Droits et obligations à l'ère numérique : protection de la vie privée *

Daniel Le Métayer † et Guillaume Piolle ‡

30 septembre 2010

Mots-clés Vie privée, protection des données personnelles, législation, réglementation, outils informatiques, traçabilité, identité numérique, anonymat, droit à l'oubli, droit à l'image, propriété intellectuelle, responsabilité, conflits de lois.

1 Introduction

L'apparition de nouvelles techniques suscite toujours des manifestations variées et leur adoption par la société résulte généralement d'un processus d'ajustements successifs. L'informatique n'échappe évidemment pas à la règle et sa présence toujours croissante dans tous les aspects de notre vie quotidienne l'expose nécessairement, à des réactions de crainte, voire de méfiance, à la mesure de ses succès. Les atteintes à la vie privée figurent parmi les premières sources d'inquiétude à son égard. Pour ce qui est de l'actualité récente, on peut penser notamment aux remous qui ont suivi la publication du décret instituant le fichier EDVIGE (Exploitation documentaire et valorisation de l'information générale) [1], aux pratiques de sociétés comme Google ou Facebook qui se voient régulièrement rappelées à l'ordre par les autorités de contrôle de l'Union européenne et leur émanation, le Groupe 29 [2], ou encore au déploiement massif des caméras de surveillance, cartes à puce, puces RFID ou instruments de géo-localisation.

Différents types d'attitudes sont possibles face à cette situation : certains, encouragés en cela par des acteurs notables de l'industrie informatique dont le modèle commercial repose sur l'accumulation de données personnelles [3], en viennent à considérer qu'il faut simplement abandonner tout espoir de protection de la vie privée à l'ère numérique ; d'autres refusent de troquer un droit qu'ils considèrent comme fondamental contre des services jugés superflus ; d'autres encore, les plus nombreux, refusent de se couper du monde numérique et utilisent ses services en tentant de préserver tant bien que mal leur vie privée, ou de n'en dévoiler que ce qu'ils souhaitent. Mais ces craintes sont-elles réellement justifiées ? Le droit à la vie privée est-il un droit fondamental ou un bien négociable ? Quelles protections effectives peut-on offrir aux citoyens aujourd'hui ? On peut convenir que le développement et la diffusion extraordinairement rapides des nouvelles techniques du numérique ont au moins le mérite de poser des questions de fond qui doivent être abordées en prenant en compte toutes les facettes du problème (informatique, juridique, sociale, économique, politique, etc.).

---

*Ce document est une version « auteur » de l'article [17].

†INRIA, 655 avenue de l'Europe, Montbonnot, 38334 Saint-Ismier Cedex, France – daniel.le-metayer@inria.fr

‡Supélec, CS 47601, Avenue de la Boulaie, 35576 Cesson-Sévigné Cedex, France – guillaume.piolle@supelec.fr

1. Remous qui ont conduit au retrait dudit décret et à son remplacement par un nouveau fichier baptisé EDVIRSP...

2. Groupe de travail indépendant institué dans l'article 29 de la Directive 95/46/CE, le Groupe de l'Article 29 rassemble les représentants des autorités de protection des données des pays de l'Union européenne.

3. Voir par exemple les déclarations récentes des PDG de Google et de Facebook.

Dans cet article, nous rappelons brièvement les origines des droits à la vie privée et à la protection des données personnelles (Partie 2.1), nous analysons les menaces nouvelles posées par les technologies de l'information (Partie 2.2) et suggérons des voies à suivre pour répondre à ces menaces (Partie 2.3). Dans la partie suivante (Partie 3), nous présentons de manière plus détaillée les solutions informatiques, en insistant sur les techniques existantes et utilisables dès à présent pour améliorer la protection de sa vie privée.

2 Vie privée : nouvelles menaces, nouvelles protections

2.1 Origine du droit à la vie privée et du droit des données personnelles

Le droit au respect de la vie privée est mis en exergue dans de nombreux textes internationaux. Par exemple, l'article 12 de la Déclaration universelle des droits de l'homme précise que « Nul ne sera l'objet d'immixtions arbitraires dans sa vie privée, sa famille, son domicile ou sa correspondance, ni d'atteintes à son honneur et à sa réputation ». Cependant, le sens même de la notion de vie privée est sujet à interprétations diverses et les protections offertes par le droit varient selon les latitudes et les époques.

Une des premières références juridiques au respect de la vie privée se trouve dans l'article de deux juges américains, Warren et Brandeis, intitulé « *The right to privacy* » et publié dans le Harvard Law Review en 1890 [31]. Il est d'ailleurs intéressant de dresser un parallèle avec l'époque actuelle puisque c'est le développement des nouvelles techniques d'alors, comme la photographie, qui a amené les juges à s'inquiéter du fait qu'il n'existait aucune disposition juridique permettant de protéger les citoyens contre ces nouvelles possibilités d'intrusion dans sa vie privée. Depuis, la *common law* américaine reconnaît une responsabilité civile de droit commun pour atteinte à la vie privée : quiconque s'ingère dans la vie privée d'autrui est responsable des dommages causés par son fait. Elle distingue quatre types d'atteintes à la vie privée : (1) l'intrusion dans l'intimité ou la solitude de l'individu, (2) l'appropriation de l'image d'une personne ou de sa ressemblance avec autrui, (3) la révélation au public de faits qui relèvent manifestement de la vie privée de l'individu à condition que ces faits soient de nature à choquer toute personne raisonnable et qu'ils ne présentent pas d'élément de nature à éveiller un intérêt légitime dans le public et (4) la publication d'informations fausses et inexactes faisant apparaître l'individu sous un jour défavorable.

Force est de constater que le droit au respect de la vie privée existe aux États-Unis, mais il n'a pas la même portée qu'en Europe puisqu'il doit s'incliner devant les exigences constitutionnelles et plus particulièrement devant le premier amendement qui prévoit que « Le Congrès ne fera aucune loi (. . .) qui restreindrait la liberté d'expression ou de presse ». Le résultat est que, le plus souvent, un conflit entre le droit à la vie privée et le droit à la liberté d'expression se solde par la défaite du premier et la victoire du second. Ainsi lorsqu'une information sur la vie privée d'une personne est vraie, qu'elle a été obtenue par des moyens légaux et qu'elle présente un caractère d'intérêt public, la Cour suprême estime qu'un état ne peut pas en interdire la publication sans raison impérieuse.

Le droit des États-Unis contient, par ailleurs, une protection de toute intrusion de la puissance publique dans la vie privée des citoyens, particulièrement lorsque l'ingérence étatique prend la forme d'une intrusion au domicile de la personne, notamment par le biais du quatrième amendement sur l'interdiction des perquisitions et saisies déraisonnables.

Pour ce qui concerne les informations et données à caractère personnel, le gouvernement fédéral et les autorités administratives sont soumises aux dispositions du *Privacy Act* de 1974, qui les oblige en particulier à ne stocker que les informations pertinentes et nécessaires pour accomplir leurs fonctions, à tenir des fichiers exacts, complets, à jour et pertinents et à en garantir la sécurité par des contrôles administratifs, physiques et techniques. Le *Privacy Act* interdit le transfert à d'autres

agences fédérales des données à caractère personnel qui sont contenues dans les fichiers informatisés si ce n'est à la demande ou avec le consentement de l'intéressé. Parallèlement au *Privacy Act*, il existe de nombreuses autres lois sectorielles (banque, santé, protection des mineurs, etc.) qui apportent certaines protections des données personnelles dans le secteur privé, mais ces lois restent parcellaires et tranchent avec la démarche globalisante adoptée en Europe.

Pour ce qui est du Conseil de l'Europe, l'article 8 de la Convention européenne des droits de l'homme de 1950 énonce que « Toute personne a droit au respect de sa vie privée et familiale, de son domicile et de sa correspondance ». La Cour européenne des droits de l'homme, qui est chargée d'appliquer la Convention européenne des droits de l'homme, a développé une jurisprudence très protectrice en matière de vie privée. Pour sa part, l'Union européenne a adopté en 2000 une Charte des droits fondamentaux qui énonce notamment que « Toute personne a droit au respect de sa vie privée et familiale, de son domicile et de ses communications » (article 7) et que « Toute personne a droit à la protection de ses données personnelles. Ces données doivent être traitées loyalement, à des fins déterminées sur la base du consentement ou en vertu d'un autre fondement légitime prévu par la loi. Toute personne a le droit d'accéder aux données collectées la concernant et d'en obtenir la rectification. Le respect de ces règles est soumis à une autorité indépendantes. » (article 8). Ces principes sont traduits dans plusieurs directives européennes comme la Directive 95/46/EC relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données [27], et la Directive 2002/58/EC concernant le traitement des données à caractère personnel et la protection de la vie privée dans le secteur des communications électroniques [28] (modifiée par la Directive 2009/136/CE dans le cadre du « paquet télécoms » adopté en 2009 [29]).

En France, ce n'est que depuis la loi du 17 juillet 1970 que le Code civil français comprend un article 9 qui énonce que « chacun a doit au respect de sa vie privée ». Cependant aucun des textes ayant force contraignante en droit français ne définit précisément ce qu'il faut entendre par « vie privée ». En effet, même dans les textes législatifs les plus spécifiques, comme ceux ayant trait à la protection des données à caractère personnel tels que la loi du 6 janvier 1978, dite « Informatique et libertés » (modifiée en 2004 lors de la transposition de la Directive 95/46/EC), ne figure aucune disposition précisant le contenu de la notion de « vie privée ». La loi du 6 janvier 1978 s'applique aux traitements automatisés et non automatisés de données à caractère personnel contenues ou appelées à figurer dans des fichiers. Dans son article 2, elle définit les données à caractère personnel de manière extensive : « Constitue une donnée à caractère personnel toute information relative à une personne physique identifiée ou qui peut être identifiée, directement ou indirectement, par référence à un numéro d'identification ou à un ou plusieurs éléments qui lui sont propres. Pour déterminer si une personne est identifiable, il convient de considérer l'ensemble des moyens en vue de permettre son identification dont dispose ou auxquels peut avoir accès le responsable du traitement ou toute autre personne. » Cette dernière précision est particulièrement significative et prend un relief particulier, comme nous le verrons dans la partie suivante, à une époque où les techniques d'inférence de données sont d'une efficacité de plus en plus redoutable. D'un point de vue pratique, c'est la loi de 1978 qui a institué la Commission Nationale de l'Informatique et des Libertés (CNIL), autorité administrative indépendante qui est chargée de l'application de la loi et comporte, entre autres missions, celles d'informer et de sensibiliser, de recevoir les plaintes des citoyens et d'exercer, pour leur compte, l'accès aux fichiers intéressant la sûreté de l'État, la défense et la sécurité publique. C'est également à la CNIL que doivent être adressées les déclarations de traitement de données personnelles et, pour les traitements qui présentent des risques particuliers (données sensibles comme les données médicales), les demandes d'autorisation. Notons enfin que la CNIL dispose de moyens de contrôle et d'un pouvoir de sanction : même si elle a jusqu'à présent utilisé ces moyens de manière assez parcimonieuse, elle peut infliger des sanctions pécuniaires (dans la limite de 300 000 €) et prononcer

des injonctions de cesser des traitements de données personnelles.

Pour conclure cette introduction aux aspects juridiques, deux remarques s'imposent :

- Tout d'abord, la notion de vie privée résiste à toute définition précise. En Europe, on la rattache parfois aux droits de la personnalité, droits ayant pour caractéristiques d'être extra-patrimoniaux, incessibles et insaisissables. Aux Etats-Unis, certains la considèrent plutôt comme une extension du droit de propriété et donc, dans une certaine mesure, cessible, négociable, monnayable. Certains auteurs ont également analysé les différences d'attitude entre l'Europe et les Etats-Unis en terme d'importances relatives accordées aux valeurs de dignité et de liberté. Par ailleurs, la frontière entre privé et public a tendance à s'estomper. Ainsi le droit au respect à la vie privée du salarié est de plus en plus souvent invoqué dans le litige opposant le salarié à son employeur. Au-delà des États-Unis et de l'Europe, l'importance même que l'on accorde à la vie privée varie largement selon les cultures, les époques, et l'état des techniques elles-mêmes...
- Comme l'ont montré d'éminents juristes, il faut se garder du travers commun de réduire la protection de la vie privée à une question purement individuelle. Il s'agit d'un droit fondamental, d'une condition nécessaire à l'exercice d'autres droits fondamentaux, comme celui d'exercer sa liberté de penser, d'acquérir une autonomie, de tenir un rôle en société. C'est pourquoi certains en sont allés jusqu'à qualifier ce droit de « droit fondamental fondamental » et à le redéfinir comme un droit à l'auto-détermination, primordial dans un régime démocratique.

2.2 Société numérique : des menaces nouvelles pour la vie privée

Comme on l'a vu dans la partie précédente, la notion de vie privée est assez insaisissable, et on lui accorde généralement une grande importance sans pouvoir la caractériser précisément. De fait, les opinions et les comportements des citoyens traduisent cette incertitude : quand on les interroge, ceux-ci se montrent de plus en plus préoccupés par les atteintes à leur vie privée ; quand on observe leurs comportements, force est de constater qu'ils se montrent assez accommodants, voire même qu'ils participent délibérément à la mise en lumière de pans importants de leur vie privée. De telles attitudes traduisent parfois une incompréhension des risques, parfois elles reflètent des désirs contradictoires : échanger des informations, se mettre en valeur, jouer un rôle social tout en maintenant certains aspects de sa vie à l'abri de certains regards. Pour analyser plus précisément les nouvelles menaces que les techniques du numérique font peser sur la vie privée, il convient de distinguer différentes étapes dans ce qu'on pourrait appeler « le cycle de vie » des données personnelles : obtention, utilisation et effacement.

2.2.1 Obtention des données personnelles

Avant de traiter des données personnelles, il faut les avoir obtenues, d'une manière ou d'une autre. Les risques sont déjà nombreux à ce stade :

- Le premier risque est la collecte d'information à l'insu de la personne concernée, ou sans qu'elle en prenne vraiment conscience. De ce point de vue, la frontière de la légalité est parfois floue : l'information préalable à l'utilisateur peut être modérément visible ou compréhensible (cas des caméras de surveillance, des déclarations sur la protection de la vie privée sur Internet) et son consentement peut être plus ou moins « libre et éclairé » : on sait notamment qu'une demande de consentement présentée à un internaute pressé d'accéder à un service sur Internet a toutes les chances d'être acceptée sans véritable réflexion. Dans d'autres situations, l'internaute ou le citoyen peu averti n'ont simplement pas conscience des traces qu'il peuvent laisser de leur navigation sur Internet ou de leur usage de dispositifs comme des cartes de fidélité, puces

RFID ou dispositifs GPS. Il arrive également que la collecte soit clairement affichée, mais disproportionnée en regard des finalités des traitements (et donc en infraction à la loi de 1978) : c'est très souvent le cas des formulaires électroniques qui conditionnent l'accès aux services en ligne.

– Le deuxième risque est l'utilisation illégale par des tiers, des acteurs différents du collecteur de données initial. Pour accéder à ces données, deux cas de figure sont possibles : d'une part, la divulgation délibérée, contre rémunération, des données personnelles par le collecteur ; d'autre part, l'exploitation de failles de sécurité ou de négligences de la part du collecteur de données : depuis les CD comportant les données fiscales de 25 millions de contribuables britanniques égarées par la poste en novembre 2007, jusqu'aux failles de sécurité de Facebook et de Google récemment, il ne se passe guère de semaine sans que des fuites majeures de données personnelles ne soient révélées par la presse.
– Le troisième risque, encore largement sous-estimé, est lié aux progrès effectués en matière d'inférence de données et de techniques de « désanonymisation » [19]. De nombreuses études et expériences réelles ont montré comment des données apparemment anonymes pouvaient être analysées ou recoupées avec d'autres données disponibles pour retrouver les personnes concernées (affaires AOL, Netflix, GIC, études sur les réseaux sociaux, etc.). De fait, de simples résultats statistiques établissent par exemple que 87 % des citoyens des Etats-Unis peuvent être identifiés de manière unique à partir de la seule connaissance de leur code postal, date de naissance et sexe. Par ailleurs, la masse des données actuellement disponibles sur Internet, notamment à travers les réseaux sociaux, rend potentiellement personnelle toute donnée relative à une personne, même si cette donnée a été fournie de manière anonyme ou « anonymisée » ultérieurement.

2.2.2 Utilisation des données personnelles

Quand les données ont été collectées, le risque principal concerne leur utilisation. Dans les pays de l'union européenne, celle-ci doit normalement être circonscrite aux finalités déclarées lors de la collecte. Cependant la tentation est parfois grande, pour le responsable de traitement, d'excéder ces finalités et d'utiliser les données à des fins non prévues, par exemple pour des besoins de marketing ou pour cibler des offres commerciales. Le risque est d'autant plus significatif que le sujet ne dispose plus, après la collecte, de véritable moyen d'action et dépend largement de la fiabilité du responsable de traitement. L'autre catégorie de risques a trait à la mise en œuvre des mesures de sécurité qui incombent au responsable de traitement. On retrouve donc ici l'exploitation des failles de sécurité évoquées plus haut, mais à l'intérieur même de l'organisation ayant collecté les données. On sait malheureusement que les mesures minimales de sécurité ne sont pas toujours mises en œuvre par les responsables de traitement, notamment les mesures organisationnelles, ce qui permet à des personnes n'ayant aucun « besoin d'en connaître » d'accéder à des données parfois sensibles (informations médicales, bancaires, etc.).

2.2.3 Effacement des données personnelles

Le dernière étape du cycle de vie des données, leur effacement, est elle aussi une source de risques, au sens où ces données peuvent être gardées au-delà du temps nécessaire. Ce risque est loin d'être théorique puisque, les ressources en mémoire augmentant de manière exponentielle, il est souvent plus facile aujourd'hui de conserver les données que de les effacer. Par ailleurs, certains acteurs ont également intérêt à conserver ces données le plus longtemps possible afin d'en tirer le bénéfice maximum : c'est ainsi que les fournisseurs de moteurs de recherche sont en conflit permanent avec les autorités de protection des données européennes et le Groupe 29 à propos de la durée

de conservation des données de recherche. Non seulement les durées de conservation excessives contreviennent au principe si ardemment défendu à l'époque actuelle de « droit à l'oubli » mais elles aggravent également tous les autres risques répertoriés plus haut : plus longtemps une donnée est gardée, plus elle est susceptible d'être divulguée ou réutilisée de manière indue.

Enfin, au-delà des menaces sur les données personnelles qui ont été évoquées ici, il convient de poser la question du profilage, à mi-chemin entre la collecte de données et leur utilisation et qui, à grande échelle, peut devenir une manière de différencier les individus qui confine à de la discrimination.

2.3 Quelles réponses possibles ?

Les législateurs ont bien pris la mesure des nouveaux défis posés par les technologies de l'information et cherchent les meilleurs moyens d'y répondre avec les armes du droit. C'est ainsi que, à l'échelon européen, la Directive 2002/58/EC citée plus haut a finalement été révisée en 2009, à l'issue d'un processus extrêmement laborieux d'adoption du fameux « paquet télécom » et que des discussions ont lieu concernant l'avenir de la Directive 95/46/EC. Au niveau national, les sénateurs Détraigne et Escoffier ont déposé en novembre 2009 une proposition de loi « visant à mieux garantir le droit à la vie privée à l'heure du numérique ». Cette proposition modifierait la loi Informatique et Libertés (déjà modifiée en 2004) [4]. La CNIL elle-même mène une réflexion sur le sujet ainsi que le Secrétariat d'État à la prospective et au développement de l'économie numérique.

De manière générale, aussi bien la doctrine juridique que les experts de l'Union européenne, insistent de plus en plus sur l'application du principe de « *privacy by design* » ou « protection de la vie privée dès la conception ». Ainsi, le Groupe 29 évoqué plus haut a émis en décembre 2009 des recommandations à la Commission européenne, parmi lesquelles figuraient au premier plan la nécessité de « préciser les modalités d'application de certaines règles et principes clés en matière de protection des données (tels que le consentement et la transparence) ; d'actualiser le cadre par l'ajout de nouveaux principes (tels que la prise en compte du respect de la vie privée dès la conception et la responsabilité) ». Le « privacy by design » est une démarche indispensable car il est très difficile d'améliorer la protection de la vie privée dans des systèmes qui n'ont pas été conçus avec cette exigence et elle suscite de plus en plus de travaux de recherche en informatique. Parmi eux, on peut citer notamment :

- La conception de systèmes de péages routiers reposant sur la géo-localisation des véhicules tout en préservant la vie privée des conducteurs, en effectuant le calcul du tarif par le matériel de bord du véhicule [5] ;
- La conception de « cartes d'identité blanches », ou d'accréditation anonymes, permettant de prouver de nombreux attributs, comme le fait d'être membre d'une association, d'être titulaire d'un permis de conduire, d'un droit de séjour, d'un droit de vote, etc., sans pour autant divulguer son identité ;
- La conception de systèmes permettant de mieux protéger le consentement des individus avant toute divulgation de données personnelles, à travers l'assistance d'agents logiciels mettant en œuvre des politiques décidées au préalable ;
- La conception d'architectures permettant d'éviter la « pollution de données » et de garder la maîtrise de ses données personnelles.

La démarche de « privacy by design » [16] illustre également l'absolue nécessité de collaboration entre juristes et informaticiens : le législateur espère un soutien de la part des concepteurs de solutions informatique mais ceux-ci doivent comprendre précisément les exigences du droit et les traduire dans leurs produits. Au-delà des déclarations de soutien, le législateur doit également fournir des

4. Cette proposition de loi prévoit notamment d'imposer la publication d'une alerte en cas de fuite de données.

incitations au développement de solutions protectrice de la vie privée : faute de telles incitations, on sait que les techniques de protection (les « PET », ou « *Privacy Enhancing Technologies* ») peinent à trouver un véritable marché. Une manière de favoriser le développement d'un tel marché serait de créer des certificats ou labels « protection de vie privée » qui offriraient à leurs détenteurs des différentiateurs capables de susciter la confiance des utilisateurs. Le projet européen Europrise a déjà exploré cette voie, l'autorité indépendante de la région du Schleswig-Holstein en Allemagne décerne déjà des labels et la CNIL devrait lancer son programme en 2011.

Pour conclure sur les pistes de solution, il convient également d'insister sur la nécessité de renforcer les contrôles a posteriori, c'est-à-dire après la divulgation des données. Comme nous l'avons souligné plus haut, l'individu se retrouve généralement démuni quand ses données ont été communiquées. La loi de 1978 prévoit bien des droits d'accès à ces données (« la communication, sous une forme accessible, des données à caractère personnel qui la concernent ainsi que de toute information disponible quant à l'origine de celles-ci. ») mais les responsables de traitement n'offrent généralement pas de moyens simples pour exercer ces droits qui sont de fait très peu pratiqués et qui ne constituent pas une véritable protection. Il serait hautement souhaitable de développer une culture de la responsabilité pour rétablir autant que possible une forme d'équilibre des forces entre les collecteurs de données et les individus, aussi bien sur le plan juridique (obligation effective de transparence) que sur le plan technique (moyens de mise en œuvre de cette transparence). La dimension technique de la responsabilité (« accountability » en anglais) suscite de plus en plus d'intérêt dans la communauté académique mais il faut reconnaître, comme nous le verrons dans la partie suivante, que les informaticiens ont jusqu'à présent largement privilégié les outils de protection « a priori ».

3 Outils pour la protection de la vie privée des usagers

La protection de la vie privée, problème juridique complexe comme nous venons de le voir, est également délicat lorsque perçu d'un point de vue informatique, tout d'abord à cause des difficultés techniques réelles mais également parce qu'il est composé de nombreuses dimensions qui ne peuvent pas forcément être considérées indépendamment les unes des autres. Néanmoins, de nombreux outils existent qui peuvent aider l'utilisateur à protéger certains aspects de sa vie privée lors de son utilisation d'Internet ou d'un autre réseau informatique [6, 13]. En effet, c'est à cette occasion que se produisent de nombreuses « fuites » de données personnelles, et c'est surtout un cadre qu'il est le plus aisé à l'utilisateur de maintenir un certain contrôles sur ces données (par opposition aux divulgations en provenance d'institutions tierces, comme un employeur ou une administration). Les technologies associées ont donc principalement pour objet d'éviter que des données personnelles ou sensibles ne soient indûment communiquées à des tiers. Seuls de rares outils s'intéressent au contrôle de l'utilisation qui est faite de ces données une fois qu'elles sont collectées.

Nous allons, dans la mesure du possible, présenter des outils d'utilisation courante, facilement utilisables par des non-spécialistes et pour lesquels de nombreuses sources d'information sont disponibles. Nous n'évoquerons donc pas les solutions de protection de la vie privée qui relèvent davantage du domaine de l'expérimentation ou de la recherche.

3.1 Notions de cryptographie

Nombre de solutions techniques présentées ici reposent sur des moyens cryptographiques. Il est utile d'avoir une certaine compréhension de quelques concepts de base.

La cryptographie est la science de la protection des messages à l'aide de termes secrets, appelés **clés**. Le **chiffrement** est la transformation, à l'aide d'une clé, d'un **message en clair** en un message

chiffré, ou cryptogramme. Le **déchiffrement** est la restitution du message en clair à l'aide de la clé, et le **décryptage** désigne la récupération (par un attaquant) du message en clair sans l'aide de la clé. Une petite clarification lexicographique rarement superflue : les termes « cryptage » et « crypter » n'existent pas en français (on utilisera « chiffrement » et « chiffrer »), à l'inverse de « décryptage » et « décrypter » qui sont bien définis. Dans les scénarios de cryptographie, on a l'habitude de considérer deux individus, Alice et Bob, qui cherchent à échanger des messages, en assurant notamment leur confidentialité (propriété voulant que seuls les destinataires d'un message aient accès à sa version en clair).

3.1.1 Chiffrement symétrique

Le chiffrement symétrique est le principe de chiffrement le plus classique et le plus ancien. Alice et Bob partagent la connaissance d'une même clé secrète, qui servira à la fois au chiffrement et au déchiffrement. Il existe des techniques pour qu'Alice et Bob puissent se mettre d'accord sur une clé secrète même s'ils ne se sont jamais vus auparavant, et sans que cette clé puisse être devinée en écoutant leur conversation.

Les mises en œuvre techniques de chiffrement symétrique sont généralement rapides, et nécessitent des clés de taille relativement faible. RC4, DES, 3DES ou AES sont des exemples d'algorithmes courants.

3.1.2 Chiffrement asymétrique

La cryptographie asymétrique met en œuvre des paires de clés, constituées d'une **clé publique** et d'une **clé privée**. Comme leur nom l'indique, la clé privée reste secrète alors que la clé publique peut être divulguée. Dans le cas de la transmission d'un message secret entre Alice et Bob, Alice dispose de la clé publique de Bob, alors que seul Bob connaît sa clé secrète. Alice va chiffrer le message avec la clé publique, et Bob le déchiffrera avec sa clé secrète. Ce n'est donc pas la même clé qui sert pour les deux opérations, mais ces deux clés sont étroitement liées. Ce système permet d'éviter aux deux interlocuteurs de devoir se mettre d'accord sur une clé secrète, puisque la clé publique de Bob est à la disposition de tout le monde (sur un serveur de clés publiques).

D'autres utilisations des clés publiques et privées permettent également à Alice de signer son message, prouvant ainsi à Bob que c'est bien elle qui l'a écrit (**authentification**) et qu'il n'a pas été modifié entretemps (**intégrité**). Ce type de méthode est également à l'origine de la notion de **certificat**, document qui garantit l'identité de son propriétaire. Pour simplifier, un certificat est constitué d'une clé privée conservée par le propriétaire et d'une clé publique qui sera signée (donc garantie) par l'autorité de certification, et rendue publique. Les certificats peuvent éventuellement être attachés à une URL ou une adresse e-mail. Ils ont une date d'expiration, au-delà de laquelle ils devront être considérés comme invalides, et leur propriétaire peut les révoquer en cas de problème, les rendant publiquement invalides.

Les mises en œuvre techniques de cryptographie asymétrique sont généralement plus lentes, et nécessitant des clés de plus grande taille que dans le cas de la cryptographie symétrique. RSA, DSA et ElGamal sont des exemples d'algorithmes courants.

3.2 Protection des communications

Un des premiers moyens fournis à l'usager de protéger sa vie privée est d'utiliser des outils d'anonymisation [5] de sa connexion ou de protection des contenus échangés. Ces outils s'appuient

5. Les termes « anonymisation » et « anonymiser » ne semblent pas exister en français, mais ils n'ont pas d'équivalent pratique et nous nous permettrons cette licence lexicographique largement répandue.

directement sur les méthodes cryptographiques que nous venons d'évoquer. Suivant les cas, ces outils permettent de cacher le contenu de la communication aux tiers, de masquer l'adresse IP de l'usager, et/ou de masquer l'identité de ses interlocuteurs.

3.2.1 TLS/SSL

Le protocole SSL (*Secure Socket Layer*), maintenant remplacé par le protocole TLS (*Transport Layer Security*) bien que le terme « SSL » soit toujours communément utilisé dans les deux cas, est le standard de chiffrement des communications HTTP. Il contribue à sécuriser la communication des usagers avec les sites web, en rendant cette communication incompréhensible aux éventuels attaquants qui seraient capables d'écouter la ligne. Il consiste en un chiffrement symétrique de tous les messages échangés entre le terminal de l'usager et le serveur distant.

L'utilisation de ce protocole permet l'authentification du serveur (l'usager a la garantie qu'il ne s'adresse pas à un « imposteur »), la confidentialité des données échangées, leur intégrité et, dans certains cas (comme par exemple en France pour la déclaration des revenus en ligne), l'authentification de l'usager vis-à-vis du serveur. Ce protocole est largement utilisé pour les sites web (adresses en HTTPS) mais peut également être utilisé pour le transfert de fichiers (protocole FTP sur SSL/TLS) et le courrier électronique (SMTP, IMAP, POP sur SSL/TLS).

Il est important de bien comprendre ce qui est sécurisé et ce qui ne l'est pas. La communication entre client et serveur est bien chiffrée, mais le serveur aura évidemment accès aux informations en clair. TLS/SSL n'a donc d'intérêt que si l'on a confiance dans les fournisseurs du service distant. D'autre part, même si l'on visite un site en mode HTTPS, il est tout-à-fait possible de remplir alors un questionnaire qui enverra les informations soit en clair (sans chiffrement), soit à un site tiers ! Toutefois, les navigateurs sont généralement configurés par défaut pour avertir l'utilisateur dans ce cas. D'autre part, les adresses IP ne sont nullement masquées et un attaquant à l'écoute peut déterminer qu'il y a eu un échange plus ou moins important entre le client et le serveur.

Cet outil de protection de la vie privée peut être considéré comme « passif », au sens où il ne relève généralement pas du choix de l'utilisateur, mais du fournisseur de service. L'utilisateur peut parfois choisir d'accéder à certains services en utilisant TLS ou SSL, mais il faut que cette option lui soit proposée. La technologie est particulièrement intéressante en ce sens ou elle fournit une certaine protection à l'utilisateur sans qu'il ait besoin de rien faire, sinon éventuellement de réagir à des alertes ponctuelles (ce qui implique déjà de les comprendre !).

Cela devrait toutefois devenir une habitude pour l'utilisateur de vérifier que le site qu'il visite (avec SSL/TLS ou pas) est bien celui qu'il pense, et notamment que le nom de domaine dans la barre d'adresse est le bon. C'est le meilleur moyen pour éviter de devenir victime d'un site de *phishing*, fraude consistant à présenter à l'utilisateur une page web ressemblant à celle d'un site de confiance (banque, site marchand) pour l'inciter à fournir des informations sensibles, mais détournant ces informations au profit du fraudeur. Un site de *phishing* peut tout-à-fait proposer un chiffrement SSL/TLS, qui n'est alors pas une garantie en soi.

3.2.2 Serveurs mandataires

Les serveurs mandataires, ou *proxies*, sont des serveurs agissant comme des intermédiaires entre l'utilisateur et le service auquel il souhaite accéder (un site web, par exemple). L'utilisateur envoie sa requête au proxy, qui la transmet au service et sert également d'intermédiaire pour le message de retour. Les informations d'acheminement des données (adresses IP, notamment) sont modifiées par le proxy pour faire disparaître l'adresse de l'utilisateur et la remplacer par celle du proxy. De cette manière, le fournisseur de service a l'impression d'avoir affaire au proxy et non à l'utilisateur.

Le poste utilisateur est donc « masqué » derrière le proxy. Ce masquage peut n'être que partiel, un proxy pouvant, suivant sa configuration, révéler l'adresse IP de l'utilisateur (dans ce cas on utilise vraisemblablement le proxy pour d'autres raisons, qui peuvent être très variées).

Cependant, seules certaines informations sont modifiées par le proxy, et le contenu du message en lui même peut contenir une quantité d'informations identifiantes. Les proxies sont donc spécifiques à un protocole (proxies HTTP pour le web, par exemple), de manière à pouvoir anonymiser les informations qui lui sont spécifiques. Par défaut les communications ne sont pas chiffrées, mais cela peut être proposé en option.

Il convient de noter que les proxies mal administrés sont couramment utilisés pour couvrir les traces de malveillances informatiques diverses. D'autre part, il faut rester conscient du fait que les administrateurs du proxy sont en mesure d'accéder à l'intégralité des messages, de tenir des registres de toutes les activités de l'utilisateur et de procéder plus facilement à certaines attaques informatiques. Il faut donc, pour utiliser un proxy, avoir confiance dans ses administrateurs.

3.2.3 Réseaux privés virtuels

Un réseau privé virtuel, ou VPN (pour *Virtual Private Network*), est une solution technique permettant de chiffrer l'intégralité des activités réseau de l'usager, de masquer à d'éventuels attaquants observant ses communications l'identité de ses interlocuteurs et de masquer à ces interlocuteurs l'adresse IP réelle de l'usager. C'est donc en première approximation une version améliorée du proxy.

Le VPN est un service auquel l'utilisateur doit souscrire. Il peut par exemple être fourni par son employeur ou par une société commerciale. L'utilisateur devra lancer un logiciel sur son poste (un client VPN), s'authentifier auprès d'un serveur VPN, et le logiciel fera passer toutes ses connexions futures, de manière chiffrée, via le serveur VPN. Ainsi, un observateur local aura l'impression que l'utilisateur ne communique qu'avec le serveur VPN (et non avec des sites web par exemple), et un observateur distant que toutes les communications proviennent du serveur VPN (et non du poste de l'utilisateur). De plus les connexions avec le serveur sont chiffrées : on parle de tunnel entre le poste utilisateur et le serveur VPN.

Le VPN agissant à un niveau relativement bas (contrairement au proxy), ce n'est pas seulement le trafic web qui est concerné par la protection mais l'ensemble des opérations réseau1. Il faut bien réaliser cependant que les communications ne sont chiffrées qu'entre l'utilisateur et le serveur VPN. Cela signifie d'une part que le fournisseur de service VPN a accès à toutes les informations que l'on essaie de protéger (il faut donc pouvoir lui faire confiance), et d'autre part que la communication entre le serveur VPN et le service que l'on souhaite joindre s'effectue en clair. L'utilisation conjointe d'un VPN et d'un chiffrement de la connexion entre l'utilisateur et le service final (via SSL/TLS par exemple) permet de contourner ce problème. Toutefois, le fournisseur de service VPN pourra toujours identifier le service distant, à défaut de pouvoir lire le contenu des messages. L'utilisation d'un deuxième moyen d'anonymisation du trafic (comme Tor, voir plus bas) permet de contourner le problème en cas de réelle nécessité.

Le VPN s'avère une technologie très efficace, à condition d'une part de pouvoir faire confiance au fournisseur, et d'autre part que ce fournisseur dispose d'une bande passante suffisante pour ne pas ralentir la connexion. Souscrire un accès VPN avant un déplacement dans un pays connu pour filtrer ou surveiller Internet est une idée plus que judicieuse pour protéger des activités susceptibles d'être censurées ou espionnées.

3.2.4 Tor

Tor (*The Onion Router*) [30] est un réseau de routeurs interconnectés, comprenant des nœuds d'entrée et des nœuds de sortie. L'utilisateur se connecte à un nœud d'entrée, et ses messages transitent par un certain nombre de routeurs (suivant un chemin aléatoire) avant de sortir du réseau Tor par un nœud de sortie et d'être transmis au destinataire final de manière conventionnelle. Les techniques de chiffrement « par couche » utilisées (expliquant la métaphore de l'oignon) assurent principalement deux propriétés. D'une part, l'adresse de l'utilisateur est masquée à tous sauf au nœud d'entrée. Le destinataire ne la connaît pas, même s'il peut renvoyer une réponse en suivant un chemin de retour à travers le réseau Tor. D'autre part, le contenu du message et l'identité du destinataire sont masqués à tous sauf au destinataire et au nœud de sortie (les nœuds de sortie doivent donc être administrés par des gens de confiance).

Pour utiliser Tor, l'usager doit installer un logiciel de proxy spécial sur sa machine et configurer ses logiciels (principalement son navigateur web, mais possiblement tout logiciel utilisant le protocole TCP/IP) pour l'utiliser. N'importe qui peut utiliser le réseau Tor sans enregistrement préalable, mais l'accès aux routeurs d'entrée et de sortie, dont la liste est publique, peuvent être bloqués dans les pays pratiquant la censure (c'est le cas en Chine) ou bien par des services n'acceptant pas les communications anonymes. D'autre part, l'utilisation de Tor ralentit énormément les activités réseau. Il est par conséquent à réserver à des tâches ponctuelles particulièrement sensibles et/ou nécessitant peu de bande passante.

Il existe des systèmes similaires à Tor pour anonymiser l'envoi de courrier électronique tout en autorisant la réception d'une réponse (réseaux de remailers), mais leur utilisation est beaucoup moins répandue.

3.3 Gestion des identités

Anonymiser sa connexion internet est une chose, mais cela n'offre aucune garantie à l'utilisateur vis-à-vis des sites qu'il visite et avec lesquels il interagit. En effet, la plupart du temps, les gestionnaires de ces sites peuvent aisément l'identifier, conserver une trace de ses différentes activités pour construire un profil comportemental ou de consommation, et même corréler ces informations avec celles collectées par d'autres sites web. Il est vrai que l'utilisateur a souvent le besoin de s'identifier d'une manière ou d'une autre sur un site web (pour bénéficier d'un service, participer à un forum, un site communautaire, etc.). Cependant, il devrait être possible de ne dévoiler que les informations strictement nécessaires à l'activité visée, sans permettre au gestionnaire de conserver des données personnelles ou de relier entre elles les différentes activités de l'usager.

3.3.1 OpenID

La technologie OpenID [20] propose un moyen d'éviter à l'utilisateur de créer des comptes sur les sites web qu'il visite. L'utilisateur ouvre un compte uniquement chez un fournisseur d'identité OpenID. N'importe qui pouvant fournir ce service, il suffit de choisir un fournisseur en qui l'on a confiance, ou même de mettre en œuvre son propre serveur OpenID. On dispose alors d'une URL particulière (du type `username.fournisseuropenid.com`) qu'il suffit de fournir au site web visité, qui vérifiera auprès du fournisseur d'identité que le visiteur est bien qui il prétend être. L'utilisateur s'authentifie donc uniquement auprès du fournisseur d'identité OpenID. Il existe également des options pour que le fournisseur d'identité héberge des informations de profil de l'utilisateur (comme ses coordonnées), et les fournisse aux sites web sur autorisation de l'utilisateur.

Cette technologie élégante vise donc à faciliter l'expérience de l'utilisateur, en lui évitant de renseigner plusieurs fois les mêmes informations et surtout de multiplier les mots de passe, mais

elle ne contribue pas réellement à anonymiser l'activité de l'utilisateur en ligne, puisque tous les sites web concernés peuvent se référer à la même URL OpenID. La technologie est de plus en plus répandue : les comptes Google, Yahoo ou Windows Live ID constituent des identités OpenID (et de nombreux autres fournisseurs existent, comme par exemple `http://openid.net/`), et de plus en plus de sites web proposent un champ OpenID comme alternative à une authentification par mot de passe.

3.3.2 U-prove

Garantir à l'utilisateur que ses diverses activités, sur un même site web ou sur plusieurs sites web, ne pourront pas être corrélées, est un problème beaucoup plus difficile techniquement. C'est l'objectif de la technologie U-prove, relativement récente [2]. Elle a été mise au point par une société (Credentica) rachetée depuis par Microsoft, qui a publié les spécifications complètes en 2010 [18], pour faciliter son interopérabilité. L'amélioration par rapport à OpenID est l'intégration de moyens cryptographiques qui permettent à l'usager de fournir au site des garanties anonymes portant sur des informations qui ne sont pas complètement divulguées. Par exemple, l'utilisateur pourrait ainsi prouver qu'il a payé pour un service en ligne, sans pour autant donner aucune information identifiante. Il n'existe pas encore réellement de produits et services disponibles pour l'utilisateur final.

3.4 Outils de chiffrement et de signature

Les méthodes cryptographiques que nous avons évoquées plus haut peuvent être utilisées quasiment telles quelles pour faire bénéficier l'utilisateur de propriétés standard de la sécurité informatique, comme la confidentialité des données, leur intégrité ou leur authenticité, ces propriétés pouvant contribuer à l'amélioration de la vie privée.

3.4.1 Chiffrement de données sur disque

Un des aspects les plus évidents de la protection de la vie privée est la capacité à chiffrer les documents que l'on estime sensibles et qui sont stockés sur un poste de travail, afin qu'une tierce personne ne puisse pas y accéder. Cela peut concerner l'intégralité des données stockées.

On pourrait considérer que le fait d'utiliser un mot de passe sûr pour accéder à son compte sur un ordinateur est un moyen de protéger ses informations. C'est effectivement le cas, si l'attaquant se réduit à une petite sœur curieuse ou à un voisin de bureau indélicat. Si l'on imagine un attaquant un minimum motivé, disposant d'un peu de temps (quelques minutes suffisent) ou capable de voler le matériel, c'est largement insuffisant. En effet, les droits d'accès attachés aux documents ne sont efficaces que si le système d'exploitation accepte de les respecter. Si l'on fait démarrer un autre système d'exploitation avec un accès au disque considéré, il pourrait accéder aux données sans aucune restriction. C'est ainsi que sur un ordinateur en double boot Windows/Linux, Linux peut accéder à l'intégralité des partitions NTFS, quels que soient les droits d'accès spécifiés sous Windows. Pour éviter ce désagrément, il faut chiffrer les données, de manière à ce que seule l'activation du compte par la saisie du mot de passe sous le système d'exploitation prévu permette son déchiffrement. Les trois grandes familles de systèmes d'exploitation proposent des méthodes de chiffrement du système de fichiers : EFS (Encrypted File System) sous Windows, FileVault sous Mac OS X, des paquetages du type cryptfs ou dm-crypt sous Linux. Dans tous les cas, c'est un chiffrement symétrique qui est mis en œuvre.

Bien entendu, de nombreux logiciels gratuits ou commerciaux existent également pour chiffrer vos données sur le disque. Ils ne se valent pas tous, le détail de la mise en œuvre technique d'un

algorithme de cryptographie étant un aspect essentiel du niveau de sécurité final (en d'autres termes, un bon algorithme mal codé peut présenter des vulnérabilités). Il convient donc de se renseigner sur le logiciel que l'on prévoit d'utiliser. Par exemple, il est maintenant de notoriété publique que le chiffrement du système de fichiers de l'iPhone, bien que s'appuyant sur l'algorithme AES-256 réputé très sûr, n'offre aucune protection d'aucune sorte et peut être contourné très facilement et très rapidement par un non-expert [11].

3.4.2 Chiffrement et signature de données électroniques

Une application plus répandue des outils cryptographiques concerne le courrier électronique. Les messages envoyés peuvent être chiffrés, pour garantir que seuls les destinataires prévus pourront les lire, et/ou signés, pour prouver l'identité de l'auteur du message. On s'appuie ici sur des méthodes de chiffrement asymétrique. En effet, si Alice veut s'assurer que seul Bob lira son message, elle devra le chiffrer avec la clé publique de ce dernier, pour qu'il puisse la déchiffrer avec sa clé privée. Si elle veut signer son message, elle ajoutera à son envoi un « résumé » cryptographique du message (le condensat, ou *hash*), qu'elle aura chiffré avec sa clé privée (le tout constituant sa signature). Bob, de son côté, déchiffrera le condensat à l'aide de la clé publique d'Alice, et vérifiera que le condensat qu'il peut calculer de son côté est bien identique à celui qui a été chiffré par Alice. On peut donc envoyer un message signé à un destinataire qui ne dispose pas de clé personnelle, mais on ne peut envoyer un message chiffré qu'à un destinataire dont on connaît la clé publique. Le type de clés et de méthodes pouvant être utilisés se divise en deux familles.

Par défaut, les clients de messagerie comme Thunderbird sont capables de chiffrer et signer les messages en utilisant des certificats personnels, signés par des autorités de certification. Par exemple, l'employeur d'Alice a pu lui fournir un certificat à son nom (ou elle a pu en acheter un auprès d'une société spécialisée) après que son identité a été vérifiée. L'employeur a signé le certificat d'Alice avec son propre certificat, lui-même signé par une autre autorité, et ainsi de suite jusqu'à une autorité racine. Les logiciels faisant confiance à un certain nombre de ces autorités racines (qui sont en général des sociétés spécialisées), ils sont à même de déterminer automatiquement si un certificat est digne de confiance ou pas (en vérifiant qu'il n'est ni périmé ni révoqué et qu'il existe une chaîne de certification remontant jusqu'à une autorité racine). Il est également possible d'ajouter des autorités racines dans son logiciel, de manière à prendre en compte les employeurs qui ne font pas signer leurs certificats par une de ces autorités (qui font payer le service). C'est le cas d'un certain nombre d'administrations. Ces certificats personnels, qui correspondent à des couples clé publique / clé privée, sont conçus pour servir soit uniquement à signer, soit à chiffrer et signer (en fonction de leur « qualité » cryptographique).

Ces certificats « institutionnels » sont facilement utilisables, mais ils font dépendre l'utilisateur de son employeur ou du fournisseur de certificat d'une manière générale. De plus, ils sont attachés à une identité réelle et ne peuvent être utilisés avec une adresse e-mail anonyme. Pour bénéficier du chiffrement et de la signature tout en échappant à ces inconvénients, on peut s'appuyer sur un autre système cryptographique, nommé PGP (*Pretty Good Privacy*) et dont GnuPG (*Gnu Privacy Guard*) [12] constitue la mise en œuvre technique la plus courante. GnuPG est un programme qui doit être installé sur le poste de l'utilisateur. Une extension doit également être installée dans le logiciel de messagerie pour ajouter les fonctionnalités à l'interface. Pour Thunderbird, il s'agit de l'extension Enigmail [26]. Une fois les logiciels nécessaires installés et configurés, l'utilisateur pourra lui-même créer son couple de clés, en choisissant lui-même l'algorithme et la taille de la clé (de manière à pouvoir à la fois signer et chiffrer les messages). La clé publique pourra alors être publiée sur un serveur dédié à cet effet, pour permettre aux tiers de vérifier la signature de l'utilisateur ou de lui envoyer des messages chiffrés. Étant donné qu'ici la clé publique n'est pas signée par une autorité

racine, le système repose sur la confiance que l'on peut avoir dans le fait qu'une clé soit effectivement rattachée à un interlocuteur donné. Enigmail permet de préciser un niveau de confiance pour chaque clé observée, ainsi que de signer soi-même la clé publique d'un interlocuteur pour fournir aux tiers une garantie sur son authenticité.

Les outils de chiffrement et de signature de messagerie électronique sont relativement répandus, ils sont simples à utiliser et très efficaces (si correctement utilisés). Il arrive toutefois (rarement) que certains webmails ou logiciels clients aient du mal à afficher des messages signés.

3.5 Contrôle de l'usage des données personnelles

La majeure partie des outils que nous avons abordés ont pour but d'éviter que des données personnelles ou confidentielles ne soient indûment diffusées. Il existe aussi des méthodes pour se protéger de certaines des utilisations qui peuvent ensuite être faites des données, notamment en matière de profilage et de publicité.

3.5.1 Configuration des navigateurs web

Dire que l'utilisation d'un navigateur web laisse des traces exploitables est devenu un poncif et nous ne nous attarderons pas sur la question. De nombreux sites web permettent aux utilisateurs de se rendre compte de la quantité d'information que leur navigateur est susceptible de transmettre à un site web. Les choix de configuration à la disposition de l'utilisateur [9] peuvent alors relever soit du contrôle de la diffusion des informations, soit du contrôle de leur utilisation.

Sans même parler de transfert d'information, il peut être utile de limiter la quantité d'information que le navigateur conserve, et qui pourraient être consultable soit par simple accès à la machine, soit à distance dans certains cas. Il est par exemple de bon ton de se demander si le besoin de conserver trois mois d'historique de visites et de téléchargements est réel, ou si l'on doit vraiment mémoriser tous les mots de passe saisis et tous les champs de formulaire renseignés. Les réponses doivent notamment être apportées en fonction des personnes ayant accès à l'ordinateur et de l'usage qui en est fait. En outre, on ne saurait trop recommander de protéger la base des mots de passe par un « mot de passe maître », comme cela est par exemple possible dans les réglages de Firefox. On pourra également s'intéresser aux outils de « nettoyage » souvent inclus dans les navigateurs pour effacer ses traces (sur la machine locale) après une session de navigation.

À tort ou à raison, les cookies (petits fichiers laissés dans le navigateur par le site web visité et récupérables lors d'une visite ultérieure) ont largement été pointés comme un vecteur d'intrusion dans la vie privée des internautes. Les cookies peuvent être utilisés pour éviter à l'utilisateur d'avoir à rentrer de nouveau un mot de passe sur un site récemment visité, ou encore à stocker un petit nombre de préférences (comme la langue d'affichage préférée, par exemple). Cependant, la plupart du temps, il ne contiennent qu'un identifiant abstrait qui lie le navigateur à un profil d'utilisateur complet dans la base de données du gestionnaire de site web, profil dont l'internaute ne sait rien. Les cookies servent également aux plates-formes de régies publicitaires à tracer l'utilisateur d'un site à l'autre et à construire son profil comportemental de manière à cibler les propositions qui lui sont faites (profil qui pourra éventuellement être revendu). Dans cette optique, les cookies servent effectivement à la fois à faciliter la collecte d'informations personnelles, à l'insu de l'utilisateur, et à exploiter ces données de profil en proposant du contenu personnalisé (ce qui peut être perçu par l'utilisateur comme un service ou comme une intrusion). Les navigateurs modernes possèdent des options d'effacement des « traces » permettant de supprimer ces cookies. Tous permettent également d'accepter ou de refuser leur utilisation. L'option permettant de refuser les cookies déposés par des sites tiers (ciblant principalement les cookies des régies publicitaires) est à recommander tout particulièrement si l'on souhaite se tenir au maximum à l'écart du profilage comportemental.

3.5.2 Filtrage des publicités

Si le refus des cookies peut permettre d'éviter ou de limiter le profilage comportemental, il existe des moyens de bloquer complètement et de manière transparente les bannières de publicité, si celles-ci constituent une nuisance pour l'utilisateur. L'utilitaire gratuit AdBlock, notamment proposé sous la forme d'une extension Firefox, est très efficace dans ce domaine.

Outre la publicité, régies publicitaires et spécialistes du profilage utilisent en outre souvent les capacités des navigateur à exécuter automatiquement le code Javascript publié sur une page web, celui-ci pouvant par exemple aller collecter des informations très fournies sur la machine de l'utilisateur pour les transmettre à un serveur (qui n'est pas forcément celui du site web visité). Il est toujours possible de désactiver complètement Javascript, mais de plus en plus de sites web s'appuient sur cette technologie : la dénomination « web 2.0 » désigne des contenus très dynamiques permis par cette technologie. L'expérience utilisateur peut donc en souffrir. En revanche, des utilitaires comme NoScript permettent de filtrer les scripts qui seront exécutés, en refusant au site web la permission d'exécuter certaines opérations.

3.5.3 Communications électroniques non sollicitées

Les communications électroniques non sollicitées, communément appelées *spam* ou *junk mail*, résultent d'une divulgation indue ou imprudente de données personnelles (en l'occurrence l'adresse électronique, mais d'autres peuvent y avoir été associées). La « fuite » de ces données ayant déjà eu lieu, la seule défense qu'il reste à l'utilisateur est une certaine forme de contrôle sur l'utilisation qui en est faite, à savoir le filtrage de ces courriers électroniques. Celui-ci peut avoir lieu à deux niveaux : sur le serveur et sur le poste de l'utilisateur.

Le filtrage du spam sur les serveurs peut relever de plusieurs techniques. Tout d'abord, la plupart des serveurs bien administrés (ce qui est souvent le cas des administrations d'une certaine taille) refusent les connexions provenant d'adresses IP ou de plages d'adresses IP connues pour émettre du spam, ou tout du moins pour en avoir la capacité. D'autre part, un logiciel de filtrage peut analyser le courrier avant de le transmettre aux utilisateurs, et lui attribuer une note en fonction d'un certain nombre de critères, avertissant qu'un message est potentiellement indésirable, et désactivant éventuellement certains éléments pouvant être dangereux (pièces jointes, contenu actif).

Sur le poste de l'utilisateur enfin, la plupart des logiciels clients de messagerie électronique disposent d'un filtrage intégré, aux actions configurables et capable d'un apprentissage des caractéristiques qui font qu'un message est légitime ou indésirable. Ces filtres, comme ceux installés sur les serveurs, sont également à même d'identifier les messages potentiellement frauduleux, notamment les messages de phishing destinés à tromper l'utilisateur sur l'émetteur du message en imitant, dans le même esprit que le phishing web, un message provenant d'un interlocuteur de confiance réclamant des informations sensibles.

4 Conclusion

Les quelques technologies que nous avons suggérées ici peuvent être des outils fort utiles à l'usager de moyens informatiques et en particulier d'Internet, en cela qu'ils l'aident à ne pas exposer démesurément sa vie privée et ses données personnelles lors de ses activités. On pourra trouver, sur des sites web dédiés à la protection de la vie privée, comme celui de l'*Electronic Privacy Information Center* [10], ou plus généralistes, comme celui de l'*Electronic Frontier Foundation* [9, 8], des listes de logiciels plus complètes ainsi que de nombreux conseils et guides. Comme nous l'avons suggéré, il existe également de nombreuses technologies à l'état de prototypes, de preuves de concept ou

même simplement de modèles théoriques [6]. Il ne nous est pas apparu pertinent de les présenter ici, car elles ne sont pour l'instant pas directement utilisables au quotidien. Les plus ambitieuses de ces technologies visent à assurer à l'utilisateur le plein contrôle de ses données personnelles, même lorsque celles-ci doivent être partagées, distribuées ou confiées à des tiers. L'apparente antinomie de cette formulation traduit la difficulté des problématiques de recherche actuelles.

Nous ne saurions trop mettre l'accent sur le fait que ces outils techniques n'offrent chacun qu'une protection limitée et spécifique. Il est très important d'être bien conscient de ce qui est protégé et de ce qui ne l'est pas : il est facile et dramatiquement courant de se croire en sécurité parce que l'on a installé un logiciel censé « sécuriser » les données ou la connexion, sans se rendre compte que la protection technique ne répond pas au risque principal. Par exemple, l'utilisateur aura beau transmettre ses informations sensibles au travers d'un canal de transmission ultra-protégé par des méthodes cryptographiques de haut vol, si son interlocuteur à l'identité non vérifiée s'avère être mal intentionné, chiffrer la connexion ne servira pas à grand-chose. Suivant les cas, l'aspect le plus important peut être la protection d'une donnée locale contre un accès non autorisé, le chiffrement ou l'anonymisation d'une connexion, l'authentification d'un tiers... Et la plupart du temps, c'est une combinaison de différentes méthodes qui est la plus adaptée.

Enfin, pour la majorité des utilisateurs, les données personnelles qui sont diffusées suites à des « fuites » techniques, des intrusions, des malveillances restent négligeables par rapport aux informations délivrées volontairement par l'usager. C'est une bonne idée de chiffrer ses messages électroniques sensibles ou d'utiliser Tor pour visiter des sites web risqués, mais il est encore plus vital de se poser les bonnes questions avant de remplir un formulaire en ligne : les informations qui me sont demandées sont-elles réellement nécessaires pour la fourniture du service auquel je souhaite accéder ? Ai-je la possibilité de fournir moins d'informations, ou d'en falsifier certaines, sans que la qualité du service en pâtisse ? Ai-je confiance dans le gestionnaire du service ? Cette collecte de données semble-t-elle en règle, ou du moins compatible, avec la législation française ? Et surtout, quel est le risque encouru si ces données sont conservées, transmises ou utilisées indûment ? Les utilisateurs ne sont souvent pas conscients de la valeur que peuvent avoir quelques coordonnées, ou même des informations apparemment non identifiantes, pour des tiers mal intentionnés ou des sites marchands. De plus, il est dans la nature humaine de minimiser les risques à long terme et d'exagérer la perception les avantages immédiats [33] : c'est ce qui nous fait remplir des questionnaires détaillés dans le fol espoir qu'il soit tiré au sort pour gagner un porte-clés / pointeur laser, avec l'illusion que les bulletins non élus seront sagement jetés à la poubelle.

En conclusion, l'éducation des usagers, non seulement aux outils techniques mais également aux problématiques générales de protection des données personnelles et à une gestion saine de leurs informations en ligne, est sans doute la défense la plus efficace pour la préservation de leur vie privée sur Internet.

Références

[1] L'informatique. Textes et documents pour la classe (TDC), June 2010. (59).

[2] S. Brands. *Rethinking Public Key Infrastructures and Digital Certificates ; Building in Privacy.* MIT Press, 2000.

[3] C. Castelluccia and M. A. Kaafar. Owner-centric networking : a new architecture for a pollution-free internet. *ERCIM News*, (7), 2009.

[4] A. Cavoukian. Privacy and radical pragmatism : change the paradigm. White paper, 2008. Information and Privacy Commissioner of Ontario.

[5] W. De Jonge and B. Jacobs. Privacy-friendly electronic traffic pricing via commits. In *Formal Aspects in Security and Trust (FAST'08)*, volume 5491 of *LNCS*, pages 143–161. Springer, 2008.

[6] Y. Deswarte and C. Aguilar Melchor. *Sécurité des Systèmes d'Information*, chapter Technologies de Protection de la Vie Privée sur Internet, pages 49–71. Hermès, Paris, France, 2006.

[7] Y. Deswarte and S. Gambs. Protection de la vie privée : principes et technologies. *Les technologies au service des droits : opportunités, défis, limites. Cahiers du CRID*, (32), 2010. Bruylant.

[8] Electronic Frontier Foundation. Surveillance self-defense : defensive technology, 2010. `https://ssd.eff.org/tech/`.

[9] Electronic Frontier Foundation. Surveillance self-defense : Web browsers, 2010. `https://ssd.eff.org/tech/browsers`.

[10] Electronic Privacy Information Center. Epic online guide to practical privacy tools, 1994-2009. `http://epic.org/privacy/tools.html`.

[11] R. Ferguson. iProtect, iEncrypt... iLeak. Trend Micro Inc., June 2010. `https://countermeasures.trendmicro.eu/iprotect-iencrypt-ileak`.

[12] Free Software Foundation, Inc. The GNU privacy guard, 2002-2004. `http://www.gnupg.org/`.

[13] I. Goldberg. *Digital Privacy : Theory, Technologies, and Practices*, chapter Privacy Enhancing Technologies for the Internet III : Ten Years Later. Auerbach, Germany, December 2007.

[14] J. Le Clainche and D. Le Métayer. De la protection des données personnelles à la protection des traitements personnalisés. 2010. to appear.

[15] D. Le Métayer. A formal privacy management framework. In *Formal Aspects in Security and Trust (FAST'08)*, volume 5491 of *LNCS*, pages 162–176. Springer, 2008.

[16] D. Le Métayer. Privacy by design : a matter of choice. *Computer, Privacy and Data Protection*, 2009. Springer Verlag.

[17] D. Le Métayer and G. Piolle. Droits et obligations à l'ère numérique : protection de la vie privée. In L. Calderan, B. Hidoine, and J. Millet, editors, *L'usager numérique*, pages 63–88, Anglet, France, September 2010. INRIA, ADBS éditions.

[18] Microsoft. Microsfot u-prove ctp, 2010. `https://connect.microsoft.com/content/content.aspx?contentid=12505&siteid=642`.

[19] P. Ohm. Broken promises of privacy : responding to the surprising failure of anonymization. Technical Report 09-12, University of Colorado Law Legal Studies Research Paper, 2009.

[20] OpenID Foundation. Openid : an actually distributed identity system. `http://openid.net/`.

[21] Y. Pouillet. Pour une troisième génération de réglementations de protection des données. In *Conférence internationale des commissaires à la protection des données*, Montreux, 2005.

[22] Y. Pouillet. The directive 95/46/EC : Ten years after. *Computer Law & Security Report*, (22) :206–217, 2006.

[23] A. Rouvroy and Y. Poullet. *Reinventing Data Protection ?*, chapter The right to informational self-determination and the value of self-development. Reassessing the importance of privacy for democracy. Springer, 2009.

[24] F. B. Schneider. Accountability for perfection. *IEEE Security and Privacy*, 7(2) :3–4, March-April 2009.

[25] D. J. Solove. "I've got nothing to hide" and other misunderstandings of privacy. *San Diego Law Review*, 44 :745, 2007.

[26] The Enigmail Project. A simple interface for OpenPGP email security, 2010. `http://enigmail.mozdev.org/home/index.php.html`.

[27] The European Parliament and the Council. Directive 1995/46/EC of the european parliament and of the council of 24 october 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. In European Union, editor, *Official Journal of the European Communities*, October 1995.

[28] The European Parliament and the Council. Directive 2002/58/EC of the european parliament and of the council of 12 july 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector. In European Union, editor, *Official Journal of the European Communities*, July 2002.

[29] The European Parliament and the Council. Directive 2009/136/EC of the european parliament and of the council of 25 november 2009. In European Union, editor, *Official Journal of the European Communities*, November 2009.

[30] Tor Project Inc. Tor : l'anonymat en ligne, 2010. `http://www.torproject.org/`.

[31] S. D. Warren and L. D. Brandeis. The right to privacy. *Harvard Law Review*, 4 :193–195, 1890.

[32] Weitzner, Abelson, T. Berners-Lee, Hanson, Hendler, Kagal, McGuinness, Sussman, and Waterman. Transparent accountable data mining : new strategies for privacy protection. Technical Report MIT-CSAIL-TR-2006-007, MIT CSAIL, 2006.

[33] A. Westin. "Freebies" and privacy : What net users think. Opinion Research Corporation, July 1999.

[34] J. Q. Whitman. The two western cultures of privacy : dignity versus liberty. *The Yale Law Journal*, 113, April 2004.