Perspective : la sauvegarde croisée

Et si, grâce à la redondance et au chiffrement, les réseaux pair-à-pair assuraient bientôt la sauvegarde de vos données ?

La quantité de données personnelles stockées sur nos disques durs ne cesse d'augmenter. Qu'arrive-t-il quand ils tombent, inéluctablement, en panne ? Mieux vaut prévenir que guérir, et le pair-à-pair est la technique toute indiquée pour remédier à ce type de risque.

Les logiciels de sauvegarde croisée en pair-à-pair ne sont pas encore mûrs mais ce n'est qu'une question de mois. Voyons leur fonctionnement en avant-première.

Pourquoi sauvegarder en pair-à-pair ?

Les technologies utilisées dans nos ordinateurs sont devenues extrêmement complexes : elles nous fournissent des systèmes miniaturisés très bon marché, mais surtout beaucoup plus performants. Les ordinateurs ont ainsi pris une

Chiffres

Nos ordinateurs sont de plus en plus puissants

Nos ordinateurs sont 1 000 fois plus rapides qu'il y a 20 ans, ils peuvent archiver 20 000 fois plus de données. Ainsi, à 3 giga-hertz, un processeur moderne fait trois milliards d'opérations par seconde, et un disque dur moderne peut stocker le prénom de chaque humain sur terre !

place plus importante dans notre vie de tous les jours, y compris à la maison : nous les utilisons pour gérer notre compte bancaire, correspondre avec nos amis, consulter les médias, écouter de la musique ou regarder des films, et de plus en plus, conserver et visionner des documents personnels volumineux, tels que les photos et vidéos de nos dernières vacances.

Les entreprises dépensent des sommes importantes pour sauvegarder leurs données : souvent, celles-ci sont collectées toutes les nuits, transférées dans des endroits lointains (contre l'incendie du bâtiment), puis archivées sur des supports résistant au temps (CD-Rom, bandes magnétiques).

Au contraire, nous oublions trop souvent de prendre le même soin de nos données personnelles, en oubliant que les ordinateurs sont des objets très fragiles : petits, ils peuvent être facilement volés, voire égarés ; portables, ils sont soumis à des chocs, et résistent rarement à une tasse de café ; enfin, tout simplement, ils subissent des pannes, logicielles et matérielles.

Grâce au pair-à-pair, les données personnelles pourront elles aussi bénéficier des mêmes garanties de sauvegarde et de redondance – voire de confidentialité, cela grâce à plusieurs facteurs : la taille croissante de nos disques durs, le nombre grandissant d'ordinateurs par foyer et la démocratisation des connexions permanentes grâce au haut-débit.

> Toujours plus d'espace libre dans nos ordinateurs

Les disques durs de nos ordinateurs contiennent aujourd'hui habituellement au moins 60 Go de stockage, soit 1 000 heures de musique, ou 100 000 livres de poches ! Souvent, plus d'un

quart de cet espace n'est pas utilisé, et pourrait donc servir à sauvegarder les données d'autres personnes.

> De plus en plus d'ordinateurs à la maison

Outre le portable pour travailler, la console pour jouer, il n'est plus rare de trouver aujourd'hui un disque dur dans le lecteur de DVD ou dans le décodeur numérique de la télévision.

> Nos ordinateurs sont connectés en permanence

De plus en plus d'ordinateurs restent allumés et connectés sans interruption : les ordinateurs personnels pour recevoir rapidement ses courriels, ses nouvelles, ou ses appels téléphoniques, mais aussi d'autres ordinateurs cachés, tels que les platines DVD, les décodeurs câble ou satellites à disque dur.

Comment ça marche ?

> Préparation de l'environnement

Supposons qu'une utilisatrice, Vaima, décide de sauvegarder ses données en utilisant un logiciel pair-à-pair qu'elle vient juste d'installer sur son ordinateur. Il va probablement lui être demandé :

1. de s'identifier ;
2. de choisir quoi sauvegarder ;
3. d'offrir de l'espace libre ;
4. d'entrer peut-être dans une communauté.

> S'identifier

Pour s'identifier, Vaima devra probablement choisir un pseudonyme, et un mot de passe pour protéger ses données. Il sera alors très important pour elle de ne jamais oublier ces informations, car elle en aura absolument besoin pour retrouver ses données. Elle les recopie donc sur une feuille de papier, qu'elle conserve toujours dans son sac à main. Si le logiciel utilise une clé privée (voir l'aparté page 143), il sera important qu'elle soit recopiée, sur une clé USB par exemple.

> Choisir les données à sauvegarder

Pour faire son choix, Vaima doit à la fois indiquer dans quels dossiers sont placées ses données personnelles, mais aussi le type des données qu'elle veut protéger. Elle choisit par exemple de ne pas sauvegarder sa musique (elle a les CD de ses albums), pour réserver plus de place à ses photos.

> Donner de l'espace libre aux autres utilisateurs

Les données de Vaima vont être sauvegardées sur les ordinateurs d'autres internautes. Mais il s'agit avant tout d'un échange : parce qu'ils acceptent de stocker ses données importantes, elle doit aussi accepter de stocker les leurs.

RAPPEL L'éthique du pair-à-pair

Le fonctionnement des applications pair-à-pair repose avant tout sur la générosité et réciprocité de l'échange entre utilisateurs. Lorsque le client n'est que demandeur, n'ayant rien à offrir, il doit payer pour les services rendus. Dans le cas de services gratuits, les coûts sont le plus souvent couverts par des sponsors publicitaires. Le pair-à-pair, au contraire, tire sa gratuité d'un échange permanent, le client est lui aussi fournisseur : il profite des ressources des autres pairs, et fournit une partie des siennes aux autres. Veillez donc, vous aussi, à partager vos ressources.

> Entrer dans une communauté

Si le logiciel permet de collaborer en petits groupes (entre amis par exemple), Vaima devra alors saisir le nom de la communauté, un mot de passe pour y entrer, et peut-être l'adresse de l'ordinateur de l'un de ses membres. Sinon, le logiciel placera probablement les données chez des internautes (de quelques dizaines à quelques centaines) choisis au hasard. Cette dernière solution est plus sûre, car les données se trouvent réparties chez un plus grand nombre de gens et sont donc moins vulnérables en cas de panne.

TECHNIQUE Chiffrement à clé publique

Pour protéger les données des regards indiscrets, il est nécessaire de les chiffrer, le déchiffrement n'étant possible qu'avec la *clé* associée. Soit la clé servant à chiffrer est la même que celle servant à déchiffrer – ce qui pose l'embarrassant problème de la transmission sûre de cette clé entre les interlocuteurs, soit deux clés différentes sont utilisées, une clé publique pour le chiffrement, et une clé privée gardée secrète pour le déchiffrement. On parle respectivement de chiffrement *symétrique* et *asymétrique*. Le chiffrement symétrique est extrêmement rapide. L'une de ces techniques, jugée très sûre aujourd'hui, est AES, utilisée dans plusieurs logiciels pair-à-pair (Skype, HiSpread). Le chiffrement asymétrique (dont un exemple célèbre est RSA) est plus lent mais a l'avantage que l'utilisateur possède deux clés, une *privée*, qu'il garde secrète, et une *publique*, qu'il peut diffuser à tous ses correspondants. La combinaison des deux techniques offre à la fois performance et confidentialité des données : les clés symétriques chiffrant les données sont échangées grâce à un chiffrement asymétrique...

> Sauvegarde des données

Le logiciel réserve alors une partie de l'espace libre sur le disque dur de Vaima pour les autres internautes (elle pourra probablement augmenter ou diminuer cet espace en fonction de ses besoins), puis parcourt les dossiers qu'elle désire sauvegarder à la recherche de nouveaux fichiers.

Le contenu de ces fichiers est ensuite copié dans une archive, sur le disque, avec une description de leurs propriétés (ce sont les *méta-données*, telles que la taille, le propriétaire, la date de création, etc.). Vaima peut choisir si cette archive doit être compressée afin de sauvegarder davantage de données, avec l'inconvénient que la récupération d'un fichier particulier sera beaucoup plus lente. L'archive est finalement chiffrée, en utilisant sa clé publique, puis découpée en blocs. Chacun des blocs est alors transmis à un ou plusieurs ordinateurs sur le réseau.

TECHNIQUE Redondance, réplication et codage

Pour augmenter les chances de retrouver les données sauvegardées, une technique classique consiste à répliquer ces données : chaque morceau est copié sur plusieurs ordinateurs. Il peut ensuite être récupéré dès lors qu'un de ces ordinateurs fonctionne encore. L'inconvénient est que, si les données doivent survivre, par exemple, à 4 pannes d'ordinateur (extrêmement probable sur un grand réseau en quelques mois), il faut alors les répliquer 5 fois, et donc offrir 5 Go aux autres pour sauvegarder 1 Go. Une technique intéressante pour éviter cela est le *codage réseau* (*network coding* en anglais) : chaque bloc stocké sur un ordinateur est un codage astucieux de tous les morceaux à sauvegarder. Tous les blocs sont alors équivalents, au sens qu'il suffit de récupérer arbitrairement autant de blocs que de morceaux à l'origine pour recréer tous ces morceaux. Ainsi, pour survivre à la panne de la moitié des ordinateurs du réseau, il n'y aurait plus besoin d'offrir aux autres que le double de l'espace à sauvegarder (soit 2 Go en l'occurrence). Notez que le très répandu système de redondance RAID est un exemple de codage réseau rudimentaire.

Pour que Vaima puisse un jour retrouver chez qui ses données sont hébergées, la liste des pairs concernés est aussi placée sur le réseau, soit dans une base de données centralisée, soit en utilisant un mécanisme distribué, *à la* Overnet.

> Récupération des données

La restauration est presque exactement l'opération inverse : Vaima doit d'abord retrouver chez qui ses données sont hébergées, puis télécharger chaque bloc pour reconstituer l'archive. Son mot de passe et sa clé privée sont alors utilisés pour déchiffrer le contenu, puis l'archive est décompressée : elle peut alors choisir les fichiers à restaurer sur son disque.

Bien sûr, cette opération peut être longue : certains des hôtes hébergeant ses données peuvent avoir éteint leur ordinateur, et il faudra attendre qu'ils le redémarrent... sauf si elle les connaît et peut le leur demander !

Enfin, si certains de ces ordinateurs sont tombés en panne, elle ne pourra récupérer ses données que si celles-ci ont été suffisamment répliquées dans le système (voir l'aparté).

Avantages

La sauvegarde croisée a plusieurs avantages : elle est automatique, gratuite, elle garantit la confidentialité des données, et finalement, elle est assez rapide.

> Une sauvegarde automatique

Même un utilisateur consciencieux a du mal à sauvegarder ses données tout le temps. Les sauvegardes sur CD-Rom prennent du temps et de l'espace (chaque session de gravage utilise au minimum un 30[e] du CD-Rom, même pour un minuscule fichier) : cela incite à attendre pour sauvegarder beaucoup à la fois, ce qui expose les données en attente, perdues au moindre incident.

Heureusement, la sauvegarde croisée apporte une solution à ce problème, puisque, dès que votre ordinateur est connecté, il peut sauvegarder ces données sans action de votre part !

> Une sauvegarde gratuite

Fondée sur un échange d'espace inutilisé entre pairs, la sauvegarde croisée peut être entièrement gratuite – à l'exclusion des coûts de connexion.

> Les données sauvegardées restent confidentielles

Les données sauvegardées sont chiffrées en utilisant une clé que seul votre ordinateur connaît, puis elles sont placées sur d'autres ordinateurs. Alors qu'une sauvegarde non chiffrée sur CD-Rom peut être lue par toute personne ayant accès au disque, l'utilisation de votre ordinateur *et* la connaissance de votre mot de passe sont indispensables pour restaurer les fichiers que vous avez sauvegardés en pair-à-pair.

> Sauvegarde et récupération sont rapides

La sauvegarde en pair-à-pair peut être relativement rapide. Avec une connexion ADSL ou câble (10 Mbit/s entrant, 512 kbit/s sortant), il est possible de sauvegarder entre 2 et 6 Go par jour, soit plus de 2 000 photos. Réciproquement, la restauration des fichiers peut être extrêmement rapide si les ordinateurs hébergeurs sont immédiatement accessibles.

Inconvénients

La sauvegarde en pair-à-pair présente aussi quelques inconvénients : elle est de courte durée, elle est sensible aux vers et les logiciels ne sont pas encore mûrs.

> Une sauvegarde à court terme

La sauvegarde croisée reste une protection de courte durée pour les données. Alors qu'un CD-Rom ou un DVD peut résister à peu près cinq années, une sauvegarde en pair-à-pair est très sensible aux départs des autres internautes, qu'ils soient délibérés ou dus à des pannes logicielles et matérielles. Ainsi, une fois les fichiers effacés de votre disque, la possibilité de les récupérer sur le réseau diminue considérablement après quelques semaines. Il est donc conseillé de continuer à effectuer des sauvegardes régulières sur support durable (CD-Rom, DVD...) tout en utilisant la sauvegarde croisée pour sa rapidité et son confort.

> Le danger des vers

Le ver (voir aparté) est l'ennemi de la sauvegarde croisée. En effet, lorsqu'un ver se propage sur l'Internet, il provoque des pannes sur des milliers d'ordinateurs : ceux-ci doivent souvent être déconnectés du réseau assez longtemps, voire subir une réinstallation de leur système. Cela signifie que vos données hébergées sur ces ordinateurs sont temporairement, voire définitivement inaccessibles, au moment où vous avez le plus besoin d'y avoir accès pour restaurer vos propres données perdues !

TERMINOLOGIE Les vers

Un ver (w*orm* en anglais) est un virus qui se propage sans action humaine, contrairement aux autres virus, qui s'activent souvent lorsque l'utilisateur lance un programme ou lit un courrier électronique contaminé. Les vers sont extrêmement dangereux, car ils peuvent provoquer la panne simultanée de centaines de milliers d'ordinateurs sur Internet – ô combien catastrophique pour un système de sauvegarde pair-à-pair. Dans le passé, un ver tel que Blaster a contaminé près de 500 000 ordinateurs, tandis qu'un autre, Slammer, doublait le nombre de machines qu'il avait contaminées toutes les 8,5 secondes, rendant toute communication entre ordinateurs impossible dans de nombreuses entreprises.

> Les logiciels encore immatures

Peu de logiciels existent actuellement pour la sauvegarde croisée, et ceux qui existent sont encore peu connus, peu testés et peu utilisés. Il est donc assez probable qu'ils contiennent encore des erreurs. Néanmoins, utilisés prudemment, c'est-à-dire en complément des sauvegardes classiques régulières, ces systèmes vont progressivement s'améliorer.

Sécurité

La sauvegarde croisée en pair-à-pair semble dangereuse pour la sécurité de votre ordinateur, parce que vos données personnelles se retrouvent hébergées sur les ordinateurs d'inconnus, tandis que votre disque dur contient des fichiers appartenant à ces inconnus. Pourtant, les risques associés sont très faibles.

> Confidentialité des données

Vos données sont transmises sur les autres ordinateurs après avoir été chiffrées puis découpées : chaque ordinateur n'a donc, d'une part, qu'une toute petite partie de vos données, et d'autre part, ne connaît pas la clé à utiliser pour les déchiffrer. Or, une méthode de chiffrement telle que AES, souvent utilisée par ce genre de logiciels, est aujourd'hui considérée comme

pratiquement sûre, tant que la clé utilisée reste secrète. Cette méthode est aussi utilisée abondamment par les banques et autres organismes financiers. Finalement, seule une personne ayant accès à votre ordinateur peut avoir accès à votre clé privée... mais celle-ci est normalement aussi chiffrée sur votre disque dur, en utilisant votre mot de passe secret, que vous ne devez taper que pour restaurer vos données ! Ces données sont donc bien protégées des regards étrangers...

> Fichiers dangereux (virus)

La prudence est de règle lorsque des fichiers d'origine inconnue – et même connue – sont téléchargés sur votre ordinateur. Ainsi, la moitié des ordinateurs connectés à Internet seraient infectés par des logiciels espions (*spyware*) ou des virus, ralentissant leur fonctionnement ou endommageant leurs données, et leur origine est la plupart du temps l'exécution de fichiers téléchargés sur Internet. Néanmoins, dans le cas de la sauvegarde croisée, le risque est nul : même si le contenu d'un fichier est dangereux, celui-ci est chiffré, sa lecture directe est donc impossible, et la probabilité pour qu'un contenu chiffré soit dangereux pour votre ordinateur est infime.

Quelques systèmes de sauvegarde croisée

Il existe déjà plusieurs projets de logiciels permettant la sauvegarde croisée en pair-à-pair.

> HiSpread

Ce logiciel (http://www.hispread.com/), payant mais avec un code source ouvert, fonctionne uniquement sous Windows, et permet des sauvegardes croisées avec réplication et codage réseau limité, entre les membres d'un « groupe ». Le groupe, identifié par un nom et un mot de passe, peut être réparti sur un réseau local, ou sur Internet, la connaissance de l'adresse d'un autre membre étant alors requise pour y accéder.

> AllMyData

Ce tout nouveau logiciel (http://www.allmydata.com/), également uniquement disponible sous Windows, propose des sauvegardes croisées globales, gratuitement, moyennant un échange de dix méga-octets d'espace libre pour chaque méga-octet sauvegardé.

> Palabre

Palabre (http://backup.palabre.net/) est un projet de logiciel libre de l'INRIA visant à fournir, en un seul logiciel multi-plate-forme, un ensemble de services pair-à-pair pour le particulier, allant de la publication de photographies à la sauvegarde croisée, en utilisant les technologies les plus avancées dans ce domaine.

ATTENTION Rien n'est jamais sûr

Le chiffrement n'est pas une garantie en lui-même, car les protocoles utilisés peuvent contenir des failles de sécurité importantes. De même, la réplication ne fait *qu'améliorer* les chances de retrouver ses données.

ALTERNATIVES Sauvegarde par systèmes centralisés

Il existe des services payants de sauvegarde centralisée sur serveur, susceptibles d'offrir des garanties plus fortes sur l'intégralité de la restauration de vos données, mais plus faibles quant à leur confidentialité. Ces systèmes fournissent un logiciel qui sauvegarde régulièrement sur un serveur, avec une tarification variable, reposant soit sur un loyer, soit sur la facturation de la récupération des données – en fonction, dans les deux cas, de la quantité de données à protéger.