Zeitschrift: Reihe Kriminologie / Schweizerische Arbeitsgruppe für Kriminologie = Collection criminologie / Groupe suisse de travail de criminologie

Herausgeber: Schweizerische Arbeitsgruppe für Kriminologie

Band: 24 (2006)

Artikel: Identification des diffuseurs de fichiers illégaux dans les réseaux de partage de fichiers "peer-to-peer"

Autor: Seeger, Pascal

DOI: https://doi.org/10.5169/seals-1051082

PASCAL SEEGER

IDENTIFICATION DES DIFFUSEURS DE FICHIERS ILLÉGAUX DANS LES RÉSEAUX DE PARTAGE DE FICHIERS «PEER-TO-PEER»

Résumé

Le «peer-to-peer» est un ensemble de réseaux de partage de ressources décentralisé, communautaire et volontaire. Il permet l'échange de tous types de fichiers informatiques. Il est dit «décentralisé» car chaque internaute, utilisant ce type de programme, devient un émetteur de fichiers à partager pour les autres utilisateurs, lesquels deviennent à leur tour des émetteurs des fichiers stockés sur leur ordinateur. Il est dit «communautaire» car l'ensemble des internautes utilisateurs transitent par des serveurs dans lesquels ils trouvent l'information nécessaire afin de se connecter directement à l'ordinateur d'un autre internaute (adresse Internet ponctuelle, liste des fichiers en partage, etc). Il est dit «volontaire» car d'une part l'internaute choisit les fichiers qu'il entend mettre à disposition de la communauté de partage et d'autre part le téléchargement de fichiers désirés est le résultat d'une démarche intentionnelle de sa part quant à leurs natures et à leurs noms. Il existe plusieurs systèmes de peer-to-peer disponibles sur Internet, utilisant des protocoles de communication différents (eMule, eDonkey, Kazaa, Grokster, Bit-torrent, etc). Dans la phase initiale de recherches, la personne chargée d'une investigation utilise un logiciel public (GNU) ou propriétaire, spécialement adapté à l'identification de diffuseurs de fichiers illégaux sur le P2P. Ces recherches s'effectuent sur la base de mots-clés préalablement définis et connus comme étant susceptibles d'être contenus dans le nom du document, de l'image ou de la vidéo (ex: pedo ou childlover). Une fois le résultat de recherche acquis, l'enquêteur demande la mise en téléchargement de ces fichiers. Lors de ce processus, des informations d'identification et constitutives de preuve sont collectées, principalement l'adresse IP (Internet Protocol) de l'émetteur, la date et l'heure de début et de fin de téléchargement et le fichier téléchargé. Il est de rigueur de procéder à un filtrage des adresses IP des fournisseurs d'accès Internet pour se concentrer uniquement sur les internautes se trouvant dans la juridiction de l'enquêteur. Aussi, il est possible d'empêcher le logiciel servant à l'investigation de diffuser lui-même les contenus illicites téléchargés. De telles solutions techniques d'investigations sont utilisées par les forces de l'ordre mais également par le secteur privé pour la défense de la propriété intellectuelle et des droits d'auteur.

Identifikation von Personen, die illegale Dateien in elektronischen Tauschbörsen verbreiten (Peer-to-Peer)
«Peer-to-Peer» stellt die Gesamtheit von Tauschnetzen mittels dezentralisierter, gemeinschaftlicher und freiwilliger Ressourcen dar. Es erlaubt den Austausch jeder Art von Datei. Dezentralisiert, weil jeder Nutzer, der ein solches Programm benutzt, gleichzeitig ein Anbieter von Dateien für andere Nutzer wird, die ihrerseits zu Anbietern von Dateien werden, die auf ihrem Computer gespeichert sind. Gemeinschaftlich, weil die Nutzer über Server geleitet werden, wo sie die notwendige Information finden, um sich mit einem anderen Nutzer direkt zu verbinden (genaue Internet-Adresse, Dateiliste etc.). Freiwillig, weil der Nutzer einerseits die Dateien wählt, die er zum Tausch zur Verfügung stellt, und andererseits das Herunterladen von Dateien hinsichtlich Namen und Charakter von seinem Willen abhängt. Es existieren mehrere Peer-toPeer-Netze auf dem Internet, die unterschiedliche Kommunikationsprotokolle verwenden (eMule, eDonkey, Kazaa, Grokster, Bit-torrent, etc). In der Anfangsphase benutzt die mit einer Untersuchung beauftragte Person eine Public Domain Software (GNU) oder ein anderes Programm auf die Identifizierung von Verbreitern illegaler Dateien via P2P spezialisiertes Programm. Die Untersuchung basiert auf vorgängig definierten Schlüsselwörtern, die als Bezeichnung entsprechender Dateien bekannt sind (z.B. Pädo oder Childlover). Einmal aufgefunden, versucht der Ermittler, die Datei herunterzuladen. Am Rahmen dieses Prozesses werden Informationen zur Identifikation gesammelt, v.a. die IP-Adresse (Internet Protocol) des Anbieters, das Datum und die Uhrzeit von Beginn und Ende des Downloads und die heruntergeladene Datei. Notwendigerweise müssen die IP-Adressen sodann nach Zugangsprovidern gefiltert werden, um sich ausschliesslich auf jene Nutzer zu konzentrieren, die sich im Anwendungsbereich des nationalen Rechts des Ermittlers befinden. Möglich ist, die Software, die zur Ermittlung dient, so zu konfigurieren, dass sie selbst keine illegalen Dateien anbietet. Solche technischen Lösungen werden von den Strafverfolgungsbehörden, aber auch von Privaten zum Schutz von Immaterialgüterrechten genutzt.

Introduction

A l'époque où la légalisation du téléchargement et de la diffusion de musiques au format MP3, qualifiées de «copies privées», sur les réseaux de peer-to-peer est largement débattue en France, il ne faut pas perdre de vue qu'il reste d'autres natures de fichiers qui sont définitivement illégales comme l'illustration d'actes sexuels commis sur des enfants, les documents d'appel à la haine raciale ou au terrorisme. Identifier les ordinateurs, et bien sûr leurs utilisateurs, propagateurs

de ces fichiers illicites à travers le monde est aujourd'hui possible par le biais de solutions techniques mises en place dans plusieurs services de polices en Europe. Des sociétés privées ne sont également pas en reste et proposent des prestations quasi identiques.

Cet article donne une vision non exhaustive de la structure de ce média, des acteurs qui l'animent et des moyens d'enquêtes sur ces réseaux P2P. Il a pour ambition de démontrer l'ampleur d'un phénomène grandissant ainsi que la facilité de la commission d'un délit tout en affirmant qu'il est possible d'agir et de punir les diffuseurs de ces fichiers illégaux.

Notions de base

Pour une bonne compréhension, il est important de connaître la notion d'adresse IP. Sur Internet, constitué d'un ensemble de multiples réseaux interconnectés, les ordinateurs communiquent entre eux grâce au protocole TCP/IP[1] qui utilise des adresses numériques, appelées adresses IP. Elles sont composées de quatre nombres entiers entre 0 et 255, séparés par un point. Ces adresses servent aux ordinateurs du réseau pour se localiser afin d'acheminer les paquets de données. Utilisons une métaphore sous la forme d'une enveloppe affranchie pour expliquer l'acheminement des données entre ordinateurs sur Internet. Il sera aussi plus simple par la suite de comprendre le filtrage des adresses IP dans ce type d'investigation sur Internet. Prenons par exemple l'adresse IP no 62.50.74.249. Avec les deux premiers nombres 62.50, il est possible de déterminer le réseau concerné ou pour une adresse conventionnelle, il s'agirait du pays et la ville d'origine. Les nombres suivants 74.249 permettent de localiser l'ordinateur dans ce réseau, sur l'enveloppe, il s'agit du domicile et du nom du destinataire. Le timbre serait le fournisseur d'accès Inter-

1 «Transmission Control Protocol/Internet Protocol» représente d'une certaine façon l'ensemble des règles standardisées de communication sur Internet et se base sur la notion adressage IP. http://www.commentcamarche.net/internet/tcpip.php3

net rémunéré pour son travail de transmission et le tampon de la poste représenterait la date et l'heure d'utilisation. Pour revenir à un aspect plus réel et plus technique, il faut savoir que les fournisseurs d'accès Internet (FAI), au début de l'exploitation de leur entreprise, obtiennent des lots (range) d'adresses IP de l'organisme RIPE NCC (réseaux IP Européens). Cette organisation est sous l'égide de L'ASO (Address Supporting Organization) qui gère également l'ARIN (American Registry for Internet Numbers) et l'APNIC (Asia-Pacific Network Information Center). L'ASO est quant à elle administrée par l'ICANN[2] qui s'assure de la coordination de tous les standards d'Internet.

Dans la grande majorité des cas, un internaute s'adresse à un FAI de son pays ou de sa région pour obtenir un accès à Internet. De ce fait, il s'acquitte d'un abonnement payant, ce qui favorise indubitablement son identification éventuelle par les autorités. L'adresse IP peut être fixe (généralement les entreprises mais de plus en plus de particuliers) ou dynamique ce qui est encore le plus courant. Dans ce dernier cas, le FAI alloue temporairement une adresse IP pour avoir une existence sur Internet et bénéficier de tous les services. C'est un peu comme un ticket d'entrée pour un parc d'attractions que l'usager conserve jusqu'au moment de quitter les lieux et qui sera remis au client suivant.

Ces fournisseurs d'accès sont tenus de détenir et de contrôler l'accès aux informations de connexion sur la base des textes régissant dans chaque pays l'exploitation d'une telle activité de communication. En Suisse, ces journaux d'activités, appelés «logs», sont conservés pen-

2 L'ICANN (Internet Corporation for Assigned Names and Numbers) a été créée en octobre 1998 à l'initiative du gouvernement américain. Sa mission est de succéder au gouvernement américain dans l'administration de l'Internet. Elle doit traiter à ce titre des questions relatives aux noms de domaines, mais aussi aux adresses IP et aux protocoles permettant aux machines de communiquer entre elles. L'ICANN est une organisation internationale: ses équipes et ses dirigeants sont des personnes de tous pays et disposant d'un large éventail de compétences. Elle est la plus haute autorité internationale pour toutes les questions liées aux noms de domaines, adresses et protocoles. Juridiquement, l'ICANN est une société à but non lucratif fonctionnant selon les lois de l'Etat de Californie. http://www.gouvernance-internet.com.fr/information/faq-icann.html

dant 6 mois dans leur système informatique. Il faut savoir que l'architecture du réseau Internet ne permet qu'une seule connexion par adresse IP à un moment donné.

Néanmoins, il est très facile de nos jours d'être anonyme sur Internet, sans de grandes connaissances techniques. Il suffit de se rendre dans un cybercafé ou d'employer simplement des connexions sans fil (WIFI) ouvertes et gratuites[3]. De telles connexions libres sont disponibles depuis chez soi également si son voisin a installé un point d'accès sans fil sans en sécuriser son utilisation par une protection (clé WEP ou WPA). Il est aussi difficile d'effectuer un tri par pays si l'internaute utilise un opérateur multinational comme AOL.

Définition

Le P2P est un système *décentralisé, communautaire* et *volontaire* de partage de ressources. Il permet l'échange de tout type de fichiers informatiques tels que musiques, vidéos, logiciels ou textes.

Il est dit *«décentralisé»* car chaque internaute, utilisant ce type de programme, devient un émetteur de fichiers à partager pour les autres utilisateurs, lesquels deviennent à leur tour des émetteurs de ces fichiers stockés sur leur propre ordinateur. C'est le principe de l'échange.

Qualifié de *«communautaire»* puisque l'ensemble des internautes utilisateurs transitent par des serveurs dans lesquels ils trouvent l'information nécessaire afin de se connecter directement à l'ordinateur d'un autre internaute (adresse Internet ponctuelle, liste des fichiers en partage, etc.), il répond au principe de la communauté d'échange.

3 http://www.freespots.ch

Il est déclaré *«volontaire»* attendu que d'une part l'internaute choisit les fichiers qu'il entend mettre à disposition de la communauté de partage et d'autre part le téléchargement de fichiers désirés est le résultat d'une démarche intentionnelle de sa part quant à leurs natures et à leurs noms.

Les époques successives du peer-to-peer

Historiquement, le peer-to-peer découle d'une volonté de partager des ressources de calcul. Prenons le cas du projet SETI@home[4] qui est une expérience scientifique employant les ordinateurs de particuliers connectés à Internet. Ces ordinateurs personnels, une fois en mode veille, sont utilisés pour la recherche d'intelligence extraterrestre par l'analyse de données d'un radiotélescope, envoyées et reçues par segments depuis un serveur central.

C'est véritablement Napster en 1999 qui a lancé la mode du peer-to-peer grand public. S'il a été au départ développé dans une autre optique par son créateur, SHAWN FANNING, un américain de 18 ans qui souhaitait simplement échanger de la musique avec ses amis, ce système devient rapidement dédié au téléchargement de médias musicaux sur Internet en s'appuyant sur la technologie P2P centralisée. Dès la première semaine de lancement du service, 15 000 personnes ont téléchargé le logiciel puis 23 millions en juillet 2000. L'augmentation fulgurante du nombre d'utilisateurs a provoqué un accroissement considérable du nombre de chansons disponibles. Cette vague a déferlé rapidement sur Internet comme un raz-de-marée sur les surfeurs du Web et le phénomène s'est imposé.

En comblant certaines lacunes techniques et consécutivement aux déboires juridiques[5] de Napster, le logiciel Kazaa s'assure un succès

4 http://www.seti.org
5 La Saga Napster http://www.journaldunet.com/dossiers/musique/napster1.shtml

et une popularité tout aussi rapides. La possibilité de reprendre un téléchargement interrompu et le fait de pouvoir télécharger un fichier depuis plusieurs sources afin d'augmenter la vitesse sont les atouts majeurs de cette nouvelle génération. L'évolution de l'Internet haut débit (ADSL) favorise considérablement cette expansion. Les internautes laissent leur PC connecté en permanence ce qui provoque une augmentation de la disponibilité temporelle des fichiers mais aussi quantitative. Le peer-to-peer mute ainsi vers le partage de ressources de stockage.

En 2003, le programme eDonkey2000 et ses descendants comme eMule et Overnet surpassent Kazaa et prennent le relais dans les habitudes des utilisateurs. Les développeurs de ces logiciels reprennent la technique du fractionnement des fichiers, dès qu'un téléchargement est commencé que la partie récupérée est déjà disponible à l'envoi. Une multitude d'autres logiciels, plus ou moins clonés, apparaissent sur le réseau (Kazaa Lite, Bearshare, WinMX, LimeWire, Shareaza).

C'est à cette époque que l'on assiste à un changement des mentalités de certains utilisateurs. Ils se sentent de plus en plus acteurs et ont vocation à alimenter le réseau et à être reconnus pour cela. Ils signent de leur pseudonyme les fichiers mis à disposition et se regroupent en équipes. Ces signatures deviennent un gage de qualité des fichiers pour les «téléchargeurs» et les «teams» gagnent en prestige. Certains en viennent même à acheter des cédéroms originaux pour avoir le bénéfice de les mettre en premier à disposition sur le peer-to-peer après en avoir retiré les éventuelles protections.

La dernière génération de logiciels représentée par Bittorent et Grabit optimise la bande passante[6] en envoi et réception pour un débit maximal en flux continu. Ces nouvelles applications découpent

6 Débit d'informations d'un media de communication généralement mesuré en octets (bytes) par seconde ou en bits par seconde (bit/s ou bps). Terme plus communément utilisé par les fournisseurs d'accès Internet pour donner le débit maximum d'un abonnement.

systématiquement les fichiers qui sont par conséquent moins lourds et plus rapides à télécharger.

Issus généralement des milieux soucieux de ne pas être pris à défaut par Big Brother[7], des réseaux comme Freenet utilisent la cryptographie pour garantir l'anonymat des utilisateurs et de leurs données. Dotés de systèmes de chiffrement variés, ces logiciels garantissent à leurs utilisateurs une confidentialité presque parfaite dans leurs échanges. Ainsi les autorités ne peuvent que très difficilement remonter le réseau pour arrêter les éventuels coupables. Le chiffrement se fait sur un système de clé publique et privée[8]. Heureusement pour l'instant, ces réseaux sont peu utilisés par les adeptes du peer-to-peer car ils sont tout de même contraignants à mettre en place.

Au fil du temps, essentiellement les majors de la musique ont œuvré, procès après procès pour condamner cette activité. Du fait de la pression exercée par ceux-ci, on remarque qu'aucun des logiciels de peer-to-peer ne perdure vraiment, exceptés les logiciels libres puisqu'ils sont développés par une communauté indépendante et disséminée sur le globe. Au mieux, ces programmes rentrent dans le rang des logiciels commerciaux et légaux. La majorité des utilisateurs délaissent les logiciels dans le collimateur de la justice pour d'autres moins exposés.

Venus tardivement dans le monde des applications développées pour Internet, il faut relever que ces logiciels sont très intuitifs et de fait faciles d'emploi pour le béotien en informatique. En quelques clics, l'internaute intègre une communauté d'utilisateurs du peer-to-peer, ce qui lui donne accès à une gigantesque bibliothèque de fichiers à télécharger et à distribuer.

---

7 Big Brother est un personnage fictif, créé par le romancier britannique GEORGE ORWELL pour son roman 1984. Big Brother est devenu la représentation de l'État policier et inquisiteur.

8 Dans un système de chiffrement à clé publique, les utilisateurs choisissent une clé aléatoire qu'ils sont seuls à connaître (il s'agit de la clé privée). A partir de cette clé, ils déduisent chacun automatiquement un algorithme (il s'agit de la clé publique). Les utilisateurs s'échangent cette clé publique au travers d'un canal non sécurisé. http://www.commentcamarche.net/crypto/clepublique.php3

Les différents types de réseaux

Les réseaux numériques que les données des logiciels de P2P empruntent ont des dénominations particulières telles que FastTrack, Gnutella, Bit-Torrent ainsi que FreeNet. Ils se répartissent en plusieurs grandes catégories selon leur architecture.

Dans l'architecture centralisée quelque peu controversée sur son appartenance au P2P, le logiciel de l'utilisateur se connecte à un serveur unique qui gère les partages et la recherche. Les fichiers transférés ne passent cependant pas par le serveur central. C'est la solution la plus fragile puisque le serveur est indispensable au réseau. Ainsi, s'il est supprimé, à la suite d'une action en justice par exemple, comme ce fut le cas avec Napster et Audiogalaxy, tout le réseau s'effondre.

L'architecture décentralisée permet de résister à de telles mesures puisqu'un logiciel P2P ne se connecte pas à un unique serveur mais à d'autres programmes comme le sien qui ont le double rôle de client et de serveur. C'est d'ailleurs pour cela que dans ce type de réseau, le logiciel est appelé «servent» parce qu'il demande et fournit des fichiers. Le système est ainsi plus robuste mais la recherche d'informations est plus fastidieuse. Dans des réseaux comme Gnutella, cette dernière nécessite un nombre de requêtes élevé, proportionnel au nombre d'ordinateurs du réseau et exponentiel suivant la profondeur de recherche. Néanmoins, des protocoles optimisés ont pu être mis en place, basés sur les tables de hachage distribuées[9], permettant de réaliser des recherches en un nombre de messages croissant en relation avec le nombre d'éléments du réseau, comme CAN, Chord, Freenet, GNUnet, Tapestry, Pastry et Symphony.

9 Une table de hachage distribuée (ou DHT pour Distributed Hash Table), est une technologie permettant l'identification et l'obtention, dans un système réparti, comme certains réseaux P2P, d'une information. L'ensemble de la table de hachage est constituée virtuellement par tous ces constituants répartis sur tous les éléments du réseau, qui en possèdent chacun une partie. http://fr.wikipedia.org/wiki/Table_de_hachage_distribu%C3%A9e

Une autre solution a été imaginée: les «superpeers» ou «super-nœuds». Ce sont des éléments du réseau sélectionnés en fonction de leur puissance de calcul et de leur bande passante qui réalisent diverses fonctions comme l'indexation des informations et le rôle d'intermédiaire dans les requêtes. Cette solution employée dans les réseaux FastTrack, comme Kazaa, rendent ceux-ci un peu moins robuste car les cibles à «détruire» pour que le réseau devienne inopérant sont moins nombreuses que dans un réseau de type Gnutella. Du même ordre d'idée, le réseau eDonkey2000 utilise des serveurs spécialisés, vulnérables mais nombreux, qui interagissent d'une façon similaire aux super-nœuds de FastTrack. Prenons le cas de tels serveurs comme Razorback 2.0 et 2.1 qui, avant l'arrestation en février 2006 de l'animateur principal, un Suisse de 36 ans, indexaient plus de 100 millions de fichiers en partage (musiques, images, films anciens ou récents, logiciels, jeux, émissions de télévision, etc.). Jusqu'à 1,3 million de personnes se connectaient pour référencer le contenu de leur répertoire partagé.

Dans le cadre du réseau DirectConnect, il y a une multitude de serveurs, nommés les Hubs, auxquels les clients référencés se connectent. Tout un chacun peut héberger un Hub, ce qui a mené le réseau à une notoriété rapide. Les Hubs hébergent entre deux et plusieurs centaines d'utilisateurs sans dépasser le millier généralement. Les Hubs requièrent la plupart du temps un quota de fichiers à respecter afin de pouvoir entrer alors que certains sont strictement privés et demandent une inscription. En plus de proposer un échange de fichiers anonyme comme le proposent de nombreux autres programmes, DirectConnect veut fonder une communauté autour de l'échange. Ainsi, chaque Hub propose un dialogue par un Chat central. La recherche, bien qu'existante, laisse la place à une navigation dans les fichiers d'autrui, segmentée par dossiers. Les Hubs cherchent donc à se regrouper par centres d'intérêts, ce qui orientera les utilisateurs vers un ensemble de fichiers qu'ils aimeront probablement.

Il existe principalement deux logiciels clients pour accéder aux Hubs DirectConnect: l'officiel de Neo-Modus et DC++. Outre ces deux programmes, MLdonkey peut également se connecter bien que ce client soit souvent très mal admis par les utilisateurs et les hébergeurs de Hubs.

Il existe aussi des systèmes mixtes qui utilisent des protocoles comme OpenNap permettant de se connecter simultanément à plusieurs autres réseaux.

Vous trouverez ci-dessous un résumé des différents réseaux et des logiciels de P2P.

Réseaux	**Logiciels (servent)**
BitTorrent	BitTorrent Générique, Azureus, ABC, BitComet, BT++, eXeem, PTC, BitTornado (Shadow's Experimental), TorrentStorm, Transmission, Shareaza, WinMobile Torrent
Gnutella	Limewire, Shareaza, Acquisition (Mac), BearShare, Cabos (Mac: Aquisition + Limewire), Gnucleus, Morpheus, mlmac, Poisoned, PeerCast, Phex, Swapper, XoloX
Napster	OpenNap, mlmac, Poisoned
FastTrack	Kazaa, Grokster, iMesh, gIFT, mlmac, Poisoned
eDonkey2000	eDonkey2000 (regroupement eDonkey2000 – Overnet), mlDonkey, eMule, xMule (eMule pour linux), aMule (multiplateforme), Shareaza
MP2P	Piolet, Blubster, RockItNet
Freenet	Frost, Fuqid, Freemail, Mute, Spider, Winny
Direct Connect	Direct Connect, DC++, BlackDC, oDC, rmDC, DC Pro
Ares Galaxy	Ares (Galaxy ou Lite), Warez P2P, FileCroc
Autres réseaux	Akamai, Alpine, ANts_P2P, bwa, CAN, Carracho, Chord, Dexter, Evernet, Filetopia, Groove, Hotwire, IRC, JXTA, Kademlia, MojoNation, Nodezilla, FreePastry, Scribe, Soulseek, Swarmcast, Symphony, Tapestry, TribalWeb, WinMX.

Les principes de fonctionnement du logiciel «servent»

Lors de l'installation d'un logiciel de P2P, un répertoire de l'ordinateur de l'utilisateur est partagé pour tous les autres utilisateurs du même logiciel ou du même réseau. Basé sur l'échange, chacun doit

donner pour recevoir. Ainsi, un système de crédits récompense les utilisateurs qui émettent. La quantité de données transférées détermine la quantité de crédits. Les crédits ne sont pas globaux, ils ne peuvent être utilisés que chez le client qui vous les a accordés. Les crédits sont un modificateur majeur lors du calcul de la progression dans la file d'attente d'un autre client. Plus vous avez de crédits, plus vous avancez vite.

Chaque utilisateur a un pseudonyme par défaut qu'il peut changer dans les paramètres du logiciel. Nous verrons par la suite que cette donnée d'identification personnelle est complétée par une signature encore plus forte, à savoir une empreinte numérique de l'utilisateur, qui n'est pas modifiable.

Si nous prenons le cas du logiciel eMule, comme la plupart des logiciels de P2P, chaque programme se connecte à un premier serveur pour accéder au réseau. Lorsque le servent établit une connexion avec un serveur, ce dernier vérifie si les autres clients peuvent librement communiquer avec lui. Si la réponse est affirmative, le serveur assigne au client ce que l'on appelle une ID[10] forte (high ID). Si la communication est bloquée, le serveur lui assigne une ID faible (low ID). Une fois l'ID attribuée, le client eMule envoie au serveur une liste de tous ses fichiers partagés dans le répertoire local de l'ordinateur. Le serveur ajoute les noms des fichiers et leurs valeurs de hachage dans sa base de données. Tous les fichiers reçoivent une telle valeur qui est une combinaison de chiffres et de lettres qui permet l'identification unique de chaque fichier. Un fichier peut avoir de nombreux noms, mais cela ne modifie en rien sa valeur de hachage. Nous verrons cette notion plus loin dans le détail. Chaque utilisateur a donc la possibilité de trouver toutes les sources d'un fichier précis, quels que soient les noms qui lui sont donnés.

10 L'ID est une valeur calculée à partir de l'adresse IP du client. Elle est assignée par le serveur, lorsqu'e Mule se connecte à lui avec succès. L'ID indique s'il est possible ou non de mettre deux clients en communication directe.

Lorsqu'un internaute désire télécharger un fichier sur le P2P, le processus suit un certain nombre d'étapes. Tout d'abord, il effectue une recherche par mot-clé sur l'un des serveurs communautaire et obtient un nombre variable de réponses lui décrivant la disponibilité du fichier recherché (nombre de sources) ainsi que de multiples informations comme le nom du fichier, sa taille, son type (vidéo par exemple) et son format informatique (MP3, MPEG, AVI, etc.) et encore d'autres détails. Chaque source correspond à un ordinateur détenant le fichier en question dans son répertoire partagé.

Ensuite, l'internaute choisit un ou plusieurs des fichiers proposés et demande sa mise en téléchargement par un simple clic. Il obtient ainsi du serveur la liste et l'adresse IP des internautes émetteurs. Relevons que l'adresse IP est généralement cachée à l'utilisateur dans l'interface de visualisation. Le système effectue une demande de téléchargement sur chacun des «ordinateurs sources». Chaque nouvel internaute nouvellement connecté qui détient le fichier requis est automatiquement rajouté comme une source. Le demandeur est ainsi placé sur une liste d'attente en fonction de l'ordre d'arrivée de sa requête sur chacun des ordinateurs «sources».

Il est important de comprendre que le téléchargement en lui-même n'est pas affecté par le choix du réseau. La topologie du réseau est seulement relative à la recherche des fichiers et des clients qui en sont les sources. Une fois la source trouvée, le logiciel la contacte. Elle vous réserve alors une place dans sa file d'attente pour ce fichier précis. Lorsque la première place de cette file est attribuée à votre logiciel, après un certain temps d'attente, il est autorisé à recevoir des données. Une fenêtre des transferts donne différentes informations sur les taux courants de téléchargement et d'émission. Dans l'interface de visualisation du logiciel, la partie supérieure de la fenêtre est réservée aux téléchargements. La partie inférieure affiche clairement les morceaux de fichiers émis (upload) à un ou plusieurs autres demandeurs. Il est par conséquent difficile qu'un internaute évoque qu'il ne savait pas que le logiciel utilisé diffusait des fichiers à des tiers.

Tous les fichiers sont découpés (hachés) en morceaux identifiés et numérotés afin de pouvoir télécharger simultanément plusieurs parties distinctes de ce fichier auprès de sources différentes. Une fois que tous les morceaux ont été téléchargés, le système reconstitue l'original complet qui sera alors utilisable par l'internaute demandeur.

L'empreinte numérique

De nombreux réseaux de P2P utilisent la technique du hachage (ou hashing). Le hachage d'un fichier peut être comparé à une empreinte digitale. Cette technique utilise des algorithmes qui peuvent être publics (MD4, MD5, CRC32 ou SHA-1) ou encore privés comme par exemple la Théraographie développée par la société Advestigo SA. Ces résultats mathématiques permettent une identification, unique et certaine, du contenu de ce fichier, en n'utilisant qu'une faible quantité de données. Grâce à la Théraographie, il est possible d'effectuer une comparaison d'empreintes par similitude si le fichier a été modifié ou altéré d'une quelconque manière.

Le protocole eDonkey/eMule utilise l'algorithme MD4 (message digests 4), dont la longueur est fixe (128 bits). Quelle que soit la taille du fichier, 10 Mo (méga octets[11]), 200 Mo ou 1,5 Go (giga octets), la valeur de hachage fait toujours la même longueur. Il faut bien préciser que deux fichiers de même taille mais dont les contenus ou les formats diffèrent ont des valeurs de hachage différentes. Le hachage du fichier représente donc l'identificateur unique de son contenu.

A la base, les protocoles Gnutella et FastTrack prévoient que pour être mis en partage, chaque fichier doit être «découpé» en morceaux élémentaires. Ces entités élémentaires ont une taille de 9,28 Mo. Seuls les morceaux complets et non corrompus sont partagés. Ainsi,

11 L'octet et ses multiples sont généralement utilisés comme mesure de la capacité de mémorisation de la mémoire informatique. http://fr.wikipedia.org/wiki/Octet

chaque fois qu'eMule termine le téléchargement d'un morceau, le logiciel s'assure de l'absence de corruption des données. S'il est valide, il est proposé en partage. Le morceau, tout comme n'importe quel fichier, est reconnu non par son nom (un morceau n'a pas de nom), mais par sa valeur de hachage. eMule applique l'algorithme MD4 sur les données constitutives de chaque morceau élémentaire complet. Le résultat de cette opération est une valeur de hachage unique. C'est la valeur obtenue en appliquant un algorithme MD4 aux valeurs de hachage de l'ensemble des morceaux qui le constituent.

Lorsque l'on ajoute un téléchargement à eMule, la première source qui envoie des données expédie pour commencer la table de hachage (ou hash set) du fichier qui est l'ensemble des valeurs de hachage de tous les morceaux qui constituent le téléchargement.

De plus, l'algorithme MD4 est aussi utilisé pour générer d'autres identifiants nécessaires à eMule. Ainsi, chaque utilisateur du réseau est reconnu par la valeur de hachage qui lui est attribuée une fois pour toute lors du premier lancement d'eMule et qui est sauvegardée dans un fichier local. Par ce moyen, les autres clients le reconnaissent infailliblement quel que soit le pseudo utilisé.

En outre, les services de police calculent, à partir des images ou fichiers illégaux en leur possession, des empreintes numériques pour chacun d'eux. Ces empreintes sont cataloguées dans des bases de données locales. Ainsi, il est possible de vérifier si un fichier téléchargé est bien de nature illégale en comparant sa signature, sans devoir le visionner, ou encore l'écarter des résultats si sa signature fait partie des «faux positifs». Si le fichier a été modifié d'une manière ou d'une autre (taille, format, insertion de logo, etc.), la valeur de l'empreinte va forcément être modifiée. Pour pallier ce problème, la société Advestigo SA, parmi d'autres, a développé son format d'empreinte «intelligente» selon le principe de la «reconnaissance visuelle». Cette technologie brevetée permet l'extraction d'empreintes numériques de contenus multimédia variés. Il ne s'agit pas de tatouage ou «wa-

termarking» de fichier. Les contenus élémentaires tels que les sons, images ou textes sont caractérisés au moyen d'outils spécifiques à chaque média. Les contenus composites, c'est-à-dire composés de plusieurs médias comme les documents audiovisuels, les documents bureautiques, les animations Flash, sont décomposés en médias élémentaires représentés au moyen de leurs empreintes spécifiques. Un document «Word» est ainsi représenté par une ou plusieurs empreintes qui vont permettre de le retrouver en entier ou des parties au sein d'autres documents malgré des modifications substantielles apportées au contenu ou à son environnement. La technologie d'Advestigo est orientée sur la détection d'au moins un facteur de similarité et non sur la constitution d'une hypothétique «distance de similarité»[12] entre documents. Ce sont les «caractéristiques communes» qui constituent le plagiat et non les différences introduites qui constituent l'originalité.

Cette technologie est applicable «a posteriori», c'est-à-dire même après que le contenu a été diffusé, puisqu'il fonctionne comme une empreinte digitale ou une empreinte ADN. Pour savoir si un document suspect est un plagiat d'une œuvre de référence, il suffit de calculer l'empreinte de l'œuvre de référence, l'empreinte du suspect et de les comparer. La technologie d'Advestigo permet également de créer des bases d'œuvres originales pour simplifier et accélérer la recherche de tels plagiats par rapport à de grandes bases d'œuvres originales.

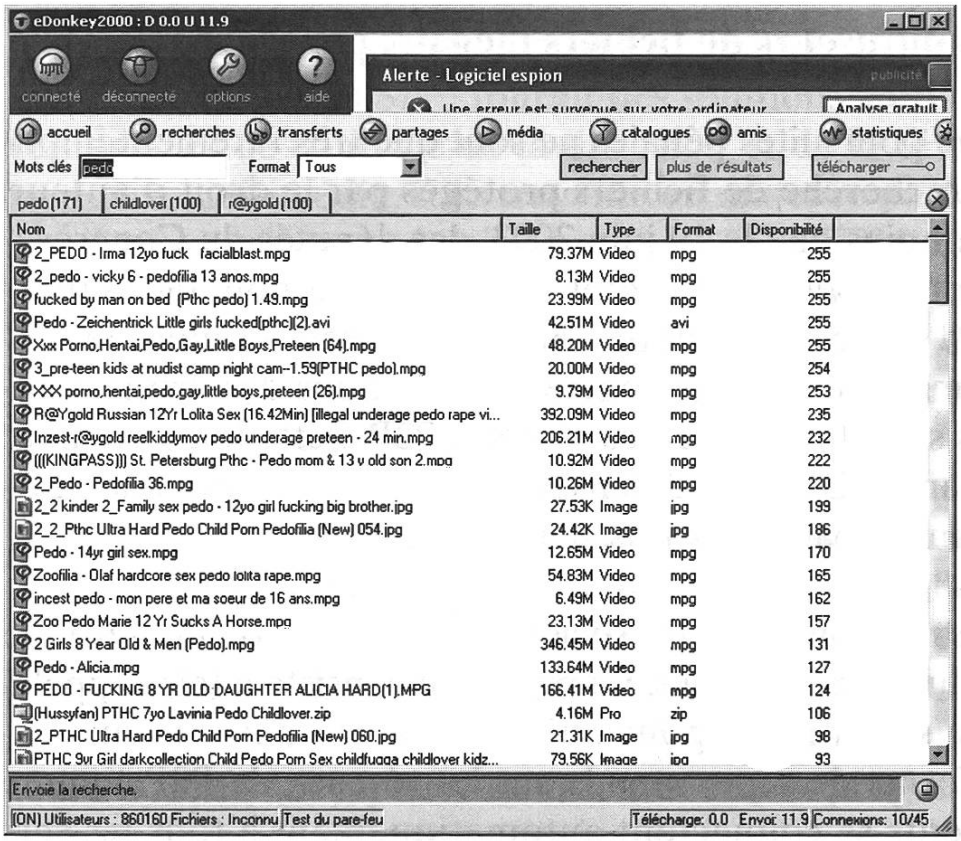
Recherche de contenus illégaux

Il est très aisé de rechercher des contenus illégaux grâce au peer-to-peer avec un succès immédiat et presque garanti. Il suffit d'employer des mots-clés pertinents selon la catégorie de fichiers désirés. Par ex-

12 Une méthode de réconciliation sémantique pour l'extraction de connaissances http://lbdwww.epfl.ch/e/publications_new/articles.pdf/ATT01110.pdf

emple, il suffit de taper «pedo» ou «childlover» dans eDonkey pour obtenir une liste d'une centaine de fichiers à télécharger, de la simple image au film vidéo de plusieurs minutes. Une recherche effectuée le matin ou l'après-midi sera généralement moins fructueuse que le soir, sachant qu'il y a vraisemblablement moins d'ordinateurs connectés à cette période de la journée. De ce fait, le résultat d'une même requête peut varier du simple au double. Il est possible de relancer la recherche plusieurs fois pour étendre la zone d'interrogation des serveurs et ainsi augmenter ses chances de résultats.

Figure 1: Aperçu du résultat d'une recherche avec le mot-clé «pedo».



En avril 2006, au moyen d'un logiciel expérimental, développé par un programmeur indépendant voulant garder l'anonymat, trois requêtes ont été faites sur 144 serveurs du réseau eDonkey avec les mots-clés suivants: childlover, r@ygold et pedo. Voici les résultats édifiants, obtenus en moins d'une minute:

Mot-clé	Diffuseurs	Fichiers
childlover	18661	2993
r@ygold	32602	8926
pedo	68214	14952

Dans un contexte plus préventif, il y a lieu de savoir que certains internautes inqualifiables renomment des films de pornographie dure avec des intitulés pour enfants. Dès lors, un chérubin qui souhaite obtenir un film de Walt Disney pourrait fortement se retrouver confronté malgré lui à de la zoophilie par exemple.

La traque des diffuseurs de fichiers illégaux par les majors

Les autorités de poursuites pénales ne sont que très rarement impliquées dans la recherche de fichiers protégés par le droit d'auteur. Relevons néanmoins que le 19 juin 2003, des députés du Congrès à Washington ont présenté un projet de loi sous le nom de «Piracy Deterrence and Education Act of 2003»[13] qui visait à rendre le FBI compétent en matière de lutte contre les violations des droits d'auteur, y compris sur l'Internet et les réseaux P2P. Cette loi, fortement controversée par les ONG de défenses des libertés individuelles, évidemment soutenues par les majors, n'a pas été retenue.

Pour lutter contre le piratage, les sociétés de distribution de musique ont recourt à plusieurs solutions. La première consiste à faire appel à des sociétés de statistiques pour qu'elles placent des serveurs espions, sous la forme de super-nœuds, sur les réseaux de P2P comme eMule. Les clients se connectent automatiquement à tous les serveurs placés dans leur liste, et donc également aux serveurs espions, pour envoyer leurs statistiques de téléchargement et la liste des fichiers mis en partage. Cela permet d'épingler des internautes en flagrant délit pour ensuite les assigner en justice.

13 http://thomas.loc.gov/cgi-bin/query/z?c108:H.R.2517:

Une autre possibilité, qui retourne le système contre les adeptes du peer-to-peer, est de faire appel à d'autres sociétés pour qu'elles fabriquent et polluent les réseaux avec de faux fichiers (appelés des *fakes*). Ces «faux» répondent aux mots-clés liés aux originaux mais ils ne contiennent rien ou sont illisibles. L'effet escompté est de décourager les téléchargeurs et les résigner à se tourner vers les plateformes légales qui proposent du contenu sûr et de qualité, moyennant évidemment une légère contribution financière. Cependant, il ne reste que peu ou plus de sociétés spécialisées dans ce genre d'actions. Des sociétés comme OverPeer Inc. ou MediaDefender Inc. ont simplement fait faillite.

La bataille n'est pas pour autant terminée. Par exemple, la société française Advestigo SA[14] fournit des solutions automatisées de surveillance et de contrôle de la diffusion et de la dissémination des contenus multimédia. Elle travaille en majeure partie pour les majors de la musique et du cinéma comme SACEM, SDRM et SCPP. Cette société propose deux solutions modulaires qui vont de l'évaluation de la menace de piratage sur des morceaux de musique à la mise en place de moyens pour lutter contre la piraterie sur les réseaux peer-to-peer. Il s'agit de:

1. AdvestiWATCH™ est une étude qui permet de déterminer dans quelle phase de piratage se trouve l'œuvre surveillée: phase de latence (contenu et sources sont en manque par rapport aux demandeurs); phase «explosive» de diffusion des copies pirates (sources vidéos convenables disponibles, très forte augmentation de la demande); phase de ralentissement progressif de la diffusion; phase de diffusion marginale.

2. AdvestiSEARCH_warning™ est un service automatique de protection de droit d'auteur permettant aux fournisseurs de contenus numériques de surveiller, d'évaluer et de combattre l'utilisation

14 http://www.advestigo.com

illicite de leurs capitaux numériques. Il comprend une surveillance complète en temps réel, une fonction d'alerte détaillée, un service en ligne de rapport précis et entièrement documenté. Ce service peut être complété par l'envoi massif de messages d'avertissement au fraudeur.

Les techniques utilisées par les services de police

Pour ce qui est des services de police, plusieurs approches ont été faites. La première est simple, l'enquêteur utilise indifféremment un ou plusieurs logiciels de peer-to-peer afin d'effectuer des requêtes sur les réseaux communautaires. Ces recherches s'effectuent sur la base de mots-clés préalablement définis et connus comme étant susceptibles d'être contenus dans le nom des images ou vidéo. Ces mots sont récoltés au travers des ordinateurs saisis lors de perquisitions ou simplement par une veille sur d'autres protocoles d'Internet comme les newsgroups (Usenet[15]). Comme un utilisateur ordinaire de peer-to-peer, l'enquêteur demande ensuite la mise en téléchargement de ces fichiers. Selon les logiciels utilisés de P2P, il est déjà possible de lire l'adresse IP de l'internaute émetteur et de ce fait, il est envisageable d'effectuer une identification de l'origine de cette adresse et de ne pas télécharger ce fichier si elle n'est pas détenue par un fournisseur d'accès Internet suisse. Cette technique a été choisie notamment par le Service national de coordination de la lutte contre la criminalité sur Internet[16] (SCOCI/KOBIK) qui a modifié le code source d'un logiciel de peer-to-peer pour en faire un logiciel de traque, couplé à des bases de données d'empreintes numériques. Ce service de la police fédérale helvétique effectue quotidiennement une veille sur

15 Usenet est un ensemble de protocoles servant à générer, stocker et récupérer des «articles» (des messages qui sont proches, dans leur structure, des courriels), et permet l'échange de ces articles entre les membres d'une communauté qui peut être répartie sur une zone potentiellement très étendue. http://fr.wikipedia.org/wiki/Newsgroup

16 Le SCOCI www.scoci.ch constitue le point de contact central en Suisse pour les personnes souhaitant signaler l'existence de contenus suspects sur Internet.

le P2P et dénonce les cas positifs aux polices cantonales et à Interpol.

Une autre solution a été trouvée en plaçant un autre ordinateur en amont de celui qui établit les requêtes et qui reçoit les fichiers demandés. LogP2P fonctionne grâce à une astucieuse méthode d'analyse, très confidentielle, couplée à un système de surveillance (monitoring). Il analyse les flux du protocole et décèle les éléments d'identification comme l'adresse IP encapsulée dans l'ID[17] par exemple. Il se positionne entre l'enquêteur et Internet et identifie, en temps réel et de manière catégorique, les diffuseurs de fichiers. Ce logiciel fait partie intégrante de la solution AntiPedoFiles de l'ONG Action Innocence. En outre, cet outil permet de catégoriser les résultats (images) et de générer des empreintes numériques avant leur transmission à un serveur central.

La police hollandaise a également développé son propre logiciel de traque. Geocholone fonctionne par l'analyse du contenu des fichiers temporaires (extension .dat) générés par le logiciel Kazaa lors de la demande de téléchargement du fichier. L'adresse IP est mise en évidence avec la date et l'heure du déclenchement de la session. L'illicéité du fichier est déterminée grâce à une base de données locale d'empreintes numériques.

En Norvège et en Italie, la police a mis sur pieds une solution globale de traque très complexe, nécessitant d'énormes ressources humaines et techniques. Peu d'informations sont disponibles pour le grand public à ce sujet. Ces systèmes ont générés d'énormes quantités de résultats qui ont été diffusés auprès de plusieurs polices en Europe et dans le monde.

17 Lors de la connexion à un serveur P2P, ce dernier attribue un numéro (ID) calculé à partir de l'adresse IP de l'ordinateur. C'est l'identification utilisée sur le réseau pour communiquer avec les autres utilisateurs et les serveurs. Les logiciels P2P ainsi que les serveurs sont capables de retrouver cette adresse IP à partir de l'ID et vice versa.

Les opérations de police au plan international

En mai 2004, l'opération «Enea» a été conduite par la police criminelle norvégienne (Kripos). Elle découle de l'utilisation d'un système automatisé de recherche des diffuseurs de fichiers à caractère pédophile sur le peer-to-peer. Des centaines d'internautes ont été interpellés en Norvège, en Suède, en Finlande et au Danemark.

«Callidus» est une vaste opération de police contre la pédopornographie sur le P2P, menée conjointement par huit pays de l'Union européenne du 2 au 6 mai 2005. Elle a permis l'interpellation d'une centaine de suspects. Les huit pays concernés étaient la Suède, la Grande-Bretagne, le Danemark, la France, les Pays-Bas, Malte, la Norvège et la Pologne. Cette opération est le résultat de la coopération d'un groupe européen de polices créé aux Pays-Bas en octobre 2004. Le projet a été baptisé «COSPOL» et vise à renforcer la coopération européenne dans les domaines de la criminalité liée à Internet et de la pornographie infantile.

Le 22 avril 2004, le Département de la Justice américain a annoncé les résultats de l'opération baptisée «Fastlink» qui s'est déroulée dans 27 états aux USA et dans once autres pays: la Belgique, le Danemark, la France, l'Allemagne, la Hongrie, Israël, les Pays-Bas, Singapour, la Suède, la Grande-Bretagne et l'Irlande du Nord. Elle visait les groupes Warez[18] les plus actifs comme Fairlight, Kalisto, Echelon, Class, APC ou encore Project X. 200 ordinateurs ont été saisis, dont 30 servaient exclusivement au stockage et à la distribution de contenus piratés. L'opération a été supervisée par la BSA (défense des éditeurs de logiciels), l'ESA (jeux-vidéo), la MPAA (films), et la RIAA (musique). Ce sont en tout près de 100 individus

18 Issu du suffixe «ware» de software, freeware ou shareware, le warez désigne l'ensemble des logiciels habituellement payants qui sont rendus disponibles en version complète par piratage. Le «z» vient quant à lui de la terminaison originale «wares» (prononcé «wairz» en anglais). De la même façon, on utilise les termes «gamez» pour les jeux vidéos ou «moviez» pour les films piratés. http://www.dicodunet.com/definitions/internet/warez.htm

qui ont été identifiés lors de l'opération, dont beaucoup seraient selon le Département de la Justice «des leaders de membres de haut niveau de différentes organisations internationales de piratage».

Ces quelques exemples démontrent bien la volonté des autorités d'enrayer ce phénomène de téléchargement sur le peer-to-peer et d'agir de manière coordonnée au plan international. Toutefois, les résultat ont de ces opérations, bien que très médiatiques, a un effet négligeable sur la masse de téléchargements effectuée au quotidien. Chaque mois, près de 900 millions de morceaux musicaux sont échangés illégalement sur le P2P. La fédération internationale de l'industrie phonographique (IFPI), représentant mondial de l'industrie du disque, évalue la perte de revenus potentiels due au piratage des œuvres à près de 4.6 milliards de dollars à l'échelle mondiale[19]. Les pertes de revenus potentiels étaient estimées à 2.1 milliards de dollars en 2004.

La recherche d'éléments d'identification

A la différence d'un internaute quelconque, l'enquêteur va analyser et stocker les données relatives à l'identification de chacun des émetteurs, à savoir son adresse Internet IP, la ou les parties du fichier téléchargé, la date et l'heure de début et de fin de téléchargement, et le fichier téléchargé qui est enregistré sur son disque dur pour être visualisé. L'enquêteur dispose ainsi des éléments de preuve (fichier image et «données de connexion» des émetteurs) nécessaires à la poursuite de l'infraction dans son ressort de compétences territoriales et légales.

Sur la base de ces informations collectées, l'enquêteur peut procéder à la rédaction d'une requête aux fins de poursuites pénales qu'il adressera au FAI concerné. Il obtient en retour l'identification nomi-

19 IFPI Commercial piracy report 2005 http://www.ifpi.org/site-content/library/piracy2005.pdf

native du titulaire de la connexion (nom de l'abonné) au moment du transfert.

Rigueur obligatoire

Par souci d'efficacité et de parfaire le dossier menant à une poursuite pénale, il y a lieu de prendre des mesures rigoureuses. Comme nous l'avons vu, il est possible d'effectuer un filtrage des adresses IP pour se concentrer uniquement sur les internautes utilisant un opérateur national. Des plages d'adresses IP peuvent être introduites dans le logiciel P2P utilisé par l'investigateur lors de la recherche ou, l'ensemble des résultats sont transmis au logiciel et le filtrage a lieu par la suite (rejet des adresses IP étrangères). De plus, il est de rigueur d'être sûr que l'internaute a bien téléchargé l'intégralité du fichier, démontrant ainsi sa capacité non seulement de le visionner mais aussi de le diffuser complètement. Pour être réellement professionnel, l'emploi de bases de données locales d'empreintes numériques, soit d'utilisateurs et/ou de fichiers, permet d'établir une liste de fichiers diffusés par un internaute et de valider la nature du fichier mis à disposition sur le peer-to-peer. Typiquement, la solution de traque utilisée par la Police fédérale, précisément par son service de veille sur Internet (SCOCI), est conforme à ces exigences.

Conclusion

Techniquement, le peer-to-peer a fait sauter une restriction technique importante qui résultait des systèmes centralisés. Autrefois, plus un contenu devenait populaire, moins le serveur devenait accessible pour des raisons de saturation. Avec le P2P, l'augmentation du succès d'un contenu se traduit par une augmentation de la facilité de disponibilité. Evidemment, cette faculté a inexorablement attiré les amateurs de fichier illégaux, détournant encore une fois à des fins criminelles une extraordinaire invention.

La lutte contre la diffusion de fichiers illicites s'organise et il est possible d'obtenir des résultats probants. Nous ne pouvons que nous en réjouir. Des solutions techniques sont disponibles à des fins judiciaires et civiles, répondant à un cahier des charges précis pour rester dans la légalité. L'identification des diffuseurs de contenus illégaux sur le P2P est réalisée par plusieurs forces de police dans différents pays d'Europe. Force est de constater que le nombre de résultats positifs est bien souvent trop important pour une action en justice de tous les cas, par défaut de ressources humaines essentiellement, pour conduire les auteurs devant les tribunaux.

Références

Poste à poste
http://fr.wikipedia.org/wiki/Peer-to-peer
Guide utilisateur eMule
http://www.emule-project.net/home/perl/help.cgi?l=13&rm=show_topic&topic_id=160
Histoire d'Internet et de l'informatique à partir de 1980
http://perso.club-internet.fr/pboursin/bonus4b.htm
Kazaa légalisé par la justice hollandaise, 01net., 29/03/2002
http://www.01net.com/article/180618.html
«Le peer to peer en baisse, les plaintes en hausse», Nouvelobs. com, 20/01/2004
http://archquo.nouvelobs.com/cgi/articles?ad=multimedia/20041007.OBS8521.html&host=http://permanent.nouvelobs.com/
Société britannique d'analyses Internet Cachelogic, 2004
http://www.cachelogic.com/research/p2p2004.php
Joelle Farchy, Internet et le droit d'auteur: La culture Napster
http://www.freescape.eu.org/biblio/article.php3?id_article=132
Clay Shirky, Kelly Truelove, Rael Dornfest & Lucas Gonze, *The Emergent P2P Platform of Presence, Identity, and Edge Resources,* O'reilly, 2001.
«Bile666», le Suisse qui fait trembler Hollywood, L'Hebdo, 23 mars 2006, page 76.