Université de Montréal

Priv-C : une politique de confidentialité personnalisable

par
Oluwa Sosso Lawani

Département d'Informatique et de Recherche Opérationnelle

Faculté des arts et des sciences

Mémoire présenté à la Faculté des arts et des sciences en vue de l'obtention du grade de Maitrise en Sciences en Informatique

Janvier 2016

Résumé

Les politiques de confidentialité définissent comment les services en ligne collectent, utilisent et partagent les données des utilisateurs. Bien qu'étant le principal moyen pour informer les usagers de l'utilisation de leurs données privées, les politiques de confidentialité sont en général ignorées par ces derniers. Pour cause, les utilisateurs les trouvent trop longues et trop vagues, elles utilisent un vocabulaire souvent difficile et n'ont pas de format standard.

Les politiques de confidentialité confrontent également les utilisateurs à un dilemme : celui d'accepter obligatoirement tout le contenu en vue d'utiliser le service ou refuser le contenu sous peine de ne pas y avoir accès. Aucune autre option n'est accordée à l'utilisateur.

Les données collectées des utilisateurs permettent aux services en ligne de leur fournir un service, mais aussi de les exploiter à des fins économiques (publicités ciblées, revente, etc). Selon diverses études, permettre aux utilisateurs de bénéficier de cette économie de la vie privée pourrait restaurer leur confiance et faciliter une continuité des échanges sur Internet.

Dans ce mémoire, nous proposons un modèle de politique de confidentialité, inspiré du P3P (une recommandation du W3C, World Wide Web Consortium), en élargissant ses fonctionnalités et en réduisant sa complexité. Ce modèle suit un format bien défini permettant aux utilisateurs et aux services en ligne de définir leurs préférences et besoins. Les utilisateurs ont la possibilité de décider de l'usage spécifique et des conditions de partage de chacune de leurs données privées. Une phase de négociation permettra une analyse des besoins du service en ligne et des préférences de l'utilisateur afin d'établir un contrat de confidentialité.

La valeur des données personnelles est un aspect important de notre étude. Alors que les compagnies disposent de moyens leur permettant d'évaluer cette valeur, nous appliquons dans ce mémoire, une méthode hiérarchique multicritères. Cette méthode va permettre également à chaque utilisateur de donner une valeur à ses données personnelles en fonction de l'importance qu'il y accorde.

Dans ce modèle, nous intégrons également une autorité de régulation en charge de mener les négociations entre utilisateurs et services en ligne, et de générer des recommandations aux usagers en fonction de leur profil et des tendances.

Mots-clés : Politique de confidentialité, personnalisation, options, négociation, économie de la vie privée, incitations économiques.

Abstract

Privacy policies define the way online services collect, use and share users' data. Although they are the main channel through which users are informed about the use of their private data, privacy policies are generally ignored by them. This is due to their long and vague content, their difficult vocabulary and their no standard format.

Privacy policies also confront users to a dilemma. Indeed, they must agree to all their content in order to use the service or reject it, and in this case they do not have access to the service. No other alternative is given to the user.

Online services process data collected from users to provide them a service, but they also exploit those data for economic purposes (targeted advertising, resale, etc.). According to various studies, allowing users *to benefit from the use of their data* could restore their trust towards online services and facilitate data exchanges on the Internet.

In this work, we propose a new model of privacy policy, inspired by the P3P (a World Wide Web Consortium - W3C Recommendation) but increasing its functionalities and reducing its complexity. This model defines a specific structure allowing users and online services to define their preferences and needs. Users have the opportunity to decide for each of their private data, specifying how it will be used and shared. A negotiation phase will allow a needs analysis of the online service and preferences of the user to establish a confidentiality agreement.

The value of personal data is also an important aspect of our study. While companies have resources allowing them to rate this value, we apply in this thesis, a *hierarchical multi-criteria method*. This method will allow each user to give value to his personal data according to the importance he attaches to it.

In this model, we also integrate a regulation authority. It is in charge of conducting *negotiations* between users and online services, and generate *recommendations* to users based on their profile and current trends.

Keywords: Privacy Policy, personalisation, options, negotiation, economics of privacy, economic incentives.

Table des matières

Liste des tableaux

Liste des figures

À

mes parents et à toute ma famille

Remerciements

J'adresse mes premiers remerciements à Professeure Esma Aïmeur, ma directrice de recherche, pour son encadrement durant cette année de recherche. Ses enseignements, commentaires, critiques ainsi que sa disponibilité m'ont permis de toujours m'améliorer et avancer dans mes travaux. Merci !

J'adresse également mes remerciements à Professeure Kimiz Dalkir, qui, malgré la distance, a contribué à la réalisation et aboutissement de mes travaux de recherche.

Je remercie mes collègues du laboratoire HERON avec qui j'ai partagé des moments chaleureux dans une ambiance riche en expériences. Leurs commentaires lors et hors de nos réunions d'équipe m'ont permis d'améliorer mes axes de recherche et me familiariser avec divers concepts.

Je remercie enfin ma famille qui, de loin, m'a toujours apporté son soutien et encouragements. Un merci particulier à mon frère, ancien étudiant de l'Université de Montréal, grâce à qui j'ai pu y entamer mes études supérieures.

1 Chapitre 1 Introduction

1.1 Contexte

La technologie submerge aujourd'hui tous les domaines de la vie courante. Des objets autrefois traditionnels, deviennent connectés et sont en mesure de traiter des informations et de les transmettre via Internet. Les téléphones intelligents sont dotés de capteurs leur permettant d'analyser des données telles que : le *regard* de l'utilisateur, les *empreintes digitales*, la *localisation* et la *proximité*. Les montres et bracelets connectés sont dotés de cardio-fréquencemètre et peuvent analyser le rythme cardiaque, le nombre de pas effectués, etc. Grâce au NFC (Near Field Communication) embarqué sur les téléphones intelligents, ceux-ci peuvent être directement utilisés pour effectuer des achats comme si l'usager avait une carte bancaire ou une carte de fidélité. Cette mondialisation de la technologie a tendance à prendre de l'ampleur et facilite la vie aux consommateurs. La question qui se pose alors est : « À quel prix ? ».

En effet, cela entraine un risque de plus en plus important sur la vie privée des usagers. On peut se questionner sur les données collectées par ces appareils: comment sont-elles réellement utilisées ? Qui peut y avoir accès ? Et à qui profite réellement l'exploitation de toutes ces informations ? Selon Daniel Therrien, Commissaire à la protection de la vie privée du Canada, le respect de la vie privée constitue de plus en plus un indicateur du degré d'imputabilité des organisations et, par le fait même, de la confiance des citoyens et des consommateurs[1].

1.2 Problématique

La protection de la vie privée des utilisateurs de services en ligne est devenue une problématique de plus en plus importante, affectant les usagers et l'échange de données sur Internet. Selon un sondage de Pew Research[2] en 2014, 91% des internautes américains sont d'accord pour dire que les consommateurs ont perdu le contrôle sur la

[1] https://www.priv.gc.ca/parl/2015/parl_20150525_f.asp, accédé le 24/12/2015
[2] http://www.pewinternet.org/2014/11/12/public-privacy-perceptions/, accédé 17/12/2015

manière dont sont collectées et utilisées leurs informations privées par les entreprises. D'après ce même sondage, 81% des internautes se sentent surveillés lorsqu'ils partagent des informations confidentielles sur les réseaux sociaux.

Par conséquent, ce mémoire s'intéresse plus particulièrement aux politiques de confidentialité, premier accord entre un utilisateur et un service en ligne sur la protection (ou non) de ses informations personnelles. Nous abordons ici deux volets fondamentaux concernant les politiques de confidentialité.

Le premier point est *le désintéressement* des utilisateurs relatif au contenu des politiques de confidentialité. Ceci s'explique par le fait qu'elles ont en général des contenus vagues ou difficiles à comprendre (O. o. t. P. C. o. Canada, 2013). Raisons principales pour laquelle, d'après une étude publiée en 2013, 55 % des répondants déclarent ne jamais lire les politiques de confidentialité (dos Santos Brito, Cardoso Garcia, Araujo Durao, & Romero de Lemos Meira, 2013). Comme nous l'étudierons dans ce mémoire, d'autres raisons expliquent ce phénomène plutôt inquiétant, sachant que les politiques de confidentialité sont pratiquement le seul moyen informant les utilisateurs des pratiques relatives à leurs données personnelles.

Le second point évoqué dans ce mémoire est *l'absence absolue de dialogue* entre l'utilisateur et le service en ligne en ce qui concerne les politiques de confidentialité. Dans le monde actuel des objets connectés, il est difficile pour les utilisateurs de savoir lesquelles de leurs données sont collectées, utilisées ou revendues, et ce, malgré la présence de politiques de confidentialité quand elles existent. Le New York Times révélait dans un article[3] l'astuce d'une compagnie de collecte de données qui utilisait les signaux Wi-Fi des consommateurs, sans les en informer. Grace à ces signaux, ils pouvaient savoir combien de temps le consommateur est resté dans un magasin, les rayons qu'il a visités, etc. Ceci pouvait être raffiné à partir de la vidéosurveillance pour

[3] http://www.nytimes.com/2013/07/15/business/attention-shopper-stores-are-tracking-your-cell.html?_r=0, accédé 21/1/2015

déduire d'autres informations telles que le sexe, etc. L'avis explicite de l'usager devrait donc être demandé pour toute collecte et usage de ses informations personnelles (Richards & King, 2014). Ce dernier se voit actuellement contraint d'accepter le contenu d'une politique pour pouvoir utiliser le service, sans aucune possibilité de discuter ou de négocier des clauses.

1.3 Objectifs du mémoire

Nos objectifs, pour répondre aux problématiques précédentes, se définissent autour des conditions de collecte de données et de leur utilisation. Notre contribution s'articule ainsi autour de quatre principaux points.

1.3.1 Quelles options pour l'utilisateur ?

L'accord de l'utilisateur est très souvent implicite en matière de politiques de confidentialité. Notre premier objectif est donc de résoudre ce problème de choix offerts aux utilisateurs. Ces derniers sont en effet contraints d'accepter ou de refuser entièrement une politique. Un faux dilemme les obligeant le plus souvent à adhérer à un service en ligne sans être d'accord avec les pratiques de ce dernier. D'ailleurs, il va de soi que les utilisateurs ne ressentent pas le besoin de lire les politiques de confidentialité vu qu'ils ne peuvent pas les discuter.

Nous visons donc d'établir une meilleure relation de confiance avec les consommateurs, en leur signifiant plus clairement et de façon explicite les enjeux sur leur vie privée et en leur offrant diverses options sur la collecte et l'usage de chacune de leurs données.

La loi québécoise stipule à ce sujet[4] :

> *Le consentement à la communication ou à l'utilisation d'un renseignement personnel doit être manifeste, libre, éclairé et être donné à des fins spécifiques. Ce consentement ne vaut que pour la durée nécessaire à la réalisation des fins pour lesquelles il a été demandé.*

[4] http://www.cai.gouv.qc.ca/entreprises/protection-des-renseignements-personnels-1/, accédé 02/01/2016

1.3.2 La valeur des données personnelles

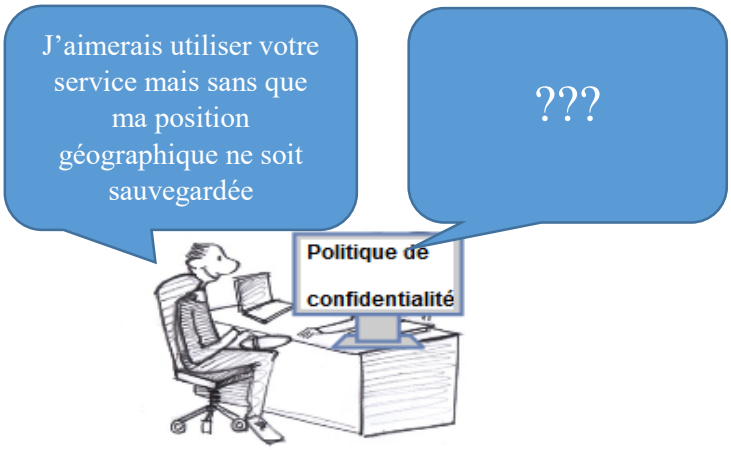
Les solutions aux problèmes de confidentialité ont été étudiées sous divers aspects (politique, social,…). L'aspect économique est également abordé par de nombreuses études, et semble être un moyen efficace d'obtenir un compromis plus ou moins équitable pour l'échange de données entre consommateurs et fournisseurs de services (Chorppath & Alpcan, 2013; Osothongs & Sonehara, 2014). Mesurer la vie privée d'un usager (Nepali & Wang, 2013; Wang, Nepali, & Nikolai, 2014) ou la valeur de ses données privées sont dès lors des sujets critiques et fort importants pour parvenir à cette équité. Alors que de nombreuses entreprises ont déjà déterminé un coût aux données privées, d'autres chercheurs pensent que la valeur des données privées est relative, et varie en fonction de nombreux facteurs (Acquisti, John, & Loewenstein, 2013).

Nous étudions ici les politiques de confidentialité, avec une orientation économique, pour concevoir notre modèle. Il vise à permettre aux utilisateurs, à partir d'un modèle mathématique, d'évaluer la valeur des données personnelles, en fonction de l'importance qu'ils y accordent. Cette valeur sera donc utilisée durant les échanges avec les services en ligne.

1.3.3 La négociation des termes

L'utilisateur donne son accord (une forme de signature) à la politique de confidentialité, elle-même signée préalablement par le service en ligne. Il peut aussi poursuivre le service en ligne si celui-ci venait à violer une clause de sa politique de confidentialité. Considérant ces facteurs, celle-ci apparait donc comme une forme de contrat entre les deux parties. Toutefois, dans la majorité des contrats, les signataires ont la possibilité de discuter sur les termes dudit contrat. Par exemple, avant de signer un contrat de travail, l'employé discute souvent avec l'employeur de certains points tels que le salaire, le temps de travail, etc. Des négociations peuvent alors se produire, par exemple moins d'heures de travail contre une légère baisse de la rémunération. En ce qui concerne les politiques de confidentialité, il n'y a pas de discussion possible car il n'y a aucun interlocuteur face à l'utilisateur.

Contrat de travail



Contrat avec un service en ligne

1.3.4 L'automatisation des tâches

Notre objectif final est d'automatiser toutes les opérations décrites ci-dessus. Divers modèles ont déjà vu le jour mais ne répondent toujours pas entièrement aux enjeux actuels. Le système que nous mettons en place doit permettre aux usagers et aux services en ligne de conclure des contrats de confidentialité avec le minimum d'interventions humaines possibles. Les usagers doivent également être certains, sans avoir lu la politique de confidentialité, que le contrat qui en résulte est personnalisé en fonction de leur préférence de confidentialité.

Organisation du mémoire

Ce mémoire est organisé comme suit : le chapitre 2 présente l'état de l'art. Nous y abordons trois volets fondamentaux que sont les **politiques de confidentialité**, la **vie privée** et **l'économie de la vie privée**. Dans le chapitre 3, nous présentons la méthodologie et la solution que nous proposons. Le chapitre 4 est consacré à l'implémentation de notre solution qui est validée dans le chapitre 5. Nous y présentons aussi les résultats issus de la comparaison de notre approche avec d'autres solutions existantes. L'évaluation du modèle a été effectuée à l'aide de diverses expérimentations,

avec plus de 1000 participants sondés. Les résultats ont fait objet de deux publications dans la conférence CRiSIS (The International Conference on Risks and Security of Internet and Systems) 2015 et dans le journal CHB (Computers in Human Behavior).

2 Chapitre 2 État de l'art

Ce chapitre fait le point sur l'état actuel des politiques de confidentialité. Il s'agira des problèmes actuels rencontrés avec les politiques de confidentialité mais aussi avec la protection de la vie privée en général. Nous faisons le lien entre la vie privée et l'économie de la vie privée, en mettant un accent sur la valeur des données privées. Ce dernier point est en effet important car nous tenterons dans notre solution d'y apporter une réponse.

2.1 Les politiques de confidentialité

L'objectif principal de ce mémoire étant les politiques de confidentialité, cette section fait une présentation de même qu'un état des lieux. Différents points de vue seront étudiés, tant celui des utilisateurs que celui des services en ligne.

2.1.1 Histoire et définition

Les politiques de confidentialité décrivent aux utilisateurs comment les services en ligne collectent et utilisent leurs données (voir exemples en Figures 2.1 et 2.2). La quasi-totalité des services en ligne disposent d'une politique de confidentialité. Bien que divers formats existent, la plupart se présentent sous forme d'un texte, structuré ou non. Le contenu, différent d'un service à un autre, décrit en général les données qui sont recueillies des utilisateurs, comment elles sont collectées, utilisées et partagées.

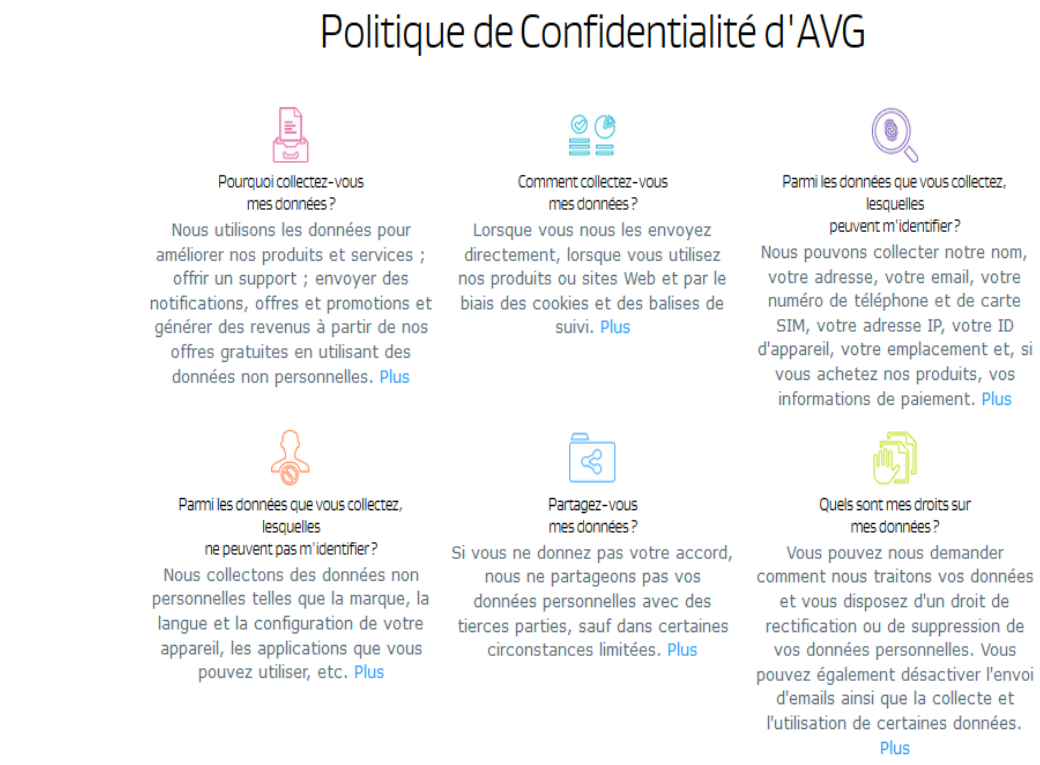


Figure 2.1 : Politique de confidentialité d'AVG-Antivirus

Figure 2.2 : Politique de confidentialité de Google

Les premiers textes relatifs aux données des utilisateurs remontent aux années 1970. Aux États-Unis, la première véritable loi relative aux données privées est le **Privacy Act** établi en 1974. Il établit un code de pratiques équitables qui régit la collecte, l'entretien, l'utilisation et la diffusion des informations des individus qui sont conservées dans des dossiers par les organismes fédéraux (Justice, 2015). Bien que certains pays se soient manifestés plus tôt en Europe, l'une des premières lois européennes adoptées remonte à 1981. Il s'agit de la *Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel*. Son but, tel que spécifié dans la convention, est de garantir, sur le territoire de chaque partie, à toute personne physique, quelles que soient sa nationalité ou sa résidence, le respect de ses droits et de ses libertés fondamentales, notamment de son droit à la vie privée, à l'égard du traitement automatisé des données à caractère personnel la concernant (l'Europe, 1981).

Les lois se sont au fur et à mesure renforcées, obligeant parfois les entreprises privées à s'y conformer au mieux possible. Le Canada dispose de deux lois fédérales sur la protection de la vie privée : la Loi sur la Protection des Renseignements Personnels (LPRP), qui régit les pratiques de traitement des renseignements personnels des ministères et organismes fédéraux, et la Loi sur la Protection des Renseignements Personnels et les Documents Électroniques (LPRPDE), qui s'applique au secteur privé (P. d. Canada, 2015).

Bien qu'étant régis par des lois, les politiques de confidentialité changent beaucoup en fonction des fournisseurs de services en ligne. Toutefois, on y retrouve en général un certain nombre d'informations, notamment :

- **La collecte des données** : les données qui sont collectées des utilisateurs de même que les méthodes de collecte
- **L'utilisation des données** : comment les informations sont exploitées par le service et dans quel but

- **Le partage de données** : comment et avec qui les données des utilisateurs sont partagées
- **La conservation des données** : méthodes de stockage, de modifications ou suppression des données
- **Les contacts** : comment contacter le service pour des renseignements, plaintes, etc.

Certaines politiques se veulent très brèves, négligeant parfois certains détails importants. On retrouve également des politiques assez longues, couvrant d'autres aspects que ceux cités ci-dessus. La politique de confidentialité de Microsoft[5] présente par exemple des sections décrivant comment Microsoft s'engage à protéger la confidentialité de ses utilisateurs et des sections décrivant chacun des services qu'il offre (Skype, MSN, Xbox, Bing, etc.).

Bien qu'étant le seul moyen informant les internautes de l'utilisation de leurs données privées, diverses études montrent que ceux-ci ne les lisent pas en général.

2.1.2 État des lieux des politiques de confidentialité

En 2013, le Commissariat à la Protection de la Vie Privée du Canada publie un sondage (O. o. t. P. C. o. Canada, 2013) effectué auprès des Canadiens sur les enjeux liés à la protection de la vie privée. Le sondage révèle que les canadiens savent qu'il est important que les sites Internet informent, par le biais de politiques de confidentialité, les types de renseignements personnels qu'ils recueillent et l'utilisation qu'ils en font. Cependant, fort est de constater que seulement 21 % des répondants affirment lire toujours ou souvent ces politiques de confidentialité. Plus de la moitié disent ne jamais les lire. D'après le sondage, cela s'explique par le fait que les canadiens pensent que les politiques de confidentialités des sites Internet ne sont pas assez claires. En effet, 62 % des répondants pensent que les politiques de protection de la vie privée des sites web sont quelque peu

[5] https://www.microsoft.com/fr-fr/privacystatement/default.aspx?intsrc=client-_-windows-_-7.16-_-go-privacy&setlang=fr, accédé le 02/01/2016

vagues ou très vagues lorsqu'il s'agit de leur donner de l'information sur ce que l'entreprise fera de leurs données personnelles. Ce sentiment a tendance à augmenter puisque lors du précédent sondage, ils étaient 52 % à le penser.

Dans une autre étude réalisée en 2013 (dos Santos Brito et al., 2013), 55 % des répondants ont effectivement affirmé n'avoir jamais lu les politiques de confidentialité des sites qu'ils visitent. Seuls 4 % déclarent toujours les lire, alors que 39 % disent les lire quelques fois. Cette étude a été conduite avec 900 participants. Il se trouve aussi que beaucoup de services en ligne de nos jours n'affichent pas, de façon systématique, les politiques de confidentialité aux utilisateurs. Ceux-ci doivent simplement cocher une case pour donner leur accord sur le contenu de la politique qu'il est possible d'afficher en suivant un lien. Une étude montre que lorsque les politiques de confidentialité sont affichées systématiquement, les utilisateurs ont tendance à les lires attentivement tandis que la majorité ne les lisent pas lorsqu'ils ont l'option de cocher une simple case (Steinfeld, 2016).

Il s’avère aussi que les attitudes des utilisateurs n’ont pas évolué à travers le temps. Une étude effectuée en 2014 (Williams, Agarwal, & Wigand, 2015) a comparé les résultats de leur sondage à un autre effectué en 2005 par le Centre de la Politique Publique Annenberg (Annenberg Public Policy Center). Ils ont analysé le comportement des utilisateurs par rapport aux politiques de confidentialité, de même que les changements dans ces comportements durant la dernière décennie. Il en ressort que ces attitudes n’ont pas changé. Selon les utilisateurs, les politiques de confidentialité sont toujours aussi longues, complexes et servent principalement à protéger les organisations.

En 2014, une étude analyse les politiques de confidentialité de 75 grandes compagnies afin de voir si elles contiennent assez d'informations pour aider la prise de décision des utilisateurs (Cranor, Hoke, Leon, & Au, 2014). Après comparaison de ces politiques de vie privée, l'étude trouve que de nombreuses compagnies ne sont pas très bavardes par rapport à la collecte et à l'utilisation des données personnelles des utilisateurs. Une comparaison de ces politiques avec les réglementations en vigueur dévoile que de

nombreuses politiques ne sont pas conformes auxdites réglementations. D'après les auteurs, les réglementations sont trop généralistes, ce qui permet facilement aux compagnies de s'y conformer sans pour autant fournir la transparence nécessaire aux consommateurs. Il serait difficile aux utilisateurs d'évaluer les politiques pour prendre des décisions adéquates à cause d'un manque de terminologies bien établies et adoptées par l'ensemble des politiques sur Internet

Outre les facteurs évoqués plus haut, certaines études révèlent que la prise de décision d'un utilisateur quelconque est difficile à cause de l'interprétation faite par ce dernier du contenu des politiques de confidentialité. Des experts et des non-experts ont été soumis à une série de questions relative à des politiques de confidentialité qu'ils ont préalablement lues (Reidenberg et al., 2014). Leur objectif était de déterminer s'ils comprennent suffisamment le contenu des politiques de confidentialité pour prendre des décisions relatives à leur vie privée. Les résultats montrent un écart important d'interprétation en ce qui concerne surtout les questions relatives au partage des données des utilisateurs. Ce manque de compréhension s'observe encore plus pour les services en ligne intégrant un échange de données avec des réseaux sociaux.

La langue est également un problème dans la compréhension des politiques de confidentialité par les utilisateurs. Une étude menée sur les plus grands réseaux sociaux sur Internet compare le degré de traduction de ces sites par rapport à la traduction de leurs politiques de confidentialité (Ur, Sleeper, & Cranor, 2012). Les résultats diffèrent véritablement en fonction des réseaux sociaux. Google+ était alors disponible en 40 langues à part l'anglais. Le site est traduit entièrement en 40 langues et sa politique de confidentialité en 39 langues. Au contraire, Facebook était disponible en 67 langues alors que sa politique de confidentialité est traduite uniquement en 36 langues, dont 26 sont faites partiellement. Il en ressort donc qu'un grand nombre d'utilisateurs de Facebook, qui ne comprennent pas l'anglais, ne sont pas en mesure de lire la politique de confidentialité du site. Il en est de même pour d'autres réseaux populaires tels que

Twitter, disponible en 22 langues tandis que sa politique de confidentialité a été traduite entièrement en trois langues (soit 14%) et partiellement en 17 langues (soit 77%). Ces chiffres viennent confirmer le nombre important d'utilisateurs abonnés à des sites sans avoir lu ou compris entièrement le contenu de leur politique de confidentialité. L'article affirme que cette responsabilité revient aux propriétaires de sites Internet de proposer une traduction des politiques dans toutes les langues dans lesquelles le site est disponible. Alors que certaines politiques sont difficiles à comprendre et d'autres non traduites dans toutes les langues, certains sites ou applications ne disposent carrément pas de politiques de confidentialité. Une étude (Sunyaev, Dehling, Taylor, & Mandl, 2014) examine 600 applications de santé (Mobile Health) parmi les plus utilisées et les mieux notées sur les systèmes Android et iOS. Il ressort que seul 30.5% (183) d'entre elles ont une politique de confidentialité. Parmi elles, deux tiers (66.1%) ne traitent pas spécifiquement de l'application elle-même. Il est important de noter qu'il s'agit d'applications de santé, recueillant des données très personnelles des utilisateurs.

Toutes ces études montrent un problème pertinent concernant les politiques de confidentialité des services Internet. Il ressort que la plupart des utilisateurs ne les comprennent pas ou ne les lisent carrément pas. Elles sont pourtant l'élément essentiel pour la protection de la vie privée de l'utilisateur.

2.2 Vie privée : état des lieux

Même s'il en est de plus en plus question de nos jours, la protection de la vie privée n'est pas un problème récent. En 1948, 58 pays réunis à Paris adoptent la déclaration universelle des droits de l'homme dont l'article 12 stipule : « *Nul ne sera l'objet d'immixtions arbitraires dans sa vie privée, sa famille, son domicile ou sa correspondance, ni d'atteintes à son honneur et à sa réputation. Toute personne a droit à la protection de la loi contre de telles immixtions ou de telles atteintes.* »

Définir la vie privée elle-même peut se révéler parfois difficile, car elle a des significations différentes selon le contexte, le domaine d'étude et même les individus.

En 1981, Posner, définit la vie privée comme paix et tranquillité des usagers (ne pas se faire déranger par des coups de téléphone publicitaires par exemple), et également d'autonomie (Posner, 1981). Dans une tentative de définition de la vie privée sous divers axes (Lancelot-Miltgen & Gauzente, 2006), la vie privée est évoquée comme droit à être laissé seul. Cette vision se rapproche un peu de celle de l'ermitage et se veut plus restrictive que celle de Posner. Dans leur définition, ils évoquent également la vie privée comme un accès limité à un individu et aussi que le droit au secret. Dans ce dernier cas, une violation intervient lorsque le secret de l'individu est brisé. Le point central de la vie privée est la possibilité de garder un contrôle sur la dissémination de ses données d'identification (Hermalin & Katz, 2006). Il s'agit là d'une théorie assimilable à la dissimulation d'information. De nos jours, la vie privée fait aussi référence à la liberté de penser, la protection de la réputation de soi, la non invasion etc.

Bien plus qu'un droit pour les utilisateurs, la protection de la vie privée est une obligation légale. Toutefois, la restriction des lois est différente en fonction des législations et des pays. Il faut également une certaine adaptation de ces lois en fonction de l'évolution des technologies. Aux États Unis par exemple, il y a eu le Bank Secrecy Act en 1970, le Privacy Act en 1974, Right to Financial Privacy Act en 1978, Cable TV Privacy Act en 1984, Electronic Communications Privacy Act en 1994, etc.

Malgré ces législations, des écarts sont constatés tant au niveau des gouvernements que des entreprises. Suite aux attentats terroristes de Septembre 2001, le *Patriot Act* est signé en octobre de la même année. Il permet en l'occurrence l'accès aux données informatiques des citoyens sans leur autorisation préalable (Chapman, 2015).

En 2013, un évènement important a marqué le monde et éveillé la conscience des utilisateurs quant à la protection de leurs données privées. Un informaticien américain, ancien employé de la CIA (Central Intelligence Agency) et de la NSA (National Security Agency), Edward Snowden, fait des révélations troublantes sur les méthodes de surveillance et de renseignements des services secrets américains. Il publie, par le biais

des grands journaux, The Gardian et le Washington Post, des documents secrets illustrant l'accès par les services secrets aux données des plus grandes entreprises technologiques, l'interception des données transmises via les fibres optiques, etc. Il parle notamment de PRISM, un programme de vingt millions de dollars, donnant accès aux services secrets aux serveurs de grandes firmes tels que Google, Apple, Microsoft, Facebook. Les documents de Snowden montrent que ces grandes firmes ont parfois contribué à ces programmes, violant la confidentialité de leurs utilisateurs. Les services secrets investissent chaque année des millions de dollars, en collaboration avec les entreprises, pour affaiblir les protocoles de cryptographie et de sécurité, afin de rendre les accès faciles[6].

D'autres incidents en matière de confidentialité ont également fait surface depuis quelques années. Récemment en 2015, The MarketWatch révèle qu'un employé de Morgan Stanley a été licencié pour pertes de données confidentielles. Ce dernier aurait accédé illégalement aux données de 350000 clients[7]. En 2014, la grande banque JPMorgan annonce avoir été victime d'une attaque de pirates, attaque au cours de laquelle les données (noms, adresses, numéros de téléphones, courriels) d'environ 83 millions de comptes ont été divulguées[8].

L'après Snowden marque une étape nouvelle en matière de sécurité et de confidentialité des données, de même que les différents incidents qui ont lieu au cours des dernières années. Les individus sont plus soucieux de l'utilisation de leurs données par les entreprises et les gouvernements. Les grandes firmes en sont conscientes et prennent donc des mesures et des résolutions pour renforcer la confiance de leurs utilisateurs. Apple annonce qu'à partir de la version 8 du système d'exploitation iOS (utilisé sur sa marque de téléphone iPhone), il n'aurait plus accès aux mots de passe de ses utilisateurs.

[6] http://www.theguardian.com/world/interactive/2013/nov/01/prism-slides-nsa-document, accédé 02/01/2016
[7] http://www.marketwatch.com/story/morgan-stanley-fires-financial-adviser-after-data-theft-2015-01-05, accédé le 12/12/2015
[8] http://www.reuters.com/article/us-jpmorgan-cybersecurity-idUSKCN0HR23T20141003, accédé 12/12/2015

Google, pour sa part, renforce la sécurité de Gmail et annonce que la prochaine version de son système d'exploitation Android sera chiffrée par défaut.

Ces mesures prises par ces grandes firmes ont pour but de redonner confiance à leurs utilisateurs afin de ne pas perdre des parts de marchés. La vie privée a en effet une valeur économique, d'où de nombreuses études et recherches effectuées sur l'économie de la vie privée.

2.3 L'Économie de la vie privée

Tout comme la vie privée, son économie est l'objet de nombreuses études et définitions. Acquisti (Acquisti et al., 2013) définit l'économie de la vie privée comme visant à étudier les coûts et les bénéfices associés à la protection et à la divulgation des informations personnelles; pour l'individu concerné, pour le détenteur des données de même que pour la société dans son ensemble.

Beaucoup d'intérêts se portent vers cet aspect économique, après que diverses tentatives de solutions n'aient pas abouties. Une étude s'interroge sur le fait que peut être que le problème serait résolu si les utilisateurs payaient pour garder le contrôle de leur vie privée, alors que les lois et les gouvernements tardent ou ne parviennent pas à trouver des mesures adéquates (Lesk, 2012).

2.3.1 Les courants de pensées

Deux principaux courants de pensées relatifs à l'économie de la vie privée ont vu le jour.

- **Approche du libre marché**

Cette approche, soutenue principalement par l'école de Chicago, atteste que le marché des informations personnelles devrait fonctionner exactement comme celui des produits et services. L'efficacité de ce marché serait atteinte si l'information était entièrement accessible à tous les participants du marché (Calzolari & Pavan, 2006). Ainsi, certains affirment que la protection de la vie privée crée une inefficacité dans le marché (Posner, 1981). En réduisant par exemple la quantité d'informations sur les acheteurs dans le

marché du travail (les employés), cela réduit considérablement l'efficacité de ce marché et nuit donc à l'évolution des vendeurs (employeurs). Cet exemple peut être généralisé à divers types de marchés. En considérant les échanges entre deux entreprises par exemple, une étude montre que le partage d'informations personnelles des consommateurs entre deux entreprises contribue à améliorer le bien-être social (Calzolari & Pavan, 2006). Ils trouvent également que ce partage n'entrave pas l'équilibre des prix et serait donc bénéfique pour le consommateur. Dans ce même ordre d'idée, Stigler pense que l'ingérence des gouvernements dans le marché de la vie privée restera inefficace pour celui-ci (Stigler, 1980). Aucun besoin de régulation n'est donc nécessaire pour ce marché.

- **Régulation de la vie privée**

Diverses théories se sont érigées contre les pensées de l'Université de Chicago. Certains chercheurs affirment que la protection des données a des effets positifs sur le bien-être économique (Hermalin & Katz, 2006). Révéler qu'un individu fume mène à des polices d'assurances moins avantageuses. L'usage des données des utilisateurs peut entraîner des coûts à ces derniers. L'exemple de l'usage secondaire des données laissées par un utilisateur à un fournisseur contre des services peut être évoqué (Varian, 1996). Ces données seront revendues à une firme tierce qui enverrait par exemple des spams à l'utilisateur. Dans cette logique, il est aussi question de la discrimination des prix : les firmes se servent en effet des données des utilisateurs, notamment de leurs achats antérieurs, pour fixer des prix en ligne en fonction du comportement et du pouvoir d'achat de chaque consommateur (Acquisti & Varian, 2005). Ce tracking effectué par les entreprises serait, selon eux, efficace si les données recueillies des clients servent uniquement à leur proposer des services améliorés.

Toutes ces théories spéculent sur l'intérêt de révéler ou dissimuler ses données privées tant pour les individus que pour les firmes qui exploitent de ces données. Protéger ou divulguer ses informations personnelles a donc des avantages et des inconvénients tant pour les utilisateurs que pour les services en ligne.

2.3.2 Dissimuler ou divulguer : avantages et inconvénients

Nous évoluons dans un contexte où de plus en plus de services souhaitent atteindre un équilibre entre divulgation de données et services offerts (Domingo-Ferrer, 2010). Dans son étude, Acquisti (Acquisti, 2010), souligne des effets bénéfiques pour un utilisateur de divulguer ses données. Il parle en effet de bénéfices monétaires résultant d'offres promotionnelles ou de rabais sur des services dont bénéficie l'usager. Outre cela, la personnalisation des interfaces et des services permet aux usagers de mieux tirer profit des services offerts. Dans son analyse, il observe que les usagers peuvent bénéficier de réduction sur des tarifs lorsque les entreprises dépensent moins pour leur marketing. La discrimination des prix peut être avantageuse pour l'usager dans certaines conditions (Acquisti & Varian, 2005).

Divulguer ses données privées peut aussi entraîner des coûts, économiques ou non, aux usagers. Des fraudes d'identité sont possibles lorsque les usagers laissent beaucoup d'informations privées sur Internet (Brown, 2013). Ces fraudes peuvent nuire à l'usager en endommageant sa crédibilité financière, lui rendant donc un accès difficile au crédit. Il souligne également l'invasion de la vie privée par des publicités, par l'exploitation de données parfois très intimes de l'utilisateur. Ces données peuvent être liées à sa santé, sa sexualité, son historique de crédit.

Qu'il s'agisse de bénéfices ou de coûts, comment évaluer la valeur des données personnelles d'un utilisateur ? Dans la partie qui suit, il s'agira de montrer la valeur accordée par les utilisateurs à leur confidentialité.

2.4 Quelle est la valeur de la vie privée ?

Selon une étude de BCG (Boston Consulting Group) réalisée en 2013[9], le marché des données personnelles devrait atteindre un montant colossal de 330 milliards d'euros en

[9] https://www.bcgperspectives.com/content/articles/digital_economy_consumer_insight_value_of_our_digital_identity/, accédé 04/01/2016

Europe d'ici 2020, ce qui représenterait 22 % de la croissance annuelle du continent. D'après cette étude, seuls 30 % des consommateurs sont conscients de la nature des informations collectées, de ceux qui les collectent et de comment elles sont utilisées. Ils ne sont d'ailleurs pas conscient non plus des risques encourus lorsqu'ils partagent des informations sensibles sur les réseaux sociaux (Srivastava & Geethakumari, 2013).

Grâce au Global Consumer Sentiment Survey, un sondage réalisé par la même compagnie en 2013, d'environ 10 000 personnes dans 18 pays du monde, il ressort que 75 % des consommateurs ont le souci de l'utilisation faite de leurs données. Seulement 7 % des consommateurs affirment se sentir à l'aise avec l'utilisation de leurs données en dehors de l'objectif initial pour lequel elles ont été collectées. Par contre 70 % des consommateurs acceptent l'utilisation de leurs données pour leur faire des propositions marketing personnalisées. Le consommateur est alors plus à l'aise lorsqu'il perçoit également un bénéfice dû à l'utilisation de ses données par les entreprises. Les consommateurs ne se sont pas exprimés quant à la valeur de leurs données, mais il faut noter que l'intérêt accordé aux données personnelles varie selon les régions. Les consommateurs asiatiques se soucieraient moins de l'utilisation faite de leurs données comparativement aux consommateurs européens.

Ces différents sondages montrent en effet que les utilisateurs sont conscients du fait que leurs données ont une valeur. Toutefois, sont-ils en mesure de l'évaluer et dans quelles circonstances sont-ils prêts à céder leurs informations personnelles contre des incitations économiques.

2.4.1 Circonstances où l'usager échange ses données contre des incitations économiques

Une étude (dos Santos Brito et al., 2013) montre le désir des consommateurs d'être rémunéré. À la question de savoir s'ils voudraient un nouveau type de réseaux sociaux qui intégreraient des marchés de données privées, 32 % des sondés pensent que ce serait le scénario idéal, 43 % ont répondu qu'ils ne croient pas que cela pourrait arriver de nos

jours. Par contre 20 % pensent que ce serait contraignant de payer pour utiliser des services qui sont actuellement gratuits.

La volonté des consommateurs d'être rémunérés pour leurs données privée intéresse en 2007 des chercheurs qui montrent qu'un très grand nombre d'utilisateurs sont prêts à révéler leurs données contre un montant financier (Grossklags & Acquisti, 2007). Mieux encore, l'écart est vraiment considérable entre les utilisateurs acceptant de recevoir de l'argent contre ceux prêts à payer pour garder leur confidentialité.

Avec des incitations économiques, les utilisateurs sont plus disposées à partager certaines informations personnelles (Chorppath & Alpcan, 2013). Cette étude montre qu'en récompensant l'utilisateur, ce dernier partage avec plus de précision sa position géographique, ce qui améliore également les services du fournisseur de l'application mobile. Cette étude peut être généralisée à bien d'autres domaines que la géolocalisation. En effet, le manque de partage d'informations pourrait résulter à moins d'avantages tant pour le consommateur que le fournisseur de services (Osothongs & Sonehara, 2014).

Afin d'étudier la volonté qu'ont les individus de payer pour leur vie privée, une étude (Beresford, Kübler, & Preibusch, 2012) confronte un groupe d'utilisateurs à l'achat d'un DVD sur deux plateformes électroniques distinctes. Dans la première expérience, le DVD coûte 1 Euro moins cher sur la plate-forme N°1 qui exige néanmoins plus d'informations personnelles. A l'issue de cette expérience, les DVD ont plus été achetés sur la plate-forme 1 que sur la 2. Dans une seconde expérience, le DVD a le même prix sur les deux plateformes avec la plate-forme 1 qui exige toujours plus d'informations personnelles. Dans ce cas, un achat à parts égales sur les deux plateformes est observé. Il ressort de cette étude que les usagers qui s'inquiètent beaucoup de l'abus des compagnies en ce qui concerne leurs données privées, sont ceux qui les divulguent souvent pour peu ou même gratuitement.

Les résultats contrastent en fonction des études. En effet, une expérimentation quelque peu similaire (Egelman, Felt, & Wagner, 2013) réalisée un an plus tard, vise à comprendre dans quelle circonstance les utilisateurs sont **prêts à payer pour protéger**

leur vie privée. Elle étudie notamment comment le choix de l'architecture affecte la volonté des utilisateurs de smartphones d'installer ou non des applications exigeant des autorisations différentes. L'objectif est de proposer des améliorations pour les architectures de téléphones intelligents qui, de nos jours, ne supportent pas cette possibilité de payer pour restreindre l'accès à ses données. L'expérience propose l'achat de deux applications aux fonctionnalités similaires, avec l'une demandant un accès à plus d'informations privées. Les résultats montrent qu'un quart des utilisateurs ont accepté payer 1,5 dollars pour l'application demandant moins d'accès à la vie privée. Ils concluent donc que de nombreux utilisateurs sont prêts à payer pour protéger leurs données sensibles, mais que les plateformes mobiles ne leurs offrent malheureusement pas cette possibilité.

2.4.2 Peut-on estimer la valeur des données privées ?

Dans un article paru en 2013 (Times), le Financial Times fournit un outil permettant à tout internaute de calculer la valeur de ses informations personnelles. Selon le journal, les informations générales tels que l'âge, le sexe, la localisation valent environ 0.50 $ pour 1000 personnes. Certaines étapes dans la vie d'un individu représentent plus de valeur pour les entreprises, tels que le mariage, un déménagement, l'achat d'une voiture, un divorce etc. Toujours selon l'article, pour 0.26$ par personne, des entreprises peuvent avoir accès aux données sur la santé des individus. Toutes ces informations ont permis au journal de créer un calculateur de valeur de données confidentielles utilisable par tout internaute.

Une étude (Acquisti et al., 2013) cherche également la valeur accordée par les usagers à leurs données en leur proposant diverses expériences. L'objectif est de savoir à quel prix ceux-ci seraient prêts à perdre leur confidentialité ou combien seraient-ils prêts à payer pour garder celle-ci. Il en ressort que la question sur la valeur de la vie privée dépend non seulement de chaque individu, mais aussi de comment elle est posée. L'étude montre également que la valeur est fonction de la direction de l'échange: elle est plus grande

lorsque l'usager reçoit une compensation en échange de ses données, mais plus petite lorsque l'usager doit payer pour garder le contrôle de ses données privées.

Ce résultat s'illustre également dans des travaux qui font l'expérience de des ventes aux enchères pour un échange de données privées (Ghosh & Roth, 2013). Ils simulent la volonté d'acquisition de données privées par un agent qui en a fortement besoin. Les propriétaires de données ont la possibilité de faire des enchères pour parvenir à une entente de prix avec l'acquéreur. Les résultats montrent qu'il s'avère difficile d'indemniser les utilisateurs pour la perte de leurs confidentialité, vue la valeur que ces derniers accordent à leurs informations personnelles.

Même si elle n'a pas une valeur universelle, elle est désormais une monnaie d'échange à cause de son potentiel à créer de la valeur ajoutée aux entreprises et aux individus (Spiekermann, Acquisti, Böhme, & Hui, 2015). Nombreux sont ceux qui en tirent donc profit.

2.4.3 Tirer profit des données personnelles

Les utilisateurs, propriétaires des données personnelles, sont de plus en plus nombreux à désirer des récompenses en échange de leurs informations. Dans un article[10] paru en 2014, le quotidien The Gardian s'interroge sur la valeur des données privées des consommateurs, vu l'engouement vers ce commerce. Il évoque deux points de vue. Pour les uns, le partage d'informations personnelles serait le prix à payer pour l'utilisation d'Internet, un service large et gratuit. Par contre, d'autres pensent que l'utilisation des données personnelles par les entreprises est une violation de la vie privée et que les consommateurs devraient être payés en retour. Dans les lignes du journal, est évoqué le cas d'un étudiant Néerlandais qui a décidé de vendre aux enchères ses données personnelles sur Internet. Shawn Buckles aurait reçu 288£ pour la vente de ses données, allant de ses historiques de navigation aux contenus de ses mails. Le quotidien s'interroge

[10] http://www.theguardian.com/news/datablog/2014/apr/22/how-much-is-personal-data-worth, accédé le 30/12/2015

alors si tous les internautes pourraient recevoir autant de la part des entreprises qui font le commerce de données privées.

De nombreuses compagnies se sont également lancées dans le commerce de données privées. Citons ici Datacoup[11], une entreprise Américaine basée à New York. Elle rémunère d'environ 8$ ses utilisateurs pour accéder à diverses informations provenant de leurs réseaux sociaux de même que certaines informations bancaires. Ces informations sont ensuite revendues à de grandes firmes. Une autre entreprise Américaine basée à San Diego, Luth Research[12], propose environ 100$/mois pour des informations allant des transactions bancaires à la position géographique, historiques de navigations, etc.

Les gouvernements commencent également à s'intéresser à ce que pourraient leur rapporter le commerce de données personnelles. En avril 2014, le quotidien The Gardian[13] rapporte que le trésor Anglais pourrait permettre au HMRC (Her Majesty Revenue and Customs) de vendre à des compagnies privées les informations personnelles des millions de contribuables du Royaume-Uni. Ces données incluraient les revenus de particuliers, leurs taxes et historiques de paiement, un arrangement qui pourrait donc permettre d'identifier les propriétaires des données. Cette annonce a provoqué un soulèvement des sociétés de protection de vie privée. Le directeur du HMRC affirmait effectivement en 2012 qu'ils possédaient plus de données que la British Library.

Les différentes études et expérimentations présentées dans cette section montrent que les usagers ont le souci de leurs informations personnelles, mais certains sont néanmoins prêts à les révéler contre un certain gain. Toutefois, certains facteurs viennent modifier ce comportement. En 2010, une étude (Tsai, Egelman, Cranor, & Acquisti, 2011) teste une plate-forme d'achat en ligne où les politiques de confidentialité sont présentées à

[11] http://datacoup.com, accédé le 12/07/2015

[12] http://luthresearch.com/, accédé le 16/07/2015

[13] http://www.theguardian.com/news/datablog/2014/apr/22/how-much-is-personal-data-worth, accédé le 30/12/2015

l'usager de façon plus claire et compréhensible au moment de l'achat. L'expérience montre que les usagers ont plus tendance à effectuer leurs achats sur des sites qui protègent mieux leur confidentialité. Il en ressort donc que lorsque les politiques de confidentialité sont montrées plus clairement aux usagers, ceux-ci préfèrent parfois utiliser des services premiums payants pour mieux protéger leur confidentialité. Ce résultat fort intéressant montre en effet que la valeur accordée par les utilisateurs à leurs données dépend de comment la question ou l'expérience est formulée.

Cela soulève la question des politiques de confidentialité, qui comme on le voit dans l'expérience, influencent le choix de l'usager lorsqu'elles sont plus visibles et plus compréhensibles par ce dernier. Dans la partie qui suit, nous allons effectuer un état des recherches et avancées apportées dans le cadre de l'amélioration des politiques de confidentialité.

2.5 Des solutions aux politiques de confidentialité

Afin de résoudre les différents problèmes liés aux politiques de confidentialité présentées plus haut, un standard a été proposé.

2.5.1 Un standard : le P3P

Le W3C, conscient des problèmes majeurs liés aux politiques de confidentialité en standardise un modèle. Le P3P (Platform for Privacy Preferences Project) est lancé en 1997 et a pour objectif de standardiser les politiques de confidentialité des sites web. En créant un format structuré des politiques de confidentialité, il devrait désormais être possible à des logiciels (fureteurs par exemple) d'effectuer automatiquement des actions en fonction du contenu de ladite politique et des préférences de l'utilisateur. Ce format structuré peut prendre une forme compacte sous forme d'énumérations de codes ou une forme complète en format XML interprétable par un navigateur (http://www.p3ptoolbox.org/guide/section2.shtml, 2015).

Le P3P propose un vocabulaire qui permet d'exprimer les besoins suivant :

- Qui collecte les données

- Quelles données sont collectées
- Dans quels buts les données sont collectées
- Y a-t-il possibilité d'accepter ou refuser l'utilisation de certaines données ?
- Qui reçoivent les données collectées
- Quelles est la période de rétention des données
- Comment les conflits seront-ils résolus

L'exemple 1 ci-dessous illustre une section de la politique P3P déclarant que le service en ligne utilise un cookie (témoin de navigation) qui sauvegarde les informations démographiques de l'usager.

```
<DATA-GROUP>
   <DATA ref="#dynamic.cookies">
<CATEGORIES><demographic/><online/><physical/></CATEGORIES>
   </DATA>
</DATA-GROUP>
```

L'exemple 2 illustre un extrait d'une politique de confidentialité P3P spécifiant que le service en ligne utilise des cookies et que ces cookies sont associés à la politique P3P nommée « Policies.xml#cookies ».

```
<META xmlns="http://www.CatalogShopExample.com/2002/P3Pv1">
 <POLICY-REFERENCES>
   <POLICY-REF about="/P3P/Policies.xml#cookies">
     <COOKIE-INCLUDE name="*" value="*" domain="*" path="*"/>
   </POLICY-REF>
 </POLICY-REFERENCES>
</META>
```

La version du P3P faisant office de recommandation (version 1.0) remonte à Avril 2002. La version 1.1, élaborée en 2006, n'est pas entrée en application. Internet Explorer est le premier navigateur à l'avoir implémenté et l'un des seuls malheureusement. Mozilla Firefox propose un outil de gestion de cookies, permettant, entre autres, d'autoriser ou

bloquer automatiquement l'utilisation des cookies à un site web en se basant sur sa politique au format P3P.

En ce qui concerne les sites web, peu de concepteurs ont adopté le P3P. Il a été lentement implémenté par de grandes firmes du domaine de l'Internet, toutefois il ne demeure pas la priorité de la plupart des concepteurs de sites web. Il faudrait en effet à ceux-ci plus de budget et de temps pour le mettre en place vu sa complexité. Certains utilisateurs n'ont aucune idée de ce qu'est un témoin de navigation ni du rôle qu'il joue. Il semble alors difficile pour ces derniers de définir des préférences sur ce genre de données. Aussi, beaucoup s'interrogent sur l'efficacité réelle du P3P. Le principe même du P3P repose plus le fait d'obtenir le consentement de l'usager pour une collecte de données plutôt que d'aider ce dernier à les protéger (Salvas, 2002).

Un problème de confiance se pose également avec le P3P. Aucun système de contrôle n'est mis en place pour certifier les politiques des sites web et garantir que ceux-ci respectent les termes de ces politiques.

Alors que la négociation fait partie des recommandations faites par l'équipe de développement du P3P en version de 1997, elle n'est pas du tout abordée par le dernier groupe de travail en 2002. Il s'agit là d'un point important de notre étude.

Le schéma ci-dessous décrit brièvement une simple transaction http avec le P3P.

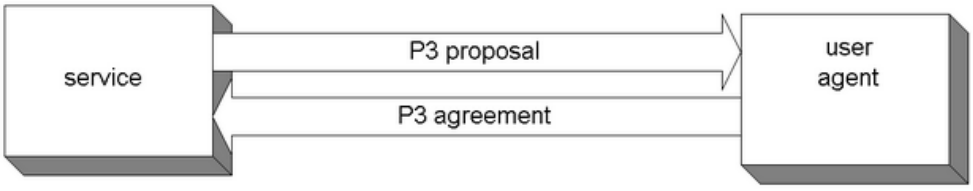


Figure 2.3 : Architecture basique du modèle P3P

2.5.2 Le langage naturel

Afin de résoudre le problème de formats non standards et de la difficulté de compréhension des politiques de confidentialité, une méthode d'extraction des pratiques faites sur les données dans les politiques de vie privée a été testée (Sadeh et al., 2014;

Schaub, Breaux, & Sadeh, 2014). Les sites et les applications mobiles sont supposés informer les internautes de l'utilisation faite de leurs données par le biais des politiques de confidentialité. Toutefois, celles-ci sont écrites dans des formats non standards, ce qui rend leur compréhension difficile par les consommateurs. Ils conçoivent dans leur article un flux de tâches permettant d'extraire, des politiques de confidentialité, l'utilisation des données privées afin d'attirer plus d'attention des utilisateurs sur l'usage de leurs données tels que le partage des contacts ou des données bancaires avec des tiers. L'extraction suit quatre étapes : l'acquisition de la politique de confidentialité, la segmentation de la politique en sections de 120 mots afin d'en faciliter la lecture par l'usager, l'extraction de l'utilisation des données par le biais de mots clés tels que « collecte », « partage avec des tiers », etc. La quatrième étape consiste en une analyse des informations extraites afin d'évaluer l'utilisation abusive ou non des données par le site Internet. Ce modèle, présenté dans la Figure 2.4, qui peut être assez pratique pour les utilisateurs, n'est cependant pas entièrement automatisé.

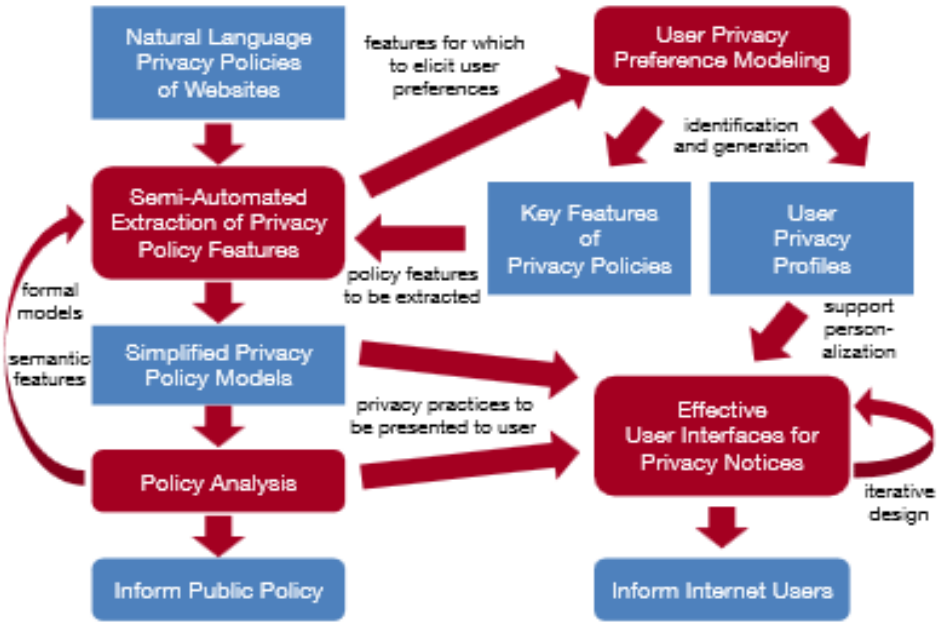


Figure 2.4 : Modèle d'extraction des pratiques des politiques de confidentialité

Une étude (Zimmeck & Bellovin, 2014) propose en 2014 un système basé également sur le traitement du langage naturel pour analyser les termes essentiels des politiques de

confidentialité et effectuer des classifications automatiques. Ils affirment que les politiques de nos jours notifient uniquement les utilisateurs et leur demandent ensuite d'indiquer un choix. Or la difficulté de les comprendre fait que de nombreux utilisateurs adhèrent aux services sans lire le contenu des politiques. Ils proposent pour cela **Privee**. Privee est un outil qui a été testé sous forme d'extension ajoutée à un navigateur Internet. Le logiciel récupère les résultats d'analyse des politiques à partir d'un référentiel disponible en ligne, et effectue des classifications automatiques si aucun résultat n'est présent dans le référentiel. Vu que le logiciel utilise des techniques de traitement du langage naturel, les résultats expérimentaux indiquent des performances limitées. Toutefois, les auteurs sont confiants quant aux potentialités de l'outil, en soulignant que l'ambiguïté du langage naturel diminuera au fur et à mesure que les politiques adopteront des formalismes communs. Il est à noter que Privee demeure un outil pratique pour faciliter les notifications aux utilisateurs et les aider dans leurs choix.

2.5.3 Autres solutions pratiques

P2U (Purpose-to-Use) (Iyilade & Vassileva, 2013) est un système développé pour permettre à l'usager d'avoir le contrôle sur le partage de ses données. En effet, avec l'émergence du « cloud » et des téléphones intelligents, les données des utilisateurs sont souvent partagées entre différentes applications mobiles ou sites Internet, afin de fournir des services personnalisés à celui-ci. Cela permet également à l'usager de ne pas saisir plusieurs fois la même information. Par exemple, lorsqu'il installe une nouvelle application, celle-ci accède à des données déjà existantes auprès d'une autre application, un gain de temps donc pour l'usager. Toutefois cela pose des problèmes de confidentialité. Il est difficile aux utilisateurs de contrôler quelles données sont partagées et entre quelles applications. Pour ce faire, P2U propose un marché dans lequel s'effectue l'échange de données. Tel que présenté dans la Figure 2.5, quatre acteurs y interviennent : *l'utilisateur*, *les fournisseurs de services* (applications qui collectent les données auprès de l'utilisateur), *les consommateurs de données* (applications qui ont besoin des données de l'utilisateur auprès d'autres applications) et le Framework qui

conduit les négociations. Les négociations portent principalement sur *les types de données* à partager, *la durée de rétention* et *le prix*, l'utilisateur pouvant en effet tirer profit de l'utilisation de ses données. Tout ceci se fait par le biais d'un langage prédéfini, permettant à l'usager de classifier ses données selon leur pertinence, et au fournisseur de service de spécifier les différents usages et conditions dans lesquelles les données sont partagées.

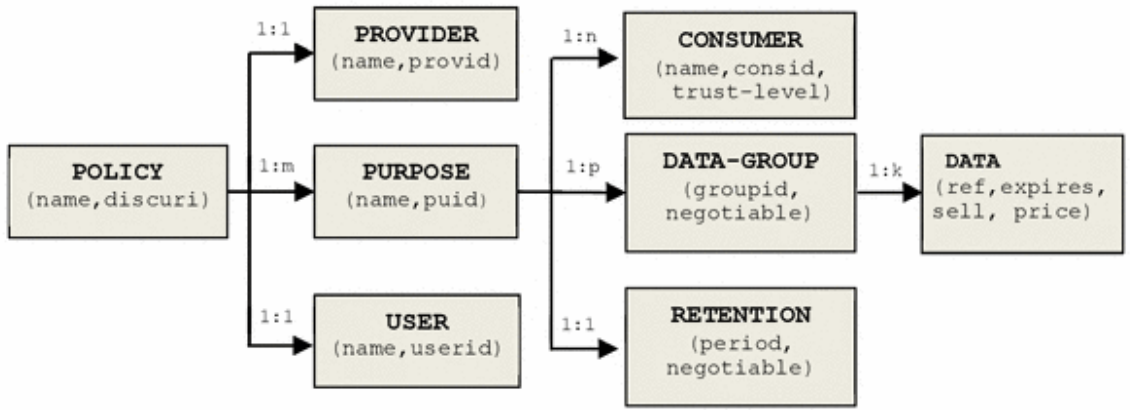


Figure 2.5 : Principaux composants de P2U

Bien qu'offrant plus de contrôle aux utilisateurs, P2U se limite uniquement au partage de données entre applications. Le premier niveau, la collecte des données, n'est pas pris en compte. Il est à noter également qu'il est difficile, voire impossible à l'usager de contrôler réellement l'usage qui est fait de ses données. En effet, les données sont partagées directement entre les applications, aucun tiers n'est présent lors de cet échange.

D'autres études s'orientent des modèles intelligents capables de générer automatiquement des politiques de confidentialité. Une parmi elles propose une solution qui analyse les habitudes de navigation de l'utilisateur pour lui proposer une politique de confidentialité personnalisée (Apolinarski, Handte, & Marron, 2015). Leur application, dont l'architecture est présentée dans la Figure 2.6, a été testée sous Android. Elle extrait les contenus partagés par l'utilisateur de même que les paramétrages de partage qui y étaient associés. L'application détecte également le contexte de partage et

propose à l'utilisateur une politique de confidentialité en fonction de ses habitudes antérieures.

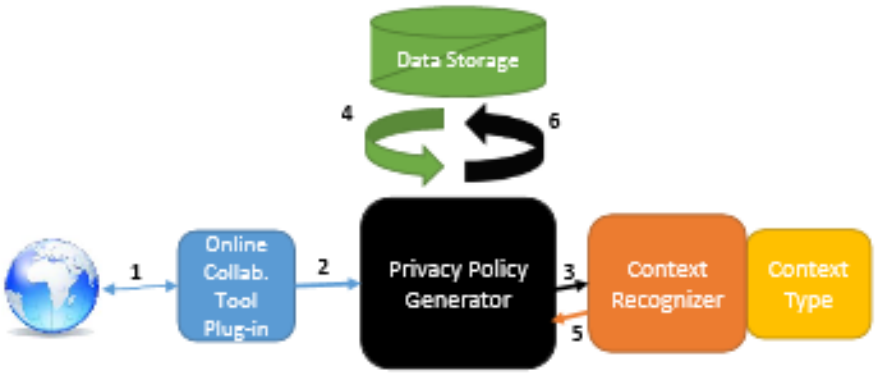


Figure 2.6. Architecture de génération de politique de confidentialité

Pouvant être avantageux pour des utilisateurs assez restrictifs sur le partage de leurs informations personnelles, cette solution ne se montre toutefois pas protectrice pour un utilisateur moins rigoureux. En effet, elle aura tendance à lui proposer des politiques assez invasive si l'utilisateur a eu des habitudes laxistes par le passé. Cela pourrait permettre à des applications d'abuser des données de certains utilisateurs.

L'une des difficultés rencontrées également par les utilisateurs est le changement et les mises à jour fréquentes de politiques de confidentialité par les sites. Afin d'aider les utilisateurs à suivre et réadapter leurs paramètres en fonction de ces changements, la compagnie **AVG** (Anti-Virus of Grisoft), vendeur de logiciels anti-virus, propose **PrivacyFix** en 2014. Il s'agit d'un logiciel gratuit permettant de gérer les paramètres de confidentialité des différents réseaux sociaux à partir d'un seul tableau de bord. Selon AVG le logiciel permettrait également de bloquer les différents logiciels qui traquent les données de navigation des utilisateurs pour leur proposer des publicités ciblées. PrivacyFix est disponible sous forme d'extension pour navigateurs ou d'applications mobiles. Il notifierait l'utilisateur dès qu'une politique de confidentialité est mise à jour et l'oriente vers le paramétrage adéquat. PrivacyFix serait également en mesure

d'informer l'utilisateur lorsqu'il se connecte à un site qui revend les données privées des utilisateurs, et serait capable de détecter lorsque les données confidentielles de l'utilisateur sont partagées pour l'en informer.

Bien qu'informant les utilisateurs à partir de notifications et d'alertes, cette solution n'offre toutefois pas un contrôle sur la collecte et l'usage de ses données, ce que nous l'envisageons dans notre étude.

2.6 Synthèse et conclusion

Les politiques de confidentialité sont la première étape à franchir pour l'inscription à un service ou l'abonnement à un site Internet. Parce qu'elles contiennent l'ensemble des pratiques effectuées par le site Internet sur les données des consommateurs, elles se doivent d'être rédigées de façon claire et compréhensible pour tous les utilisateurs. Pourtant, comme nous l'avons observé dans ce chapitre, de nombreux problèmes font que la majorité des utilisateurs ignorent leur contenu. Rappelons l'étendue et la complexité de leur contenu, qui justifient selon une étude (McDonald & Cranor, 2008), un temps minimum moyen de dix minutes pour lire une politique de taille moyenne. Il se pose également un problème de traduction des politiques dans les diverses langues des consommateurs. Peu de sites ont leurs politiques de confidentialité traduites en autant de langues que l'est le site lui-même. Nous avons également souligné le manque d'un formalisme ou d'un langage standard, qui pose parfois des ambiguïtés dans le contenu des politiques.

D'un point de vue technologique, diverses solutions ont été proposées à l'instar du P3P, un standard du W3C. D'autres se veulent plus innovatrices en effectuant automatiquement des comparaisons du contenu de la politique avec les préférences de l'utilisateur. D'autres encore proposent l'extraction d'informations clés contenues dans les politiques en se basant sur le traitement du langage naturel (Schaub et al., 2014; Zimmeck & Bellovin, 2014). Nombre de ces solutions se sont heurtées à des problèmes de complexité, de difficulté d'implémentation ou de traitement de langage naturel.

Les législations aussi ont essayé de proposer des solutions pour des politiques de confidentialités plus compréhensibles. Des associations et partis politiques s'y sont néanmoins fermement opposés en mettant en avant la liberté et l'indépendance de l'Internet par rapport aux gouvernements.

Il est certain que la confidentialité a un prix. Il a été effectivement question des pertes engendrées par les gouvernements et les grandes firmes après les révélations d'Edward Snowden. Ces événements ont éveillé la conscience des utilisateurs quant à la valeur qu'ils peuvent donner à leurs données privées. Les tendances s'orientent désormais vers une plus grande attention portée par les consommateurs quant à l'utilisation de leurs données par des firmes à des fins commerciales. On observe également plusieurs services qui proposent désormais aux utilisateurs des compensations financières contre une plus grande intrusion dans leur vie privée.

Nombreuses sont les études qui ont donc tenté d'estimer cette valeur qu'accordent les utilisateurs à leurs données. Il faut tout de même noter que cette valeur varie énormément en fonction des utilisateurs, des régions du monde, du sens des échanges, etc.

Malgré toutes ces recherches et avancées, il reste encore d'énormes progrès à faire pour trouver une solution qui serait applicable à tout type de service sur Internet et qui serait un compromis entre les utilisateurs et les fournisseurs de services. Peu de recherches se sont attardées sur la question d'offrir aux utilisateurs plus de flexibilité dans l'adhésion à une politique de confidentialité. Les modèles qui existent aujourd'hui sont basés sur une approche du tout ou rien : l'utilisateur accepte entièrement la politique ou la refuse entièrement. Aucune possibilité de négociation n'est offerte pour trouver un juste compromis entre les préférences de l'utilisateur et les besoins du service Internet.

Ces différents facteurs sont pris en compte dans notre étude pour élaborer un modèle de politique de confidentialité que nous présentons dans le chapitre suivant.

3 Chapitre 3 Méthodologie

Ce chapitre décrit en détails la solution que nous proposons afin de résoudre les différents problèmes révélés dans le chapitre 2. Nous décrivons le modèle de politique de confidentialité proposé de même que les interactions qui y figurent.

3.1 Notre approche de détermination de la valeur des données privées

L'idée d'un marché d'échange de données personnelles entre utilisateurs et services en lignes a déjà été évoquée par plusieurs recherches. Nous reprenons ici ce concept avec une approche particulière, dans le but de l'intégrer à notre solution.

3.1.1 Présentation du concept

L'idée d'un marché des données privées a été également dans une étude effectuée en 2013. Dans cette approche, le détenteur du marché récolte les données privées auprès des utilisateurs, ainsi que leurs préférences de prix et de confidentialité. Le détenteur du marché reçoit ensuite les demandes des acheteurs. Il leur fixe les prix en fonction des préférences des utilisateurs, puis récompense ces derniers (Huberman & Aperjis, 2013).

Une étude effectuée en 2015 (Gkatzelis, Aperjis, & Huberman, 2015) simule un marché dans lequel des acheteurs pourraient avoir accès à des échantillons impartiaux de données privées tout en récompensant les propriétaires des données en fonction de leurs habitudes de vie privée. Dans leur approche, l'étude montre qu'en regroupant les acheteurs en un point central, cela peut contribuer à diminuer le prix que ces derniers ont à payer pour les données des utilisateurs.

Bien d'autres modèles ont été étudiés et expérimentés. L'objectif premier de notre étude n'est pas d'établir un marché de données privées. Toutefois, nous l'intégrons dans notre solution afin de permettre aux utilisateurs de pouvoir (s'ils le souhaitent) bénéficier de l'utilisation de leurs données privées. En effet, Allan Westin classe dans son étude 55% des internautes comme des pragmatiques. Cette catégorie d'internautes analyse les bénéfices proposés en échange de leurs données ainsi que les risques encourus et les

garanties offertes. Ensuite ils prennent leur décision. Dans une étude similaire, Jensen, Potts et Jensen les estiment à 43%. Basée sur des variables descriptives notamment les variables sociodémographiques, les variables d'expérience, les variables attitudinales et les variables comportementales, Caroline Lancelot-Miltgen effectue également une classification des internautes (Lancelot-Miltgen & Gauzente, 2006). Dans ses résultats, elle classe 18.3% des participants comme des négociateurs. Ces négociateurs, représentant le second plus grand groupe sur les 4 qu'elle obtient, considèrent l'information comme une monnaie d'échange et ne les révèlent donc que pour une raison valable. Elle montre dans ses analyses que ce groupe fait partie des plus sensibles en termes de perception de risques liés à la divulgation des données personnelles.

Il s'avère donc important d'établir un cadre pour permettre à cette catégorie d'internautes, jugée le plus à risque, de mieux se protéger. En effet, les attitudes de divulgation des données diffèrent selon les utilisateurs et les services en ligne pourraient adapter leur pratique de confidentialité en fonction des attitudes de ceux-ci (Knijnenburg, Kobsa, & Jin, 2013).

Ainsi, notre conception est proche de celles présentées ci-dessus :

- **L'utilisateur** définit ses **préférences** pour chacune de ses données. Le prix pour l'utilisation des données fait partie des préférences de l'utilisateur.
- **Le service en ligne** définit également ses **besoins** en données privées, en fixant un prix pour certains types d'utilisation de données.
- **Un agent de régulation** est chargé de mener les **négociations** entre les deux parties afin de trouver le prix idéal pour chacune des parties. L'agent de régulation est en mesure de calculer une valeur pour une donnée en fonction de divers paramètres.

La difficulté dans cette approche est donc de déterminer le véritable prix d'une donnée personnelle. En effet, il pourrait y avoir un grand écart entre le prix fixé par un usager et

celui du service en ligne. Dans la partie qui suit, nous proposons une approche pour déterminer la valeur des données privées en fonction de divers paramètres.

3.1.2 Estimation de la valeur des données privées

Comme nous l'avons évoqué dans le chapitre précédent, la valeur des données personnelles dépend des usagers eux-mêmes, des services en ligne, mais aussi de la direction de l'échange (Acquisti et al., 2013), les régions du monde, etc. Dans le modèle que nous proposons, les utilisateurs ont la possibilité de recevoir des récompenses en fonction de l'utilisation qu'ils autorisent de leurs données privées. Comment évaluer alors la valeur de la donnée de l'utilisateur afin de lui attribuer ses récompenses ? Nous voulons donc que notre système soit en mesure d'offrir un échange plus ou moins équitable entre utilisateurs et services en ligne.

Différentiation des approches

Spécifions la différence de notre approche de celles évoquées plus haut. Dans ces concepts, le détenteur du marché reçoit les données de nombreux utilisateurs. À la demande d'un ou de plusieurs acheteurs, il effectue une sélection des utilisateurs dont les données seront revendues aux acheteurs. Cette sélection est basée sur divers critères tels que les préférences des utilisateurs, les prix qu'ils ont fixés etc. Dans cette logique, les utilisateurs fixant les prix les plus bas sont ceux qui ont le plus de chance de vendre leurs données aux acheteurs.

Dans notre approche, il n'y a aucune compétition entre les utilisateurs pour vendre leurs données aux services en ligne. Les négociations sont menées entre un et un seul utilisateur et un et un seul service en ligne. Pour cela nous allons nous baser sur un principe de détermination de prix en fonction des préférences de confidentialité de l'utilisateur, et non en fonction de l'offre et de la demande.

Principe de détermination de la valeur des données privées : Analyse Multicritère Hiérarchique (AMCH) (Saaty, 2008)

Afin de déterminer la valeur des données privées des utilisateurs, nous procédons à l'évaluation, par ces derniers, de l'importance qu'ils accordent à chacune de leurs données à partir d'une méthode d'analyse multicritères. Nous nous sommes orientés vers ce choix afin de répondre à la problématique évoquée plus haut. En effet, habituellement, la valeur attribuée aux données personnelles dépend du rendement qu'elles rapportent aux services en ligne, ou de la quantité de la demande sur le marché des données personnelles. Entre autres avantages de la méthode AMCH, sa capacité de structurer un problème complexe de façon hiérarchique. Elle est employée dans divers domaines. En 2001, elle a servi à déterminer le site de réinstallation de la ville Turque Adapazari dévastée par un tremblement de terre. Entre autres exemples, elle a également servi à l'État de Caroline du Nord aux États-Unis pour élaborer des critères d'évaluation et assigner des notes aux fournisseurs, menant à la sélection d'un meilleur rapport qualité-fournisseur acceptable pour les décideurs (Saaty, 2008).

Par cette méthode, nous permettons aux utilisateurs d'avoir une valeur de leurs données personnelles, en fonction de leurs préférences et de l'importance qu'ils accordent à ces données.

Dans son étude (Saaty, 2008), Thomas Saaty définit le processus d'analyse hiérarchique (AHP) comme une théorie de la mesure à travers des comparaisons par paires. Des échelles prioritaires sont établies à partir de jugements d'experts. Les mesures sont faites en se basant sur la façon dont ils servent aux décideurs.

Le processus de prise de décision, tel que défini par Saaty, peut se décomposer en 4 étapes :

- **Définir le problème** et déterminer le type de connaissance recherché
- **Construire une hiérarchie** avec l'objectif à atteindre au sommet, et les objectifs intermédiaires dans les nœuds et sous-nœuds.
- **Construire un ensemble de matrices de comparaisons par paires**. Chaque élément d'un niveau supérieur est utilisé pour comparer des éléments dans le niveau immédiatement inférieur par rapport à lui.

- **Utiliser les priorités obtenues** à partir des comparaisons pour donner un poids aux priorités dans le niveau immédiatement inférieur. Faire cela pour chaque élément. Ensuite, pour chaque élément dans le niveau inférieur, ajouter ses poids et obtenir sa priorité globale.

Suivons ce cheminement en 4 étapes dans le cas de notre étude.

Étape 1 : Définition du problème et détermination du type de connaissance recherché

Le problème est, comme nous l'avons souligné, la détermination de la valeur des données privées des utilisateurs en fonction de l'importance qu'ils y accordent. Le type de connaissance auquel nous voulons aboutir est une pondération des données de l'utilisateur, selon le degré de risque d'atteinte à sa vie privée lorsque la donnée en question est compromise. Plus le degré de pondération est élevé, plus la donnée a de l'importance pour l'usager et plus elle aura de la valeur sur le marché d'échange.

Étape 2 : Hiérarchie des objectifs

Il existe différentes méthodes pour une prise de décision. La plus répandue est la méthode BOCR (Bénéfices, Opportunités, Coûts, Risques). T.Saaty évoque également d'autres méthodes notamment SWOT (Strengths, Weaknesses, Opportunities and Threats).

La méthode **BOCR**, que nous utilisons ici, consiste à voir l'objectif global sous quatre points de vue différents. L'objectif que nous fixons ici est la révélation d'une donnée par un utilisateur. La pondération qu'il donnera aux différents sous-critères permettra de donner une valeur au désir ou non de l'utilisateur de révéler cette donnée.

L'objectif principal est subdivisé donc en 4 branches, elles-mêmes décomposées en sous-branches :

- ***Bénéfices*** : Qualité du service, personnalisation, rémunérations financières (Acquisti, 2010)
- ***Opportunités*** : Services exclusifs, Rabais et offres promotionnelles
- ***Coûts***

- ***Risques*** : Invasion de la vie privée, Vol d'identité (Brown, 2013), discrimination des prix (Acquisti, 2010)

La Figure 3.1 décrit l'arbre de prise de décision relatif à notre objectif, identifié à l'étape 1.

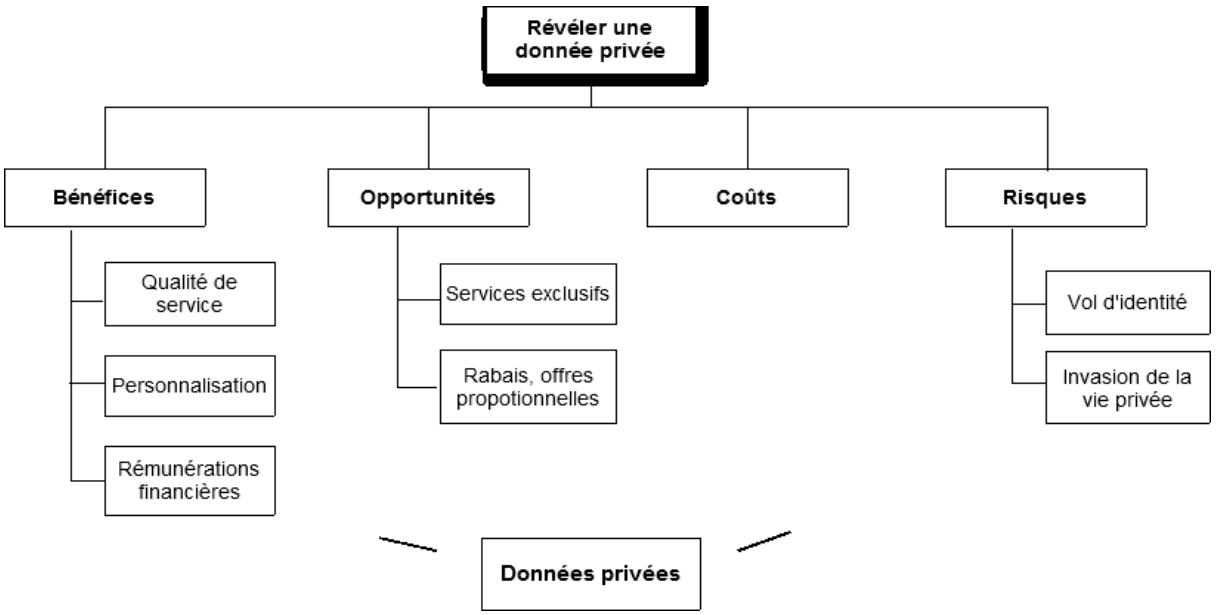


Figure 3.1. Arbre de prise de décision

Ce schéma devrait donc être appliqué à chaque donnée privée. Afin de simplifier le processus (vu le nombre sans cesse croissant de données privées possibles), nous allons dans la partie qui suit, classifier les données privées en groupes cohérents.

Classification des données privées

La norme ISO/IEC 27001, norme la plus répandue parmi les normes de management de la sécurité de l'information, spécifie que la classification de l'information est importante dans le but de la protéger.

Pour ce faire et afin d'effectuer l'analyse multicritères hiérarchique, nous allons procéder à la classification des données privées des utilisateurs. Cette classification va permettre aux usagers de donner une pondération par groupe de données, non pas pour chacune de leurs données privées.

Différentes études proposent une classification des données privées en prenant en compte des critères différents (Petkos, Papadopoulos, & Kompatsiaris, 2015; Rochelandet, 2010). Fabrice Rochelandet (Rochelandet, 2010) propose une classification des données personnelles en 2 grandes catégories : les données objectives et les données subjectives. Une étude (Petkos et al., 2015) réalisée récemment propose également une classification en huit catégories : données démographiques, psychologiques, profil sexuel, attitudes politiques, croyances religieuses, facteurs et conditions de santé, localisation, profil de consommateurs.

Après comparaison des deux classifications, nous avons opté pour celle de Rochelandet pour diverses raisons. La seconde étude ne comporte pas les données d'identification d'un utilisateur, que nous trouvons très utiles dans notre étude. Le profil financier de l'utilisateur n'y apparait pas non plus. La première classification inclut toutes les données de la seconde et se veut plus complète. Elle effectue un excellent regroupement des données des utilisateurs d'un point de vue économique. Cela correspond donc aux objectifs de notre étude.

Tableau 3.1. Classification des données privées

Données objectives

Données d'Identification et de Contact (DIC)	Identité civile; adresse postale; numéro de téléphone; numéro d'immatriculation du véhicule; numéro du permis de conduire; adresse électronique; références bancaires; numéro de carte bancaire; numéro de sécurité sociale; pseudonyme sur Internet…

Données Physiques et de Signalement (DPS)	Taille; poids; couleur des yeux; état de santé; maladies contagieuses; vaccins; empreintes digitales; photographie; voix…
Données Sociodémographiques (DSD)	Date de naissance; âge; genre; statut marital; nombre d'enfants; niveau d'études; profession…
Données Juridiques (DJ)	Capacité juridique; amendes; casier judiciaire…
Données Financières et Foncières (DFF)	Solde du compte bancaire; crédit et dettes; fréquence des impayés; interdits bancaires; allocations sociales; pensions alimentaires versées/reçues; propriété foncières; hypothèques…
Données Subjectives	
Préférences et Centres d'Intérêts (PCI)	Orientation sexuelles; préférences alimentaires, vestimentaires et culturelles; destinations touristiques préférées…
Opinions et Activités Politiques, Religieuses et Syndicales (PRS)	Opinions politiques émises publiquement; adhésion à un syndicat; croyances intimes déclarées ou manifestées publiquement; signature d'une pétition…
Données Comportementales (DC)	Apparence; types d'chats et établissements fréquentés; pratiques et fréquences des loisirs; consommation

	d'électricité et d'eau; utilisation d'un service public; données de connexion et navigation sur Internet; consommation d'alcool, tabac et drogues…
Données Géographiques (DG)	Lieu de retrait d'argent; identification Bluetooth; position GPS…
Données Relationnelles (DR)	Taille du réseau social; fréquence des rencontres; nombre d'amis, de partenaires sexuels; participation à un club ou à une association; appels téléphoniques; emails; SMS reçus et envoyés…

Une fois la hiérarchie des objectifs et des données privées réalisée, nous pouvons procéder à l'étape suivante de l'analyse.

Étape 3 : Construire un ensemble de matrices de comparaisons par paires. Chaque élément d'un niveau supérieur est utilisé pour comparer des éléments dans le niveau immédiatement inférieur par rapport à lui.

Les matrices suivantes doivent être élaborées à cette étape :

- 1 matrice pour les 4 critères par rapport à l'objectif global (Exemple dans le Tableau 3.2)
- 3 matrices pour les sous critères : 1 pour ceux de « Bénéfices », 1 pour ceux de « Opportunités » et 1 pour ceux de « Risques ». Le Tableau 3.3 présente un exemple de l'une de ces matrices.
- En considérant n le nombre de groupes de données privées (toutes représentées dans le schéma par « Données privées » pour des raisons de lisibilité), 8n matrices pour chaque groupe de données privées. Dans notre schéma, 8 représentent tous

les sous critères et critères qui n'en n'ont pas. Dans notre cas précis, nous aurons 80 matrices en considérant les 10 groupes de données privées décrites ci-dessus. Le Tableau 3.4 présente un exemple de l'une de ces matrices.

Les valeurs numériques dans les matrices représentent le degré auquel un élément est important par rapport à un autre, à l'égard du critère auquel ils sont comparés. La valeur 1 signifie par exemple que les 2 éléments s'équivalent, 5 signifierait qu'un élément à une grande importance par rapport à l'autre et 9 que l'importance est la plus grande et évidente possible. L'échelle va donc de 1 à 9. Nous comprendrons donc que si un individu attribue 5 à l'élément A par rapport à l'élément B, il est automatique que la comparaison de B par rapport à A vaudra 1/5.

Tableau 3.2 Matrice de comparaison par paire des principaux critères à l'égard de l'objectif

	Bénéfices	Opportunités	Coûts	Risques	Priorités
Bénéfices	1	3	4	1/3	**0.25**
Opportunités	1/3	1	3	1/5	**0.12**
Coûts	1/4	1/3	1	1/8	**0.06**
Risques	3	5	8	1	**0.57**

Tableau 3.3 Matrice de comparaison par paire des sous critère à l'égard du critère "Bénéfices"

	Qualité de service	Personnalisation	Rémunération financière	Priorités
Qualité de service	1	5	3	**0.60**
Personnalisation	1/5	1	1/6	**0.08**
Rémunération financière	1/3	6	1	**0.32**

Tableau 3.4 Matrice de comparaison de toutes les alternatives à l'égard du sous-critère "Rémunérations financières"

	DIC	DPS	DSD	DJ	DFF	PCI	PRS	DC	DG	DR	Priorités
DIC	1	6	5	1	2	8	3	4	3	2	**0.21**
DPS	1/6	1	4	1/4	1/5	3	1/4	1/3	1/3	1/3	**0.04**
DSD	1/5	1/4	1	1/3	1/5	3	1/3	1/2	1/4	1/4	**0.03**
DJ	1	4	3	1	1/2	3	2	3	3	3	**0.15**
DFF	1/2	5	5	2	1	3	2	5	4	4	**0.20**
PCI	1	1/3	1/3	1/3	1/3	1	1/2	1/3	1/2	1/4	**0.05**
PRS	1/3	4	3	1/2	1/2	2	1	1	1/2	1/3	**0.07**
DC	1/4	3	2	1/3	1/5	3	1	1	1/3	1/2	**0.06**
DG	1/3	3	4	1/3	1/4	2	2	3	1	1/2	**0.08**
DR	1/2	3	4	1/3	1/4	4	3	2	2	1	**0.11**

Pour déterminer les priorités dans chaque matrice, trois calculs sont nécessaires :

- Faire la somme de tous les éléments de chaque colonne de la matrice

Soient $M = (a_{i,j})$ la matrice de comparaison, sa dimension notée (m, n), C_j la somme des éléments de la colonne j.

$$C_j = \sum_{k=1}^{m} a_{k,j}$$

- Normaliser la matrice en divisant chaque entrée de la matrice par le total de sa colonne

Notons M' = $(a'_{i,j})$ la matrice résultant de la normalisation (Voir tableau 3.5)

$$a'_{i,j} = \frac{a_{i,j}}{C_j}$$

- Faire une moyenne arithmétique des lignes de la matrice normalisée

Notons p_i la priorité correspondante à chaque ligne i de la matrice normalisée.

$$p_i = \frac{1}{n}\sum_{k=1}^{n} a'_{i,k}$$

Tableau 3.5. Calcul de priorités de la comparaison de toutes les alternatives à l'égard du sous-critère "Rémunérations financières"

											Priorité ($\boldsymbol{p_i}$)
DIC	0.19	0.20	0.16	0.16	0.37	0.25	0.20	0.20	0.20	0.16	**0.21**
DPS	0.03	0.03	0.13	0.04	0.04	0.09	0.02	0.02	0.02	0.03	**0.04**
DSD	0.04	0.01	0.03	0.05	0.04	0.09	0.02	0.02	0.02	0.02	**0.03**
DJ	0.19	0.14	0.10	0.16	0.09	0.09	0.13	0.15	0.20	0.25	**0.15**
DFF	0.09	0.17	0.16	0.31	0.18	0.09	0.13	0.25	0.27	0.33	**0.20**
PCI	0.19	0.01	0.01	0.05	0.06	0.03	0.03	0.02	0.03	0.02	**0.05**
PRS	0.06	0.14	0.10	0.08	0.09	0.06	0.07	0.05	0.03	0.03	**0.07**
DC	0.05	0.10	0.06	0.05	0.04	0.09	0.07	0.05	0.02	0.04	**0.06**
DG	0.06	0.10	0.13	0.05	0.05	0.06	0.13	0.15	0.07	0.04	**0.08**
DR	0.09	0.10	0.13	0.05	0.05	0.13	0.20	0.10	0.13	0.08	**0.11**

Dans le 1er tableau de comparaison, nous observons que le facteur qui a la plus grande priorité pour l'utilisateur est le facteur « Risques ». De même, dans le 3ème tableau de comparaison, nous pouvons observer que relativement à la rémunération financière, les facteurs qui ont le plus d'importance pour l'usager sont les Données d'Identification et de Contrôle (DIC) et les Données Financières et Foncières (DFF). On obtient donc une hiérarchie des données des utilisateurs en fonction de la valeur (priorité) qu'il y accorde.

- **Étape 4** : Utiliser les priorités obtenues à partir des comparaisons pour donner un poids aux priorités dans le niveau immédiatement inférieur. Faire cela pour chaque élément. Ensuite, pour chaque élément dans le niveau inférieur, ajouter ses poids et obtenir sa priorité globale.

3.1.3 Un danger pour la vie privée ?

L'objectif de cette étude n'est pas, bien entendu, de vulgariser le commerce des données personnelles par les utilisateurs. Toutefois, nous pensons que mettre en place un système où les services en ligne auraient à payer pour collecter les données des usagers réduiraient considérablement la collecte et l'usage abusifs de données personnelles. Il est probable que de nombreux utilisateurs voudront y tirer de nombreux intérêts. Toutefois, comme nous l'avons évoqué plus haut, plusieurs études (Lancelot-Miltgen & Gauzente, 2006) ont décelé un important pourcentage d'internautes révélant facilement leurs données personnelles pour peu ou rien. Un cadre réglementé permettrait donc de mieux protéger ces utilisateurs ou de leur permettre de mieux contrôler l'échange de données avec les services en ligne.

3.2 Le modèle de politique de confidentialité

En s'inspirant des travaux du P3P et en intégrant plus de contrôle et de simplification, nous proposons un modèle de politique de confidentialité répondant à la majeure partie des problèmes identifiées dans ce mémoire.

3.2.1 Une vision globale du système

Une vue globale du modèle issu de notre travail est représentée dans la Figure 3.2. Les apports essentiels sont *l'introduction d'une autorité de régulation* ainsi que de *contrats* signés constituant la politique de confidentialité personnalisée pour chaque utilisateur. Ce modèle intègre également la certification préalable des politiques de confidentialité, permettant de garantir le respect des clauses par le service en ligne de même que des sanctions éventuelles en cas de violation. Ceci est une faille non encore résolue du P3P. Ce modèle illustre donc un scénario dans lequel les préférences de l'utilisateur de même que les besoins du service en ligne sont envoyés à une autorité de régulation. Ce dernier dispose d'un négociateur chargé de conduire les discussions et d'un recommandeur en mesure de générer des recommandations à l'utilisateur. Des ajustements peuvent être faits en cas de conflits et le contrat final est envoyé aux 2 parties. L'utilisateur peut donc

fournir ses données au service en ligne conformément aux clauses contenues dans le contrat de confidentialité.

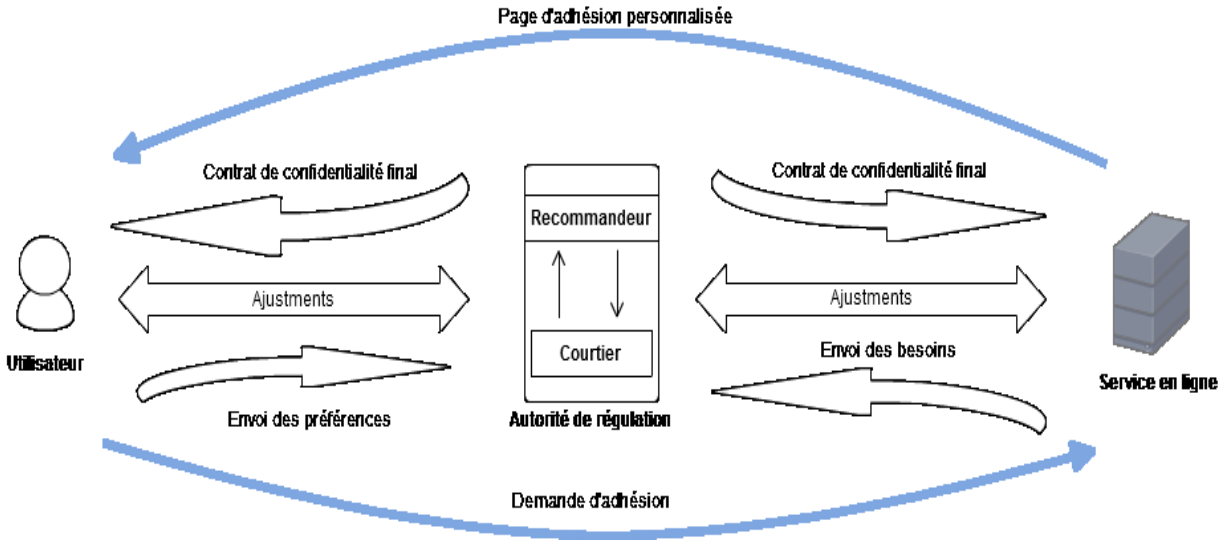


Figure 3.2: Modèle de la politique de confidentialité

3.2.2 Modèle détaillé

La Figure 3.3 décrit les différents processus ainsi que leurs enchainements.

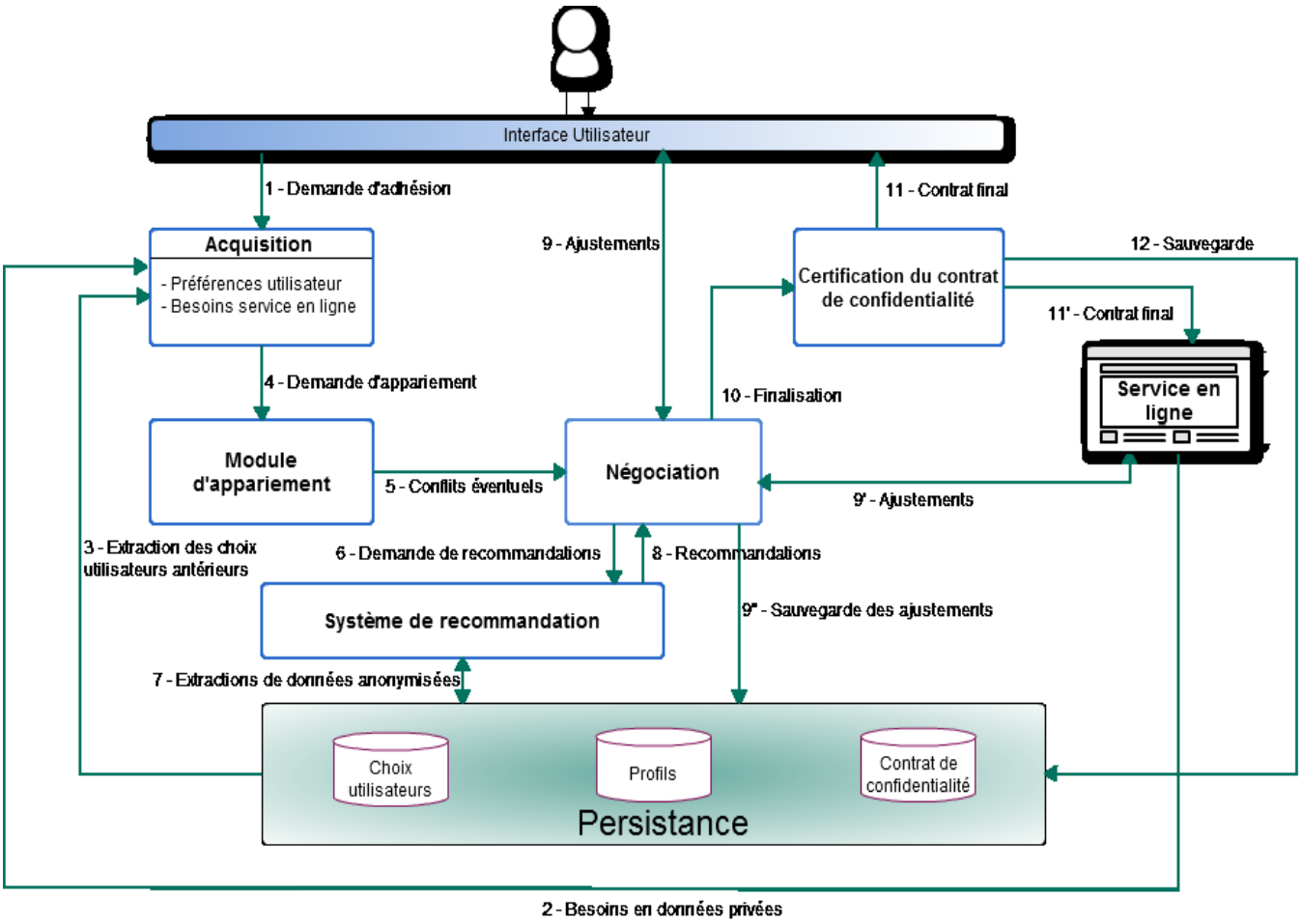


Figure 3.3 : Modèle détaillé de la politique de confidentialité

Toute personne utilisant ce service procède dans un premier temps à un paramétrage de ses préférences de confidentialité.

Lorsqu'il désire adhérer à un service en ligne (**1**), ses préférences sont envoyées à au système **Priv-C** ainsi que, éventuellement, ses choix antérieurs (**3**). Le service en ligne envoie également ses besoins (**2**). Une fois les informations obtenues et validées, le module d'appariement effectue une comparaison des entrées dans le but d'établir un contrat entre l'utilisateur et le service en ligne (**4**). Cette phase d'appariement peut générer des conflits (**5**), qui sont pris en charge par le module de négociation. Celui échange avec le module de recommandations (**6**, **7**, **8**) pour obtenir des suggestions à faire à l'utilisateur. Il génère de nouvelles options, et effectue la négociation entre

l'utilisateur et le service en ligne jusqu'à résolution ou non du conflit (**9**, **10**). Une fois tous les conflits traités, un contrat, s'il y a lieu, est généré, certifié et envoyé aux différents acteurs (**11**, **11'**).

Dans les sections suivantes, nous donnons des détails sur les entrées et les différents processus présentés dans notre modèle ci-dessus.

3.3 Les entrées

Le système reçoit en entrée les préférences de l'utilisateur de même que les besoins du service en ligne. Ces deux fichiers respectent une nomenclature normalisée que nous décrivons ci-dessous.

3.3.1 Les préférences de l'utilisateur

Les préférences de l'utilisateur sont un ensemble de paramétrages portant sur chacune de ses données privées. Ces paramétrages contiennent essentiellement :

- **La donnée** : il s'agit de la donnée privée sur laquelle porte un paramétrage précis.
- **L'usage** : elle désigne l'objet de l'utilisation de la donnée personnelle. Elle peut prendre les valeurs suivantes :
    - *La qualité du service* : la donnée de l'utilisateur peut être utilisée uniquement dans le cadre de l'amélioration du service auquel il adhère. Le service n'a pas le droit de partage sur la donnée. Par exemple recueillir sa position géographique uniquement pour lui indiquer un chemin sur une carte.
    - *Usage commercial* : la donnée peut être utilisée à des fins commerciales et uniquement par le service auquel l'utilisateur adhère. Le service n'a pas le droit de partage sur cette donnée. Par exemple, la position géographique pourrait servir à proposer à l'usager la publicité d'un magasin proche de lui.
    - *Partage pour qualité du service* : la donnée peut être partagée avec d'autres services uniquement dans le but de la qualité du service. Par

exemple, un magasin pourrait partager la position géographique d'un de ses clients avec une application de géolocalisation afin que cette dernière propose un itinéraire au client pour venir au magasin.

- *Partage pour usage commercial* : la donnée peut être partagée avec d'autres services à des fins commerciales. Par exemple, un réseau social peut partager les intérêts d'un utilisateur avec d'autres services de vente, afin de lui proposer des articles qui y sont relatifs.
- *Partage pour raisons légales/pénales* : la donnée peut être partagée avec des institutions gouvernementales. Par exemple, lorsque la loi le permet, les habitudes de consommation d'un utilisateur peuvent être divulguées à la justice pour des raisons pénales.

- **L'action** : l'action désigne ici la décision de l'usager. Elle peut prendre les valeurs suivantes :
  - *Accepter* : l'utilisateur donne son accord pour que la donnée personnelle soit utilisée pour un usage donné.
  - *Refuser* : l'utilisateur refuse l'accès à sa donnée pour un usage précis.
  - *Négocier* : l'utilisateur est prêt à négocier à propos d'un usage précis.
- **Les conditions** : l'utilisateur peut définir des conditions à respecter avant que l'action qu'il a définie pour un usage donné ne soit appliquée.

  Les conditions sont sous la forme **Si *Expression* Alors *Action.***
- **Les exceptions** : l'utilisateur peut définir des exceptions auxquelles l'action qu'il a définie pour un usage donné ne soit pas appliquée.

  Les exceptions se présentent sous la forme : **À Moins que *Expression* Faire *Action***
- **La valeur de la donnée personnelle** : l'utilisateur définit la valeur estimée de sa donnée personnelle.
- **Les critères de résolution de conflits** : l'utilisateur définit les critères de résolution de conflits. Un conflit se produit lorsqu'il n'y a pas concordances entre

les préférences de l'utilisateur et les besoins du service en ligne. Par exemple, lorsque l'utilisateur refuse un usage précis d'une de ses données alors que celle-ci est définie comme étant obligatoire par le service en ligne. La résolution des conflits peut prendre les formes suivantes :

- *Demander explicitement mon avis* : une liste de proposition sera envoyée à l'usager pour qu'il y fasse un choix.
- *Suivre le recommandeur* : le choix de l'usager sera celui recommandée par le Recommandeur. Aucune question ni liste d'options ne sera donc proposée à l'utilisateur.
- *Choisir l'option avec le maximum de points* : le choix de l'usager sera celui qui lui rapporte le maximum de points dans la liste des options. Aucune question ni liste d'options ne sera donc proposée à l'utilisateur.
- *Je ne change pas de position sur ma décision (au risque de ne pas utiliser le service)* : l'utilisateur maintient sa position. Dans ce cas il ne peut avoir contrat et donc l'utilisateur n'accèdera pas au service.

Pour illustrer ce formalisme, voici, dans le Tableau 3.6 un exemple de préférences pour la position géographique de l'usager **Bob**.

Tableau 3.6. Exemple de préférences de l'usager Bob

Donnée	**Usage**	**Action**	**Conditions**	**Exceptions**	**Critères conflit**	**Valeur**
Position géographique	Qualité de service	Accepter	Aucune	Aucune	Aucune	15
	Commercial	Refuser	Si recompense > valeur Alors Accepter	À moins que non accès à tous les services	Suivre le recommandeur	25
	Partage pour qualité	Accepter	Si Duree Rétention > 90 jours Alors Refuser	Aucune	Suivre le recommandeur	15
	Partage commercial	Refuser	Aucune	À moins que non accès à au moins un service	Suivre le recommandeur	30

	Partage pour raison légale	Accepter	Aucune	Aucune	Aucune	0

Dans cet exemple Bob spécifie comment il voudrait que sa position géographique soit utilisée par le service en ligne. Bob aimerait que sa position géographique soit utilisée uniquement pour des raisons de qualité de service. Toutefois, dans le cas où la donnée doit être partagée avec une autre entité pour la qualité du service, il accepte à la condition que cette entité ne conserve pas l'information plus de 90 jours (Condition *Si Durée Rétention > 90 jours Alors Refuser*).

Dans le cadre commercial, Bob n'aimerait pas que sa donnée soit utilisée. Il ajoute toutefois une exception où sa position géographique pourrait être utilisée pour des raisons commerciales. L'exception stipule en effet que la donnée ne doit pas être utilisée pour un partage commercial excepté si cela le prive de l'accès à toutes les fonctionnalités du service en ligne.

La valeur des données dans la dernière colonne n'est pas fournie par Bob. C'est le système qui la génère, suite à l'analyse multicritères hiérarchique.

Tout comme les utilisateurs, les services en ligne disposent aussi d'un cadre pour l'expression de leurs besoins en données personnelles.

3.3.2 Les besoins du service en ligne

Le service en ligne définit ses besoins en données privées sous forme de paramètres préétablis. Ils comprennent essentiellement :

- **La donnée** : il s'agit de la donnée privée sur laquelle porte un paramétrage précis.
- **L'usage :** (idem que dans les préférences de l'utilisateur – voir 3.3.1)
- **Contraintes sur données :** une contrainte désigne la nécessité ou non d'obtenir une donnée fonctionnelle pour effectuer un usage précis. Par exemple, la position

géographique pourrait être une donnée obligatoire pour une application de géolocalisation alors qu'elle serait optionnelle pour un réseau social.

La contrainte peut prendre deux valeurs différentes :

- *Obligatoire* : l'utilisateur est obligé de fournir la donnée pour utiliser le service.
- *Optionnelle* : l'utilisateur peut fournir ou non la donnée.

- **La durée de rétention** : la durée de rétention désigne la période pendant laquelle le service en ligne conserve la donnée de l'utilisateur dans ses fichiers. Elle peut prendre les valeurs suivantes :
  - *Pendant l'utilisation du service* : la donnée n'est conservée que lorsque l'utilisateur est connecté au service.
  - *X jours* : la donnée est conservée pendant un nombre X de jours que précise le service en ligne.
  - *Jusqu'à la suppression du compte* : la donnée est conservée par le service en ligne jusqu'à la suppression de son compte par l'utilisateur.
  - *Sur demande* : la donnée est toujours conservée par le service en ligne à moins que l'utilisateur ne fasse une demande explicite de suppression.
  - *Jamais* : la donnée est conservée par le service en ligne même après suppression de son compte par l'utilisateur
- **Les options** : les options sont les possibilités offertes à l'utilisateur sur chaque usage de la donnée. Ce dernier aura donc à choisir l'une d'entre elles à partir des « actions » qu'il aura définies dans ses préférences.
- **Data value :** le service en ligne définit la valeur estimée de la donnée personnelle de l'utilisateur.
- **Les critères de résolution de conflits** : le service en ligne définit les critères de résolution de conflits. Ces critères sont de façon à avoir le moins d'interactions possible – voire aucune – entre le service en ligne et le système. Pour donc

permettre ce minimum d'interactions, ces critères contiennent également un ensemble d'options acceptables par le service en ligne en cas de négociation.

Nous illustrons, dans le tableau 3.7, les besoins d'un service en ligne pour la position géographique d'un utilisateur.

Tableau 3.7. Exemple de besoins d'un service en ligne

Donnée	**Usage**	**Contrainte**	**Rétention**	**Options**	**Valeur**	**Critères conflits**
Position géographique	Qualité de service	Obligatoire	Fermeture du compte	Accepter	Aucune	Ok si duree ≤ duree_sur_marché Ok si Choix recommandeur
			Demande de l'utilisateur	Accepter	50	
			Jamais	Accepter	Badge Premium	
			Aucune	Refus	Non accès à service TOUS	
	Commercial	Optionnel	Fin de la session	Accepter	30	
			Fermeture du compte	Accepter	80	
			Aucune	Refus	-10	
	Partage pour qualité	Optionnel	Fin de la session	Accepter	Accès service Partage photos	
			Fermeture du compte	Accepter	30 + Accès service partage photos	
			Aucune	Refus	Aucune	
		Optionnel	Fin de la session	Accepter	50	

	Partage commercial		Fermeture du compte	Accepter	100	
			Aucune	Refus	Aucune	
	Partage pour raison légale	Obligatoire	Aucune	Accepter	Aucune	

Les besoins du service en ligne portent la position géographique de l'utilisateur. Pour chaque usage de la donnée, le service définit plusieurs périodes de rétention possibles. Chaque période de rétention donne droit à des récompenses ou non. Enfin, dans la section « Résolution de conflits », le service en ligne définit les options qu'il pourrait accepter si jamais un conflit se produisait et que l'utilisateur n'acceptait aucune des options qu'il propose dans la section des options.

Cette dernière information se présente sous la forme : **Ok si *Expression*.**

Exemple 1 : Ok si valeur ≤ valeur_sur_marché : cela signifie que le service en ligne pourrait céder pour une valeur proposée inférieure ou égale à la valeur de la donnée sur le marché.

Exemple 2 : Ok si valeur ≤ valeur_service : ici cela signifie que le service en ligne ne souhaite pas négocier si la valeur proposée par l'utilisateur est supérieure à la sienne.

Exemple 3 : Ok si valeur ≈ valeur_marché + 10 : cela signifie que le service en ligne pourrait céder pour une valeur proposée supérieure à celle du marché de 10 points maximum.

3.3.3 Des entrées optionnelles

Ces entrées concernent principalement les choix précédents de l'utilisateur. Ils sont optionnels et ne sont pris en compte que si ces 2 conditions sont remplies :

- *L'utilisateur n'est pas nouveau* : il serait en effet impossible d'extraire et exploiter un quelconque historique des choix d'un utilisateur lorsque celui-ci utilise le système pour la première fois.
- *L'utilisateur désire sauvegarder son historique* : l'utilisateur a la possibilité de choisir si le système doit ou non sauvegarder son historique pour des besoins ultérieurs.

L'intérêt de la sauvegarde de l'historique des choix de l'utilisateur est non seulement de lui fournir de meilleures recommandations, mais aussi d'accélérer le processus en prenant en compte les choix antérieurs d'un usager par rapport à une donnée. Il ne sera plus alors toujours nécessaire de redemander l'avis de l'usager.

3.4 Les processus

Le modèle de politique de confidentialité comprend un ensemble d'opérations qui sont effectuées par le système. Elles sont détaillées dans les sections qui suivent.

3.4.1 L'acquisition

Cette première étape du processus consiste à:

- Obtenir les préférences de l'utilisateur et les besoins du service en ligne
- Obtenir optionnellement l'historique de l'utilisateur
- Vérifier la demande d'adhésion de l'utilisateur au service en ligne
- Vérifier l'authenticité du fichier de configuration de l'utilisateur
- Vérifier l'identité du service en ligne et l'authenticité de son fichier de paramètres
- Valider syntaxiquement et sémantiquement les entrées

Vérification de la demande d'adhésion de l'utilisateur au service en ligne

Cette vérification a pour but d'éviter au serveur de traiter des requêtes visant à le surcharger (voir exemple en Figure 3.4). Il faut en effet s'assurer que la demande

d'adhésion est faite effectivement par un utilisateur et non un robot et qu'elle est fraiche (qu'il ne s'agit pas d'une redite).

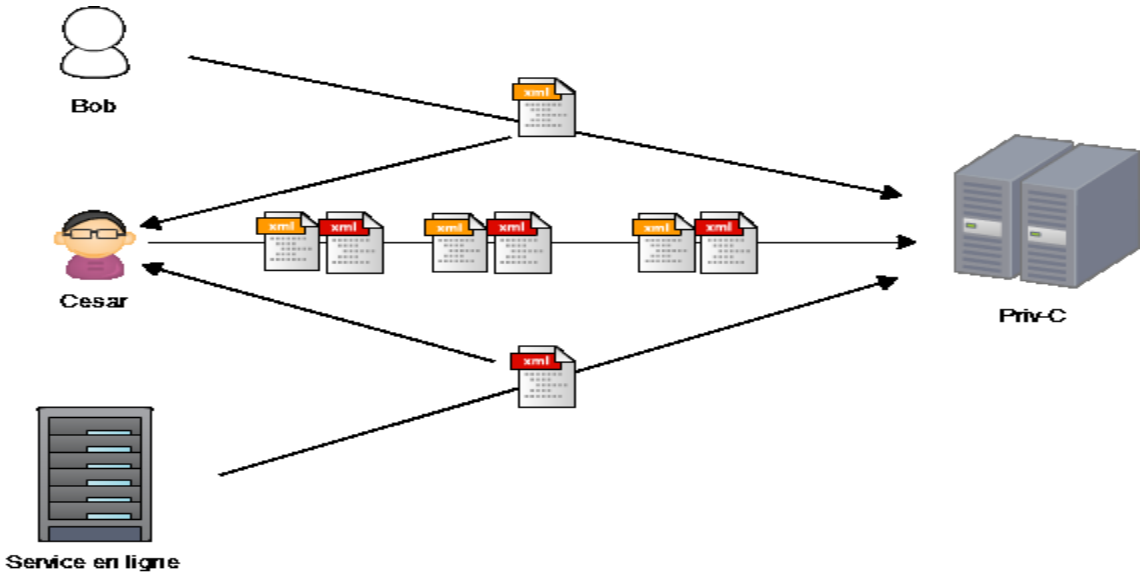


Figure 3.4 Attaque de l'homme du milieu

Il existe plusieurs façons de se protéger des attaques par redite. Dans notre étude, nous avons opté pour *l'horodatage*. Il s'agit d'ajouter la date et l'heure au message puis de le signer. Le récepteur s'assure que le message envoyé est récent (temps inférieur à x minutes). L'homme du milieu ne peut modifier la date et l'heure car il lui est impossible d'y apposer la signature du service en ligne.

L'identité du service en ligne

La vérification de l'identité du service en ligne permet de s'assurer qu'il ne s'agit pas d'un service fictif. Dans cette étude, nous avons opté pour le système de certificats numériques.

Pour établir une connexion SSL avec un utilisateur, chaque service en ligne dispose d'un certificat numérique délivré par une autorité de certification. Ce certificat numérique contient des informations tels qu'un ID identifiant de façon unique la compagnie, son nom, son adresse, sa clé publique de même que la signature électronique du certificat par l'autorité de certification. Le client (navigateur par exemple) vérifie la signature de

l'autorité de certification à partir de la clé publique de ce dernier. Si la signature est valide, le client vérifie la validité du certificat, sa date de validité et dispose de ce fait de la clé publique de la compagnie afin de vérifier les messages signés que cette dernière enverra durant la session.

Ainsi, similairement, le système procédera à la validation de l'identité du service en ligne à partir du certificat de ce dernier.

L'authenticité du fichier de configuration du service en ligne

L'échange s'effectuant par une connexion sécurisée, nous prenons comme hypothèse qu'il est impossible à l'homme du milieu de modifier le contenu de ce fichier, la connexion SSL étant considérée sure. Il existe toutefois un moyen supplémentaire au serveur de vérifier le fichier de configuration du service en ligne à partir de la clé publique de ce dernier.

Validation syntaxique et sémantique

Les préférences de l'utilisateur de même que les besoins du service en ligne doivent être validés syntaxiquement et sémantiquement par le système avant toute opération. Ces informations étant reçues sous formats XML, l'utilisation d'un schéma XML permet de vérifier la syntaxe de leurs contenus, la présence de toutes les informations obligatoires, les formats des valeurs attendues etc. Les formats des fichiers XML ainsi que leurs schémas de validation seront présentés dans le chapitre 4.

Une fois acquis par Priv-C, les différentes entrées sont transférées au module d'appariement dont le fonctionnement est présenté dans la section qui suit.

3.4.2 Module d'appariement

Ce module est chargé d'analyser le contenu des fichiers reçus, de faire des comparaisons par paires et d'isoler les éventuels conflits.

La comparaison par paire consiste à analyser les spécifications de l'utilisateur et du service en ligne pour une donnée et pour un usage particulier. Pour ce faire il suffit d'appliquer un filtre sur les données reçues pour extraire les informations spécifiques à une donnée et à un usage (Voir Figure 3.5). Lorsque les spécifications ne concordent pas, Priv-C les marque avec un label « conflit » pour traitement ultérieur. En cas de non conflit, ces informations deviennent une clause du futur contrat de confidentialité.

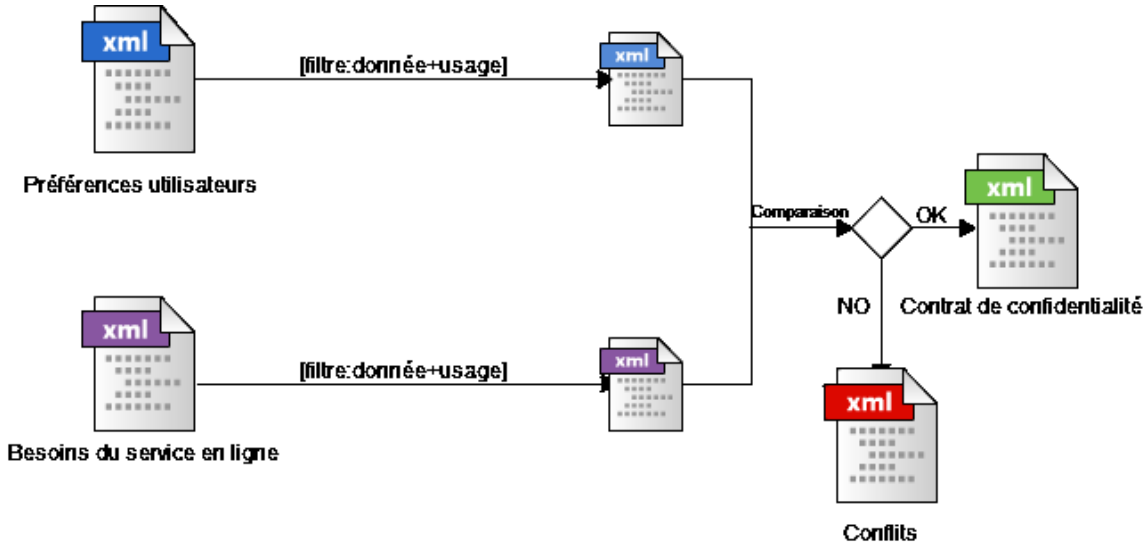


Figure 3.5 : Appariement

L'appariement est suivi selon notre modèle par le module de négociation qui est une pièce centrale du modèle.

3.4.3 Négociation

Le module de négociation est le module d'analyses et de résolution de conflits. La négociation est en effet un processus permettant d'améliorer les accords sur des points de vue communs ou des plans d'action (Durfee & Lesser, 1989). En fonction des conflits éventuels reçus du module d'appariement, le module de négociation peut effectuer les opérations suivantes :

- Demande de recommandations au module approprié (détaillé dans la section suivante)
- Propositions de nouvelles clauses aux deux parties (utilisateur et service en ligne)

- Recommandations accompagnant les nouvelles clauses pour un choix éclairé de l'utilisateur
- En fonction des paramétrages faits par l'utilisateur, il peut prendre des décisions pour celui-ci

Protocoles de négociations

Un protocole de négociation régit l'ensemble des règles qui dirigent l'interaction. Ceci inclut en général les acteurs, les états de la négociation, les événements qui font passer d'un état à un autre et les actions valides et acceptables de la part des participants (Amraoui, Benmammar, Krief, & Bendimerad, 2012). Différents protocoles de négociations existent notamment *les enchères*, *la négociation heuristique*, la *négociation par argumentation*. Celle que nous employons ici est la négociation heuristique, signifiant une interactivité des acteurs avec des refus de propositions et des contre-propositions.

Les composantes de la négociation

Nous avons modélisé le processus de négociation en prenant en compte les aspects suivants :

- **Les acteurs** : outre le système en charge de mener les négociations, deux acteurs sont présents : l'utilisateur et le service en ligne.
- **L'objet de la négociation** : la négociation ici porte sur les données privées des usagers, leur valeur de même que leur durée de rétention.
- **Les processus de décision** : nous décrivons ici les algorithmes de négociation sur la valeur des données privées (Algo 1) et celle de génération de nouvelle valeur pour une donnée (Algo 2).
- Les actions valides et les propositions acceptables par les participants sont toutes paramétrées dans les fichiers de préférences, notre objectif étant ici de rendre le processus de négociation le plus rapide et le plus automatisé possible.

La phase de négociation consiste donc à des ajustements faits par l'utilisateur et le service en ligne. Ces ajustements concernent les clauses du contrat marquées du label « conflit ». Le module renvoie donc les conflits de même que des propositions de solutions au conflit à l'usager puis au service en ligne pour que ceux-ci prennent une décision. Toutefois, cette succession d'opérations de va-et-vient peut être automatisée en fonction du paramétrage des préférences de l'utilisateur et du service en ligne. Ceux-ci peuvent donc spécifier que la résolution du conflit soit effectuée par le système lui-même. Dans ce cas, le système choisit la solution la plus modérée et équitable pour les 2 parties, en se basant sur les recommandations.

Comme mentionnée ci-dessus, la négociation peut porter sur la durée de rétention des données de même que sur leur valeur. Le pseudocode suivant décrit le processus de négociation sur la valeur des données de l'utilisateur :

Algo 1 : Négociation sur la valeur d'une donnée

```
Début
Conflit_resolu = 0
Répéter
        Si valeur_utilisateur < valeur_service_en_ligne Alors
                Valeur = valeur_service_ligne
        Sinon
        Valeur=Generer_nouvelle_valeur(valeur_utilisateur) – [Voir Algo 2]
        Fin si
        Générer options avec Valeur
        Choisir option à recommander
        Envoyer options à utilisateur
        Si choix_utilisateur == conflit_resolu Alors
                Conflit_resolu = 1
                Sauvegader_choix
        Sinon
                //L'utilisateur souhaite renégocier, alors augmenter la valeur de la
donnée
                Valeur_utilisateur = valeur + valeur_incrementation
        Fin si
Tant que conflit_resolu ==0 et Fin_negociation = Faux
Fin
```

À partir de cet algorithme, le système résout les conflits en proposant de nouvelles options aux utilisateurs. Cette opération de négociation se répètera tant que le conflit n'aura pas été résolu. L'utilisateur peut donc à chaque fois choisir une option pour continuer la négociation. Alors la boucle reprend et de nouvelles options sont générées. Le conflit est résolu lorsque l'utilisateur choisit une option parmi celles qui lui sont proposées ou lorsque la négociation n'est plus possible, en fonction des paramètres du service en ligne. L'utilisateur a également toujours le choix d'une option mettant fin aux négociations au risque de ne pas utiliser le service. Dans la génération des options, le système génère à chaque fois une nouvelle valeur pour l'échange de la donnée. Le système génère cette valeur de façon maximale pour l'utilisateur, selon le pseudocode suivant :

Algo 2 : Génération d'une nouvelle valeur pour une donnée

```
Début
Si valeur_utilisateur < valeur_sur_marché Alors
        Si valeur_sur_marché OK pour service_en_ligne Alors
                Nouvelle_valeur = valeur_sur_marché
        Sinon
                Si valeur_utilisateur OK pour service_en_ligne Alors
                        Nouvelle_valeur = valeur_utilisateur
                Sinon
                        Nouvelle_valeur = valeur_limite_service

Sinon //valeur_utilisateur > valeur_sur_marché
        Si valeur_utilisateur Ok pour service_en_ligne Alors
                Nouvelle_valeur = valeur_utilisateur
        Sinon
                Si valeur_sur_marché Ok pour service_en_ligne Alors
                        Nouvelle_valeur = valeur_sur_marché
                Sinon
                        Nouvelle_valeur = valeur_limite_service
//Si en sortant de cette fonction, la valeur de la donnée vaut moins que la valeur que
l'utilisateur désire, cela signifie qu'il n'y a plus de négociation possible
Si nouvelle_valeur < valeur_utilisateur alors
        Fin_negociation = vrai
Fin
```

Cette fonction génère une nouvelle valeur pour la donnée de façon maximale. La variable « valeur_sur_marché » représente la tendance globale de la donnée chez l'ensemble des utilisateurs et des services en ligne. La fonction « Ok pour service_en_ligne » vérifie si la valeur générée est acceptée par le service en ligne. Cela est vérifié dans le fichier de configuration de ce dernier dans la balise « options acceptables ».

La figure 3.6 ci-dessous présente le processus de négociations tel que géré par notre modèle. On y observe une itération du processus de générations d'options et de choix de l'utilisateur jusqu'à résolution de conflit. Ce choix peut toutefois être automatisé en fonction des préférences définies par l'utilisateur.

L'itération prend fin lorsqu'une des conditions est remplie :

- L'utilisateur ne fait aucun choix. Cela signifie qu'il ne veut plus poursuivre les négociations. Le conflit est enregistré comme « Non résolu » et est fermé.
- Le choix de l'utilisateur satisfait au service en ligne, en fonction des paramétrages de celui-ci. Le conflit est enregistré comme « résolu » et est fermé.

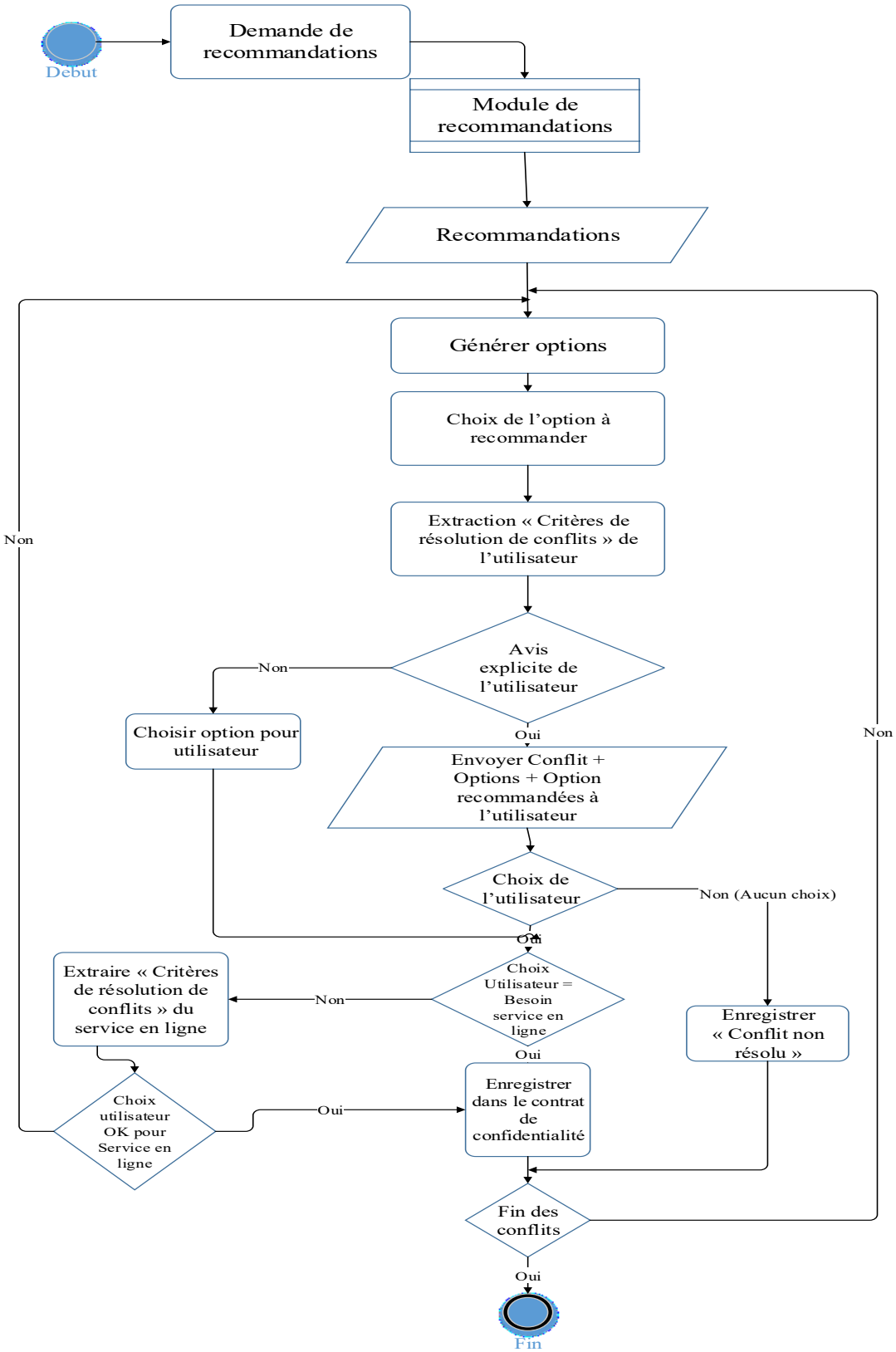


Figure 3.6. Module de négociations

Le module de négociations communique en permanence avec celui chargé des recommandations. Chaque nouvelle série d'options générées pour être envoyée à l'utilisateur demande en effet une analyse et une suggestion faite l'entité de recommandations.

3.4.4 Génération des recommandations

Les systèmes de recommandations représentent les préférences de l'utilisateur dans le but de lui faire des suggestions pour un achat ou une analyse (Burke, 2002). Il existe de nombreuses classifications des systèmes de recommandations. Burke se base sur la source des données utilisées pour classer les systèmes de recommandations en 5 catégories : *le filtrage collaboratif*, *le filtrage sur le contenu*, le *filtrage démographique*, le *filtrage à base d'utilité* et le *filtrage à base de connaissance*. De nouvelles méthodes sont expérimentées et voient le jour tel que le « group-based elicitation » qui vient étendre le « item-based preferences » (Chang, Harper, & Terveen, 2015). Ces 2 méthodes peuvent être regroupées parmi les méthodes de filtrage collaboratives.

Utiliser l'une ou plusieurs de ces méthodes dépend des informations traitées par un système. Un système qui ne connait pas encore ses utilisateurs pourrait opter pour le « group-based illicitation » afin de connaitre les préférences généralistes de l'usager. Un tel système pourrait également opter pour le filtrage collaboratif, en se basant sur le comportement des utilisateurs ou la notation des contenus pour faire des recommandations.

Dans notre système, l'utilisateur a le choix ou non de sauvegarder son historique. Nous optons pour le filtrage collaboratif, avec la possibilité d'y ajouter d'autres méthodes ultérieurement pour le rendre hybride.

Le filtrage collaboratif permet de proposer des recommandations à un utilisateur en fonction des estimations données dans son voisinage. Ce filtrage se base soit sur une similarité entre les utilisateurs soit sur une similarité entre les articles. La similarité entre les articles demande que chacun des usagers aient déjà attribué des notes à des articles.

Le système recommande alors à un utilisateur des articles similaires à celui ou ceux qu'il a noté positivement. Il se pose ici un problème de « démarrage à froid » lorsque l'utilisateur n'a encore noté aucun article. Bien qu'il existe des méthodes pour résoudre ce problème, nous avons opté dans cette étude pour le filtrage collaboratif basé sur les utilisateurs.

Le filtrage collaboratif basé sur les utilisateurs (à base de mémoire ou heuristique) recherche pour un utilisateur donné ceux qui lui sont le plus similaires. Dans le cas où l'utilisateur à qui le système veut faire des recommandations n'a encore noté aucun article, le système sélectionne les articles qui sont les mieux notés par ses voisins pour lui faire des recommandations. Dans le cas où l'utilisateur a déjà noté des articles, la recommandation peut être plus affinée. Le système peut en effet sélectionner ceux qui ont des notations similaires à l'utilisateur pour lui faire des recommandations.

Une échelle de notation

Une échelle de notation permet donc aux usagers d'attribuer des notes aux différentes propositions qu'il a reçues durant la phase de négociation. L'objectif étant de recommander l'option la mieux notée aux futurs usagers, en se basant aussi sur leur proximité.

Nous avons pour cela établi une échelle de notation de 1 à 5. Le 1 désigne la plus haute expression de refus de la part de l'usager tandis que le 5 représente la plus haute expression d'acceptation ou d'appréciation.

Calcul de la proximité

L'une des méthodes de similarité les plus populaires permettant à un système de rechercher pour un usager ceux qui lui sont le plus similaires est la corrélation de Pearson (Breese et *al.*, 1998).

$$w\,(A,B) = \frac{\sum_j (v_{A,j} - \overline{v_A})(v_{B,j} - \overline{v_B})}{\sqrt{\sum_j (v_{A,j} - \overline{v_A})^2 \sum_j (v_{B,j} - \overline{v_B})^2}}$$

Formule de la corrélation de Pearson.

j: nombre d'objets ayant été voté à la fois par A et B
$v_{A,j}$: Vote de A pour l'item j
$\overline{v_A}$: Moyenne des votes de A

Prédiction

Une fois les usagers similaires à l'utilisateur repérés, le système utilise leur notation pour générer une recommandation sur le conflit concerné. Le système sélectionne parmi les similaires ceux dont les notations ressemblent à celles de l'utilisateur, et effectue une moyenne pondérée de leur notation.

La formule utilisée pour le calcul de la prédiction est présentée ci-dessous, où « *sim* » est une mesure de similarité, par exemple la fonction *w* ci-dessus.

$$P_{Aj} = \bar{v}_A + \frac{\sum_{i=1}^{n} sim(A,i) * (v_{i,j} - \bar{v}_i)}{\sum_{i=1}^{n} |sim(A,i)|}$$

Formule de calcul de la prédiction, réadaptée de Naak (Naak, 2009)

n représente le nombre d'usagers voisins de A, ayant voté pour l'option j

$v_{i,j}$ représente le vote de l'usager j pour l'option j

$\bar{v}_i$ représente la moyenne des votes de l'usager i

Illustrons tout ceci par un exemple. Le système a sélectionné les utilisateurs similaires à Alice afin de lui faire des recommandations sur chacune des options générées à la suite d'un conflit.

Tableau 3.8. Exemple de sélection d'utilisateurs similaires

	Option 1	Option 2	Option 3	$\bar{v}_i$
Alex	4	3	1	2.67
Alice	**1.06**	**-0.24**	**-0.1**	?
Bob	5	3	2	3.33
Camille	3	3	2	2.67

Si la similarité entre Alice et Alex vaut 0.7, entre Alice et Bob elle vaut 0.65 et entre Alice et Camille 0.8 **alors** la prédiction pour Alice pour chacune des options se détermine comme suit :

$$P_{Alice,option\ 1} = \frac{0.7 * (4 - 2.67) + 065 * (5 - 3.33) + 0.8 * (3 - 2.67)}{0.7 + 0.65 + 0.8} = \frac{2.28}{2.15} = 1.06$$

$$P_{Alice,option\ 2} = \frac{0.7 * (3 - 2.67) + 065 * (3 - 3.33) + 0.8 * (2 - 2.67)}{0.7 + 0.65 + 0.8} = \frac{-0.52}{2.15} = -0.24$$

$$P_{Alice,option\ 3} = \frac{0.7 * (1 - 2.67) + 065 * (2 - 3.33) + 0.8 * (2 - 2.67)}{0.7 + 0.65 + 0.8} = \frac{-0.23}{2.15} = -0.1$$

La prédiction de l'option 1 à Alice vaut donc 1.06 et reste la plus élevée, ce qui parait logique vues les notations des autres usagers. Les prédictions d'Alice ne sont pas très élevées car Alice n'a encore fait aucune notation sur chacune des options.

À l'issue des recommandations et de la phase de négociation, le contrat de confidentialité, s'il y a lieu est prêt à être envoyé à l'utilisateur et au service en ligne. Une dernière opération est toutefois effectuée.

3.4.5 Certification du contrat de confidentialité

Cette dernière étape du processus consiste à apposer une signature électronique de l'autorité de régulation. Cette signature permettra à quiconque, notamment à l'utilisateur et au service en ligne de vérifier l'authenticité du contrat. Elle permettra également de vérifier que le contrat n'a pas été modifié, puisque la signature s'appliquera sur le résultat du hachage du contrat de confidentialité (sous format XML).

L'utilisateur peut disposer d'un agent intelligent sur son navigateur capable de vérifier l'authenticité d'un contrat de confidentialité. Il suit alors les étapes suivantes :

- Vérifier que le hachage du contrat de confidentialité correspond bien au hachage original effectué par l'autorité de régulation.
- Vérifier que la signature du hachage est bien celle de l'autorité de régulation, à partir de la clé publique de ce dernier.

Il s'agit du principe « hache et signe » reconnu sûr (Gennaro, Halevi, & Rabin, 1999; Hohenberger & Waters, 2009), car modifier le fichier original contraint l'adversaire à hacher de nouveau le fichier et le signer. Ce qu'il s'avère impossible puisque ce dernier ne dispose pas de la clé privée de l'autorité de régulation.

3.5 La sortie : le contrat de confidentialité

Le contrat de confidentialité se présente sous la forme d'un fichier XML. Il est envoyé tant à l'utilisateur qu'au service en ligne. Il est composé de clauses, chaque clause comprend les informations d'accord entre l'utilisateur et le service en ligne à l'issue des négociations.

3.6 Conclusion

Nous avons conçu un modèle et un formalisme pour décrire les préférences de l'utilisateur et les besoins des services en ligne. Ces besoins intègrent de nombreuses

options à l'utilisateur, lui permettant de choisir pour chacune de ses données, comment elle sera utilisée et partagée, sa durée de rétention, et des récompenses éventuelles. Le modèle proposé permet également la négociation entre l'utilisateur et le service en ligne sur des paramètres tels que la donnée, sa valeur, sa durée de rétention, les services offerts, etc. Pour finir, tous ces processus peuvent être automatisés. L'utilisateur définit ses préférences une seule fois, ainsi que le service en ligne. Des options sont prévues dans les paramètres pour résoudre les conflits et effectuer les négociations sans aucune intervention humaine. Le chapitre qui suit décrit cette automatisation du système.

4 Chapitre 4 Implémentation

Dans cette section, nous présentons le développement du système Priv-C pour répondre aux problématiques et aux spécifications décrites dans les chapitres précédents. Nous présentons les fonctionnalités de la solution développée. Nous discutons également du choix des différentes méthodes de conception et de développement et présentons différents cas de figures du fonctionnement de notre système.

PRIV-C est un système orienté vers les technologies du web. Des pages web permettent à l'utilisateur d'interagir avec le système. Les technologies XML permettent le stockage et l'échange d'informations entre les différents acteurs. Afin que cet échange soit possible, les données sont organisées suivant une syntaxe et sémantique bien définies.

Les préférences et les contrats de l'usager sont stockés sur son ordinateur personnel. Les contrats sont également stockés auprès du service en ligne et du système Priv-C. Dans la figure qui suit, nous voyons que l'utilisateur stocke en interne ses préférences de même que le service en ligne en ce qui concerne ses besoins en données privées. Les deux entités communiquent avec Priv-C par le biais d'une API (Application Programming Interface), à laquelle ils envoient des requêtes et reçoivent des réponses en retour.

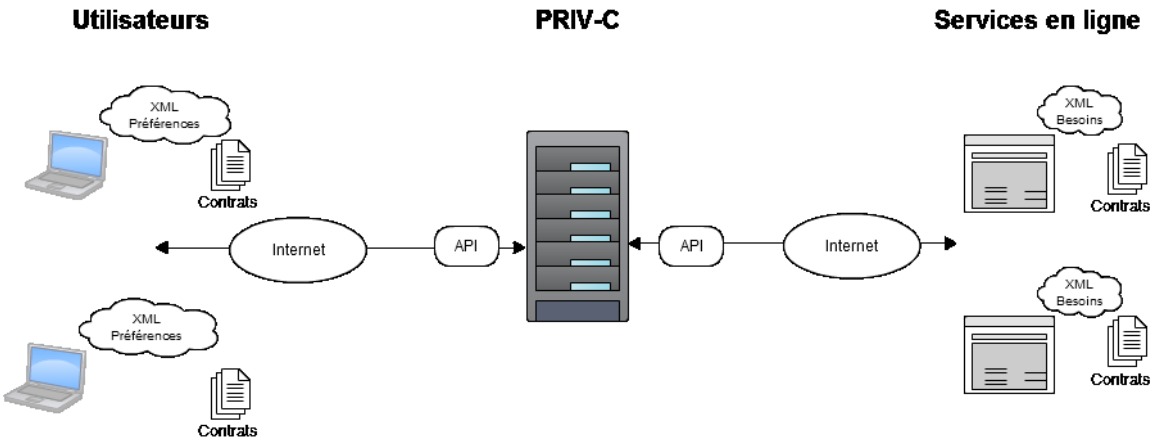


Figure 4.1. Architecture logicielle de Priv-C

4.1 Les fonctionnalités du système

La figure suivante présente l'arborescence des fonctionnalités offertes par Priv-C chez un utilisateur.

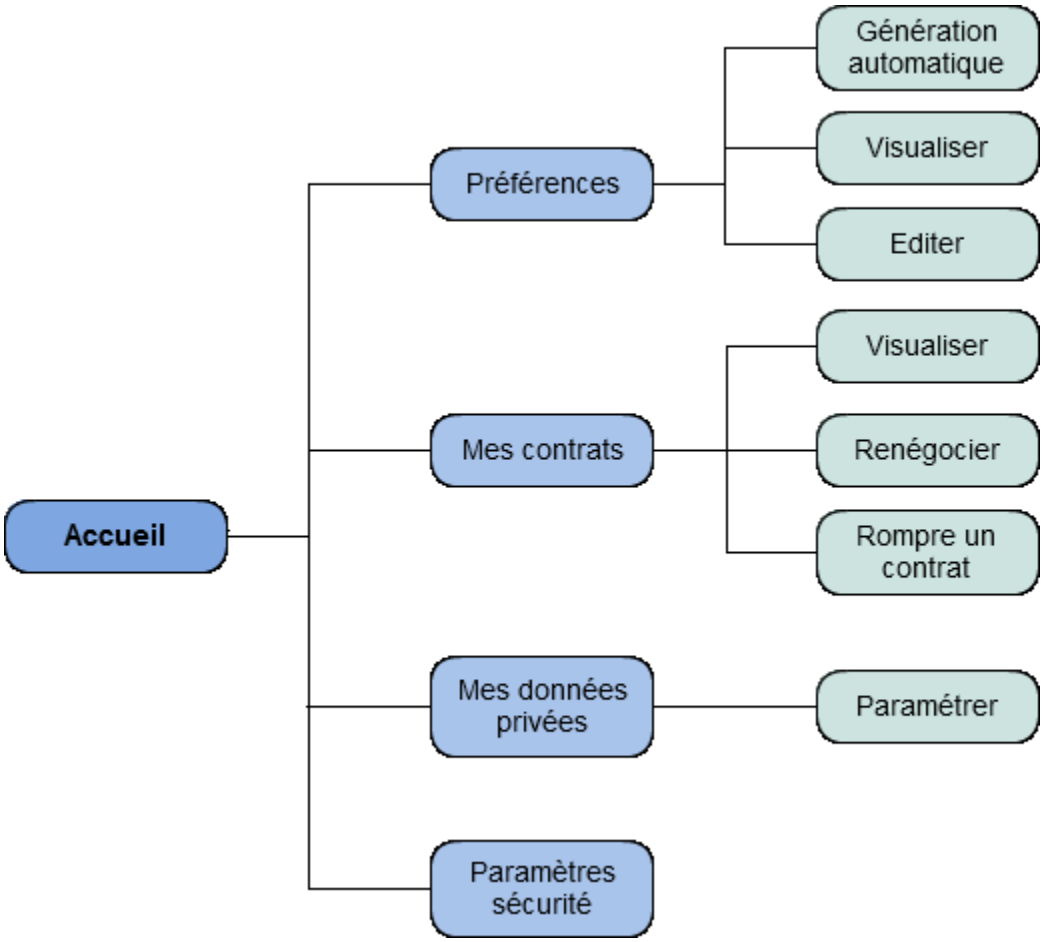


Figure 4.2. Fonctionnalités de Priv-C

4.1.1 Les préférences

Cette fonctionnalité permet à l'utilisateur de visualiser et modifier ses préférences en matière de confidentialité. Les préférences se présentent tel que spécifiées dans la section 3.3.1.

Les préférences de l'utilisateur sont sauvegardées dans un fichier XML. La validité du fichier sémantique est assurée par un fichier XSD (XML Schema) qui en contrôle la syntaxe, la sémantique et le contenu.

L'exemple qui suit présente le paramétrage des préférences pour les données d'identification de l'utilisateur, de même que son équivalent selon notre schéma XML. Tels que spécifié dans le chapitre précédent, la préférence est relative à un groupe de données (identifié ici par GD1) et à un usage particulier (ici Qualité de service). Les autres informations y sont aussi représentées, telles que les conditions, les exceptions et les critères de résolution de conflit.

```
<preference idGroupeDonnees="GD1" usageDonnee="Qualité Service">
    <action>OUI</action>
    <conditions>
        <condition>
            <expression>
                <condition>Si</condition>
                <terme>Durée de rétention</terme>
                <symbole>>=</symbole>
                <valeur><recompense>90</recompense></valeur>
            </expression>
            <action>NON</action>
        </condition>
    </conditions>
    <resolutionconflit>Suivre le recommandeur</resolutionconflit>
</preference>
```

Figure 4.3.Extrait des préférences utilisateurs en XML

Figure 4.4. Extrait des préférences utilisateurs sur une page web

La génération automatisée des préférences utilisateurs

Telle que spécifiée, définir ses préférences par un utilisateur revient à paramétrer pour chaque groupes de données personnelles et pour chaque usage, des actions, des conditions et des exceptions. Ce paramétrage peut être fait différemment en fonction des types de services en ligne. En effet, la façon dont un utilisateur souhaite protéger ses données peut être différent face à un réseau social ou à un site de commerce en ligne. Tel que présentée sur la figure 4.4, la phase de paramétrage des préférences utilisateurs peut être longue et fatiguant, car elle doit être faite pour chacune des données et les différents usages possibles.

Pour résoudre cela et rendre cette tâche aisée aux utilisateurs, PRIV-C inclut une option de génération automatisée des préférences utilisateurs. Cette option permet à l'usager, à l'aide de quelques choix, de générer des préférences personnalisées, qu'il peut ensuite modifier ou sauvegarder comme telles.

Étant donné que chaque service en ligne propose des fonctionnalités, l'usager choisit les fonctionnalités auxquelles il souhaite impérativement avoir accès, celles dont il pourrait se passer en fonction des exigences en données privées. Le système génère alors des préférences pour cet usager avec le maximum de protection de ses données privées.

Exemple 1 : Alice aimerait avoir accès aux services de « Messagerie » et de « Achats en ligne » tout en gardant un maximum de confidentialité.

Voici les préférences générées par le système pour Alice :

```
<preferences>
    <preference idGroupeDonnees="GD1" usageDonnee="Qualité Service">
        <action>OUI</action>
        <conditions>
            <condition>
                <expression>
                    <condition>Si</condition>
                    <terme>Durée de rétention</terme>
                    <symbole>>=</symbole>
                    <valeur><recompense>90</recompense></valeur>
                </expression>
                <action>NON</action>
            </condition>
        </conditions>
        <resolutionconflit>Suivre le recommandeur</resolutionconflit>
    </preference>
    <preference idGroupeDonnees="GD1" usageDonnee="Commercial">
        <action>NON</action>
        <exceptions>
            <exception>
                <expression>
                    <condition>À moins que</condition>
                    <terme>Non accès aux services</terme>
                    <symbole>==</symbole>
                    <valeur>
                        <services>
                            <service>MESSAGERIE</service>
                            <service>ACHAT EN LIGNE</service>
                        </services>
                    </valeur>
                </expression>
```

```
        </exception>
    </exceptions>
</preference>
<preference idGroupeDonnees="GD1" usageDonnee="Partage pour qualité">
    <action>OUI</action>
    <conditions>
        <condition>
            <expression>
                <condition>Si</condition>
                <terme>Durée de rétention</terme>
                <symbole>>=</symbole>
                <valeur>
                    <recompense>90</recompense>
                </valeur>
            </expression>
            <action>NON</action>
        </condition>
    </conditions>
    <exceptions>
        <exception>
            <expression>
                <condition>À moins que</condition>
                <terme>Non accès aux services</terme>
                <symbole>==</symbole>
                <valeur>
                    <services>
                        <service>MESSAGERIE</service>
                        <service>ACHAT EN LIGNE</service>
                    </services>
                </valeur>
            </expression>
```

Nous pouvons remarquer qu'excepté pour les usages « Qualité de service » et « Partage pour qualité », le système ne souhaite pas donner accès aux données de l'usager, tout en ajoutant une exception : « À moins que cela ne prive l'usager des services de Messagerie et d'Achat en ligne », vu que ce sont les services que l'usager tient à utiliser.

Exemple 2 : Bob aimerait avoir accès à tous les services, tout en obtenant des profits quand cela est possible. Il aimerait malgré tout que ses données ne soient pas largement partagées avec d'autres services.

Voici les préférences générées par Priv-C pour Bob :

```
<preferences>
    <preference idGroupeDonnees="GD1" usageDonnee="Qualité Service">
        <action>OUI</action>
    </preference>
    <preference idGroupeDonnees="GD1" usageDonnee="Commercial">
        <action>OUI</action>
    </preference>
    <preference idGroupeDonnees="GD1" usageDonnee="Partage pour qualité">
        <action>OUI</action>
    </preference>
    <preference idGroupeDonnees="GD1" usageDonnee="Partage commercial">
        <action>NON</action>
        <conditions>
            <condition>
                <expression>
                    <condition>Si</condition>
                    <terme>Valeur récompense</terme>
                    <symbole>>=</symbole>
                    <valeur><recompense>100</recompense></valeur>
                </expression>
                <action>OUI</action>
            </condition>
        </conditions>
        <exceptions>
            <exception>
                <expression>
                    <condition>À moins que</condition>
                    <terme>Non accès aux services</terme>
                    <symbole>==</symbole>
                    <valeur>
                        <services>
                         <service>TOUS</service>
                        </services>
                    </valeur>
                </expression>
            </exception>
    </exceptions>
</preference>
<preference idGroupeDonnees="GD1" usageDonnee="Partage pour raison légal">
    <action>OUI</action>
</preference>

<preference idGroupeDonnees="GD2" usageDonnee="Qualité Service">
    <action>OUI</action>
</preference>
<preference idGroupeDonnees="GD2" usageDonnee="Commercial">
    <action>OUI</action>
</preference>
<preference idGroupeDonnees="GD2" usageDonnee="Partage pour qualité">
    <action>OUI</action>
</preference>
<preference idGroupeDonnees="GD2" usageDonnee="Partage commercial">
    <action>NON</action>
    <conditions>
        <condition>
            <expression>
                <condition>Si</condition>
                <terme>Valeur récompense</terme>
                <symbole>>=</symbole>
                <valeur><recompense>100</recompense></valeur>
            </expression>
            <action>OUI</action>
```

Le système génère des préférences permettant à l'usager d'accéder à tous les services et de bénéficier de profits quand cela est possible. Toutefois, en ce qui concerne le partage, le système génère une condition et une exception quant à l'usage « Partage commercial » pour respecter les restrictions de l'usager.

4.1.2 Les besoins du service en ligne

Les besoins du service en ligne suivent un format modélisé en XML. La validité du fichier XML est assurée par un fichier XSD (« XML Schema »).

L'utilisateur n'a pas vraiment besoin de visualiser la politique de confidentialité du service en ligne, étant donné que le système se charge de générer automatiquement le contrat de confidentialité en fonction des préférences de l'utilisateur. Toutefois, Priv-C offre à l'utilisateur la possibilité de visualiser d'une façon claire et concise les besoins du service en ligne relativement à chacune de ses données privées.

Menu

Accueil
Préférences
Mes contrats
Mes données privées
Paramètres sécurité

Besoins du service en ligne

Données	Usage	Utilisation	Durée de rétention	Options
Données d'identification et de contact	Qualité Service	Obligatoire	1 - Fermeture du compte 2 - Demande de l'utilisateur 3 - Jamais	1 - ACCEPT 2 - ACCEPT 50 3 - ACCEPT BAGDE PREMIUM - REFUS TOUS
	Commercial	Optionnel	1 - Fin de la session 2 - Fermeture du compte	1 - ACCEPT 30 2 - ACCEPT 80 - REFUS
	Partage pour qualité	Optionnel	1 - Fin de la session 2 - Fermeture du compte	1 - ACCEPT PARTAGE DE POSITION 2 - ACCEPT 30 PARTAGE DE POSITION - REFUS PARTAGE DE POSITION
	Partage commercial	Optionnel	1 - Fin de la session 2 - Fermeture du compte	1 - ACCEPT 25 2 - ACCEPT 25 SMS GRATUITS - REFUS SMS GRATUITS
	Partage pour raison légal	Obligatoire	1 - 1 annees	- ACCEPT - REFUS TOUS
				1 - ACCEPT

Figure 4.5. Besoins du service en ligne sur une page web

L’équivalent XML des besoins de ce service en ligne se présente comme suit :

```xml
<besoin idGroupeDonnees="GD1" usageDonnee="Qualité Service">
    <utilisation>Obligatoire</utilisation>
    <dureeretention>
        <valeurindeterminee ordre="1">Fermeture du compte</valeurindeterminee>
        <valeurindeterminee ordre="2">Demande de l'utilisateur</valeurindeterminee>
        <valeurindeterminee ordre="3">Jamais</valeurindeterminee>
    </dureeretention>
    <options>
        <option action="ACCEPT" ordreDureeRetention="1">
            <contrepartie>
                <aucune/>
            </contrepartie>
        </option>
        <option action="ACCEPT" ordreDureeRetention="2">
            <contrepartie>
                <contre>
                    <valeur>50</valeur>
                </contre>
            </contrepartie>
        </option>
        <option action="ACCEPT" ordreDureeRetention="3">
            <contrepartie>
                <contre>
                    <services>
                        <service>BAGDE PREMIUM</service>
                    </services>
                </contre>
            </contrepartie>
        </option>
        <option action="REFUS">
            <contrepartie>
```

Figure 4.6. Besoins du service en ligne en format XML

4.1.3 Mes contrats

Ce module permet à l’usager de visualiser tous les contrats avec les services en ligne qu’il utilise. C’est un outil très important car il offre une vision globale à l’usager de toutes ses données privées, des services en ligne qui y ont accès et de comment chacune de ses données est utilisée. L’utilisateur peut donc décider de renégocier un contrat (ne

plus permettre l'accès à une donnée par exemple), de rompre un contrat avec un service qu'il n'utilise plus souvent par exemple.

La figure 4.8 présente une vue des contrats d'un utilisateur. Les contreparties en couleur rouge indiquent une redevance dues par l'utilisateur, celles en vert indiquent un gain ou aucune redevance pour l'usager.

```
<clauses>
    <clause idGroupeDonnees="GD1" usageDonnee="Qualité Service">
        <accept>
            <dureeretention>
                <valeurindeterminee>Fermeture du compte</valeurindeterminee>
            </dureeretention>
            <contrepartie>
                <aucune/>
            </contrepartie>
        </accept>
    </clause>
    <clause idGroupeDonnees="GD1" usageDonnee="Commercial">
        <refus>
            <contrepartie>
                <aucune/>
            </contrepartie>
        </refus>
    </clause>
    <clause idGroupeDonnees="GD1" usageDonnee="Partage pour qualité">
        <refus>
            <contrepartie>
                <contre>
                    <services>
                        <service>PARTAGE DE POSITION</service>
                    </services>
```

Figure 4.7. Contrat de confidentialité au format XML

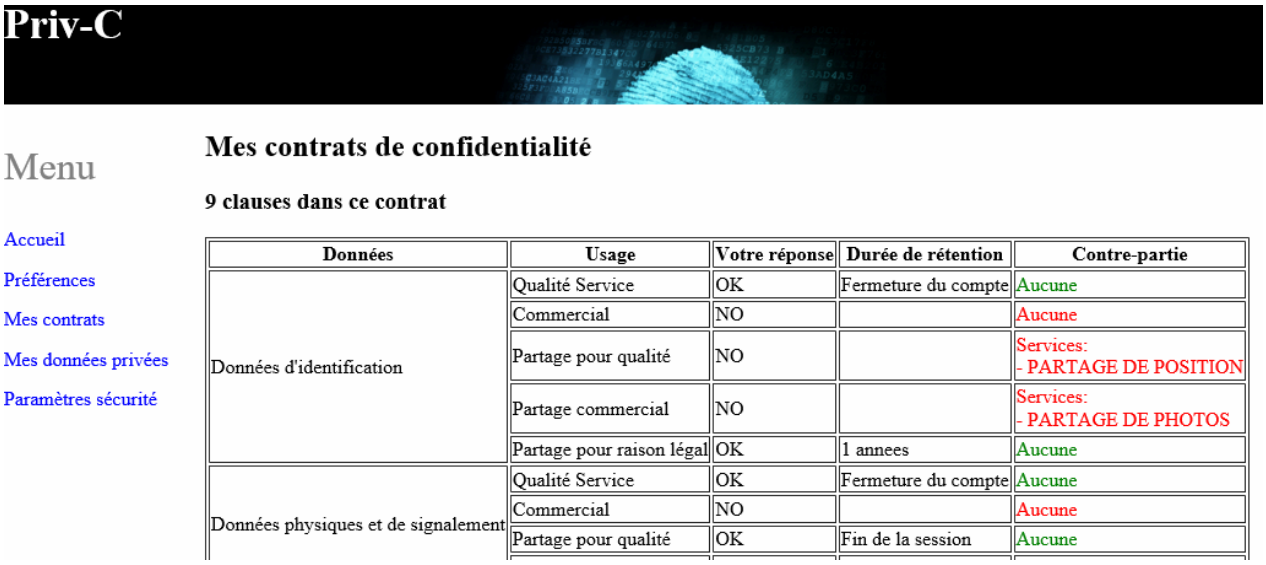


Figure 4.8. Contrat de confidentialité sur une page web

4.2 Quelques scénarios de fonctionnement

Cette section présente quelques cas de fonctionnements du système, en l'occurrence la détection et la résolution de conflits.

4.2.1 Détection des conflits

Il se produit un conflit durant le processus lorsque les besoins du service en ligne ne correspondent pas aux préférences de l'usager et qu'aucune option présente dans les deux fichiers ne permet au système de trouver un point d'accord.

```xml
<besoin idGroupeDonnees="GD1" usageDonnee="Qualité Service">
    <utilisation>Obligatoire</utilisation>
    <dureeretention>
        <valeurindeterminee ordre="1">Fermeture du compte</valeurindeterminee>
        <valeurindeterminee ordre="2">Demande de l'utilisateur</valeurindeterminee>
        <valeurindeterminee ordre="3">Jamais</valeurindeterminee>
    </dureeretention>
    <options>
        <option action="ACCEPT" ordreDureeRetention="1">
            <contrepartie>
                <aucune/>
            </contrepartie>
        </option>
        <option action="ACCEPT" ordreDureeRetention="2">
            <contrepartie>
                <contre>
                    <valeur>50</valeur>
                </contre>
            </contrepartie>
        </option>
        <option action="ACCEPT" ordreDureeRetention="3">
            <contrepartie>
                <contre>
                    <services>
                        <service>BAGDE PREMIUM</service>
                    </services>
                </contre>
            </contrepartie>
        </option>
        <option action="REFUS">
            <contrepartie>
```

```xml
<preferences>
    <preference idGroupeDonnees="GD1" usageDonnee="Qualité Service">
        <action>OUI</action>
        <conditions>
            <condition>
                <expression>
                    <condition>Si</condition>
                    <terme>Durée de rétention</terme>
                    <symbole>>=</symbole>
                    <valeur><recompense>90</recompense></valeur>
                </expression>
                <action>NON</action>
            </condition>
        </conditions>
        <resolutionconflit>Suivre le recommandeur</resolutionconflit>
    </preference>
```

Un conflit se produit car l'usager accepte l'utilisation de ses données GD1 en vue de la qualité de service à la condition que la durée de rétention ne soit pas supérieure ou égale à 90 jours. Or comme on le constate, toutes les durées de rétention proposées par le

service en ligne sont supérieures à 90 jours, et en plus l'obtention de cette donnée est obligatoire pour l'accès par l'usager à tous les services proposés.

4.2.2 Résolution d'un conflit

Le conflit présenté dans la section précédente porte sur les données « GD1 » et l'usage « Qualité de service ».

Le système génère un ensemble de nouvelles options. Parmi ces options, il en recommande une à l'usager.

Options:

- Donner accès à la donnée aux conditions du service en ligne pour avoir accès au service TOUS
- Donner accès avec condition Durée de rétention et contre x (x représentant la valeur de cette donnée) points pour avoir accès aux services TOUS
- Poursuivre la négociation : générer de nouvelles options
- Ne pas changer d'avis, vous n'aurez pas accès aux services TOUS

4.2.3 La notation des options

Comme précisé au chapitre 3, nous utilisons ici un système de recommandations collaboratif basé sur les utilisateurs (à base de mémoire ou heuristique).

Pour y parvenir, à la fin de chaque processus, la liste des nouvelles propositions générées par le système est envoyée à l'utilisateur, afin que celui-ci attribue des notations à chacune d'entre elles. L'objectif est de recommander l'option la mieux notée par les usagers aux futures usagers, en se basant aussi sur leur proximité.

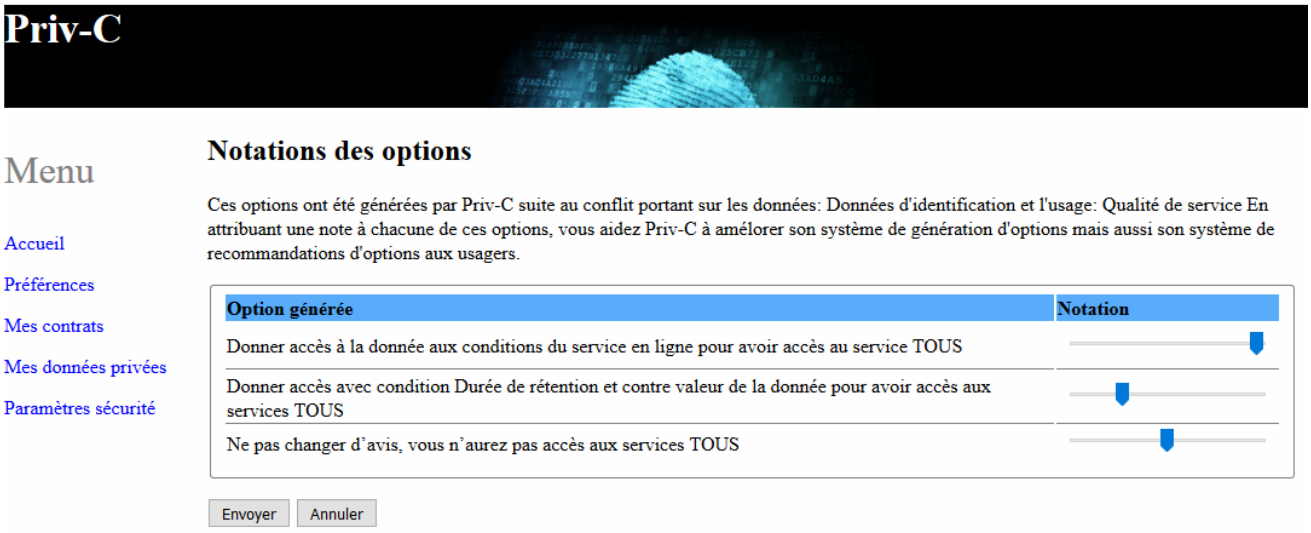


Figure 4.9. Formulaire de notations des options générées

Priv-C est un système simple d'utilisation et assez autonome. Une fois les paramètres définis à la première utilisation, toutes les autres tâches sont prévues pour être gérées en toute rapidité.

De nombreux autres modèles de politiques de confidentialité existent. Une évaluation de Priv-C par rapport à ces modèles va nous permettre de distinguer les nouvelles contributions de ce modèle personnalisable de politique de confidentialité.

5 Chapitre 5 Validation

Ce chapitre est consacré à validation de notre système. Pour ce faire, nous le comparons à des systèmes semblables existants pour une évaluation qualitative. Ensuite, nous effectuons également des expérimentations incluant des sondages afin de recueillir et évaluer les commentaires et notations des utilisateurs.

5.1 Comparaison aux systèmes existants

Dans cette comparaison, nous prenons en compte les principaux travaux en matière de politique de confidentialité présentés dans le chapitre 2. Bien que le P3P ne soit plus d'actualités, nous l'incluons dans ce tableau comparatif car c'est le seul standard du W3C, en matière de politiques de confidentialité, produit de nos jours.

Tableau 5.1. Comparaison de Priv-C aux systèmes existants

	Préférences automatisées	Multiple options par données	Négociation des termes	Vue globale de l'utilisation de ses données	Système de recommandation	Calcul de la valeur des données privées	Notifications
P3P							X
P2U			X				
PRIVEE	X						
PrivacyFix							X
Priv-C	X	X	X	X	X	X	

Du point de vue du paramétrage des préférences, seul notre système Priv-C de même que PRIVEE offre la possibilité de générer automatiquement des préférences. PRIVEE permet de le faire en se basant sur les habitudes de navigation de l'utilisateur tandis que Priv-C génère les préférences en fonction des fonctionnalités que l'utilisateur souhaite utiliser auprès du service en ligne et du degré de confidentialité qu'il souhaite avoir.

Seul Priv-C offre la possibilité d'avoir pour chaque donnée et pour chaque usage, de multiples options à l'utilisateur. Notre système est également le seul à proposer une vue

globale à l'utilisateur de ses données. Cela lui permet de mieux suivre les accès à ses informations, de rompre des contrats avec des services en ligne ou de les renégocier avec de nouvelles clauses. Aucun des systèmes présentés ici n'offre également des recommandations aux utilisateurs, excepté Priv-C. Cela est très important dans un contexte où la technologie avance très vite. Les utilisateurs ont besoin d'être guidé afin d'assurer un meilleur contrôle de leurs informations.

La négociation des termes d'une politique est offerte par Priv-C et P2U. Notons que P2U se limite uniquement à la durée de rétention et au prix des données tandis que Priv-C intègre aussi la négociation des options et des fonctionnalités offertes par un service en ligne.

Aucun modèle de confidentialité présenté, excepté Priv-C, ne propose de module de détermination de la valeur des données des utilisateurs afin de les intégrer dans les négociations. Enfin, une future amélioration que pourrait intégrer notre système serait une fonction de notifications, pour envoyer des alertes à des usagers.

5.2 Évaluation par les utilisateurs

Afin d'évaluer le niveau de confiance des utilisateurs pour un service Internet qui utilise ce modèle de politique de confidentialité, nous avons procédé à deux expérimentations :

- La première est un sondage de 363 utilisateurs d'Internet ayant adhéré à un service en ligne avec une politique de confidentialité suivant notre modèle. L'expérimentation ainsi que les résultats a été présenté dans un article publié à la conférence CRISIS 2015 :

  Oluwa Lawani, Esma Aïmeur, Kimiz Dalkir, "Improving users' trust through friendly privacy policies: an empirical study", *The International Conference on Risks and Security of Internet and Systems*, Lesvos, 2015.
- La seconde a consisté en deux sondages : le premier interrogeant 359 internautes utilisant un service avec une politique de confidentialité classique, le second avec

358 internautes ayant adhéré à un service avec une politique de confidentialité suivant notre modèle. Tous les détails de l'expérimentation, des résultats et ces analyses sont présentés dans un article, accepté pour publication dans le journal CHB - Computer in Human Behavior.

Esma Aïmeur, Oluwa Lawani, Kimiz Dalkir, When changing the look of privacy policies affects user trust: An experimental study, *Computers in Human Behavior*, vol 58, pp 368-379.

Nous présentons dans les sections qui suivent les éléments clés relatifs à cette seconde expérimentation.

5.2.1 L'expérimentation

Le point dominant de cette étude est de donner aux usagers le contrôle de leurs données personnelles en leur offrant diverses options sur chaque clause d'un contrat de confidentialité. Pour ce faire, nous avons soumis 2 groupes d'utilisateurs à deux tests : le premier groupe adhère à une politique de confidentialité conventionnelle et le second à une politique suivant le modèle décrit dans ce mémoire. Répondant ensuite à une même série de questions, nous avons évalué, par le biais d'un modèle et d'une analyse factorielle confirmatoire, la confiance des utilisateurs vis-à-vis des deux services en ligne.

Pourquoi avons-nous choisi deux groupes d'utilisateurs au lieu de soumettre les deux politiques à un même groupe ? Il existe en effet deux principaux types de sondage pour mener une étude comparative : le sondage de type « within-subjects » où chaque répondant est exposé à divers expériences et le sondage de type « between-subjects » où chaque individu est exposé à une seule expérience. Les deux types de sondage ont leurs mérites et le choix de l'un par rapport à l'autre devrait être considéré avec soin en termes de mise en œuvre pratique de l'étude de recherche (Charness, Gneezy, & Kuhn, 2012). La conception « within-subjects » peut avoir de meilleurs effets. Leur validité interne ne dépend pas d'une assignation aléatoire. Cependant, la conception « between-subjects »

est meilleure dans un environnement où l'individu est susceptible de faire face à une seule décision. Il a aussi l'avantage que le comportement des utilisateurs dans les futures expériences n'est pas influencé par les précédentes.

Dans cette étude, nous avons opté pour la méthode « between-subjects ». Le principal défi était alors d'avoir deux groupes de participants avec la meilleure ressemblance en termes de caractéristiques. Comme nous le verrons plus loin dans Tableau 5.2, ce défi a été bien surmonté. Grâce à Mechanical Turk (MTurk), service de Amazon par lequel le sondage a été réalisé, nous avons d'abord sélectionné tous nos répondants du Canada et des États-Unis. Les deux groupes de répondants sont presque également répartis dans toutes les catégories. En outre, nous nous sommes assurés de ne pas avoir un participant dans les deux groupes en empêchant, grâce à MTurk, tous les participants de la première expérience de participer à la seconde.

5.2.2 Le modèle de recherche

Nous avons conçu un modèle (voir Figure 5.1) et défini différentes hypothèses dans le but d'étudier les facteurs affectant la confiance des utilisateurs vis-à-vis des services en ligne.

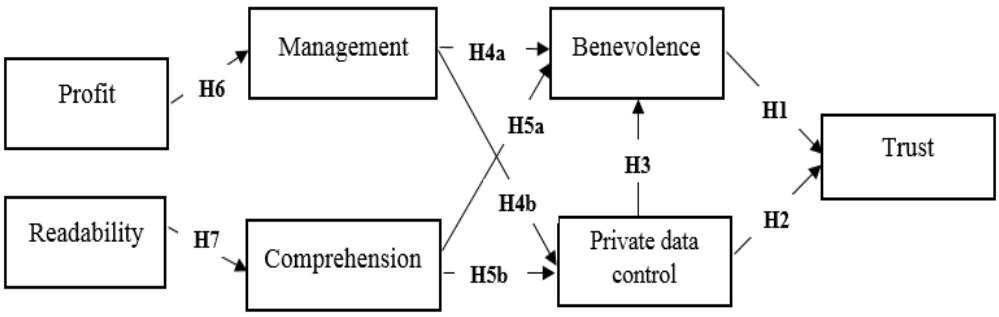


Figure 5.1. Modèle de recherche

Ce modèle comprend 7 facteurs qui sont reliés entre eux par des hypothèses. Nous présentons ici 3 de ces facteurs, dont il a été le plus question dans ce mémoire.

Benevolence (Bienveillance : BEN) :

La confiance est un facteur au cœur des affaires et du commerce sur Internet (Urban, Sultan, & Qualls, 2000). Elle est également la clé du succès des réseaux sociaux (Sherchan, Nepal, & Paris, 2013). La confiance comprend trois axes: *la compétence, l'intégrité et la bienveillance* (Kim, Ferrin, & Rao, 2008). Lorsque l'utilisateur estime que ces trois facteurs sont présents, il accorde sa confiance au service en ligne. Selon le but de notre étude, nous considérons la bienveillance comme un facteur influençant de façon considérable la confiance. La bienveillance signifie que le service en ligne ne considère pas seulement ses intérêts, mais se soucie aussi de ceux de ses utilisateurs. Le service en ligne doit donc faire de son mieux pour l'utilisateur, en mettant de côté son égoïsme et agir dans le meilleur intérêt de l'utilisateur (Raimondo, 2000). Cela nous permet d'émettre cette hypothèse :

H1: La bienveillance est positivement liée à la confiance.

Le contrôle des données personnelles (Private data control : PDC)

En 2012, la Commission européenne a proposé de renforcer la protection des utilisateurs en ligne. Les nouvelles directives devraient permettre aux utilisateurs de prendre le contrôle de leurs données. Les changements indiquent, entre autres choses, que l'autorisation de l'utilisateur doit être explicitement demandée et que tous les paramètres des sites doivent être orientés vers la protection de la vie privée de l'utilisateur (Rotenberg & Jacobs, 2013). Selon ses directives, les utilisateurs peuvent faire confiance aux entreprises qui fournissent une meilleure information sur la façon dont leurs données seront utilisées et protégées. Dans ce contexte, nous émettons les deux hypothèses suivantes :

H2: Le contrôle des données personnelles est positivement lié à la confiance.
H3: Le contrôle des données personnelles est positivement lié à la bienveillance.

Le profit

Le profit motive les utilisateurs à partager des données personnelles (Acquisti et al., 2013; Chorppath & Alpcan, 2013). Lorsqu'ils sont conscients que leurs données privées ont une valeur, ils sont plus engagés dans l'échange avec un service en ligne. Sachant que toutes les données n'ont pas la même valeur pour tous les utilisateurs (Acquisti et al., 2013), nous nous attendons à ce que chaque utilisateur évalue les risques réels et attribue une valeur à ses données privées avant de décider de les divulguer ou de les protéger. Nous émettons donc l'hypothèse suivante :

H6: Le profit est positivement lié à la gestion.

Les autres hypothèses de ce modèle sont stipulées comme suit (voir Figure 5.1):

H4a: La gestion des données est positivement liée à la bienveillance.

H4b: La gestion des données est positivement liée au contrôle des données personnelles.

H5a: La compréhension est positivement liée à la bienveillance.

H5b: La compréhension est positivement liée au contrôle des données privées.

H7: La lisibilité est positivement liée à la compréhension.

5.2.3 Le sondage

Le sondage a été effectué sur Amazon Mechanical Turk (MTurk - www.mturk.com), en Mai 2015. MTurk est un service offert par Amazon permettant de recruter de façon optimale des participants pour divers types de sondages.

Les participants ont eu à souscrire à une politique de confidentialité d'un service fictif, appelé **Ikrani**[14], puis répondre à une série de questions sur l'expérience. Une rémunération de 1,25 $ a été offerte à chaque participant.

Le questionnaire, présenté en Annexe 1, a été conçu à partir des facteurs du modèle de recherche. Chacun des facteurs a été décomposé en questions, sur la base de diverses études antérieures. Par exemples, les questions relatives au facteur « Confiance » permettent de mesurer la capacité et l'intégrité des services en ligne (Zhou, 2013). Celles du facteurs « Bienveillance » mesurent le fait que le service en ligne s'occupe de ses utilisateurs et entretient une relation amicale avec eux (Raimondo, 2000).

Un total de 359 réponses ont été prises en compte dans l'analyse du premier scénario et 358 dans le second. Rappelons que le premier scénario concerne les participants ayant adhérés à la politique de confidentialité conventionnelle (voir Annexe 2), le second ceux ayant souscrit à la politique de confidentialité suivant notre modèle (voir Annexe 3). Ils devaient tous répondre à des qualifications élevées pour être admis au sondage, par exemple une approbation ≥ 95%, un nombre de succès approuvés ≥ 100.

Les caractéristiques démographiques des participants sont présentées dans le tableau suivant.

Tableau 5.2. Caractéristiques démographiques des participants

		Scenario 1	**Scenario 2**
Critères	**Categories**	**Fréquence**	**Fréquence**
Genre	Féminin	42,9 %	41,1 %
	Masculin	56,3 %	58,9 %
	Je préfère ne pas répondre	0,8 %	0 %
Age	16-24	18,1 %	19 %
	25-34	42,3 %	48,9 %
	35-44	24,5 %	17,9 %
	45-54	8,9 %	8,7 %
	55+	6,1 %	5,6 %

[14] Ikrani est un mot arabe signifiant "Lis-moi"

	Je préfère ne pas répondre	0 %	0%
Éducation	Primaire	0,3 %	0 %
	École secondaire	29,5 %	27,4 %
	École technique/commercial	17,5 %	15,4 %
	Licence	42,6 %	43,9 %
	Maitrise	5,8 %	8,9 %
	Doctorat	1,7 %	1,4 %
	Je préfère ne pas répondre	1,1 %	1,4 %
	Autre	1,4 %	1,7 %
Revenu annuel	Aucun revenu	1,9 %	2,2 %
	Moins de $ 10,000	18,4 %	14 %
	$10,001 – 25,000	20,3 %	19,3 %
	$25,001 – 35,000	18,7 %	20,4 %
	$35,001 – 45,000	13,4 %	13,7 %
	$45,000+	24 %	25,7 %
	Je préfère ne pas répondre	3,3 %	4,7 %

5.2.4 Les résultats

L'outil utilisé pour conduire cette étude est AMOS 22.0.0, un produit de IBM permettant, entre autres, d'effectuer des Analyses Factorielles Confirmatoires (CFA).

Dans une première étape, nous avons évalué, nos instruments de mesure de l'étude, afin de tester la fiabilité et la validité.

Le tableau 5.3 présente les résultats avec le coefficient de saturation, la variance moyenne extraite (AVE), la fiabilité (CR) et les valeurs alpha de Cronbach. Tous les AVE sont supérieurs à 0,5 (sauf un), tous les CR supérieurs à 0,7, ce qui indique que l'échelle a une bonne validité convergente (Gefen, Straub, & Boudreau, 2000). Comme suggéré (Numally, 1978), toutes les valeurs alpha sont supérieures à 0,7, indiquant une bonne fiabilité.

Le tableau 5.4 montre que, pour presque tous les facteurs, la racine carrée de l'AVE est plus grande que ses coefficients de corrélation avec d'autres facteurs, indiquant une bonne validité discriminante (Fornell & Larcker, 1981; Gefen et al., 2000).

La deuxième étape, consiste à mesurer la qualité globale de l'ajustement pour le modèle de recherche. Comme présenté dans le tableau 5.5, la comparaison des indices d'ajustement avec les valeurs recommandées démontre des résultats acceptables.

Tableau 5.3. Analyse de la fiabilité et de la validité

		Scenario 1				Scenario 2			
Facteur	**Objet**	**Standardized item loading**	**AVE**	**CR**	**Alpha value**	**Standardized item loading**	**AVE**	**CR**	**Alpha value**
Readability (**REA**)	REA1	1,000	0,66	0,78	0,717	1,000	0,68	0,80	0,740
	REA2	0,561				0,594			
Profit (**PRO**)	PRO1	0,430	0,59	0,72	0,600	0,449	0,60	0,73	0,610
	PRO2	1,000				1,000			
Comprehension (**COM**)	COM1	0,724	0,61	0,82	0,814	0,778	0,63	0,84	0,836
	COM2	0,752				0,811			
	COM3	0,856				0,795			
Management (**MAN**)	MAN1	0,827	0,66	0,79	0,792	0,928	0,59	0,73	0,677
	MAN2	0,793				0,560			
Private data control (**PDC**)	PDC1	0,678	0,57	0,80	0,788	0,759	0,45	0,71	0,694
	PDC2	0,863				0,735			
	PDC3	0,714				0,494			
Benevolence (**BEN**)	BEN1	0,766	0,70	0,87	0,865	0,693	0,63	0,83	0,829
	BEN2	0,907				0,798			
	BEN3	0,825				0,876			
Trust (**TRU**)	TRU1	0,886	0,66	0,80	0,785	0,907	0,77	0,87	0,869
	TRU2	0,736				0,847			

Tableau 5.4. Racine carré de AVE et coefficients de corrélation des facteurs

	Scenario 1							Scenario 2						
	PDC	**REA**	**COM**	**PRO**	**MAN**	**BEN**	**TRU**	**PDC**	**REA**	**COM**	**PRO**	**MAN**	**BEN**	**TRU**
PDC	**0,756**							**0,673**						
REA	0,052	**0,811**						0,300	**0,822**					
COM	0,094	0,752	**0,779**					0,455	0,750	**0,795**				
PRO	0,451	0,115	0,092	**0,770**				0,334	0,109	0,236	**0,775**			
MAN	0,924	0,031	0,092	0,475	**0,810**			0,629	0,325	0,470	0,180	**0,766**		
BEN	0,771	0,162	0,250	0,530	0,745	**0,835**		0,699	0,311	0,422	0,455	0,434	**0,793**	
TRU	0,688	0,115	0,160	0,503	0,559	0,749	**0,814**	0,717	0,365	0,465	0,434	0,420	0,785	**0,878**

Tableau 5.5. Mesure de l'ajustement du modèle

		Chi² /DF	**GFI**	**AGFI**	**CFI**	**NFI**	**RMSEA**
	Valeur recommandée	< 3	> 0.90	> 0.80	> 0.90	> 0.90	< 0.08
Scenario 1	CFA	2,76	0,92	0,88	0,94	0,92	0,070
	SM	2,71	0,92	0,88	0,94	0,91	0,069
Scenario 2	CFA	2,99	0,91	0,86	0,93	0,90	0,075
	SM	3,27	0,89	0,84	0,92	0,89	0,080

5.2.5 Tests des hypothèses

Des neuf hypothèses, 7 sont supportées dans le premier scénario et 8 dans le second, comme représenté sur la Figure 5.2 et la Figure 5.3, et dans le tableau 5.6.

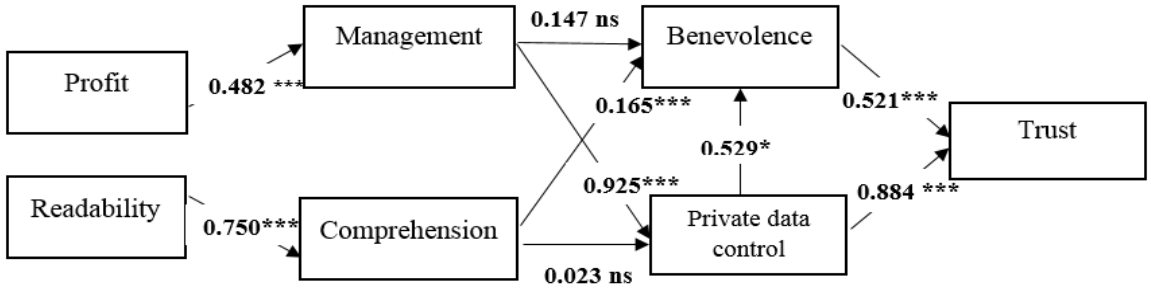


Figure 5.2. Tests des hypothèses – Scenario 1

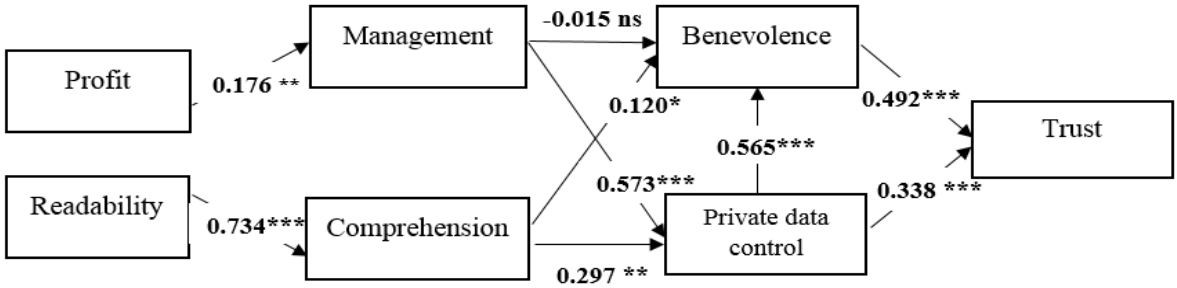


Figure 5.3. Test des hypothèses – Scenario 2

Tableau 5.6. Tests des hypothèses

Attributs	Hypotheses	Scenario 1			Scenario 2		
		Estimés	p	Statuts	Estimés	p	Statuts
BEN → TRU	H1	0,521	< 0,001	Supportée	0,492	< 0,001	Supportée
PDC → TRU	H2	0,884	0,009	Supportée	0,338	< 0,001	Supportée
PDC → BEN	H3	0,529	0,015	Supportée	0,565	< 0,001	Supportée
MAN → BEN	H4a	0,147	0,505	Non Supportée	-0,015	0,844	Non Supportée
MAN → PDC	H4b	0,925	< 0,001	Supportée	0,573	< 0,001	Supportée
COM → BEN	H5a	0,165	< 0,001	Supportée	0,120	0,036	Supportée
COM → PDC	H5b	0,023	0,734	Non Supportée	0,297	0,003	Supportée
PRO → MAN	H6	0,482	< 0,001	Supportée	0,176	0,002	Supportée
REA → COM	H7	0,750	< 0,001	Supportée	0,734	< 0,001	Supportée

Le Bénéfice affecte positivement la gestion des données personnelles (H6), hypothèse supportée dans les deux scénarios.

Comme nous pouvons le constater, l'hypothèse **H4a**, stipulant que la gestion des données personnelles affecte positivement la bienveillance, n'est pas prise en charge dans les deux scénarios. H4b est quant à elle vérifiée dans les deux scénarios (la gestion influe positivement le contrôle des données privées).

L'hypothèse **H1** (la bienveillance affecte positivement la confiance), **H2** (le contrôle des données privées influe positivement sur la confiance) et **H3** (le contrôle des données privées affecte positivement la bienveillance) sont toutes confirmées dans les deux scénarios.

Nos résultats montrent clairement que définir de nouveaux formats de politiques de confidentialité permettant aux utilisateurs de comprendre leur contenu, de gérer l'accès à chacune de leurs données et de bénéficier de l'utilisation de ses données, change le niveau de confiance des utilisateurs et rend les systèmes d'exploitation plus dignes de confiance. Ce nouveau format tient compte du format de la politique de confidentialité, de l'organisation et de la séquence des informations. Cela permet aux utilisateurs de comparer les alternatives sur chaque clause de la politique avant de prendre une décision.

De cette étude expérimentale et comparative, il apparaît que les politiques de confidentialité actuelles ne prennent pas considérablement en compte la gestion des données, leur contrôle et la compréhension des clauses. Ce sont en effet les points clés et les contributions de notre modèle de politique de confidentialité.

5.3 Conclusion

La validation de Priv-C montre de manière qualitative, les fonctionnalités et les nouvelles contributions par rapport aux systèmes actuels. Priv-C a également été testé sur plus de 1000 utilisateurs afin de recueillir et analyser leurs impressions. Les résultats montrent que les objectifs visés par Priv-C, notamment le contrôle et la gestion des

données personnelles, sont les facteurs majeurs affectant la confiance des utilisateurs envers les services en ligne.

6 Chapitre 6 Conclusion

Les recherches en matière de protection des données personnelles sont nombreuses et s'orientent sur différents axes : la cryptographie, l'anonymat, le comportement social, la quantification des risques, etc. Ce mémoire s'est orienté vers les politiques de confidentialité, premier accord entre utilisateurs et services en ligne relatif à la collecte et à l'utilisation des données personnelles. La problématique abordée est l'absence de choix aux utilisateurs qui sont contraints d'accepter entièrement le contenu d'une politique de confidentialité dans le but d'utiliser un service.

L'état de la littérature, dans le chapitre 2, a permis de faire le point sur les différents travaux en matière de politiques de confidentialité. Il en ressort d'abord un manque de standard dans le domaine. Le P3P, seule recommandation du W3C, n'a pas su satisfaire les usagers et peine à évoluer depuis des années. Bien d'autres modèles sont proposés, sans pour autant résoudre entièrement le problème, apportant en général des fonctionnalités limitées à de simples notifications. Dans cette revue de la littérature, il a été observé également que de plus en plus d'études et de systèmes proposent d'offrir des récompenses aux utilisateurs en échange de l'utilisation de leurs données par les services en ligne.

Nous avons présenté dans le chapitre 3 notre modèle de politique de confidentialité. Ce modèle intègre de multiples options faites à l'utilisateur sur chaque clause de la politique de confidentialité. Il lui permet également de négocier avec le service en ligne sur l'utilisation qui sera faite de ses données personnelles mais aussi sur la valeur de celles-ci et leur durée de rétention. Une méthode d'analyse multicritères a été utilisée pour évaluer les données personnelles des utilisateurs en fonction de l'importance qu'ils y accordent. Une fois cette valeur déterminée, elle est utilisée durant les négociations, pour permettre à l'utilisateur, s'il le souhaite, de recevoir des récompenses en échange de l'utilisation de ses données personnelles. Une option de recommandation est également

intégrée, permettant au système de faire des suggestions à l'usager en fonction de ses préférences et des habitudes des utilisateurs qui lui sont similaires.

Ce modèle a été implémenté et présenté dans le chapitre 4. Essentiellement basé sur les technologies XML, Priv-C permet à l'utilisateur et au service en ligne de définir leurs préférences et besoins en matière de données personnelles. Un module d'appariement effectue une évaluation et isole les conflits éventuels qui sont traités par un module de négociation qui génère de nouvelles options et des recommandations à l'utilisateur.

Pour évaluer le système, nous avons, dans le chapitre 5, comparé Priv-C aux principaux systèmes existants. Une expérimentation a permis aux utilisateurs de donner leur avis, par le biais de 3 sondages effectués avec plus de 1000 participants. Deux articles ont été rédigés à la suite de ces expérimentations et publiés. L'étude empirique qui a été faite dans ces articles montre que les utilisateurs ont plus tendance à faire confiance aux services en ligne leur permettant d'avoir un contrôle sur l'utilisation de leurs données personnelles et également d'en tirer profit.

Notre objectif est d'offrir aux utilisateurs des politiques de confidentialités souples et équitables. Priv-C est un système offrant de multiples options sur chaque clause de la politique de confidentialité, la possibilité de négocier sur les termes et des recommandations aux utilisateurs. Le contrat de confidentialité qui en résulte est donc différent pour chaque utilisateur, en fonction des préférences de ce dernier.

L'une des contributions majeures de ce mémoire est donc le contrôle offert aux utilisateurs sur leurs données privées. Pour une fois, ils ont la possibilité de décider quelle donnée partager avec un service en ligne, du temps de rétention de la donnée, de son utilisation. Ce contrôle est essentiel pour regagner la confiance des utilisateurs, et permettre une continuité des échanges sur Internet.

Cette solution résout également l'épineuse question de la lecture des politiques de confidentialité, en général longues, difficiles à lire, et parfois mal traduites. L'utilisateur

définit ses préférences et le système se charge de générer automatiquement un contrat de confidentialité personnalisé. L'utilisateur peut également visualiser de façon concrète, sous forme d'un tableau, la politique de confidentialité du service en ligne indiquant la façon précise dont chacune de ses données sera collectée et utilisée.

Un autre apport essentiel de ce mémoire est la possibilité offerte à l'utilisateur de visualiser d'une façon globale tous les contrats de confidentialité qu'il a signés avec des services en ligne. Il peut de cette façon voir tous les services qui collectent ses données, comment ils les utilisent et ce qui lui revient comme récompense. L'utilisateur peut décider de rompre un contrat ou d'en renégocier les termes.

Priv-C reste un prototype et de nombreuses améliorations pourraient y être apportées notamment dans les algorithmes de négociation et de recommandation.

Annexe 1: Questionnaire relatif au modèle de recherche

Trust (adapted from (Zhou, 2013))

TRU1: This online service is trustworthy.

TRU2: This online service will keep their commitments.

Benevolence (adapted from (Raimondo, 2000))

BEN1: Through this privacy policy, I feel close to the online service.

BEN2: Through this privacy policy, I think this online service cares about my concerns.

BEN3: This online service keeps customers' interests in mind.

Private data control

PDC1: I know that my private data will not be disclosed to a third party without my permission.

PDC2: I can decide who has access to my private data.

PDC3: I can change my mind about my privacy settings whenever I want.

Management (adapted from (Song, Kunjithapatham, & Messer, 2006))

MAN1: This online service allows me to choose which data I want to share.

MAN2: This online service offers multiple choices on each of the terms of the privacy policy.

Comprehension (adapted from (Sumeeth, Singh, & Miller, 2012))

COM1: The content of this policy makes sense to me.

COM2: Important information is easily identifiable.

COM3: I understand all the issues related to my privacy.

Profit

PRO1: This privacy policy allows me to get benefits.

PRO2: Incentives motivate me to deal with the online service.

Readability (adapted from (Sumeeth et al., 2012))

REA1: Important points of this privacy policy are easily remembered.

REA2: This privacy policy can be read quickly.

Annexe 2: Politique de confidentialité conventionnelle de Ikrani

This policy describes what information we collect and how it is used and shared. How do we use your information? Ikrani can compare your email address and those of your contacts provided to a website or a third party online service to see if you can benefit from the combined offering. Your profile picture and your name may be associated with an advertisement to show your Ikrani activity (for example, if you subscribe to the Starbucks Page). We use services from other companies to help us derive a general geographic area based on your IP address in order to customize certain features or deals in your area. We can obtain additional information about you, such as demographic data we purchase from other companies. Our automated systems analyse your content (including e-mail) to offer customized product features such as personalized search results, customized advertisements and spam detection and malware. The information you provide on your profile Ikrani can be included in the search directory to enable products that interact with Ikrani to search you and to get in touch with you. We and our partners use various technologies to collect and store information when you visit a Ikrani service, and this may include sending one or more cookies or anonymous identifiers to your device. We also use cookies and anonymous identifiers when you interact with services we offer to our partners, such as advertising services or Ikrani features that may appear on other sites.

Annexe 3 : Politique de confidentialité de Ikrani selon notre modèle

Your email address

Ikrani can compare your email address and those of your contacts provided to a website or a third party online service to see if you can benefit from the combined offering.

Mark only one oval.

- I agree (I earn 100 points)
- I do not agree (0 point)

Profile name and picture

Your profile picture and your name may be associated with an advertisement to show your Ikrani activity (for example, if you subscribe to the Starbucks Page).

Mark only one oval.

- I agree (I earn 100 points)
- I do not agree (0 point)

Profile information

The information you provide on your profile Ikrani can be included in the search directory to enable products that interact with Ikrani to search you and to get in touch with you.

Mark only one oval.

- I agree (I earn 100 points)
- I do not agree (0 point)

Your Geolocation

We use services from other companies to help us derive a general geographic area based on your IP address in order to customize certain features or deals in your area.

Mark only one oval.

- I agree (I earn 100 points)
- I do not agree: I do not want any geolocation, even if it meant I would not benefit from a better quality of service (0 points)
- I do not agree: I wish my physical location to be collected only when I give my explicit agreement and only to be used for my service (I lose 100 points)

Demographics

We can obtain additional information about you, such as demographic data we purchase from other companies.

Mark only one oval.

- I agree (I earn 100 points)
- I do not agree (0 point)

Email content

Our automated systems analyse your content (including e-mail) to offer customized product features such as personalized search results, customized advertisements and spam detection and malware.

Mark only one oval.

- I agree (I earn 100 points)
- I do not agree: I do not want any analysis of my emails (0 point)
- I do not agree: I want the analysis of my emails to only be used to detect spam and malware (I lose 100 points)

Références

Acquisti, A. (2010). The economics of personal data and the economics of privacy. *OECD Joint WPISP-WPIE Roundtable, 1*, 50.

Acquisti, A., John, L. K. & Loewenstein, G. (2013). What is privacy worth? *The Journal of Legal Studies, 42*(2), 249-274.

Acquisti, A., & Varian, H. R. (2005). Conditioning Prices on Purchase History. *Marketing Science, 24*(3), 367-381.

Amraoui, A., Benmammar, B., Krief, F. & Bendimerad, F. T. (2012). *Négotiations à base d'enchères dans les réseaux radio cognitive.* NOTERE (Nouvelles Technologies de la Répartition)/CFIP (Colloque francophone sur l'ingénierie des protocoles).

Apolinarski, W., Handte, M., & Marron, P. J. (2015). Automating the Generation of Privacy Policies for Context-Sharing Applications. *Intelligent Environments (IE)*.

Beresford, A. R., Kübler, D., & Preibusch, S. (2012). Unwillingness to pay for privacy: A field experiment. *Economics Letters, 117*(1), 25-27.

Brown, I. (2013). The economics of privacy, data protection and surveillance. *Handbook on the Economics of the Internet, Cheltenham: Edward Elgar*.

Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User modeling and user-adapted interaction, 12*(4), 331-370.

Calzolari, G., & Pavan, A. (2006). On the optimality of privacy in sequential contracting. *Journal of Economic Theory, 130*(1), 168-204.

Canada, O. o. t. P. C. o. (2013). Survey of Canadians on Privacy-Related Issues. https://www.priv.gc.ca/information/por-rop/2013/por_2013_01_e.asp. Accédé le 27 mars 2015.

Parlement du Canada (2015). Les lois fédérales du Canada sur la protection de la vie privée. http://www.parl.gc.ca/Content/LOP/ResearchPublications/prb0744-f.htm. Accédé le 02 oct. 2015.

Chang, S., Harper, F. M., & Terveen, L. (2015). *Using Groups of Items to Bootstrap New Users in Recommender Systems. 18th ACM Conference on Computer Supported Cooperative Work & Social Computing,* 1258-1269.

Chapman, E. T. (2015). PATRIOT Act: Implications for Colleges and Universities. *Oklahoma Academy of Science*, 65-71.

Charness, G., Gneezy, U., & Kuhn, M. A. (2012). Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior & Organization, 81*(1), 1-8.

Chorppath, A. K., & Alpcan, T. (2013). Trading privacy with incentives in mobile commerce: A game theoretic approach. *Pervasive and Mobile Computing, 9*(4), 598-612.

Cranor, L. F., Hoke, C., Leon, P. G., & Au, A. (2014). Are they worth reading? An in-depth analysis of online advertising companies' privacy policies. *Telecommunications Policy Research Conference*.

Domingo-Ferrer, J. (2010). Rational privacy disclosure in social networks *Modeling Decisions for Artificial Intelligence* (255-265): Springer.

dos Santos Brito, K., Cardoso Garcia, V., Araujo Durao, F., & Romero de Lemos Meira, S. (2013). How people care about their personal data released on social media. *11th Annual International Conference on Privacy, Security and Trust (PST), 111-118*.

Durfee, E. H., & Lesser, V. R. (1989). Negotiating task decomposition and allocation using partial global planning. *Distributed Artificial Intelligence, 2*(1), 229-244.

Egelman, S., Felt, A. P., & Wagner, D. (2013). Choice architecture and smartphone privacy: There's a price for that. *The Economics of Information Security and Privacy* (211-236): Springer.
Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of marketing research*, 39-50.
Gefen, D., Straub, D., & Boudreau, M.-C. (2000). Structural equation modeling and regression: Guidelines for research practice. *Communications of the Association for Information Systems, 4*(1), 1-70.
Gennaro, R., Halevi, S., & Rabin, T. (1999). *Secure hash-and-sign signatures without the random oracle.* Paper presented at the Advances in Cryptology—EUROCRYPT'99.
Ghosh, A., & Roth, A. (2013). Selling privacy at auction. *Games and Economic Behavior*.
Gkatzelis, V., Aperjis, C., & Huberman, B. A. (2015). Pricing private data. *Electronic Markets*, 1-15.
Grossklags, J., & Acquisti, A. (2007). When 25 Cents is too much: An Experiment on Willingness-To-Sell and Willingness-To-Protect Personal Information. *Workshop on the Economics of Information Security*.
Hermalin, B. E., & Katz, M. L. (2006). Privacy, property rights and efficiency: The economics of privacy as secrecy. *Quantitative Marketing and Economics, 4*(3), 209-239.
Hohenberger, S., & Waters, B. (2009). Realizing hash-and-sign signatures under standard assumptions. *Advances in Cryptology-EUROCRYPT 2009* (333-350): Springer.
http://www.p3ptoolbox.org/guide/section2.shtml. (2015). What is P3P and How Does it Work? . Accédé le 12 oct. 2015
Huberman, B., & Aperjis, C. (2013). Creating a market for unbiased private individual data: Google Patents.
Iyilade, J., & Vassileva, J. (2013). A Framework for Privacy-Aware User Data Trading. *User, Modeling, Adaptation and Personalization Conférence*, 310-317.
The United States Departement of Justice (2015). Privacy Act of 1974. http://www.justice.gov/opcl/privacy-act-1974, Accédé le 06 oct. 2015.
Kim, D. J., Ferrin, D. L., & Rao, H. R. (2008). A trust-based consumer decision-making model in electronic commerce: The role of trust, perceived risk, and their antecedents. *Decision support systems, 44*(2), 544-564.
Knijnenburg, B. P., Kobsa, A., & Jin, H. (2013). Dimensionality of information disclosure behavior. *International Journal of Human-Computer Studies, 71*(12), 1144-1162.
Conseil de l'Europe (1981). Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel. http://conventions.coe.int/Treaty/fr/Treaties/Html/108.htm, Accédé le 05 oct. 2015.
Lancelot-Miltgen, C., & Gauzente, C. (2006). Vie privée et partage de données personnelles en ligne: une approche typologique.
Lesk, M. (2012). Your Memory Is Now a Vendor Service. *IEEE Security & Privacy, 10*(1), 88-90.
McDonald, A. M., & Cranor, L. F. (2008). The cost of reading privacy policies. *A Journal of Law and Policy for the Information Society, 4*, 543.
Naak, A. (2009). Papyres: un système de gestion et de recommandation d'articles de recherche.
Nepali, R. K., & Wang, Y. (2013). Sonet: A social network model for privacy monitoring and ranking. Distributed Computing Systems Workshops (ICDCSW), 2013 IEEE 33rd International Conference, 162-166.

Numally, J. C. (1978). Psychometric Theory. NY: McGraw-Hill.
Osothongs, A., & Sonehara, N. (2014). A proposal of personal information trading platform (PIT): A fair trading between personal information and incentives. 4th International Conference on Digital Information and Communication Technology and its Applications (DICTAP), 269-274.
Petkos, G., Papadopoulos, S., & Kompatsiaris, Y. (2015). PScore: A Framework for Enhancing Privacy Awareness in Online Social Networks. *10th International Conference Availability, Reliability and Security (ARES)*, 592-600.
Posner, R. A. (1981). The economics of privacy. *The American Economic Review*, 405-409.
Raimondo, M. A. (2000). The measurement of trust in marketing studies: a review of models and methodologies. *16th Industrial Marketing and Purchasing-conference*, Bath, UK.
Reidenberg, J. R., Breaux, T., Cranor, L. F., French, B., Grannis, A., Graves, J. T., . . . Ramanath, R. (2014). Disagreeable Privacy Policies: Mismatches between Meaning and Users' Understanding. Telecommunications Policy Research Conference.
Richards, N. M., & King, J. H. (2014). Big data ethics. *Wake Forest Law Review*, 40.
Rochelandet, F. (2010). *Économie des données personnelles et de la vie privée*: La Découverte.
Rotenberg, M., & Jacobs, D. (2013). Updating the law of information privacy: The new Framework of the European union. *Harv. JL & Pub. Pol'y, 36*, 605.
Saaty, T. L. (2008). Decision making with the analytic hierarchy process. *International journal of Services Sciences, 1*(1), 83-98.
Sadeh, N., Acquisti, A., Breaux, T. D., Cranor, L. F., McDonald, A. M., Reidenberg, J., . . . Schaub, F. (2014). *Towards Usable Privacy Policies: Semi-automatically Extracting Data Practices From Websites' Privacy Policies*.
Salvas, B. (2002). La protection de la vie privée sur le Web avec P3P: l'arrimage incertain du technique et du juridique.
Schaub, F., Breaux, T. D., & Sadeh, N. (2014). Crowdsourcing the Extraction of Data Practices from Privacy Policies. *AAAI Conference on Human Computation and Crowdsourcing*, 56-57.
Sherchan, W., Nepal, S., & Paris, C. (2013). A survey of trust in social networks. *ACM Computing Surveys, 45*(4), 47.
Song, Y., Kunjithapatham, A., & Messer, A. (2006). Method and apparatus for user centric private data management: Google Patents.
Spiekermann, S., Acquisti, A., Böhme, R., & Hui, K.-L. (2015). The challenges of personal data markets and privacy. *Electronic Markets*, 1-7.
Srivastava, A., & Geethakumari, G. (2013). Measuring privacy leaks in online social networks. *Advances in Computing, Communications and Informatics* (ICACCI), 2095-2100.
Steinfeld, N. (2016). "I agree to the terms and conditions":(How) do users read privacy policies online? An eye-tracking experiment. *Computers in human behavior, 55*, 992-1000.
Stigler, G. J. (1980). An introduction to privacy in economics and politics. *The Journal of Legal Studies*, 623-644.
Sumeeth, M., Singh, R., & Miller, J. (2012). Are Online Privacy Policies Readable? *Optimizing Information Security and Advancing Privacy Assurance: New Technologies: New Technologies*, 91.
Sunyaev, A., Dehling, T., Taylor, P. L., & Mandl, K. D. (2014). Availability and quality of mobile health app privacy policies. *Journal of the American Medical Informatics Association*, amiajnl-2013-002605.

Times, F. How much is your personal data worth? http://www.ft.com/cms/s/2/927ca86e-d29b-11e2-88ed-00144feab7de.html#axzz3nRvkMfus, Accédé le 02 fév. 2015.

Tsai, J. Y., Egelman, S., Cranor, L., & Acquisti, A. (2011). The effect of online privacy information on purchasing behavior: An experimental study. *Information Systems Research, 22*(2), 254-268.

Ur, B., Sleeper, M., & Cranor, L. F. (2012). *{Privacy, Privacidad, Приватност} policies in social media: Providing translated privacy notice. 1st Workshop on Privacy and Security in Online Social Media*.

Urban, G. L., Sultan, F., & Qualls, W. J. (2000). Placing trust at the center of your Internet strategy. *Sloan Management Review, 42*(1), 39-48.

Varian, H. R. (1996). Economic aspects of personal privacy. *Privacy and Self-regulation in the Information Age*.

Wang, Y., Nepali, R. K., & Nikolai, J. (2014). Social network privacy measurement and simulation. *Computing, Networking and Communications Conference*, Honolulu, Hawaii, USA.

Williams, T. L., Agarwal, N., & Wigand, R. T. (2015). Protecting Private Information: Current Attitudes Concerning Privacy Policies. *ASE BigData/SocialInformatics/PASSAT/BioMedCom*, Harvard University.

Zhou, T. (2013). An empirical examination of continuance intention of mobile payment services. *Decision support systems, 54*(2), 1085-1091.

Zimmeck, S., & Bellovin, S. M. (2014). Privee: An architecture for automatically analyzing web privacy policies. *23rd USENIX Security Symposium*.