Le visiteur visité
Quand les éditeurs de logiciel Internet passent subrepticement à travers les mailles du filet juridique

Jean-Marc DINANT(*)

***Lex Electronica,* vol. 6, n°2, Hiver/Winter 2001**

Synopsis

This article is a plea for the integration of measures devised to protect privacy in the very heart of technology. This adaptation is an indispensable pre-requisite if one wishes to raise the consumer's confidence in the World Wide Web. With no reliable vehicle on the Internet, one which would be free from latent defects, it seems illusive for someone to hope to attract and keep electronic clients in the near future.

Résumé

Cet article est un plaidoyer pour l'intégration de mesures protectrices de la vie privée au cœur même de la technologie. Cette adaptation est un pré-requis indispensable à la confiance du consommateur dans le réseau des réseaux. Sans véhicule Internet fiable et libéré de ses vices cachés, il demeure illusoire d'espérer, à moyen terme, attirer et conserver des clients électroniques.

Table des matières

Introduction

1. Hiérarchie des protocoles de télécommunication

a. La couche physique

b. La couche TCP/IP

c. La couche services

2. Hiérarchie des périphériques de communication

3. Hiérarchie des acteurs de télécommunications

4. Les auteurs de software comme acteurs omniprésents d'Internet

5. Vers une nouvelle définition de la donnée à caractère personnel

6. Par l'absurde

7. Du marketing conforme à la protection des données

8. De la responsabilité des auteurs de logiciels en matière de vie privée

Introduction

1. En 1998, " Les traitements invisibles sur Internet "[1] a reçu un accueil enthousiaste dans le monde juridique et a été maintes fois cité dans de nombreuses publications juridiques. En 1999, le groupe 29 qui regroupe l'ensemble des commissions de protection de la vie privée en Europe a publié la *Recommandation 1/99 sur le traitement invisible... des données à caractère personnel sur l'Internet*[2]. Si, à l'aube du troisième millénaire, j'ai jugé utile de procéder à une refonte complète de la description essentiellement technique des traitements invisibles sur le grand réseau, c'est parce que je suis mu par une triple conviction :

- les juristes travaillant dans le domaine des nouvelles technologies de l'information et de la communication ont un besoin croissant d'information technique exacte, fiable et compréhensible ;

- le grand public possède d'Internet une vision encore naïve étroitement liée à ce qu'il voit sur l'écran. Sur Internet, " L'essentiel est invisible pour les yeux " et l'essentiel sort de la machine bien souvent à l'insu de son utilisateur ;

- les informaticiens eux-mêmes mesurent rarement l'ampleur des effets de certaines de leurs décisions techniques sur la protection des données à caractère personnel.

2. Cet article se veut donc une explication raisonnablement exhaustive des aspects techniques de la protection des données sur Internet ainsi qu'une typologie des risques existant actuellement. Au niveau technique, sa lecture sera probablement frustrante pour un ingénieur en télécommunication. J'ai volontairement omis de détailler certaines techniques et d'autres concepts ont été drastiquement simplifiés.

1. Hiérarchie des protocoles de télécommunication

3. " *Dividere ad regnandum* " pourrait sans conteste être un proverbe phare de l'industrie du *software* en général et du *software* Internet en particulier. Pour parvenir à maîtriser la complexité d'un processus de communication se déroulant à un niveau mondial et en temps réel, les informaticiens ont spécialisé les différents programmes de transmission de l'information. Il existe ainsi trois " couches " de programmes de télécommunication que je nommerai par la suite la couche physique, la couche TCP/IP et la couche service. Chaque couche se compose des programmes qui communiquent avec les programmes des couches de niveau inférieur ou supérieur par le biais d'interfaces. Concrètement, ces interfaces sont constituées d'un ensemble de programmes permettant d'effectuer certaines opérations simples et élémentaires.

4. À chaque niveau, chaque programme incarne un certain aspect d'un protocole, c'est-à-dire qu'il suit certaines règles conventionnelles et publiques. Le respect des protocoles est une condition *sine qua non* pour que deux machines puissent communiquer. Typiquement, les protocoles de télécommunication définiront les règles de préséance (quel appareil peut parler le premier ; peut-on parler et écouter en même temps ?), la résolution des conflits

(que faire si deux machines émettent de l'information simultanément sur le même fil ?) et la gestion des erreurs (que faire si une information est abîmée durant son transport ?). Ces protocoles sont définis dans des normes internationales et publiées. Néanmoins, un même protocole peut aboutir à des programmes différents parce que, d'une part, certains aspects peuvent être sujets à interprétation et, d'autre part, les protocoles ne règlent pas de manière obligatoire dans les moindres détails tous les aspects de la télécommunication[3].

5. Le réseau Internet est composé d'ordinateurs clients (typiquement un " *browser* ") et d'ordinateurs serveurs (classiquement un serveur Web). Les postes clients sont en général utilisés par des personnes privées (dans le cadre d'une activité qui ne l'est pas nécessairement) tandis que les ordinateurs serveurs demeurent en général l'apanage des entreprises et des administrations. *A priori*, les clients envoient des requêtes pour obtenir de l'information aux serveurs et ceux-ci donnent l'information, suite aux requêtes qui leur sont adressées.

a. La couche physique

6. La couche la plus basse est constituée de programmes qui permettent l'envoi et la réception de *bits* sur un support de transmission (fil téléphonique, fibre optique, câble Ethernet, etc...) à l'aide d'un appareil particulier. Ces programmes apparaissent bien souvent sous la forme d'un pilote (*driver* en anglais). Ce pilote est spécifique à une famille de périphériques d'un même constructeur. Par exemple, un *driver* 3C5X9 permettra de piloter certaines cartes réseau de la marque *3Com* ; un autre pilote offrira des programmes permettant l'envoi et la réception de *bits via* un *modem V90* de la marque *Zoom*.

7. Il est important de noter que ces *drivers* ne sont pas standards et ne sont donc pas interchangeables. Ils sont liés à une marque particulière et à un type précis de périphérique au sein de cette marque particulière. Par contre**,** ils offrent toujours une interface standard par rapport à la couche de niveau supérieur, celle de niveau intermédiaire. Cela signifie que la migration d'un *modem* vers une carte réseau suppose l'installation de nouveaux pilotes mais n'entraîne théoriquement aucune modification des programmes des couches supérieures.

b. La couche TCP/IP

8. La couche intermédiaire permet d'envoyer des paquets d'information d'un ordinateur quelconque vers un autre ordinateur. Elle utilise donc nécessairement la couche inférieure qui offre des programmes permettant d'envoyer des *bits* d'une machine à une autre en utilisant un support de transmission et des périphériques particuliers. Les programmes composant cette couche ignorent tout, tant de la couche inférieure de la machine sur laquelle ils fonctionnent que sur la couche inférieure de l'ordinateur destinataire des paquets envoyés. Corollairement, les programmes de cette couche intermédiaire ignorent quel est le service (terme que j'utiliserai pour désigner les programmes de la couche supérieure) qui fait appel à eux. Impossible donc pour un programme du niveau intermédiaire de savoir si le paquet qu'il envoie représente une partie d'une image fixe, d'un son, d'une page Web, d'un courriel ou d'un clip vidéo. La transparence se situe donc

au niveau des interfaces alors qu'un secret entoure la manière donc chaque programme fonctionne.

9. Quel que soit le type de périphérique physique réellement utilisé et de pilote présent dans la couche basse, quelque puisse être le service utilisé dans la couche haute (par exemple, service Web, serveur FTP, envoi de courriels, etc...), le seul et unique protocole de la couche intermédiaire sera toujours TCP/IP[4].

10. Afin de pouvoir envoyer des paquets d'information d'un ordinateur à un autre, éventuellement situé de l'autre côté de la planète, en utilisant une multitude de supports de transmission, il faut tout d'abord, à l'instar du courrier postal classique, convenir d'un système mondial d'adressage, c'est-à-dire d'un identifiant unique et public pour chaque ordinateur du réseau à un moment donné ainsi que des mécanismes permettant de s'assurer de l'unicité des adresses données.

11. Au niveau de la couche TCP/IP, chaque ordinateur est identifié sur le réseau Internet par un numéro unique noté conventionnellement A.B.C.D., où A,B,C et D représentent un nombre entier compris entre 0 et 255[5]. Théoriquement[6], il peut donc y avoir 256*256*256*256 soit environ quatre milliards d'ordinateurs connectés simultanément au réseau Internet. Pour éviter que deux ordinateurs distincts ne se retrouvent avec une adresse identique, les adresses IP sont distribuées par bloc (appelés classes) au niveau mondial par l'ICANN[7]. Ces adresses sont attribuées aux gros fournisseurs d'accès Internet ou à de grosses institutions qui peuvent à leur tour en attribuer certaines tranches à des entreprises ou organisations de taille plus réduite.

12. En règle générale, cette allocation peut se réaliser de deux manières différentes :

. Adresse IP dynamique : un client (personne physique ou personne morale représentée par une personne physique) utilise un réseau public de télécommunication (ligne téléphonique analogique ou numérique, téléphone mobile) pour se connecter chez un fournisseur d'accès Internet. Ce fournisseur va lui attribuer pour la durée de sa connexion à Internet un numéro IP unique choisi au hasard au sein de la classe d'adresses qu'il a obtenu directement de l'ICANN, ou par le biais d'un intermédiaire. Cette attribution se déroule normalement dans le cadre d'un contrat sur la base duquel l'utilisateur aura obtenu un identifiant d'utilisateur et un mot de passe. Généralement le fournisseur d'accès va inscrire[8] dans un journal de bord l'heure et les coordonnées du client à qui il a donné une adresse IP particulière.

Il est également possible et courant que les adresses IP soient distribuées au sein d'un serveur local par le protocole DHCP ou BOOTP. Dans ce cas, l'ordinateur, lors de son démarrage reçoit une adresse choisie automatiquement par un serveur d'adresse situé sur le réseau. Là aussi, ce serveur peut stocker dans un journal de bord le numéro IP attribué ainsi que le numéro *MAC* de la carte réseau. Ce numéro *MAC*, unique au monde est gravé dans le *silicium* de toute carte réseau de type Ethernet, protocole de bas niveau généralisé dans les réseaux locaux.

. Adresse IP statique. Dans le cadre d'un réseau local possédant une connexion permanente à Internet, il est courant que les ordinateurs soient dotés d'une adresse statique. Celle-ci est attribuée une fois pour toutes par le responsable informatique lors de la connexion des ordinateurs au réseau Internet.

13. Le mécanisme DNS (*Domain Name Service*) permet d'associer des noms de domaines aux adresses IP qui sont numériques. Concrètement, la traduction entre un nom de domaine particulier et son adresse (ou vice-versa) TCP/IP s'opère par le biais de programmes présents au niveau de la couche intermédiaire. Cette traduction s'opère sur le réseau par le biais de serveurs DNS, de manière hiérarchique. Le nom de domaine se compose :

- d'un nom de domaine supérieur (*top domain*) qui est soit géographique (par exemple " .ca " pour Canada, " .be " pour Belgique, etc.), soit générique (par exemple " .org " pour les organisations sans but lucratif, " .com " pour les sociétés commerciales). Ici encore, c'est l'ICANN qui est responsable de la création des noms de domaine du niveau supérieur ;

- précédé d'un ou de plusieurs sous domaines. Ici encore l'ICANN va déléguer la gestion d'un nom de domaine particulier à des organisations spécialisées. Par exemple, la gestion du domaine national ".be " est confiée à une ASBL dénommée " DNS BE ". Cette ASBL contrôle le serveur DNS ".be " qui contient les adresses des serveurs de domaine inférieur. Le sous domaine " gov.be " est géré par le gouvernement belge qui peut créer des sous domaines comme par exemple " just.fgov.be " qui désigne le ministère de la justice du gouvernement belge. Ce ministère possède lui aussi un serveur DNS et peut créer des noms de domaines comme par exemple " www.just.fgov.be " qui est un synonyme de l'adresse IP du serveur abritant le site du Ministère de la Justice en Belgique.

14. Il n'est pas, techniquement parlant, indispensable pour un client ou un serveur Internet d'avoir un nom de domaine. Corollairement, rien n'empêche une machine particulière de posséder plusieurs noms de domaine différents. Une machine Internet peut donc avoir zéro, un ou plusieurs noms de domaines mais, à un moment donné, un nom de domaine particulier renvoie toujours à une machine unique.

15. L'attribution d'un nom de domaine se fait *via* une procédure administrative, à l'aide d'un formulaire spécifique, éventuellement accessible en ligne. Les détenteurs d'un nom de domaine sont inscrits dans une banque de données accessible à tous via le réseau Internet.

16. Sur la quasi-totalité des systèmes d'exploitation, l'utilisateur dispose de programmes lui permettant de connaître l'adresse TCP/IP correspondant à un nom de domaine particulier et inversement. La conclusion de ce qui précède est qu'il est de manière générale possible et facile pour n'importe qui sur le réseau de trouver le propriétaire d'une adresse IP. Ce propriétaire sera en général capable, sans déployer des moyens disproportionnés, de faire le lien entre cette adresse IP à une certaine période et un utilisateur particulier, qu'il s'agisse du client d'un *ISP* ou du travailleur au sein d'une institution. L'identification peut

généralement s'opérer par un tiers, lui-même facilement identifiable à l'aide du réseau, sans déployer des efforts importants.

17. Ceci est à mettre en parallèle avec le considérant 26 de la Directive 95/46[9] qui pose que " *pour déterminer si une personne est identifiable, il convient de considérer l'ensemble des moyens susceptibles d'être raisonnablement mis en oeuvre, soit par le responsable du traitement, **soit par une autre personne**, pour identifier ladite personne* ". Le fait que cette autre personne n'aie pas la volonté de communiquer cette information (de son propre chef ou sur base d'un code de conduite ou d'un usage) importe peu, c'est l'existence concrète d'outils techniques effectivement disponibles auprès d'un responsable de traitement qui s'avère décisive.

18. Ce point reste toujours controversé et la loi anglaise[10] censée transposer la Directive 95/46 n'hésite pas à définir comme données à caractère personnel "*data which relate to a living individual who can be identified (a) from those data, or (b) from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller* ". Si tel est le cas, la liste des pages visitées sur un site, des mots-clés tapés sur un moteur de recherche ou des articles lus dans un journal en ligne, associées à l'adresse IP de la personne concernée peuvent librement être publiées et circuler dans le monde entier, même et surtout vers les pays n'offrant aucune protection légale des données personnelles : la loi ne s'applique pas. Il serait donc possible d'assister à l'émergence de profils complets basés sur une adresse IP particulière. Dans le cas d'une adresse IP statique, un site quelconque pourrait donc connaître au sujet d'un visiteur particulier l'historique de ses déplacements sur le Web et notamment les pages téléchargées sur Internet, les mots-clés tapés sur les moteurs de recherche et les articles lus sur les journaux en ligne. Mais que le lecteur paranoïaque se rassure. Il reste néanmoins parfaitement anonyme au regard du site visité puisqu'il ne lui donne que son adresse IP et ne lui communique pas son nom.

c. La couche services

19. La couche TCP/IP, alliée à la couche physique permet d'envoyer un paquet d'information d'un bout à l'autre de la planète, d'un ordinateur à un autre, chacun d'entre eux étant identifié par une adresse IP unique et éventuellement par un ou plusieurs noms de domaine. La couche " services " permet d'installer sur Internet les véritables programmes dont l'utilisateur a besoin. Ici encore, chaque programme particulier implémente un ou plusieurs protocoles spécialisés. Les protocoles les plus courants sont :

- HTTP (*HyperText Transport Protocol*) utilisé pour " surfer sur le Web ". C'est un protocole général qui permet de transférer de nombreux types d'informations : pages écrites en HTML, sons, images animées ou non, clip vidéo... ;

- FTP (*File Transfert Protocol*) permet de transférer des fichiers d'un ordinateur à un autre. Typiquement FTP pourra être utilisé pour télécharger des programmes ou pour mettre à jour un site Web ;

- SMTP (*Simple Mail Transport Protocol*) constitue le protocole utilisé pour envoyer des courriers électroniques ;

- POP3 (*Post Office Protocol 3*) permet de lire le contenu d'une boîte aux lettres électronique ;

- un autre protocole similaire est IMAP mais ne sera pas abordé ici ;

- NNTP (*News Network Transport Protocol*) est un protocole utilisé dans les groupes de discussions (*News*).

20. Un " *browser* " moderne est capable d'utiliser les cinq protocoles détaillés. Toutefois certains programmes spécialisés permettent l'utilisation de certains de ces protocoles seulement. On trouve ainsi des programmes de courrier électronique, de transfert de fichiers, etc...

2. Hiérarchie des périphériques de communication

21. *Prima facie*, l'Internet fonctionne donc grâce à une hiérarchie de protocoles standardisés. Cela signifie que chaque programme au sein de chaque couche est transparent au niveau de ses interfaces (les fonctionnalités qu'il offre) mais conserve le secret de sa cuisine interne (comment il effectue les opérations demandées). Parallèlement à cette hiérarchie, des protocoles on trouve une hiérarchie des périphériques de communication. Ainsi, un modem travaille au niveau de la couche physique et ne sait s'il transporte de la voix numérisée ou des paquets TCP/IP. Un routeur est un appareil extrêmement stratégique et permet de diriger les paquets sur l'un ou l'autre lien TCP/IP, sur base de l'adresse de destination. Au-delà de ses fonctions d'aiguillage, le routeur va par ailleurs servir de pare-feu (*firewall*) notamment en empêchant que des paquets avec une mauvaise adresse d'origine ou de destination ne puissent être transmis sur le réseau Internet. Les routeurs assurent les interconnections entre les différents fournisseurs d'accès.

3. Hiérarchie des acteurs de télécommunications

22. Dans une deuxième analyse, on peut mettre en concordance cette hiérarchie des protocoles et des périphériques de communication avec une hiérarchie des acteurs. Logiquement, chaque acteur se concentrera sur une couche particulière. Je distinguerai donc les opérateurs de télécommunications, les fournisseurs d'accès Internet[11], les fournisseurs de services Internet, les fournisseurs de contenu, les tierces parties fournissant des services sur Internet et l'utilisateur.

23. L'opérateur de télécommunication fournit un support de télécommunication point à point. Il s'agit en général de connections numérisées conçues pour transporter de l'information binaire. L'opérateur ne connaît pas et n'a pas à connaître la nature de l'information circulant sur les supports mis à disposition. Il doit néanmoins pouvoir s'assurer que les équipements terminaux ne sont pas susceptibles d'endommager ou

d'empêcher le bon fonctionnement de son réseau. Il offre des services de transports de bits de point à point (couche matérielle).

24. Le fournisseur d'accès Internet fournit une connectivité TCP/IP à ses clients en utilisant les services offerts par les opérateurs de télécommunication. Cela se concrétise par une ou plusieurs adresses TCP/IP attribuées pour une période plus ou moins longue. L'IAP reçoit les paquets IP de ces clients et les faire aboutir, autant que possible, à l'adresse IP de destination. Inversement, il relaie les paquets IP destinés à ses clients vers l'adresse mentionnée sur le paquet. Accessoirement, il peut offrir des services de transit par rapport à d'autres IAP en redirigeant certains paquets sur le réseau même si ces paquets ne proviennent ni ne sont destinés à certains de ses clients. Bien souvent, dans le cadre de cette connectivité TCP/IP un serveur de nom de domaine sera offert. Il permettra toujours de traduire les adresses numériques vers de noms de domaine et inversement. Dans certains cas, il permettra même de donner un nom de sous-domaine à certaines adresses IP particulières. Le fournisseur d'accès Internet offre donc un service de transport de paquet d'un point du réseau vers un autre point du réseau (couche TCP/IP).

25. Le fournisseur de service[12] offre des ordinateurs serveurs équipés de protocoles de haut niveau : HTTP pour un serveur Web, SMTP et POP3 pour un serveur de courrier électronique, NNTP pour un serveur de *news* ou encore FTP pour un serveur de fichiers (couche service).

Hiérarchie des protocoles, du hardware et des acteurs

Couche Services	**Protocole**	HTTP Surfer	SMTP envoyer un courriel	POP3 recevoir un courriel	NNTP accéder aux news	FTP télécharger des fichiers	etc ... d'autres protocoles présents ou à venir
	Hardware	Client/Serveur HTTP	Client/serveur de courrier électronique		Client/serveur de news	Client/serveur TFP	
	Acteurs	Utilisateur/fournisseur de service					
Couche TCP/IP	**Protocole**	TCP/IP					
	Hardware	Routeur					
	Acteur	Fournisseur d'accès Internet					
Couche matérielle	**Protocole**	PPP utilisé sur lignes téléphoniques analogiques	X-75 utilisé sur lignes numériques	ADSL Utilisé sur lignes téléphoniques classiques	EHERNET utilisé sur les réseaux locaux	etc… D'autres protocoles présents ou à venir	
	Hardware	Modem	Adaptateur de terminal	Modem ADSL	Carte réseau Ethernet		
	Acteur	Opérateur de télécommunication					

26. Le fournisseur de contenu est celui qui utilise un service pour transmettre ou mettre à disposition du contenu sur Internet. Il est responsable de la licité du contenu offert et

pourrait être condamné pour, par exemple, avoir envoyé de fausses informations ou des informations calomnieuses par courrier électronique, publié sur un site du matériel pédophile, permis la copie illégale de software, effectué la vente de médicaments ou de drogue en ligne, etc... De nombreuses affaires récentes ont tenté, parfois avec succès, de mettre sur la sellette des fournisseurs de services pour avoir hébergé des contenus illicites. Il semble actuellement acquis que les fournisseurs de service n'ont pas l'obligation générale de surveiller les informations qu'ils hébergent ou transportent, sauf dans la mesure où ils connaissent le caractère illicite de cette information et dans la limite où ils n'ont pas fait diligence pour la faire disparaître[13].

27. Le tiers fournisseur de contenu est un acteur qui peut paraître surprenant et n'apparaît que rarement dans les analyses juridiques d'Internet. Il s'agit typiquement du cybermarketeur qui peut injecter, en temps réel, une bannière sur une page Web. Il s'agit d'un tiers car il n'est normalement pas partie prenante à la communication entre un utilisateur et un site déterminé, mais il peut en capter certains éléments clés comme, par exemple, les mots-clés tapés sur un moteur de recherche ou l'article précis lu dans un journal en ligne. Les firmes fournissant des compteurs de visites sont aussi des tiers fournisseurs de services et reçoivent les mêmes renseignements, ce qui permet d'ailleurs leur financement.

28. Le dernier acteur est l'utilisateur. Il ne voit bien souvent que la partie supérieure de l'iceberg, c'est-à-dire la fenêtre de son *browser*. Tous les acteurs énumérés *supra* tombent sous le coup d'une ou plusieurs Directives européennes[14].

29. La hiérarchie développée plus haut ne tient que si tous les acteurs possèdent une maîtrise totale de l'information à tous les niveaux. Or, en pratique, cette manipulation de l'information s'opère toujours par le biais de programmes que les utilisateurs n'ont que rarement écrit. Certes, leurs spécifications sont publiques mais chaque programme possède ses secrets.

4. Les auteurs de software comme acteurs omniprésents d'Internet

30. Techniquement parlant, tant que les codes sources n'ont pas été publiés[15], chaque programme possède l'autonomie suffisante pour mener à bien, parfois subrepticement, nombre de calculs ou de transmissions qui demeurent inconnues, en bref des*traitements invisibles*. Toujours techniquement parlant, rien n'empêche généralement un programme de lecture de CD musicaux censé ne pas accéder au réseau d'envoyer sur celuici le détail des morceaux écoutés par l'utilisateur.

31. Je m'attacherais donc dans ce qui suit, non pas à décrire les risques posés par le jeu des acteurs décrits *supra* non pas que ces risques soient inexistants mais parce qu'ils sont en général bien perçus et pris en compte par plusieurs Directives européennes mais à analyser les traitements invisibles effectués par le *browser* HTTP hébergé sur le disque dur de l'internaute. Je m'intéresse à ce qui arrive par défaut à un utilisateur moyen. Il est clair que l'ingénieur en sécurité informatique possède une connaissance et des outils permettant de s'exonérer de la plupart de ces traitements invisibles.

32. Ces traitements invisibles ne sont pas de la science fiction et la deuxième partie de cet article se contente de les décrire pour le seul protocole HTTP[16], au niveau du poste client. Dans ce cadre volontairement limité mais représentatif des autres protocoles, j'aborderai les traitements invisibles suivants : les *ET software*, les hyperliens invisibles, le bavardage des *browser*, la redirection HTTP, les *cookies*, le *JavaScript* et les *applets Java*. Chacun de ces traitements invisibles n'est que relativement (peu) privacide. C'est lorsqu'ils sont combinés qu'ils deviennent des outils redoutables utilisés quotidiennement des centaines de millions de fois par les entreprises de *cybermarketing*.

> Les ET software

33. Les *ET software* (par exemple *RealJukeBox Player*[17] ou le *wizard* d'enregistrement en ligne de *Microsoft*[18]) sont pas des traitements invisibles redoutables, dans la mesure où ils ont été réparés et où ils n'étaient pas spécifiés publiquement par leurs auteurs. Ces *ET software* transféraient des informations profitables vers leurs auteurs respectifs, ce qui rend la Directive européenne 95/46 applicable quant à son principe, mais difficilement applicable dans les faits, les auteurs étant outre atlantique. Par ailleurs on voit mal le particulier lésé intenter une action individuelle contre *Microsoft* ou *RealNetworks* sur Internet, les dommages sont bien souvent gigantesques pas leur ampleur (le nombre d'individus concernés), mais relatifs pour un individu particulier. Il n'y a donc que le parquet qui puisse se saisir d'une telle affaire.

> Les hyperliens implicites et invisibles

34. La notion d'hyperlien n'a plus à être détaillée. Elle constitue un élément essentiel de la navigation sur Internet. Toutefois, aux hyperliens explicites et visibles qui nécessitent une action objective de l'internaute pour voyager dans le cyberespace, s'opposent les hyperliens implicites et invisibles, s'exécutant sans intervention de l'utilisateur et à son insu. Ces hyperliens invisibles permettent d'inclure dans une page Web des images issues d'autres sites. Cette possibilité technique est largement utilisée pour la diffusion de bannières publicitaires stockées sur un seul site du Réseau mais rendues visibles sur des milliers d'autres pages d'autres serveurs disséminés à travers le monde entier. Concrètement, il suffit d'inclure dans une page HTML la balise <IMG SRC= http://www.ailleurs.com/...> pour que le browser téléchargeant la page en question ouvre automatiquement une nouvelle session HTTP vers un site tiers (ici www.ailleurs.com) à l'insu de l'internaute qui a l'illusion de croire que tout le contenu apparaissant à l'écran provient du site qu'il est entrain de visiter.

> Le " bavardage "

35. Le " bavardage " des programmes de navigation constitue une deuxième caractéristique apparaissant comme d'avantage privacide. Lorsqu'un programme de navigation demande une page à un site Internet, il communique de manière cachée et systématique certaines informations relatives à la machine du demandeur et notamment[19] :

- le type de système d'exploitation de la machine (*Windows* 95 ou 98, *Macintosh* ou *Unix*) ;

- le type et la langue du programme de navigation (par exemple *Netscape Navigator 4.01 FR*) ;

- le type de processeur (*Intel*, *Mac* ou *PowerMac*…) ;

- la langue parlée par l'internaute (par exemple le Français de Belgique... ou du Canada) ;

- la page référente. Dans le cas d'un hyperlien classique, il s'agit donc de l'URL de la page visitée avant la page précédente (en fait la page où se trouvait l'hyperlien menant à la page actuelle). Dans le cas d'un hyperlien invisible, il s'agit de la référence de la page où sera affichée l'image issue d'un site tiers. Dans le cas d'une page fournissant les résultats d'une recherche, le nom de la page référente contient les critères de recherche (les mots-clés). Dans le cas d'un article de journal mis en ligne, la page référente contient la référence précise de l'article lu. En d'autres termes, un site de *cybermarketing* qui transmet une bannière destinée à être affichée dans les résultats d'un moteur de recherche ou sur la page d'un journal en ligne connaît, **avant de transmettre la bannière**, ce que l'utilisateur recherche ou la référence précise de l'article lu.

Cette deuxième caractéristique, associée à la première, permet à des sociétés invisibles du Réseau et inconnues du grand public de capter en temps réel le comportement de dizaines de millions d'internautes[20]. Toutefois, cette possibilité reste éphémère car l'internaute ne communique aucun élément qui l'identifie de manière stable. Il communique certes son adresse TCP/IP mais cette adresse est attribuée dynamiquement par le fournisseur d'accès lors d'une connexion par modem. Il s'ensuit que cette adresse n'est pas un identifiant stable de l'utilisateur résidentiel.

> Les cookies

36. Les *cookies* sont des informations qu'un ordinateur serveur (accédé par un hyperlien classique *ou invisible*) peut stocker, lire ou effacer de manière permanente[21] sur le disque dur de l'internaute, généralement[22] à l'insu de celui-ci. Il s'agit en fait d'un code barre souvent inintelligible que n'importe quel site peut coller, supprimer ou modifier sur le dos de ses visiteurs, éventuellement par hyperlien invisible interposé.

37. L'avantage principal du *cookie* est qu'il est stable dans le temps et lié à une machine particulière. Il permet donc de s'affranchir du caractère dynamique de l'adresse TCP/IP de l'utilisateur résidentiel et de profiler une machine unique pour plusieurs dizaines d'années[23]. En d'autres termes, si un ordinateur aujourd'hui doté d'une adresse TCP/IP A.B.C.D. a reçu un *cookie* d'un site particulier (typiquement d'une entreprise de *cybermarketing* par hyperlien invisible), le même ordinateur transmettra systématiquement ce code barre à son insu à ce site (typiquement lors du téléchargement d'une bannière par hyperlien invisible) même s'il possède demain une adresse TCP/IP E.F.G.D ou R.T.U.V. Les sociétés de *cybermarketing* utilisent classiquement des*cookies* rémanents dans des hyperliens invisibles.

> La redirection HTTP

38. Lorsqu'un site Web reçoit une requête HTTP portant sur une certaine page Web, il peut répondre en envoyant un code de redirection[24] vers une autre page Web. À ce moment, le *browser* va télécharger la nouvelle page sans en informer l'utilisateur. Cette technique est souvent utilisée pour masquer des pages Web malicieuses. Ici encore, comme dans le cas des hyperliens invisibles, la redirection peut s'effectuer vers une autre page du site visité mais aussi vers une page située sur un site Web quelconque. Il semble que ce mécanisme ait été utilisé par certains moteurs de recherche pour rapporter en temps réel les résultats intéressant les visiteurs[25].

> JavaScript

39. Lorsqu'un navigateur reçoit une page d'un site, celle-ci peut contenir certaines instructions en *JavaScript*. Ces instructions sont interprétées en temps réel par un interpréteur *Java* intégré dans ou appelé par le navigateur. Typiquement, l'exécution de ces microprogrammes écrits en langage *JavaScript* a pour effet de dynamiser la page affichée en effectuant certaines animations ou en liant l'exécution de certains microprogrammes à certains boutons particuliers présents dans la page.

40. La possibilité offerte par défaut par les *browsers* classiques de permettre à des sites distants d'exécuter des programmes conçus par eux à l'insu de l'internaute reste dangereux, même si certains gardes-fous sont censés empêcher ces scripts d'accéder aux données personnelles de l'utilisateur. À titre d'exemple, le trou de Cuartango permet[26] à un serveur malicieux de lire le contenu d'un fichier quelconque situé sur le disque[27] du visiteur, à condition d'en connaître le nom. Dans le cas précis du trou de Cuartango, la rectification logicielle apportée par *Microsoft* n'était pas satisfaisante. Une variante permettait facilement de passer outre et il aura donc fallu deux " *patches*[28] " à *Microsoft* pour résoudre, au niveau théorique[29], ce trou de sécurité.

41. Par ailleurs, le langage *JavaScript* permet, dans de nombreux cas, de s'affranchir du filtrage opéré par les *proxy* serveurs au niveau des cookies et du masquage de l'adresse IP.

> Les applets Java

42. Les *applets Java* constituent une évolution des scripts *Java* à ne pas confondre avec ces derniers. Dans le cas des *applets*, les instructions *Java* sont préalablement traduites dans un pseudocode par les soins d'un précompilateur installé sur le serveur. Lors de l'exécution d'un *applet*, le navigateur charge le PCODE lié à l'applet et l'exécute par le biais d'un interpréteur. La différence fondamentale entre les scripts *JavaScript* et les *applets* est que les premiers sont transmis en langage clair au navigateur tandis que les seconds sont préalablement convertis en PCODE et donc inintelligible, même pour un programmeur moyen. Dans les deux cas de figure, la transmission s'effectue de manière souterraine mais, ici encore, la possibilité existe, dans les navigateurs récents, d'inhiber l'exécution des scripts ou des *applets*.

43. Ici encore arriva ce qui devait arriver. Le *Brown Orifice Vulnerability*, trou présent dans *Netscape Communicator*[30], permet de transformer ce type de *browser* en un serveur

FTP ou HTTP, permettant à quiconque sur Internet d'accéder à la totalité des informations présentes sur un disque[31] de l'utilisateur. Dans ce cas, c'est la liste des fichiers contenus sur le support qui est rendue accessible et non pas seulement le contenu des fichiers dont l'attaquant connaît le nom. *Netscape* a produit une nouvelle version de *Communicator* qui est censée boucher ce trou[32].

44. Les deux trous de sécurité explicités *supra* n'ont valeur que d'exemple. Sur Internet, ce genre de bogue est fréquent, largement documenté et accessible. Dans le meilleur des cas, ces trous de sécurité font l'objet d'un *patch* efficace. Dans le pire, ces trous sont méconnus ou circulent sous le manteau dans les cercles privés des *hackers*, reçoivent des modifications logicielles elles-mêmes bogués ou restent inconnus. Il est clair que l'internaute de la rue est généralement inconscient du danger que son*browser* troué fait courir à ses données personnelles. On peut aussi déplorer que les auteurs des *browsers* gratuits n'aient pas pour habitude d'utiliser l'adresse électronique de leurs clients afin de les avertir des dangers liés à un défaut de leur produit et de la manière d'y remédier. Dans ce contexte, la mention systématique, lors de la requête HTTP, du type et de la version du *browser* utilisé, constitue une information précieuse pour le site Web malicieux qui pourra alors choisir un trou de sécurité adapté au*browse*r de l'internaute pour accéder aux contenu des disques du visiteur.

5. Vers une nouvelle définition de la donnée à caractère personnel

45. Les sociétés de *cybermarketing* se défendent de collecter des données à caractère personnel. Cet argument n'est guère sérieux. La Directive, comme la loi belge qui la transpose, prévoit explicitement que l'identification puisse se réaliser à l'aide d'un numéro. Il est clair que si un numéro TCP/IP ou un *cookie* identifiant permet à quelqu'un de déduire des informations relatives à celui qui le porte, nous nous trouvons en présence d'une données relative à un individu particulier. Il est regrettable que la loi utilise la notion de " *personne identifiable* ", tant l'identification, dans l'esprit des juristes, nécessite la connaissance des données d'état civil d'une personne en particulier.

46. Sur Internet, la connaissance des noms, prénoms et date de naissance d'un individu n'est que rarement nécessaire pour pouvoir l'identifier comme étant " celui qui ". Internet permet de s'affranchir de la non-connaissance de l'état civil d'une personne particulière dans le cadre d'un profilage particulièrement pointu, impossible à réaliser dans le monde réel.

6. Par l'absurde

47. Il est clair que le législateur a voulu, par le biais de la Directive 95/46, éviter des discriminations abusives basées sur des données excessives, sans le moindre recours de la personne concernée. Le simple fait, par exemple, de stocker dans un *cookie* la « race » de la personne qui " surfe ", constitue pour nous une donnée à caractère personnel dans la mesure où cette donnée apparaîtra automatiquement lors de chaque connexion HTTP. Dans le cas contraire, il faudra admettre qu'Internet puisse effectuer de nombreuses et subtiles discriminations basées sur l'âge, le sexe, l'appartenance ethnique, le niveau d'instruction,

les revenus, les goûts politiques, l'état de santé, etc..., sans que jamais la personne victime de ces discriminations ne possède le droit d'en comprendre la raison, ne puisse accéder à son profil (et éventuellement le rectifier) ou s'opposer à l'utilisation de ces informations. À l'aube du troisième millénaire, un *cookie* peut avoir la même signification que l'infamante étoile jaune que les nazis collèrent sur le dos des juifs, non pas pour les identifier comme des personnes mais pour les discriminer comme des juifs. Doit être considérée comme donnée à caractère personnel toute donnée qui se rapporte à une personne particulière car, sur Internet, il y a quasiment toujours mille et une manière de pouvoir identifier une personne particulière et l'identification classique n'y est pas le préalable indispensable à la discrimination.

7. Du marketing conforme à la protection des données

48. L'application de la loi sur la protection des données à des données de profilage relative à une personne particulière mais dont l'état civil demeure inconnu ne signifie pas que tout marketing direct soit interdit. Ce marketing est autorisé moyennant :

· L'information lors de la collecte : la personne concernée doit savoir quelle société de *cybermarketing* collecte quelles données (et notamment les mots-clés qu'elle tape sur les moteurs de recherche et la référence des articles qu'elle lit en ligne) pour quelle finalité. Serait-il impensable que la première bannière de publicité " *one to one* " mentionne ce fait ?

· Dans le cas d'une finalité de marketing direct, elle possède un droit d'opposition au traitement de ces données.

· Dans tous les cas, la personne concernée peut prendre connaissance des données qui la concernent et, le cas échéant, en demander la rectification.

· Est-il utile de préciser que nous sommes bien loin du compte actuellement dans la mesure où même les sites plaçant des hyperliens vers des firmes de *cybermarketing* sur Internet ignorent généralement le fait que, par là-même, vu l'état actuel de la technique, ils communiquent foule de détails sur leurs visiteurs ?

8. De la responsabilité des auteurs de logiciels en matière de vie privée

49. Pour importer une automobile sur le réseau routier européen, il y a lieu de répondre à certaines exigences essentielles et de passer un test de conformité. Sur les autoroutes de l'information, les constructeurs de véhicules de l'information ne sont soumis à aucune réglementation *a priori* ou *a posteriori*. Il devient de plus en plus inquiétant de constater que l'industrie informatique responsable dans les faits d'un nombre impressionnant et sans cesse croissant de violations des grands principes de protection des données, passent entre les mailles du filet juridique. Cela est probablement dû à une vision aujourd'hui désuète de l'outil informatique. L'outil s'est progressivement transformé en une gigantesque usine à gaz dont le fonctionnement ponctuel ou global échappe à l'utilisateur. L'obstacle principal reste peut-être la faiblesse des connaissances techniques de l'internaute moyen, des législateurs… et des *Webmestres* eux-mêmes. Ainsi, il y a peu, un site belge de vente en

ligne publiait très probablement en toute bonne foi dans ses " questions souvent posées " qu'il n'utilisait pas de *cookies* rémanents, ce qui était faux. En fait, il envoyait un *cookie* expirant en l'an 2009. La technologie serait-elle d'égale opacité pour les entreprises qui l'utilisent en payant, pour les juristes qui tentent de la réguler et pour les internautes qui la subissent gratuitement ? À quand une loi *anti smog* technologique ?

Notes

(*) Informaticien, Chargé de recherche au Centre de Recherche Informatique et Droit, Facultés Universitaire Notre Dame de la Paix, Namur (Belgique). Email : jmdinant@fundp.ac.be

[1] Jean-Marc Dinant, "Les traitements invisibles sur Internet", Namur, Bruylant, *Cahiers du CRID*, n°16, 1999, p. 277-.302. Disponible en ligne : <http://www.droit.fundp.ac.be/cv/jmdinant.html>.

[2] Recommandation 1/99 sur le traitement invisible et automatique des données à caractère personnel sur l'Internet effectué par des moyens logiciels et matériels. Disponible en ligne : <http://europa.eu.int/comm/internal_market/en/media/dataprot/wpdocs/wp17fr.pdf>.

[3] Nous verrons plus loin à quel point cela pose des problèmes importants de vie privée dans le cadre du protocole HTTP.

[4] *Transport Control Protocol/Internet Protocol*. Il s'agit en fait d'une paire de protocoles. Le protocole IP permet l'envoi de paquets d'une adresse IP à une autre mais ne garantit pas qu'ils arriveront à destination sans erreur ni dans l'ordre d'envoi. Pour aboutir à un envoi sans erreur et dans le respect de la séquence d'émission, IP est souvent associé au protocole TCP.

[5] Les connaisseurs auront compris qu'une adresse IP constitue un ensemble de quatre octets.

[6] Certaines adresses ou catégories d'adresses sont réservées à des fins techniques et ne peuvent donc être attribuées en tant que telles à un ordinateur particulier.

[7] *The Internet Corporation for Assigned Names and Numbers* (ICANN) est une association sans but lucratif qui a été constituée pour assumer la responsabilité de l'espace IP, <http://www.icann.org>. En Europe, l'allocation de cet espace se réalise par le biais du RIPE (*Réseaux IP Européens*), <http://www.ripe.net>.

[8] Cela se fait bien sûr automatiquement par le programme qui attribue les adresses IP.

[9] Directive 95/46 du parlement européen et du conseil du 24 octobre 1995 relative à la protection des individus à l'égard du traitement de données personnelles et à la libre circulation de ces données, JOCE, 23 nov. 1995 No L. 281 p. 31.

[10] *Data Protection Act* de 1998, <http://www.legislation.hmso.gov.uk/acts/acts1998/80029--a.htm>.

[11] C'est délibérément que j'emploie ce terme. Dans le vocable commun, le fournisseur de service Internet (*Internet Service Provider* est tout d'abord un fournisseur d'accès et offre une connectivité TCP/IP. En pratique, il offre également d'autres services : classiquement il offrira du contenu sur lui-même et il hébergera les pages de ses clients. Le mot acteur doit donc être compris comme "jouant un rôle", un acteur particulier pouvant jouer plusieurs rôles. C'est le rôle joué qui conditionne l'applicabilité des différentes lois.

[12] Service signifie dans ce cadre un protocole de niveau supérieur.

[13] Directive 2000/31/CE du Parlement européen et du Conseil du 8 juin 2000 relative à certains aspects juridiques des services de la société de l'information, et notamment du commerce électronique, dans le marché intérieur ("Directive sur le commerce électronique"), Journal officiel n° L 178 du 17/07/2000 p. 0001 0016. Art 15 : "*Absence d'obligation générale en matière de surveillance. Les États membres ne doivent pas imposer aux prestataires ...une obligation générale de surveiller les informations qu'ils transmettent ou stockent, ou une obligation générale de rechercher activement des faits ou des circonstances révélant des activités illicites.*"

[14] Citons en vrac : 95/46/EC (protection de la vie privée en général) ; 97/66/EC (protection de la vie privée par les opérateurs de télécommunication), Directive 1999/93/EC (signature électronique), 97/7/EC (protection du consommateur et vente à distance), 2000/31/EC (commerce électronique).

[15] La publication des codes sources n'est pas la panacée universelle. Elle permet toutefois aux experts d'analyser les opérations réellement effectuées ou effectuables par un programme donné. Dans le monde de la cryptographie, cette publication est une condition *sine qua non* pour apprécier la fiabilité d'une solution de cryptage.

[16] Dans l'article "Les traitements invisibles" cité *supra*, note 1, certains risques liés au protocole TCP/IP sont détaillés. Ils restent d'actualité, principalement en ce qui concerne *l'opacité du routage, c'est-à-dire l'identité des acteurs susceptibles de prendre connaissance des paquets TCP/IP*.

[17] L'image, extraite du film de *Steven Spielberg* est parlante. L'extraterrestre ET "téléphone" en cachette de temps en temps à la maison, pour raconter ce qui s'est passé sur la terre. Un *ET software* est donc un programme qui communique *via* Internet des détails sur le comportement de son utilisateur. Un exemple célèbre est le cas de *RealJukeBox Player*, *software* d'écoute de CD musicaux diffusé à plus de treize millions d'exemplaires qui rapportait régulièrement à la société mère (*RealNetworks*) le détails de CD insérées

dans le lecteur, de manière encryptée, <http://www.tiac.net/users/smiths/privacy/realjb.htm>. À la suite d'un article paru dans le *New York Times*, *RealNetworks* a modifié son logiciel.

[18] Plus de détails sur <http://www.truste.org/users/ms_process.ppt>.

[19] Le lecteur curieux pourra tester le bavardage de son propre *browser* sur <http://www.droit.fundp.ac.be/crid/privacy/WhatIknow.htm> => *show the full http header my browser is sending.*

[20] Serge Gauthronet, "Les services en ligne et la protection des données, annexe au rapport annuel 1998 du groupe de travail établi par la Directive 95/46", Commission européenne, 1998 : <http://europa.eu.int/comm/dg15/en/media/dataprot/studies/servint.htm>, p. 92 cite le chiffre de 17.000.000 pour une seule entreprise américaine. Deux ans plus tard, ce même chiffre si situe à hauteur de 500.000.000 de bannières quotidiennes <http://www.doubleclick.net/company_info/investor_relations/financials/analyst_metrics.htm>.

[21] Les *cookies* de session ne possèdent aucune date limite et ne sont pas stockés sur le disque dur de l'utilisateur.

[22] Les utilisateurs avertis peuvent paramétrer leur *browser* de manière à inhiber les *cookies* ou à accepter au cas par cas mais ces mesures restent largement inappropriées parce que

1.- l'utilisateur moyen ignore ce qu'est un *cookie* et quels sont ses risques en matière de vie privée ;

2.- certains sites refusent les visiteurs qui refusent les *cookies* ;

3.- certains sites envoient plusieurs *cookies* et provoquent harcèlement et fatigue chez l'internaute ;

4.- les mécanismes d'opposition empêchent la réception de nouveaux *cookies* mais pas toujours l'envoi des cookies précédemment reçus et enregistrées ;

5.- la distinction entre *cookies* rémanents ou non, issue de sites tiers ou non n'est pas faite.

[23] Il n'est pas rare que les entreprises de *cybermarketing* envoient des *cookies* conçus pour durer jusqu'en l'an 2035… ou plus.

[24] Le lecteur pourra tester cette redirection en visitant <http://www.droit.fundp.ac.be/cv/jmdinant.htm> qui le renverra automatiquement vers <http://www.droit.fundp.ac.be/cv/jmdinant.html>.

[25] Le moteur de recherche donne comme hyperlien l'adresse de la firme de *cybermarketing* avec un code détaillant l'URL précise de la page recherchée. Recevant

ce message, la société de *cybermarketing* redirige immédiatement le visiteur vers le lien voulu... en notant au passage son *cookie,* les mots-clés recherchés ou l'article lu en ligne et son intérêt pour un hyperlien particulier.

[26] Plus de renseignements sur <http://pages.whowhere.com/computers/cuartangojc/cuartangoh1.html>. L'auteur a personnellement testé ce trou de sécurité et affirme qu'il est réel.

[27] Qu'il s'agisse du disque dur local, de la disquette, du CD-ROM ou d'une partie d'un réseau local montée sur un disque virtuel.

[28] Portion de code que l'on peut réinjecter dans un programme précis. Cela permet de modifier une erreur contenue dans un logiciel sans avoir à réinstaller un système complet.

[29] Pour que ce trou soit bouché au niveau pratique, il a fallu que les millions d'utilisateurs de *Microsoft* Internet explorer 4.01 téléchargent et installent avec succès le deuxième patch délivré par *MicroSoft*.

[30] Rapporté par le Cert (*Computer Emergency Response Team*). Voir <http://www.cert.org/advisories/CA-2000-15.html>. Concernerait les versions 4.4 jusque 4.74. L'auteur n'a pas eu l'occasion de tester personnellement ce trou de sécurité.

[31] Voir note 28.

[32] Voir <http://www.netscape.com/security>.

ISSN 1480-1787