Page de Garde

Des Technologies pour protéger la vie privée sur Internet

Yves Deswarte
Directeur de Recherche CNRS

LAAS-CNRS
7, avenue du Colonel Roche
31077 Toulouse cedex 4
France

tél. +33 561 336 288
fax : +33 561 336 411

mél : Yves.Deswarte@laas.fr

Des Technologies pour protéger la vie privée sur Internet

Yves Deswarte

LAAS-CNRS — Université de Toulouse — Yves.Deswarte@laas.fr

Résumé : Alors que dans notre société, protéger sa vie privée est reconnu comme l'un des droits fondamentaux, les citoyens et les consommateurs considèrent que les nouvelles technologies représentent un danger pour la protection de leur vie privée, et ce sentiment général est un frein au développement de la société de l'information. Cet article présente deux principes de base pour protéger sa vie privée, ainsi que quelques technologies (*Privacy-Enhancing Technologies* ou *PET*) qui permettent d'améliorer la confiance des utilisateurs envers les services disponibles sur Internet.

Mots-clés : sécurité informatique, Internet, vie privée.

Abstract: While in our society, privacy is considered as a fundamental right, citizens and consumers feel that new technologies endanger their privacy, and this feeling is detrimental to the development of the Information SocietyThis paper presents two basic privacy principles and some Privacy-Enhancing Technologies (PETs) that can contribute to give users more confidence on the protection of their privacy.

Keywords: security, Internet, privacy

1. Introduction

Dans notre société, protéger sa vie privée est considéré comme l'un des droits primordiaux, reconnu entre autres par la charte européenne des droits fondamentaux [Charte 2000]. Cette considération est à l'origine de nombreuses législations internationales sur le stockage, le traitement et la transmission de données personnelles. En 1978, la France était l'une des premières nations à se doter d'une loi protégeant les données personnelles, la loi dite « Informatique et Libertés » [Loi 1978], révisée en août 2004 [Loi 2004]. En septembre 1980, l'Organisation de Coopération et de Développement Économiques (OCDE) a publié un guide sur la protection et les mouvements transfrontières des données personnelles [OCDE 2002]. En décembre 1990, l'Assemblée générale des nations unies a elle aussi adopté un guide sur les fichiers informatiques de données personnelles. En 1995, la directive européenne 95/46/EC protégeant les données à caractère personnel a été adoptée, et elle est maintenant transposée dans les législations nationales de tous les pays membres de l'Union européenne. Malgré cette réglementation, les citoyens et les consommateurs considèrent que les nouvelles technologies représentent un danger pour leur vie privée, et ce sentiment général est un frein au développement de la *société de l'information*. La réglementation est vue comme mal appliquée, comme le montrent de fréquents incidents qui font la manchette des journaux. Il est donc nécessaire de développer des technologies qui permettent de garantir la protection des données personnelles, et ainsi regagner la confiance du public.

À première vue, la vie privée devrait pouvoir être protégée par des moyens classiques de sécurité informatique : au fond, il ne s'agit de garantir que la confidentialité de données ou de méta-données[1] personnelles, et il existe de nombreuses techniques pour protéger la confidentialité. Mais "*le diable se niche dans les détails*" des techniques de sécurité : pour contribuer à prouver qu'une action est légitime (ou non), il faut collecter et conserver beaucoup d'informations qui peuvent nuire à la vie privée. Ainsi le développement actuel des techniques de traçabilité et d'authentification forte se justifie par un souci d'améliorer la sécurité, mais il met en danger la vie privée des utilisateurs. Il faut donc protéger la vie privée par des technologies adaptées.

Les *critères communs* [Critères 2006] définissent la protection de la vie privée (*privacy*) comme une classe de fonctionnalité, avec quatre exigences principales :

- l'anonymat (*anonymity*) exige que d'autres utilisateurs ou sujets soient incapables de déterminer l'identité d'un utilisateur associé à un sujet ou à une opération ;
- la possibilité d'agir sous un pseudonyme (*pseudonymity*) exige qu'un ensemble d'utilisateurs ou de sujets soit incapable de déterminer l'identité d'un utilisateur associé à un sujet ou à une opération, mais que cet utilisateur réponde quand même de ses actions ;
- l'impossibilité d'établir un lien (*unlinkability*) exige que des utilisateurs ou des sujets soient incapables de déterminer si le même utilisateur a déclenché certaines opérations spécifiques dans le système ;
- la non-observabilité (*unobservability*) exige que des utilisateurs ou des sujets ne puissent pas déterminer si une opération est en cours d'exécution.

Ce sont ces exigences que devraient satisfaire les technologies de protection de la vie privée (*Privacy-Enhancing Technologies* ou PET) pour améliorer la confiance des utilisateurs de services Internet. Mais l'implémentation de ces technologies ne doit bien sûr pas aller à l'encontre d'autres exigences critiques, comme la lutte contre la criminalité. Par exemple,

---

[1] Les méta-données sont des informations nécessaires au fonctionnement des systèmes informatiques et des réseaux de communications qui ne sont pas directement gérées par les utilisateurs. Certaines de ces méta-données peuvent avoir un caractère personnel, par exemple, les adresses IP.

l'usage de pseudonymes (avec les propriétés indiquées ci-dessus) doit être préféré à un anonymat total : les PET devraient protéger les utilisateurs légitimes vis-à-vis d'entreprises ou d'agences qui tenteraient de violer leur vie privée, sans pour autant aider des criminels à perpétrer des actions illégales en toute impunité. Cela est possible si les PET sont fournies et maintenues sous le contrôle de citoyens honnêtes, qui coopèrent avec les autorités judiciaires dûment habilitées, mais qui ne contribueraient pas à un usage abusif par d'autres parties.

Pour protéger la vie privée des utilisateurs d'Internet, on peut considérer plusieurs catégories de PET, qui sont brièvement décrites dans les sections suivantes : gestion des identités, communications et accès anonymes, schémas d'autorisation respectant la vie privée. La conclusion introduira quelques exemples d'autres PET, qui ne sont pas directement reliés à l'Internet.

Mais avant de présenter ces technologies, il convient de poser les principes de base de la protection de la vie privée.

2. Principes de la protection de la vie privée

Les données personnelles appartiennent à la personne à laquelle elles se rapportent, elles n'appartiennent pas au propriétaire du système qui stocke ces données. Ainsi, on reconnaît généralement, au moins en Europe, que le dossier médical appartient au patient, et non pas au médecin qui le crée ou le met à jour, ni à l'hôpital qui le conserve dans ses fichiers. En tant que propriétaire de ses données personnelles, tout individu doit pouvoir exercer sa souveraineté sur ces informations et les garder sous son contrôle direct, autant que possible. Cela signifie que les données personnelles devraient être stockées de façon permanente seulement sur des dispositifs directement sous son contrôle, plutôt que sur des serveurs qu'il ne contrôle pas, et ces données ne devraient être divulguées à un tiers que selon le principe du *besoin d'en connaître* :

- les données divulguées devraient se limiter à celles strictement nécessaires pour l'accomplissement de la tâche décidée ou acceptée par la personne (principe de *minimisation des données personnelles*) ;
- le tiers ne devrait garder ces informations confidentielles et ne devrait y accéder que pour réaliser la tâche désignée par la personne ; il devrait donc les effacer dès qu'elles ne sont plus nécessaires à l'accomplissement de cette tâche (principe de *souveraineté*).

Ces deux points sont analysés dans les sous-sections suivantes.

2.1. Minimisation des données personnelles

La première mesure de minimisation consiste, pour l'utilisateur, à ne divulguer que les informations qui sont réellement nécessaires pour remplir l'objectif prévu pour la transaction, et ces informations ne devraient être divulguées qu'aux seules parties qui en ont besoin. Prenons un exemple simplifié d'une transaction d'achat sur Internet, impliquant un client, un marchand, une entreprise de livraison, la banque du client et celle du marchand. Le marchand a besoin de savoir ce qui est acheté, et il doit être sûr qu'il sera payé pour cet achat (validité du moyen de paiement), mais il n'a pas besoin de connaître l'identité de l'acheteur, la banque de cet acheteur, l'adresse de livraison, etc. L'entreprise de livraison doit connaître l'adresse de livraison et éventuellement l'identité du destinataire (qui n'est pas nécessairement l'acheteur), ainsi que les caractéristiques physiques de l'objet acheté (poids, dimension, si c'est fragile ou pas, etc.), mais elle n'a pas à connaître le prix d'achat, qui l'a acheté, etc. La banque du marchand doit savoir quel montant doit être viré et depuis quelle banque, mais n'a pas à connaître l'identité de l'acheteur ni même son numéro de compte, ce qui a été acheté, etc. La banque de l'acheteur doit connaître le montant à virer ainsi que la banque et le numéro de

compte du marchand à créditer, mais pas ce qui a été acheté, l'adresse de livraison, etc. Dans cet exemple, aucune des parties n'a besoin de collecter toutes les informations personnelles relatives à cette transaction ; elles doivent être fragmentées et distribuées entre les parties, et ceci contribue à la confidentialité des données personnelles [Fabre & Deswarte 1996].

Même lorsque l'identité d'une personne est nécessaire et doit être vérifiée (authentification), cette personne doit pouvoir sélectionner laquelle de ses identités elle désire divulguer (voir section 3), de façon à se prémunir contre des liens qui pourraient être établis abusivement. Mais très souvent, des données personnelles sont collectées, traitées et redistribuées alors qu'il n'y a pas d'utilité à les relier à une personne identifiée. C'est le cas, par exemple, lorsque des données médicales sont rassemblées dans un but de recherche sur de nouveaux traitements, ou pour protéger la population contre des épidémies, ou encore pour améliorer les ratios efficacité/coûts des soins médicaux. De telles données médicales sont des informations personnelles très sensibles, et doivent donc être anonymisées pour protéger la vie privée des patients : l'étude qui utilise ces données peut avoir besoin d'informations très détaillées sur les trajectoires de soins, voire les antécédents familiaux, etc., mais pour autant il n'est pas nécessaire de connaître la personne à qui ces données se rapportent. Dans certains cas, les données d'un même patient peuvent devoir être collectées par des sources différentes (hôpitaux, cabinets médicaux...), et à différents moments, et pourtant devoir être identifiées comme reliées au même patient, sans pour autant divulguer l'identité du patient. Cela nécessite la création d'un *identifiant anonyme* unique pour chaque patient mais commun aux différents points et instants de collecte des informations. Un tel identifiant anonyme devrait être généré de façon à empêcher toute tentative d'inverser le processus d'anonymisation, c'est-à-dire de faire correspondre une personne d'une population donnée à un identifiant anonyme. De plus, dans certains cas, il peut être utile, ou même vital, de "désanonymiser" l'identifiant pour réidentifier le patient, par exemple lorsqu'un nouveau traitement a été découvert pour une maladie identifiée à partir des données médicales anomymisées. Une technique d'anonymisation [Abou El Kalam *et al.* 2004] a été développée pour contrôler la possibilité ou non d'établir un lien entre données anonymisées correspondant à un même patient et pour donner au patient le contrôle sur le processus de désanonymisation.

Les données personnelles anonymisées doivent elles-mêmes parfois être minimisées, en particulier pour empêcher des attaques par inférence, qui permettraient d'identifier une personne à partir de données anonymisées. Par exemple, la connaissance des semaines de naissance de deux frères ou sœurs suffit pour identifier de façon unique leur mère dans une grande population (celle de la France, par exemple). Il peut donc être nécessaire de réduire la précision de données personnelles anonymes, par appauvrissement des données ou filtrage. Par exemple, il peut être souhaitable de remplacer une date de naissance par un âge ou une tranche d'âge, un code postal par un code de région, etc. Bien sûr, il faut établir un bon compromis entre la protection de la vie privée et la nécessité d'avoir des informations suffisamment précises pour réaliser l'étude pour laquelle elles sont collectées (selon le principe du besoin d'en connaître). Le filtrage et l'appauvrissement peuvent aussi s'appliquer à des données personnelles à caractère nominatif, de façon à permettre à un utilisateur de ne transmettre que la quantité d'information nécessaire pour réaliser une transaction acceptable.

2.2. Souveraineté sur les données personnelles

Lorsque des données personnelles se trouvent sur un site distant, c'est-à-dire une machine qui n'est pas sous le contrôle direct de la personne concernée (typiquement, un serveur d'une entreprise ou administration), soit pour un court moment (par exemple l'exécution d'une simple transaction), soit pour plus longtemps (par exemple des dossiers médicaux dans un hôpital), l'accès à ces données devrait être strictement limité à l'usage

souhaité par leur propriétaire, c'est-à-dire la personne correspondant à ces données. Cela signifie que le propriétaire des données doit pouvoir imposer une politique de protection de la vie privée sur ses données et que le serveur qui conserve et traite ces données doit mettre en œuvre cette politique par des mécanismes de contrôle des accès à ces données. La politique en question peut définir des permissions et des interdictions précisant qui peut ou ne peut pas réaliser quelle opération sur ces données personnelles, mais aussi des obligations précisant, par exemple, que les données expirent (et donc doivent être effacées) après un délai donné suivant la terminaison de la transaction, ou que la divulgation de ces données à un tiers doit être notifiée au propriétaire par courriel, etc. Bien sûr, la politique de vie privée imposée par le propriétaire des données doit être compatible avec la politique de sécurité qui protège les biens de l'entreprise et gouverne l'exécution de l'application, et donc les accès effectifs aux données. La compatibilité entre ces deux politiques doit être vérifiée avant la divulgation par l'utilisateur de ses données personnelles (voir le projet P3P précédemment cité).

Suivant ces principes, le propriétaire du serveur est responsable de la sécurité des données personnelles qu'il héberge, et peut être poursuivi en cas d'abus de ces données. Il est donc important que le serveur implémente des mécanismes de sécurité capables de mettre en œuvre efficacement les exigences de vie privée des utilisateurs et la politique de sécurité de l'entreprise. Cela peut se faire en utilisant des mécanismes de contrôle d'accès conventionnels et une conception rigoureuse des logiciels d'applications et des procédures d'opération. Mais cela peut aussi être facilité par des mécanismes dédiés, tels que des médiateurs d'accès aux données, capables de mettre en œuvre des *politiques indétachable*s (*sticky policie*s) associées à chaque donnée personnelle élémentaire, sous forme d'une étiquette indétachable. L'article [Casassa Mont 2003] présente une telle architecture, utilisant un chiffrement basé sur l'identité (*Identity-Based Encryption*).

3. Gestion des identités virtuelles

Pour qu'une personne puisse protéger sa vie privée, il est important de cacher ou de réduire autant que possible les liens entre cette personne et les actions et données correspondantes. Par exemple, si une personne est le seul utilisateur d'un ordinateur connecté à Internet avec une adresse IP fixe[2], il est possible pour un observateur d'associer à cette personne toutes les informations émises depuis cette adresse IP : l'adresse IP peut alors être considérée comme un identifiant unique, c'est-à-dire qu'il est propre à une seule personne. De tels identifiants uniques permettent d'établir un lien fort entre différentes actions indépendantes réalisées par la même personne, ou entre des ensemble d'informations liées à la même personne. C'est donc une menace directe contre la vie privée, et en particulier vis-à-vis de la troisième exigence des critères communs présentés précédemment (*unlinkability*).

Un moyen pour réduire les risques d'établissement de tels liens consiste à utiliser des communications anonymes et des accès anonymes aux services (voir la section 3). Mais bien souvent c'est insuffisant, puisque pour obtenir un service personnalisé, l'utilisateur doit se faire reconnaître avec une *identité*. L'identité peut être définie comme la représentation d'une personne pour un service. Cette fois encore, si une personne accède à plusieurs services sous la même identité, il est possible d'établir un lien entre ces accès. Aussi est-il souhaitable d'avoir des identités virtuelles (ou *pseudonymes*) multiples pour accéder à des services multiples. Bien sûr, chaque personne doit pouvoir sélectionner quelle identité utiliser pour chaque service, et doit pouvoir gérer la validité temporelle de ses identités : si la même identité est utilisée pour plusieurs accès à un ou plusieurs services, il est possible d'établir une

[2] Le fait qu'un utilisateur a toujours (ou souvent) la même adresse IP peut s'observer facilement, par exemple dans les en-têtes des courriels qu'il envoie.

correspondance entre ces accès, et cette correspondance peut être plus ou moins sensible du point de vue de la vie privée. L'utilisateur devrait donc pouvoir définir une date d'expiration pour chacune de ses identités, les deux choix extrêmes étant une *identité valide une fois* (une nouvelle identité doit être générée à chaque accès) et une identité *permanente*.

Différentes identités peuvent aussi être utilisées pour différents niveaux d'exigences vis-à-vis de la vie privée. Par exemple, certaines identités virtuelles ne servent qu'à permettre d'enregistrer certaines préférences de l'utilisateur, sans que ce soit des données personnelles directement identifiantes (ou *à caractère nominatif*). Pour un service de météo par exemple, ces préférences porteront sur le choix de la ville, ou des unités (degrés Celsius ou Fahrenheit, miles ou kilomètres, etc.). En revanche, d'autres identités pourraient être dédiées aux accès à des services sensibles, tels que ceux de déclaration d'impôts ou de vote électronique, etc. La vérifiabilité des identités doit être directement liée à la sensibilité des services : aucune authentification n'est utile pour accéder à un service non sensible qui ne stocke ni ne gère aucune donnée personnelle à caractère nominatif, alors qu'une authentification forte devrait être exigée pour des services sensibles, pour empêcher toute usurpation d'identité. De plus, un utilisateur devrait sélectionner des identités différentes pour accéder à des services sensibles différents. Par exemple, il faudrait utiliser des identités différentes, sans lien direct entre elles mais toutes deux avec une forte authentification, pour déclarer ses impôts et pour s'inscrire sur des listes électorales, de façon à se prémunir contre des abus éventuels de certaines administrations ou gouvernements. Ainsi, il serait dangereux d'utiliser un unique certificat à clé publique (par exemple stocké sur une carte d'identité nationale électronique) pour l'accès à tous les services publics. Au contraire, il est souhaitable d'utiliser des certificats différents, même s'ils sont tous émis par des services de l'État, à condition bien sûr qu'il ne puisse y avoir collusion entre ces services.

La gestion, par un utilisateur, de ses multiples identités est un problème qui peut être complexe [Identity Management Systems 2003], et pourtant elle doit être rendue suffisamment facile pour être utilisable et compréhensible par chaque citoyen, quelles que soient ses aptitudes techniques. Ceci est l'un des défis majeurs du projet européen PRIME (*Privacy and Identity Management for Europe*)[3].

4. Communications et accès anonymes

L'écoute passive de communications est une menace importante contre la vie privée puisque, même si le contenu d'une communication peut être chiffré, la simple observation des adresses source et destination (dans les paquets IP, par exemple) peut révéler des informations sensibles.

4.1. Définitions et exemples

On peut définir plusieurs types d'anonymat pour les communications. L'*anonymat d'émissio*n est obtenu par un utilisateur face à un attaquant lorsque celui-ci est incapable de détecter l'émission de messages par l'utilisateur. De manière similaire, l'*anonymat de réception* est obtenu par un utilisateur face à un attaquant lorsque celui-ci est incapable de détecter la réception de messages par l'utilisateur. L'*anonymat relationnel* est obtenu par un groupe d'utilisateurs face à un attaquant lorsque l'attaquant est incapable de savoir si deux membres du groupe communiquent entre eux ou pas. De ces définitions, il découle de façon immédiate que si l'anonymat d'émission (ou de réception) est garanti à chaque utilisateur d'un groupe, le groupe possède la propriété d'anonymat relationnel.

---

[3] <http://www.prime-project.eu/>

Il est important de remarquer que les anonymats d'émission, de réception et relationnel ne procurent pas par eux-mêmes la confidentialité du contenu des messages : il faut pour cela utiliser des techniques appropriées, comme le chiffrement. À l'inverse, le chiffrement du contenu des messages en soi ne procure pas l'anonymat des communications.

Ces différents types d'anonymat sont intimement liés. Par exemple si un utilisateur n'a que l'anonymat d'émission, et pas d'anonymat de réception, l'augmentation soudaine du nombre de messages reçus par lui pourrait dévoiler l'existence d'une communication et donc indirectement qu'il est en train d'envoyer des messages. La cohérence des propriétés d'anonymat dans un système doit donc être analysée de façon rigoureuse.

4.2. Les MIX et les réseaux de MIX

En 1981 David Chaum a introduit le problème de l'analyse de trafic, en soulignant que la vie privée des utilisateurs était en danger dès lors qu'un observateur peut déterminer l'existence d'une communication ou identifier deux personnes qui communiquent entre elles [Chaum 1981]. Pour empêcher l'analyse de trafic, il a proposé un protocole utilisant des routeurs qu'il a appelés *MIX*. Les MIX sont des routeurs qui cachent le lien entre les messages entrants et sortants. Un attaquant peut essayer de rétablir ce lien grâce à deux techniques simples. La première consiste à analyser les contenus des messages pour reconnaître un message entrant parmi les messages sortants. Cette attaque peut être passive (par simple lecture et analyse des messages), ou active, par l'introduction de faux messages. Par exemple, un attaquant peut comparer les tailles des messages entrants et sortants ou réémettre un message qui a été précédemment envoyé de façon légitime, pour reconnaître parmi les messages sortants lequel est répété, identifiant ainsi la route du message original (attaque par rejeu). Pour éviter de telles attaques, les MIX utilisent du bourrage et un chiffrement aléatoire : en fonction de l'implémentation, les messages entrants peuvent être déchiffrés puis chiffrés à nouveau (comme dans la figure 1), ou simplement déchiffrés si plusieurs couches de chiffrement sont appliquées au message original.

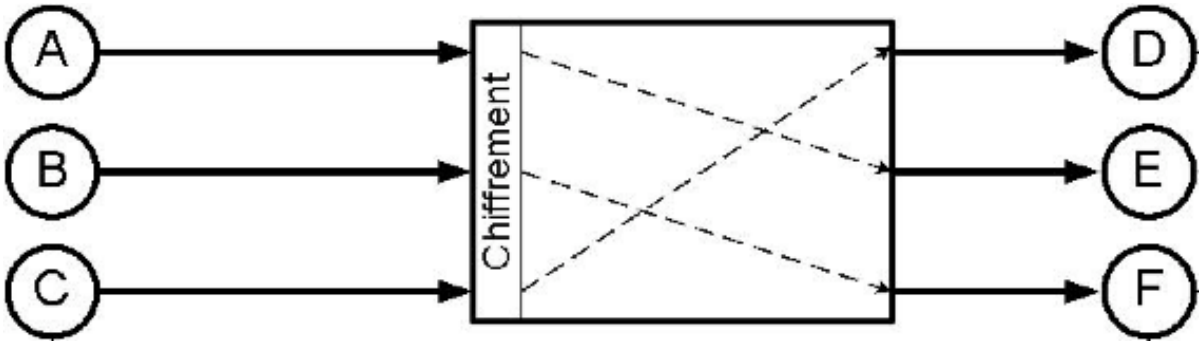

Figure 1 : Un MIX simple

Un deuxième type d'attaques est basé sur le fait qu'un message entrant et un message sortant ont d'autant plus de chance d'être liés que le temps qui sépare la réception de l'un de l'émission de l'autre est court. Pour contrer cette analyse temporelle, un MIX peut utiliser deux techniques différentes. S'il reçoit un trafic important, le MIX peut retarder les messages pendant un laps de temps variable et aléatoire. De cette façon, un message donné sera caché parmi les messages reçus juste avant et juste après sa propre réception. Le nombre de messages qui couvriront un message donné dépend de la quantité de trafic et du temps moyen d'attente. Une deuxième option pour le MIX est d'attendre jusqu'à avoir reçu un nombre `n` de messages et les envoyer tous d'un coup. Chaque message sera ainsi caché parmi les `n-1` autres messages, indépendamment du trafic ; cependant, s'il y a peu de trafic pendant une

certaine période, les délais introduits par un MIX de ce type peuvent être excessifs. Pour diminuer les délais de transit, des faux messages peuvent être envoyés au MIX, voir insérés par le MIX lui-même pour assurer que les tampons se remplissent assez fréquemment.

On peut imaginer d'autres attaques contre un MIX, et diverses parades ont été élaborées pour les contrer. Pour une présentation plus détaillée des MIX récents, le lecteur est invité à consulter l'étude présentée dans [Dìaz & Preneel 2004].

Lorsque l'on utilise un MIX, un attaquant qui réalise une écoute passive des communications peut détecter que quelqu'un communique, mais il ne peut savoir avec qui, puisque les informations entrantes et sortantes d'un MIX ne peuvent être reliées, sauf par celui qui contrôle le MIX (celui qui contrôle le MIX peut bien sûr établir un lien entre tout paquet entrant et sortant, et donc pourrait identifier l'origine et la destination de toute communication le traversant). Pour empêcher cela, les messages peuvent être envoyés à travers deux MIX. Dans ce cas, le premier MIX peut identifier une communication entre l'émetteur du message et le second MIX, alors que le second MIX peut identifier une communication entre le premier MIX et le destinataire, mais aucun des deux MIX ne peut établir de lien entre l'émetteur et le destinataire, sauf s'ils collaborent tous les deux pour réaliser cette attaque. En généralisant cette approche, si le message est envoyé à travers un grand nombre de MIX, il est impossible de relier l'émetteur et le destinataire, sauf si absolument tous les MIX du chemin qui les relie collaborent pour établir ce lien : l'anonymat relationnel est garanti entre l'émetteur et le récepteur si au moins l'un de ces MIX refuse de collaborer avec les autres. Cette structure de MIX est appelée *cascade* (figure 2).

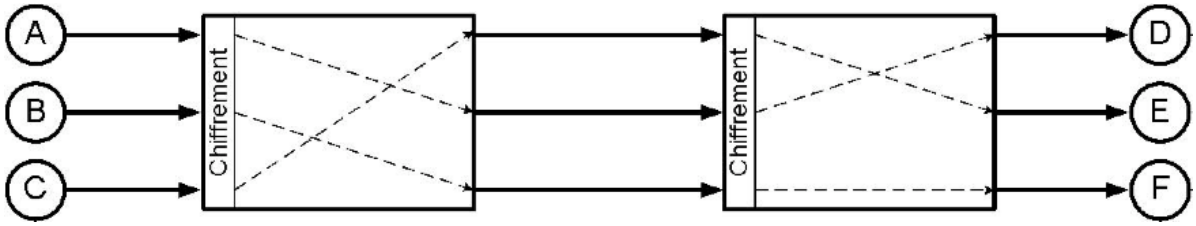

Figure 2 : Exemple de cascade

Le principal problème avec les cascades centralisées [Berthold *et al*. 2000, Goldschlag *et al*. 1999], c'est que lorsqu'un groupe de gens décide de créer une cascade ou un ensemble de cascades, il est difficile pour un utilisateur d'être sûr que les membres de ce groupe ne vont pas collaborer contre lui. Et même s'ils ne collaborent pas, il faut garder à l'esprit qu'un ensemble donné de MIX peut être piraté ou forcé d'une autre manière à collaborer pour révéler le routage des informations qu'ils transmettent ou ont transmis précédemment, ainsi que l'a prouvé l'incident[4] subi par JAP, il y a quelques années.

Pour ces raisons, un effort important de recherche a été mené ces dernières années sur les structures de réseaux de MIX pair à pair, tels que Tarzan [Freedman & Morris 2002] et Morphix[5], où n'importe quel utilisateur peut devenir un MIX et participer à une cascade. Dès lors, pour transmettre un message, on choisit aléatoirement des MIX dans un large ensemble de volontaires répartis partout dans le monde, ce qui réduit considérablement les risques de collusion : à la fois le grand nombre et la rapide évolution de l'ensemble de MIX garantit que ce serait une tâche colossale que de les pirater ou les contraindre à révéler leurs informations de routage. Ces systèmes sont appelés des réseaux de MIX (voir figure 3). Si les utilisateurs sont des nœuds du réseau de MIX, les messages transitant par chaque nœud (plus les faux messages, si nécessaire) forment un *trafic de couverture* pour les messages réellement émis

[4] <http://www.*datenschutzzentrum*.de/material/themen/presse/anon-bka_e.htm>
[5] <http://www.morphix.org/>

ou reçus par l'utilisateur correspondant à ce nœud, et donc l'émission et la réception ne peuvent être détectées par une simple écoute passive. Cela garantit les anonymats d'émission, de réception et relationnel.

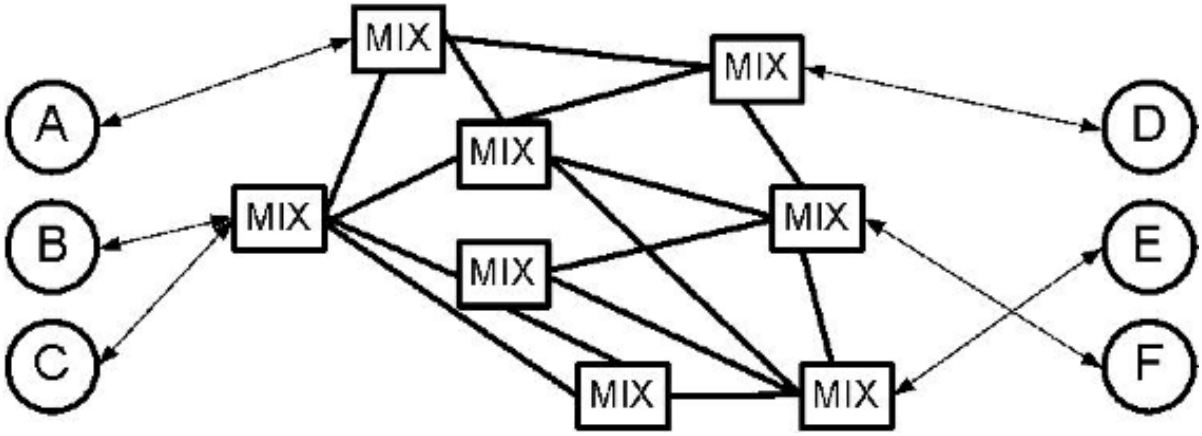

Figure 3 : Réseau de MIX

4.3. Accès anonymes aux services Internet

Fournir des communications anonymes ne suffit pas pour obtenir un accès anonyme à un service : les messages envoyés au fournisseur de service peuvent contenir des informations identifiantes, qu'il faut effacer ou transformer par un *mandataire* (*proxy*) avant qu'elles ne soient transmises au fournisseur. Cette transformation dépend de la sémantique du message (c'est-à-dire de la signification de son contenu), et la tâche peut donc être très ardue. Si le mandataire est dédié à un service spécifique, il est relativement aisé d'analyser la syntaxe des en-têtes, par exemple, pour éliminer une partie des informations sensibles. Cependant, la structure des messages requis par la plupart des services peut être très variable, et donc très difficile à anonymiser.

Bien sûr, utiliser un seul mandataire pour accéder à un service suppose que l'on ait confiance dans ses administrateurs, puisqu'ils peuvent enregistrer des informations sensibles sur l'application comme sur les communications. La façon la plus sûre de naviguer consiste sans doute à combiner un relais d'anonymat local au niveau application avec un réseau de MIX au niveau communication. Néanmoins cette solution est coûteuse et peut parfois être remplacée par une solution basée sur un mandataire distant.

5. Schémas d'autorisation respectant la vie privée

Les services Internet peuvent être classés en deux catégories : les services d'accès public, et les services d'accès restreint. Les services d'accès public devraient être accessibles sans authentification, et donc tout utilisateur devrait pouvoir y accéder sans se faire identifier, puisqu'aucune autorisation spécifique n'est nécessaire. En revanche, les services restreints ne sont fournis que selon des règles strictes d'autorisation. Dans la plupart des cas, l'autorisation est basée sur le modèle client-serveur : le serveur décide d'accorder ou de refuser ses services au client, en fonction de règles définies par le serveur lui-même. Généralement, le serveur exige que le client soit enregistré, identifié et authentifié (par exemple, en échangeant un mot de passe au travers d'une adresse de courriel valide). Le serveur stocke aussi les données de la transaction pour servir de preuve en cas de litige. Pour cela, le serveur doit rassembler de nombreuses informations personnelles sur le client pour chaque transaction, et ces données

pourraient être utilisées abusivement à d'autres fins que la résolution de litiges : profilage de clients, marketing direct, y compris par *spam*, chantage et autres. Dans certains cas, ces données peuvent être commercialisées auprès d'autres entreprises, voire être utilisées pour usurper l'identité d'utilisateurs vis-à-vis d'autres serveurs[6]. Le projet *Platform for Privacy* Preferences[7] (P3P) du *World-Wide-Web Consortium* vise à réduire ces risques, en vérifiant que les exigences de respect de la vie privée du client sont compatibles avec la politique de sécurité et de préservation de la vie privée proclamée par le serveur. Mais en fait P3P ne garantit pas que la politique proclamée est bien mise en œuvre par le serveur, avec des mécanismes de sécurité suffisamment efficaces, en particulier pour la gestion des données personnelles dans les procédures non routinières comme la gestion des sauvegardes pour des systèmes de secours, les évolutions technologiques des serveurs, la transmission de fichiers clients à d'autres entreprises, la faillite ou la fusion d'entreprises, etc.

Ce schéma asymétrique client-serveur s'adapte mal à de nombreuses applications sur Internet : non seulement il est dangereux pour la vie privée, mais en plus il est peu pratique lorsqu'il y a plus de deux parties prenantes dans une transaction. Par exemple, une transaction typique d'achat sur Internet implique non seulement un acheteur et un marchand, mais aussi un établissement de carte de crédit, la banque du client et celle du marchand, une entreprise de livraison, sans parler des fournisseurs d'accès à Internet ou les opérateurs de télécommunications. Pour résoudre ce problème dans le modèle client-serveur, l'une des parties, par exemple le marchand, sert généralement d'intermédiaire avec toutes les autres : le marchand joue le rôle de serveur vis-à-vis de l'acheteur, mais aussi de client (mandataire) vis-à-vis des autres parties (serveurs). Pour cela, le marchand collecte généralement d'abord auprès de l'acheteur suffisamment d'information personnelle (y compris sa ou ses identités) pour pouvoir exécuter la transaction auprès des autres parties. Cela est en contradiction avec le principe de *minimisation des données personnelles* (voir section2). Mais ce schéma donne aussi plus de privilèges à l'une des parties (ici le marchand) qu'aux autres. Puisque certaines de ces parties peuvent avoir des intérêts divergents et se méfier les unes des autres, il serait préférable d'adopter un schéma d'autorisation plus équilibré, où chaque partie jouerait un rôle équivalent.

Mais, bien sûr, l'aspect le plus important pour qu'un schéma d'autorisation respecte la vie privée est de séparer l'autorisation de l'authentification. En particulier, il ne devrait pas toujours être nécessaire de s'authentifier (c'est-à-dire de s'identifier et de fournir une preuve d'identité) pour obtenir des privilèges. Par exemple, l'accès à un magazine en ligne peut être restreint à ceux qui ont payé un abonnement. Une façon d'implémenter cela, tout en préservant la vie privée, pourrait être pour l'utilisateur de souscrire son abonnement en payant avec de la monnaie électronique (*e-Cash*) auprès d'un kiosque en ligne, qui lui transmettrait une preuve individuelle d'abonnement (avec une date de validité), et avec cette preuve, il pourrait récupérer depuis le site de l'éditeur du magazine tous les numéros qui couvrent la période de son abonnement. S'il paie par *e-Cash*, et si ses communications sont anonymes (voir section 3), son identité n'est transmise ni au kiosque, ni à l'éditeur. Dans cet exemple, l'éditeur accorde ou refuse l'accès, en fonction de la preuve présentée par l'utilisateur. Cette preuve d'abonnement est donc une *garantie anonyme* (*anonymous credential*), et en tant que telle, la preuve doit présenter certaines caractéristiques, dont l'*infalsifiabilité* (l'éditeur doit être sûr que l'abonnement a bien été payé), l'intransférabilité (seul l'abonné peut utiliser l'abonnement), etc. On peut imaginer d'utiliser des garanties anonymes pour prouver de nombreuses autres propriétés, comme l'adhésion à une association, la citoyenneté, le permis

---

[6] C'est aussi une des raisons pour lesquelles il vaut mieux utiliser des identités différentes (et des mots de passe différents) pour accéder à des services différents.

[7] <http://www.w3.org/TR/P3P/>

de conduire, le droit de vote, etc. Dans certaines circonstances, l'anonymat doit pouvoir être révoqué par l'émetteur de la garantie, par exemple en cas de fraude ou de falsification. Dans ces cas-là, la garantie est plus "pseudonyme" qu'anonyme. Il existe plusieurs exemples d'implémentation de telles garanties anonymes, dont IDEMIX[8].

6. Conclusion

Cet article a présenté quelques technologies qui permettraient d'améliorer la confiance des utilisateurs dans le respect porté à leur vie privée lorsqu'ils se connectent à l'Internet : techniques de gestion des identités, communications et accès anonymes, autorisation préservant la vie privée. D'autres PET ont été proposés pour des applications non directement liées à Internet. Par exemple, le développement des applications liées à la localisation soulève d'importantes inquiétudes, puisque la localisation d'une personne à un moment donné peut être une information privée très sensible. Des PET sont actuellement développés spécifiquement pour ce type d'applications. Ils se basent généralement sur une gestion d'identités multiples (pseudonymes) pour chaque utilisateur, chaque pseudonyme n'étant connu que par l'une des nombreuses parties prenantes dans le service (par exemple, un opérateur de téléphonie mobile, des fournisseurs de services lié à la localisation, des fournisseurs d'accès Internet, etc.). Le lien entre les différents pseudonymes d'une même personne doit alors être contrôlé par l'utilisateur lui-même ou par une tierce partie de confiance. De même le développement de l'informatique ubiquitaire, des réseaux de capteurs, des réseaux *ad hoc*, de routage et de stockage pair à pair, et toutes les autres technologies nouvelles d'information et de communication peuvent créer de nouveaux défis pour la protection de la vie privée, qu'il faudrait analyser avant que ces technologies ne soient déployées. Après leur déploiement, il sera peut-être trop tard pour développer des solutions pratiques acceptables, et cela risque de conduire à la frustration des usagers : comment un citoyen pourrait-il abdiquer ce qu'il considère comme un droit fondamental, pour les bénéfices souvent illusoires des nouvelles technologies ?

Remerciements

Cette étude a été partiellement soutenue par le projet PRIME (*Privacy and Identity Management for Europe*) IST-507591 <http://www.prime-project.eu/> du 6ème programme-cadre de la Commission Européenne. Une version étendue de cet article a paru dans la Revue de l'Électricité et de l'Électronique (REE), n°9, octobre 2006, pp. 65-74.

Références

A. Abou El Kalam, Y. Deswarte, G. Trouessin, E. Cordonnier, Une démarche méthodologique pour l'anonymisation de données personnelles sensibles, *Actes du 2ème Symposium sur la Sécurité des Technologies de l'Information et des Communications* (SSTIC 2004), Rennes (France), 2-4 juin 2004, pp. 91-115.

O. Berthold, H. Federrath, M. Köhntopp, Project Anonymity and Unobservability in the Internet, *Proceedings of the Workshop on Freedom and Privacy by Design* / Conference on Computers, Freedom and Privacy 2000, Toronto, Canada, 4-7 avril 2000.

M. Casassa Mont, S. Pearson, P. Bramhall, *Towards Accountable Management of Identity and Privacy: Sticky Policies and Enforceable Tracing Services*, HP Laboratories Bristol Report HPL-2003-49, 2003.

[8] <http://www.zurich.ibm.com/security/idemix/>

Charte des Droits Fondamentaux de l'Union Européenne, Journal officiel des communautés européennes (2000/C 364/01), 18 décembre 2000. <http://www.europarl.eu.int/charter/default_fr.htm>

Critères communs pour l'évaluation de la sécurité des technologies de l'information, Partie 2 : Exigences fonctionnelles de sécurité, septembre 2006, version 3.1, révision 1, CCMB-2006-09-002, disponible sur <http://www.ssi.gouv.fr/fr/confiance/methodologie.html>

D. Chaum, Untraceable Electronic Mail, Return Addresses, and Digital Pseudonyms, *Communications of the ACM*, 24(2), pp. 84-88, février 1981.

C. Dìaz, B. Preneel, Taxonomy of Mixes and Dummy Traffic, *Information Security Management, Education and Privacy*, Proc. of the 3rd Working Conf. on Privacy and Anonymity in Networked and Distributed Systems (I-NetSec04), Toulouse, France, août 2004, Kluwer Academic Publishers, pp. 216-232.

J.-C. Fabre, Y. Deswarte, L. Blain, Tolérance aux fautes et sécurité par fragmentation redondance-dissémination, *Technique et Science Informatiques* (TSI), Vol.15(4), pp. 405-427, 1996.

M.J. Freedman, R. Morris. Tarzan: A Peer-to-peer Anonymizing Network Layer, in *Proceedings of the 9th ACM Conference on Computer and Communications Security* (CCS 2002), pp. 193-206, Washington, DC, novembre 2002.

D.M. Goldschlag, D.M. Reed, P. Syverson, Onion Routing for Anonymous and Private Internet Connections, *Communications of the ACM*, 42(2), pp. 84-88, 1999.

Identity Management Systems (IMS): Identification and Comparison Study, Independent Centre for Privacy Protection (ICPP) / Unabhängiges Landeszentrum für Datenschutz (ULD) Schleswig-Holstein and Studio Notarile Genghini (SNG), 2003-09-07. <http://www.datenschutzzentrum.de/idmanage/study/ICPP_SNG_IMS-Study.pdf>

Information Technology — Open Systems Interconnection — Security frameworks for open systems: Access control framework, International Standard ISO/IEC 10181-3, 1st edition, 1996-09-15.

Loi n°78-17 du 6 janvier 1978, Loi relative à l'informatique, aux fichiers et aux libertés.

Loi n°2004-801 du 6 août 2004, Loi relative à la protection des personnes physiques à l'égard des traitements de données à caractère personnel.

OCDE, *Guidelines on the Protection of Privacy and Transborder Flows of Personal Data*, ISBN 9264197192, 2002, 66 pp.