

Interstices

Découvrir

Les réseaux de pair à pair

Par Laurent Viennot

Odile Lausecker

NIVEAU DE LECTURE **Intermédiaire** • • •

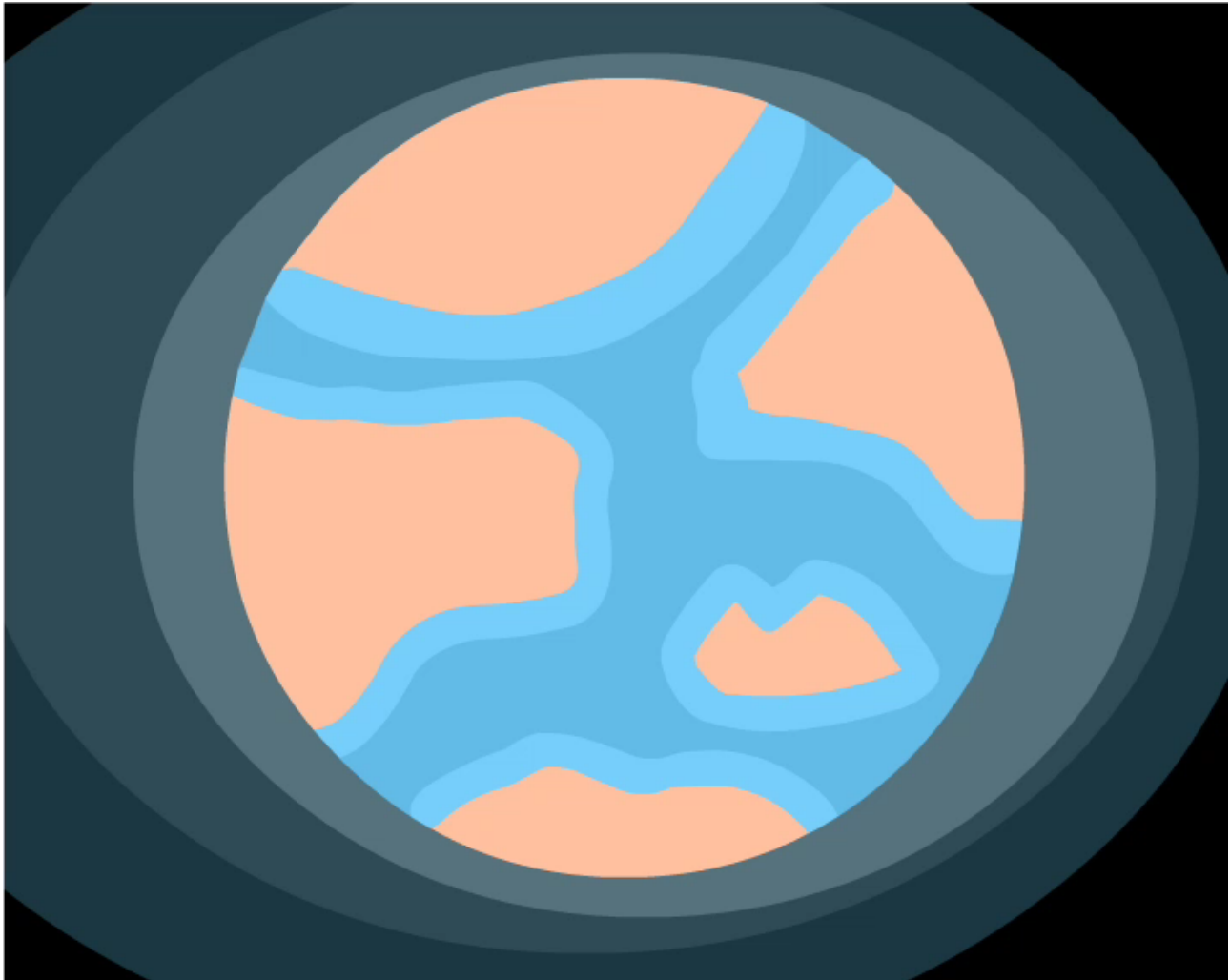
Publié le 04/02/2005

En 1990, bien avant de connaître la diffusion qui est la sienne aujourd'hui, le web a été créé par des informaticiens pour répondre aux besoins de communication des physiciens. Depuis, les utilisateurs se sont appropriés Internet, et c'est de cette appropriation que sont nés les réseaux de pair à pair (appelés en anglais « peer to peer » ou P2P), qui permettent le partage des fichiers entre plusieurs ordinateurs connectés. Comment ces réseaux se sont-ils développés et comment fonctionnent-ils ?

1. Qu'est-ce qu'un réseau de pair à pair ?

Pour échanger des fichiers, il faut que les ordinateurs communiquent entre eux. Le mode de communication le plus répandu est la connexion dite « client serveur ».

Dans le cas de l'échange de fichiers sur des réseaux « pair à pair », les ordinateurs des internautes qui participent à l'échange sont tour à tour demandeur et donneur, client et serveur. Ce sont des « pairs » : la communication s'effectue sur un pied d'égalité.



Visionner la vidéo (Durée : 2 minutes 15).

Pour qu'un réseau pair à pair existe et fonctionne, il faut tout d'abord que chaque internaute participant installe sur son ordinateur un logiciel qui lui permette de se connecter au réseau.

Il faut aussi que le réseau soit doté de règles de fonctionnement communes à tous : c'est ce qu'on appelle le protocole du réseau. C'est lui qui spécifie comment et selon quel format les ordinateurs s'échangent des informations. Par exemple : qui possède tel ou tel fichier ? comment initier un transfert de données ? Les protocoles identifient l'ordinateur qui propose le fichier, et contrôlent le transfert des données.

Pour comprendre le fonctionnement des réseaux pair à pair, le plus simple est encore de voir comment, depuis cinq ans, différentes applications se sont succédées. En effet, comme ces réseaux sont créés par des utilisateurs, chaque nouvelle application a apporté des réponses aux problèmes rencontrés par les applications précédentes.

2. Premier problème : trouver l'adresse

Pour obtenir un fichier, il faut trouver une personne dont l'ordinateur contient le fichier recherché et qui accepte la connexion pour le partager. Le premier problème est donc de trouver « l'adresse IP » d'un ordinateur sur lequel le fichier recherché est accessible.

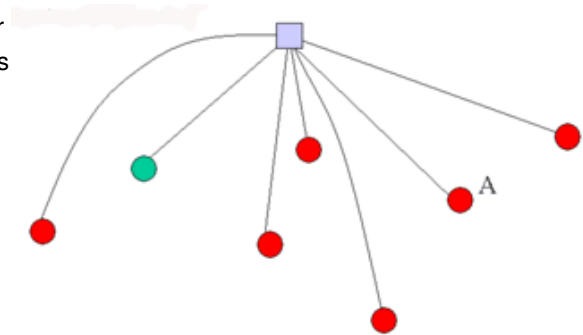
Première génération d'applications : l'index central (Napster)

La première génération d'applications de partage de fichiers comprenait un serveur central qui servait d'index. Les exemples les plus connus sont Napster et Audiogalaxy.

Dans ce type de système, lorsqu'un internaute souhaite partager un fichier, il le déclare au serveur central. Celui-ci répertorie alors son adresse IP.

Tout autre internaute qui souhaite obtenir ce fichier interroge l'index central, comme il interrogerait un moteur de recherche. L'index central lui communique une adresse IP. Ensuite, le logiciel de partage lui permet de se connecter directement sur l'ordinateur qui propose le fichier.

Le fichier en lui-même ne transite pas par le serveur central, qui fonctionne simplement comme un annuaire : une fois l'adresse IP trouvée, les ordinateurs peuvent se connecter directement entre eux.



Principe de fonctionnement de Napster.

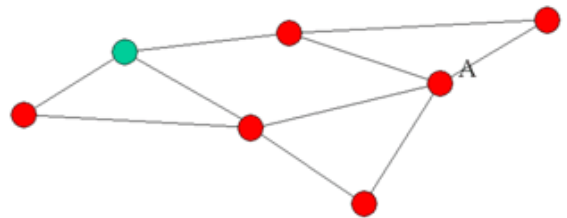
Mais il ne s'agit pas non plus d'un réseau totalement « pair à pair », dans la mesure où sans le serveur central, le réseau ne peut pas fonctionner.

Deuxième génération : chaque internaute indexe ses propres fichiers (réseau Gnutella - version 0.4)

Dans cette deuxième génération, le serveur central n'est plus nécessaire à la connexion. Chaque internaute indexe lui-même ses propres fichiers. Ceux qui sont à la recherche d'un fichier interrogent, de proche en proche, tous les ordinateurs du réseau. L'exemple le plus connu est le réseau Gnutella.

Ici, le premier problème à résoudre est de trouver, lors de la première connexion, d'autres internautes qui participent au partage.

Dès que l'on rencontre un ordinateur connecté au réseau, celui-ci peut communiquer les adresses IP d'autres participants. Pour trouver ce premier internaute, on utilise les « Gwebcache ». Ce sont des ordinateurs du réseau Gnutella qui fonctionnent également comme des serveurs web. Ainsi, ils sont répertoriés par les moteurs de recherche : il suffit donc de taper « Gwebcache » sur l'un de ces moteurs pour obtenir des listes d'ordinateurs connectés.



Principe de fonctionnement de Gnutella (version 0.4).

Lors des connexions ultérieures, la liste ainsi établie sera automatiquement réutilisée.

Mais chaque internaute ne peut se connecter qu'à un nombre limité d'autres participants. Ainsi, de proche en proche, se constitue ce qu'on appelle un « réseau logique ». C'est un « overlay network » qui se greffe sur le réseau global.

En pratique, quand on recherche un fichier, c'est sur ce réseau logique qu'on diffuse sa requête. Celle-ci est d'abord adressée aux ordinateurs voisins, puis peu à peu propagée jusqu'à trouver un ordinateur possédant le fichier en question. La réponse, constituée par l'adresse IP de cet ordinateur, suit le même chemin en sens inverse. Ce système garantit donc une certaine forme d'anonymat : celui qui détient le fichier ne connaît pas l'identité de celui qui le lui demande.

Contrairement aux applications de première génération, le réseau Gnutella (version 0.4) est lui un véritable réseau pair à pair, puisqu'on peut se passer totalement d'entité centrale.

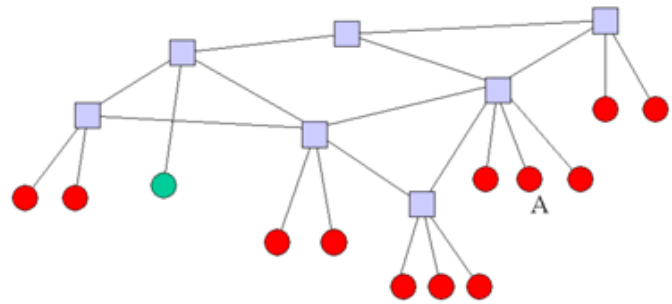
Mais cette deuxième génération a bien vite périclité, car le réseau était constitué à la fois d'ordinateurs connectés en haut débit et d'ordinateurs connectés par modem. Du coup, ces derniers imposaient leur vitesse de connexion à l'ensemble du réseau, et la propagation des données se trouvait fortement ralentie.

Troisième génération : les superpairs (réseau FastTrack et logiciel de partage Kazaa, Gnutella - version 0.6)

Cette troisième génération opère une distinction entre deux niveaux de pairs : ceux qui ont une connexion haut débit et ceux qui ont une connexion par modem.

Les ordinateurs disposant d'une connexion par modem se relient à un ordinateur ayant une connexion haut débit. Ce dernier devient dès lors un « superpair ». Chaque superpair indexe alors les fichiers des pairs bas débit qui lui sont rattachés, comme le faisait autrefois le serveur central des réseaux de première génération.

Entre deux superpairs en revanche, le système continue à fonctionner comme les réseaux de seconde génération. Mais la propagation des données est plus rapide, puisqu'elle n'utilise plus que les connexions haut débit.



Principe de fonctionnement de Kazaa.

Une fois l'adresse IP communiquée à l'ordinateur d'origine, une connexion directe s'établit entre les deux pairs, quel que soit leur niveau.

Il s'agit donc d'une solution hybride entre les réseaux de première et de seconde génération. Le réseau FastTrack, associé au logiciel de partage Kazaa, constitue un bon exemple de réseau pair à pair de troisième génération.

3. Deuxième problème : s'adapter aux caractéristiques de l'ADSL

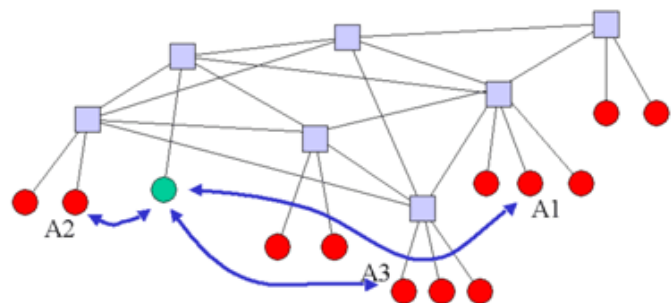
Le développement de l'ADSL a posé un nouveau problème pour le fonctionnement des réseaux pair à pair. En effet, le débit fourni par les systèmes ADSL est asymétrique (c'est ce que signifie d'ailleurs le « A » de l'abréviation). Puisque les utilisateurs, qui sont les « clients », n'envoient que des requêtes de faible poids, le débit ascendant (« upload ») sera faible, tandis qu'à l'inverse, pour recevoir les grosses quantités d'informations que les serveurs envoient aux clients en guise de réponse, le trafic descendant (« download ») sera dix à vingt fois plus élevé. Cette asymétrie va à l'encontre du fonctionnement des réseaux pair à pair, puisque ce faible débit d'upload freine l'ensemble du réseau : si rapide que soit sa vitesse de connexion en download, un ordinateur ne peut pas recevoir un fichier à un débit plus rapide que celui imposé par l'ordinateur qui lui envoie les données !

Pour s'adapter à cet environnement hostile à un trafic symétrique, eDonkey introduit le téléchargement multiple.

Le téléchargement multiple (réseau eDonkey et logiciels de partage eDonkey, eMule, mlDonkey)

Pour compenser le caractère asymétrique de l'ADSL, des internautes ont eu l'idée de rendre possible le téléchargement à partir de plusieurs fichiers, et donc de plusieurs ordinateurs. Le fichier est « découpé » en petites tranches et peut ensuite être téléchargé morceau par morceau à partir de plusieurs pairs.

Le trafic descendant se nourrit donc de plusieurs trafics ascendants. Évidemment, cette solution n'est que partielle : elle ne peut pas permettre à tous les ordinateurs d'augmenter en même temps leur vitesse de téléchargement, puisqu'au total, les capacités cumulées de download et d'upload du réseau restent déséquilibrées. Par ailleurs cette fragmentation en tranches favorise le partage partiel du fichier, un coup de pouce pour la coopération.



Principe du téléchargement multiple par eDonkey.

Développer la coopération

Le développement de la coopération est en effet une problématique essentielle des réseaux pair à pair. Une étude a montré que très peu d'utilisateurs de réseaux de partage de données les utilisent effectivement pour partager des

données ; bien plus nombreux sont ceux qui les utilisent pour acquérir des données, mais n'en proposent pas aux autres.

En quoi consiste la coopération ? Pendant qu'un internaute est en train de télécharger un fichier, son ordinateur peut immédiatement redistribuer les fragments déjà obtenus à d'autres utilisateurs : les différents clients s'entraident donc automatiquement. Par contre, lorsqu'un internaute a fini de télécharger un fichier, il a souvent tendance à l'effacer du dossier partagé ou bien à se déconnecter du réseau pair à pair.

Ce comportement est doublement préjudiciable pour le fonctionnement du réseau : d'une part, c'est une source de téléchargement en moins pour les autres utilisateurs désireux d'acquérir ce fichier, d'autre part, c'est autant de capacité d'upload indisponible pour compenser l'asymétrie. Pour inciter à la coopération, les réseaux pair à pair ont donc mis en place un système de points, qui marchent sur le modèle du bonus/malus : les données fournies en upload permettent d'acquérir des points donnant droit à des capacités de download supplémentaires.

Le protocole BitTorrent, présenté un peu plus loin, propose une solution élégante au problème en utilisant le principe d'« un prêté pour un rendu » : il faut redonner des morceaux du fichier pour pouvoir en acquérir de nouveaux plus rapidement.

Gérer un réseau dynamique

L'un des enjeux fondamentaux des réseaux pair à pair, c'est de gérer la « dynamique » du réseau. En effet, il ne faut pas oublier que le réseau n'est pas figé : à chaque instant, de nouveaux utilisateurs font leur apparition, tandis que d'autres se déconnectent.

Avant les téléchargements multiples, les échanges se faisaient uniquement entre deux pairs : l'inconvénient résidait dans le fait que si le donneur se déconnectait en cours de transfert, le demandeur devait trouver une autre source et recommencer tout le téléchargement depuis le début.

Grâce au système des donneurs multiples, les échanges sont devenus bien plus ergonomiques, puisque de nouveaux pairs peuvent prendre le relais des pairs qui se déconnectent. Un fichier peut ainsi rester disponible dans le réseau alors même que plus personne ne le propose dans son intégralité ! Le téléchargement multiple induit ainsi une persistance des fichiers dans le temps extrêmement utile pour le partage.

4. Troisième problème : optimiser les protocoles

Le protocole d'un réseau englobe toutes les règles organisant les échanges de messages entre pairs. Toutes les activités sont ainsi encadrées : recherche des adresses IP, gestion et contrôle des fragments de fichiers échangés, etc.

Mais la mise en place de ces protocoles génère un trafic d'informations très abondant sur le réseau pair à pair : ainsi, dans le réseau Gnutella (version 0.4), le volume des données liées au protocole était aussi important que celui des fichiers échangés. À l'origine, le protocole HTTP (qui permet la navigation sur internet) a été optimisé. Le trafic lié à ce protocole ne représente que quelques pour cent du trafic des données. À l'heure actuelle, les protocoles pair à pair, eux, sont loin d'être optimisés. L'optimisation de ces protocoles est donc un thème crucial de recherche qui permettrait d'améliorer de façon considérable le fonctionnement global du réseau.

Réseau Overnet : les tables de hachage distribuées pour optimiser les protocoles

Le réseau Overnet est un réseau précurseur en matière d'optimisation.

Pour trouver un fichier dans un réseau pair à pair classique, on propage la requête de pair en pair, en interrogeant tous les ordinateurs connectés jusqu'à ce qu'on trouve un ordinateur qui le détient.

Pour optimiser les échanges, Overnet utilise un algorithme qui s'appuie sur une table de hachage distribuée : « Kademlia ».

Le principe des tables de hachage consiste à utiliser une fonction de hachage pour faire correspondre à chaque nom de fichier son empreinte, une sorte de « résumé crypté » du fichier, sous forme d'une chaîne de bits (une succession

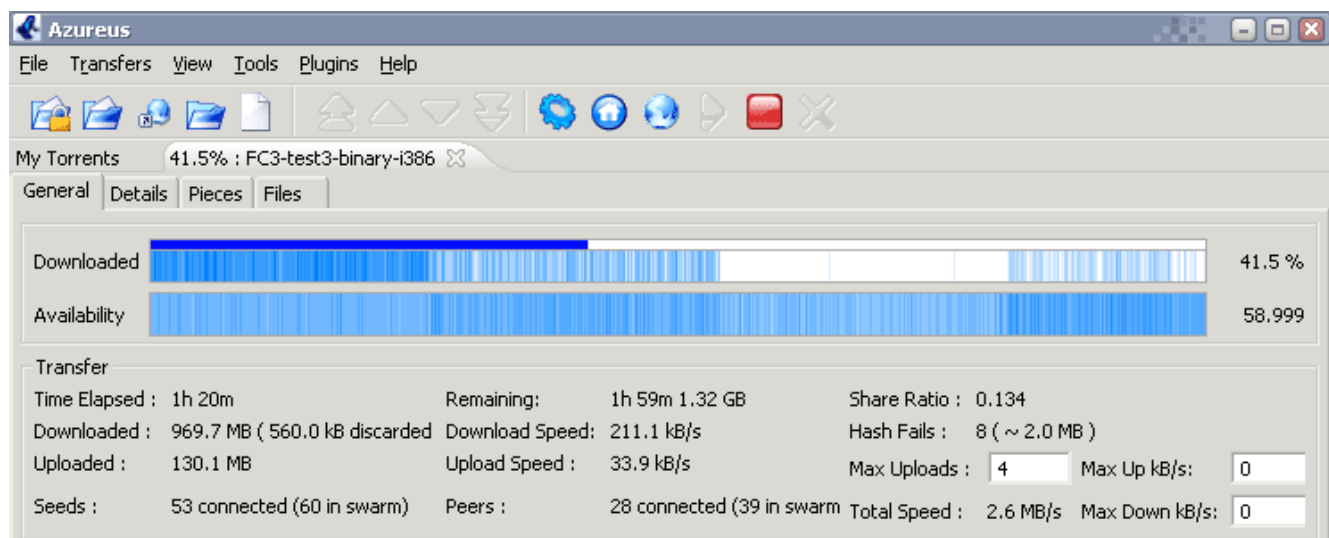
de 0 et de 1) de longueur fixée (souvent 128 ou 160 bits). Ce nombre devient l'« identifiant » du fichier. Les nœuds en charge d'une empreinte sont ceux dont l'identifiant en est très proche.

L'approche classique de recherche par diffusion consiste à propager la requête dans tout le réseau, ce qui engendre un trafic proportionnel au nombre de nœuds (ou de super-nœuds pour les réseaux hiérarchiques). Dans l'approche par tables de hachage distribuées, grâce à la structuration des connaissances, on peut trouver les nœuds en charge de l'empreinte en routant de proche en proche en un nombre d'étapes qui est logarithmique par rapport au nombre de nœuds. La recherche est donc beaucoup plus efficace : typiquement, une dizaine d'échanges de messages au lieu de quelques centaines voire quelques milliers si la donnée cherchée n'est pas répliquée.

Pour signifier le partage d'un fichier, son possesseur doit contacter le nœud en charge de l'empreinte du fichier pour lui indiquer son adresse IP. Un pair qui recherche le fichier n'aura qu'à contacter ce même nœud pour retrouver l'adresse IP du pair (ou des pairs) le partageant.

Le logiciel BitTorrent

Le logiciel BitTorrent, quant à lui, est un autre système d'échange de fichiers. Il se concentre sur le téléchargement d'un fichier donné. L'indexation des fichiers est prise en charge par un autre protocole. Il reprend le principe du téléchargement multiple, mais il permet également de développer la coopération en utilisant le principe d'« un prêté pour un rendu » : pour cela, un algorithme d'incitation à la coopération inspiré de la théorie des jeux est utilisé. Chaque pair BitTorrent, en tant que serveur, donne en priorité aux clients qui lui ont récemment envoyé des morceaux du fichier. Ainsi un pair, en tant que client, a intérêt à donner ce qu'il possède pour recevoir le plus souvent possible des morceaux de fichier.



Téléchargement avec BitTorrent.

À l'instar du réseau eDonkey, BitTorrent utilise la capacité d'« upload » des ordinateurs connectés pour redistribuer le fichier et ainsi désengorger le serveur. Des fichiers de taille importante peuvent donc être distribués par de petits serveurs.

Le logiciel BitTorrent est notamment utilisé pour distribuer certaines versions de Linux. Certaines sociétés éditrices de jeux l'utilisent également pour la diffusion de leurs logiciels et de leurs versions d'essai.

5. Au-delà de la technique, des enjeux de société

Enjeux éthiques

Les réseaux pair à pair sont fréquemment utilisés pour partager des fichiers de façon illicite, sans tenir compte du copyright.

Certes, cet aspect juridique, bien qu'important, ne relève pas de la recherche informatique. Mais celle-ci peut néanmoins contribuer à proposer des solutions techniques à ces problèmes éthiques, permettant de garantir la pérennité des réseaux pair à pair. On pourrait par exemple imaginer un protocole respectueux des droits de propriété

intellectuelle, qui confierait directement la gestion de la rémunération des auteurs aux fournisseurs d'accès.

D'autre part, les réseaux pair à pair peuvent contribuer à la diffusion de contenus illégaux à l'insu de leurs utilisateurs. C'est le cas lorsque l'on aide à partager des fichiers cryptés sans en connaître l'origine ni la destination. Sans le savoir, en téléchargeant un film ou un album de musique, on peut servir de relais à la diffusion d'images pédophiles. C'est un problème qui se pose sur des réseaux qui prônent un anonymat total, comme Freenet par exemple.

Quelle pourra être l'influence de l'organisation de la répression et de sa médiatisation sur l'évolution des usages ? Cela semble difficile à quantifier.

Du reste, les applications des réseaux pair à pair ne fonctionnent pas toutes en marge de la légalité. Ainsi, le logiciel BitTorrent est utilisé de façon purement professionnelle par des sociétés éditrices pour diffuser des logiciels à partir de petits serveurs. De même, une application de téléphonie sur IP utilise les concepts issus de Kazaa. Des groupes de musique peu connus utilisent également les réseaux pair à pair pour diffuser leurs œuvres et tenter ainsi de percer.

Enjeux économiques

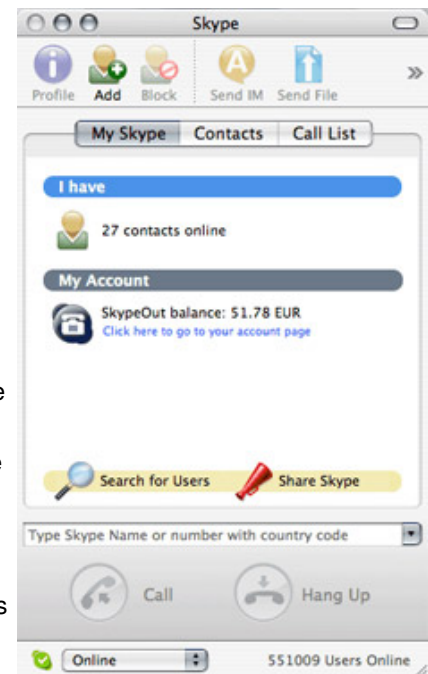
Le développement des réseaux pair à pair a souvent été communautaire, loin de toute démarche mercantile. Kazaa est toutefois une exception, puisqu'il a été développé par une société commerciale. Ses ressources sont issues pour partie de la publicité. Le logiciel installe aussi sur les ordinateurs des internautes des « spyware », logiciels espions susceptibles de recueillir des informations sur l'internaute pour en établir un profil. Ces informations de profils pourraient éventuellement être commercialisées, mais on ne sait pas très bien dans quelles conditions. Ce modèle économique qui vise à rentabiliser l'exploitation de réseaux pair à pair pose de graves problèmes au regard du droit à la protection de la vie privée et des données personnelles.

Enfin, si les logiciels de partage sont souvent proposés gratuitement, des versions plus complètes, mais payantes, sont également proposées aux utilisateurs. Mais peut-on espérer que ce système soit assez attractif pour les internautes et qu'il devienne vraiment rentable ?

Les formidables possibilités offertes par les réseaux pair à pair peuvent encore être améliorées, et ne manqueront sûrement pas de l'être. Mais c'est tout une nouvelle économie qu'il faut penser pour les mettre en phase avec les exigences du droit et de la morale.

Tags

- Réseau
- Utilisateur



Interface de l'application de téléphonie Skype.