elle peut utiliser celui qu'elle veut du moment qu'il remplit son rôle (qu'il fournit les services prévus). Elle peut également changer de protocole sans que cela affecte le fonctionnement des couches supérieures.

Ces idées s'accordent tout à fait avec celles plus modernes de la programmation orientée objet. Un objet, comme une couche, comprend un ensemble de méthodes (opérations) qui peuvent être invoquées par des processus extérieurs. La sémantique de ces méthodes définit l'ensemble des services offerts par l'objet. Les paramètres et les résultats des méthodes forment l'interface de l'objet. Le code interne à l'objet est son protocole et n'est pas visible du dehors.

Le modèle TCP/IP, quant à lui, n'a jamais clairement fait la distinction entre les services, les interfaces et les protocoles, en dépit de certaines tentatives pour le rapprocher du modèle OSI. Par exemple, les seuls services véritablement offerts par la couche internet sont l'envoi d'un paquet IP (**SEND IP PACKET**) et la réception d'un paquet IP (**RECEIVE IP PACKET**). De ce fait, les protocoles du modèle OSI sont mieux masqués que ceux du modèle TCP/IP et peuvent assez facilement être remplacés à mesure que les technologies évoluent. Cette capacité d'apporter des changements de manière transparente était d'ailleurs l'un des principaux effets recherchés avec les protocoles en couches.

Le modèle OSI ayant été élaboré *avant* les protocoles, il ne pouvait pencher en faveur d'aucun ensemble particulier de protocoles. Il présente ainsi l'avantage d'être très général. En revanche, l'inconvénient fut que les concepteurs ne possédaient pas une grande expérience du sujet et n'avaient pas une idée très claire de quelles fonctionnalités placer dans chaque couche.

Par exemple, la couche liaison de données ne pouvait gérer initialement que les réseaux point-à-point. Avec l'arrivée des réseaux à diffusion, une nouvelle sous-couche a dû être ajoutée au modèle. De plus, lorsque des ingénieurs ont entrepris de construire des réseaux en utilisant ce modèle et les protocoles existants, ils ont découvert qu'ils ne correspondaient pas aux spécifications de services requises. Il a donc fallu « greffer » des sous-couches de convergence au modèle pour remédier à ces différences. Enfin, les concepteurs du modèle OSI ayant pensé que chaque pays disposerait d'un réseau unique, exploité par une administration et utilisant les protocoles OSI, ils ne s'étaient pas vraiment intéressés à l'interconnexion des réseaux. Mais les choses ne se sont pas déroulées ainsi.

Avec TCP/IP, cela a été exactement l'inverse : les protocoles sont arrivés en premier, suivis du modèle qui ne faisait que les décrire. Les protocoles étaient en parfaite adéquation avec le modèle. Le seul problème était que le *modèle* n'était adapté à aucune autre pile de protocoles. Il n'était donc pas très utile pour décrire d'autres réseaux.

Pour en revenir à des considérations moins théoriques, une différence flagrante entre ces deux modèles est le nombre de couches : sept dans le modèle OSI et quatre dans le

modèle TCP/IP. Les couches réseau (internet pour TCP/IP), transport et application sont communes aux deux modèles, tandis que les autres couches diffèrent.

Une autre différence concerne le support des modes avec connexion et sans connexion. Le modèle OSI autorise les deux types de communication dans la couche réseau, mais uniquement le mode avec connexion dans la couche transport (ce qui est important car le service de transport est visible des utilisateurs). Le modèle TCP/IP ne prend en charge qu'un mode dans la couche réseau (sans connexion), mais deux dans la couche transport, ce qui donne le choix aux utilisateurs. Ce choix est particulièrement important pour les protocoles de type demande-réponse.

1.4.5 Critique du modèle et des protocoles OSI

Ni le modèle OSI et ses protocoles, ni le modèle TCP/IP et ses protocoles ne sont parfaits. Certaines critiques peuvent être, et ont été, soulevées à leur encontre. Nous en examinerons certaines aux sections suivantes, en commençant par OSI.

À l'époque où la deuxième édition de ce livre a été publiée (1989), il était clair pour de nombreux experts du domaine que le modèle OSI et ses protocoles allaient s'imposer et tout balayer sur leur passage. Cela ne s'est pas produit. Pourquoi ? Un retour sur le passé nous permettra de le comprendre, en explicitant les principales causes :

1. moment inopportun ;
2. mauvaise technologie ;
3. mauvaises implémentations ;
4. mauvaise stratégie.

Moment inopportun

Le moment auquel une norme s'impose est crucial pour son succès, comme le montre la théorie illustrée à la figure 1.24, émise par David Clark du MIT.

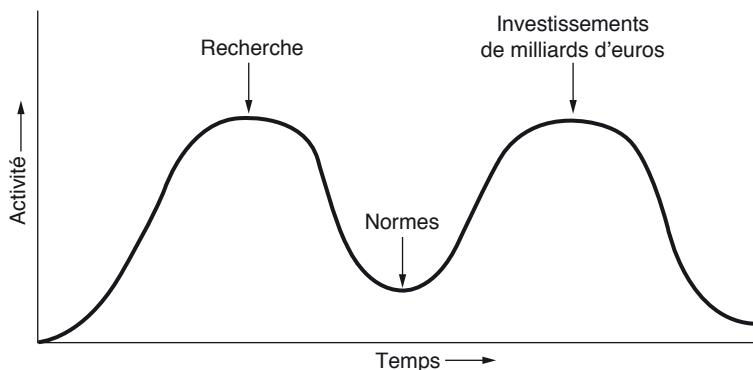


Figure 1.24 • Activités autour d'un sujet en fonction du temps.

Cette figure illustre l'activité qui entoure un nouveau sujet. Un sujet qui vient d'être découvert provoque une explosion de débats, de documents, de réunions, etc. Après un certain temps, l'activité se calme, les entreprises s'intéressent au sujet, et l'on assiste au déferlement de milliards d'investissements.

Il est essentiel qu'une norme s'impose dans la fenêtre de temps située entre les deux bosses. Si elle est publiée trop tôt (avant que les résultats des recherches ne soient bien établis), le sujet peut ne pas être suffisamment compris et la norme présentera des lacunes. Si elle est développée trop tardivement, les entreprises auront déjà tellement investi dans diverses façons de procéder qu'elle sera de fait ignorée. Si l'intervalle entre les deux courbes est très réduit (tout le monde ayant hâte de démarrer), les responsables de la rédaction des normes risquent d'être un peu bousculés.

C'est ce qui semble s'être produit pour les protocoles OSI. Les protocoles de la pile TCP/IP étaient déjà largement déployés au sein des universités de recherche lorsque ceux du modèle OSI sont apparus. Si la vague d'investissements n'avait pas encore eu lieu, le marché universitaire était suffisamment important pour que certains constructeurs commencent prudemment à proposer des produits TCP/IP. Lorsque le modèle OSI s'est présenté, ils ne voulaient pas prévoir de support pour une seconde pile de protocoles tant qu'ils n'y étaient pas forcés. Il n'y a donc pas eu d'offre initiale. Les entreprises ont pratiqué l'attentisme, aucune ne voulant prendre le risque d'être la première, et OSI a fait long feu.

Mauvaise technologie

La deuxième raison de l'absence d'adoption des protocoles OSI est que tant le modèle que les protocoles présentent des imperfections. Le choix des sept couches a été plus politique que technique, et deux d'entre elles – session et présentation – sont pratiquement vides, alors que deux autres – liaison de données et réseau – sont au contraire trop pleines.

Le modèle OSI, de même que les définitions de services et les protocoles qui l'accompagnent, est incroyablement complexe. Si l'on empile les documents constitutifs de la norme, on obtient pratiquement un mètre de papier. Leur implémentation est tout aussi difficile et leur fonctionnement inefficace. Il nous vient alors à l'esprit cette devinette de Paul Mockapetris :

Q. : Qu'obtient-on en croisant un gangster et une norme internationale ?

R : Quelqu'un qui vous fait une proposition que vous ne pouvez pas comprendre.

Outre le fait d'être incompréhensible, un autre problème avec OSI est que certaines fonctions, telles que l'adressage, le contrôle de flux et le contrôle d'erreurs, réapparaissent dans chaque couche. Des études ont montré que, par exemple, pour être efficace, le contrôle d'erreurs doit être réalisé dans la couche la plus haute, si bien que le répéter à chaque niveau est souvent inutile et inefficace.

Mauvaises implémentations

Étant donné l'énorme complexité du modèle et des protocoles, il n'est pas surprenant que les implémentations initiales aient été volumineuses, lourdes à manipuler et lentes. Tous ceux qui s'y sont frottés s'en sont rendu compte à leurs dépens. Dans ce contexte, il n'a pas fallu longtemps pour que OSI devienne synonyme de piètre qualité. Bien que les produits se soient améliorés au fil du temps, l'image a subsisté.

Par contraste, l'une des premières implémentations de TCP/IP, qui faisait partie de l'UNIX de Berkeley, était tout à fait réussie (et gratuite de surcroît). Elle a rapidement fait de nombreux adeptes, ce qui a conduit à des améliorations, puis à l'élargissement de la communauté d'utilisateurs, qui a entraîné elle-même de nouvelles améliorations, et le cercle vertueux s'est ainsi amorcé.

Mauvaise stratégie

Suite à la première implémentation, nombreux sont ceux, en particulier dans le milieu universitaire, à avoir pensé que TCP/IP faisait partie d'UNIX. Et le milieu universitaire des années 1980 voyait UNIX comme le *nec plus ultra*.

D'un autre côté, OSI a été vu comme le produit des administrations européennes des télécommunications, puis de la Communauté européenne, et plus tard du gouvernement des États-Unis. Cette croyance n'était que partiellement vraie, mais l'idée même qu'une bande de technocrates pouvait essayer d'imposer de force une norme technique inférieure aux pauvres chercheurs et programmeurs confrontés quotidiennement aux problèmes de conception de réseaux informatiques, n'a pas servi la cause du modèle OSI. Certains ont jeté sur ce développement le même regard que sur IBM, qui annonçait en 1960 que PL/I était le langage du futur, ou que sur le DoD, qui annonça ensuite pour corriger le tir que ce serait en fait Ada.

1.4.6 Critique du modèle de référence TCP/IP

Le modèle TCP/IP et ses protocoles ont aussi leurs problèmes. Premièrement, le modèle ne différencie pas clairement les concepts de services, d'interfaces et de protocoles. Les bonnes pratiques du génie logiciel ont montré qu'il était essentiel de séparer la spécification de l'implémentation, ce que le modèle OSI fait soigneusement, mais que le modèle TCP/IP ne fait pas. Par conséquent, on peut difficilement s'appuyer sur ce dernier pour concevoir de nouveaux réseaux mettant en œuvre de nouvelles technologies.

Deuxièmement, ce modèle n'étant pas du tout général, il ne permet pas réellement de décrire des piles de protocoles autres que TCP/IP. Tenter de s'en servir pour décrire Bluetooth, par exemple, est totalement impossible.

Troisièmement, la couche liaison n'est pas du tout une couche au sens où on l'entend habituellement dans le contexte des protocoles en couches, mais une interface (entre les couches réseau et liaison de données). Cette distinction entre une interface et une couche est essentielle, mais le modèle TCP/IP n'en tient pas compte.

Quatrièmement, le modèle TCP/IP ne distingue (ou n'évoque même) pas les couches physique et liaison de données qui sont pourtant complètement différentes. La couche physique est concernée par les caractéristiques de transmission du câble en cuivre, de la fibre optique et des canaux radio, tandis que la couche liaison de données est chargée de déterminer le début et la fin des trames et de les transmettre entre deux extrémités avec le degré de fiabilité souhaité. Un modèle correct devrait inclure ces deux couches en les séparant nettement, chose que ne fait pas le modèle TCP/IP.

Enfin, alors que les protocoles IP et TCP ont été soigneusement conçus et bien implémentés, le développement des autres protocoles de la suite a été improvisé et généralement confié à des doctorants. Leurs implémentations étant distribuées gratuitement, ces protocoles se sont largement imposés au point qu'il est encore difficile aujourd'hui de les remplacer. À titre d'exemple, le protocole de terminal virtuel, Telnet, a été conçu pour un terminal Télécopie mécanique à 10 caractères par seconde et ignore donc tout des interfaces graphiques et de la souris. Malgré cela, il est toujours largement utilisé trente ans plus tard.

1.5 Exemples de réseaux

L'interconnexion des ordinateurs donne lieu à de nombreux types de réseaux différents, plus ou moins étendus et plus ou moins bien connus. Chacun d'eux se distingue par l'objectif qu'il sert, par sa taille et par les technologies qu'il met en œuvre. Les sections suivantes examinent plusieurs exemples de réseaux afin de vous donner une idée de leur diversité.

Nous commencerons par le réseau le plus connu de tous, l'Internet, et nous présenterons son histoire, son évolution et son architecture. Nous verrons ensuite le réseau de téléphonie mobile, techniquement très différent de l'Internet. Nous présenterons ensuite IEEE 802.11, la norme dominante pour les réseaux locaux sans fil. Enfin, nous aborderons RFID et les réseaux de capteurs, deux technologies qui étendent la portée du réseau pour inclure le monde physique et les objets de la vie quotidienne.

1.5.1 L'Internet

L'Internet n'est pas véritablement un réseau, mais un immense regroupement de différents réseaux qui ont en commun certains protocoles et offrent certains services similaires. C'est un système inhabituel, au sens où il n'a pas été planifié par qui-conque et que personne ne le contrôle. Pour mieux le comprendre, voyons comment il s'est développé et pour quelles raisons.

ARPAnet

Tout a commencé à la fin des années 1950. Au plus fort de la guerre froide, le ministère de la Défense des États-Unis (DoD, *Department of Defense*) souhaitait disposer d'un réseau capable de résister à une attaque nucléaire. À cette époque, toutes les communications militaires empruntaient le réseau téléphonique public qui était considéré comme vulnérable. Pour comprendre les causes de cette vulnérabilité, examinons la

figure 1.25(a). Chaque point noir représente un commutateur téléphonique, auquel étaient reliés des milliers de téléphones. Ces commutateurs locaux étaient eux-mêmes connectés à des commutateurs centraux pour former une hiérarchie nationale, avec une redondance minimale. Le problème était que la destruction de quelques-uns seulement de ces centres pouvait conduire à l'éclatement du système, qui serait alors fragmenté en de nombreux îlots isolés.

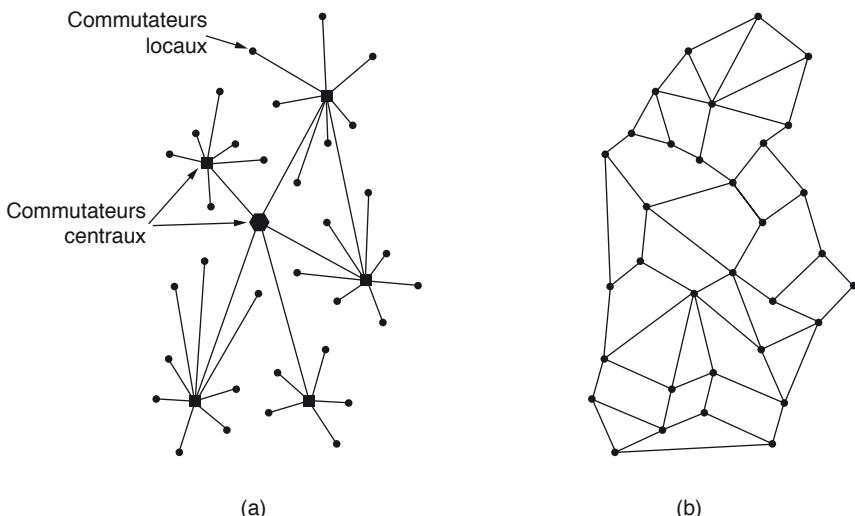


Figure 1.25 • (a) La structure du système téléphonique. (b) Le système de commutation distribué proposé par Baran.

Vers 1960, le DoD passa un contrat avec la RAND Corporation pour trouver une solution. L'un de ses employés, Paul Baran, imagina une architecture hautement distribuée et tolérante aux pannes, illustrée à la figure 1.25(b). La distance entre deux commutateurs étant trop importante pour que les signaux analogiques puissent parcourir sans distorsion, Baran proposa d'employer une technologie numérique de commutation de paquets. Il rédigea plusieurs rapports à l'intention du DoD, décrivant en détail son idée. Le Pentagone la trouva intéressante et demanda à AT&T, qui détenait à l'époque le monopole des communications téléphoniques sur le territoire des États-Unis, de développer un prototype. AT&T rejeta d'emblée le concept de Baran. Ce n'était pas un jeune employé qui allait apprendre à l'entreprise la plus grande et la plus riche du monde à concevoir un système téléphonique. Baran s'enthousiaqua finalement dire que son idée de réseau n'était pas réalisable, et elle fut donc abandonnée.

Plusieurs années s'écoulèrent sans que le système du DoD change. Pour comprendre ce qui advint par la suite, il faut revenir en octobre 1957, date à laquelle l'Union soviétique remporta une victoire sur les États-Unis dans la course vers l'espace, avec le lancement du premier satellite artificiel, Spoutnik. En tentant de déterminer

les causes de ce retard, le président Eisenhower fut consterné de découvrir que les armées de terre, de l'air et la marine, se disputaient le budget de recherche du Pentagone. Il réagit immédiatement avec la création d'une seule unité de recherche de la Défense, l'**ARPA** (*Advanced Research Projects Agency*). Cette agence n'employait aucun scientifique et ne possédait pas non plus de laboratoires : en fait, elle ne disposait que d'un bureau et d'un budget modeste (selon les standards du Pentagone). Sa tâche consistait à octroyer des subventions et des contrats aux universités et aux entreprises dont les projets lui semblaient prometteurs.

Au cours de ses premières années d'existence, l'ARPA tenta de mieux définir quelle devait être sa mission. En 1967, l'attention de Larry Roberts, un directeur de programme qui cherchait un moyen d'accéder à des ordinateurs à distance, se tourna vers les réseaux. Parmi les différents experts contactés se trouvait Wesley Clark qui suggéra de créer un sous-réseau à commutation de paquets, avec un routeur associé à chaque hôte (voir figure 1.10).

Après quelques hésitations, Roberts accepta l'idée et présenta un projet assez vague à ce sujet lors du symposium SIGOPS (*Special Interest Group on Operating Systems*) organisé par l'ACM (*Association for Computing Machinery*) à Gatlinburg dans le Tennessee, fin 1967. À sa grande surprise, une autre communication présentée à cette conférence décrivait un système similaire, qui avait été non seulement conçu mais encore entièrement implémenté au Royaume-Uni, sous la direction de Donald Davies du NPL (*National Physical Laboratory*), et qui se référait en outre au travail de Baran. Ce système n'était pas développé à l'échelle nationale (il connectait seulement quelques ordinateurs du campus du NPL), mais il démontrait la faisabilité d'un réseau à commutation de paquets. Roberts revint de Gatlinburg bien décidé à mettre en œuvre ce qui allait devenir l'**ARPAnet**.

Le sous-réseau serait formé de mini-ordinateurs appelés **IMP** (*Interface Message Processor*), reliés par des lignes de transmission à 56 kbit/s. Pour une fiabilité maximale, chaque IMP serait connecté au moins à deux autres IMP. Le sous-réseau opérerait en mode datagramme, afin de pouvoir rerouter automatiquement les messages sur d'autres chemins en cas de défaillance d'une ligne ou d'un IMP.

Chaque noeud du réseau consisterait en un IMP et un hôte, situés dans la même pièce et reliés par un câble court. L'hôte pourrait envoyer des messages de 8 063 bits maximum à son IMP, qui les diviserait alors en paquets de 1 008 bits maximum et les transmettrait indépendamment vers leur destination. Chaque paquet serait entièrement reçu avant d'être retransmis. Il s'agirait ainsi du premier réseau électronique à commutation de paquets en mode différé (*store-and-forward*).

L'ARPA lança donc un appel d'offres pour la construction du sous-réseau. Parmi les douze entreprises soumissionnaires, c'est la BBN, une société de conseil basée à Cambridge dans le Massachusetts, qui fut chargée en décembre 1968 de développer le sous-réseau et d'en écrire le logiciel. Elle choisit d'utiliser comme IMP des mini-ordinateurs Honeywell DDP-316 spécialement modifiés, avec une mémoire centrale de 12 K mots de 16 bits. Ils étaient dépourvus de disques, les composants mobiles étant considérés comme peu fiables, et étaient interconnectés au moyen de liaisons

à 56 kbit/s louées à des opérateurs de télécommunications, débit qui était ce qu'il y avait de mieux à l'époque.

Le logiciel fut divisé en deux parties : sous-réseau et hôte. La partie sous-réseau incluait l'extrémité IMP de la connexion hôte-IMP, le protocole IMP-IMP, et un protocole de l'IMP source vers l'IMP de destination prévu pour améliorer la fiabilité. La figure 1.26 présente le réseau ARPAnet à ses débuts.

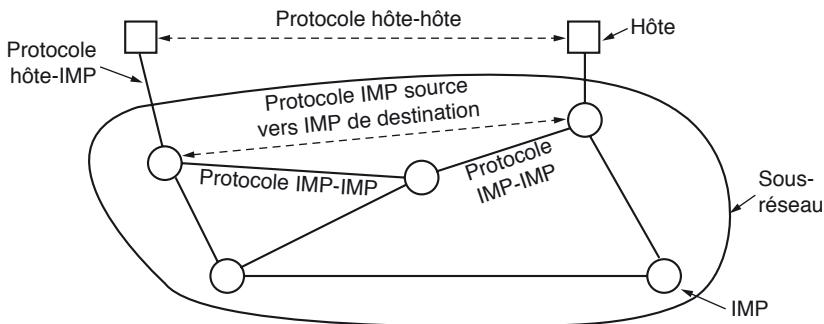


Figure 1.26 • Conception initiale du réseau ARPAnet.

En dehors du sous-réseau, il fallait également du code pour l'extrémité hôte de la connexion hôte-IMP, pour le protocole hôte-hôte et pour le logiciel applicatif. Il apparut rapidement que la BBN pensait que son rôle se bornait à faire en sorte qu'un message soit accepté sur la liaison hôte-IMP, puis placé sur la liaison hôte-IMP de destination.

Pour résoudre ce problème de logiciel du côté hôte, Larry Roberts (de l'ARPA) convia des chercheurs en réseau, pour la plupart des étudiants en troisième cycle, à se réunir à Snowbird, dans l'Utah, durant l'été 1969. Ceux-ci s'attendaient à ce qu'un expert leur décrive la conception du réseau et de son logiciel puis demandaient à chacun d'en implémenter une partie. À leur grande surprise, ils constatèrent qu'il n'y avait ni expert ni conception pour les guider, et qu'il leur faudrait déterminer eux-mêmes comment procéder.

C'est ainsi qu'un réseau expérimental vit le jour en décembre 1969, formé de quatre nœuds situés à l'université de Californie à Los Angeles (UCLA), à l'université de Californie à Santa Barbara (UCSB), au Stanford Research Institute (SRI) et à l'université d'Utah. Ces quatre institutions furent choisies en raison du grand nombre de contrats passés par chacune avec l'ARPA, mais aussi pour la diversité et l'incompatibilité de leurs ordinateurs respectifs (ce qui accentuait le défi). Le premier message d'hôte à hôte avait été envoyé deux mois plus tôt depuis le nœud de l'UCLA par l'équipe qu'animait Len Kleinrock (un pionnier de la commutation de paquets) au nœud du SRI. Le réseau se développa rapidement avec la livraison et l'installation de nouveaux IMP et finit par couvrir tous les États-Unis. La figure 1.27 illustre la rapide croissance de l'ARPAnet au cours de ses trois premières années.

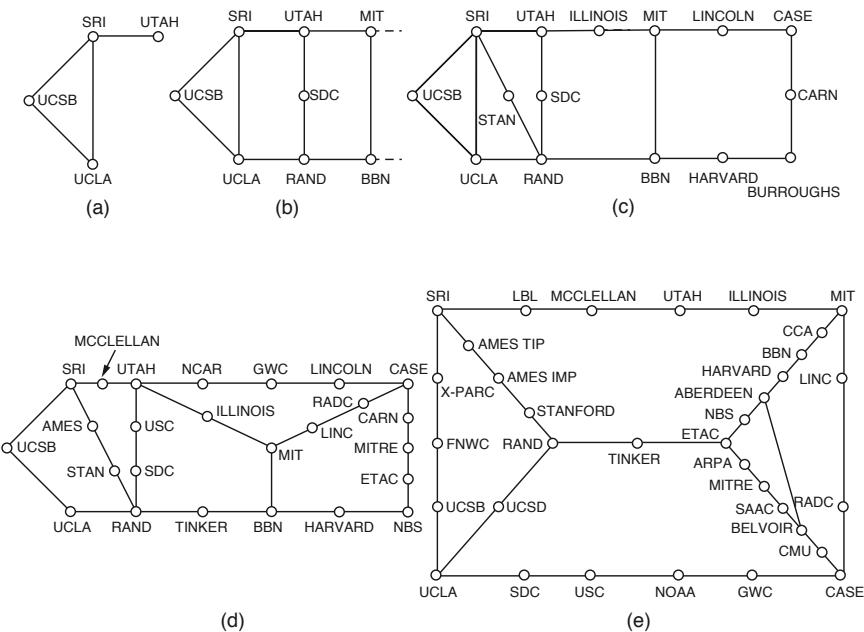


Figure 1.27 • Croissance d'ARPAnet. (a) Décembre 1969. (b) Juillet 1970. (c) Mars 1971. (d) Avril 1972. (e) Septembre 1972.

En plus de soutenir la croissance du tout jeune réseau ARPAnet, l'ARPA a aussi financé des recherches sur les réseaux satellitaires et les réseaux radio mobiles par paquets. Dans l'une de ses expériences devenue célèbre, un chercheur parcourant la Californie à bord d'un camion utilisait un réseau radio par paquets pour envoyer des messages au SRI, qui les transmettait à travers l'ARPAnet vers la côte est, après quoi ils étaient acheminés vers l'University College de Londres *via* un satellite. Le chercheur pouvait ainsi utiliser un ordinateur situé à Londres pendant que lui-même se déplaçait de l'autre côté de l'Atlantique.

Cette expérience a également démontré que les protocoles ARPAnet existants n'étaient pas adaptés aux transmissions entre réseaux différents. Cette observation a entraîné la poursuite des recherches sur les protocoles, conduisant à l'invention du modèle TCP/IP et de ses protocoles. Ce modèle a été conçu précisément pour gérer les communications sur des interréseaux, répondant à une nécessité croissante due au raccordement de réseaux toujours plus nombreux à l'ARPAnet.

Pour encourager l'adoption de ces nouveaux protocoles, l'ARPA attribua plusieurs contrats pour implémenter TCP/IP sur différentes plates-formes, notamment les systèmes IBM, DE CET HP, ainsi que pour l'UNIX de Berkeley. Des chercheurs de l'université de Californie à Berkeley réécrivirent TCP/IP avec une nouvelle interface de programmation, les **sockets** pour leur nouvelle version d'UNIX, BSD 4.2. Ils écrivirent également de nombreux programmes d'application, utilitaires et outils

de gestion pour montrer à quel point il était pratique d'utiliser le réseau avec des sockets.

Le moment était très bien choisi. En effet, bon nombre d'universités venaient d'acquérir un deuxième ou un troisième ordinateur VAX, ainsi qu'un LAN pour les relier, mais ne disposaient pas encore de logiciel de réseau. Le système UNIX BSD 4.2, avec TCP/IP, les sockets et tous ses utilitaires, fut adopté dès son apparition. En outre, grâce à TCP/IP, les universités pouvaient aisément connecter leur LAN à l'ARPAnet, ce qu'elles ne manquèrent pas de faire.

Au cours des années 1980, davantage de réseaux, surtout des LAN, furent raccordés à l'ARPAnet. Mais plus celui-ci s'étendait, plus il devenait coûteux de localiser ses hôtes. En réponse à ce problème, le système de gestion des noms de domaine, **DNS** (*Domain Name System*), fut développé de façon à organiser les machines en domaines et à associer des noms d'hôtes aux adresses IP. Depuis, DNS est devenu un système de base de données distribué généralisé stockant une grande variété d'informations relatives au nommage. Nous l'étudierons en détail au chapitre 7.

NSFNET

À la fin des années 1970, la NSF (*National Science Foundation*) constata à quel point l'impact de l'ARPAnet sur la recherche universitaire était important, puisqu'il permettait aux scientifiques de partager des données et de travailler sur des projets communs à travers tout le pays. Toutefois, pour pouvoir participer à ce réseau, une université devait être sous contrat de recherche avec le DoD, ce qui n'était pas le cas pour beaucoup d'entre elles. La réponse initiale de la NSF fut de financer le **CSNET** (*Computer Science Network*) en 1981. Celui-ci connectait des départements d'informatique et des laboratoires de recherche appliquée à l'ARPAnet *via* des connexions par modem et des lignes louées. À la fin des années 1980, la NSF alla plus loin et décida de construire un successeur de l'ARPAnet qui serait ouvert à toutes les instances de recherche universitaire.

La NSF décida donc de créer un réseau fédérateur pour connecter ses six centres de calcul basés à San Diego, Boulder, Champaign, Pittsburgh, Ithaca et Princeton. À chacun des six supercalculateurs était associé un micro-ordinateur LSI-11 ou **fuzzball**. Ces fuzzballs étaient reliés au moyen de lignes louées à 56 kbit/s et formaient le sous-réseau, à l'image de la technologie matérielle utilisée par l'ARPAnet. Toutefois, la technologie logicielle était différente : les fuzzballs supportaient dès le départ TCP/IP, faisant de ce sous-réseau le premier WAN TCP/IP.

La NSF finança également plusieurs réseaux régionaux (une vingtaine au total), qui furent reliés à l'épine dorsale pour permettre aux utilisateurs de milliers d'universités, de laboratoires de recherche, de bibliothèques et de musées d'accéder à n'importe quel supercalculateur et de communiquer entre eux. Le réseau complet, incluant l'épine dorsale et les réseaux régionaux, fut nommé **NSFNET** et raccordé à l'ARPAnet au moyen d'une liaison entre un IMP et un fuzzball dans la salle des machines de l'université Carnegie-Mellon. La figure 1.28 illustre l'épine dorsale d'origine, superposée à une carte des États-Unis.

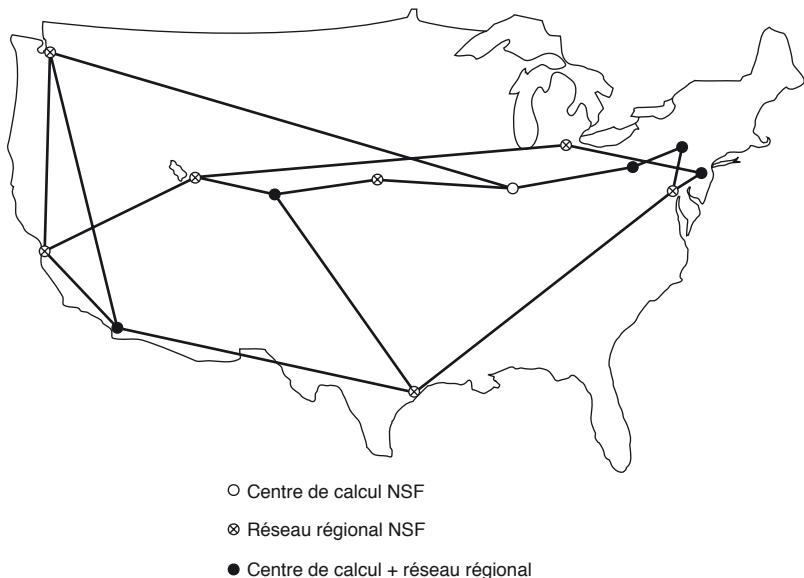


Figure 1.28 • L'épine dorsale du NSFNET en 1988.

Victime de son succès, NSFNET fut surchargé dès le début. La NSF ne tarda pas à réagir en chargeant le consortium MERIT dans le Michigan de développer le successeur de ce réseau. Des canaux en fibre optique à 448 kbit/s furent donc loués auprès de MCI (qui a depuis fusionné avec WorldCom) pour former la version 2 de l'épine dorsale, avec des PC-RT d'IBM pour routeurs. Ce réseau fut lui aussi rapidement saturé et, en 1990, son débit passa à 1,5 Mbit/s.

À mesure que l'expansion du réseau continuait, la NSF se rendit compte que le gouvernement ne pourrait pas la financer indéfiniment. En outre, les statuts de la fondation empêchaient les entreprises commerciales d'utiliser un réseau dont elle assurait seule le financement. En conséquence, elle encouragea MERIT, MCI et IBM à créer une organisation à but non lucratif, l'**ANS** (*Advanced Networks and Services*), qui serait la première étape sur la voie de la commercialisation. En 1990, l'**ANS** prit en charge le réseau NSFNET et fit passer son débit à 45 Mbit/s pour former le réseau **ANSNET**. Ce dernier fonctionna pendant cinq ans avant d'être cédé à America Online. Entre-temps, l'offre de services IP commerciaux s'était étoffée, et il était devenu évident que le gouvernement pouvait se retirer.

Pour faciliter la transition et s'assurer que tous les réseaux régionaux pourraient communiquer entre eux, la NSF signa des contrats avec quatre opérateurs différents en vue d'établir des points d'accès au réseau ou **NAP** (*Network Access Point*). Ces opérateurs étaient PacBell à San Francisco, Ameritech à Chicago, MFS à Washington D.C. et Sprint à New York (plus exactement à Pennsauken, dans le New Jersey).

Chaque opérateur devait se raccorder à tous les NAP pour pouvoir offrir des services d'épine dorsale aux réseaux régionaux. Cela signifiait qu'un paquet provenant de

n'importe lequel de ces réseaux avait le choix entre plusieurs opérateurs pour aller du NAP source vers le NAP de destination. Les opérateurs se disputèrent donc les marchés régionaux à coups d'offres de services et de prix, ce qui était bien entendu l'effet recherché. Le concept d'une seule épine dorsale par défaut fut ainsi remplacé par une infrastructure commerciale où la concurrence était vive. Nombreux sont ceux qui reprochent au gouvernement fédéral son manque d'innovation, mais il ne faut pas oublier que ce sont le DoD et la NSF qui ont ouvert la voie à l'Internet en créant une infrastructure de réseau dont ils ont ensuite confié l'exploitation à l'industrie.

Au cours des années 1990, de nombreux autres pays et régions du monde mirent en place des réseaux de recherche nationaux, souvent inspirés des modèles ARPAnet et NSFNET. EuropaNET et EBONE sont des exemples européens qui ont débuté à 2 Mbit/s pour atteindre 34 Mbit/s. En Europe, l'infrastructure de réseau a fini elle aussi par passer aux mains de l'industrie.

L'Internet a beaucoup changé depuis cette époque. Sa taille a explosé avec l'émergence du World Wide Web (WWW) à la fin des années 1990. Des données récentes de l'Internet Systems Consortium estiment le nombre d'hôtes présents sur l'Internet à plus de 600 millions. Ce chiffre est probablement sous-estimé, mais il dépasse de loin les quelques millions de machines existantes lorsque la première conférence sur le WWW s'est tenue au CERN en 1994.

La façon dont nous utilisons l'Internet a aussi changé radicalement. À l'origine, les applications dominantes étaient la messagerie pour les universitaires et les chercheurs, les groupes de discussion, la connexion à distance et le transfert de fichiers. Puis nous sommes passés au courrier électronique pour tout le monde, au Web et à la distribution de contenus P2P, comme Napster, désormais fermé. Actuellement, ce sont la distribution multimédia en temps réel, les réseaux sociaux (comme Facebook) et le « microblogging » (comme Twitter) qui prennent leur envol. Ces changements ont introduit des types de médias plus riches et génèrent donc beaucoup plus de trafic. En réalité, le trafic dominant semble changer régulièrement : par exemple, de nouvelles façons améliorées de traiter la musique ou les films pourraient se répandre très bientôt.

Architecture de l'Internet

L'architecture de l'Internet a aussi beaucoup changé, au fur et à mesure de son explosion. Au fil de cette section, nous entreprenons de donner un aperçu de son aspect actuel. Le tableau est un peu compliqué, du fait des bouleversements dus aux compagnies téléphoniques, aux câblo-opérateurs et aux fournisseurs d'accès, et il est parfois difficile de savoir qui fait quoi. L'une des raisons de ces bouleversements est la convergence des télécommunications, dans laquelle on emploie un réseau à d'autres usages que ceux pour lesquels il était prévu. Par exemple, dans une offre triple service (*triple play*), un opérateur vous vend des services de téléphonie, de télévision et d'accès à l'Internet sur la même connexion, arguant du fait que vous réaliserez des économies. Par conséquent, notre description sera nécessairement un peu plus simple que la réalité, et ce qui est vrai aujourd'hui pourrait ne plus l'être demain.

La figure 1.29 est une image globale de l'architecture de l'Internet. Étudions-la en détail, en commençant par l'ordinateur domestique (sur les bords de la figure). Pour se relier à l'Internet, l'ordinateur est connecté à un **fournisseur d'accès à l'Internet**, ou simplement **FAI**, auquel l'utilisateur achète de l'**accès à l'Internet**, autrement nommé **connectivité**. Cela lui permet d'échanger des paquets avec tous les autres hôtes accessibles sur le réseau. Qu'il envoie des paquets pour surfer sur le Web ou pour une autre des mille utilisations possibles n'a aucune importance. Il existe de nombreuses formes d'accès à l'Internet, que l'on distingue généralement en fonction de leur débit et de leur coût.

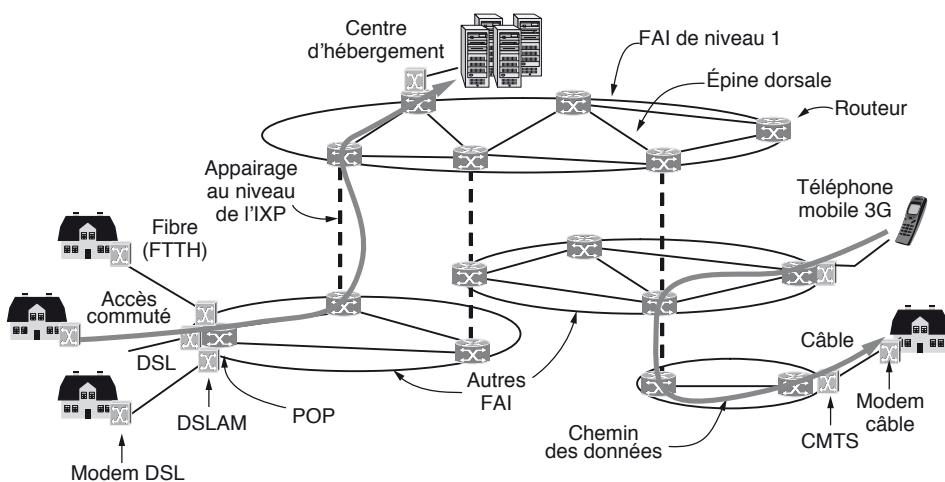


Figure 1.29 • Aperçu de l'architecture de l'Internet.

Une façon courante de se connecter à un FAI consiste à passer par votre ligne de téléphone. La technologie **DSL** (*Digital Subscriber Line*) réutilise cette ligne pour la transmission des données numériques. L'ordinateur est relié à un modem DSL, qui convertit les paquets numériques en signaux analogiques capables de circuler sans encombre sur la ligne. À l'autre bout, un équipement nommé **DSLAM** (*Digital Subscriber Line Access Multiplexer*) convertit les signaux en paquets. La figure 1.29 illustre d'autres types de connexion courants à un FAI. DSL est une technologie à haut débit qui utilise la ligne de téléphone locale pour transmettre des bits en lieu et place de la voix dans un appel traditionnel : il s'agit d'un accès communiqué (*dial-up*), réalisé au moyen d'un type de modem différent aux deux extrémités. Le mot **modem**, abréviation de « modulateur-démodulateur », désigne tout appareil qui effectue une conversion entre signaux numériques (bits) et signaux analogiques.

Une autre méthode consiste à transmettre des signaux *via* le système de télévision câblée. Comme DSL, c'est une façon de réutiliser l'infrastructure existante, en l'occurrence les canaux TV autrement inutilisés. L'équipement domestique est un modem câble et celui situé à la tête de réseau est le **CMTS** (*Cable Modem Termination*

System). Les technologies DSL et câble fournissent un accès à l'Internet à des débits allant d'une petite fraction de mégabit par seconde à plusieurs mégabits par seconde, selon le système. Ces débits sont très supérieurs à ceux de l'accès commuté traditionnel, qui sont limités à 64 kbit/s en raison de l'étroitesse de la bande passante servant aux appels vocaux. Cet accès à l'Internet à haut débit est qualifié parfois d'accès large bande, appellation qui renvoie à une bande passante plus large plutôt qu'à un débit particulier.

Les méthodes d'accès mentionnées jusqu'ici sont limitées par la bande passante du « dernier kilomètre », le dernier tronçon de ligne. Équiper les résidences en fibre optique peut permettre d'obtenir des débits de l'ordre de 10 à 100 Mbit/s. C'est la technologie nommée **FTTH** (*Fiber to the Home*). Les entreprises situées dans des zones d'activité importantes peuvent préférer des lignes louées rapides entre leurs bureaux et le FAI le plus proche. Par exemple, en Amérique du Nord, une ligne T3 fonctionne à environ 45 Mbit/s.

Les technologies sans fil permettent également d'accéder à l'Internet. Nous verrons plus loin l'exemple des réseaux de téléphonie mobile 3G. Ceux-ci offrent des services de données à des débits de 1 Mbit/s ou plus aux abonnés mobiles et fixes dans leur zone de couverture.

Il est maintenant possible de transférer des données entre une résidence et un FAI. On appelle l'endroit où les paquets du client pénètrent dans le réseau du FAI le **point de présence**, ou **POP** (*Point of Presence*), du FAI. Nous expliquerons plus loin comment les paquets sont transférés entre les points de présence de différents FAI. À partir de ce point, c'est un système à commutation de paquets entièrement numérique.

Les réseaux de FAI peuvent être régionaux, nationaux ou internationaux. Nous avons déjà vu que leur architecture est constituée de lignes de transmission longue distance qui interconnectent les routeurs au niveau des points d'accès dans les différentes villes que le FAI dessert. On appelle cette installation le **réseau fédérateur** (*backbone*) du FAI. Si un paquet est destiné à un hôte desservi directement par le FAI, il est routé sur le réseau fédérateur et remis à l'hôte. Sinon, il doit être transmis à un autre FAI.

Pour échanger du trafic, les FAI connectent leurs réseaux à des **points d'échange Internet** ou **IXP** (*Internet eXchange Point*). On dit alors que les FAI sont appairés (*peered*). Il existe de nombreux IXP dans différentes villes du monde. La figure 1.29 les représente verticalement, parce que, géographiquement, les réseaux de FAI se chevauchent. En substance, un IXP est une pièce pleine de routeurs, et il en existe au moins une par FAI. Comme un LAN connecte tous les routeurs de cette pièce, les paquets peuvent être transmis entre deux réseaux fédérateurs de FAI quelconques. Les IXP peuvent être des installations de grande envergure détenues de façon indépendante. L'une des plus grandes est l'Amsterdam Internet Exchange, auquel des centaines de FAI sont connectés et par l'intermédiaire duquel ils échangent des centaines de gigabits par seconde.

L'appairage (*peering*) qui a lieu au niveau de l'IXP dépend des relations commerciales entre les FAI, qui peuvent prendre de nombreuses formes. Par exemple, un petit FAI peut acheter de la connectivité à un plus grand pour atteindre des hôtes distants, un

peu comme un client achète les services d'un fournisseur Internet. Dans ce cas, on dit que le petit FAI paie pour le transit. Ou bien, deux grands FAI peuvent décider d'échanger du trafic directement, de façon à éviter de payer pour le transit. C'est là l'un des nombreux paradoxes de l'Internet : des FAI qui se disputent les clients en public coopèrent souvent en privé pour l'appairage.

Le chemin qu'emprunte un paquet à travers l'Internet dépend des choix d'appairage des FAI. Si le FAI qui émet le paquet est appairé avec le FAI de destination, il peut le lui remettre directement. Sinon, il peut router le paquet vers le prochain point auquel il est relié à un fournisseur de transit payant, pour que celui-ci puisse le livrer. La figure 1.29 illustre deux exemples de chemins. Souvent, le chemin emprunté n'est pas le plus court.

Au sommet de la chaîne, on trouve une poignée de grands opérateurs, comme AT&T et France Télécom, qui exploitent de grands réseaux fédérateurs internationaux comprenant des milliers de routeurs connectés par des liens en fibre optique à haut débit. Ces FAI ne paient pas le transit. On les appelle généralement des **FAI de niveau 1**, et l'on dit qu'ils forment l'épine dorsale de l'Internet, puisque tous les autres doivent se connecter à eux pour atteindre le reste du réseau.

Les sociétés qui fournissent beaucoup de contenus, comme Google et Yahoo!, placent leurs ordinateurs dans des **centres d'hébergement** (*data centers*) qui sont bien connectés au reste de l'Internet. Ces centres sont conçus pour les ordinateurs, non pour les humains, et abritent des machines empilées en racks qui constituent des **fermes de serveurs**. Les centres d'hébergement ou de **clocation** permettent aux clients de placer des équipements comme des serveurs au niveau des points de présence des FAI, afin de pouvoir établir des connexions courtes et rapides entre leurs serveurs et les réseaux fédérateurs des FAI. L'industrie de l'hébergement Internet se « virtualise » de plus en plus, si bien qu'il est désormais courant de louer une machine virtuelle qui s'exécute sur une ferme de serveurs au lieu d'installer un ordinateur physique. La taille de ces centres est telle (ils peuvent contenir des dizaines ou des centaines de machines) que l'alimentation électrique représente un coût majeur, et on les construit parfois dans des zones où le courant est bon marché.

Cela termine notre brève présentation de l'Internet. Au cours des prochains chapitres, vous en apprendrez beaucoup plus sur ses différents composants, ainsi que sur leur conception, leurs algorithmes et leurs protocoles. Un autre point utile à mentionner ici est que la signification de « être sur l'Internet » est en train de changer. À l'origine, une machine « était sur l'Internet » si (1) elle exécutait la pile de protocoles TCP/IP, (2) elle possédait une adresse IP et (3) elle pouvait envoyer des paquets IP à toutes les autres machines sur l'Internet. Toutefois, les FAI réemploient souvent les adresses IP selon les ordinateurs utilisés à un moment donné, et les réseaux domestiques partagent souvent une adresse entre plusieurs ordinateurs. Cette pratique contredit la deuxième condition. En outre, des mesures de sécurité comme les pare-feu peuvent empêcher des ordinateurs de recevoir des paquets, ce qui contredit la troisième condition. Malgré ces difficultés, il est logique de considérer que ces machines sont sur l'Internet quand elles sont connectées à leur FAI.

Il peut aussi être utile de mentionner en passant que certaines entreprises ont interconnecté tous leurs réseaux internes existants, souvent en utilisant la même technologie que celle de l'Internet, formant ainsi des **intranets**. Ceux-ci ne sont généralement accessibles que dans les locaux de l'entreprise ou à partir de portables lui appartenant, mais fonctionnent néanmoins de la même manière que l'Internet.

1.5.2 Réseaux de téléphonie mobile de troisième génération

Les gens aiment encore plus parler au téléphone que surfer sur l'Internet, et c'est pourquoi le réseau de téléphonie mobile est celui qui rencontre le plus de succès dans le monde entier, avec plus de quatre milliards d'abonnés. Pour mettre ce chiffre en perspective, sachez qu'il représente environ 60 % de la population mondiale, et plus que le nombre d'hôtes de l'Internet additionné au nombre de lignes téléphoniques fixes.

L'architecture de ces réseaux a énormément changé ces quarante dernières années, à mesure qu'ils se développaient de manière vertigineuse. Les systèmes de première génération transmettaient les appels vocaux sous forme de signaux (analogiques) variant continûment et non de séquences de bits (numériques). **AMPS** (*Advanced Mobile Phone System*), déployé aux États-Unis en 1982, était un système de première génération, tout comme Radiocom 2000 lancé en France en 1986. Puis les systèmes de deuxième génération passèrent à la transmission numérique des appels, pour augmenter la capacité, améliorer la sécurité et offrir la messagerie texte. **GSM** (*Global System for Mobile communications*), déployé à partir de 1991 et qui est devenu le système de téléphonie mobile le plus utilisé au monde, est un système de deuxième génération, ou 2G.

Les systèmes de troisième génération, ou 3G, ont été déployés à partir de 2001 et offrent à la fois des services numériques de voix et de données à haut débit. Ils s'accompagnent aussi d'un jargon considérable et de nombreux standards différents parmi lesquels il faut choisir. 3G est défini de façon assez vague par l'UIT (un organisme international de normalisation que nous verrons à la prochaine section) comme un système fournissant des débits d'au moins 2 Mbit/s aux utilisateurs stationnaires ou nomades et de 384 kbit/s dans un véhicule en mouvement. **UMTS** (*Universal Mobile Telecommunications System*), également appelé **WCDMA** (*Wideband Code Division Multiple Access*), est le principal système 3G, en cours de déploiement rapide dans le monde entier. Il offre jusqu'à 14 Mbit/s en liaison descendante et près de 6 Mbit/s en liaison montante. Les futures versions utiliseront plusieurs antennes pour fournir des débits encore plus élevés.

La ressource rare dans les systèmes 3G, comme dans les systèmes 2G et 1G avant eux, c'est le spectre radio. Les gouvernements accordent aux opérateurs de téléphonie mobile le droit d'en utiliser une partie, souvent au moyen d'un système d'enchères. Une telle pratique facilite la conception et la gestion d'un réseau, puisque personne d'autre n'est autorisé à émettre sur cette portion du spectre, mais elle est parfois très coûteuse. Au Royaume-Uni par exemple, cinq licences 3G ont été vendues aux enchères en 2000, pour un total d'environ 40 milliards de dollars.

C'est cette rareté qui a conduit à la solution du **réseau cellulaire**, illustré à la figure 1.30, maintenant utilisé pour la téléphonie mobile. Pour gérer les interférences radio entre les utilisateurs, la zone de couverture est divisée en cellules. À l'intérieur d'une cellule, les canaux affectés aux utilisateurs n'interfèrent pas entre eux et ne causent pas trop d'interférences avec les cellules adjacentes. Cette technique permet une bonne réutilisation du spectre, ou **réutilisation des fréquences**, dans les cellules voisines, ce qui augmente la capacité du réseau. Dans les systèmes 1G, qui transportaient chaque appel vocal sur une bande de fréquences spécifique, ces fréquences étaient soigneusement choisies afin d'éviter les conflits avec les cellules voisines. De cette manière, une fréquence donnée ne pouvait être réutilisée qu'une fois dans plusieurs cellules. Les systèmes 3G modernes permettent à chaque cellule d'utiliser toutes les fréquences, mais de telle sorte que le niveau d'interférence résultant soit tolérable pour les cellules voisines. Il existe différentes variantes de la conception cellulaire, avec notamment l'emploi d'antennes directionnelles ou sectorielles sur les stations de base pour réduire encore les interférences, mais le principe demeure le même.

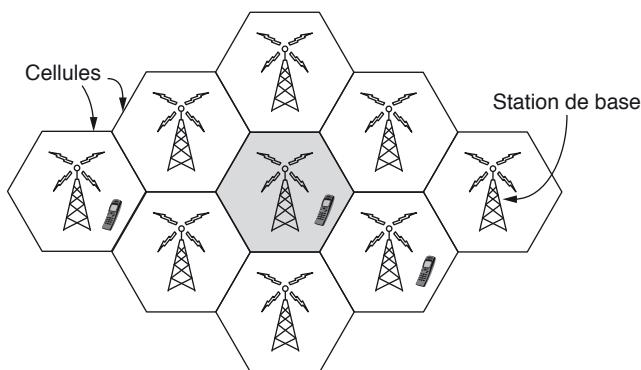


Figure 1.30 • Conception cellulaire des réseaux de téléphonie mobile.

L'architecture du réseau de téléphonie mobile est très différente de celle de l'Internet. Elle se compose de plusieurs parties, comme le montre la version simplifiée de l'architecture UMTS représentée à la figure 1.31. Tout d'abord, il y a l'**interface radio**, terme utilisé pour désigner le protocole de radiocommunication utilisé entre un équipement mobile (un téléphone par exemple) et une **station de base**. Les progrès réalisés en la matière ces dernières décennies ont considérablement augmenté les débits des réseaux sans fil. L'interface radio du système UMTS est basée sur la technologie **CDMA** (*Code Division Multiple Access*), que nous étudierons au chapitre 2.

La station de base cellulaire et son contrôleur forment le **réseau d'accès radio**. Cette partie est le côté sans fil du réseau de téléphonie mobile. Le nœud contrôleur, ou **RNC** (*Radio Network Controller*), contrôle la façon dont le spectre est utilisé. La station de base met en œuvre l'interface radio. On l'appelle le **nœud B**, une désignation temporaire qui a perduré.

Le reste du réseau de téléphonie mobile transporte le trafic pour le réseau d'accès radio. On l'appelle le **cœur de réseau**. Le cœur de réseau d'UMTS a évolué à partir de celui du système GSM de deuxième génération qui l'a précédé. Toutefois, il s'y passe quelque chose de surprenant.

Depuis que les réseaux existent, les partisans des réseaux en mode paquet (sous-réseaux sans connexion) s'opposent aux adeptes des réseaux en mode circuit (sous-réseaux avec connexion). Les principaux tenants des paquets sont issus de la communauté Internet. Dans le mode sans connexion, chaque paquet est routé indépendamment des autres. En conséquence, en cas de défaillance d'un routeur, aucun dommage n'est subi tant que le système est capable de se reconfigurer lui-même dynamiquement et de trouver une autre route pour acheminer le reste des paquets vers leur destination.

Le camp favorable au mode circuit est issu du monde de la téléphonie. Dans le système téléphonique, un appelant doit composer le numéro de son interlocuteur et attendre d'être mis en relation avec lui avant de pouvoir parler ou envoyer des données. Le processus d'établissement de connexion détermine une route au travers du réseau qui est maintenue le temps de l'appel. Tous les mots ou tous les paquets suivent le même chemin. Si une ligne ou un commutateur sur ce chemin connaît une défaillance, la communication est interrompue : le réseau est donc moins tolérant aux pannes qu'un réseau sans connexion.

L'avantage des réseaux de circuits est qu'ils prennent en charge la qualité de service plus facilement. En établissant une connexion à l'avance, le sous-réseau peut réservier des ressources comme de la bande passante, de l'espace tampon et de la puissance processeur. Si une tentative d'appel a lieu et que les ressources disponibles soient insuffisantes, la demande est rejetée et l'appelant reçoit un signal d'occupation. De cette façon, lorsqu'une connexion est établie, elle est assurée de bénéficier d'un service de qualité.

Sur un réseau en mode sans connexion, un routeur qui reçoit un trop grand nombre de paquets à la fois se bloque et peut en perdre. L'émetteur s'en apercevra tôt ou tard et enverra de nouveau les paquets, mais la qualité de service sera médiocre et inadaptée à la transmission d'informations audio ou vidéo, à moins que le réseau ne soit très peu chargé. Inutile de préciser que les opérateurs de télécommunications se soucient tout particulièrement de la qualité audio, ce qui explique leur préférence pour le mode avec connexion.

La surprise de la figure 1.31, c'est que le réseau cœur contient à la fois des équipements de commutation de paquets et de circuits. Elle représente le réseau de téléphonie mobile en transition, avec des opérateurs pouvant mettre en œuvre l'une des deux solutions, ou parfois les deux. Les anciens réseaux utilisaient un cœur à commutation de circuits, dans le style du réseau téléphonique traditionnel, pour transporter les appels vocaux. Cet héritage est perceptible dans le réseau UMTS, avec le centre de commutation mobile ou **MSC** (*Mobile Switching Center*), le centre de commutation mobile de transit ou **GMSC** (*Gateway Mobile Switching Center*) et la passerelle média ou **MGW** (*Media GateWay*), qui établissent des connexions sur un réseau cœur à commutation de circuits comme le **réseau téléphonique public commuté (RTC)**.

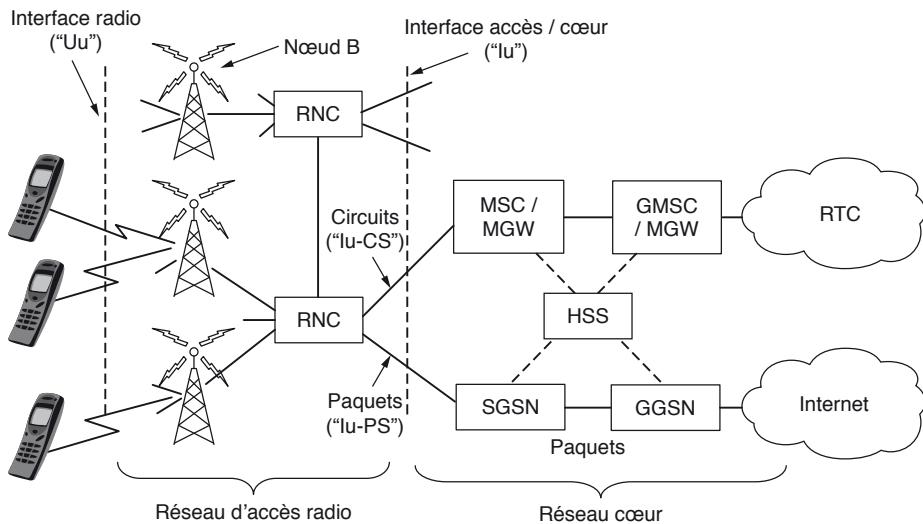


Figure 1.31 • Architecture du réseau de téléphonie mobile UMTS 3G.

Les services de données ont pris une place beaucoup plus importante qu'auparavant dans le réseau de téléphonie mobile, en commençant par la messagerie texte et les premiers services de transmission par paquets comme le **GPRS** (*General Packet Radio Service*) dans le système GSM. Ces anciens services avaient des débits de plusieurs kilobits par seconde, mais les utilisateurs en voulaient plus. Les réseaux plus récents transportent des paquets à plusieurs mégabits par seconde. Pour la comparaison, le débit d'un appel vocal est de 64 kbit/s, généralement de trois à quatre fois plus avec compression.

Pour transporter toutes ces données, les nœuds du réseau cœur UMTS sont directement connectés à un réseau à commutation de paquets. Le nœud de support du GPRS de desserte, ou **SGSN** (*Serving GPRS Support Node*), et le nœud de support du GPRS passerelle, ou **GGSN** (*Gateway GPRS Support Node*), transmettent les paquets de données entre les mobiles et ont une interface avec des réseaux de paquets externes, comme l'Internet.

Cette transition est vouée à se poursuivre dans les réseaux de téléphonie mobile actuellement planifiés et déployés. Des protocoles internet sont même utilisés sur les mobiles afin d'établir des connexions pour les appels vocaux sur un réseau de données en mode paquet, à la manière de VoIP. IP et les paquets sont utilisés tout du long, depuis le réseau d'accès radio jusqu'au réseau cœur. Bien entendu, la façon dont les réseaux IP sont conçus change également, pour offrir une meilleure qualité de service. Dans le cas contraire, les sons hachés et les vidéos saccadées n'impressionneraient pas les clients payants. Nous reviendrons sur ce sujet au chapitre 5.

Une autre différence entre les réseaux de téléphonie mobile et l'Internet traditionnel est la mobilité. Lorsqu'un utilisateur sort de la portée d'une station de base et entre dans celle d'une autre, le flux de données doit être rerouté de l'ancienne station vers

la nouvelle. Cette technique, connue sous le nom de **transfert intercellulaire automatique** ou **handover** ou encore **handoff**, est illustrée à la figure 1.32.



Figure 1.32 • Transfert intercellulaire pour un téléphone mobile (a) avant, (b) après.

Soit l'équipement mobile, soit la station de base peut demander un transfert intercellulaire quand la qualité du signal baisse. Dans certains réseaux cellulaires, habituellement ceux basés sur la technologie CDMA, il est possible de se connecter à la nouvelle station de base avant de se déconnecter de l'ancienne. La qualité de la connexion s'en trouve améliorée pour le mobile, parce qu'il n'y a pas d'interruption dans le service : l'appareil est réellement connecté aux deux stations pendant un bref instant. On parle alors de **soft handover**, par opposition au **hard handover**, dans lequel le mobile se déconnecte de l'ancienne station de base avant de se connecter à la nouvelle.

Une question apparentée est celle de savoir comment trouver un mobile lorsqu'il y a un appel entrant. Chaque réseau de téléphonie mobile possède un **serveur d'abonné résidentiel**, ou **HSS** (*Home Subscriber Server*), situé dans le réseau cœur, qui connaît l'emplacement de chaque abonné, ainsi que d'autres éléments de profil utilisés pour l'authentification et l'autorisation. De cette manière, il est possible de repérer chaque mobile en contactant le HSS.

Enfin, une dernière question concerne la sécurité. Historiquement, les opérateurs téléphoniques ont pris la sécurité beaucoup plus au sérieux que ceux de l'Internet, en raison de la nécessité de facturer le service et d'éviter les fraudes (au paiement). Malheureusement, cela ne veut pas dire grand-chose. Néanmoins, dans le passage des technologies 1G à 3G, les compagnies ont pu déployer certains mécanismes de sécurité de base pour les mobiles.

Depuis le système GSM, un téléphone mobile se compose de l'appareil lui-même et d'une puce amovible contenant l'identité de l'abonné et d'autres informations sur son compte. On appelle familièrement cette puce une **carte SIM** (*Subscriber Identity Module*). On peut l'insérer dans différents appareils pour les activer, et elles fournissent une base pour la sécurité. Les clients GSM en voyage d'agrément ou d'affaires emportent souvent leur téléphone, mais ils achètent en arrivant une nouvelle carte SIM pour une somme modique, afin de pouvoir passer des appels locaux sans facturation supplémentaire. Pour réduire les possibilités de fraude, les informations contenues dans la carte SIM sont également utilisées par le téléphone mobile pour

authentifier l'abonné et vérifier qu'il est autorisé à accéder au réseau. Dans le système UMTS, le mobile utilise aussi les informations de la carte SIM pour vérifier qu'il converse avec un réseau légitime.

Un autre aspect de la sécurité est la confidentialité. Comme les signaux sans fil sont diffusés à tous les récepteurs situés à proximité, des clés cryptographiques intégrées à la carte SIM servent à chiffrer les transmissions pour empêcher les indiscrets d'épier les conversations. Cette approche procure une confidentialité bien supérieure à celle des systèmes 1G, qui étaient faciles à pirater, mais elle n'est pas une panacée, en raison des faiblesses des algorithmes de chiffrement.

Les réseaux de téléphonie mobile sont appelés à jouer un rôle central dans les futurs réseaux. L'intérêt se porte désormais plus sur les applications mobiles à haut débit que sur les appels vocaux, et cela a des conséquences majeures pour les interfaces radio, l'architecture des réseaux cœur et la sécurité. Des technologies 4G, plus rapides et meilleures, sont à l'étude sous le nom de **LTE** (*Long Term Evolution*), même si la conception et le déploiement de réseaux 3G continuent. D'autres technologies sans fil offrent également un accès Internet à haut débit aux clients fixes et mobiles, notamment les réseaux 802.16, plus couramment appelés **WiMAX**. Il est tout à fait possible que LTE et WiMAX entrent en collision, et il est difficile de prédire comment les deux systèmes vont évoluer.

1.5.3 LAN sans fil : 802.11

Dès l'apparition des premiers ordinateurs portables, nombreux sont ceux qui s'imaginaient déjà pouvoir pénétrer dans un bureau et être instantanément connectés à l'Internet. Divers groupes de travail ont donc entrepris de rechercher des moyens d'atteindre cet objectif. La méthode la plus pratique a consisté à équiper les ordinateurs fixes et portables d'émetteurs et de récepteurs radio de faible portée pour leur permettre de communiquer.

De nombreuses sociétés se sont alors lancées dans la commercialisation de produits sans fil, sans se soucier des problèmes de compatibilité. La prolifération des standards signifiait qu'un ordinateur équipé d'un système radio de marque X ne pouvait pas fonctionner comme prévu dans un bureau où une station de base de marque Y avait été installée. Au milieu des années 1990, l'industrie décida finalement qu'il était temps de définir une norme unique de LAN sans fil, et la tâche fut confiée au comité de l'IEEE qui s'occupait déjà de la normalisation des LAN filaires.

La première décision était la plus facile : comment l'appeler. Puisque toutes les autres normes portaient des numéros comme 802.1, 802.2 et 802.3 jusqu'à 802.10, on la baptisa 802.11. On l'appelle familièrement **Wi-Fi** (*Wireless Fidelity*), mais comme c'est une norme importante qui mérite le respect, nous lui donnerons son nom officiel, 802.11.

Le reste fut plus difficile. Le premier problème était de trouver une bande de fréquences adéquate et disponible, de préférence dans le monde entier. L'approche adoptée fut l'inverse de celle des réseaux de téléphonie mobile. Au lieu d'utiliser de coûteuses portions du spectre sous licence, les systèmes 802.11 opèrent dans

des bandes non soumises à licence, comme les bandes **ISM** (industriel, scientifique et médical) définies par l'UIT-R (par exemple 902-928 MHz, 2,4-2,5 GHz, 5,725-5,825 GHz). Tous les équipements sont autorisés à utiliser ce spectre, pourvu qu'ils limitent leur puissance d'émission afin de laisser les autres coexister. Bien entendu, cela signifie que les radios 802.11 peuvent se trouver en conflit avec des téléphones sans fil, des commandes d'ouverture de porte de garage et des fours à micro-ondes.

Les réseaux 802.11 sont constitués de clients, comme des ordinateurs portables et des téléphones mobiles, et d'équipements d'infrastructure installés dans les bâtiments, les **points d'accès** ou AP (*Access Points*), parfois appelés **stations de base**. Ces derniers sont connectés au réseau filaire, et toutes les communications entre les clients passent par l'un d'eux. Des clients à portée radio les uns des autres, par exemple deux ordinateurs situés dans un même bureau, peuvent communiquer sans point d'accès : c'est ce que l'on appelle un **réseau ad hoc**. Ce mode est utilisé beaucoup moins souvent que le mode infrastructure. La figure 1.33 illustre les deux.

La transmission 802.11 est compliquée du fait des conditions qui varient au moindre changement dans l'environnement sans fil. Aux fréquences utilisées par 802.11, les signaux radio peuvent être réfléchis par des objets solides, si bien que plusieurs échos d'une émission peuvent atteindre un récepteur en empruntant des chemins différents. Ces échos peuvent s'annuler ou se renforcer mutuellement, causant une fluctuation considérable du signal. Ce phénomène, appelé **atténuation due aux trajets multiples** (*multipath fading*), est illustré à la figure 1.34.

La principale solution pour surmonter ces conditions variables est la **diversité de trajets** (*path diversity*), autrement dit l'envoi d'informations sur plusieurs chemins indépendants. De cette manière, ces informations devraient être reçues, même si l'un des chemins est impraticable en raison d'une trop forte atténuation. Ces chemins indépendants sont généralement intégrés au schéma de modulation numérique, au niveau de la couche physique. Plusieurs techniques sont possibles : employer des fréquences différentes dans la bande autorisée, suivre des chemins spatiaux différents entre différentes paires d'antennes ou répéter les émissions de bits pendant des laps de temps différents.

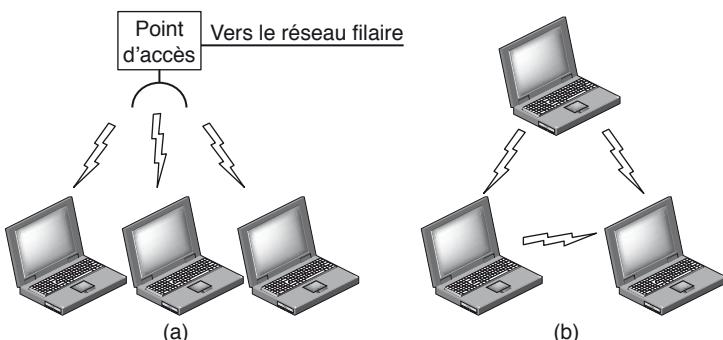


Figure 1.33 • (a) Réseau sans fil avec point d'accès. (b) Réseau ad hoc.

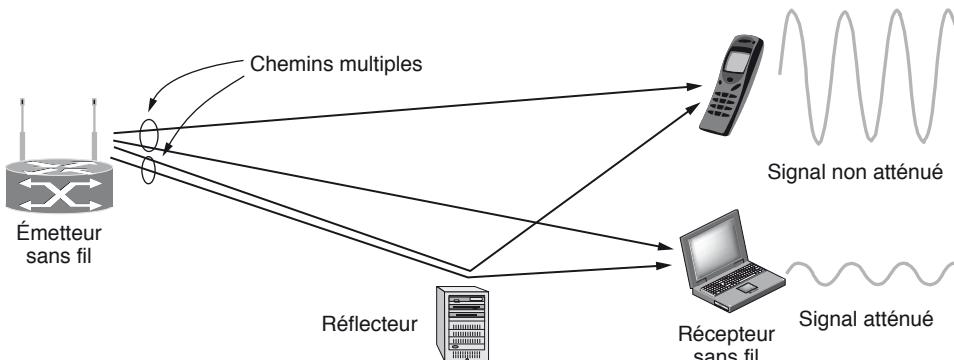


Figure 1.34 • Atténuation due aux trajets multiples.

Les différentes versions de 802.11 ont utilisé toutes ces techniques. La norme d'origine (publiée en 1997) définissait un LAN sans fil qui fonctionnait à 1 ou à 2 Mbit/s, en sautant d'une fréquence à l'autre ou en étalant le signal sur le spectre permis. Les utilisateurs ne tardèrent pas à se plaindre d'une lenteur excessive, et les travaux s'orientèrent vers plus de rapidité. La technique d'étalement du spectre fut améliorée et donna lieu à la norme 802.11b (1999) avec des débits atteignant 11 Mbit/s. Les normes 802.11a (1999) et 802.11g (2003) passèrent à une autre technique de modulation, nommée **OFDM** (*Orthogonal Frequency Division Multiplexing*), soit multiplexage par répartition en fréquences orthogonales. Elle divise une large bande du spectre en nombreuses tranches étroites sur lesquelles les différents bits sont envoyés en parallèle. Cette amélioration, que nous étudierons au chapitre 2, a permis aux réseaux 802.11a/g d'atteindre des débits allant jusqu'à 54 Mbit/s. C'était une augmentation significative, mais les utilisateurs en voulaient encore plus pour prendre en charge des applications toujours plus exigeantes en débit. La dernière version de la norme est 802.11n (2009). Elle utilise des bandes de fréquence plus larges et jusqu'à quatre antennes par ordinateur, pour obtenir des débits pouvant atteindre 450 Mbit/s.

Comme le sans-fil est, par nature, un média à diffusion, les radios 802.11 doivent également prendre en compte le problème représenté par plusieurs émissions simultanées qui entrent en collision et peuvent interférer avec la réception. Pour ce faire, 802.11 utilise la méthode **CSMA** (*Carrier Sense Multiple Access*), inspiré des idées de l'Ethernet filaire classique, qui, ironiquement, a repris lui-même les concepts d'un ancien réseau radio développé à l'université de Hawaï et nommé **ALOHA**. Dans celle-ci, les ordinateurs attendent un court intervalle aléatoire avant d'émettre et diffèrent leur émission s'ils se rendent comptent que quelqu'un d'autre émet déjà. Ainsi, deux ordinateurs sont moins susceptibles d'émettre en même temps. Toutefois, cela ne fonctionne pas aussi bien dans le cas des réseaux filaires. Pour comprendre pourquoi, examinez la figure 1.35. Supposez que l'ordinateur A envoie des données à l'ordinateur B mais que la portée radio de son émetteur soit trop courte pour atteindre l'ordinateur C. Si C souhaite émettre en direction de B, il peut écouter avant de commencer, mais le fait qu'il n'entende rien ne signifie pas pour autant que l'émission réussira. L'incapacité de C à entendre A peut faire que des collisions se

produisent. Malgré ce problème et quelques autres, cette technique fonctionne assez bien dans la pratique.

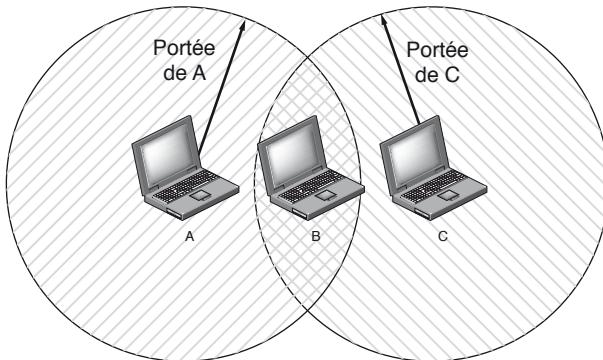


Figure 1.35 • La portée d'un émetteur peut ne pas couvrir l'ensemble du système.

Un autre problème est celui de la mobilité. Si un client mobile s'éloigne du point d'accès qu'il utilise et entre dans la portée d'un autre point d'accès, il faut trouver un moyen d'assurer le relais. La solution réside dans un réseau 802.11, lequel peut être constitué de plusieurs cellules, chacune dotée de son propre point d'accès, et d'un système de distribution qui connecte les cellules. Ce système de distribution est souvent en Ethernet commuté mais il peut utiliser n'importe quelle autre technologie. Quand les clients se déplacent, ils peuvent trouver un autre point d'accès offrant un meilleur signal que celui auquel ils sont associés actuellement et changer d'association. De l'extérieur, le système entier est perçu comme un seul LAN filaire.

Cela dit, la mobilité est moins intéressante dans les réseaux 802.11 que dans les réseaux de téléphonie mobile. Généralement, 802.11 est utilisé par des clients nomades, qui se rendent d'un point fixe à un autre, plutôt que pendant le déplacement. La mobilité n'est pas réellement nécessaire pour l'usage nomadique. Même si cette possibilité est exploitée, elle ne concerne qu'un seul réseau 802.11, qui ne peut couvrir au plus qu'un grand bâtiment. De nouvelles méthodes devront permettre la mobilité entre des réseaux différents et des technologies diverses (par exemple 802.21).

Enfin se pose le problème de la sécurité. Comme les transmissions sans fil sont des diffusions, des ordinateurs proches peuvent facilement recevoir des paquets d'informations qui ne leur étaient pas destinés. Pour empêcher cela, la norme 802.11 incluait un protocole de chiffrement nommé **WEP** (*Wired Equivalent Privacy*). L'idée était d'offrir aux réseaux sans fil une sécurité égale à celle des réseaux filaires. Si cette idée était bonne, son implémentation fut imparfaite et elle fut bientôt craquée. WEP a depuis été remplacé, avec la norme 802.11i, par un nouveau mécanisme, **WPA** (*WiFi Protected Access*), maintenant devenu **WPA2**.

802.11 a provoqué dans l'univers des réseaux sans fil une révolution qui ne fait que commencer. En dehors des bâtiments, on en installe de plus en plus dans les trains,

les avions, les bateaux et les automobiles, pour que les utilisateurs puissent surfer sur l'Internet partout où ils se trouvent. Les téléphones mobiles et toutes sortes de produits électroniques grand public, des consoles de jeu aux appareils photo numériques, peuvent communiquer avec. Nous y reviendrons en détail au chapitre 4.

1.5.4 RFID et les réseaux de capteurs

Les réseaux que nous avons étudiés jusqu'ici étaient composés d'équipements électroniques faciles à reconnaître, des ordinateurs aux téléphones mobiles. Avec la **radio-identification**, ou **RFID** (*Radio Frequency IDentification*), les objets de tous les jours peuvent également faire partie d'un réseau informatique.

Une étiquette RFID, ou radio-étiquette, ressemble à un autocollant de la taille d'un timbre-poste que l'on peut fixer sur un objet (ou implanter dans cet objet) pour l'identifier ou le localiser. L'objet peut être une vache, un passeport, un livre ou une palette de manutention. L'étiquette est constituée d'une micropuce dotée d'un identifiant unique et d'une antenne qui reçoit des émissions radio. Des lecteurs RFID installés à des points de contrôle détectent les étiquettes quand elles sont à portée, et les interrogent pour lire les informations qu'elles contiennent (voir figure 1.36). Les applications sont nombreuses, allant de la vérification d'identité à la gestion de la chaîne logistique, en passant par le chronométrage des courses et le remplacement des codes-barres.

Il existe de nombreuses formes de RFID, chacune dotée de propriétés différentes, mais l'aspect le plus fascinant est peut-être que la plupart des radio-étiquettes ne possèdent ni prise électrique, ni batterie : toute l'énergie nécessaire à leur fonctionnement est fournie sous forme d'ondes radio par les lecteurs RFID. On qualifie cette technologie de **radio-identification passive**, pour la distinguer de la **radio-identification active** (moins courante) dans laquelle l'étiquette contient une source d'alimentation.

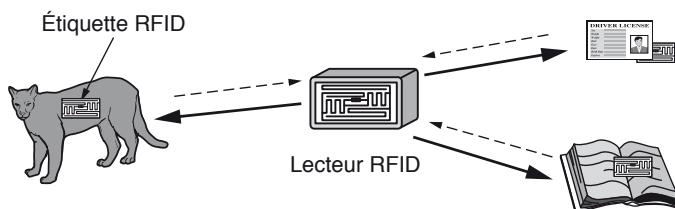


Figure 1.36 • RFID et mise en réseau d'objets de la vie quotidienne.

Une forme courante de RFID est la technologie nommée **UHF RFID** (*Ultra-High Frequency RFID*). On l'utilise sur les palettes de manutention et sur certains permis de conduire nord-américains. Les lecteurs émettent des signaux dans la bande des 902-928 MHz aux États-Unis. Les étiquettes communiquent à des distances de plusieurs mètres en changeant la façon dont elles réfléchissent les signaux des lecteurs,

et les lecteurs sont capables d'intercepter ces réflexions. Ce mode de fonctionnement s'appelle la **rétrodiffusion** (*backscatter*).

Une autre variante répandue est la technologie **HF RFID** (*High Frequency RFID*). Fonctionnant à 13,56 MHz, elle est susceptible d'être présente dans les passeports, les cartes de crédit, les livres et les systèmes de paiement sans contact. HF RFID a une courte portée, généralement d'un mètre ou moins, parce que le mécanisme physique est basé sur l'induction et non sur la rétrodiffusion. D'autres formes de RFID utilisent d'autres fréquences, par exemple **LF RFID** (*Low Frequency RFID*), développée avant HF RFID et qui est utilisée pour les animaux. Votre chat en est peut-être déjà équipé.

Les lecteurs RFID doivent d'une façon ou d'une autre résoudre le problème des étiquettes multiples se trouvant à portée. Cela signifie que celles-ci ne peuvent pas se contenter de répondre lorsqu'elles entendent un lecteur, sous peine que leurs signaux entrent en collision. La solution est comparable à la méthode adoptée dans la norme 802.11 : les étiquettes attendent un court laps de temps aléatoire avant de répondre en s'identifiant, ce qui permet au lecteur de les isoler et de les interroger.

La sécurité est un autre problème. La capacité des lecteurs RFID à localiser facilement un objet, et donc la personne qui le porte, peut conduire à une atteinte à la vie privée. Malheureusement, les radio-étiquettes sont difficiles à sécuriser, parce qu'elles ne disposent pas de la puissance de calcul et de communication nécessaire pour exécuter des algorithmes de chiffrement forts. On utilise donc des mécanismes plus faibles, comme les mots de passe (qui sont faciles à craquer). Si un policier peut facilement lire à distance votre carte d'identité à un poste frontière, qu'est-ce qui empêchera d'autres personnes de localiser cette même carte à votre insu ? Pas grand-chose.

Après avoir été de simples puces d'identification, les étiquettes RFID se transforment rapidement en ordinateurs à part entière. Par exemple, nombre d'entre elles possèdent une mémoire qui peut être mise à jour et interrogée plus tard, si bien qu'il est possible d'y stocker des informations sur l'historique des objets. Des chercheurs en ont démontré les risques : ces étiquettes sont vulnérables à tous les problèmes posés par des logiciels malveillants, et votre passeport ou votre chat pourraient un jour servir à propager un virus.

Un pas supplémentaire rendu possible par la radio-identification est le **réseau de capteurs**. Ces réseaux sont déployés pour surveiller des aspects du monde physique. Jusqu'ici, ils ont surtout été employés dans le cadre de la recherche scientifique, comme l'observation des populations d'oiseaux, de l'activité des volcans ou de la migration des zèbres, mais des applications industrielles telles que les équipements médicaux, les systèmes de surveillance des vibrations et le suivi des produits surgelés, réfrigérés ou autrement périssables ne sauraient certainement tarder.

Les noeuds capteurs sont de petits ordinateurs, souvent de la taille d'un porte-clés, capables de percevoir des températures, des vibrations, etc. De nombreux noeuds sont placés dans l'environnement à surveiller. Ils fonctionnent généralement sur batterie, bien que certains puissent tirer leur énergie des vibrations ou du soleil. Comme pour la radio-identification, le fait de disposer de suffisamment d'énergie est une difficulté importante, et les noeuds doivent communiquer soigneusement pour pouvoir

transmettre les informations perçues par les capteurs à un point de collecte externe. Une stratégie courante passe par leur auto-organisation pour relayer les messages entre eux, comme le montre la figure 1.37. Cette organisation s'appelle un **réseau multi-saut** (*multihop networks*).

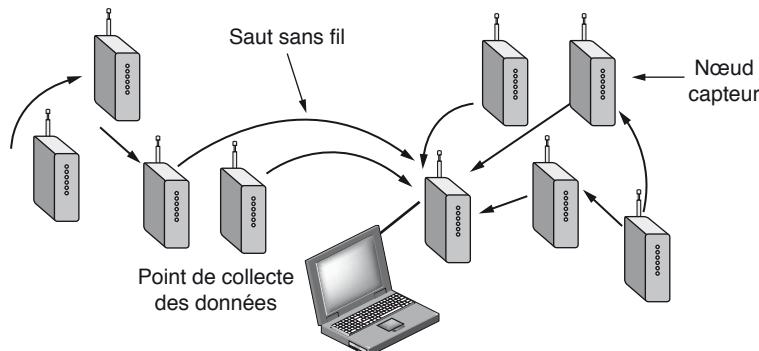


Figure 1.37 • Topologie multi-saut d'un réseau de capteurs.

La radio-identification et les réseaux de capteurs sont certainement voués à un bel avenir. Des chercheurs ont déjà combiné le meilleur des deux technologies, en créant un prototype de plate-forme associant des étiquettes RFID programmables et des capteurs de lumière, de mouvement et d'orientation.

1.6 Normalisation des réseaux

Il existe de nombreux fabricants et fournisseurs de réseaux, chacun possédant ses propres idées sur les façons de procéder. Sans coordination, ce serait le chaos le plus total, au grand dam des utilisateurs. Pour éviter cela, il faut se mettre d'accord sur des normes. De bonnes normes ne permettent pas seulement à des ordinateurs différents de communiquer : elles entraînent aussi un élargissement du marché pour les produits qui y adhèrent et, du même coup, une production en masse, des économies d'échelle lors de la fabrication, de meilleures implémentations et d'autres avantages qui font baisser les prix et favorisent l'acceptation des produits.

Cette section présente le monde important mais mal connu de la normalisation internationale. Mais voyons d'abord ce qui appartient à une norme. Une personne raisonnable pourrait penser qu'une norme explique comment un protocole doit fonctionner, pour que l'on puisse l'implémenter de manière satisfaisante. Elle aurait tort. Les normes définissent les conditions de l'interopérabilité : ni plus ni moins. Cela permet l'émergence d'un marché plus vaste, dans lequel les entreprises sont en concurrence sur des critères de qualité. Par exemple, la norme 802.11 définit plusieurs débits d'émission mais ne dit pas quand un émetteur doit utiliser quel débit, ce qui est un facteur essentiel pour de bonnes performances. Cette décision incombe au fabricant du produit.

L'interopérabilité est souvent difficile à atteindre dans ces conditions, puisque les choix d'implémentation sont nombreux et que les normes définissent généralement quantité d'options. 802.11 a posé tant de problèmes que, dans une stratégie devenue monnaie courante, un groupement professionnel, la **Wi-Fi Alliance**, a été lancé pour travailler sur l'interopérabilité dans le cadre de la norme.

De même, une norme définira la manière dont un protocole fonctionne sur un média, mais pas comment implémenter l'interface de service, sauf pour aider à expliquer le protocole. Les interfaces de service réelles sont souvent propriétaires. Par exemple, la façon dont TCP est interfacé avec IP dans un ordinateur n'a pas d'importance quand il s'agit de converser avec un hôte distant. La seule chose qui compte, c'est que l'hôte en question « parle » TCP/IP. En fait, TCP et IP sont couramment implantés ensemble sans aucune interface distincte. Cela dit, les bonnes interfaces de services, comme les bonnes API, sont précieuses pour l'adoption des protocoles, et les meilleures (comme les sockets de Berkeley) peuvent devenir très courantes.

On distingue les standards et les normes. Les standards, appelés parfois normes *de facto* (de fait), découlent d'une situation établie et non d'une approbation officielle. HTTP a commencé par être un standard *de facto*. Il faisait partie des premiers navigateurs développés par Tim Berners-Lee au CERN et a pris son essor avec la croissance du Web. Bluetooth est un autre exemple. Il a été développé à l'origine par Ericsson, mais tout le monde l'utilise désormais.

Les vraies normes, appelées aussi normes *de jure* (de droit), bénéficient quant à elles d'une reconnaissance officielle de la part d'organismes de normalisation officiels, qui se divisent en deux groupes : ceux fondés en vertu de traités signés par plusieurs gouvernements et ceux issus d'organisations indépendantes. Dans le domaine des réseaux, on trouve plusieurs organisations de chaque type, notamment l'UIT, l'ISO, l'IETF et l'IEEE, dont il sera question ci-après.

Dans la pratique, les relations entre les normes, les entreprises et les instances de normalisation, sont compliquées. Les standards se transforment souvent en normes, surtout s'ils connaissent le succès. C'est ce qui s'est passé dans le cas de HTTP, qui a été rapidement repris par l'IETF. Les organismes de normalisation ratifient souvent les normes des autres, en se congratulant mutuellement, pour élargir le marché d'une technologie. De nos jours, de nombreuses alliances commerciales ponctuelles se forment autour d'une technologie particulière et jouent un rôle significatif dans le développement et l'amélioration des normes réseau. Par exemple, le **3GPP** (*Third Generation Partnership Project*) est un projet collaboratif qui réunit des organismes de standardisation des télécommunications et travaille sur les normes de téléphonie mobile UMTS 3G.

1.6.1 Le Who's Who des télécommunications

Le statut juridique des opérateurs de télécommunications varie considérablement d'un pays à l'autre. À un extrême, les États-Unis totalisent plus de 2 000 opérateurs privés (majoritairement très petits). Quelques-uns se sont ajoutés en 1984, à la suite du démantèlement d'AT&T (qui était alors la plus grande entreprise au monde et

assurait à elle seule environ 80 % du service téléphonique), puis avec le *Telecommunications Act* de 1996, qui a remanié la réglementation pour encourager la concurrence.

À l'autre extrême, on trouve les pays dont le gouvernement détient le monopole de toutes les télécommunications : courrier électronique, télégraphe, téléphone, et souvent radio et télévision. Une grande partie du monde tombe dans cette catégorie. Dans certains cas, l'opérateur de télécommunications est une entreprise publique. Dans d'autres, c'est une branche de l'administration de type P&T (postes et télécommunications). Actuellement, la tendance mondiale est à la libéralisation, c'est-à-dire à la libre concurrence et à la fin du monopole d'État. Dans la plupart des pays européens, le secteur des télécommunications a déjà été privatisé (parfois seulement en partie), et ailleurs le mouvement ne prend que lentement de l'essor.

Étant donné cette diversité de fournisseurs de services, une compatibilité à l'échelle mondiale est nécessaire pour que les personnes (et les ordinateurs) d'un pays puissent communiquer avec leurs homologues d'un autre pays. En fait, ce besoin n'est pas nouveau. En 1865, déjà, des représentants de nombreux gouvernements européens se sont réunis pour fonder ce qui est aujourd'hui l'UIT (Union internationale des télécommunications) en lui donnant pour mission la normalisation des télécommunications internationales qui se limitaient à l'époque au télégraphe. Il était clair que si certains pays utilisaient le morse et que d'autres utilisaient un code différent, cela poserait un problème. Lorsque le téléphone est devenu international, l'UIT s'est également chargée de sa normalisation. Depuis 1947, l'UIT est une agence des Nations unies.

L'UIT compte environ deux cents gouvernements membres, dont presque tous les pays des Nations unies. Les États-Unis ne possédant pas d'administration des télécommunications pour les représenter, c'est le département d'État qui a été choisi pour s'y substituer, probablement en raison du fait que les affaires étrangères sont sa spécialité et que l'UIT avait justement affaire à des pays étrangers. Elle compte également sept cents membres sectoriels et associés, parmi lesquels des opérateurs de télécommunications (AT&T, Vodafone, Sprint...), des constructeurs d'équipements de télécommunications (Cisco, Nokia, Nortel...), des industriels de l'informatique (Microsoft, Agilent, Toshiba...), des fabricants de circuits intégrés (Intel, Motorola, Texas Instruments...) et d'autres entreprises intéressées (Boeing, CBS, VeriSign...).

L'UIT comprend trois secteurs principaux. Nous nous intéresserons surtout à l'UIT-T, le secteur de la normalisation des télécommunications, qui s'occupe des systèmes de téléphonie et de transmission de données. Avant 1993, ce secteur s'appelait le CCITT (Comité consultatif international télégraphique et téléphonique). L'UIT-R, le secteur des radiocommunications, est chargé de coordonner la façon dont les groupes d'intérêts concurrents utilisent les fréquences radio dans le monde entier. Le dernier, l'UIT-D, est le secteur du développement des technologies de l'information et de la communication. L'un de ses objectifs est de réduire la « fracture numérique » entre les pays qui accèdent massivement à ces technologies et les pays en voie de développement.

Le rôle de l'UIT-T est d'émettre des recommandations techniques concernant les interfaces téléphoniques, télégraphiques et de transmission de données.

Ces recommandations deviennent souvent des normes internationales, bien qu'il s'agisse techniquement de suggestions que les gouvernements sont libres d'accepter ou d'ignorer. Rien n'empêche un pays d'adopter une norme de téléphonie différente de celle utilisée partout ailleurs, mais cela le couperait du reste du monde.

Au sein de l'UIT-T, dix commissions d'études, réunissant parfois quatre cents participants et couvrant des sujets allant de la facturation téléphonique à la sécurité en passant par les services multimédias, se partagent le travail. Par exemple, la CE 15 normalise les technologies DSL et optiques d'accès à l'Internet. Elles sont divisées en groupes de travail qui sont, à leur tour, divisés en groupes d'experts pouvant eux-mêmes être organisés en groupes *ad hoc*.

Depuis sa création, l'UIT-T a produit plus de trois mille recommandations totalisant approximativement soixante mille pages, dont beaucoup sont largement mises en application, par exemple, la recommandation H.264 (qui est aussi la norme ISO MPEG-4 AVC) pour la compression vidéo, ou la norme X.509 qui spécifie notamment les certificats à clé publique utilisés pour sécuriser la navigation web et pour la signature numérique des courriers électroniques.

Avec le passage des télécommunications de la sphère nationale à la scène mondiale, un processus entamé dans les années 1980, les normes vont jouer un rôle de plus en plus important et un nombre toujours croissant d'organisations vont vouloir s'engager dans leur élaboration.

1.6.2 Le Who's Who de la normalisation internationale

Les normes internationales sont produites et publiées par l'**ISO** (Organisation internationale de normalisation), une organisation indépendante fondée en 1946. Ses membres sont les instances nationales de normalisation de 157 pays membres, parmi lesquelles l'ANSI (États-Unis), le BSI (Royaume-Uni), l'AFNOR (France) et le DIN (Allemagne).

Ces normes couvrent de très nombreux domaines qui vont des écrous et des boulons aux revêtements des poteaux téléphoniques – sans parler des fèves de cacao (ISO 2451), des filets de pêche (ISO 1530), des sous-vêtements féminins (ISO 4416) et de toutes sortes d'autres sujets que l'on n'aurait pas imaginés se prêtant à la normalisation. En matière de télécommunications, l'**ISO** et l'**UIT-T** coopèrent souvent (l'**ISO** est membre de l'**UIT**).

Plus de 17 000 normes ont déjà été établies, dont les normes OSI. L'**ISO** compte plus de deux cents comités techniques (TC), numérotés dans leur ordre de création et traitant chacun d'un sujet spécifique. Par exemple, le comité TC1 s'occupe des filetages, et le JTC1 (*Joint Technical Committee 1*) des technologies de l'information, notamment des réseaux, des ordinateurs et des logiciels. C'est le premier (et jusqu'ici le seul) comité technique mixte, créé en 1987 par la fusion du TC97 et des activités de la Commission électrotechnique internationale, un autre organisme de normalisation. Chaque TC possède des sous-comités (SC) qui sont divisés en groupes de travail (WG).

La plus grande partie du travail est réalisée au sein des WG par plus de cent mille bénévoles à travers le monde. Ces « bénévoles » sont en fait des employés d'entreprises dont les produits sont en cours de normalisation. Participant également à ces travaux des fonctionnaires d'États qui souhaiteraient que leurs propres normes soient approuvées sur le plan international. Des chercheurs universitaires s'impliquent aussi dans de nombreux groupes de travail.

La procédure d'adoption mise en place par l'ISO vise à recueillir le consensus le plus large possible. Elle débute lorsqu'une instance nationale de normalisation exprime le besoin de disposer d'une norme internationale dans un certain domaine. Un groupe de travail est alors formé, dans le but de produire un avant-projet, ou **CD** (*Committee Draft*), ensuite transmis à tous les membres, qui disposent de six mois pour l'étudier. Si une majorité d'entre eux l'approuve, une version révisée appelée **DIS** (*Draft International Standard*) leur est soumise pour commentaires et vote. Si cette étape est concluante, le texte final de la norme, l'**IS** (*International Standard*), est préparé, approuvé et publié. En cas de contestation importante, plusieurs versions d'un CD ou d'un DIS peuvent être proposées jusqu'à obtenir un nombre suffisant de voix, ce qui peut prendre des années.

Aux États-Unis, le **NIST** (*National Institute of Standards and Technology*) – autrefois *National Bureau of Standards* –, qui fait partie du ministère du Commerce, publie des normes s'appliquant obligatoirement à tout produit acheté par le gouvernement. Le DoD définissant ses propres normes, il échappe à celles du NIST.

Un autre acteur majeur du monde de la normalisation est l'**IEEE** (*Institute of Electrical and Electronics Engineers*), la plus grande organisation professionnelle au monde. En plus de publier de nombreuses revues et d'organiser annuellement des centaines de conférences, elle possède une section chargée de développer des normes dans les domaines de l'électronique et de l'informatique. Par exemple, le comité IEEE 802 a normalisé plusieurs types de LAN. Nous étudierons certains de ses travaux plus loin dans ce livre. La figure 1.38 présente les groupes de travail qui œuvrent dans le cadre de ce comité et montre un taux de réussite plutôt bas : disposer d'un numéro 802.x n'est pas une garantie de succès. Pourtant, l'impact des « success stories » (comme 802.3 et 802.11) sur l'industrie et sur le monde a été énorme.

Figure 1.38 • Groupes de travail du comité 802.

Numéro	Sujet
802.1	Vue d'ensemble et architecture des LAN.
802.2 ↓	Contrôle de liaison logique (LLC).
802.3 *	Ethernet.
802.4 ↓	Bus à jeton (utilisé un temps dans les installations industrielles).
802.5	Anneau à jeton (LAN d'IBM).
802.6 ↓	DQDB (premier MAN).
802.7 ↓	Groupe consultatif sur les technologies à large bande.

Numéro	Sujet
802.8 †	Groupe consultatif sur les technologies à fibre optique.
802.9 ↓	LAN isochrones (pour les applications en temps réel).
802.10 ↓	LAN virtuels (VLAN) et sécurité.
802.11 *	LAN sans fil (Wi-Fi).
802.12 ↓	Demande de priorité (AnyLAN de Hewlett-Packard).
802.13	Numéro non affecté (les scientifiques sont parfois superstitieux).
802.14 ↓	Modem-câble (disparu. Un consortium industriel a tiré le premier).
802.15 *	Réseaux personnels (PAN) [Bluetooth].
802.16 *	Sans fil à large bande (WiMAX).
802.17	RPR (<i>Resilient packet ring</i>).
802.18	Groupe consultatif sur les questions de réglementation radio.
802.19	Groupe consultatif sur la coexistence de toutes ces normes.
802.20	Sans fil à large bande mobile (analogique à 802.16 ^e).
802.21	Transfert indépendant du média (pour la mobilité entre différentes technologies).
802.22	Réseau régional sans fil.

* : groupes les plus importants. ↓ : groupes en suspens. † : groupes ayant abandonné.

1.6.3 Le Who's Who de la normalisation de l'Internet

L'Internet possède ses propres mécanismes de normalisation qui sont très différents de ceux de l'UIT-T et de l'ISO. En simplifiant à l'excès, on peut dire qu'il y a d'un côté ceux qui assistent aux réunions en costume cravate et, de l'autre, ceux qui portent des jeans (ou carrément des shorts lorsqu'ils se réunissent à San Diego). En fait, les réunions organisées par l'UIT-T et l'ISO sont fréquentées par des représentants d'entreprises et de gouvernements dont la normalisation est le travail et qui sont intimement convaincus de son utilité. Quant aux tenants de l'Internet, ils préfèrent l'anarchie tout en sachant qu'un minimum de normes est nécessaire pour que les centaines de millions d'internautes puissent communiquer. Les normes sont donc souvent vues comme un mal nécessaire. David Clark a dit un jour, dans une remarque maintenant célèbre, que les normes Internet découlent « d'un large consensus et d'un code qui tourne ».

Lors de la création de l'ARPAnet, le DoD a constitué un comité informel pour en assurer la surveillance. En 1983, ce comité a été rebaptisé **IAB** (*Internet Activities Board*) et a reçu comme tâche supplémentaire de faire coopérer autant que possible les chercheurs impliqués dans le développement de l'ARPAnet et ceux travaillant sur l'Internet. L'acronyme IAB a pris par la suite le sens de *Internet Architecture Board*.

L'IAB comptait une dizaine de membres qui dirigeaient chacun un groupe de travail consacré à un aspect important. Ils se réunissaient plusieurs fois par an pour examiner les résultats des travaux et en rendre compte au DoD et à la NSF qui finançaient en grande partie ce comité. Lorsqu'une norme était nécessaire (par exemple, un nouvel algorithme de routage), ils en débattaient puis annonçaient le changement de façon que les étudiants en troisième cycle, sans qui nombre de développements logiciels n'auraient pas été possibles (en tout cas pas aussi rapidement), puissent l'implémenter. La communication se faisait par le biais d'une série de rapports techniques appelés **RFC** (*Request For Comments*). Il en existe aujourd'hui plus de cinq mille, numérotés selon leur ordre chronologique. Ils sont stockés en ligne et consultables à l'adresse www.ietf.org/rfc. Nous nous référerons à de nombreuses RFC au cours de ce livre.

En 1989, l'Internet avait pris une telle ampleur que cette façon très informelle de procéder était devenue inadaptée. De nombreux fabricants offraient alors des produits TCP/IP et n'étaient pas prêts à les modifier sous prétexte qu'une dizaine de chercheurs avaient pensé à une meilleure idée. Cet été-là, l'IAB fut donc réorganisé de nouveau. Les chercheurs furent répartis en deux groupes : l'**IRTF** (*Internet Research Task Force*), qui dépendait de l'IAB, et l'**IETF** (*Internet Engineering Task Force*), et des représentants d'organisations extérieures à la recherche ont rejoint l'IAB. Ce comité était au départ renouvelé par cooptation, chaque membre étant élu pour un mandat de deux ans. Plus tard, l'**ISOC** (*Internet Society*) fut créée, composée de professionnels intéressés par l'Internet. En un sens, elle est comparable à l'ACM et à l'IEEE. Elle est gouvernée par des administrateurs élus qui désignent les membres de l'IAB.

L'idée de cette répartition était que l'IRTF se concentre sur les recherches à long terme et l'IETF sur les développements à court terme. Des groupes de travail furent constitués au sein de l'IETF, chacun chargé de résoudre un problème spécifique. Les sujets traités incluaient les nouvelles applications, l'information des utilisateurs, l'intégration OSI, le routage et l'adressage, la sécurité, la gestion des réseaux et la normalisation. À l'origine, les présidents des différents groupes se réunissaient en comités de pilotage pour diriger l'effort de développement. Au bout d'un moment, les groupes étaient devenus si nombreux (plus de soixante-dix) qu'ils furent regroupés en zones, dont les présidents formèrent le comité de pilotage.

En outre, un autre processus de normalisation fut adopté, plus formel celui-là, et calqué sur le modèle de l'ISO. Il prévoit deux étapes principales avant l'acceptation d'une norme : une proposition ou **PS** (*Proposed Standard*) et un avant-projet ou **DS** (*Draft Standard*). Pour qu'une idée fasse l'objet d'une proposition, elle doit d'abord être détaillée dans une RFC et présenter un intérêt suffisant pour la communauté. Pour passer à l'étape d'avant-projet, une implémentation opérationnelle doit subir des tests stricts sur deux sites indépendants pendant au moins quatre mois. Si l'IAB est convaincu que l'idée est viable et que le logiciel fonctionne correctement, il déclare la RFC comme une **norme Internet** (*Internet Standard*). Le DoD a fait siennes certaines de ces normes (MIL-STD), les rendant incontournables pour ses fournisseurs.

En ce qui concerne le Web, le **W3C** (*World Wide Web Consortium*) développe des protocoles et des lignes directrices pour faciliter sa croissance à long terme. Il s'agit d'un consortium supervisé par Tim Berners-Lee, fondé en 1994 lorsque le Web a

réellement commencé à prendre son essor. Il compte maintenant plus de trois cents membres dans le monde entier et a produit plus de cent recommandations W3C, nom donné à ces standards, couvrant des sujets tels que HTML et la protection de la vie privée sur le Web.

1.7 Système métrique

Ce livre, comme toute l'informatique en général, utilise le système métrique. Les principaux préfixes sont répertoriés à la figure 1.39 mais ils apparaissent généralement dans leur forme abrégée, limitée à la première lettre. Par exemple, si l'on prend l'octet comme unité de mesure, également abrégé en « o », on obtient : ko, Mo, Go, etc. Le préfixe (k, M, G) est généralement une majuscule, à l'exception, pour des raisons historiques, du « k » de kbit/s (kilobits par seconde). Notez que dans les expressions du genre 64 kbit/s ou 10 Mbit/s, « bit » est aussi une abréviation, raison pour laquelle il ne prend pas de « s » au pluriel. Notez également que « μ » (la lettre grecque mu) signifie « micro ».

Vous avez probablement remarqué qu'employés avec « bit », les préfixes prennent chacun leur valeur habituelle, alors qu'associés à « octet » (pour mesurer la taille d'une mémoire, d'un disque, d'un fichier, d'une base de données, etc.), ils prennent des valeurs différentes. Par exemple, kilo ne signifie pas 10^3 (1 000) mais 2^{10} (1 024), parce que l'espace mémoire est mesuré en puissances de 2. Donc 1 Ko de mémoire contient 1 024 octets et non 1 000 octets. De la même manière, 1 Mo de mémoire contient 2^{20} (1 048 576) octets, 1 Go contient 2^{30} (1 073 741 824) octets, et une base de données de 1 To contient 2^{40} (1 099 511 627 776) octets. En revanche, une ligne de transmission à 1 kbit/s transmet 1 000 bits par seconde et un LAN à 10 Mbit/s fonctionne avec un débit de 10 000 000 bit/s, car ces valeurs ne sont pas issues de puissances de 2. Beaucoup ont tendance à confondre ces deux systèmes, surtout pour les tailles de disque. Nous utilisons donc dans ce livre les symboles Ko, Mo, Go et To pour 2^{10} , 2^{20} , 2^{30} et 2^{40} octets respectivement, et les symboles kbit/s, Mbit/s, Gbit/s et Tbit/s pour 10^3 , 10^6 , 10^9 et 10^{12} bits par seconde respectivement.

Figure 1.39 • Les principaux préfixes du système métrique.

Puissance	Explicite	Préfixe	Puiss.	Explicite	Préfixe
10^{-3}	0,001	milli	10^3	1 000	kilo
10^{-6}	0,000 001	micro	10^6	1 000 000	méga
10^{-9}	0,000 000 001	nano	10^9	1 000 000 000	giga
10^{-12}	0,000 000 000 001	pico	10^{12}	1 000 000 000 000	téra
10^{-15}	0,000 000 000 000 001	femto	10^{15}	1 000 000 000 000 000	péta
10^{-18}	0,000 000 000 000 000 001	atto	10^{18}	1 000 000 000 000 000 000	exa
10^{-21}	0,000 000 000 000 000 000 001	zepto	10^{21}	1 000 000 000 000 000 000 000	zetta
10^{-24}	0,000 000 000 000 000 000 000 001	yocto	10^{24}	1 000 000 000 000 000 000 000 000	yotta

1.8 Aperçu de la suite de cet ouvrage

Ce livre décrit les réseaux à la fois du point de vue théorique et pratique. La plupart des chapitres débutent par une présentation de certains principes, qui sont ensuite illustrés à l'aide de nombreux exemples. Ceux-ci s'inspirent généralement de deux types de réseaux importants mais aussi très différents : l'Internet et les réseaux sans fil comme les systèmes de téléphonie cellulaire. Des exemples évoquant d'autres réseaux seront aussi présentés si nécessaire.

La structure du livre suit le modèle hybride de la figure 1.23. Nous commencerons dès le chapitre 2 à explorer la hiérarchie de protocoles à partir du bas en abordant les fondements de la transmission de données, tant dans les systèmes de transmission filaires que sans fil. Les informations présentées concernent essentiellement les aspects architecturaux de la couche physique, ignorant ses aspects matériels. Plusieurs exemples relatifs à la couche physique, tels que le réseau téléphonique public, le réseau de téléphonie mobile et la télévision par câble, sont aussi présentés.

Les chapitres 3 et 4 traitent de la couche liaison de données en deux parties. Le chapitre 3 présente le problème de la transmission de paquets sur un lien, avec notamment la détection et la correction des erreurs. Nous étudierons DSL (utilisé pour l'accès Internet à haut débit sur une ligne téléphonique) comme exemple réel de protocole de niveau liaison de données.

Le chapitre 4 aborde la sous-couche d'accès au support, qui fait partie de la couche liaison de données et gère la façon de partager un canal entre plusieurs ordinateurs. Les exemples que nous verrons concernent des réseaux sans fil, comme les réseaux 802.11 et RFID, et des réseaux filaires comme les réseaux Ethernet classiques. Les commutateurs de cette couche, qui servent à connecter des LAN comme les réseaux Ethernet commuté, sont également décrits.

Le chapitre 5 est consacré à la couche réseau et, en particulier, au routage. De nombreux algorithmes de routage statiques et dynamiques sont expliqués. Toutefois, même avec de bons algorithmes de routage, des paquets seront retardés ou perdus si le réseau reçoit plus de trafic qu'il ne peut en traiter. Nous étudierons ce problème, de la façon d'empêcher la congestion à celle de garantir une certaine qualité de service. Les nombreux problèmes liés à l'interconnexion de réseaux hétérogènes en interréseaux sont aussi traités. Ce chapitre se termine par un examen approfondi de la couche réseau de l'Internet.

Le chapitre 6 traite de la couche transport, en mettant l'accent sur les protocoles avec connexion et la fiabilité qui sont nécessaires à de nombreuses applications. Les protocoles de transport de l'Internet, UDP et TCP, ainsi que les questions liées à leurs performances font l'objet d'une présentation détaillée.

Le chapitre 7 examine la couche application, ses protocoles et ses applications. Il traite en premier du DNS, qui est en quelque sorte l'annuaire de l'Internet, avant de poursuivre avec le courrier électronique et ses protocoles. Puis nous passerons au Web, avec les contenus statiques et dynamiques et le fonctionnement côté client et côté serveur. Nous continuerons par une présentation des applications multimédias,

avec entre autres le streaming audio et vidéo. Enfin, nous aborderons les réseaux de fournitures de contenus, avec notamment les applications P2P.

Le chapitre 8 est consacré à la sécurité sur l'Internet. Étant donné que ce sujet touche toutes les couches, il était préférable de le traiter à la fin, après avoir analysé en détail chacune d'elles. Il débute par une introduction à la cryptographie, puis illustre son rôle dans la sécurisation des communications, du courrier électronique et du Web. Il conclut par une présentation de domaines dans lesquels la sécurité touche, non sans conflits éventuels, à des sujets sensibles tels que le respect de la vie privée, la liberté d'expression, la censure et d'autres aspects sociaux.

Une courte bibliographie qui propose quelques références en français sur le domaine clôt l'ouvrage.

1.9 Résumé

Les réseaux d'ordinateurs offrent de nombreux services, tant aux entreprises qu'aux particuliers. Les entreprises les utilisent pour partager des informations, généralement selon le modèle client-serveur, dans lequel les stations de travail des employés accèdent à de puissants serveurs situés dans une salle informatique. Aux particuliers, les réseaux offrent l'accès à une variété de ressources, de l'information au divertissement, ainsi qu'un moyen d'acheter et de vendre des biens et des services. Ils accèdent souvent à l'Internet en se connectant à un FAI, une compagnie téléphonique ou un câblo-opérateur, le sans-fil étant de plus en plus utilisé avec les ordinateurs et les téléphones portables. Les progrès technologiques donnent lieu à de nouvelles applications mobiles et à des réseaux formés d'ordinateurs embarqués dans toutes sortes d'appareils. Ces mêmes progrès ne sont pas sans poser des questions de société, comme les problèmes de respect de la vie privée.

On peut dire sommairement que les réseaux se répartissent en LAN, MAN, WAN et interréseaux. Les LAN couvrent un immeuble et opèrent à des hauts débits, les MAN couvrent généralement une ville (un exemple est le système de télévision par câble que beaucoup utilisent aujourd'hui pour accéder à l'Internet) et les WAN peuvent couvrir un pays ou un continent. Certaines des technologies employées sont point-à-point (par exemple les liaisons par câble) alors que d'autres sont à diffusion (par exemple dans les réseaux sans fil). Des réseaux peuvent être interconnectés au moyen de routeurs pour former des interréseaux, dont l'Internet est le plus grand et le plus connu. Les réseaux sans fil, comme les réseaux 802.11 et les réseaux de téléphonie mobile, se répandent de plus en plus.

Les logiciels de réseau sont construits autour de protocoles, c'est-à-dire de règles permettant à des processus de communiquer. La plupart des réseaux supportent des hiérarchies de protocoles, chaque couche fournissant des services à la couche supérieure et isolant celle-ci des détails des protocoles utilisés dans les couches inférieures. Ces piles de protocoles s'inspirent habituellement du modèle OSI ou du modèle TCP/IP, qui ont en commun les couches réseau (internet pour TCP/IP), transport et application mais diffèrent en ce qui concerne les autres couches. Les aspects à prendre en

compte lors de la conception comprennent notamment la fiabilité, l'allocation des ressources, l'évolutivité et la sécurité. Une bonne partie du livre traite de ces protocoles et de leur conception.

Les réseaux fournissent différents services à leurs utilisateurs. Ces services peuvent aller de la livraison de paquets sans connexion en mode « au mieux » (*best effort*) à la livraison avec connexion et garantie. Dans certains réseaux, un service sans connexion est assuré par une couche et un service avec connexion est fourni par la couche supérieure.

Les réseaux les plus connus sont l'Internet, le réseau de téléphonie mobile 3G et le LAN sans fil 802.11. L'Internet a évolué à partir de l'ARPAnet, auquel sont venus s'ajouter d'autres réseaux pour former un interréseau. L'Internet d'aujourd'hui est en fait constitué d'un ensemble de milliers de réseaux qui utilisent la pile de protocoles TCP/IP. Le réseau de téléphonie mobile 3G fournit un accès sans fil à l'Internet, à des débits de plusieurs mégabits par seconde, ainsi bien sûr que des services voix. Les réseaux sans fil basés sur la norme IEEE 802.11, déployés dans nombre d'immeubles et de lieux publics, peuvent offrir une connectivité à des débits dépassant les 100 Mbit/s. De nouveaux types de réseaux sont également en train d'émerger, comme les réseaux de capteurs embarqués et ceux basés sur la technologie RFID.

Pour permettre à de nombreux ordinateurs différents de communiquer, un travail de normalisation important, tant sur le plan matériel que logiciel, doit être mené. Des organisations telles que l'UIT-T, l'ISO, l'IEEE et l'IAB se chargent chacune d'un aspect de ce processus de normalisation.

Exercices

- Imaginez que vous avez dressé Bernie, votre saint-bernard, pour qu'il puisse transporter une boîte de trois cartouches de 8 mm à la place d'un tonnelet de whisky (lorsque le disque de votre ordinateur est plein, vous considérez cela comme une urgence). Chaque cartouche contient 7 Go de données. Bernie peut vous rejoindre où que vous soyez à une vitesse de 18 km/h. Pour quelles distances le débit de Bernie est-il plus élevé qu'une ligne de transmission à 150 Mbit/s (surcharge de service exclue) ? Quelle est votre réponse si (i) la vitesse de Bernie est doublée, (ii) la capacité de chaque bande est doublée, (iii) le débit de la ligne de transmission est doublé ?
- Une alternative à un LAN est simplement un gros système en temps partagé avec un terminal par utilisateur. Citez deux avantages d'un système client-serveur utilisant un LAN.
- Les performances d'un système client-serveur sont fortement influencées par deux caractéristiques majeures : la bande passante du réseau (le nombre de bits qu'il peut transporter par seconde) et sa latence (le nombre de secondes que met le premier bit pour aller du client au serveur). Donnez deux exemples de réseaux, un ayant une bande passante et une latence élevées et un autre ayant une bande passante et une latence faibles.
- Outre la bande passante et la latence, quelle est l'autre caractéristique nécessaire à la définition de la qualité de service offerte par un réseau servant (i) au transport de la voix numérisée, (ii) au transport de vidéos, (iii) à des transactions financières ?
- Un facteur de retard sur un système à commutation de paquets en mode différé est le temps nécessaire à un commutateur pour recevoir complètement un paquet et le retransmettre. Un délai de commutation de $10 \mu\text{s}$ affecterait-il beaucoup le temps de réponse sur un système client-serveur dont le client se trouve à New York et le serveur en Californie, en admettant que la vitesse de propagation sur cuivre ou sur fibre soit égale à $2/3$ de la vitesse de la lumière dans le vide ?
- Un système client-serveur utilise un réseau satellitaire. Les satellites se trouvent à une altitude de 40 000 km. Quel peut être le meilleur temps de réponse à une requête ?
- Dans le futur, lorsque tout le monde disposera chez lui d'un terminal relié à un réseau d'ordinateurs, des référendums instantanés sur des questions importantes pourront avoir lieu. Il se peut même que l'on assiste par la suite à la disparition de la fonction de député au profit de l'expression directe de la volonté populaire. Les aspects positifs d'une telle démocratie directe sont évidents ; décrivez certains des aspects négatifs.
- Cinq routeurs doivent être interconnectés dans une configuration de sous-réseau point-à-point. Entre les paires de routeurs, les concepteurs ont le choix entre

installer une ligne à haut débit, une ligne à moyen débit, une ligne à bas débit ou pas de ligne du tout. S'il faut 100 ms de temps machine pour générer et inspecter chaque topologie, combien de temps faudrait-il pour les inspecter toutes ?

9. Un inconvénient des réseaux à diffusion est une perte de capacité lorsque plusieurs hôtes tentent d'accéder simultanément au canal. Prenons un exemple simple dans lequel le temps est divisé en intervalles discrets et où chacun des n hôtes tente avec une probabilité p d'accéder au canal durant chaque intervalle. Quelle sera la proportion d'intervalles gaspillés en raison des collisions ?
10. Donnez deux raisons pour lesquelles on utilise des protocoles en couches. Citez l'un de leurs inconvénients possibles.
11. Le président de la société Peintures & Co envisage de s'associer avec un brasseur local pour produire une canette de bière invisible (à titre de mesure antipollution). Il demande conseil à son service juridique, lequel contacte ensuite les ingénieurs pour qu'ils commencent à travailler sur le projet. L'ingénieur en chef appelle son homologue de l'autre société pour discuter des aspects techniques. Ils remettent ensuite un rapport à leurs départements juridiques respectifs, qui prennent contact par téléphone pour traiter des aspects légaux. Pour finir, les deux présidents négocient l'aspect financier de l'association. Quel principe d'un protocole multicouche, au sens du modèle OSI, ce mécanisme de communication enfreint-il ?
12. Deux réseaux offrent un service avec connexion, l'un sous la forme de flots d'octets fiables et l'autre sous la forme de flots de messages fiables. Sont-ils identiques ? Si oui, pourquoi opère-t-on cette distinction ? Si non, illustrez en quoi ils diffèrent par un exemple.
13. Quel est le sens du terme « négociation » lorsqu'il est question de protocoles réseau ? Donnez un exemple.
14. La figure 1.19 illustrait un service. Comprend-elle d'autres services implicites ? Si oui, à quel endroit ? Si non, pourquoi ?
15. Sur certains réseaux, la couche liaison de données gère les erreurs de transmission en demandant que les trames endommagées soient retransmises. Si la probabilité qu'une trame soit altérée est p , quel est le nombre moyen de transmissions requises pour envoyer une trame, sachant que les acquittements ne sont jamais perdus ?
16. Un système dispose d'une hiérarchie de n protocoles. Les applications génèrent des messages de M octets, et chaque couche ajoute un en-tête de h octets. Quelle est la portion de la bande passante occupée par ces en-têtes ?
17. Quelle est la principale différence entre TCP et UDP ?
18. La figure 1.25(b) présentait un réseau conçu pour pouvoir résister à une guerre nucléaire. Combien faudrait-il de bombes pour scinder le système en deux zones déconnectées ? Supposez que chaque bombe soit capable de détruire un nœud et toutes ses liaisons.
19. L'Internet double approximativement de taille tous les dix-huit mois. En l'absence de chiffres sûrs, on estimait le nombre de ses hôtes à 600 millions en 2009.

Sur la base de ces données, évaluez leur nombre en 2018. Pensez-vous que cela soit réaliste ? Justifiez votre réponse.

20. Lorsqu'un fichier est transféré entre deux ordinateurs, deux stratégies d'acquittement sont envisageables. Après que le fichier a été découpé en paquets, soit ce sont les paquets individuels qui sont acquittés par le destinataire mais pas le fichier entier, soit c'est le fichier qui est acquitté mais pas les paquets individuellement. Comparez ces deux approches.
21. Les opérateurs de téléphonie mobile ont besoin de savoir où se trouvent les téléphones (et en conséquence leurs détenteurs). Expliquez les avantages et les inconvénients pour les utilisateurs.
22. Quelle longueur en mètres occupe un bit dans la norme 802.3 d'origine ? Partez d'un débit de transmission de 10 Mbit/s et supposez que la vitesse de propagation dans le câble coaxial soit égale à $2/3$ de la vitesse de propagation de la lumière dans le vide.
23. En partant d'une image non compressée de 1 024 pixels sur 768 et en comptant 3 octets par pixel, combien de temps durerait sa transmission sur une liaison par modem à 56 kbit/s ? Sur une liaison par modem-câble à 1 Mbit/s ? Sur une liaison Ethernet à 10 Mbit/s ? Sur une liaison Ethernet à 100 Mbit/s ? Sur une liaison Gigabit Ethernet ?
24. Ethernet et les réseaux sans fil présentent des ressemblances et des différences. Une caractéristique d'Ethernet est qu'une seule trame peut être transmise à la fois. La norme 802.11 partage-t-elle cette propriété ? Justifiez votre réponse.
25. Citez deux avantages et deux inconvénients des normes internationales en matière de protocoles réseau.
26. Lorsqu'un système comprend une partie fixe (lecteur de CD-ROM, par exemple) et une partie amovible (le CD-ROM), sa normalisation est importante afin que ces parties, si elles sont fabriquées par des constructeurs différents, puissent fonctionner ensemble. En dehors de l'informatique, citez trois domaines dans lesquels de telles normes internationales ont été définies et citez-en trois autres dans lesquels il n'en existe pas.
27. Supposez que les algorithmes utilisés pour implémenter les opérations de la couche k aient changé. Quel serait l'impact sur les opérations des couches $k-1$ et $k+1$?
28. Supposez qu'il y ait un changement dans le service (l'ensemble d'opérations) fourni par la couche k . Quel serait l'impact sur les services des couches $k-1$ et $k+1$?
29. Citez plusieurs raisons pour lesquelles le temps de réponse d'un client peut être supérieur au délai minimal.
30. Dressez une liste des activités que vous accombez quotidiennement et qui mettent en œuvre des réseaux d'ordinateurs. En quoi votre vie se trouverait-elle changée si ces réseaux cessaient subitement d'exister ?

31. Déterminez quels types de réseaux sont déployés sur votre lieu d'études ou de travail et décrivez-les. Décrivez également leur topologie et les méthodes de commutation utilisées.
32. La commande *ping* permet d'envoyer un paquet de test vers une destination donnée, et de connaître le temps total pris par l'émission du paquet et son écho. Utilisez-la pour connaître le temps requis pour atteindre différents sites. À partir des informations obtenues, déterminez la durée de transit sur l'Internet dans une direction en fonction de la distance. Privilégiez les sites universitaires, car l'emplacement de leurs serveurs est connu très précisément. Par exemple, *berkeley.edu* se trouve en Californie, *mit.edu* au Massachusetts, *vu.nl* à Amsterdam, *www.usyd.edu.au* à Sydney et *www.uct.ac.za* au Cap.
33. Visitez le site de l'IETF, à l'adresse www.ietf.org, pour en apprendre davantage sur ses activités. Choisissez un projet qui vous intéresse et rédigez un rapport d'une demi-page sur le problème et la solution proposée.
34. L'Internet est composé d'un très grand nombre de réseaux, et sa topologie est déterminée par leur agencement. Quantité d'informations sont disponibles en ligne à ce sujet. Utilisez un moteur de recherche pour y accéder, et rédigez un bref rapport sur ce que vous aurez découvert.
35. Effectuez des recherches en ligne pour savoir quels sont les principaux points d'appairage actuellement utilisés pour router les paquets sur l'Internet.
36. Écrivez un programme pour implémenter un flot de messages allant de la couche supérieure à la couche inférieure du modèle de protocoles en sept couches. Votre programme doit comprendre une fonction de protocole distincte pour chaque couche. Les en-têtes de protocoles sont des séquences de 64 caractères au plus. Chaque fonction accepte deux paramètres : un message transmis par le protocole de couche supérieure (un tampon de caractères) et la taille du message. Elle attache l'en-tête au début du message, affiche le nouveau message sur la sortie standard, puis appelle la fonction de protocole de la couche inférieure. L'entrée du programme est un message d'application (une suite de 80 caractères ou moins).