

Cahiers **GUT** *enberg*

☞ SYSTÈMES D'INFORMATION SUR INTERNET
☞ Luc OTTAVI

Cahiers GUTenberg, n° 19 (1995), p. 3-26.

<http://cahiers.gutenberg.eu.org/fitem?id=CG_1995__19_3_0>

© Association GUTenberg, 1995, tous droits réservés.

L'accès aux articles des *Cahiers GUTenberg*

(<http://cahiers.gutenberg.eu.org/>),

implique l'accord avec les conditions générales

d'utilisation (<http://cahiers.gutenberg.eu.org/legal.html>).

Toute utilisation commerciale ou impression systématique

est constitutive d'une infraction pénale. Toute copie ou impression

de ce fichier doit contenir la présente mention de copyright.

Systèmes d'information sur Internet*

Luc OTTAVJ

INRIA Sophia-Antipolis,
2004 route des Luciolles, F-06565 Valbonne cedex
`luc.ottavj@inria.fr`

1. Introduction

Internet se présente comme un réseau de réseaux interconnectant à travers les cinq continents plusieurs millions de machines. Ces interconnexions sont réalisées avec une grande variété de médias à des débits diversifiés.

On trouve sur Internet un nombre impressionnant de services et d'applications dans des domaines tels que la messagerie, l'accès à des machines distantes, le transfert de fichiers, la découverte d'informations et le travail en groupe. Les logiciels correspondants sont, pour la plupart, placés dans le domaine public et disponibles gratuitement avec les sources des programmes et la documentation associée.

La présence sur le réseau de l'essentiel du potentiel de recherche publique mondiale et l'existence d'un réel esprit d'entraide et de coopération font que ces services, souvent préfigurateurs, sont évolutifs, fiables et maintenus de façon coopérative.

La viabilité d'Internet repose sur le respect de normes élaborées en son sein selon une procédure simple, ouverte et efficace.

2. Architecture des logiciels

Les informations transitant sur Internet vont traverser une architecture en cinq couches. Dans cette organisation, maintenant traditionnelle en télé-informatique, chaque couche demande et reçoit des services de la couche immédiatement inférieure et correspond avec son homologue dans l'équipement distant selon un protocole définissant les dialogues et le format des messages échangés.

*. Cette note correspond à l'exposé fait lors de la *journée GUTenberg* sur la diffusion électronique des documents (Nanterre, 19 janvier 1995), exposé lui-même basé sur un cours donné lors de l'école Inria sur le traitement électronique du document (Aix-en-Provence, octobre 1994). Cet article est déjà paru sous le titre « Internet et outils de recherche documentaire » dans *Le traitement électronique du document* (textes réunis par Jean-Claude Le Moal et Bernard Hidoine), ADBS Editions, 1994, pages 117-148. Il est reproduit ici avec l'aimable autorisation de ADBS Editions que nous tenons à remercier {NDLR}.

Cette architecture modulaire permet de gérer efficacement l'hétérogénéité. Ainsi, par exemple, il n'y aura aucune modification à faire dans les logiciels d'une application selon que les machines la mettant en œuvre sont raccordées au réseau téléphonique commuté ou bien à un réseau local ou bien, enfin, l'une au réseau téléphonique et l'autre au réseau local.

2.1. Couches physique et liaison de données

Ces couches ont pour but d'acheminer l'information entre deux ou plusieurs stations situées sur le même support physique.

Les réseaux locaux rendent ce type de service sur une portée géographique limitée et avec une grande diversité de protocoles, de médias, d'architectures et de débits. Pour interconnecter ces réseaux locaux, on utilise des services ou des médias tels que les liaisons spécialisées, le réseau téléphonique commuté analogique ou numérique, des transmissions satellite, hertziennes, radio ou laser.

Sur le réseau téléphonique commuté, deux modes sont possibles selon que l'on a un abonnement normal ou Numéris.

- Sur un abonnement normal, on pourra utiliser des coupleurs asynchrones standard que l'on connectera à des modems analogiques. On trouve couramment des modems respectant la norme de modulation V32bis qui permet d'atteindre 14 400 bits/s lorsque la ligne est de bonne qualité. La disponibilité de modems agréés en Europe pour V34 est imminente, le débit sera alors de 28 400 bits/s. Les modems actuels permettent de faire de la correction d'erreurs et de la compression de données qui permet d'augmenter ces débits jusqu'à un facteur quatre, au prix d'une augmentation des délais de transmission.
- Sur un abonnement Numéris qui coûte environ 400 F/mois, le débit possible est de 2 fois 64 kbits/s. On peut utiliser des coupleurs spécifiques RNIS ou bien des coupleurs standard synchrones ou asynchrones, à condition de se doter de boîtiers d'adaptation. Pour les modems numériques, on pourra utiliser des coupleurs synchrones.

À ce niveau, on utilise des protocoles tels que HDLC (*High Data Link Control*), SLIP (*Serial Line IP*), ou PPP (*Point to Point Protocol*).

HDLC est le protocole de niveau 2 utilisé conformément à la norme CCITT X25 pour l'accès aux réseaux publics de transmission de données par paquets comme Transpac. Dans le but d'améliorer les performances, ce sont souvent des versions simplifiées de HDLC qui sont utilisées sur Internet.

SLIP est un protocole sommaire pour transporter des informations venant de la couche réseau IP, sur des liaisons série ; il est particulièrement utilisé pour relier un micro-ordinateur à Internet par un coupleur asynchrone, lui-même

relié à un modem analogique connecté au réseau téléphonique commuté. Le principal avantage de SLIP est sa simplicité. Ce protocole tend vers la désuétude en raison de limitations fonctionnelles.

PPP propose un service d'encapsulation pour n'importe quel protocole de niveau 3 (Réseau). Il marche sur coupleurs synchrones ou asynchrones et utilise un protocole de transport de type HDLC : même format de trames, détection d'erreurs par FCS, mais pas de correction des erreurs par retransmission, pas de détection de perte ou de duplication de blocs, pas de contrôle de flux. PPP propose deux modes d'authentification des correspondants : PAP (*PassWord Authentication Protocol*) et CHAP (*Challenge Handshake Authentication Protocol*).

Le correspondant qui veut prouver son identité en mode PAP envoie à l'initialisation de la connexion son identifiant et un mot de passe associé. Le correspondant qui doit s'identifier en mode CHAP reçoit à intervalles réguliers une chaîne de caractères variable et doit retourner le résultat du cryptage de cette chaîne avec une clé secrète connue par les deux correspondants et non transmise sur la ligne.

Grâce à LCP (*Link Control Protocol*), PPP permet aux deux extrémités de négocier certains paramètres comme la longueur des trames, le mode d'authentification et, en mode asynchrone, la liste de caractères de contrôle, comme *Xon*, *Xoff* dont il faut assurer la transparence lorsqu'ils sont présents dans les données.

Un protocole de contrôle est associé à chaque protocole de niveau réseau transporté par PPP. Le protocole de contrôle associé à IP est IPCP. Ce protocole permet la négociation des adresses IP et la compression de certains champs des en-tête IP ou TCP.

Ces diverses possibilités d'authentification et de négociation d'options facilitent l'administration, rendent viable l'utilisation du réseau téléphonique commuté et permettent l'extension d'Internet à des applications et des utilisateurs domestiques. Cette grande diversité de débits au niveau des couches basses explique la grande disparité dans les performances selon la destination atteinte.

2.2. Internet Protocol : IP

Le protocole IP est un service de niveau réseau. Orienté datagramme, il permet d'envoyer des messages de longueur variable à un correspondant distant identifié par son adresse. Les messages, expédiés sans opération préalable de connexion, sont acheminés de façon indépendante en suivant éventuellement des chemins différents, si bien qu'ils peuvent arriver dans le désordre. IP ne prévoit aucun mécanisme de détection ou de correction d'erreurs. Seul un dispositif de *checksum* sur l'en-tête des datagrammes permet d'accorder foi aux informations qu'elle contient. De même, rien n'est prévu à ce niveau pour détecter ou corriger d'éventuelles pertes de messages. Un champ « Durée de vie », décrémenté lors de la traversée de chaque routeur, permettra de limiter la diffusion du datagramme. L'existence de

ce champ permet notamment d'éviter les congestions en cas d'incohérence dans les informations de routage. En revanche, IP assure deux services nécessaires à la résolution du problème de l'interconnexion : l'adressage et la fragmentation.

Les adresses IP sont codées sur 4 octets et se décomposent en une partie identifiant le réseau et une partie identifiant la machine sur le réseau. La partie réseau va être codée sur un, deux ou trois octets selon que l'on est en classe A (premier octet inférieur à 128), B (premier octet compris entre 128 et 191) ou C (premier octet compris entre 192 et 223). Ce qui laisse trois, deux ou un octet pour coder l'adresse de la machine sur le réseau. L'unicité des adresses IP sur le plan mondial est facilitée par leur caractère hiérarchique. Au tout début, un service central localisé aux Etats-Unis, le NIC (*Network Information Center*), attribuait des adresses réseau à la demande. Actuellement ce service a été décentralisé sur les divers continents et pays à qui l'on a confié la gestion d'une plage d'adresses. Ce rôle de *Network Information Center* est assuré en France actuellement par l'INRIA.

Ayant obtenu du NIC une adresse réseau unique, chaque administrateur assure l'unicité des adresses pour les machines de son réseau. Les adresses dont le premier octet est supérieur ou égal à 224 sont dites de classe D. Elles correspondent à des adresses de diffusion sur Internet : les informations postées à ces adresses sont diffusées à toutes les machines d'Internet situées dans un rayon plus ou moins grand. Ces adresses sont notamment utilisées par les services multimédia de multiconférence.

L'autre service de base rendu par IP est celui qui consiste à fragmenter les informations pour prendre en compte les limitations sur la longueur des trames imposées par les couches basses. Ces limitations sont en effet incontournables : ainsi envoyer sur FDDI des trames de plus de 4 500 octets, c'est s'exposer à ce que des stations n'arrivent pas à répéter les messages sans perdre de bits. De même, transmettre des trames d'un million de bits sur une ligne où la probabilité d'erreur au niveau bit est de un millionième, c'est s'exposer à avoir, à coup sûr, une erreur par trame et donc à bloquer un protocole comme HDLC qui fait de la détection d'erreur au niveau trame et de la correction par retransmission de la trame.

Au cours de sa traversée d'Internet, un datagramme IP va parcourir des réseaux, remonter dans des routeurs qui, en fonction de l'adresse de destination et après avoir consulté leurs tables de routage, l'enverront vers telle ou telle interface en le découpant en datagrammes de taille compatible avec le réseau à traverser. Les informations de fragmentation contenues dans les divers datagrammes permettront à la couche IP chez le destinataire de reconstituer le datagramme original.

Ce mode de cheminement explique que, sur Internet, la connectivité est transitive. Pour être raccordé à ce réseau, il suffit, muni d'une adresse IP officielle, d'établir une liaison avec quelqu'un qui est lui-même raccordé. Bien entendu, il

faudra que les routeurs traversés acceptent de l'être et véhiculent les informations de routage concernant le réseau nouvellement raccordé.

2.3. Couche transport : TCP/UDP

À ce niveau sont essentiellement fournis, d'une part, un service d'adressage complémentaire à l'adresse IP permettant d'atteindre une application sur la machine et, d'autre part, des services permettant de pallier les inconvénients de IP. L'adressage se matérialise par un numéro de port codé sur deux octets. Ces numéros de port sont alloués de façon statique ou distribués dynamiquement par des serveurs annuaires qui sont accessibles sur des numéros de port connus. Deux services existent au niveau transport.

UDP (*User Datagram Protocol*) : c'est, tout comme IP, un *Datagram Protocol* ; pour communiquer avec son correspondant, on n'a pas à établir de connexion préalable. Ceci permet l'utilisation des adresses de diffusion générale où le même message est envoyé en une fois à l'ensemble des machines du réseau.

La présence d'un champ de contrôle permet au destinataire de détecter d'éventuelles erreurs de transmission.

TCP (*Transmission Control Protocol*) : c'est un protocole orienté connexion. Pour pouvoir communiquer avec un correspondant distant, il va falloir établir une connexion avec lui. En cas d'acceptation, une liaison bidirectionnelle va être établie. Cette liaison va permettre des échanges simultanés dans les deux sens. La circulation d'acquittements va permettre la retransmission de blocs d'informations erronés ou perdus, ainsi que la mise en place d'un mécanisme de contrôle de flux. On voit qu'ainsi les insuffisances de IP sont comblées par un service où les contrôles sont effectués de bout en bout.

2.4. Couche application

À ce niveau, on trouve une grande variété de protocoles qui vont s'appuyer sur TCP ou UDP et travailler la plupart du temps en mode client/serveur. Chaque application va échanger des messages formatés en respectant un protocole particulier. Ces protocoles sont en règle générale d'une grande simplicité. Les applications bâties dans le cadre du modèle OSI s'intègrent aussi dans cette architecture par empilage de la couche transport OSI sur TCP/UDP. Actuellement le modèle a tendance à se simplifier par connexion directe de la couche application OSI sur TCP ou UDP.

3. Organisation d'Internet

Internet n'est pas seulement un réseau physique ; c'est aussi une organisation permettant la définition de normes et l'échange de toute information utile à la communication.

À la base on trouve *Internet Society*, association à laquelle chacun peut s'affilier. L'IETF (*Internet Engineering Task Force*) – qui a pour but d'élaborer la standardisation, de servir de forum pour l'échange d'opinions et d'expériences et qui enfin émet des recommandations à l'organe décisionnaire de l'IAB – est organisée en neuf groupes de travail dont les animateurs constituent l'IESG (*Internet Engineering Steering Group*). L'IETF tient des réunions plénières plusieurs fois par an. L'IAB (*Internet Architecture Board*) vient compléter l'édifice et constitue la structure décisionnaire. Sa fonction essentielle est d'entériner les standards proposés par l'IETF. Actuellement la présidence de l'IAB est assurée par Christian Hùitema, directeur de recherche à l'INRIA.

Les normes et recommandations élaborées par Internet sont contenues dans des RFC (*Request for Comments*) qui sont identifiés par un numéro. Les RFC sont élaborés selon une procédure décrite dans le RFC1602. Certains RFC décrivent des standards Internet et constituent le noyau dur de la normalisation du réseau ; d'autres sont classés informationnels, prototypes, expérimentaux ou historiques.

Pour ce qui concerne l'élaboration de standards, la procédure permet à quiconque ayant identifié un problème ayant trait à Internet de proposer à l'IESG la création d'un groupe de travail ; il devra rédiger une charte définissant l'objet des travaux et proposant un responsable de groupe. Le groupe rattaché à un secteur de l'IESG travaillera par messagerie électronique, lors des réunions trisannuelles de l'IETF. Le groupe sera librement ouvert, les archives des messages et des documents concernant ses travaux seront accessibles sur le réseau.

Dès que ses travaux auront convergé, le groupe proposera au responsable de groupe de l'IETF un document qui pourra être un *Internet Draft* ou un *RFC expérimental*. Ce document sera relu, expertisé et soumis aux questions des membres de l'IETF pendant une période de quinze jours. L'IESG votera ensuite l'élévation du document au rang de *Proposed Standard*. Dans une période de six mois à deux ans, le document, éventuellement affecté de modifications mineures, passera à l'état de *Draft Standard*, à condition qu'au moins deux implémentations du protocole ou de l'algorithme soient proposées. Il retournera à l'état de *Proposed Standard* si les modifications sont majeures.

Au bout d'une deuxième période de six mois à deux ans, le document sera élevé au rang de *Standard* à condition que plusieurs implémentations interopérantes aient été réalisées. Tout comme le passage en *Proposed Standard*, l'élévation à l'état de *Draft* et de *Standard* est soumis aux mêmes procédures d'examen par des experts, d'appel général aux observations et de vote par l'IESG.

Tout *Draft Standard* n'accédant pas au rang de *Standard* est déchu et classé comme historique.

4. Internet en France

Internet est accessible en France à travers les offres de divers opérateurs.

4.1. Réseaux régionaux

Depuis les années 90, certaines régions ont mis en place des réseaux régionaux ouverts aux organismes ayant des activités de recherche et de développement, les buts poursuivis étant de faciliter les relations entre centres de recherche et les transferts de technologie avec le tissu industriel. Cela s'est traduit par la mise en place d'infrastructures de communication régionales à 2 mégabits/s, en général, sur lesquelles les abonnés viennent se raccorder à des débits allant de 64 kbits/s à plusieurs mégabits/s.

Ces réseaux utilisent IP comme protocole de base et sont exploités actuellement par France Télécom. Les coûts sont forfaitaires, dépendent du débit de raccordement et sont indépendants des volumes transmis et de la durée d'utilisation.

Les collectivités territoriales financent ces réseaux soit en intervenant en tiers payeurs sur les coûts d'abonnement, soit, comme en région PACA, en finançant l'épine dorsale du réseau qui constitue une véritable autoroute électronique régionale et dont l'existence et le développement sont vus en termes d'aménagement du territoire.

Dans tous les cas, la part qui incombe à l'utilisateur représente sensiblement la moitié des coûts réels. Compte tenu de différences structurelles entre les divers réseaux, et du fait que les négociations avec le fournisseur ont eu des succès divers, les coûts sont variables selon les régions et peuvent aller du simple au double : ainsi la connectivité régionale à 64 kbits coûte en région PACA 21 000 F d'installation et 3 600 F d'abonnement mensuel ; elle peut atteindre 15 000 F ailleurs.

Dans certaines régions comme Provence-Alpes-Côte-d'Azur (R3T2), précurseur en la matière, le réseau est vu comme un facteur de développement économique et doit être largement ouvert aux entreprises privées. Des actions d'accompagnement ont été prises, telles que la mise en place de deux supercalculateurs et d'un service infothèque visant à promouvoir l'essor de bases d'information et leur utilisation à travers des interfaces normalisées.

Ces réseaux régionaux ont longtemps tiré leur connectivité internationale des possibilités offertes par certains abonnés qui, comme l'INRIA, ont largement accordé un droit de transit sur leurs liaisons internationales vers l'Europe et les Etats-Unis (réseau de la NSF).

4.2. Le réseau Renater

Parallèlement à ces efforts régionaux, un certain nombre d'organismes comme le CEA, le CNES, le CNRS, EDF, l'Éducation nationale et l'INRIA ont convenu

de remplacer leurs réseaux privés par une infrastructure commune : le réseau national de télécommunications pour la technologie, l'enseignement et la recherche (RENATER).

Réunis dans un groupement d'intérêt public, ces organismes ont confié à France Télécom la mise en place d'une infrastructure de communication interconnectant les plaques régionales existantes et amenant l'infrastructure nationale dans un site représentatif pour les régions non pourvues d'un réseau régional. De nombreux grands organismes ont rejoint le GIP, ou sont en voie de le faire.

Cette épine dorsale nationale est en général à 2 mégabits/s, l'axe Paris-Lyon-Marseille- Montpellier étant à 34 mégabits/s. Dans le même temps un nœud de transit international a été mis en place qui assure des liaisons à 3,5 mégabits/s vers les États-Unis (NSFNET) et à 2,3 mégabits/s vers l'Europe dans le cadre du réseau Ebone. Ce réseau européen Ebone, bâti de façon pragmatique par des initiatives des diverses communautés intéressées par le développement de la connectivité IP en Europe, est actuellement en voie de régression au profit d'Europenet, réseau soutenu et largement financé par la CEE, enfin ralliée à des solutions basées sur IP. Ebone et Europenet sont bien entendu interconnectés.

Le réseau Renater et ses prolongements internationaux sont accessibles sans coûts supplémentaires par tous les organismes relevant des membres fondateurs et raccordés aux plaques régionales. Le réseau est aussi ouvert à tout site raccordé à un réseau régional qui, après agrément par le GIP, peut soit entrer dans celui-ci, soit acquitter une redevance mensuelle. À titre d'exemple, en 1994 il est demandé 5 000 F par mois pour la connectivité nationale et internationale d'un site non-membre du GIP Renater et raccordé à 64 kbits/s à une plaque régionale. Cette redevance s'ajoute bien entendu à la redevance d'abonnement à la plaque régionale qui couvre, elle, le trafic régional.

4.3. L'offre privée

À côté de cette offre reposant sur des financements publics et plutôt orientée vers les hauts débits, il existe une offre émanant d'opérateurs privés qui est à l'heure actuelle caractérisée par des débits plus faibles et des services modulables. Les débits offerts varient en général de 1 200 bits/s à 64 kbits/s. Ces services sont modulables depuis le simple accès en mode asynchrone à une machine abritant des boîtes à lettres, jusqu'à la connectivité IP avec tout Internet sans limitation de volume, en passant par la connectivité IP avec taxation des échanges au volume ou à la durée.

FNET : ce réseau est proposé par l'AFUU (Association française des utilisateurs d'Unix). Depuis longtemps, l'AFUU proposait un service de messagerie à couverture mondiale basé sur des liaisons en mode asynchrone, le plus souvent sur des liaisons téléphoniques commutées et utilisant le protocole UUCP. Depuis quelques années, FNET utilise IP et propose une gamme

de services allant de l'échange de messages à la connectivité complète sur Internet.

Le réseau est bâti en étoile autour d'un nœud central situé en région parisienne. De là part une liaison vers Amsterdam qui est le point de concentration européen des réseaux Unix nationaux fédérés dans EUNET. Une liaison Amsterdam États-Unis permet le raccordement à UUNET qui assure la fédération mondiale de ces réseaux.

Ce réseau est interconnecté à Internet à Paris, Amsterdam et aux États-Unis. Les abonnés à FNET acquittent des redevances qui sont fonction du service et des volumes échangés. Ils ont en outre à supporter les frais liés au matériel de raccordement (routeur, modem) et les frais téléphoniques pour rejoindre le nœud central parisien. Ces liaisons entre l'abonné et FNET peuvent se faire par RTC (réseau téléphonique commuté), Numéris, Transpac ou par liaison spécialisée.

Oleane : depuis 1993, cette société du groupe Apysoft associée à la société britannique PIPEX propose des services d'accès à Internet. Les abonnés accèdent, à leurs frais, au point de présence le plus proche (Vincennes...) par réseau téléphonique commuté analogique ou numérique, liaison spécialisée, ou par Transpac. Ils ont accès à tout site raccordé à Internet moyennant le paiement d'une redevance mensuelle fixe pour les accès permanents ou modulée en fonction du temps d'utilisation pour les accès par réseau commuté.

France-Télécom Transpac : depuis l'été 1994, Transpac propose un service d'accès à Internet. Les débits possibles sont de 9 600, 14 400 et 19 200 bits/s.

Deux modes de facturation sont prévus. Le premier mode est forfaitaire. Moyennant le paiement d'une somme fixe à l'installation et d'une redevance forfaitaire mensuelle, on bénéficie de l'installation d'un routeur, d'un numéro IP et d'une liaison spécialisée reliée au nœud le plus proche du réseau.

On a le droit d'échanger les volumes que l'on veut avec tout site raccordé à Internet. Le deuxième mode est lié aux consommations effectives. Moyennant le paiement de frais de raccordement et d'une redevance mensuelle, on pourra échanger du trafic avec Internet. Ce trafic sera taxé au volume. Dans les deux modes, l'utilisateur bénéficie, dans le cadre de l'abonnement, de la fourniture et de l'installation d'un routeur, d'un numéro IP et d'une liaison avec le plus proche point d'accès au réseau. Des conditions particulières existent pour ceux qui sont déjà équipés d'un accès X25 à Transpac.

Autres : un certain nombre d'autres sociétés comme Calvacom, French Data Networks, Francenet, Compuserve, ATT, proposent des services d'accès à Internet à bas débit, via RTC, ou Transpac. La taxation se fait en général à la durée avec abonnement mensuel. Une mention particulière revient au service kiosque Internet, proposé par Francenet, accessible par le RTC (Slip/PPP) à des numéros Audiotel (3668) où les seuls coûts supportés par l'utilisateur

sont ceux versés à France Télécom (2,19 F TTC/mn), ce dernier reversant à Francenet une partie des sommes perçues.

4.4. Comment s'y retrouver ?

Devant le foisonnement naissant de l'offre, il est important de bien connaître les caractéristiques exactes du service proposé et notamment de se poser les questions suivantes :

- s'agit-il d'un accès IP à tout Internet, ou bien seulement à quelques abonnés ou à quelques applications?
- cet accès est-il limité à une machine ou ouvert à toutes les machines du client localisées sur le site? sur d'autres sites?
- cet accès pourra-t-il être partagé par d'autres entités juridiques localisées sur le même site? sur des sites différents?
- quel est le débit?
- la taxation est-elle forfaitaire? au volume? à la durée?
- est-ce que les coûts proposés incluent les dépenses télé-informatiques pour la liaison entre le client et le point d'accès au réseau? Si non, quels sont ces coûts? Où est localisé ce point d'accès?
- est-ce que les coûts d'installation incluent la fourniture d'un routeur? de modems?
- est-ce que des services d'assistance au démarrage existent? couvrent-ils aussi l'aide à la configuration d'applications comme la messagerie?
- est-ce qu'un numéro IP et un nom de domaine officiels sont fournis?

5. Applications traditionnelles

5.1. Messagerie

Deux types de systèmes d'information basés sur l'envoi de messages cohabitent sur Internet. Ils permettent d'envoyer des messages dans des boîtes à lettres nominatives ou bien au contraire dans des journaux publics propagés sur Internet.

5.1.1. Messagerie par boîtes

Ce mode permet l'envoi de messages à des adresses électroniques qui sont de la forme **Personne@domaine**, où **Personne** est en général le nom d'un utilisateur et **domaine** le nom d'une machine. Le message sera déposé dans la boîte ouverte à la personne donnée sur la machine donnée. L'utilisateur propriétaire de la boîte consultera son courrier à l'aide d'outils appropriés (Mmail, mh, xmh, eudora, popmail, etc.) et pourra répondre aux émetteurs, faire suivre à des tiers, classer ou effacer le message.

Le protocole d'application utilisé pour acheminer les messages s'appelle SMTP (*Simple Mail Transmission Protocol*) et est défini dans le RFC821. Il s'appuie sur un service TCP au niveau transport. SMTP permet l'échange de messages dont le format initialement défini dans le RFC822 ne permettait de transporter que du texte ASCII. Ces limitations ont été levées par les extensions MIME (*Multimedia Internet Mail Extensions*) définies dans le RFC 1341. Un message en format MIME peut être composé d'une ou plusieurs parties contenant du texte, de l'image, du son numérisé ou des images animées. À défaut des données elles-mêmes, c'est un pointeur (**Machine, Chemin, Mode d'accès**) vers ces données qui sera précisé. Le format propre à chaque partie est absolument quelconque. Il est annoncé dans le champ **Contents-Type**. Il sera possible d'encoder **uuencode**, **base64**, etc. chaque partie de façon à pouvoir la transporter dans des messages standard sur 7 bits.

Il existe des protocoles permettant d'accéder à des boîtes aux lettres distantes qui s'appuient sur TCP. POP (*Post Office Protocol*) permet de transférer le contenu d'une boîte à lettres sur un ordinateur distant. IMAP (*Interactive Mail Access Protocol*) permet une consultation interactive d'une boîte à distance avec possibilité de détruire, classer ou répondre aux messages. Ces protocoles permettent à des postes connectés épisodiquement à Internet d'avoir accès à la messagerie. Ils supposent qu'une boîte à lettres au nom de l'utilisateur soit ouverte sur une machine connectée de façon permanente. POP est typiquement adapté aux postes nomades comme les micro-ordinateurs portables reliés au réseau téléphonique commuté; la connexion en TCP/IP à la machine abritant la boîte se faisant juste pour vider celle-ci et injecter les réponses. Le travail de lecture, de classement du courrier et de préparation des réponses ne nécessite pas de connexion à Internet et peut se faire en n'importe quel lieu.

Il est possible à chacun d'établir des listes de diffusion et de leur affecter une adresse électronique. Ces listes sont largement utilisées dans le cadre de groupes de travail et de clubs divers.

Les messages reçus à certaines adresses électroniques peuvent être traités par un programme. Cela permet de réaliser des fonctions automatiques de répondeur, de distributeur de fichiers et de documents électroniques ou d'inscription à des listes de diffusion.

Il est aussi possible d'atteindre par messagerie des abonnés au fax, ainsi que des destinataires accessibles dans d'autres systèmes comme X400, BITNET. En effet,

des programmes passerelles faisant le relais entre ces divers mondes existent dans le domaine public et la majorité des opérateurs des réseaux constituant Internet assurent ce service sur des machines raccordées aux divers systèmes.

5.1.2. Les news

Le système de *news* constitue sans doute l'application la plus importante d'Internet en termes d'information, de coopération et d'assistance mutuelle. Ce système donne la possibilité de lire et de poster des messages qui vont être classés dans des forums ou *news-groups*. Chaque *news-group* est relatif à un sujet technique, philosophique, culturel ou d'actualité. Les articles postés sont typiquement des opinions, des demandes d'aide, des annonces de produits, de solutions, de manifestations, de documents. Il existe actuellement environ 3 000 *news-groups* dans lesquels arrivent 2 Go d'informations nouvelles par mois, mais ces chiffres sont en perpétuelle augmentation.

Les articles sont stockés sur un serveur et consultés à l'aide de logiciels clients tournant sur le serveur ou sur des postes de travail distants. Le protocole utilisé au niveau application est alors NNTP ; il s'appuie sur la couche TCP. Tout article posté par un client arrive sur le serveur de *news* qui le stocke sur disque et le propage éventuellement à des serveurs voisins. Les articles se propagent sur Internet de serveur en serveur. Chaque article est muni d'un identifiant unique qui permet aux serveurs de rejeter les articles déjà reçus. Le caractère hiérarchique des noms de *news-groups* est exploité dans ce processus de diffusion, et certains *news-groups* sont propagés au niveau du site, de l'organisme, de la région, du pays, du continent ou du monde entier.

La création de *news-groups* nationaux et internationaux est soumise à discussion et vote préalable. Chaque *news-group* est doté d'une charte qui en définit le but. Il existe aussi des règles générales d'éthique définissant les droits et devoirs des utilisateurs. Pour la plupart des *news-groups* techniques, une personne ou un groupe se charge de publier périodiquement un article FAQ (*Frequently Asked Questions*), faisant le point sur les principaux sujets abordés. Ces articles constituent un état de l'art sur le sujet.

Les envois dans certains *news-groups* sont filtrés par une personne, le modérateur. Les messages postés dans ces *news-groups* sont en fait envoyés par messagerie au modérateur, qui effectue le postage après en avoir contrôlé le contenu.

5.2. Accès à distance

Ces services donnent la possibilité à un utilisateur de lancer des applications sur des machines distantes et d'interagir avec elles.

5.2.1. Protocole TELNET

Cette application permet à un client d'établir une connexion avec une machine distante et d'y ouvrir une session de travail. Le client sera vu comme un terminal alphanumérique. Le protocole TELNET défini dans le RFC854 s'appuie sur TCP. Des clients et serveurs TELNET sont disponibles sur la quasi-totalité des systèmes d'exploitation.

5.2.2. X-WINDOW

Ce produit du domaine public écrit par le MIT donne la possibilité à un programme tournant sur une machine distante d'effectuer des affichages graphiques (mode bitmap) sur un écran distant et de recevoir des interactions venant du clavier et de la souris attachés au poste de travail.

Un serveur tourne sur le poste de travail et gère l'écran, le clavier et la souris. Il communique avec des programmes locaux ou distants : les clients X-Window. Ces derniers peuvent faire des affichages graphiques dans des fenêtres ouvertes sur les écrans distants et recevoir des interactions venant du clavier ou de la souris attachés au poste de travail. La liaison entre le client et le serveur s'établit en TCP. Un gestionnaire de fenêtres s'exécutant sur le poste de travail ou sur une machine distante va permettre de déplacer, agrandir, occulter, créer et détruire des fenêtres. X-Window a été développé sur la plupart des systèmes d'exploitation. Des serveurs sont notamment disponibles sur des matériels dédiés : des terminaux X-Window, mais aussi sur PC, Mac et machines Unix. Il existe une multitude de clients X-Window dans le domaine public et la grande majorité des logiciels clients pour les applications décrites dans ce chapitre existent en version X-Window, sous Unix principalement.

5.3. Transferts de fichiers

FTP (*File Transfer Protocol*) permet à un client d'établir une connexion TCP avec une machine distante sur laquelle s'exécute un serveur FTP. Après s'être « authentifié », le client va pouvoir lancer des commandes qui vont lui permettre de naviguer dans les hiérarchies de fichiers locales et distantes et d'effectuer des transferts de fichiers dans les deux sens.

On peut accéder à un serveur FTP en mode anonyme : le client ne pourra alors accéder qu'à une partie de l'espace disque et acceptera l'ouverture de sessions distantes sans authentification préalable.

FTP est largement utilisé sur Internet pour échanger des programmes, des données, des documents électroniques. Le mode anonyme est utilisé pour diffuser des informations de toutes sortes, placées dans le domaine public. Le nombre de serveurs accessibles en mode anonyme est pléthorique et des services comme Archie facilitent la navigation dans cette jungle. Archie est un service consistant à bâtir sur une machine un index avec tous les noms de fichiers stockés sur les

serveurs anonymes notoires. Il est possible à un client d'interroger l'un des rares serveurs Archie sur Internet : il lui suffit de donner une chaîne de caractères pour obtenir la liste des serveurs contenant des fichiers dont le nom contient la chaîne indiquée, et rapatrier le fichier de son choix à partir d'un des serveurs indiqués.

Un autre système notable est Alex. Ce système permet de factoriser les transferts opérés depuis les machines d'un site. Alex donne aux machines locales une vision NFS de l'espace accessible sur des serveurs distants, en FTP anonyme notamment. Les postes de travail se déplacent dans l'espace Alex avec les commandes Unix traditionnelles et voient cet espace comme un disque local. La machine Alex va interpréter ces demandes et établir des connexions FTP avec les machines distantes pour récupérer les répertoires ou fichiers distants. Les informations ainsi récupérées vont être stockées un certain temps sur la machine Alex. Avec ce mécanisme de cache, on voit que si le même fichier ou répertoire est demandé plusieurs fois, ils se trouvera dans le cache et ne sera transféré qu'une fois en FTP.

6. Outils d'aide à la découverte d'informations

Avec la prolifération de sources d'informations, on a vu fleurir au cours de ces dernières années des outils permettant de découvrir, rapatrier et visualiser des informations localisées sur Internet. Les systèmes les plus remarquables sont :

- *Gopher* qui permet une navigation par menus ;
- *WWW* qui permet une navigation en mode hypertexte ;
- *Wais* qui permet de retrouver des documents à partir des mots qu'ils contiennent.

Il existe des liens entre ces trois systèmes : ainsi, un lien Gopher peut pointer vers un index Wais, un lien hypertexte peut pointer vers un serveur Gopher ou un index Wais, et un index Wais peut être constitué avec des documents WWW ou des entrées de menus Gopher.

6.1. *Wide Area Information Servers* : WAIS

À l'origine, on trouve une initiative de Thinking Machines Corporation, Apple Computers et Dow Jones, qui a consisté à mettre dans le domaine public des logiciels permettant de :

- constituer des index à partir de mots contenus dans des fichiers de formats divers (texte ASCII, PostScript, DVI, MIME, SGML, GIF, TIFF...);
- interroger ces index à distance en fournissant une liste de mots, puis en élargissant au besoin la recherche aux documents voisins par un mécanisme dit de *relevance feedback*; ce mécanisme consiste à inclure dans la question des parties de documents déjà trouvées ;

- rapatrier les documents sélectionnés et les visualiser sur son poste de travail avec les filtres de son choix.

Les logiciels d'interrogation sont bâtis sur le modèle clien/serveur et s'appuient sur une variante du protocole ANSI Z39.50 avec un transport TCP. Ils sont implémentés sur divers systèmes d'exploitation (MS-DOS, MacOS, VMS, Unix) et sous diverses interfaces homme-machine : Terminal ASCII, X-Window, Motif, Sunview, Next, Windows, Mac-OS.

De nombreux organismes utilisent Wais pour rendre accessibles des informations sur Internet. Actuellement on compte plus de 500 sources annoncées dans l'annuaire mondial et concernant les domaines et formats les plus divers tels que :

- catalogues de bibliothèques ou de bibliographies ;
- textes intégraux de rapports techniques, de documents de normalisation, d'ouvrages ;
- annuaires de personnes, de réseaux, de domaines ;
- archives des serveurs FTP notoires et de certains news-groups ;
- programmes sources ;
- images ;
- documents multimédia en format MIME ou HTML.

À toutes ces sources déclarées dans le répertoire des serveurs s'ajoutent toutes celles qui sont développées à des fins internes telles qu'archives de messages, documentation interne, sources de programmes, etc.

Les sources Wais sont accessibles par d'autres logiciels d'aide à la découverte d'informations comme Gopher et WWW, ou par d'autres moyens comme les environnements Emacs ou plus simplement la messagerie électronique.

Le développement du produit est actuellement assuré par le CNIDR (*Clearinghouse for Network Information Discovery and Retrieval*) de la NSF.

6.1.1. Protocole Z39.50

Le protocole Z39.50 a été défini en 1988 (version 1) par le NISO qui est un institut de normalisation américain accrédité par L'ANSI. Ce protocole a été redéfini en 1992 (version 2), afin d'être d'une part compatible avec le standard ISO *Search and Retrieval* qu'il est censé englober et, d'autre part, de recevoir des améliorations fonctionnelles. Le protocole utilisé par Wais aujourd'hui est une version étendue du protocole Z39.50-88. L'évolution vers la version 2 (Z39.50-92) est prévue par le CNIDR.

Ce protocole, conçu pour uniformiser les dialogues entre un client et un serveur dans le cadre d'applications documentaires, s'inscrit, dans la philosophie du modèle OSI de l'ISO, comme un protocole d'application. Il s'appuie sur des dialogues en mode connecté et fournit des services pour :

- l'initialisation avec contrôle de la communication client/serveur INIT ;

- le déclenchement de recherches (*Search*);
- la récupération de résultats (*Present*);
- la terminaison de la communication.

Les échanges d'informations correspondant à ces divers services se font par des messages (APDU) normalisés, dont le format est décrit en ASN1 et le contenu encodé selon les règles basiques d'encodage BER ASN1 (ISO 8825).

Z39.50 voit une base de données comme un ensemble de fichiers contenant des documents. Chaque document peut être structuré en champs. Z39.50 permet le codage de questions de divers types. Certains types autorisent l'emploi d'opérateurs booléens et de proximité. Les opérandes employés dans ces questions vont être des points d'accès à la base (champ ou valeur ; exemple : **Auteur = Dupont**), ou bien des sous-ensembles de résultats issus de questions précédentes.

Une autre notion importante dans le protocole Z39.50 est celle d'*Attribute Set*. Cette notion consiste à définir par un nom générique comme BIB1 la structure générale de la base :

- quels sont les noms de champ et comment ils sont codés (exemple : **Titre=4**) ;
- quelles sont les relations possibles pour associer des valeurs à ces noms de champ : **LT, LE, EQ, GT, GE, NE** (exemple : **EQ=3**) ;
- comment sont codées des directives relatives à la position du mot (exemple : **First in field, Any position in field**) ;
- la structure du champ (phrase, mot, liste de mots, date) ;
- la façon d'indiquer des troncatures (droite, gauche...).

L'*Attribute Set* BIB1 propose une telle nomenclature adaptée aux bases bibliographiques. Les versions ultérieures du protocole pourront proposer d'autres nomenclatures.

La version 2 de Z39.50 a pris en compte les caractéristiques de SR (*Search and Retrieval*), le standard ISO pour ce domaine (ASN1, ensemble de résultats multiples, négociation du format des résultats), et apporté des fonctionnalités supplémentaires (contrôle d'accès, contrôle de ressources, types de questions diversifiés).

Le protocole utilisé par Wais actuellement est une variante de Z39.5088 (version 1). Il utilise des questions de Type-3 non prévues dans la norme. Ces questions sont des suites de mots cités explicitement ou contenus dans des portions de documents. C'est un « ou » logique qui est pris comme opérateur implicite.

L'ensemble de résultats est effacé à chaque question de façon à rendre le serveur *stateless*. Le service de présentation des documents appartenant à un sous-

ensemble est remplacé par une question de Type-1 ayant comme opérandes des points d'accès de deux nouveaux types :

- **System-Control-Number=xxx**. où **xxx** est l'identifiant (Doc-ID) du document ;
- des parties de documents (**System-Control-Number=xxx**. et **rang.GT. X** et **rang.LT. Y**) **X** et **Y** étant des rangs d'octets, de lignes ou de paragraphes dans le document.

Le CNIDR prévoit la mise en service de logiciels mettant en œuvre le protocole Wais et Z39.50V2.

6.1.2. Logiciels Wais

Les logiciels Wais se composent d'une partie client qui tourne sur la machine émettant les requêtes, d'une partie serveur qui tourne sur la machine abritant la base de données et d'une partie indexation qui permet de constituer la base d'informations. Plusieurs versions du logiciel sont disponibles, notamment :

Wais-b5.1 (ftp quake.think.com) : c'est la dernière version produite par TMC ; elle est stable mais n'est plus maintenue ;

Wais-b5.1+ BIO+filters+Inria (ftp zenon.inria.fr) obtenue en intégrant à la précédente des fonctions permettant l'interrogation booléenne et avec troncatures, l'indexation par filtres externes, l'élimination de mots vides ;

FreeWais-0.3 (ftp ftp.cnidr.org) : c'est la dernière version produite par le CNIDR ; elle comprend les fonctions de la B5 et apporte des extensions : les interrogations booléennes, l'indexation par filtres externes, la prise en compte de synonymes, l'interrogation par radicaux (anglais) ou par troncatures, l'élimination de mots vides et la prise en compte de documents en format compressé ou en représentation multiple.

Des modifications ont aussi été apportées pour permettre l'indexation et l'interrogation en alphabet français.

À ces versions du domaine public s'ajoutent des versions commerciales, proposées par la société Wais Inc. (**ftp.Wais.com**).

6.1.3. Indexation

Le logiciel d'indexation Wais-Index est un programme C qui va parcourir une hiérarchie de fichiers contenant les documents à indexer et constituer un index

avec tous les mots de deux caractères et plus. Lors de l'indexation, on va préciser un type qui va déterminer :

- les règles de découpage des fichiers en documents ; ainsi un document sera, par exemple, une ligne pour le type **one-line**, ou un paragraphe pour le type **para**, ou un message pour le type **mail-digest** ;
- la façon de constituer la ligne d'en-tête (*Header line*) ; cette ligne est rendue comme première réponse aux questions ; par exemple pour le type **mail**, la ligne d'en-tête contiendra les champs **From** et **Subject** ;
- le format du document : texte, DVI, GIF, TIFF, etc. ; ce type sera fourni au client lors de la visualisation du document ; ceci lui permettra d'appliquer le filtre de visualisation de son choix.

Il existe un nombre important de types prévus en standard et il est facile de rajouter de nouveaux types en modifiant deux modules du programme d'indexation. La version officielle indexe les mots contenus dans le document pour tous les types contenant du texte (**mail**, **news**, **para**, **one-line**, ...) et se contente d'indexer les mots contenus dans le nom du fichier pour les autres types (DVI, PostScript, GIF...).

L'index comporte une entrée par mot. Chaque entrée contient le poids et les positions du mot dans les divers documents. Le poids d'un mot est calculé de la façon suivante : 5 points pour la première occurrence, plus 1 point à chaque nouvelle occurrence du mot dans le document, ou 10 points à chaque occurrence du mot dans la ligne d'en-tête.

Il est possible de maîtriser complètement l'indexation en utilisant des extensions réalisées par Simon E. Spero qui consistent à utiliser un filtre externe pour l'indexation. Lors de l'indexation, on précise à Wais-Index (**argument-filter**) le nom d'un filtre qui va être un programme ou un *script* lancé par Wais-Index. Les communications entre Wais-Index et le filtre vont se faire par le *Standard Input* et le *Standard Output*. Wais-Index va communiquer au filtre les noms des fichiers à indexer et va attendre du filtre, pour chaque document, la ligne d'en-tête, la date du document et les mots clés avec leur poids et leur position.

Avec cette technique, on va pouvoir maîtriser totalement l'indexation des fichiers texte et réaliser des indexations de fichiers qui sont dans d'autres formats (DVI, PostScript, images...) en tirant les mots clés de fichiers explicatifs associés ou bien en lançant dans le filtre des convertisseurs en ASCII des fichiers à indexer (tels que DVI2tty, post-ASCII ou un module OCR). Une récente amélioration permet d'indexer des documents stockés sous plusieurs formats, par exemple texte, PostScript, DVI, etc. L'index sera réalisé à partir de la forme ASCII du document, mais contiendra des pointeurs vers les autres formats du document. Le client pourra retrouver les documents par l'index et récupérer à sa guise la version de celui qui aura le format de son choix.

Les tailles des index dépendent essentiellement de la nature des fichiers indexés. Cette taille peut varier de 0.03 fois (cas de fichiers PostScript avec images) à 3 fois (cas de références bibliographiques) la taille des fichiers à indexer.

La vitesse d'indexation dépend de facteurs multiples tels que la puissance du CPU, la taille mémoire et l'architecture de la machine. Sur la base de références bibliographiques de l'INRIA, on indexe 4 000 références de 0,5 ko par minute sur un Sun SS10. Ces performances d'indexation élevées permettent d'envisager des indexations totales et de pallier les déficiences du mode mise à jour qui ne permet que de rajouter des documents nouveaux.

6.1.4. Client

Les index Wais vont être consultés à distance par un logiciel client. Diverses versions du logiciel client existent. Elles fonctionnent sur le principe suivant ; l'utilisateur dispose de trois répertoires dans son environnement :

Wais-Sources contient les documents `.src` donnant les coordonnées de sources intéressant l'utilisateur ;

Wais-Questions contient le texte de questions déjà posées et prêtes à l'emploi ;

Wais-Documents contient les documents sauvegardés lors d'interrogations précédentes.

L'utilisateur va interroger une ou plusieurs de ces sources en donnant une liste de mots. Une fois les sources sélectionnées et les mots de la question tapés, le client va établir la connexion TCP avec le serveur, ouvrir la communication Z39.50 et soumettre la question. Le serveur va consulter l'index et déterminer les documents répondant à la question. Pour chaque document, le serveur va calculer une note en additionnant les poids des divers mots de la question et en ramenant les chiffres obtenus sur une échelle de 1 à 1 000. Dans FreeWais, le nombre d'occurrences du mot dans la base et la taille du document vont intervenir dans le calcul du poids de façon à ne pas défavoriser les documents courts. Le serveur va présenter au client une liste triée et limitée (40 par défaut) des documents les plus pertinents.

Pour chaque document, c'est la ligne d'en-tête, la note et la taille du document qui seront indiquées. Le client va pouvoir visualiser tout ou partie des documents en les sélectionnant.

Le *Type* (format) du document est communiqué au client qui pourra s'en servir pour appeler l'outil de visualisation le plus approprié. Dans la version X-Window du client, l'association des filtres aux divers types de documents se fait par une ressource X.

Le client va pouvoir modifier sa question en ajoutant ou supprimant des sources ou des mots et en mettant en œuvre un mécanisme dit de *relevance feedback* qui consiste à inclure dans une question des documents ou des parties de documents

déjà trouvés. Il n'est toutefois possible de faire du *relevance feedback* sur une base qu'avec des documents issus de cette base.

Les documents descripteurs des sources peuvent être indexés. De tels index constituent des annuaires. L'un d'entre eux, *directory-of-servers*, contient les descriptions des sources Wais connues sur Internet. Cette source, localisée sur la machine `quake.think.com port 210`, est connue par le logiciel client dès son installation.

Pour se procurer des descriptions de sources, il suffit d'interroger de tels annuaires et de sauvegarder dans Wais-Sources les documents intéressants. Bien entendu, les descripteurs de sources peuvent être créés par un éditeur de texte et échangés par messagerie électronique ou *news*.

Le logiciel client existe sous diverses formes et divers systèmes d'exploitation (Mac, PC, Unix...). Ces versions du domaine public sont accessibles en FTP anonyme.

6.2. Gopher

Gopher est un système de navigation sur Internet. Basé sur un modèle client/serveur, il utilise un protocole interne. Le client Gopher se connecte à un serveur qui va lui proposer une hiérarchie de menus. Chaque entrée dans un menu est un objet Gopher qui peut être notamment :

- un fichier contenant du texte ASCII, compressé, UUencodé, son, MIME, GIF, HTML, etc. ;
- une session TELNET ;
- un index Wais ou Next Librarian ;
- un menu ;
- un lien vers un objet Gopher distant ;
- un lien vers un espace distant accessible en FTP ;
- un lien vers un programme.

L'utilisateur Gopher voit donc tous ces objets à travers une hiérarchie de menus qui pointent sur des objets situés sur le serveur ou sur des serveurs distants. Le fait de choisir une entrée dans un menu provoque une action qui dépend du type de l'objet. Lorsqu'un objet de type fichier est sélectionné, le fichier sera transféré chez le client qui le présentera en tenant compte de son format.

Lorsqu'un objet de type lien est présenté, alors le client Gopher établira une liaison vers le serveur indiqué en pratiquant le protocole approprié (FTP, Gopher). Dans le cas de FTP, le client Gopher présentera les répertoires comme des menus et les fichiers comme des entrées dans ces menus.

Lorsqu'un objet de type index Wais est choisi, le client est invité à fournir une liste de mots qui constituent la question. Cette liste de mots est acheminée au serveur Gopher en utilisant le protocole Gopher. Le serveur Gopher se comporte

comme un client Wais et lance à son tour une requête Z39.50 au serveur Wais. Le serveur Gopher obtient en retour une liste synthétique de résultats qu'il présente à son client comme un menu.

Le fait, pour le client Gopher, de sélectionner un document va faire monter une requête d'affichage au serveur Gopher qui va la faire suivre en Z39.50 au serveur Wais. En retour, le serveur Wais va rendre le document au serveur Gopher qui, à son tour, va l'acheminer au client Gopher.

Hélas, seuls les documents Wais de type texte peuvent être ainsi visualisés. Par ailleurs, le mécanisme de *relevance feedback* ne peut être mis en œuvre à partir de Gopher. Ce mode de fonctionnement peut s'avérer aberrant quand par exemple un client Gopher localisé en France trouve sur un serveur australien un pointeur vers un index Wais localisé en France. Compte tenu du mécanisme décrit en page précédente, des dialogues qui pourraient rester localisés en France vont traverser deux fois Internet.

La recherche arborescente dans des menus pouvant s'avérer fastidieuse, il est possible de bâtir des index avec les noms des entrées des divers menus présents sur un serveur. L'utilisateur pourra directement accéder aux entrées répondant à sa question sans avoir à parcourir la hiérarchie.

Gopher est d'une mise en œuvre aisée, il suffit de lancer le serveur Gopher sur une hiérarchie de fichiers pour avoir cette hiérarchie présentée comme un ensemble de menus, dont les textes seront les noms de fichiers ou de répertoires. Il est bien entendu possible de donner des noms plus parlants aux divers fichiers, répertoires et liens.

6.3. World Wide Web : WWW ou W³

WWW englobe trois composantes.

HTML un format de description de documents hypertexte, multimédia, multi-protocole ;

HTTP un protocole permettant la récupération de documents HTML localisés sur des serveurs ;

WWW un réseau mondial de serveurs délivrant des documents en format HTML.

Il apporte un nombre appréciable d'améliorations sur plusieurs plans :

- ergonomie : on parcourt Internet en cliquant sur des mots situés dans des textes ;
- autonomie : on ne dépend plus d'un serveur et chacun peut, avec son éditeur de textes favori, rédiger un texte HTML correspondant à sa vision du monde ou d'un problème, le lire, en exploiter les liens, l'envoyer à ses correspondants par messagerie ou FTP, ou enfin le stocker sur un serveur WWW ;

- fonctionnalités : un lien HTML peut pointer vers un nombre d'objets variés (fichier, espace FTP, *news-group*, index Wais, serveur Gopher, serveur WWW, etc.).

Ce produit a été conçu et réalisé par le CERN ([ftp info.cern.ch](ftp://info.cern.ch)) et fait l'objet de développements multiples dans le monde. Il a connu un développement explosif le jour où le NCSA, un centre de calcul de la NSF, a mis dans le domaine public le logiciel Mosaic qui est un outil permettant la visualisation et l'exploitation d'hypertextes en format HTML.

6.3.1. HTML

L'*Hypertext Marking Language* est un type de document SGML. HTML permet de présenter des documents structurés comprenant un titre, du texte sous diverses tailles et formes, des listes, des points d'ancrage sur lesquels pointeront les liens internes au document. Le texte HTML est dans un format clair et lisible et peut être fabriqué manuellement à partir de n'importe quel éditeur de textes.

Comme dans tout hypertexte, on va pouvoir poser des liens sur des mots. En HTML, ces liens vont pointer vers d'autres parties du document, des fichiers ou objets situés éventuellement sur des machines distantes. À la lecture du document, les mots sur lesquels des liens ont été placés seront signalés, en vidéo inversée par exemple, et le fait d'en choisir un, en le sélectionnant à la souris par exemple, provoquera le rapatriement et l'affichage de l'objet pointé.

Les liens peuvent aussi être placés dans des images et le fait de sélectionner telle ou telle partie de l'image déclenchera telle ou telle action.

Les fichiers pointés dans des liens peuvent contenir du texte, des images fixes ou animées, du son, des séquences audio-vidéo, et des programmes exécutables. Lorsque de tels objets sont sélectionnés par une application cliente, ils lui seront communiqués, à charge pour elle de lui appliquer le traitement informatique approprié.

Les objets distants pointés peuvent aussi être des fichiers accessibles en FTP, des menus pris sur des serveurs Gopher, des index Wais, des *news-groups*, des documents pris sur un autre serveur WWW, des sessions TELNET. Pour exploiter de tels liens, un client comme Mosaic établira une session avec le serveur indiqué en pratiquant le protocole approprié, à savoir : FTP, Gopher, Z39.50-Wais, NNTP, HTTP ou TELNET. Il rapatriera le document et lui appliquera le traitement approprié (affichage avec l'outil *ad hoc*, redirection sur la prise audio pour les sons, exécution pour les programmes).

Les pointeurs utilisés dans les liens sont des URL (*Uniform Resource Locators*). Ce sont des chaînes de caractères de la forme « protocole serveur port chemin d'accès ». Les URL sont en voie de normalisation par l'IETF. Les travaux portent non seulement sur les problèmes de localisation (URL), mais aussi de façon plus générale sur l'identification de ressources avec des URI (*Uniforms*

Resources Identifiers). Les URL seraient un type d'URI précisant la localisation et le mode d'accès, les autres types d'URI étant les URN (*Uniform Resource Names*) donnant un nom générique à la ressource et les URC (*Uniform Resource Characteristics*) définissant son URN, ses URL et d'autres informations telles que son titre, sa version, ses auteurs ou responsables, son format, ses conditions d'accès, etc.

Les URI sont censés identifier des ressources au sens large ; lorsque ces ressources sont des documents, le but recherché est, connaissant un URN, de tirer d'un annuaire, d'un serveur de noms ou d'une base de données l'URC associé et d'exploiter les informations qu'il contient pour choisir la version et le format souhaités, puis de rapatrier le document grâce aux informations contenues dans l'URL.

Des extensions au format HTML (HTML+) sont en cours de discussion pour permettre notamment l'inclusion dans des textes HTML d'images, de formules mathématiques et de formulaires.

6.3.2. Protocole HTTP

Ultra simple, il ne présentait dans sa première version qu'une seule directive : GET. L'écriture de serveurs HTTP est donc d'une simplicité biblique. Actuellement, le protocole a évolué et il permet au client qui demande un objet d'annoncer les formats qu'il supporte et au serveur d'annoncer le format des données qu'il envoie. C'est la terminologie employée dans MIME qui est utilisée à cet effet. Le serveur a toute latitude pour convertir le document dans un format accepté par le client.

6.3.3. Logiciels WWW

Côté client, il existe plusieurs logiciels, développés notamment par le CERN, pour terminaux alphanumériques notamment, mais la palme revient sans conteste à Mosaic qui permet d'exploiter toutes les possibilités de HTML et des URL si bien qu'avec ce logiciel on a un client permettant, en mode hypertexte, de faire du FTP, de lire les *news*, d'interroger des index Wais, de consulter des serveurs Gopher et bien entendu des serveurs WWW.

Ce logiciel a la possibilité de lire des fichiers HTML locaux, de mémoriser la trace de pages intéressantes, de placer sur les documents des annotations écrites ou vocales. Mosaic supporte les images incluses dans les documents HTML et applique le programme de traitement choisi par l'utilisateur aux fichiers reçus.

Il supporte aussi la gestion de formulaires qui sont des parties de documents avec des cases vides, des menus, des boutons que l'utilisateur pourra remplir ou sélectionner. Un URL et un bouton envoi sont associés à chaque formulaire. Lors de la sélection du bouton envoi, l'URL associé est activé et les valeurs entrées sont passées en paramètres.

Il existe un nombre important de serveurs. Les plus notoires sont ceux proposés par le CERN et le NCSA. Ces serveurs permettent de délivrer les documents HTML autorisés. On peut associer à chaque document une liste de contrôle d'accès et limiter la diffusion de certaines pages à certains utilisateurs authentifiés.

Ces serveurs permettent en outre l'exécution de programmes, ce qui, conjointement à l'utilisation de formulaires, permet de réaliser des interfaces vers tout système d'information ou SGBD.

6.3.4. Un réseau mondial de serveurs

On compte actuellement plusieurs milliers de serveurs WWW sur Internet et la tendance est à la généralisation. Pour s'y retrouver, il existe plusieurs outils, notamment la page *Quoi de neuf au NCSA ?*, qui donne tous les jours les serveurs qui se sont signalés et tient des archives sur plusieurs années ; le répertoire thématique et géographique du CERN ; l'index des serveurs de l'université de Genève qui indexe les deux documents précédents et permet de retrouver des serveurs à partir des mots utilisés pour les décrire ; un ensemble de cartes mondiales maintenues par l'université de Buffalo pointant sur les cartes nationales (UREC pour la France).

Depuis quelques mois, des tentatives de parcours automatique de WWW ont été tentées : le jeu consiste à exploiter les liens – récursivement et automatiquement – et à bâtir des index avec les noms des pages récupérées, voire avec les mots contenus dans ces pages. Les résultats ne sont pas très probants et des règles de bon usage sont en cours de définition pour contingenter l'utilisation de tels outils.