

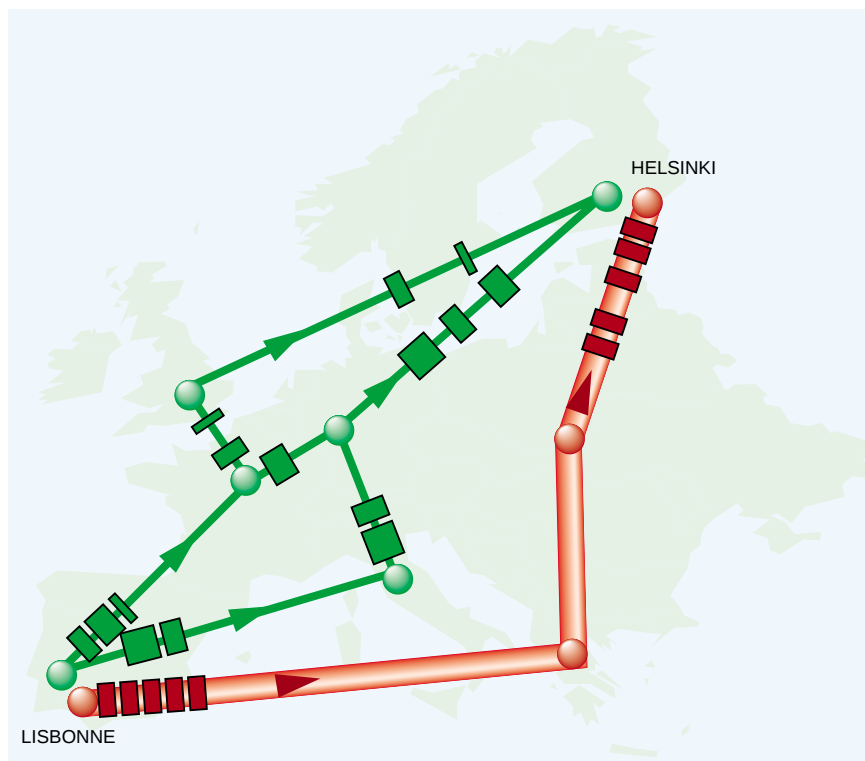
(nommé TCP/IP), le transfert de paquets s'effectue en «mode non connecté» : il n'y a pas de chemin prédéfini ni de ressources réservées, et chaque paquet d'un message se propage dans le réseau de façon autonome, indépendamment des autres paquets. C'est le récepteur final qui doit recomposer le message, en récupérant progressivement des paquets qui peuvent avoir suivi des chemins différents.

L'ATM peut être utilisé pour transporter n'importe quel type d'information. On distingue trois grandes familles : la parole, la vidéo et les données informatiques. Chacune a ses spécificités que l'ATM respecte grâce à la souplesse d'usage des cellules. On peut ainsi intégrer sur un même support physique tous les types de services, ce qui est un avantage considérable. De surcroît, comme toutes les techniques de paquets, l'ATM optimise le transfert en utilisant des cellules uniquement lorsque les sources sont actives et ont effectivement des informations à transmettre. On peut aussi faire du multiplexage statistique, c'est-à-dire profiter des périodes de silence de certaines sources pour insérer des cellules provenant d'autres sources : sur un réseau de parole, par exemple, tout le monde ne parle pas en même temps.

On comprend que, pour profiter de ces avantages, il faille connaître le comportement des sources et le trafic qu'elles engendrent. C'est relativement aisé pour les sources de parole, un peu moins pour les sources vidéo comprimées dont le débit est aléatoire, mais cela devient difficile, voire impossible, pour les sources de données qui ont une activité imprévisible et qui travaillent à des débits très variables. Ainsi, le dimensionnement et l'utilisation d'un réseau ATM pose encore quelques problèmes.

L'ATM et les services

Les concepteurs de l'ATM voulaient initialement répondre à un double besoin des opérateurs publics : offrir des services «large bande», notamment vidéo, et intégrer tous les services sur un même support. L'intérêt de l'ATM a ensuite été perçu par les informaticiens, si bien qu'aujourd'hui, ce sont eux qui poussent au déploiement de cette technique dans les réseaux privés. Inévitablement se



UN MÊME MESSAGE est envoyé de Lisbonne à Helsinki. Sur le réseau Internet (*en vert*), les données sont découpées en paquets de longueurs variables, qui sont acheminés indépendamment. Sur un réseau ATM (*en rouge*), les données, découpées en paquets de longueur fixe, voyagent comme des wagons sur des rails.

pose alors le problème de l'interconnexion des réseaux locaux d'entreprise et des réseaux publics ; il importe aussi de préserver les investissements, souvent considérables, consentis avec des techniques antérieures. La souplesse de l'ATM devrait permettre d'y parvenir et de fédérer les autres techniques de transfert en utilisant la cellule comme un conteneur universel, apte à transporter des paquets d'autres protocoles.

Le réseau Internet a été créé pour la transmission de données et l'interconnexion de réseaux locaux. À la différence de l'ATM, il opère en mode non connecté, ce qui est la technique habituelle des réseaux locaux. Fort de sa popularité, Internet cherche à transmettre la parole et la vidéo, ce pour quoi il était *a priori* mal adapté. À cette fin, des équipes élaborent de nouveaux protocoles, se rapprochant de l'ATM par certains aspects, notamment la réservation de ressources.

L'ATM, de son côté, intègre difficilement sur le même support (la file d'attente) des services aux contraintes aussi différentes que la parole, la vidéo et les données. Il évolue donc en introdui-

sant des mécanismes de priorité dans les files d'attente et des protocoles de contrôle de flux qui se rapprochent des techniques traditionnelles des réseaux de paquets.

Les deux modes ont leurs avantages et leurs inconvénients. Les services de données s'accommodent bien du mode non connecté, surtout lorsque les messages sont courts, tandis que les services «en temps réel» (parole, vidéo) sont mieux traités en mode connecté. Ce qui pose problème, avec l'une et l'autre technique, c'est le mélange de tous les types de services. À cet égard, de par la richesse de ses potentialités de traitement et de signalisation (l'ensemble des protocoles permettant à l'utilisateur de dialoguer avec le réseau pour établir une connexion), l'ATM semble mieux placé. Les deux techniques devront probablement cohabiter, mais seule l'expérience permettra de définir la part qui reviendra à chacune d'elles.

Jacques DUPRAZ a travaillé pour la Société Alcatel.

Les archives d'un réseau actif

BREWSTER KAHLE

L'archivage des documents créés et échangés sur le réseau Internet pourrait être historiquement, commercialement et politiquement important.

Au I^{er} siècle avant notre ère, les manuscrits de la bibliothèque d'Alexandrie ont disparu dans un incendie ; les premiers livres imprimés sont tombés en lambeaux ; une grande partie des tout premiers films ont été recyclés, parce qu'ils contenaient de l'argent... Malheureusement, l'histoire pourrait se répéter avec le réseau Internet.

Personne n'a tenté d'archiver intégralement les textes et les images contenus dans les documents qui apparaissent sur le Web. Comment, alors, éviter la perte de précieuses informations historiques, culturelles et scientifiques ?

Avec la diminution constante des coûts du stockage numérique, un petit groupe de techniciens, dotés d'ordinateurs personnels et de dispositifs de stockage de données, pourrait archiver en permanence les documents diffusés sur le réseau mondial. Avec quelques collègues, nous avons commencé cette tâche colossale voilà un an, dans le cadre du projet *Internet Archive*.

Lorsque ce numéro de *Pour la Science* sera en kiosque, nous aurons enregistré quasiment toutes les parties du réseau qui sont librement et techniquement accessibles. Cet ensemble de données, soit 2 000 milliards d'octets (deux téra-octets), comprendra du texte, des images et des sons (la Bibliothèque nationale de France contient l'équivalent de 20 téra-octets de texte). Dans les mois à venir, nos ordinateurs et nos supports de stockage enregistreront d'autres secteurs d'Internet, dont les informations du *Gopher* (un système

d'information antérieur au Web) et les forums de discussions d'*Usenet*. Les enregistrements faits à ce jour se sont déjà révélés très utiles pour les historiens. Dans l'avenir, ils pourraient constituer le fonds d'une bibliothèque minutieusement indexée... et consultable par le réseau.

Dix personnes qui participent au projet travaillent dans une base militaire reconvertie, à San Francisco. L'ordinateur collecteur avec lequel nous recueillons les informations est situé au Centre de calcul de l'Université, de San Diego.

Nos programmes «sillonnent» l'Internet, recueillant des documents nommés «pages», site après site. Lorsqu'une page est enregistrée, ils recherchent des références croisées, des «liens» avec d'autres pages. Ils utilisent ces liens – des adresses insérées dans la page d'un document – pour se déplacer vers les autres pages, refont alors une série de copies et cherchent d'autres liens vers de nouvelles pages. Naturellement, ils vérifient que les adresses qu'ils se proposent d'examiner n'ont pas encore été visitées à l'aide des identificateurs de sites nommés «localisateurs unifiés de ressources», ou URL. Les programmes tels qu'*AltaVista*, de la Société *Digital Equipment*, emploient eux aussi un logiciel collecteur pour indexer les sites Web.

Notre projet est possible parce que le coût de stockage de données est en constante diminution. Le prix d'un giga-octet (un milliard d'octets) d'espace sur un disque dur revient à 1 000 francs, et

le stockage sur bande magnétique revient à 100 francs par giga-octet. Pour stocker les données que les utilisateurs des archives risquent de rechercher fréquemment, nous avons choisi le stockage sur disque dur, tandis que pour les données moins utilisées, nous avons opté pour un dispositif qui charge et lit les bandes magnétiques automatiquement : en moyenne, un lecteur de disques met 15 millisecondes pour accéder aux données, contre quatre minutes pour un lecteur de bandes. L'information fréquemment consultée pourrait être composée des documents historiques ou d'un ensemble de pages qui ne sont plus en service.

Nous comptons mettre à jour nos archives tous les deux ou trois mois. Le premier archivage complet a duré environ un an, mais, lors de nos prochaines recherches, nous ne mettrons à jour que les informations modifiées depuis notre précédente inspection.

Nous n'aurons pas tous les textes, images, films et autres données du réseau, car les programmes collecteurs tels que le nôtre n'ont pas accès à des centaines de milliers de sites : les éditeurs limitent l'accès à certaines données ou stockent des documents dans un format inaccessible. Toutefois, même si notre archivage est incomplet, il donne un bon aperçu du réseau à l'époque de l'archivage.

Lorsque nous aurons stocké la partie d'Internet accessible au public, quels seront les services offerts par nos archives ? Tout d'abord, nous fournirons des documents qui ne sont plus disponibles chez les éditeurs, ce qui est un service important si le standard d'hypertexte du réseau s'impose dans le monde savant. Nos archives pourraient également se révéler très utiles pour les entreprises. En outre, les données collectées serviraient de «copies d'archives» pour les services publics ou pour d'autres institutions possédant des documents accessibles au public. Ainsi, progressivement, nous disposerons d'une bibliothèque numérique.

Retrouver les liens manquants

Des historiens utilisent déjà ces archives. David Allison, de la *Smithsonian Institution*, y a trouvé des informations pour réaliser une exposition sur l'emploi des sites Web lors des élections présiden-



INTERNET ARCHIVE a fourni au Musée d'histoire américaine du **Smithsonian** un ensemble de pages Web consacrées à l'élection présidentielle américaine de 1996. L'ordinateur qui accède aux pages a fait partie d'une exposition sur les campagnes présidentielles.

tielles : les expositions du passé présentaient des spots télévisés de campagnes électorales ; elles s'enrichissent aujourd'hui de pages Web.

La constitution d'archives pose plusieurs problèmes allant du respect de la vie privée à celui des droits d'auteur. Que se passerait-il si une étudiante créait une page montrant des photographies de son petit ami du moment et si, plus tard, elle voulait «déchirer» ces photos, alors qu'elles sont toujours présentes dans les archives ? Un homme politique pourrait-il effacer des données publiées durant ses années d'études ? Plus généralement, l'accès public aux informations collectées est-il un abus des clauses d'«usage loyal» définies par la loi sur la protection des auteurs ? (voir *Les droits d'auteur des œuvres numériques*, par Ann Okerson, *Pour la Science*, septembre 1996).

Pour résoudre ces difficiles problèmes, nous autorisons les auteurs à exclure leurs œuvres de nos archives. Nous envisageons également de communiquer aux chercheurs non pas des documents individuels, mais uniquement un recensement des grands thèmes archivés : on pourra, par exemple, compter le nombre de références aux pachydermes sur le Web, mais non regarder une page spécifique concer-

nant les éléphants. Nous espérons que ces mesures suffiront à dissiper dans un premier temps les inquiétudes concernant le respect de la vie privée et des droits de la propriété intellectuelle. En testant sur l'Internet des concepts tels que celui d'usage loyal, nos résolutions des problèmes rencontrés au fil du temps lors de la mise en place d'*Internet Archive* contribueront peut-être à faire aboutir les débats politiques sur le respect de la vie privée et sur la propriété intellectuelle.

Avec *Internet Archive*, nous cherchons également à résoudre les problèmes de pérennisation de l'information présente sur le réseau. À Washington, la Commission «préservation et accès» cherche à éviter les pertes de données résultant des changements de standard du stockage numérique. Avec d'autres équipes, nous cherchons des critères techniques qui permettraient d'attribuer un nom d'identification unique (URN, «nom de ressource unité») aux documents numériques. À la différence d'un URL, qui caractérise son emplacement sur le réseau, un URN caractérise le contenu d'un document. L'attribution d'un URN à un document permettrait de le retrouver après disparition du site qui le contenait : on estime à une quarantaine de jours la durée de vie moyenne

d'un URL. Un URN pourrait localiser d'autres URL donnant encore accès au document désiré.

D'autres projets d'archivage s'intéressent à des parties spécialisées du réseau : *DejaNews* enregistre les messages diffusés sur les forums *Usenet*, et *InReference* archive les listes de discussions par courrier électronique. Ces deux services sont financés par la publicité, ce qui pourrait également être un jour le cas d'*Internet Archive*. Aujourd'hui encore, nous finançons notre projet avec l'argent provenant de la vente d'une société de logiciels et de services pour l'Internet. De grandes entreprises d'informatique nous ont également donné du matériel.

De nombreuses années s'écouleront avant qu'une infrastructure ne préserve correctement les informations d'Internet et que les problèmes liés à la propriété intellectuelle ne soient résolus. Entre-temps, nous poursuivons notre archivage, car si trop de documents disparaissaient du réseau, nous perdriions toute trace de la naissance de ce nouveau média.

Brewster KAHLE a fondé *Internet Archive* en 1996. Il a créé, en 1989, le système de recherche de documents WAIS (*Wide Area Information Server*).

HIPPARCOS, satellite chasseur d'étoiles

FRANÇOIS MIGNARD • CHRISTIAN MARTIN

La position, le mouvement et l'éclat de 118 000 étoiles ont été mesurés avec une grande précision. Ces informations améliorent la mesure de l'Univers et permettent de tester les modèles de formation des étoiles.

Le 13 mai 1997, la mission HIPPARCOS s'achève officiellement. Après 20 années de réflexion, de mises au point, d'observations et de calculs, ce dernier-né des catalogues stellaires supplante ses prédécesseurs dans tous les domaines : il augmente à la fois le nombre d'étoiles dont la position est connue de façon absolue, par rapport à un système de coordonnées uniques, et la précision de cette position. Le catalogue HIPPARCOS contient les positions, les mouvements, les distances et les éclats de 118 000 étoiles.

Ces données seront utilisées dans toutes les branches de l'astronomie. Un premier point est d'importance : alors que les astronomes sont habituellement réduits à voir le spectacle des cieux sur une surface sphérique sans profondeur, HIPPARCOS donne une image du ciel en trois dimensions. La distance à laquelle se trouve chaque étoile du catalogue a été mesurée géométriquement : avec ces valeurs absolues, on étalonne les méthodes de mesure à plus grande distance, à partir d'hypothèses sur la physique des étoiles. Leur application progressive à des astres de plus en plus éloignés permet, en particulier, d'évaluer la taille de l'Univers.

Avec 13 millions d'observations, soit une moyenne de 110 mesures par étoile, HIPPARCOS dresse le tableau le plus complet jamais obtenu de la population d'étoiles dans un rayon de 500 années-lumière autour du Soleil. La mesure simultanée de l'éclat et de la distance donne accès à la luminosité intrinsèque des étoiles ; les variations de luminosité de plusieurs milliers d'étoiles ont été détectées et caracté-

risées ; plus de 10 000 systèmes d'étoiles doubles ont aussi été identifiés et mesurés. Ces informations nouvelles seront utiles pour l'étude de la formation et de l'évolution des étoiles. Les utilisations préliminaires du catalogue HIPPARCOS, l'étude des planètes extrasolaires, la compréhension des explosions d'étoiles ou le calcul des trajectoires d'engins spatiaux, illustrent déjà la diversité de ses usages.

En outre, à côté du catalogue HIPPARCOS, le catalogue TYCHO, réalisé avec un détecteur auxiliaire du satellite, contient plus d'un million d'étoiles, repérées avec une précision moindre. Presque toutes les étoiles qui étaient accessibles au télescope embarqué à bord d'HIPPARCOS sont ainsi répertoriées, et les astronomes pourront conduire des études statistiques sur toutes les étoiles plus brillantes que la magnitude 11,5 (la magnitude des étoiles mesure leur luminosité).

La mesure du ciel

Le recensement des étoiles et de leurs positions dans le ciel a longtemps été l'une des activités principales des astronomes. Indispensables à la navigation, ces positions l'étaient aussi, dans une moindre mesure, aux prédictions astrologiques. Au II^e siècle avant notre ère, le grec Hipparque recense les positions d'environ un millier d'étoiles et évalue leur éclat : il réalise ainsi le premier catalogue digne de ce nom. Les positions relevées par Hipparque avaient une précision de 0,3 à 0,4 degré, un peu plus que la largeur d'un quartier de Lune. Près de 1 500 ans plus tard, Ulugh

Begh, petit-fils de Tamerlan, réalise à Samarkand, au moyen d'un sextant de 48 mètres creusé dans le sol, un catalogue de 1 018 étoiles dont la précision moyenne est de l'ordre de cinq minutes d'angle (un degré contient 60 minutes d'angle, qui contiennent chacune 60 secondes d'angle).

Dans la deuxième moitié du XVI^e siècle, l'astronome danois Tycho Brahé mesure la position des étoiles à l'œil nu, à l'aide de longs quadrants muraux en bois ou en bronze. Son catalogue, édité par Johannes Kepler en 1627, renferme la position de 1 005 étoiles, avec une précision moyenne de deux à trois minutes d'angle, mais qui atteint 30 secondes d'angle pour certaines. Johannes Hévélius, astronome allemand brasseur à Dantzig, est deux fois plus précis dans son catalogue de 1 553 étoiles publié en 1690.

L'invention des lunettes et l'adaptation de systèmes de visée vers 1660 amènent progressivement la précision à quelques secondes d'angle à la fin du XVIII^e siècle et permettent d'observer des étoiles de luminosité apparente de plus en plus faible. Au XIX^e siècle, l'observation et le traitement des données deviennent plus systématiques. Ainsi, le *Bonner Sternverzeichnis*, publié en 1835 par l'allemand Friedrich Argelander, contient les positions de 458 000 étoiles. C'est le siècle des grands inventaires, où sont dressés les premiers catalogues photographiques de plusieurs centaines de milliers d'étoiles.

Les positions des étoiles contenues dans tous ces catalogues sont déterminées les unes par rapport aux autres. À l'opposé, les catalogues astrométriques