Table de matières

Liste des figures

Liste des acronymes

APPEL : A P3P Preference Exchange Language.
AT&T : American Telephone & Telegraph Company.
API :Application Programming Interface
DTD : Document Type Definition.
DOM : Document Object Model.
EPAL : Enterprise Privacy Authorization Language.
E-P3P : Platform for Enterprise Privacy Practices.
HTML : Hypertext Markup Language.
HTTP : Hypertext Transfer Protocol.
HTTPS : Hypertext Transfer Protocol Secure.
IBM : International Business Machines.
MC : Module de comparaison.
MDRP : Module de définition des règles de préférences.
MGC : Module de gestion de comportement.
MLTP : Module de localisation et de téléchargement de la politique de confidentialité.
MSN : Microsoft Network.
OCDE : Organisation de coopération et de développement économiques.
ONU : Organisation des Nations unies.
PAP : Policy Access Point.
PDP : Policy Decision Point.
PEP : Policy Enforcement Point.
PIP : Policy Information Point.
P3P : Platform for Privacy Preference.
SAX : Simple API for XML.
SSL : Secure Sockets Layer.
SSO : Single Sign-On.
UBE : Unsolicited Build Email.
URL : Uniform Resource Locator.
W3C : World Wide Web Consortium.
XACML : eXtensible Access Control Markup Language.
XML : Extensible Markup Language.
XPath : XML Path Language.
XPref : XPath-based preference language.
XUL : XML-based User interface Language

Introduction générale

Introduction générale

L'internet est un moyen de communication sociale, culturelle, pédagogique et de loisir à l'échelle mondiale. La croissance rapide de cette nouvelle technologie conduit à réévaluer l'équilibre entre les besoins sociaux en conflit, le respect pour la vie privée, la liberté d'expression, et la poursuite des activités criminelles. Même si les avances technologiques ont pour conséquences un péril croissant pour les droits de la vie privée et le danger des abus, l'internet ne crée pas en soi de nouveaux problèmes pour la vie privée. Il rend plutôt plus difficile la maîtrise des dangers déjà existants, tels que la violation du secret, de l'intégrité des données personnelles et du courrier.

L'objectif de ce mémoire est de traiter une problématique universelle et d'actualité qui est la protection de la vie privée sur internet. L'idée de base est de développer une application qui permet de savoir si un quelconque site puise des données personnelles d'un utilisateur ou non.

Afin d'aborder tous les aspects ayant une relation avec la protection de la vie privée sur internet, le travail est organisé comme suit :
Dans le Chapitre **1** nous présentons une vue générale sur la vie privée sur internet, quelques sources de menaces et d'attaques ainsi que les technologies permettant le respect de la vie privée. Le Chapitre **2** est consacré à la protection de la vie privée d'un point de vue pratique, pour cela nous exposons les langages de sa protection qui permettent aux sites web de créer leurs politiques et aux utilisateurs de définir leurs préférences, enfin le chapitre **3** qui a pour but l'explication du système mis en œuvre, et nous terminons par une brève conclusion et quelques perspectives.

Afin de faciliter la lecture de ce mémoire, nous proposons les annexes suivantes:
Annexe A décrit un guide pour protéger votre vie privée en ligne.
Annexe B définit les éléments de P3P 1.0.
Annexe C présente les éléments d'APPEL 1.0.

Chapitre I:

La vie privée sur internet

I. Introduction

Le développement des technologies numériques de l'informatique et des réseaux, et notamment de l'internet, s'est accompagné de la promesse de retombées sociales et économiques du fait de la facilitation des échanges d'informations. Cependant, l'intégration des réseaux mondiaux dans la vie quotidienne et la poursuite des innovations technologiques multipliant les possibilités de recueil de données à caractère personnel.

L'internet a changé énormément de choses, explique Jean-Claude Kauffmann[1] : **«** *C'est un nouveau monde, un nouvel univers qui s'invente, où tout n'est pas que virtuel et à de multiples implications dans le réel. Internet reformule l'ensemble de la société.* **»**

Cependant avec l'internet, les personnes peuvent laisser derrière elles des « empreintes » électroniques ou des enregistrements des « lieux » qu'elles ont visités, des sujets qu'elles ont consultés, des pensées qu'elles ont formulées, des messages qu'elles ont envoyés et des biens et services qu'elles ont achetés. Cela pose des problèmes de vie privée dans la mesure où toutes ces données à caractère personnel exploitables sur ordinateur, qu'elle aient été générées de façon automatique ou non, sont susceptibles d'être recueillies, mémorisées, détaillées, individualisées, croisées ou exploitées pour divers usages dans des lieux géographiquement dispersés partout dans le monde, éventuellement à l'insu du consommateur ou sans son consentement.

Dans ce chapitre nous présentons une vue générale sur la vie privée pour cela on commence par la définition de certains aspects ayant une relation avec la vie privée, les sources de menaces en termes de vie privée ainsi quelques attaques et enfin les technologies permettant la protection de la vie privée sur internet.

[1] Un sociologue français, spécialiste de la vie quotidienne. Il est admis au Centre national de la recherche scientifique (Centre de recherche sur les liens sociaux, Université Paris Descartes -Sorbonne) en 1977.

II. Définitions

Dans cette partie nous montrons quelques définitions pour bien comprendre l'aspect de la vie privée sur internet, nous commençons par la définition des données à caractère personnelle, leurs traitement, la vie privée dans notre vie quotidienne, la vie privée sur internet, vie privée et la législation pour savoir l'influence de la loi sur la vie privée et en termine par la définition de la surveillance et la sécurité.

II.1. Données à caractère personnel

Les données à caractère personnel sont les informations relative à une personne physique identifiée ou identifiable directement ou indirectement, par référence à un numéro d'identification ou à un ou plusieurs éléments, propres à son identité physique, physiologique, génétique, psychique, culturelle, sociale ou économique. **[1]**

II.2. Traitement des données à caractère personnel

Le traitement des données à caractère personnel est une opération ou un ensemble d'opérations portant sur des données. La loi le définit de manière large comme tout travail exercé sur la donnée : collecte, transformation, conservation, transmission, consultation, etc. Mais le texte reste neutre face à la technologie utilisée : traitement par ordinateur, par cartes à puces, serveurs Web… Cette précaution est destinée à se prémunir des avancées technologiques inconnues à ce jour. **[2]**

II.3. Vie privée

Le concept de vie privée est flou. Il recouvre une opinion personnelle et un consensus social variable dans le temps et dans l'espace. Ses limites ne sont pas les mêmes au XIXème qu'aujourd'hui. Ses limites ne sont pas les mêmes dans un petit village ou dans une grande ville. Il se définit en opposition à la vie publique.

La vie privée est l'ensemble des activités d'une personne qui relève de son intimité par opposition à la vie publique. Le droit au respect de la vie privée est proclamé par la loi. **[3]**

II.4. Vie privée sur internet

La vie privée sur Internet est une notion plus importante que celle habituellement admise dans la vie de tous les jours. Il est primordial de bien comprendre que toute information non sécurisée mise en ligne peut être accessible par tout le monde. Cette prise de conscience de l'universalité d'Internet et de sa propension à diffuser rapidement une information importante. **[4]**

Vie privée sur Internet est le désir ou le mandat de vie privé personnels concernant les transactions ou la transmission de données via Internet.

Il s'agit de l'exercice de contrôle sur le type et la quantité d'informations sur une personne et qui peut accéder à ces informations. **[28]**

On peut citer des exemples rassemblant les informations qui composent la vie privée d'un utilisateur :

• Nom, race, origine ethnique, religion, nationalité et niveau d'instruction.

• Adresse électronique[2] et adresse IP[3].

• Taille, âge, poids, dossiers médicaux, groupe sanguin, ADN, empreintes digitales et signature vocale.

• Revenus, achats, habitudes de consommation, renseignements bancaires, données sur vos cartes de crédit/débit, rapports de prêt ou de solvabilité et déclarations de revenus.

• Numéro d'assurance sociale ou autres numéros d'identification. **[5]**

II.5. Vie privée et la législation

La frontière qui sépare la vie privée de la vie publique est variable pour chaque personne, cependant, tout le monde a une vie privée. Ce droit à l'intimité de la vie privée est valable pour tous. Si l'on regarde du côté de la loi, cette notion est clairement prise en compte :

• Convention européenne des Droits de l'Homme et des libertés fondamentales :

[2] Chaîne de caractères permettant de recevoir du courrier électronique dans une boîte aux lettres informatique.
[3] Numéro qui identifie chaque ordinateur connecté à Internet.

Art. 8 *« Toutes personnes à droit au respect de sa vie privée et familiale, de son domicile et de sa correspondance »*

• Code civil français :

Art. 9 *« Chacun a droit au respect de sa vie privée. Les juges peuvent, sans préjudice de la réparation du dommage subi, prescrire toutes mesures, telles que séquestre, saisie et autres, propres à empêcher ou faire cesser une atteinte à l'intimité de la vie privée : ces mesures peuvent, s'il y a urgence, être ordonnées en référé »*

•Guides pour l'utilisation de données personnelles informatisées et leurs transmissions internationales : OCDE [4] en septembre 1980, Assemblée Générale de l'ONU [5], en décembre 1990. **[29]**

• Protection des données à caractère personnel : Convention 108 du Conseil de l'Europe (26/01/81), directives 95/46/EC (libre mouvement) et 2002/58/CE (communications électroniques) (remplaçant la directive 97/66/CE)

• Protection des données nominatives -> à caractère personnel : loi "Informatique et Libertés" du 06/01/78, révisée par loi du 6 août 2004 + loi 94-548 (recherche médicale) http://www.cnil.fr/

La loi permet donc de se protéger, et de faire respecter sa vie privée. Si ceci est assez simple avec des acteurs clairement identifiés, il en va autrement avec Internet. **[30]**

II.6. Surveillance

La loi luxembourgeoise sur la protection des personnes à l'égard du traitement des données à caractère personnel du 2 août 2002 définit la surveillance comme « *toute activité qui, opérée au moyen d'instruments techniques, consiste en l'observation, la collecte ou l'enregistrement de manière non occasionnelle des données à caractère personnel d'une ou de plusieurs personnes, relatives à des comportements, des mouvements, des communications ou à l'utilisation d'appareils électroniques et informatisés.* » **[31]**

[4] Organisation internationale d'études économiques, dont les pays membres - des pays développés pour la plupart - ont en commun un système de gouvernement démocratique et une économie de marché.

[5] Organisation internationale, Ses objectifs sont de faciliter la coopération dans les domaines du droit international, de la sécurité internationale, du développement économique, du progrès social, des droits de l'homme et la réalisation à terme de la paix mondiale.

II.7. Sécurité

La sécurité est l'état d'esprit d'une personne qui se sent tranquille et confiante. C'est le sentiment, bien ou mal fondé, d'être à l'abri de tout danger et risque; il associe calme, confiance, quiétude, sérénité, tranquillité, assurance, sûreté. La sécurité informatique, d'une manière générale consiste à assurer que les ressources matérielles ou logicielles d'une organisation sont uniquement utilisées dans le cadre prévu.

La sécurité informatique vise généralement cinq principaux objectifs :

• L'intégrité : c'est à dire garantir que les données sont bien celles que l'on croit être.

• La confidentialité : consistant à assurer que seules les personnes autorisées aient accès aux ressources échangées.

• La disponibilité : permettant de maintenir le bon fonctionnement du système d'information.

• Le non répudiation : permettant de garantir qu'une transaction ne peut être niée.

• L'authentification : consistant à assurer que seules les personnes autorisées aient accès aux ressources. **[31]**

III. Les sources de menaces en termes de vie privée

Aujourd'hui, Internet a un vaste domaine d'application : navigation, messagerie[6], commerce électronique[7],... Ce dernier s'est largement démocratisé depuis quelques temps, ce qui nécessite une évolution permanente des moyens de sécurisation.
Malheureusement, dans cette jungle qu'est Internet, les méthodes de piratage ne cessent d'évolué et des millions de personnes sont soumises à des menaces d'intimité. Les entreprises sont engagées non seulement à regarder ce que vous visitez en ligne, mais d'infiltrer les informations et envoyer des messages publicitaires en fonction de votre historique de navigation. **[32]**
Pour cela nous présentons dans cette partie les sources importantes de menaces : moteurs de recherche, photographies sur internet, les sites de réseautage social, les fournisseurs de services, …

[6] Service qui permet d'envoyer et de recevoir des courriels.
[7] Transactions effectuées par des consommateurs et des commerces par l'entremise d'un réseau, à l'aide d'ordinateurs et de systèmes de télécommunication.

III.1. Photographies sur internet

Aujourd'hui beaucoup de gens ont des appareils photo numériques et affiche leurs photos en ligne. Les personnes représentées sur ces photos pourraient ne pas vouloir les faire apparaître sur l'Internet.

Certaines organisations tentent de répondre à cette préoccupation liée à la confidentialité. Par exemple, la conférence Wikimania [8] 2005 exigeait que les photographes aient l'autorisation préalable de la population dans leurs tableaux. **[4]**

En outre, le droit de la responsabilité traditionnelle ne protège pas les personnes qui sont capturés par une photographie en public parce que ce n'est pas compté comme une invasion de la vie privée, ainsi que des photos d'autres peuvent permettre à d'autres personnes à violer la vie privée d'une personne par trouver des informations qui peuvent être utilisées pour les suivre. **[33]**

III.2. Les moteurs de recherches

Les moteurs de recherche ont la capacité de suivre les recherches d'un utilisateur. Les renseignements personnels peuvent être révélés par des recherches, y compris des articles de recherche utilisés, le temps de la recherche, et plus encore. Les moteurs de recherche ont réclamé une nécessité de conserver ces informations afin de fournir de meilleurs services, la protection contre les pressions de la sécurité, et protéger contre la fraude.

Pour le démontrer il suffit de prendre l'exemple de Google qui peut sans trop de soucis associer votre adresse IP avec l'ensemble de vos recherches, phénomène encore accentué si vous possédez un compte mail chez eux, puisqu'alors il aura accès à des informations plus privées sur vous. **[6]**

Fort de ce constat, certains ont décidé de se distinguer comme des moteurs de recherche que « respectent votre vie privée ». On peut citer Ixquick, qui vous l'annonce d'ailleurs de manière très visible sur sa page d'accueil : Ixquick Protège Votre Vie Privée !

Derrière ces annonces, qu'en est-il vraiment ?

Tout d'abord ixquick n'est pas un moteur de recherche, mais plutôt un méta moteur, c'est à dire qu'il va chercher ses résultats dans les autres moteurs de recherche. De cette façon vous ne donnez votre adresse IP associée à une recherche qu'à un seul moteur de recherche tout en bénéficiant des résultats de l'ensemble d'entre eux. Depuis janvier

[8] Le nom des conférences internationales de Wikimedia Foundation, De périodicité annuelle, elles rassemblent les contributeurs aux projets de la Wikimédia Foundation comme Wikipédia et ses projets frères.

2009 ixquick n'enregistre plus les adresses IP. Il n'utilise les cookies qu'à des fins de conservation de la configuration du moteur de recherche, lesquels sont périmés au bout de 90 jours. Il propose également de multiples services destinés à la protection de la vie privée telle que la navigation en https[9]. **[7], [8]**

III.3. Les sites de réseautage social

Les sites de réseautage social peuvent se révéler un merveilleux moyen d'établir des liens. Ils nous permettent de garder le contact avec nos amis, d'échanger des idées avec des collègues dans nos domaines de travail et de partager de l'information avec des gens dont les passetemps et les intérêts sont semblables aux nôtres. **[9]**

Les renseignements personnels affichés sur ces types de sites finissent souvent par être utilisés de manières inattendues. Les employeurs éventuels, les patrons, les parents, les enseignants, les administrateurs d'université et d'autres catégories de gens utilisent les sites de réseautage social pour se renseigner sur les gens.

Et tout cela surprend ceux qui présument que ce qu'ils affichent est du domaine privé. Des gens ont perdu leur emploi, ont raté des entrevues d'emploi et des possibilités éducatives. On les a suspendus de l'école en raison de messages instantanés, d'« affichages muraux » (sur des sites comme Facebook) et d'autres messages qu'ils croyaient à tort constitués des conversations privées avec des amis.

Voici un titre pour donner une idée du genre de choses qui se produisent :

Un étudiant de première année en génie informatique a fait face à des mesures disciplinaires après que des responsables de l'Université Ryerson de Toronto ont découvert qu'il dirigeait un groupe d'étude sur un site de réseautage social. **[10]**

III.4. Les fournisseurs de services Internet

Le fournisseur de services internet[10] est responsable de la bonne utilisation des données. Les personnes concernées peuvent lui demander quelles sont les données qu'il collecte, traite et conserve, de quelle manière et pour quelle finalité. **[4]**

[9] HyperText Transfer Protocol Secure, est un protocole réseau utilisé pour la navigation sécurisée sur le web**,** il offre des possibilités d'authentification et de chiffrement pour les sites web nécessitant un certain niveau de sécurité dans leurs échanges avec les navigateurs web.

[10] Prestataire fournissant divers services professionnels liés à Internet, Ce sont notamment l'hébergement de serveurs Web, la création complète de sites Web, la mise en place d'une boutique électronique, etc.

Le fournisseur de service doit donc respecter les procédures appropriées et les technologies qui garantissent la vie privée des personnes concernées et notamment l'intégrité et la confidentialité des données ainsi que la sécurité du réseau et des services fournis.

Le fournisseur de service ne doit pas :

- Lire, supprimer, modifier les messages envoyés à d'autre.
- Collecter, traiter, conserver des données sur les utilisateurs sans que ce soit nécessaire. Il faut une finalité explicite, déterminée et légitime.
- Conserver les données pour une période plus longue que ce qui est nécessaire pour atteindre le but du traitement.

En évoquant ces règles, il ne faut jamais perdre de vue la question de fond suivante : où se situe exactement dans la chaîne de collecte et de traitement le réel danger pour la vie privée ? **[11]**

III.5. Le commerce électronique

On utilise souvent les services de vente en ligne qui nous permettent de rester chez nous en achetant des marchandises sur l'Internet. Un internaute qui veut acheter quelque chose peut entrer dans le site Web de vente, choisir une ou plusieurs marchandises, et puis de façon quelconque, l'internaute doit faire transférer une somme d'argent à la compte bancaire du marchand. Le marchand va envoyer les marchandises choisis chez l'Internaute. Aujourd'hui, l'utilisation de carte crédit est la façon la plus populaire pour le paiement sur l'Internet.

Parmi les risques sur la vie privée de l'internaute:

- Obtention du numéro sur le poste client

Si un virus ou cheval de Troie installé sur le poste du client pourrait facilement intercepter le numéro de sa carte dès sa saisie par l'utilisateur, et l'envoie à un mal individu.

- Interception du numéro durant sa transmission

L'interception d'un numéro de carte bancaire sur l'Internet est le problème le plus fréquemment évoqué lorsqu'on interroge des internautes à s'adonner aux joies du commerce électronique.

- Obtention sur le serveur du marchand

Il y a des risques qui se situent du côté du marchand. Supposons que maintenant le marchand soit honnête, il stocke le numéro de carte bancaire sur son serveur. Si un individu a réussit à attaquer le serveur du marchand, il peut avoir des milliers de numéros de carte, y compris votre numéro.

- Découverte d'un numéro via un logiciel spécifique

Il existe des logiciels qui permettent de générer des numéros de cartes bancaires syntaxiquement valides. Avec une probabilité très petite, un individu qui utilise ce logiciel peut obtenir un numéro qui est égal à votre numéro. **[12]**

IV. Les attaques de la vie privée

IV.1. Vol d'identité

Le vol d'identité (appelé aussi l'usurpation d'identité) c'est l'utilisation de renseignements personnels volés pour usurper l'identité de quelqu'un en vue de commettre une fraude. Le vol peut être commis dans le but d'accéder à des comptes bancaires réels, d'obtenir des prêts bancaires ou à d'autres fins frauduleuses.
Il est intéressant de noter que les catégories de vol d'identité représentant le plus grand pourcentage de victimisation sont respectivement l'utilisation de carte de débit ou crédit avec 3% et les informations personnelles compromises sans fraude avec 2,5%. **[13]**

L'usurpation d'identité débute toujours par la collecte de renseignements personnels sur l'individu fraudé. Les renseignements personnels peuvent être le nom, le numéro de téléphone, la date de naissance, l'adresse, le numéro d'assurance sociale ou toute autre information permettant d'identifier la personne. La victime de l'usurpation d'identité reste vivante, et possède donc la faculté de défendre ses droits.

Les usurpateurs utilisent ensuite ces informations pour effectuer une ou des transactions en simulant l'identité de la personne fraudée. Par exemple, un fraudeur peut effectuer des appels téléphoniques ou faire des achats importants et diriger les frais vers la personne fraudée, il peut aussi retirer de l'argent du compte de banque de cette personne…. **[34]**

IV.2. Utilisation non légitime de l'information

Chacune des informations lâchées sur Internet peut paraître dérisoire, mais mises bout à bout elles permettent de dresser un profil parfois très complet d'une personne, et cela présente un risque d'atteinte à la vie privée du fait qu'il permet d'analyser avec précision le comportement des consommateurs.

Les informations collectées par les moteurs de recherches dans le cadre de leurs activités commerciales sont ensuite traitées et exploitées pour différentes finalités (offres de services, sécurité du système, publicités personnalisées, statistiques, etc.)

Un internaute inscrit sur un site de socialisation peut parfaitement diffuser des photos ou des commentaires sur un autre individu sans même que celui-ci soit inscrit sur le site. Enfin, en révélant des informations parfois très personnelles ou intimes (opinions politiques, religion, etc.) sur ces sites, les internautes s'exposent à une exploitation commerciale de leurs données par les gérants de sites ou leurs partenaires commerciaux. **[35]**

Donc comme on a vu, il y a de nombreuses manières d'utilisation malveillantes des informations touchant directement notre intimité, de ce fait, protéger sa vie privée n'est plus une question d'éthique.

V. Quelques techniques d'attaques

V.1. Les cookies (biscuits empoisonnées)

un cookie (aussi appelé plus rarement témoin) est défini par le protocole de communication HTTP[11] comme étant une suite d'informations envoyée par un serveur HTTP à un client HTTP, que ce dernier retourne lors de chaque interrogation du même serveur HTTP sous certaines conditions. **[36]**

Les cookies ont des implications importantes dans la vie privée et l'anonymat des utilisateurs du web dans lesquels peuvent être retracés des renseignements pertinents pour son commanditaire sur les comportements de l'utilisateur d'Internet.

[11] Littéralement le « protocole de transfert hypertexte », est un protocole de communication client-serveur développé pour le World Wide Web.

Ces « mini - bases de données » constituent une source d'atteinte à la vie privée de l'internaute et à son autodétermination sur les données qui la concerne.

Les cookies permettant de:

- **Identifier un navigateur**

Certaines applications des cookies permettent d'identifier votre navigateur au fil des consultations d'un même serveur et connaître précisément la liste des documents consultés.

- **Identifier un individu sur un site**

Identifier une personne ne veut pas forcément dire qui elle est. Si cette même personne est amenée à révéler son identité en remplissant un formulaire, l'administrateur du site Web considéré va non seulement savoir qui consulte son site mais, il va pouvoir le reconnaître par la suite et savoir, à chaque visite quels documents il aura consultés.

- **Traquer un individu sur plusieurs sites**

La possibilité d'insérer des publicités sur de très nombreux sites Web par le monde, lui permettant ainsi de savoir précisément le parcours d'un même internaute parmi ces différents sites. En exemple typique est l'entreprise DoubleClick [12] et ces célèbres cookies. **[12]**

V.2. Le Phishing

Le phishing ou hameçonnage, consiste pour les escrocs à envoyer au maximum d'utilisateurs un courrier électronique censément rédigé par un site Internet respectable.

Pour une raison ou pour une autre, le courrier demande à l'utilisateur de se connecter à son compte en cliquant sur le lien fourni, dont l'adresse apparente semble bien celle du site. En fait, l'adresse est un leurre et renvoie sur un faux site, copie de l'original. Quand l'utilisateur entre ses identifiants, l'escroc les récupère et peut ensuite les utiliser pour effectuer des transactions et procéder à des détournements de fonds. **[14]**

Il existe différentes variantes à l'hameçonnage. On notera le spear phishing et le in-session phishing qui sont respectivement l'hameçonnage ciblé (notamment à l'aide des réseaux sociaux) et l'hameçonnage de session (basé sur des pop-up [13] pendant la navigation).

[12] Une régie publicitaire (ciblage comportemental) sur Internet. Elle est rachetée le 14 avril 2007 par Google pour 3,1 milliards de dollars.

[13] Nouvelle Fenêtre de Navigateur s'ouvrant automatiquement au dessus de la Fenêtre de navigation actuelle de l'internaute, il utilisé en publicité pour afficher un nouveau bandeau sans surcharger une page.

Les attaques par hameçonnage sont le plus souvent dirigées vers les sites sensibles tels que les sites bancaires. Les sites de réseaux sociaux sont aujourd'hui également la cible de ces attaques. Les profils des utilisateurs des réseaux sociaux contiennent de nombreux éléments privés qui permettent aux pirates informatiques de s'insérer dans la vie des personnes ciblées et de réussir à récupérer des informations sensibles. **[15]**

V.3. Le Scam (Arnaque)

Le « scam » (« ruse » en anglais), est une pratique frauduleuse d'origine africaine, consistant à extorquer des fonds à des internautes en leur faisant miroiter une somme d'argent dont ils pourraient toucher un pourcentage. L'arnaque du scam est issue du Nigéria, ce qui lui vaut également l'appellation « 419 » en référence à l'article du code pénal nigérian réprimant ce type de pratique.

L'arnaque du scam est classique : vous recevez un courrier électronique de la part du seul descendant d'un riche africain décédé il y a peu. Ce dernier a déposé plusieurs millions de dollars dans une compagnie de sécurité financière et votre interlocuteur à besoin d'un associé à l'étranger pour l'aider à transférer les fonds. Il est d'ailleurs prêt à vous reverser un pourcentage non négligeable si vous acceptez de lui fournir un compte pour faire transiter les fonds.

En répondant à un message de type scam, l'internaute s'enferme dans un cercle vicieux pouvant lui coûter de quelques centaines d'euros s'il mord à l'hameçon et même la vie dans certains cas. **[16]**

V.4. Le logiciel espion (spyware)

Un logiciel espion (aussi appelé mouchard ou espiogiciel ; en anglais spyware) est un logiciel malveillant qui s'installe dans un ordinateur dans le but de collecter et transférer des informations sur l'environnement dans lequel il s'est installé, très souvent sans que l'utilisateur en ait connaissance. L'essor de ce type de logiciel est associé à celui d'Internet qui lui sert de moyen de transmission de données.

Un logiciel espion est composé de trois mécanismes distincts :

- Le mécanisme d'infection, qui installe le logiciel. Ce mécanisme est identique à celui utilisé par les virus, les vers ou les chevaux de Troie.

  Par exemple, l'espiogiciel Cydoor utilise le logiciel grand public Kazaa comme vecteur d'infection.

- Le mécanisme assurant la collecte d'information. Pour l'espiogiciel Cydoor, la collecte consiste à enregistrer tout ce que l'utilisateur recherche et télécharge via le logiciel Kazaa.
- Le mécanisme assurant la transmission à un tiers. Ce mécanisme est généralement assuré via le réseau Internet. Le tiers peut être le concepteur du programme ou une entreprise.

Le logiciel espion peut afficher des offres publicitaires, télécharger un virus, installer un cheval de troie (ce que fait WhenU.SaveNow, par exemple), capturer des mots de passe en enregistrant les touches pressées au clavier (keyloggers), espionner les programmes exécutés à telle ou telle heure, ou encore espionner les sites Internet visités. **[17]**

V.5. Les pixels invisibles

Ce sont des minuscules fichiers graphiques insérés dans un courriel ou dans une page Web pour surveiller un utilisateur à son insu.

En fait, un pixel invisible se met en marche lorsque nous téléchargeons une page Web. Il commence alors la collecte d'informations: adresse IP de l'ordinateur et autres informations nous concernant. Il peut alors connaître l'identité du fournisseur d'accès, puis envoyer un cookie pour analyser les habitudes de navigation.

D'ailleurs, selon une étude d'Intelytics de l'année 2009, 75% des sites commerciaux principaux utilisent les pixels invisibles pour nous traquer.

Pour cela Privacy Foundation a lancé Bugnosis, un logiciel qui permet de traquer ces drôles de pixels. **[37]**

V.6. Autres techniques d'attaques

V.6.1. Les chevaux de troie et les vers

On appelle « Cheval de Troie » (en anglais trojan horse) un programme informatique effectuant des opérations malicieuses à l'insu de l'utilisateur.

Un cheval de Troie (informatique) est donc un programme caché dans un autre qui exécute des commandes sournoises, et qui généralement donne un accès à l'ordinateur sur lequel il est exécuté en ouvrant une porte dérobée (en anglais backdoor), par extension il est parfois nommé troyen par analogie avec les habitants de la ville de Troie. **[18]**

Un cheval de Troie peut par exemple :

- voler des mots de passe ;
- copier des données sensibles ;
- exécuter tout autre action nuisible ;
- etc.

Un ver informatique est un logiciel malveillant qui se reproduit sur plusieurs ordinateurs en utilisant un réseau informatique comme Internet.

L'objectif d'un ver n'est pas seulement de se reproduire. Le ver a aussi habituellement un objectif malfaisant, par exemple :

- Espionner l'ordinateur où il se trouve.
- Offrir une porte dérobée à des pirates informatiques .
- Détruire des données sur l'ordinateur où il se trouve ou y faire d'autres dégâts.
- Envoyer de multiples requêtes vers un serveur Internet dans le but de le saturer (déni de service). **[19]**

V.6.2. Adware

Un logiciel publicitaire (adware en anglais) est un logiciel qui affiche la publicité lors de son utilisation.

Le logiciel contient habituellement deux parties :
* une partie utile (le plus souvent un jeu ou un utilitaire) qui incite un utilisateur à l'installer sur son ordinateur.
* une partie qui gère l'affichage de la publicité.
Ces logiciels espions se renseignent sur les sites visités par un utilisateur, afin de mieux cibler le type de publicités à afficher, le plus souvent à travers des fenêtres pop-up. Les informations récoltées sont parfois stockées sur des bases de données à des fins commerciales. Il existe des logiciels intégrant des adware sans en avertir l'utilisateur. **[20]**
Pour plus de protection contre ces techniques d'attaques voir Annexe A.

VI. La protection de la vie privée sur internet

Cette section explique en quoi la protection de la vie privée est un enjeu important et de quelle manière l'Internet peut mettre en place la protection des informations personnelles. **[38]**

VI.1. Les niveaux de protection de la vie privée

- L'anonymat

C'est l'impossibilité (pour d'autres utilisateurs) de déterminer le véritable nom de l'utilisateur associé à un sujet, une opération, un objet

- La pseudonymat

Idem, sauf que l'utilisateur peut être tenu responsable de ses actes, c.à.d. il peut utiliser un pseudonyme au lieu de son vrai nom.

- La non-chaînabilité

C'est l'impossibilité (pour d'autres utilisateurs) d'établir un lien entre différentes opérations faites par un même utilisateur

- La non-observabilité

C'est l'impossibilité (pour d'autres utilisateurs) de déterminer si une opération est en cours.

Chaque entité qui utilise des données privés des utilisateurs doit respecter les principes suivants :

- La minimisation des données

Ça signifie que la seule information nécessaire pour compléter une application particulière devrait être collectée/utilisée (et pas plus).

C'est une application directe du critère de légitimité défini par la directive européenne sur la protection des données personnelles (Directive 95/46/EC).

- La souveraineté des données

Ça signifie que les données liées à un individu lui appartiennent, il devrait pouvoir contrôler comment elles sont disséminées.

C'est une extension de plusieurs législations nationales sur les données médicales qui considèrent que le dossier d'un patient lui appartient, et non pas au docteur qui le crée

ou le met a jour, ni a l'hôpital qui le stocke. Difficile a réalisé dans un monde ubiquitaire.

- Le consentement explicite

Ça signifie qu'avant de collecter les données personnelles d'un individu, il faut lui demander son autorisation et lui expliquer quelle utilisation sera faite de ses données.

- La transparence

Ça signifie que le système ne doit pas être considérer comme une boite noire dans laquelle l'individu doit avoir une confiance aveugle.

- L'imputabilité

Ça signifie que l'entité qui héberge les données personnelles doit les sécuriser au meilleur de ses moyens, et le cas échéant peut être tenue responsable (par exemple devant un juge) d'un bris de vie privée.

- Le droit à l'oubli

Ça signifie que sur la demande de l'individu, ses traces doivent être effacées. **[28]**

VI.2. Technologies de protection de la vie privée

VI.2.1. Privacy by design

C'est l'intégration de la problématique du respect de la vie privée dé la conception d'un système. Considère la question de la vie privée a priori, plutôt que de réagir a posteriori une fois que le système a été déployé et qu'on constate un bris de vie privée. **[28]**

- **Les principes fondamentaux**

- Proactive et non réactif

Le Privacy by Design est une approche qui se caractérise par des mesures proactives plutôt que réactives. Il prévoit et empêche des événements de la vie privée avant qu'ils se produisent. En bref, Privacy by Design vient avant le fait, non pas après.

- La vie privée comme un réglage par défaut

Privacy by Design vise à offrir le maximum de la vie privée en faisant en sorte que les données personnelles sont automatiquement protégées dans un système d'information et de gestion. Si une personne ne fait rien, leur vie privée demeure intact. Aucune action

n'est exigée de la part de l'individu pour protéger leur vie privée, elle est établie dans le système par défaut.

- La vie privée est intégrée dans la conception

Privacy by Design est intégré dans le design, l'architecture des systèmes et les pratiques commerciales.

Le résultat est que la vie privée devient une composante essentielle du fonctionnement. La vie privée fait partie intégrante du système, sans diminuer la fonctionnalité.

- Fonctionnalité complète (à somme positive)

Privacy by Design vise à répondre à tous les intérêts légitimes et les objectifs dans un jeu à somme positive «gagnant-gagnant » et non pas par une approche à somme nulle, où inutiles.

- Protection du cycle de vie complet

Privacy by Design, ayant été intégrés dans le système avant l'assemblage du premier élément alors des mesures de sécurité solides sont essentiels à la vie privée, du début à la fin. Cela garantit que toutes les données sont bien conservées puis détruits à la fin du processus en toute sécurité.

- Visibilité et transparence

Privacy by Design vise à assurer à tous les intervenants que toutes les pratiques sont exploitables selon les promesses et les objectifs énoncés. Ses composants et les opérations restent visibles et transparentes, pour les utilisateurs et les fournisseurs.

- Respect de La vie privée de l'utilisateur

La conception exige à des architectes de conserver les intérêts de l'individu le plus élevé en offrant des mesures telles que la vie privée forte par défaut. **[39]**

VI.2.2. Privacy Enhancing Technologies (PET)

PET est un ensemble de techniques et d'applications qui permettent à un individu de protéger ses informations personnelles pendant qu'il est en ligne.
Les "technologies de protection de la vie privée" regroupent un très grand nombre d'outils, mais ceux-ci demeurent complexes, peu standardisés et au final très peu utilisés. **[28], [40]**

- **Exemples des outils PET**

1) System de gestion d'identité

Les usages divers de l'Internet ont fait naître un peu partout dans le monde des comportements atypiques, tels que la multiplication des adresses électroniques, le recours aux pseudonymes dans les blogs, aux avatars dans les mondes virtuels, etc. Ces « identités multiples » sont plus difficiles à saisir qu'un numéro de passeport, de sécurité sociale ou de compte bancaire. **[21]**
Exemples: Microsoft passport, Single Sign-On (SSO), OpenID…

a) Windows Live ID

Windows Live ID (anciennement appelé Microsoft Passport) est un service qui permet d'utiliser une adresse de messagerie et un mot de passe uniques, appelés authentifiant, pour accéder à la plupart des sites et services de Microsoft ainsi que ceux de ses partenaires choisis.
Il permet d'enregistrer ces authentifiant (adresse de messagerie et mot de passe) à un site ou un service qui utilise Windows Live ID, ou au site Web Windows Live ID. Microsoft utilise cette identité unique pour aider à améliorer l'authentification de Windows Live ID et pour la protection contre les pourriels et l'utilisation malveillante du compte. **[22]**
Windows Live ID aide à protéger la vie privée et les informations personnelles de la manière suivante :

- Le service Windows Live ID collecte et traite les informations personnelles seulement pour les raisons suivantes :
    - Pour faire fonctionner un service d'authentification.
    - Pour aider à améliorer la sécurité.

- Pour le support technique.

- Le service Windows Live ID ne contrôle ni surveille les pratiques de confidentialité de tous les sites et services sur Windows Live ID. Les pratiques de confidentialité des sites individuels peuvent varier. Toutefois, tous les sites ou services Windows Live ID doivent être d'une déclaration de confidentialité validée. **[23]**

b) Single Sign-On (SSO)

L'authentification unique (ou identification unique ; en anglais *Single Sign-On* : SSO) est une méthode permettant à un utilisateur de ne procéder qu'à une seule authentification pour accéder à plusieurs applications informatiques (ou sites Web sécurisés).

Les objectifs sont multiples :

- Simplifier pour l'utilisateur la gestion de ses mots de passe : plus l'utilisateur doit gérer de mots de passe, plus il aura tendance à utiliser des mots de passe similaires ou simples à mémoriser, abaissant par la même occasion le niveau de sécurité que ces mots de passe offrent.
- Simplifier la gestion des données personnelles détenues par les différents services en ligne, en les coordonnant par des mécanismes de type méta-annuaire.
- Simplifier la définition et la mise en œuvre de politiques de sécurité. **[24]**

c) OpenID

OpenID est un système d'authentification décentralisé qui permet l'authentification unique, ainsi que le partage d'attributs. Il permet à un utilisateur de s'authentifier auprès de plusieurs sites (devant prendre en charge cette technologie) sans avoir à retenir un identifiant pour chacun d'eux mais en utilisant à chaque fois un unique identifiant OpenID. Le modèle OpenID se base sur des liens de confiance préalablement établis entre les fournisseurs de services (sites web utilisant OpenID par exemple) et les fournisseurs d'identité (*OpenID providers*). Il permet aussi d'éviter de renseigner à chaque fois un nouveau formulaire en réutilisant les informations déjà disponibles. OpenID permet à un utilisateur d'utiliser un mécanisme d'authentification forte. **[25]**

Une faiblesse de ce système réside dans les risques de phishing ou d'hameçonnage. On peut en effet imaginer qu'une des fraudes du système OpenID consiste à détourner

l'utilisateur ou le fournisseur de service du fournisseur d'identité vers lequel il se dirige pour authentifier l'utilisateur. En dépit de ses faiblesses, OpenID, qui en est encore au stade expérimental, constitue un système d'identité numérique global très prometteur. **[41]**

2) Accès anonyme à des services

Les PETs permettant de communiquer de manière anonyme dans un réseau, c'est à dire en protégeant l'identité de l'envoyeur et/ou du receveur du message
Exemples : Mixnets, Onion Routing, Crowds, etc.

a) Mixnets

Concept introduit par Chaum en 1981 pour empêcher l'analyse de trafic. Le Mix est un routeur qui cache le lien entre les messages entrants et sortants par un mécanisme de chiffrement et de permutation des messages, pour faire face aux espions observant les communications échangées. Parmi ceux qui ont appliqué le Mixnets le Service de courriel anonyme (Mixmaster). **[42]**
Fonctionnement d'un Mix simple :

1. Reçoit en entrée plusieurs paires du type (message; adresse du destinataire) qui ont été préalablement chiffrées.
2. Déchiffre les messages.
3. Envoie en sortie les messages à leurs destinataires correspondants (possiblement chiffrés).

La figure suivante montre le fonctionnement d'un Mixnets :

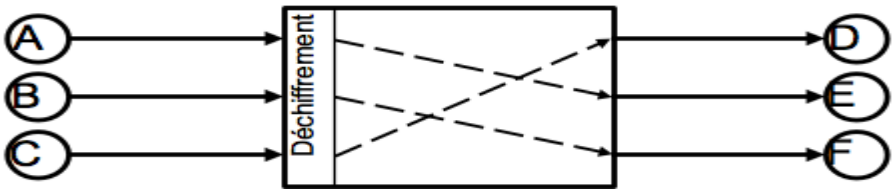

Figure I.1 : Mixnets moyenne d'accès anonyme **[42]**

b) Crowds

Protocole de communication anonyme qui protège l'anonymat de l'envoyeur d'un message en le routant de manière aléatoire vers des groupes d'utilisateurs similaires.

L'idée principale : cacher l'origine d'un message en le dispersant.

Fonctionnement de Crowds :

Initialisation : chaque nouvel utilisateur s'enregistre en tant que membre d'un groupe (appelé « Crowd ») en contactant le responsable du groupe. Quand un utilisateur rejoint un groupe, tous les membres du groupe en sont notifiés.

Le responsable du groupe est aussi chargé de la distribution des clés symétriques assurant la confidentialité entre paires de nœuds. **[42]**

c) Tor

The Onion Router ou Tor (le routage en oignon) est un réseau mondial décentralisé, organisés en couches, dont la tâche est de transmettre de manière anonyme les paquets TCP. Tout échange Internet basé sur TCP peut être anonymes en utilisant Tor. **[42]**

Tor fonctionne avec de nombreuses applications comme les navigateurs web, les clients de messagerie instantanée, les connexions à distance et tout un nombre d'application se basant sur le protocole TCP. **[43]**

3) Langages de préférence en termes de vie privée

Langages principalement basés sur le standard XML et utilisés pour permettre aux utilisateurs d'exprimer leurs préférences de confidentialité.

De même, ils facilitent la tâche des organisations pour exprimer des pratiques de confidentialité dans les serveurs Web **[44].** Le chapitre suivant détaillera ce type de langages.

VII. Conclusion

Au cours des dernières années les problèmes liés à la vie privée sur Internet sont devenus très importants aux yeux des utilisateurs.

Les usagers devraient savoir que tous les outils n'offrent pas des moyens efficaces de protéger la vie privée. Un gros désavantage tient à leur incapacité d'aborder la protection de la vie privée une fois les données sont collectées.

Pour cela de nombreux organismes internationaux se sont donc penchés sur la question, notamment le w3c (World Wide Web Consortium) qui a proposé le protocole P3P (Platform for Privacy Preference) qui sera éclairci dans le chapitre suivant.

Chapitre II:

Les langages de protection de la vie privée état de l'art

I. Introduction

La confidentialité des données est une préoccupation croissante des entreprises et autres organismes dans une variété de secteurs, tels que la santé, la finance, le e-commerce, et le gouvernement. Chaque jour, ces organisations sont chargées de la responsabilité de la gestion des renseignements personnels. Contrairement à la sécurité des données, qui se concentre principalement sur la prévention des personnes non autorisées d'obtenir des renseignements de façon inappropriée, la protection de la vie privée doit offrir aux personnes la possibilité de contrôler la façon dont leurs données sont gérées et utilisées par une organisation particulière. Pour cela de nombreux organismes proposent des solutions tel que :

- **Les langages d'expression de politique de la vie privée** : ce sont des langages qui permettent aux sites web d'informer les utilisateurs de leurs politiques vis-à-vis du respect de la vie privée.
- **Les langages de préférences en termes de vie privée :** ce sont des langages qui permettent aux utilisateurs de définir leurs préférences.
- Des outils qui comparent les politiques des sites avec les préférences des utilisateurs, et avertit l'utilisateur en cas de non-respect de ces préférences.

Dans ce chapitre nous allons voir ces langages et ces outils.

II. Les Langages d'expression de politique de la vie privée

Dans cette section nous présentons les langages d'expression de politique, ces langages devraient fournir un degré élevé de fonctionnalité, afin de couvrir tous les types définis de politique. D'ailleurs, nous considérons l'expressivité d'un langage, qui garantit la définition de toutes les parties obligatoires d'une politique.
Cette section présentera les langages XACML, EPAL et plus en détaille le P3P le standard de w3c.

II.1. P3P (The Platform for Privacy Preferences)

II.1.1. Présentation

P3P est une recommandation du w3c, qui donne la possibilité aux sites web d'informer les utilisateurs de leurs politiques vis-à-vis du respect de la vie privée. Il définit un format standardisé pour décrire ces politiques. **[27]**

P3P permet aux sites Web d'exprimer leurs politiques de confidentialité dans un format normalisé que les agents utilisateurs[14] peuvent obtenir automatiquement et interpréter aisément et cela en langage XML (eXtensible Markup Language). Les agents utilisateurs P3P permettront d'informer les utilisateurs des pratiques des sites (dans des formats lisibles à la fois par une machine et par un humain) et d'automatiser au besoin les prises de décisions en fonction de ces pratiques. Les utilisateurs n'auront donc pas besoin de lire les politiques de confidentialité de chaque site visité. . **[45]**

II.1.2. La spécification P3P 1.0

La spécification P3P1.0 définit la syntaxe et la sémantique des politiques de confidentialité P3P et les mécanismes permettant d'associer les politiques aux ressources Web. Les politiques P3P consistent en déclarations utilisant le vocabulaire P3P afin d'exprimer des pratiques touchant à la vie privée. Les politiques P3P appellent également des éléments du schéma de données de base de P3P, un jeu standard d'éléments de données que tout agent utilisateur P3P devrait reconnaître. La

[14] Application cliente utilisée avec un protocole réseau particulier ; l'expression est plus généralement employée comme référence pour celles qui accèdent au World Wide Web. Les User Agents du Web vont de la gamme des navigateurs jusqu'aux robots d'indexation, en passant par les lecteurs d'écran ou les navigateurs braille pour les personnes ayant une incapacité.

spécification P3P comprend un mécanisme permettant de définir de nouveaux éléments de données et ensembles de données et un mécanisme simple autorisant l'extension du vocabulaire de P3P. **[26]**

La figure suivante donne une vision concrète du modèle p3p.

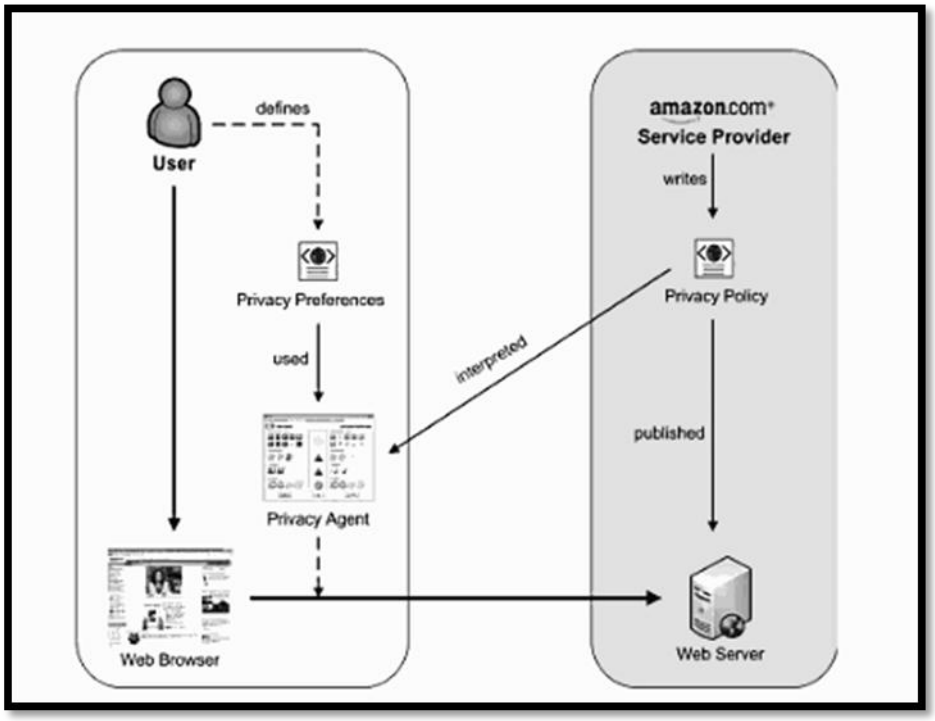

Figure II.1 : Le modèle P3P [28]

II.1.3. Les buts et les possibilités de P3P1.0

Le protocole P3P version 1.0 est conçu pour informer les utilisateurs du Web des pratiques des sites Web concernant la collecte de données. Il permet à un site Web de transcrire ses pratiques de collecte et d'utilisation des données dans un format XML, lisible par une machine, que l'on appelle *politique P3P*.

Le but de P3P version 1.0 est double :

Premièrement, permettre aux sites Web d'annoncer leurs pratiques de collecte de données de manière normalisée, lisible par une machine et facilement disponible.

Deuxièmement, permettre aux utilisateurs du Web de savoir quelles données seront collectés par les sites visités, comment ces données seront utilisées, et quels usages de ces données ces utilisateurs accepteront ou bien refuseront.

Cinq premiers topiques ont pour but d'exprimer l'intension de site Web.

- Qui est entrain de collecter cette donnée.
- Quelle information est entrain d'être collectée.
- Pour quel but.
- Quelles informations vont être partagés avec d'autre (site)
- Qui va recevoir ces informations **[26]**

II.1.4. Les politiques P3P

Les politiques P3P utilisent un codage XML avec des espaces de nommage du vocabulaire P3P afin de fournir les coordonnées de l'entité légale responsable des pratiques de confidentialité d'une politique, d'énumérer les types de données ou les éléments de données collectés et d'expliquer la destination des données en question comme il est montré dans la figure II.2. En outre, les politiques identifient les destinataires des données et divulguent d'autres informations, dont les renseignements pour la résolution des litiges et l'adresse de la politique de confidentialité d'un site écrite pour un humain. Les politiques P3P doivent couvrir tous les éléments de données et pratiques mobilisés.

Toutefois, les questions concernant le respect des demandes de renseignement légales ne sont pas traitées par cette spécification. Tout en respectant sa politique de non-redistribution des données à des tiers, un site peut être contraint de le faire par force de loi. Les déclarations P3P sont positives : les sites annoncent ce qu'ils font plutôt que ce qu'ils ne font pas. Le vocabulaire P3P est conçu pour décrire les pratiques d'un site plutôt que d'être juste un indicateur de la conformité à une loi particulière ou un code de conduite particulier. Par contre, on peut développer les agents utilisateurs de façon à tester si les pratiques d'un site sont conformes, ou non, à une loi ou un code.

Il faut remarquer que chaque politique P3P s'applique aux ressources Web spécifiques (pages Web, images, cookies, etc.) listées dans un fichier d'appel de politique. **[26]**

Pour plus d'informations sur le vocabulaire de P3P voir **Annexe B**.

```
<ENTITY>
<DATA-GROUP>
<DATA ref="#business.name">Location Provider Service</DATA>
<DATA ref="#business.contact-info.online.email">p3p@example.com</DATA>
<DATA ref="#business.contact-info.online.uri">http://www.example.com</DATA>
<DATA ref="#business.contact-info.postal.street">University Address</DATA>
</DATA-GROUP>
</ENTITY> <ACCESS><all/></ACCESS>
<DISPUTES-GROUP>
<DISPUTES resolution-type="service" service=http://www.example.com/p3p_dispute.html short-description="Dispute">
<LONG-DESCRIPTION> For any inconvenience, apply to our Customer Service (dispute@example.com) </LONG-DESCRIPTION>
<REMEDIES><correct/><money/><law/></REMEDIES>
</DISPUTES>
</DISPUTES-GROUP>
```

Figure II.2: Échantillon de politique de P3P **[26]**

II.1.5. Les agents utilisateurs P3P

Les agents utilisateurs P3P peuvent être intégrés aux navigateurs Web, aux modules d'extension des navigateurs ou aux serveurs mandataires. Ils peuvent aussi se présenter sous forme d'applets Java[15] ou de scripts JavaScript, ou être intégrés à des portefeuilles électroniques, des remplisseurs de formulaire automatiques ou à d'autres outils de

[15] Désigne les programmes développés en langage Java et que l'on trouve sur le Web. Ils fonctionnent quelque soit la plate-forme, grâce à une machine virtuelle Java (JVM), ou dans l'AppletViewer de Sun, un outil permettant de tester les applets Java. ....

gestion des données de l'utilisateur. Les agents utilisateurs P3P recherchent les appels de politiques P3P dans l'emplacement notoire, dans les en-têtes P3P des réponses HTTP et dans les balises link incorporées à un contenu HTML. Ces appels indiquent l'emplacement des politiques P3P concernées. Les agents utilisateurs peuvent récupérer la politique à l'endroit indiqué, l'analyser puis afficher des symboles, émettre des sons ou générer des invites pour l'utilisateur afin de refléter les pratiques de confidentialité P3P d'un site. Ils peuvent aussi comparer les politiques P3P aux préférences de confidentialité choisies par l'utilisateur et prendre les mesures appropriées. Un agent utilisateur n'autoriserait la délivrance des données si la politique est cohérente avec les préférences de l'utilisateur.

La spécification P3P1.0 impose peu de contraintes sur l'interface utilisateur des agents utilisateurs. Les développeurs peuvent ainsi choisir les messages et symboles à présenter à l'utilisateur pour les informer de la politique de confidentialité d'un site Web. Les développeurs ne sont pas tenus d'utiliser textuellement les définitions qui se trouvent dans cette spécification pour leurs interfaces utilisateurs. Toutefois, ils devraient s'assurer que les informations présentées à l'utilisateur, quelles qu'elles soient, représentent fidèlement les politiques P3P décrites. **[26]**

II.1.6. La mise en œuvre de P3P1.0 sur les serveurs

Les sites Web peuvent mettre en œuvre P3P1.0 sur leurs serveurs en transcrivant leurs politiques de confidentialité, lisibles par un humain, vers une syntaxe P3P puis en publiant les fichiers résultants en même temps qu'un fichier d'appel de politique qui désigne les parties du site concernées par la politique. Des outils automatisés peuvent assister les opérateurs de site dans cette traduction. **[26]**

II.1.7. La localisation du fichier de référence

Ce sont des mécanismes utilisés pour indiquer la location du fichier de référence des politiques.

- Dans une location bien connue
- Un document peut indiquer la location en utilisant la balise link du HTML
- Un document peut indiquer la location en utilisant la balise link du XHTML

- Dans l'entête de la réponse HTTP.

Les politiques sont appliquées au niveau de ressource. Une page peut se composer de plusieurs ressources, et chacun peut avoir une politique associée à lui. **[26]**

II.1.8. Les politiques compactes

Les politiques compactes sont des politiques P3P récapitulés qui fournies des conseils à des agents pour que ces agents puissent prendre des décisions vite. Les politiques compactes sont une optimisation de performance dont sa présente n'est pas obligatoire pour les agents comme les serveurs. Un agent d'utilisateur qui ne peut pas obtenir assez d'information à prendre une décision doit obtenir la politique normale.

Dans P3P 1.0, les politiques compactes contiennent des informations politiques concernant seulement les cookies. Le serveur Web doit construire les politiques compactes pour représenter des politiques sur les cookies dans la politique entière. **[26]**

Voici un exemple de politique compacte :

```
compact-policy-field = `CP="` compact-policy `"`
compact-policy = compact-token *(" " compact-token)
compact-token = compact-access |
                compact-disputes |
                compact-remedies |
                compact-non-identifiable |
                compact-purpose |
                compact-recipient |
                compact-retention |
                compact-categories |
                compact-test
```

Figure II.3: Syntaxe de la politique compacte **[26]**

II.1.9. Les Politiques complètes

Une version longue et complète de la politique de confidentialité peut être donnée au format XML, selon une forme (DTD) spécifiée par le W3C. Ce document XML, placé à un endroit « bien connu » de l'arborescence du site peut être consulté automatiquement par le navigateur afin de connaître la politique de confidentialité du site Web préalablement à la navigation.

En complément, cette politique de confidentialité sous forme informatique peut renvoyer vers une page lisible par l'internaute (une page Web). **[26]**

II.1.10. État actuel des choses

Actuellement, cette technologie n'est utilisée que par certains navigateurs et seulement :

- Pour gérer les cookies de manière «intelligente», par exemple en bloquant ceux donnant la possibilité un enregistrement abusif des actions de l'internaute.
- Pour afficher un résumé de cette politique de confidentialité à la demande de l'internaute.

D'autre part, peu de sites Web envoient une politique P3P. Ceci limite toujours la portée de cette technologie.

Etant donné que la plupart des internautes n'est guère susceptible de modifier des paramètres préconfigures sur leur logiciel de navigation, la configuration "par défaut" des choix de l'utilisateur en matière de respect de sa vie privée influera considérablement sur le niveau global de protection de la vie privée en ligne. **[26]**

II.2. XACML

II.2.1. Présentation

Est l'un des langages les plus complets qui est une extension de XML et a été conçu principalement pour permettre la restriction d'accès.

XACML (eXtensible Access Control Markup Language) est d'abord un langage basé sur XML dédié au contrôle d'accès. Il s'agit à la fois d'un langage de politique de contrôle d'accès basé sur les attributs et d'un langage protocolaire de type requêtes/réponses. De plus, la spécification fournit une architecture qui définit différentes entités impliquées dans le processus de prise de décision d'une autorisation d'accès.

Le langage de politique XACML est utilisé pour décrire les exigences générales de contrôle d'accès en termes de contraintes sur des attributs comme il est montré dans la figure II.5. Un attribut peut être n'importe qu'elle caractéristiques d'un sujet, d'une action, d'une ressource ou de l'environnement dans lequel la requête d'accès est produite. Le fait de considérer les attributs rend le langage très flexible.

De plus, XACML présente des points d'extension standards pour définir de nouveaux types de données, des fonctions additionnelles, des combinaisons de logiques, etc. **[46]**

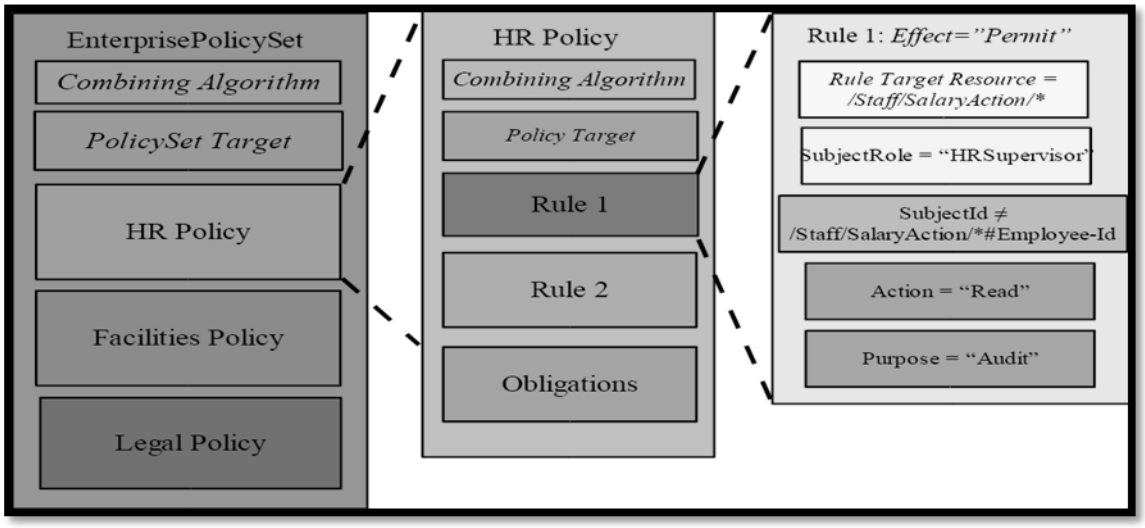

Figure II.4 : Exemple de politique en XACML **[47]**

II.2.2. Architecture de XACML

Après Une demande d'autorisation au *point de politique d'application* (PEP). Le PEP crée une demande XACML et il l'envoie au point de politique de décision (PDP), qui évalue la demande et renvoie une réponse. La réponse peut être autorisé ou refusé l'accès, avec les obligations appropriées.

Le PDP arrive à une décision après avoir évalué les politiques et les règles en leur sein. Un certain nombre de politiques peuvent être disponibles: Le PDP n'évalue pas tous, seulement ceux qui sont pertinents sont choisis pour l'évaluation, fondée sur l'objectif de la politique. L'objectif de la politique contient des informations sur le sujet, l'action, et d'autres propriétés de l'environnement.

Pour arriver à des politiques, le PDP utilise la politique Access Point (PAP), qui écrit des politiques et établit les politiques et les rend disponibles pour le PDP. Le PDP peut également invoquer le Point de politique d'information (PIP) de service pour récupérer les valeurs d'attribut liées à l'objet, la ressource, ou l'environnement. La décision d'autorisation est parvenu le PDP est envoyé à la PEP. Le PEP remplit les obligations et, en fonction de la décision d'autorisation adressée par PDP, soit autorise ou refuse l'accès. **[48]**

La figure suivante démontre ce qu'on vient de dire :

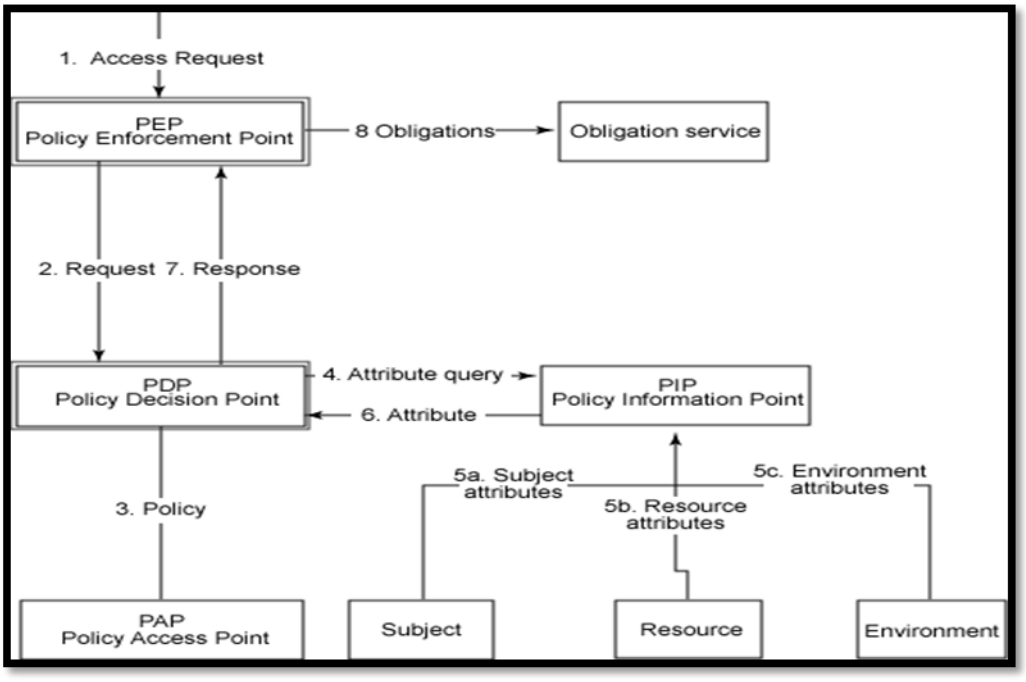

Figure II.5: Les composantes de XACML **[48]**

II.3. EPAL (Enterprise Privacy Authorization Language)

En 2002, IBM[16] publié un modèle pour la formalisation des politiques de la vie privée aux entreprises. Le modèle identifie six éléments nécessaires d'une politique de la vie privée. Basé sur ce travail, une plateforme a été proposée pour le cadre des pratiques de la vie privée d'entreprise E-p3p, qui permet à des entreprises d'imposer automatiquement des pratiques en matière de la vie privée. En conclusion, E-p3p a servi au développement d'un " langage d'autorisation de la vie privée d'entreprise" EPAL en tant qu'élément de la solution de gestion de la vie privée de l'entreprise d'IBM.

EPAL offre un langage formel comme XACML, ressemble à une solution de contrôle d'accès. Le cadre est établi sur le même modèle d'application de politique appliqué dans XACML. De même, EPAL fournit également une définition spécifique des attributs.
EPAL représente une liste de règles de priorité différente. Les règles définissent également les conditions pour considérer le contexte externe de l'information.

Nous montrons dans la figure II.7 un exemple simplifié d'une règle d'EPAL, l'exemple correspond à une politique de la vie privée qui permet de rassembler les informations des contacts pour les buts de recherches et pour se conformer aux restrictions légales strictes.

```
<rule>
  <ruling>ALLOW</ruling>
  <userCategory>organization name</userCategory>
  <action>DISCLOSE</action>
  <dataCategory>ContactInformation </dataCategory>
  <purpose>ResearchPurpose </purpose>
  <condition>Disclosure subject to strict legal restrictions </condition>
  <obligation/>
</rule>
```

Figure II.6: Exemple d'une règle EPAL **[49]**

[16] Société multinationale américaine présente dans les domaines du matériel informatique, du logiciel et des services informatiques.

Le langage EPAL permet à une organisation d'exprimer un ensemble de règles (une politique de confidentialité) relatives à l'habilitation pour l'aspect confidentialité. Autorisations et interdictions doivent contrôler l'utilisation des ressources informationnelles dans une organisation. Le modèle demande de dresser des listes hiérarchiques de catégories de données, de catégories d'utilisateurs, et des finalités de cueillette, ainsi que des ensembles d'actions, d'obligations et de conditions, toutes axées sur la confidentialité. Les actions disent comment les données sont utilisées, les obligations disent quelles actions doivent être effectuées, les conditions évaluent des éléments du contexte.

Une politique EPAL catégorise les données détenues par l'entreprise et établit des règles pour chaque catégorie. Dans la politique, l'ordre des règles est important car un principe de précédent listé (descending precedence) s'applique (rendant inopérantes les règles suivantes dans une liste). Comme une règle XACML, une règle EPAL comporte un sujet, une action et une ressource avec une indication de permission ou d'interdiction, et elle comporte en plus une mention de finalité. Une règle peut aussi contenir des conditions et des obligations.

Comme XACML, EPAL est conçu comme solution de contrôle d'accès et, par conséquent permet l'application automatique des politiques. **[49]**

III. Langages de préférences en termes de vie privée

Après une analyse des langages de politique, cette section discute les langages d'expression des préférences en termes de vie privée.
En plus d'APPEL, langage d'expression de préférence compatible avec P3P, cette section analyse XPREF et présente un langage de préférence sémantique le Rei.

III.1. APPEL (A P3P Preference Exchange Language)

APPEL (A P3P Preference Exchange Language) est un langage proche de P3P défini par le w3c. Il vise à fournir un moyen aux internautes de décrire leurs préférences personnelles en matière d'utilisation de leurs données. Ces préférences sont décrites en langage XML. Evidemment, on ne peut pas demander à un utilisateur basique d'écrire un fichier XML contenant des connecteurs logiques, des expressions régulières et des balises et attributs bien définis. Donc, il est possible d'importer des préférences par

défaut répondant aux attentes de la plupart des utilisateurs. Mais rien n'empêche un utilisateur de définir son propre fichier APPEL.

Le fichier de préférences APPEL contient un ensemble de règles (balise RULE) regroupées dans une même balise RULESET. Une règle est caractérisée par un comportement (attribut behavior) à adopter en cas de succès. Le w3c propose 3 comportements:

– **request** : la politique du site est acceptable

– **limited :** l'accès à la ressource devrait être limité

– **block :** l'accès à la ressource ne devrait pas être autorisé

Voici un extrait d'un fichier APPEL (figure qui indique que l'utilisateur accepte les cookies qui sont déposés dans le but d'adapter le contenu de la page (tailoring) et dont le destinataire sera uniquement le site lui-même (ours). La catégorie state signifie que les cookies permettent de gérer des états dans un protocole qui est sans état (http) :

```
<appel :RULE behavior="request">
      <p3p :POLICY>
            <p3p :STATEMENT>
                  <p3p :RECIPIENT appel :connective="and">
                        <p3p :ours/>
                  </p3p :RECIPIENT>
                  <p3p :PURPOSE appel :connective="non-and">
                        <p3p :tailoring/>
                  </p3p :PURPOSE>
                  <p3p :DATA-GROUP>
                        <p3p :DATA ref="#dynamic.cookies">
                              <p3p :CATEGORIES appel :connective="or">
                                    <state/>
                              </p3p :CATEGORIES>
                        </p3p :DATA>
                  </p3p :DATA-GROUP>
            </p3p :STATEMENT>
      </p3p :POLICY>
</appel :RULE>
```

Figure II.7: Exemple d'un fichier APPEL **[45]**

Cet exemple montre que le langage APPEL utilise les mêmes balises que celles utilisées dans le langage P3P. La structure du fichier des préférences utilisateur est très proche de celle d'une politique P3P, ce qui facilite la comparaison de ces deux fichiers. **[45]**

Voir **Annexe C** pour plus de détaille.

III.2. XPref (XPath-based preference language)

Le langage de préférence XPref utilise le langage d'interrogation XPath[17] de XML. Xpath définit l'adressage normalisé des parties d'un document de XML et semble un choix valable pour l'évaluation des politiques P3P basées sur XML.

Compatible à P3P, XPref offre une alternative intéressante pour APPEL. Xpref réutilise plusieurs éléments d'APPEL, tels que RULESET, RULE et behavior. Représentant le noyau de XPref.

Les conditions d'une règle sont exprimées par des expressions de XPath, contenues par l'état d'attribut. La figure suivant montre un exemple d'une règle XPref:

```
<RULESET>
  <RULE behavior="block"
    condition ="/POLICY/STATEMENT/PURPOSE/*
        [(name(.) = "contact" or
          name(.) = "telemarketing")]" />
</RULESET>
```

Figure II.8: Exemple d'une règle XPref **[49]**

L'attribut "condition" commence par un chemin qui adresse tous les nœuds enfants d'un nœud supérieur "PURPOSE" d'une politique p3p.

La condition d'une règle de XPref est satisfaite, si au moins un de ces buts s'assortit. Les opérateurs additionnels de Xpath facilitent l'évaluation des rapports multiples d'une politique de P3P.

À la différence d'APPEL, les caractéristiques de XPref permettent à des utilisateurs de définir des conditions acceptables et inacceptables, qui contribuent à l'expressivité du langage. Mais XPref ne garantit pas l'uniformité sémantique des ensembles de règle. Car Xpref suit la même orientation de syntaxe qu'APPEL, les mêmes contradictions sont susceptibles de se produire pendant l'évaluation des règles. **[49]**

[17] Un langage d'expression pour la sélection, le tri et la comparaison des données dans les documents XML, souvent utilisé avec XSLT pour la transformation et le mapping des données.

III.3. Rei

Adressant l'expressivité limitée et la contradiction sémantique des langages orientés par syntaxe (APPEL, XPref,…), un langage de politique flexible Rei est apparait.
Rei est appliqué dans le contexte des préférences d'utilisateur. Une ontologie spécifique de domaine fournit les classes et les propriétés appropriées pour la définition des préférences de la vie privée. Beaucoup d'éléments d'ontologie correspondent aux éléments d'APPEL. Les conditions préalables additionnelles permettent le filtrage des règles, avant que les conditions soient évaluées.
Comme décrit dans les langages de préférence, les conditions permettent la référence aux éléments d'une politique de P3P. Pour cela, des rapports des ontologies additionnels peuvent être employés, qui facilitent l'expression des pratiques de la vie privée.
En plus le langage permet la définition des priorités d'une règle, qui aident des utilisateurs en déterminant le comportement d'évaluation. Rei fournit également des éléments pour les spécifications des engagements qui sont liés aux actions accordées.
Comparé aux langages orientés par syntaxe, Rei qui se base sur les ontologies offre une variété maximum de préférences de la vie privée. **[49]**

IV. Quelques outils existants

IV.1. PrivacyBird

L'opérateur AT&T[18] propose un outil baptisé PrivacyBird[19] (l'oiseau de la vie privée) pour connaître d'un coup d'œil, la politique concernant les données personnelles, du site Web visité.
Disponible en version bêta, PrivacyBird s'installe comme une extension à Internet Explorer. Il repose sur la norme P3P (Platform for PrivacyPreferences) qui permet à un site Web d'inclure sous forme de code XML (non visible par l'internaute) sa politique de gestion des données personnelles.
A chaque arrivée sur un site Web, l'oiseau analyse les informations fournies par le site au format P3P et les compare aux préférences fixées par l'internaute. Un utilisateur

[18] Le leader Américain des télécommunications de tous types : vocale, vidéo, données et Internet pour les particuliers et les entreprises.
[19] Pour plus d'information, visitez le site: **www.privacybird.org/.**

peut, par exemple, choisir d'autoriser la collecte et la revente de ses données personnelles.

Selon le niveau de sécurité, PrivacyBird sonne l'alerte : de couleur verte, et accompagné d'un gazouillis pour les sites conformes en tout point au souhait de l'internaute, jusqu'au rouge et au croassement pour les sites prenant trop de liberté avec ses données personnelles.

PrivacyBird va au-delà du signal sonore et visuel. Un menu déroulant permet d'accéder à une version résumée des règles appliquées par le site visité, ainsi qu'un lien direct vers la page d'inscription/désinscription à d'éventuelles newsletters.

Si le site n'a pas traduit au format P3P sa politique sur les données personnelles, le signal passe au jaune, et le chant est moins " enjoué ". C'est d'ailleurs le cas de la majorité des sites aujourd'hui. En effet, le P3P est aujourd'hui très peu utilisé ce qui limite l'intérêt de PrivacyBird.

Enfin, l'oiseau ne fait que signaler les données récoltées par le site, sans intervenir. C'est à l'internaute de choisir s'il accepte de surfer sur un site non respectueux de sa vie privée. **[50]**

IV.2. Microsoft Internet Explorer

Le navigateur web Microsoft Internet Explorer inclut la gestion des cookies qui permet aux utilisateurs de spécifier des règles du cookie-blocage qui sont basés sur les politiques compactes P3P.

Internet Explorer est préconfiguré avec six paramètres : bloquez tous les cookies, autorisez tous les cookies, hauts, moyen-haut, moyens, et bas. Comme le montre la figure II.9, les utilisateurs peuvent sélectionnez leur cookie et voir une description courte de chaque cadre. Internet explorer est configuré par défaut au niveau moyen. **[51]**

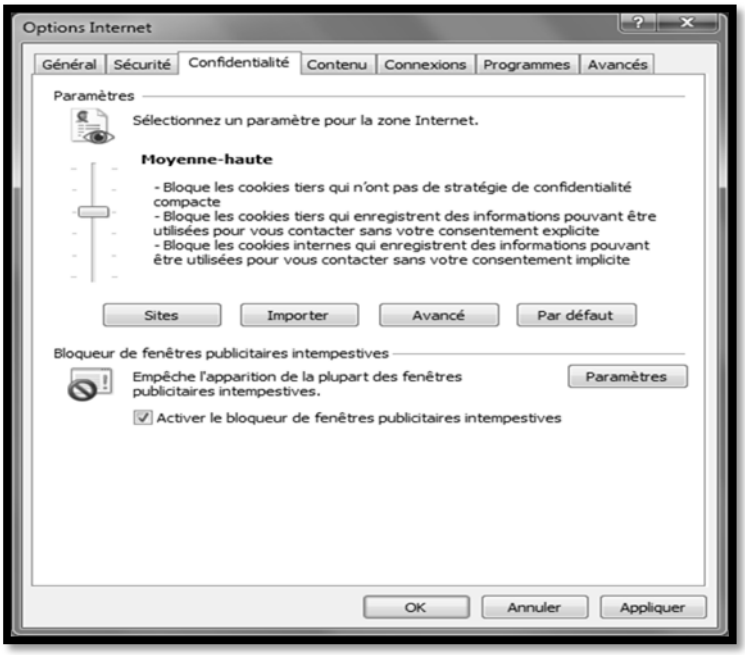

Figure II.9 : Forme de configuration d'internet explorer **[51]**

IV.3. Netscape Navigator

Netscape Navigator 7 inclut la gestion des cookies qui permet aux utilisateurs de spécifier les règles du cookie-blocage qui sont basé sur les politiques compactes de P3P qui sont semblable à ceux trouvé dans IE6.

L'interface de la spécification des préférences des utilisateurs du Netscape montré dans la figure II.10, utilise un langage similaire à celui utilisé par l'interface d'IE6. **[51]**

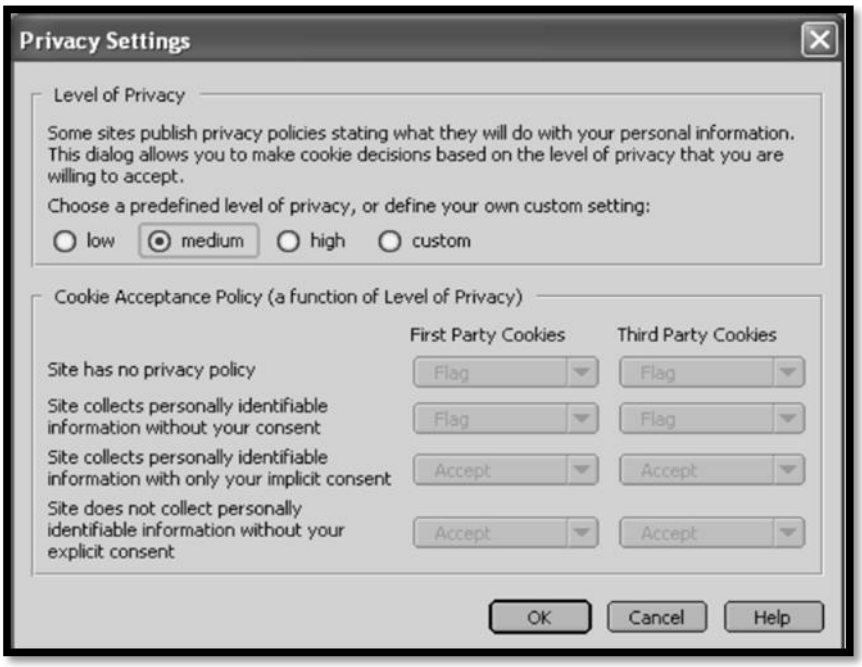

Figure II.10 : Interface de préférences de Netscape **[51]**

V. Conclusion

Bien que les langages de politiques de la vie privée fournissent un moyen technique permettant aux utilisateurs d'être informés des politiques de confidentialité avant de confier des renseignements personnels, ils n'offrent aucun mécanisme technique qui garantisse le comportement des sites conformément à leurs politiques. Les produits qui mettent en œuvre cette spécification peuvent fournir une aide dans ce sens, selon les mises en œuvre en question, mais cela n'est pas traité par cette spécification. Toutefois, le protocole P3P complète les lois et les programmes auto-règlementés définissant des mécanismes d'application. En outre, le protocole P3P ne comprend aucun mécanisme de transport des données ou de sécurisation des données personnelles en transit ou stockées. On peut intégrer P3P à des outils conçus pour faciliter le transport des données. Ces outils devraient inclure les sécurités appropriées.

Chapitre III:

Conception et implémentation

I. Introduction

Le présent chapitre décrit les détails de conception et d'implémentation du système développé.

En premier lieu, on va présenter l'architecture de notre application pour avoir une vision globale sur son but et son fonctionnement. En suite, on va décrire les tests d'utilisation, les outils de développement. Et enfin quelques expérimentations.

II. Architecture et fonctionnement

II.1. Présentation

Le but de notre application est la détection de toute utilisation non autorisée des données personnelles d'un utilisateur lors de son usage d'un site Web. Comme représente la figure III.1, l'application est conçue pour être intégrée dans un navigateur Web (les éléments 1 et 2 de la figure).

Figure III.1 : Navigateur

Après que l'utilisateur saisit l'adresse d'un site Web, et en pressant le bouton (1) l'application peut indiquer si le présent site respecte ou non ses préférences.

Un utilisateur peut définir ses préférences en terme de confidentialité[20]a traves une interface graphique accessible en cliquant sur le bouton (2) . Les détails seront présentés dans la partie description de la section III.

II.2. Architecture

Comme il est présenté dans la figure III.2 Notre système est composé de quatre modules principaux *: le module de localisation et de téléchargement de la politique de confidentialité (MLTP), le module de définition des règles de préférences (MDRP), le module de comparaison (MC)* et enfin *le module de gestion de comportement (MGC).*

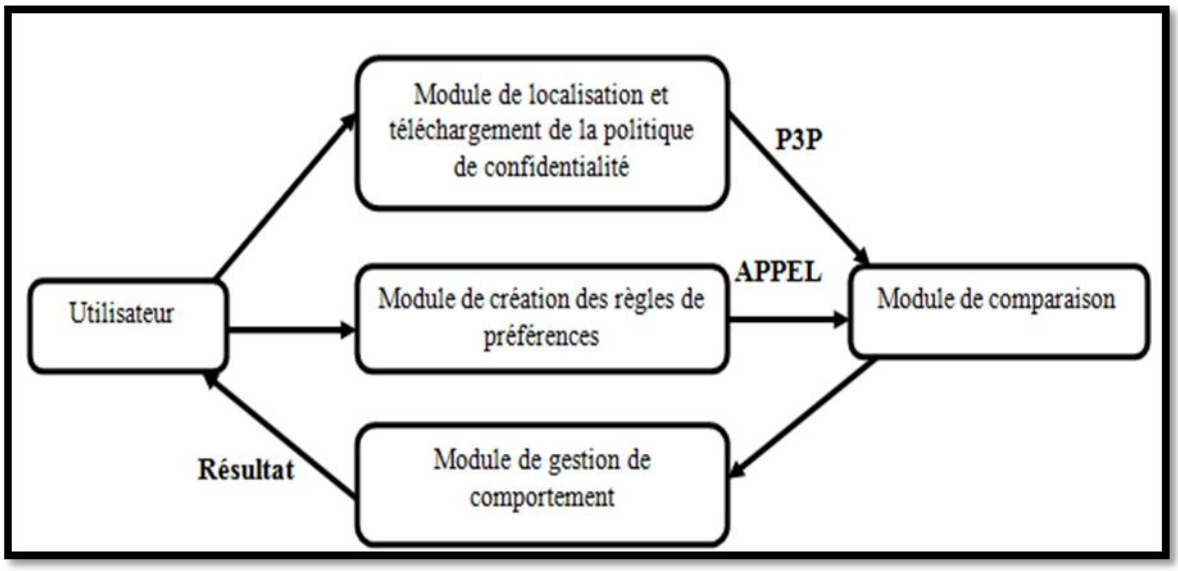

Figure III.2 : Architecture du système

1. ***le module de localisation et de téléchargement de la politique*** **(MLTP)** : son rôle est la localisation et le téléchargement du fichier P3P qui contient la politique de confidentialité du site visité par l'utilisateur. Ce fichier est envoyé par la suite au module de comparaison (MC).
2. ***Le module de définition des règles de préférences (MDRP):*** ce module fournit à l'utilisateur une interface graphique qui lui permet de définir ses préférences en termes de confidentialité (figure III.3). les choix de l'utilisateur sont transformés sous fourme de règles en utilisant le langage APPEL. Le fichier APPEL généré est envoyé par la suite au module de comparaison (MC).

[20] Le terme confidentialité possède le même sens que la vie privée.

Un fichier APPEL possède plusieurs règles, chacune d'elles:

- Décrit les usages prévus des données.
- Exprime une ou plusieurs restrictions concernant la collecte ou l'utilisation des données.
- Définit l'entité légale ou le domaine où les données peuvent être distribuées.

Il est bien de noté qu'à travers ce module l'utilisateur peut à tout moment modifier ses préférences de confidentialité.

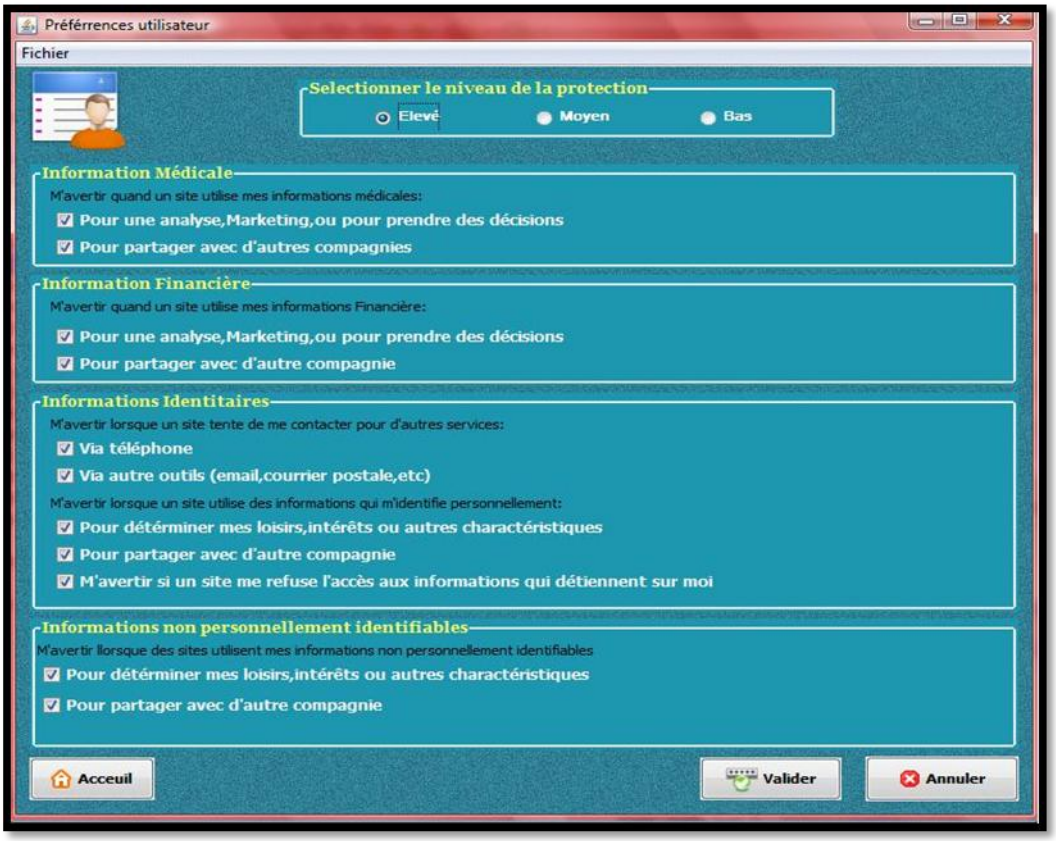

Figure III.3 : Interface de définition et de Modification des préférences de confidentialité.

3. ***le module de comparaison (MC)*** *:* ce module utilise le fichier P3P téléchargé par le *MLTP*, et le fichier APPEL généré par le *MDRP* pour faire une opération de mise en correspondance, cette opération a comme résultat l'affirmation de l'existence ou de l'absence de conformité entre la politique du site et les préférences de l'utilisateur. Le résultat de cette étape est transmit au module de gestion de comportement (*MGC*) pour déclencher l'action adéquate.

4. ***module de gestion de comportement (MGC)*** : le rôle de ce module est le déclanchement d'une action selon les résultats fournis par le MC, une action peut être un simple message textuel, visuel ou sonore pour informer l'utilisateur des résultats fournis par le MC. Les différents types d'action généré par notre application son décrite en détaille dans la partie description de la section III.

La comparaison se fait en appliquant l'algorithme suivant sur chaque élément d'une règle APPEL :

Deux expression P (d'un fichier P3P) et A (d'un fichier APPEL) se correspondent si et seulement si :

1. Les noms des deux éléments sont identique (par exemple <STATEMENT>, <POLICY>) ;
2. Touts les attributs de l'expression « A » se correspondent avec les attributs de l'expression « P ». si « P » contient des attributs non existants dans « A », alors ces attributs sont ignorés.
3. Si l'expression « A » contient un connecteur « ***OR*** », alors au moins un des expressions contenues dans « A » doit se correspondre avec l'un des expressions contenues dans « P » en appliquant les deux règles 1 et 2. si d'autres éléments existent dans P qui sont pas référencés dans « A » alors ces éléments sont ignorés.
4. Si l'expression « A » contient un connecteur « ***and*** », alors toutes les expressions contenues dans « A » doivent se correspondre avec les expressions similaires contenues dans « P » en appliquant les deux règles 1 et 2. si d'autres éléments existent dans « P » et ne sont pas référencés dans « A » alors ces éléments sont ignorés.
5. Si l'expression « A » contient un connecteur « ***or-exact***» alors au moins l'un des expressions contenues dans A doit se correspondre avec l'un des expressions contenues dans « P » en appliquant les deux règles 1 et 2. si d'autres éléments existent dans « P » et ne sont pas référencés dans « A » alors il n y a pas une correspondance entre « A » et « P ».
6. Si l'expression « A » contient un connecteur « ***and-exact***» toutes les expressions contenues dans « A » doivent se correspondre avec les expressions similaires contenues dans « P » en appliquant les deux règles 1 et 2. si d'autres

éléments existent dans « P » et ne sont pas référencés dans « A » alors il n y a pas une correspondance entre « A » et « P ».

7. Si A ne contient pas de connecteur alors on applique la règle 6.

II.3 Fonctionnement

Le diagramme de séquence suivant représente l'ordre chronologique des interactions entre les différents modules de notre système,dans le cas ou un site posséde une politique de confidentialité.

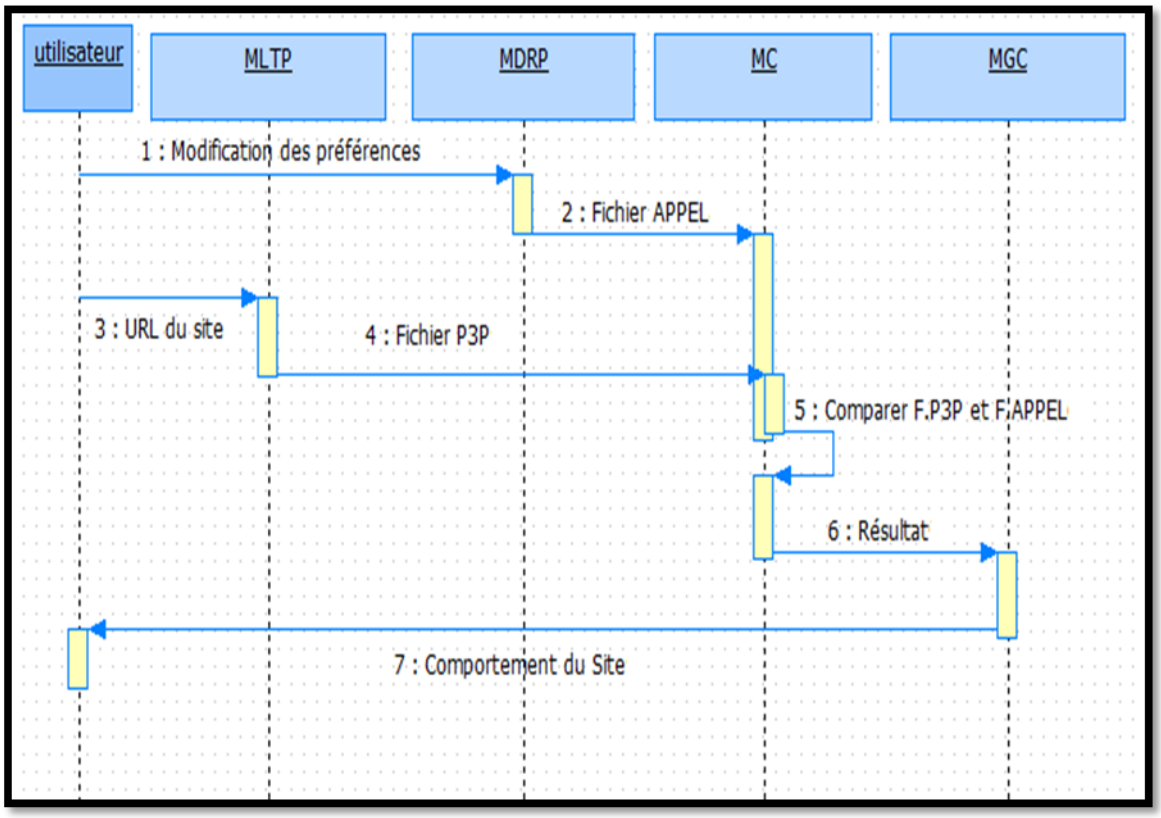

Figure III.4 : Diagramme de séquence du système.

Dans la première étape l'utilisateur doit définir ses préférences de confidentialité au niveau du « MDRP », ce dernier génère un fichier APPEL et le transmit au « MC ».

La deuxième étape est déclenchée lorsque l'utilisateur veut naviguer sur un site Web. Le module « MLTP » récupère l'adresse du site saisie par l'utilisateur, et essaie de localiser le fichier P3P qui contient la politique de confidentialité. Si le site définie une tel politique le « MLTP » la télécharge et la transmit au « MC ». Le « MC » fait une comparaison entre le fichier P3P et le fichier APPEL précédemment généré. Les

résultats de cette comparaison sont fournis au « MGC **»,** qui a son tour prévient l'utilisateur si le site respecte ou non ses préférences de confidentialité.

III. Description

Cette section présente une description des différentes facettes et fonctionnalités de notre application.

Après qu'un utilisateur saisit une adresse d'un site Web, en cliquant sur le bouton 1 de la figure III.1 (section II.1), notre application et à traves le module « MGC » retourne le résultat de mise en correspondance entre la politique du site et les préférences de l'utilisateur. On peut faire face à l'un des cas suivants:

- **Cas n°1**

Le site n'a pas de politique P3P ou notre système ne trouve pas la politique à l'emplacement prévu, l'application retourne un message textuel « impossible de récupérer la politique du site » et une icône jaune (voir la figure III.5).

Figure III.5 : politique du site introuvable

Cas n°2

La politique de sécurité du site web concorde avec les préférences d'utilisateur. L'application retourne le message montré dans la figure III.6 avec une icône verte.

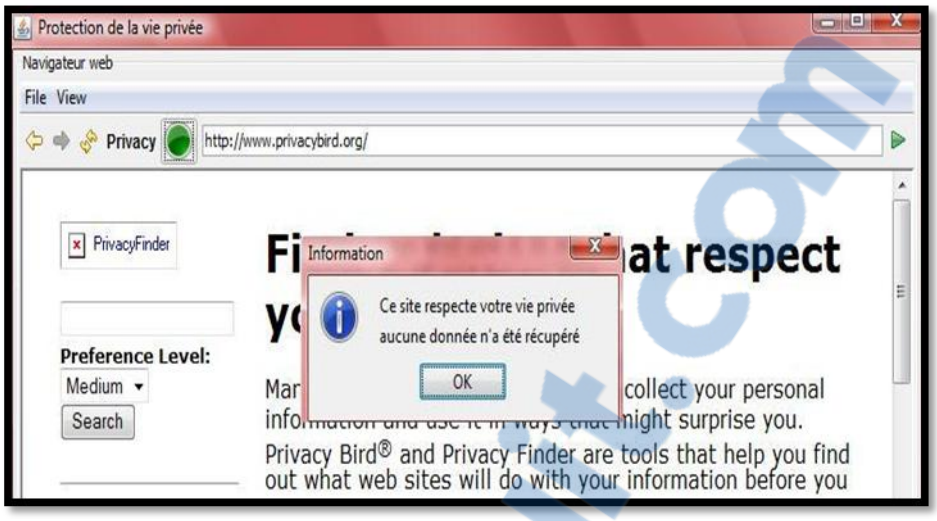

Figure III.6: Matching réussie

Cas n°3

La politique du site ne respect pas les préférences de l'utilisateur. L'application retourne le message montré dans la figure III.7 et une Icône rouge.

Figure III.7: Matching non réussie

Pour la configuration des différent paramètres de confidentialité, on clique sur le bouton (2) (figure III.1), notre application affiche une interface d'accueil (voir la figure III.8). Cette interface fournit à l'utilisateur l'accès à plusieurs fonctionnalités parmi lesquelles la définition et la modification de ses préférences (en cliquant sur le bouton (6) de la figure III.8).

Figure III.8 : Accueil de l'application

Dans le cas d'échec de la comparaison, l'utilisateur peut avoir plus de détail sur la cause du conflit en pressant le bouton (3), la figure III.9 montre un exemple.

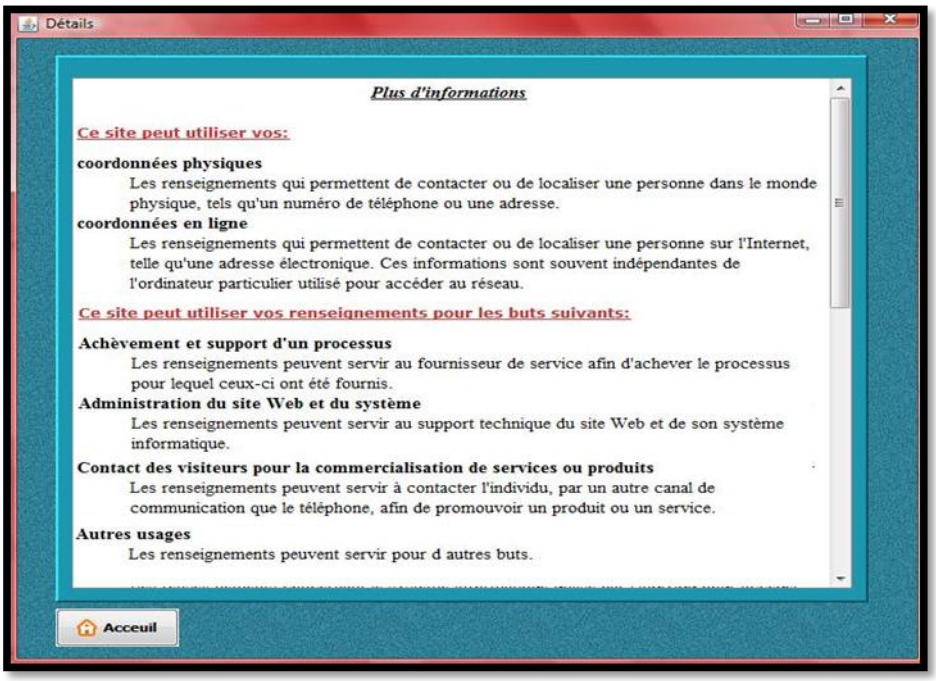

Figure III.9 : Plus d'informations sur la cause du conflit

En cliquant sur le bouton (4), on obtient plus d'informations sur la politique du site exprimé en langage naturel (voir figure III.10).

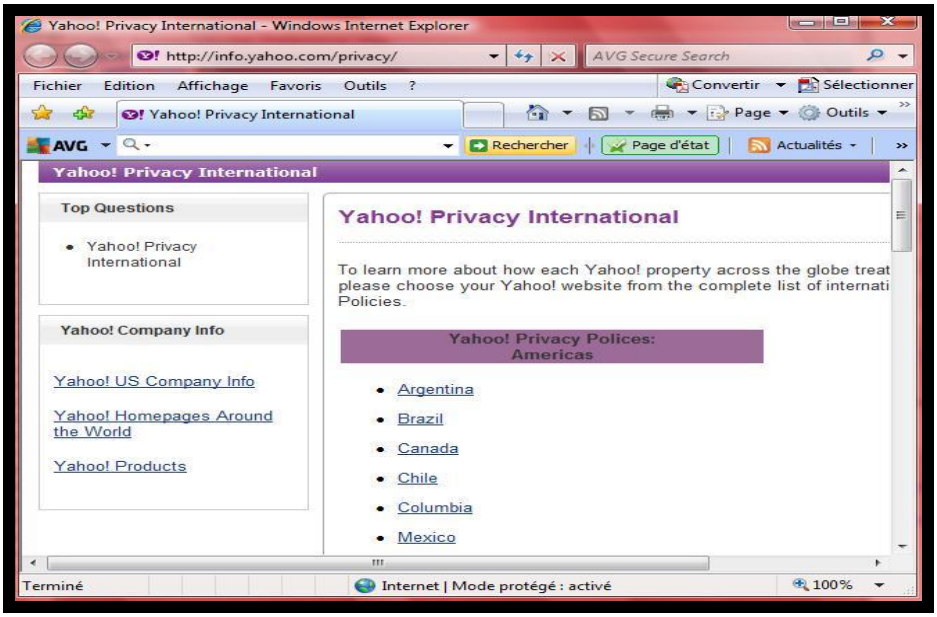

Figure III.10 : La politique complète du site

Comme on peut aussi voir la politique du site écrite en P3P en cliquant sur le bouton (5) la figure III.11, présente un exemple.

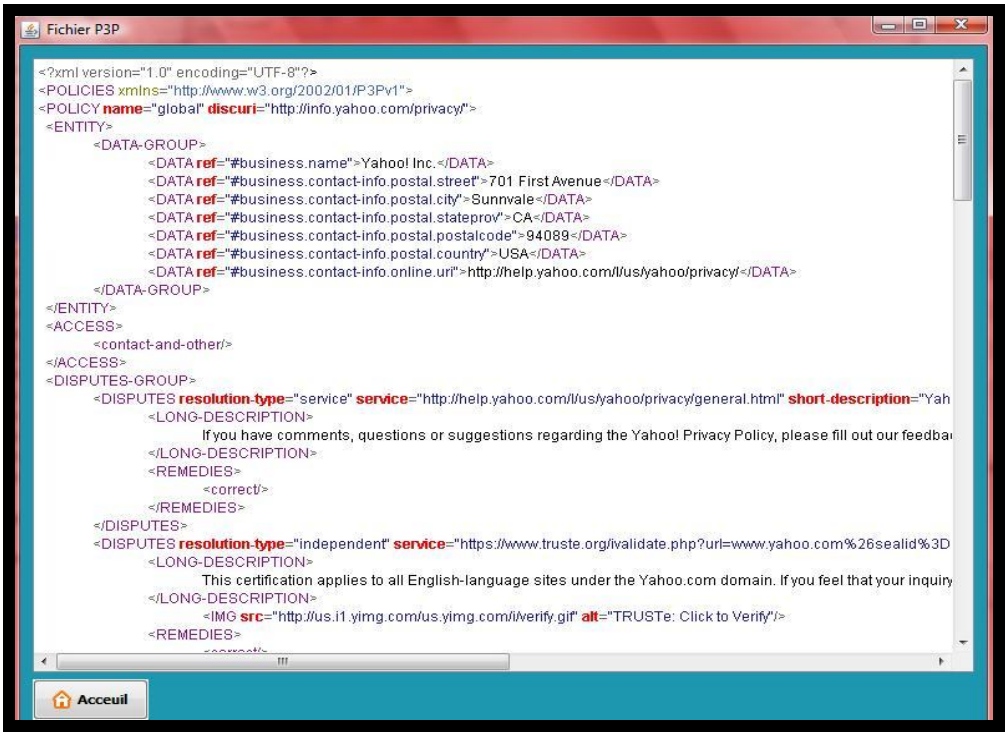

Figure III.11 : Exemple de fichier P3P.

IV. Les outils de développement

Le développement de notre système a été fait avec le langage JAVA sous l'environnement Netbeans. Plusieurs API sont utilisées dont en cite : DOM, XPath et SAX. Le présent paragraphe donne un bref aperçu de ces outils.

- L'IDE Netbeans

On a choisie l'IDE Netbeans pour réaliser notre prototype et cela pour sa simplicité et sa richesse en termes de bibliothèques.

- API DOM

DOM est l'acronyme de Document Object Model permet de modéliser, de parcourir et de manipuler un document XML. Le principal rôle de DOM est de fournir une représentation mémoire d'un document XML sous la forme d'un arbre d'objets et d'en permettre la manipulation (parcours, recherche et mise à jour).

- XPath

XPath permet de parcourir un fichier XML d'une façon à la fois simple et puissante. De la sorte, en peu de temps, un développeur peut rapidement et aisément extraire les informations qui l'intéressent.

- API SAX

Simple **A**PI for **X**ML ou SAX est une API générale pour la lecture d'un flux XML. Ce type de parseur utilise des événements pour piloter le traitement d'un fichier XML.

V. Expérimentations

Le tableau suivant représente le résultat de comparaison de notre application avec l'outil Priavcy Bird.

	L'@ du site	**Résultat de PrivacyBird**	**Résultat de notre application**
1	**http://www.google.com/**	No Privacy Policy was found	Impossible de récupérer la politique du site
2	**http://www.yahoo.com/**	Yahoo! Inc. may use some collected information	Ce site ne respect pas votre vie privée
3	**http://www.privacybird.org/**	CMU Usable Privacy and Security Lab's privacy policy matches your preferences	Ce site respect votre vie privée, aucune donnée n'a été collecté
4	**http://www.nature.com/**	Nature America's privacy policy does not match your preferences	Ce site ne respect pas votre vie privée
5	**http://www.drugs.com/**	Drugsite Trust's privacy policy does not match your preferences	Ce site ne respect pas votre vie privée
6	**http://www.bird.com/**	Privacy policy has an error in its P3P policy	Ce site ne respect pas votre vie privée
7	**http://www.democracyclub.org.uk/**	Democracy Club's privacy policy matches your preferences.	Ce site respect votre vie privée, aucune donnée n'a été collecté
8	**http://checky.mozdev.org/**	checky.mozdev.org's privacy policy does not match your preferences	Ce site respect votre vie privée, les données ont été rendus anonymes
9	**http://www.charityusa.com/**	CharityUSA, LLC.'s privacy policy does not match your preferences	Ce site respect votre vie privée, aucune donnée n'a été collecté
10	**http://www.usatoday.com/**	USATODAY.com's privacy policy does not match your preferences	Ce site ne respect pas votre vie privée
11	**http://www.ftc.gov/**	Federal Trade Commission's privacy policy matches your preferences	Ce site respect votre vie privée, aucune donnée n'a été collecté
12	**http://www.addthis.com/**	LLC's privacy policy matches your preferences.	Ce site respect votre vie privée, aucune donnée n'a été collecté

13	**https://www.toolbarn.com/**	Privacy policy has an error in its P3P policy	Ce site respect votre vie privée, aucune donnée n'a été collecté
14	**http://www.developper.com/**	No Privacy Policy was found	Impossible de récupérer la politique du site
15	**http://www.microsoft.com/1**	Microsoft Corporation's privacy policy does not match your preferences	Ce site ne respect pas votre vie privée
16	**http://www.att.com/**	AT&T's privacy policy does not match your preferences	Ce site ne respect pas votre vie privée
17	**http://www.latimes.com/**	latimes.com's privacy policy does not match your preferences	Ce site ne respect pas votre vie privée
18	**http://www.ebags.com/**	eBags Inc.'s privacy policy does not match your preferences	Ce site ne respect pas votre vie privée
19	**http://ninemsn.com.au/**	ninemsn Pty Ltd's privacy policy does not match your preferences	Ce site ne respect pas votre vie privée
20	**http://www.sky.fm/**	Privacy policy has an error in its P3P policy	Ce site respect votre vie privée, aucune donnée n'a été collecté

Tableau III.1 : Etude comparative

D'après ce tableau les résultats retournés sont presque identiques il y a une différence seulement pour les deux cas (8 et), cela nous confirme la crédibilité des résultats fourni par notre application puisque elle est comparée avec l'outil le plus complet connu dans ce domaine.

Le point fort de notre application par rapport à Privacy Bird c'est que les résultats sont fournis à l'utilisateur dans un délai plus court que Privacy Bird.

VI. Conclusion

Dans ce chapitre nous avons présenté la partie pratique de notre travail de ce fait nous avons présenté un prototype dont le but est la protection de la vie privée des utilisateurs web, ce prototype permet la vérification des politiques de confidentialité relatives aux informations recueillies lors d'une navigation.

L'un des avantages de notre application est qu'elle présente une interface utilisateur conviviale qui n'impose aucune connaissance préalable de la syntaxe de politique de confidentialité de la part des utilisateurs.

La comparaison de notre application avec Privacy Bird a montré son efficacité et sa fiabilité.

Conclusion générale

Conclusion générale

Le travail présenté dans ce mémoire s'inscrit dans le contexte de la protection de la vie privée sur Internet. Nous avons donné une vue générale sur ce domaine en introduisant la notion de la vie privée sur Internet, les risques relatifs ainsi qu'un état de l'art sur les solutions existantes. Nous avons abordé en particulier les systèmes qui se basent sur des langages d'expression de politique de vie privée. De ce fait nous avons développé une application permettant d'avertir l'utilisateur contre toute utilisation non autorisée de ses données personnelles. En manipulant la politique de confidentialité des sites Web exprimée en P3P, notre application peut détecter la non conformité entre l'utilisation des données et les préférences des utilisateurs, ces préférences sont exprimées sous forme de règles à l'aide du langage APPEL.

En comparant l'application développée avec « privacy bird », l'outil le plus complet connu dans ce domaine, notre application à démontré son efficacité et son fiabilité. Malgré ces avantages, bon nombre d'améliorations sont à envisager, cela nous amène à prévoir les perspectives suivantes :

- L'intégration de notre application comme plugin dans les navigateurs les plus utilisés comme Internet Explorer et Firefox en utilisant d'autres langages de développement tel que XUL[21].
- Améliorer l'utilité de notre application en introduisant un processus de négociation entre les sites Web et les utilisateurs ce qui permet une utilisation plus flexible de ces sites.
- L'intégration du système développé avec des mécanismes de contrôle vu que P3P et APPEL ne permettent pas aux l'utilisateur de contrôler leurs informations une fois divulguées.
- Introduire une couche sémantique au niveau du module de comparaison en utilisant des ontologies, ce la permet de faire des comparaisons sémantiques et non seulement syntaxiques, ce qui augmente pleinment l'efficacité de notre système et le rend indépendant d'un langage d'expression de politiques particulier.

[21] XUL est un langage descriptif qui permet le développement d'interfaces graphiques fondé sur XML. Il a été conçu par Mozilla pour l'ensemble de ses produits et notamment Firefox et Thunderbird.

Références bibliographiques

Références bibliographiques

[1] « Wikipedia, the free encyclopedia », consulté le 02/02/2011, disponible sur: http://fr.wikipedia.org/wiki/Donn%C3%A9es_personnelles

[2] « Correspondant CNIL (CIL) - Au service des Correspondants Informatique et Libertés » consulté le 11/02/2011, disponible sur: http://www.cabinetcilex.com/index.php?mact=News,cntnt01,detail,0&cntnt01articleid=7&cntnt01returnid=49

[3] François GRANGER, « SansFiltre », Mars 2010, consulté le 11/02/2011, disponible sur:: http://www.fgranger.com/dotclear/index.php/post/2010/02/13/Vie-priv%C3%A9e,-d%C3%A9finition

[4] « Wikipedia, the free encyclopedia » consulté le 11/11/2010, disponible sur: http://en.wikipedia.org/wiki/Internet_privacy

[5] « Office of the Privacy Commissioner of Canada Commissariat à la protection de la vie privée du Canada » : http://www.privcom.gc.ca/

[6] « Tenshy's IT Blog » consulté le 02/03/2011, disponible sur: http://tenshyit.wordpress.com/2010/06/23/moteurs-de-recherche-et-vie-privee/

[7] «Privacy Rights Clearinghouse-Empowering Consumers. Protecting Privacy- », consulté le 22/02/2011, disponible sur: http://www.privacyrights.org/fs/fs18-cyb.htm

[8] «Ixquick Web Recherche», Le moteur de recherche le plus confidentiel au monde, consulté le 11/03/2011, disponible sur: http://eu.ixquick.com/fra/protect-privacy.html

[9] « Commissariat à la protection de la vie privée du Canada », consulté le 02/02/2011, disponible sur: http://www.priv.gc.ca/speech/2008/sp-d_080909_f.cfm

[10] « Commissariat à la protection de la vie privée du Canada », consulté le 15/11/2010, disponible sur:http://www.priv.gc.ca/speech/2010/sp-d_20100520_dc_f.cfm

[11] « Université Evry Val d'Essonne » : http://leda.univ-evry.fr/pdf/infos_legales/Vie_Privee_V1

[12] Nghiem Long PHAN et Sébastien FONTAINE, « *Internet et la vie privée* », AuthSecu. Le site français de la sécurité informatique pour les entreprises, consulté le 12/03/2011, disponible sur: http://www.authsecu.com/internet-et-la-vie-prive/internet-et-la-vie-prive.php

[13] « Wapedia encyclopedia», consulté le 23/03/2011, disponible sur: http://wapedia.mobi/fr/Usurpation_d'identit%C3%A9

[14] « Wikipedia, the free encyclopedia», consulté le 05/04/2011, disponible sur: http://fr.wikipedia.org/wiki/Hame%C3%A7onnage

[15] «Tom's Hardware France», consulté le 13/04/2011, disponible sur: http://www.presence-pc.com/lexique/Phishing,03-0-415.html

[16] « Comment Ça Marche (CCM) ». Communauté informatique, consulté le 13/04/2011, disponible sur: http://www.commentcamarche.net/contents/attaques/scam.php3

[17] « Wikipedia, the free encyclopedia», consulté le 17/04/2011, disponible sur: http://fr.wikipedia.org/wiki/Logiciel_espion

[18] « Portail officiel de la sécurité informatique - ANSSI - République française », consulté le 17/04/2011, disponible sur: http://www.securite-informatique.gouv.fr/gp_article47.html

[19] « Wikipedia, the free encyclopedia», consulté le 21/04/2011, disponible sur: http://fr.wikipedia.org/wiki/Ver_informatique

[20] « Wikipedia, the free encyclopedia», consulté le 21/04/2011, disponible sur: http://fr.wikipedia.org/wiki/Publiciel

[21] « Vox Internet - Internet governance the democratic construction of standards» : http://www.voxinternet.org/IMG/pdf/IdO

[22] « Microsoft Online Privacy Notice Highlights », consulté le 25/04/2011, disponible sur: http://privacy.microsoft.com/fr-fr/windowsliveid.mspx

[23] « support.microsoft.com », consulté le 22/04/2011, disponible sur: http://support.microsoft.com/kb/916988/fr

[24] « Wikipedia, the free encyclopedia», consulté le 21/04/2011, disponible sur: http://fr.wikipedia.org/wiki/Authentification_unique

[25] « Wikipedia, the free encyclopedia», consulté le 20/05/2011, disponible sur: http://fr.wikipedia.org/wiki/OpenID

[26] « Spécification de la platforme pour les préférences de confidentialité 1.0 (P3P1.0) », consulté le 05/03/2011, disponible sur: http://www.yoyodesign.org/doc/w3c/p3p1/index.html

[27] « Wikipedia, the free encyclopedia», consulté le 05/03/2011, disponible sur: http://fr.wikipedia.org/wiki/P3P

[28] Sébastien GAMBS, *« Introduction à la protection de la vie privée* », Cours, .IRISA. Institut de Recherche en Informatique et Systèmes Aléatoires, 19 novembre 2010.

[29] Yves DESWARTE, « *Technologies pour la Protection de la vie privée sur Internet* », Cours, -ENSEEIHT- École notionnel supérieur d'électrotechnique, d'électronique, d'informatique, d'hydraulique, et des télécommunications, LAAS-CNRS, Toulouse.

[30] Bellanova Rocco et De Hert Paul, « *Protection des données personnelles et mesures de sécurité : vers une perspective transatlantique* », Cultures & Conflits, 2009/2 n° 74, p. 63-80.

[31] Laurent COLLÉE, « *Sécurité et vie privée sur les réseaux sociaux* », Mémoire de master, Université du Luxembourg, 2009.

[32] BEDARD Matthieu-BONET Nicolas-LORY Nicolas, « *Man In The Middle* », Rapport, (UMR6072-GREYC-) Groupe de Recherche en Informatique, Image, Automatique et Instrumentation de Caen, 2006-2007.

[33] OLIVIER, « *Appareils photos numériques et vie privée* », Article,Respect et protection de la vie privée -www.vieprivee.com-, lundi 12 juillet 2004.

[34] Marie-Hélène BEAULIEU-France GAIGNARD et Stéphanie POULIN, « *Usurpation d'identité* », Revue, Association des consommateurs de québec, Novembre 2003.

[35] DLARE, «*Vie privée, comment éviter le profilage sur Internet*», COAGUL (http://coagul.org/drupal), 07 Mai 2010.

[36] Kamel REZGUI, «*Protection technique des données personnelles de l'internaute* », Faculté de droit de Tunis, 26 février 2008, DRIMAN Laboratoire de droit des relations internationales, des marches et des négociations.

[37] Jacques LAVALLEE et Danielle PERRAS, « *Aspects éthiques et sociaux de l'entreprise numérique* », Cours, Université de Sherbrooke.

[38] Esma AIMEUR, « *Vie privée sur Internet* », Cours IFT CEL 6001, Novembre 2007.

[39] Ann CAVOUKIAN et Ph.D,« *Privacy by Design The 7 Foundational Principles*»,Information and Privacy Commissioner of Ontario Canada, Aôut 2009.

[40] GROUPE "INFORMATIQUE & LIBERTES 2.0 ?», « *Le nouveau paysage des données personnelles: Quelles conséquences sur les droits des individus ? »*, Janvier 2009.

[41] Olivier ITEANU, « *L'identité numérique en question* », Eyrolles, p.82

[42] Sébastien GAMBS, « *Réseaux de communication anonyme* », « IRISA ». Institut de Recherche en Informatique et Systèmes Aléatoires, 19 novembre 2010.

[43] Adrien GUINAULT, « *Retour sur le Hack Of The Year : TOR, votre meilleur ennemi* », Janvier 2008, « XMCO - Consultants en sécurité informatique (PCI QSA et ISO 27001 Lead Auditor) »

[44] Uciel FRAGOSO-RODRIGUEZ, « *Modèle de Respect de la Vie Privée dans une Architecture d'Identité Fédérée* », thèse de doctorat, Université D'EVRY-VAL D'ESSONNE Mexique, 2009.

[45] Vincent-CRIDLIG-Olivier-FESTOR-Jacques-GUYARD-PierreEtienne MOREAU, «*Formalisation et évaluation de politiques*», LORIA - Campus Scientifique -, CFIP'02, 5 mai 2003.

[46] Romain LABORDE et Thierry DESPARTS, «*Gestion de conditions stables dans XACML : intérêt d'une approche par notification*», Université Paul Sabatier-France-.

[47] Anne ANDERSON et Staff ENGIEER, « *Privacy Policy Languages: XACML vs EPAL*», 5th Annual Privacy & Security Workshop, 29 October 2004.

[48] Mark O'Neill et al, « *Web Services Security*», McGraw-Hill/Osborne 2003 (312 pages), chapitre 7.

[49] Jan Paul Kolter, « *User Centric Privacy*», JOSEF EUL VERLAG GmbH, Lohmar-Koln, 2010.

[50] Karine SOLOVIEFF, « *Vie privée : un oiseau veille au grain* », 08/04/2002, 01net.

[51] Lorrie Faith CRANOR, « *User Interfaces for Privacy Agents*», Carnegie Mellon University, 7/14/05 DRAFT – UNDER REVIEW – PAGE 1.

Annexes

Annexe A

Cinq démarches pour protéger votre vie privée en ligne

1- Limitez les données personnelles que vous divulguez

- Ne fournissez que l'information nécessaire pour effectuer la transaction. Ne fournissez aucun renseignement qui ne soit pas obligatoire, telles que les données biographiques. Si possible, utilisez un pseudonyme. Si le site vous semble demander un nombre excessif de données non essentielles, montrez votre désaccord en changeant de site.
- Empêchez vos logiciels de divulguer vos données personnelles. Parmi ces logiciels, « Navigation intelligente » de Netscape, et de Internet Explorer « AutoFill» (logiciel qui remplit automatiquement les formulaires du Web). Une option dans le menu « Préférences » permet de désactiver ces fonctions. Maints autres programmes, y compris les logiciels de traitement de texte, les jeux et les programmes Internet, renvoient souvent des informations sur l'usager à l'entreprise exploitant le site. Pour une liste des logiciels et des programmes pour les éliminer, visiter : http://www.grc.com/OptOut.htm
- Sachez que lorsque vous présentez un message dans des newsgroups (forums), des chatrooms (lieux de rencontre virtuelle pour bavarder), des sites Web ou ailleurs sur la Toile, l'information est souvent stockée et rendue publiquement disponible.

2- Ouvrez un compte e-mail indépendant

- Ouvrez des comptes e-mail gratuits indépendants de votre compte personnel ou commercial, à utiliser uniquement pour faire du commerce électronique, participer aux newsgroups, chatrooms etc. Pour un annuaire des services e-mail gratuits, visiter : http://www.emailaddresses.com/

3- Refusez d'accepter les cookies

Réglez votre navigateur afin qu'il refuse d'accepter les cookies, ou tout du moins les cookies provenant de tiers tels des entreprises comme DoubleClick. Choisissez un réglage dans le menu « Outils » d'Internet Explorer et « Préférences - Avancées » de Netscape. Lorsque vous voulez utiliser un des rares sites qui exige un cookie, activez provisoirement les cookies et effacez les une fois que vous avez terminé. Vous pouvez également utiliser un programme qui effacera ou qui modifiera vos fichiers cookies comme bon vous semble.

4- Utilisez des outils pour protéger votre vie privée

- Vous pouvez également empêcher les sites Web de rassembler vos données personnelles en utilisant les services qui permettent de surfer sur Internet en gardant l'anonymat. D'autres moyens de protéger votre vie privée sur Internet sont : le cryptage pour verrouiller vos messages ; et les utilitaires pour supprimer en permanence les fichiers et les données personnelles hébergés dans votre disque dur. Pour une liste ce ces outils, visitez : http://epic.org/privacy/tools.html

5- Renseignez vous sur vos protections juridiques

- De nombreuses juridictions intègrent des lois destinées à protéger la vie privée des consommateurs.
- Pour télécharger une copie du rapport « Privacy@net » publié par Consumers International, voir www.consumersinternational.org Vous pouvez également imprimer des copies supplémentaires de cette fiche de renseignements. Le rapport et la fiche sont disponibles en Anglais, Français et Espagnol.

Annexe B

Les éléments de P3P 1.0

P3P Element	Attribute	Plain Language Translation
POLICY	discuri (attribute of POLICY element)	Read our full privacy policy at [with link to discuri]
POLICY	opturi (attribute of POLICY element)	Find out how to opt-in or opt-out at [with link to opturi]
ENTITY		This policy is issued by: [display all entity information provided by site]
ACCESS		Your access to information about you:
ACCESS	nonident	We do not keep any information identified with you
ACCESS	all	We give you access to all of our information identified with you
ACCESS	contact-and-other	We give you access to your contact information and some of our other information identified with you
ACCESS	ident-contact	We give you access to only your contact information in our records
ACCESS	other-ident	We allow you to access some of our information identified with you, but not your contact information
ACCESS	none	We do not give you access to our information about you
DISPUTES		Ways to resolve privacy-related disputes with us include:
DISPUTES	service	[display long description and short description, if provided, with hyperlink to service URI, otherwise display "customer service" with hyperlink to service URI]
DISPUTES	independent	[display long description and short description, if provided, with hyperlink to service URI, otherwise display "independent organization" with hyperlink to service URI]
DISPUTES	court	We believe that the following authority offers recourse for disputes: [display long description and short description, if
		provided, with hyperlink to service URI, otherwise display "possible legal complaint" with hyperlink to service URI]
DISPUTES	law	We believe that the following laws or regulations provide recourse: [display long description and short description, if provided, with hyperlink to service URI, otherwise display "law" with hyperlink to service URI]

REMEDIES		[no heading - display this following corresponding disputes element]
DISPUTES	correct	We will correct any errors we make related to the commitments in our privacy policy
DISPUTES	money	We will compensate individuals if it is determined that we have violated our privacy policy
DISPUTES	law	Our privacy policy references a law that may determine remedies for breaches of our policy
NON-IDENTIFIABLE		We do not keep any information that could be used to identify you personally
PURPOSE		The ways your information may be used:
PURPOSE	current	To provide the service you requested
PURPOSE	admin	To perform web site and system administration
PURPOSE	develop	For research and development, but without connecting any information to you
PURPOSE	tailoring	To customize the site for your current visit only
PURPOSE	pseudo-analysis	To do research and analysis in which your information may be linked to an ID code but not to your personal identity
PURPOSE	pseudo-decision	To make decisions that directly affect you without identifying you, for example to display content or ads based on links you clicked on previously
PURPOSE	individual-analysis	To do research and analysis that uses information about you
PURPOSE	individual-decision	To make decisions that directly affect you using information about you, for example to recommend products or services based on your previous purchases
PURPOSE	contact	To contact you through means other than telephone (for example, email or postal mail) to market services or products
PURPOSE	historical	To aid in historical preservation as governed by a law or policy described in this privacy policy
PURPOSE	telemarketing	To contact you by telephone to market services or products
PURPOSE	other-purpose	For other uses: [include site's human, readable explanation; if site omits human-readable explanation say "not described here"]

PURPOSE	required (attribute of purpose and recipients elements)	(attribute, see below)
PURPOSE	required always	(no remark)
PURPOSE	required opt-in	[append to purpose/recipient] -- only if you request this
PURPOSE	required opt-out	[append to purpose/recipient] -- unless you opt-out
RECIPIENT		With whom we may share your information
RECIPIENT	ours	Companies that help us fulfill your requests (for example, shipping a product to you), but these companies must not use your information for any other purpose
RECIPIENT	delivery	Delivery companies that help us fulfill your requests and who may also use your information in other ways
RECIPIENT	same	Companies that have privacy policies similar to ours
RECIPIENT	other-recipient	Companies that are accountable to us, though their privacy policies may be different from ours
RECIPIENT	unrelated	Other companies whose privacy policies are unknown to us
RECIPIENT	public	People who may access your information from a public area, such as a bulletin board, chat room, or directory
RECIPIENT	required (attribute of purpose and recipients elements)	(attribute, see below)
RECIPIENT	required always	(no remark)
RECIPIENT	required opt-in	[append to purpose/recipient] -- only if you request this
RECIPIENT	required opt-out	[append to purpose/recipient] -- unless you opt-out
RETENTION		How long we may keep your information
RETENTION	no-retention	We do not keep your information beyond your current online session
RETENTION	stated-purpose	We keep your information only long enough to perform the activity for which we collected it
RETENTION	legal-requirement	We keep your information only as long as we need to for legal purposes
RETENTION	business-practices	Our full privacy policy explains how long we keep your information
RETENTION	indefinitely	We may keep your information indefinitely

CATEGORIES		We may collect the following types of information about you
CATEGORIES	physical	Name, address, phone number, or other physical contact information
CATEGORIES	online	Email address or other online contact information
CATEGORIES	uniqueid	Website login IDs and other identifiers (excluding government IDs and financial account numbers)
CATEGORIES	purchase	Information about your purchases, including payment methods
CATEGORIES	financial	Financial information such as accounts, balances, and transaction history
CATEGORIES	computer	Information about the computer you are using, such as its hardware, software, or Internet address
CATEGORIES	navigation	Which pages you visited on this web site and how long you stayed at each page
CATEGORIES	interactive	Activities you engaged in at this web site, such as your searches and transactions
CATEGORIES	demographic	Information about social and economic categories that might apply to you, such as your gender, age, income, or where you are from
CATEGORIES	content	Messages you send to us or post on this site, such as email, bulletin board postings, or chat room conversations
CATEGORIES	state	Cookies and mechanisms that perform similar functions
CATEGORIES	political	Which groups you might be a member of such as religious organizations, trade unions, and political parties
CATEGORIES	health	Health information such as information about your medical condition or your interest in health-related topics, services, or products
CATEGORIES	preference	Information about your tastes or interests
CATEGORIES	location	Information about an exact geographic location, such as data transmitted by your GPS-enabled device
CATEGORIES	government	Government-issued identifiers such as social security numbers
CATEGORIES	other-category	Other types of data: [include site's human, readable explanation; if site omits human-readable explanation say "not described here"]
CATEGORIES	optional (attribute of data elements)	• if no: the data element is required [append to data element or category] • if yes: the data element is optional [append to data element or category]

Annexe C

Les éléments d'APPEL 1.0

APPEL Element	Attribute	Plain Language Translation
appel:RULESET	This tag is the delimiter that denotes an APPEL file. It includes a sequence of one or more rules. Each rule features a certain behavior that is returned to the calling program if the expressions listed in the rule all evaluate to true.	
appel:RULESET	**crtdby**	Name or ID of the ruleset author (could be the user agent).
appel:RULESET	**crtdon**	Time & Date of ruleset creation.
appel:RULESET	**description**	A short natural language explanation that can be displayed by the user agent when the ruleset gets selected, or to help debugging a rulefile.
appel:RULESET	Contains conditions under which a certain behavior should be carried out by the calling program.	
appel:RULE	**behavior**	Behavior that should be carried out by the calling program if the expressions match the evidence.
appel:RULE	**connective**	Allows for different matching semantics of enclosed sub elements.
appel:RULE	**crtdby**	Name or ID of the rule author (could be the user agent).
appel:RULE	**crtdon**	Time & Date of rule creation.

appel:RULE	**description**	A short natural language explanation that can be displayed by the user agent when the rule gets executed, or to help debugging a rulefile. Note that a separate promptmsg should be used in case the user should be prompted for a decision.
appel:RULE	**prompt**	Indicates whether a prompt message should be displayed to the user. If this attribute is not present, no prompt message is displayed.
appel:RULE	**persona**	If the user agent supports multiple user repositories, this string identifies the data repository that should be used in case the resource is accessed (i.e. if the rule that fires features a "request" or "limited" behavior, or if a "block" rule is overridden at the prompt). If no persona is given, the user agent's default value is used.
appel:RULE	**promptmsg**	A short natural language explanation or question that can be displayed by the user agent when the user should be prompted for a decision. Note that the description field can be used to hold a brief summary of the rule for debugging or informational purposes.
appel:OTHERWISE	So called degenerate-expression, which always evaluates to true. This can be used to craft "catch-all" rules that match all cases not covered by previous rules.	
appel:REQUEST	Allows the creation of rules that only apply to a certain resource or domain.	
appel:REQUEST	**connective**	Allows for different matching semantics of enclosed sub elements.
appel: REQUEST	**uri**	The URI of the currently requested resource (not the policy URI).
appel:connective	Determines how contained expressions are matched when a rule is compared to the available evidence. APPEL supports six different kinds of connectives: or, and, non-or, non-and, or-exact and and-exact. If no appel: connective is given, APPEL's matching semantics default to an and match: *All* of the contained expressions *must* appear in the evidence, *additional* elements will be ignored.	

Résumé

L'augmentation du nombre de fraudes sur internet dues aux vols des données personnelles ont alerté les internautes sur la nécessité de respect de la vie privée. Ils sont de plus en plus nombreux à être préoccupés par l'utilisation de leurs informations personnelles, ce qui a poussé les sites internet à publier des politiques de respect de la vie privée. Ces politiques de confidentialité sont publiées grâce à P3P l'acronyme de Platform for privacy Preferences. Dans ce travail nous avons développé un système pour la protection de la vie privée des utilisateurs Web. Pour pouvoir détecter toute utilisation non autorisée des informations personnelles des utilisateurs, notre système se base sur les deux recommandations du W3C : le P3P et le langage de définition des préférences de confidentialité : APPEL.

Abstract

The increases in Internet fraud due to the theft of personal data have alerted users about the need to respect of their privacy. They are more likely to be concerned about the use of their personal information; this prompted the websites to publish policy of respect for privacy. These privacy policies are published with P3P (the Platform for privacy Preferences). In this work we have developed a system to protect the privacy of Web users. To detect unauthorized use of personal information of users, our system is based on two W3C recommendations: the P3P and the privacy preferences definition language "APPEL".

ملخص

لقد نبهت الزيادة في الاحتيال على الإنترنت بسبب سرقة البيانات الشخصية للمستخدمين على ضرورة احترام الخصوصية. و إن تزايد قلق هؤلاء المستخدمين حول كيفية استخدام معلوماتهم الشخصية أدى بالمواقع لنشر سياسة احترام الخصوصية . ويتم نشر هذه السياسات باستخدام نظا م P3P الذي يحدد كيفية استخدام مواقع الإنترنت للبيانات التي يتم جمعها عند زيارة المستخدمين للموقع. في هذا العمل قمنا بتطوير نظام يقوم على حماية البيانات الشخصية لمستخدمي مواقع الانترنيت . و يعتمد هذا النظام من أجل كشف أية تجاوزات في استخدام البيانات الشخصية على نظام P3P و لغة APPEL التي تمكن المستخدمين من تحديد قواعد عن كيفية استعمال بياناتهم الخاصة من طرف المواقع.