Ecole doctorale « Sciences, Ingénierie, Santé »

Université de Lyon - Université Jean Monnet

Les Systèmes Multi-Agents Hippocratiques

Mécanismes sociaux entre agents pour la protection de la sphère privée

Thèse

présentée et soutenue publiquement
le 12 Novembre 2009
pour obtenir le grade de

Docteur de l'Université Jean Monnet
(Spécialité Informatique)

par

Ludivine Crépin

Composition du jury

Rapporteurs
Bertrand Braunschweig.....Professeur, Agence Nationale de la Recherche
Philippe Mathieu.....Professeur, Université de Lille

Examinateurs
Thierry Bouron.....Docteur, Orange Labs
Amal El Fallah Seghrouchni.....Professeur, Université Paris VI
Réné Mandiau.....Professeur, Université de Valenciennes
Olivier Boissier..Professeur, Ecole des Mines de Saint-Etienne (co-encadrant)
Yves Demazeau.....Directeur de Recherche, CNRS (co-directeur)
François Jacquenet.....Professeur, Université Jean Monnet (directeur)

Laboratoire Hubert Curien - UMR 5516
Laboratoire d'Informatique de Grenoble - UMR 5217

Table des matières

Table des figures

Chapitre 1

Introduction

Sommaire

1.1 Motivations et cadre de recherche

De nos jours, l'évolution incessante des systèmes d'information induit un traitement automatique et massif des données des utilisateurs, et ce notamment avec l'explosion des technologies développées avec Internet. Citons seulement comme exemples le commerce électronique, les réseaux sociaux ou encore les applications de gestion décentralisée qui requièrent un grand nombre de données pour accéder aux services qu'ils proposent. Ces données, dites *sensibles* car elles touchent à des informations personnelles et nécessitent d'être protégées, représentent actuellement une valeur économique importante et entraînent donc des risques importants d'abus et de fraudes sur leurs manipulations.

Dans cette thèse, nous abordons la problématique liée à la *privacy* en termes de gestion et de protection des données sensibles des utilisateurs. Ce terme anglophone n'ayant pas de définition communément acceptée (vie privée, intimité, confidentialité...), nous nous y référons dans ce manuscrit grâce à la notion de *sphère privée* en français. La sphère privée regroupe toutes les données sensibles d'un utilisateur ainsi que l'ensemble des règles de gestion attachées afin de prendre en compte les souhaits des utilisateurs concernés par ses données.

Du fait des différentes législations en vigueur selon les états et les entreprises, nous proposons d'étudier le respect de la sphère privée en informatique

d'un point de vue moral et éthique, tout en respectant les lignes directrices prédominantes explicitées dans les textes de lois, comme la limitation de la rétention des données sensibles par exemple.

Notre cadre de recherche au sein de cette thèse se focalise sur les systèmes multi-agents centrés utilisateur [Demazeau (2003)]. En effet, dans de tels systèmes, les utilisateurs délèguent une partie de leur sphère privée à un agent autonome qui prend alors en charge la gestion et la protection de leurs données sensibles. De plus, du fait des nombreuses interactions au sein des systèmes multi-agents, les risques encourus de manipulations frauduleuses des données sensibles déléguées deviennent de plus en plus importants en termes de diffusion, d'altération et d'utilisation.

La préservation de la sphère privée représente une problématique ayant trait au cycle de vie entier des données sensibles, de leur stockage à leur suppression et aussi lors de toutes leurs manipulations. En effet, la préservation de la sphère privée requiert d'être prise en considération lors du stockage, de la transaction et du devenir des données sensibles. Cet axe de recherche relève donc aussi bien de la sécurité et du réseau, pour le stockage et la communication des données sensibles principalement, mais également de l'intelligence artificielle pour leur gestion en terme de manipulations possibles respectueuses de la sphère privée.

1.2 Objectifs et contributions

Les contributions apportées dans cette thèse sur le respect de la sphère privée concernent des mécanismes de gestion et de protection des données sensibles venant compléter les recherches développées sur cet axe de recherche en sécurité et réseaux. Ainsi nos préoccupations portent essentiellement sur la transaction et le devenir des données sensibles, en considérant le stockage de ces données comme relevant uniquement de la sécurité.

L'objectif principal de la thèse défendue dans ce manuscrit est de concevoir un modèle et des mécanismes permettant de protéger la sphère privée des utilisateurs grâce à des agents et à la société au sein de laquelle ils appartiennent en termes de gestion et de protection des données sensibles. Ces données sont déléguées par l'utilisateur à un agent autonome qui prend donc le contrôle de celles-ci et les manipule en respectant la sphère privée.

Notre contribution consiste en la définition d'un modèle, les Systèmes Multi-Agents Hippocratiques (HiMAS), prenant en compte le respect de la sphère privée. Ce modèle permet d'intégrer la sphère privée au sein des agents. Ceci s'effectue d'une part par la délégation de l'utilisateur et d'autre part par des mécanismes de raisonnement sur les règles de gestion des données sensibles

déléguées afin de ne pas violer la sphère privée au niveau des manipulations des données sensibles.

Afin de protéger les données sensibles des utilisateurs, nous proposons également des mécanismes sociaux portant sur les manipulations possibles des données sensibles contenues dans la sphère privée des agents. Ces mécanismes viennent compléter les approches, principalement cryptographiques, développées dans les domaines de la sécurité et des réseaux. Au travers de la définition de neuf principes normatifs, les HiMAS régulent les transmissions de données sensibles entre agents et leurs traitements après leur diffusion. Ainsi, le modèle propose un protocole spécifique pour ce type de communication incluant une régulation de comportement pour les agents dans une optique de préservation de la sphère privée.

1.3 Organisation du manuscrit

Les chapitres 2 et 3 se focalisent sur le développement des deux axes relatifs au sujet de thèse : la définition de la sphère privée et les mécanismes de régulation pour les comportements des agents afin de garantir le respect de la sphère privée lors des transmissions et des traitements des données sensibles.

Le chapitre 2 propose une étude de la sphère privée en termes de gestion des données sensibles au sens de [Deswarte et Melchor (2006); Spiekermann et Cranor (2009)] au travers de différents domaines ayant trait à cet axe de recherche. Ces études nous permettent d'établir une synthèse autour du concept de sphère privée en informatique qui n'est pas exclusivement constituée des données sensibles de l'utilisateur. En effet, du fait de la délégation des données, la sphère privée doit également regrouper les éléments de gestion nécessaires à ces données tout en permettant une personnalisation de celle-ci en fonction de l'utilisateur. Nous définissons également chacun des besoins pour assurer le respect de la sphère privée lors de trois phases critiques : le *stockage*, la *transaction* et le *devenir* des données sensibles.

La dernière phase critique du respect de la sphère privée nous amène à présenter, dans le troisième chapitre, l'état de l'art sur les mécanismes de régulation de comportement des agents. En effet, cette phase impose que nous considérions le comportement des agents vis-à-vis des données sensibles qu'ils recueillent après une transaction de données sensibles. Nous présentons donc les deux principales catégories de mécanismes de régulation qui peuvent être adaptés aux systèmes multi-agents : la régulation par autorité tiers et la régulation par contrôle social.

Les chapitre 4, 5 et 6 présentent notre proposition : le modèle de Systèmes Multi-Agents Hippocratiques (HiMAS). Ce modèle a pour but de préserver la sphère privée dans les systèmes multi-agents centrés utilisateur. Les HiMAS

répondent aux besoins des trois phases critiques du respect de la sphère privée à un niveau Agent (ou niveau individuel) complété par un niveau Interaction (ou niveau social) afin de prendre en compte l'ensemble des contraintes imposées par la préservation de la sphère privée lors du raisonnement des agents et de la diffusion des données sensibles.

Le chapitre 4 présente notre définition de la sphère privée ainsi que les fondements du modèle HiMAS. Nous proposons de définir la sphère privée comme l'ensemble des données sensibles d'un agent, que nous nommons éléments de la sphère privée, ainsi que les règles de gestion de ces données (un ensemble d'autorisations, un ensemble de règles et un ensemble de normes portant sur la manipulation des éléments de la sphère privée). Nous proposons aussi à travers cette définition les caractéristiques de la sphère privée : personnelle, personnalisable et contextuelle. Ce chapitre présente ensuite les neuf principes normatifs d'un HiMAS qui garantissent une préservation de la sphère privée à un niveau individuel et social. Nous concluons ce chapitre par le centrage utilisateur des HiMAS en proposant la création de profils pour la délégation des données sensibles des utilisateurs.

Les chapitres 5 et 6 répondent aux questions soulevées par le niveau social des systèmes multi-agents hippocratiques, ou plus précisément les questions relatives aux deux dernières phases critiques du respect de la sphère privée (la *transaction* et le *devenir* des données sensibles) en représentant les principes des HiMAS liés au niveau social, ou plus précisément comment l'étude des interactions entre agents peut aider à la préservation de la sphère privée. Nous présentons les moyens à mettre en œuvre pour appliquer chacun des principes relatifs au raisonnement des agents au travers d'un protocole dédié aux transactions de données sensibles intégrant un contrôle social hippocratique.

Le chapitre 7 présente une validation de notre modèle par le biais des expérimentations menées sur le contrôle social hippocratique. Ce travail nous permet de présenter dans le chapitre 8 l'implémentation de notre modèle dans une application multi-agent de gestion décentralisée d'agenda [Demazeau *et al.* (2006)].

Pour finir, nous exposons nos conclusions et perspectives sur le respect de la sphère privée dans les systèmes multi-agents centrés utilisateur.

Chapitre 2

Sphère privée

Sommaire

L'évolution actuelle de l'informatique, notamment avec l'expansion d'Internet, conduit de plus en plus à un traitement automatique massif des données sensibles des utilisateurs. Le respect de la sphère privée devient donc un problème clé d'un point de vue moral et légal pour les technologies à venir car il

revient à protéger les utilisateurs de certaines activités nuisibles et dangereuses à son égard [Solove (2006)]. Nous proposons dans ce chapitre d'apporter une première définition de la sphère privée à partir de différents travaux relatifs à divers domaines de recherche.

D'un point de vue technologique, le problème de la préservation de la sphère privée sur Internet peut être décomposé en cinq classes de problèmes [Deswarte et Melchor (2006)] (i) protection des adresses IP, (ii) protection de la localisation, (iii) gestion de l'anonymat, (iv) autorisations respectant la vie privée et (v) accès et gestion des données.

Notre thèse relevant de cette dernière classe de problèmes dans le cadre des systèmes multi-agents, nous présentons ici un panorama des différents travaux traitant de l'accès et la gestion de données privées.

Nous commençons ce chapitre en nous tournant vers les sciences sociales et humaines, qui sont à l'origine de l'étude de la sphère privée, afin d'établir une première définition informelle de la sphère privée et cerner ainsi les enjeux possibles de son respect dans les systèmes multi-agents. Ensuite nous abordons la préservation de la sphère privée d'un point de vue opérationnel en informatique. Les sections 2.2 et 2.3 s'intéressent à deux domaines où les utilisateurs sont au centre des technologies du fait de la collecte et du traitement automatique de leurs données sensibles : Internet et les systèmes de gestion de bases de données. Notre panorama se termine en quatrième section par l'étude des différents travaux dans le domaine des systèmes multi-agents.

2.1 Des sciences humaines et sociales à l'informatique

Les sciences humaines et sociales telles que le droit, la sociologie ou la philosophie portent un intérêt particulier à la délimitation de la vie privée. Ces domaines considèrent la notion de *sphère privée*, comme étant souvent relative à la vie privée des utilisateurs. Nous proposons dans cette section d'établir un bref constat des travaux relatifs à la sphère privée dans ces domaines afin de caractériser la sphère privée d'une manière non formelle.

Afin de différencier les données de la sphère privée des autres données, nous les qualifions de *sensibles*.

Définition 2.1 (*Donnée sensible*)

Une donnée sensible est une donnée qui doit être protégée en termes d'accès, de rétention, de diffusion et d'utilisation.

La première caractéristique de la sphère privée est qu'elle est attachée à un sujet et *personnalisable*. En effet, chaque individu décide sur quoi porte sa

propre sphère et comment et quand il veut la gérer, comme le définit A. Westin [Westin (1967)] :

> *"Privacy is the right of individuals to determine for themselves when, how and to what extent informations about them is communicated to other."*

Cette définition est renforcée par [Demeulenaere (2002)] qui reprend le caractère personnel et personnalisable de la sphère privée en établissant le fait que l'individu délimite lui-même sa sphère privée en regard du domaine public selon ses souhaits et ses besoins. Demeulenaere définit donc la délimitation de cette sphère comme totalement subjective selon les personnes et les situations.

Cette dépendance entre la sphère privée et le contexte nous permet également de considérer la *contextualité* comme deuxième caractéristique de la sphère privée, comme le remarque par exemple [Bellotti et Sellen (1993); Palen et Dourish (2003)].

Une sphère privée est donc *personnelle* mais personnelle à qui ? Nous soulevons ici la question de la propriété : qui est propriétaire d'une donnée sensible ? D'un point de vue légal, comme par exemple dans [Thomson (1975)], les droits de la sphère privée sont associés aux droits à la propriété : nous sommes d'un point de vue légal les propriétaires de nos données personnelles, sans pour autant en être les propriétaires physiques.

Exemple 2.1

En France, une photo prise d'un passant est considérée comme une donnée personnelle de ce passant et non du photographe. De ce fait, les droits sur cette photo appartiennent au passant et non au photographe.

Associer des droits de propriété à la sphère privée nous amène à étudier son *contrôle* comme dernière caractéristique. Par exemple, selon [Westin (1967); Lessig (2000)], le contrôle en termes d'accès et de diffusion est propre à chacun et ne doit pas dépendre d'une personne non propriétaire de la donnée. A cette définition, s'ajoute celle de [Warren et Brandeis (1985)] qui détermine la possibilité d'interdire la diffusion de ces informations en étendant ce contrôle aux différentes entités qui possèdent des données sensibles d'un tiers.

Adaptant ces propositions à l'usage des nouvelles technologies telles qu'Internet, S. Baase réunit l'ensemble de ces critères dans l'ouvrage "*A gift of fire*" [Baase (2002)]. Tout en incluant la présence de faits et de données personnelles dans la sphère privée des utilisateurs comme par exemple leur identité, Baase donne à la protection de la sphère privée trois significations possibles :

1. L'absence d'intrusion dans l'intimité,
2. Le contrôle des données nous concernant,
3. L'absence de surveillance dans notre intimité.

Cette définition est confirmée en informatique par [Müller (2006)] qui définit la sphère privée comme suit :

"Privacy is the possibility to control the distribution and use of personal data."

Ces travaux nous permettent de définir de manière informelle la vision de la sphère privée que nous considérons tout au long de cette thèse.

Définition 2.2 (*Sphère privée*)

La sphère privée est constituée de toutes les données qu'un utilisateur estime être sensibles et qui doivent donc être protégées lors de leur utilisation et de leur diffusion. Elle intègre également les éléments de gestion de ces données afin d'assurer leur contrôle lors de la protection de ces données.

A l'instar de [Westin (1967); Thomson (1975); Warren et Brandeis (1985); Bellotti et Sellen (1993); Lessig (2000); Baase (2002); Demeulenaere (2002); Palen et Dourish (2003); Müller (2006)], nous adoptons l'idée selon laquelle la sphère privée possède comme caractéristiques le fait d'être :

- *personnelle et personnalisable : l'entité concernée juge de ce qu'elle doit contenir,*
- *contextuelle : elle dépend du contexte courant dans lequel l'entité interagit avec les autres entités,*
- *contrôlable : l'entité concernée contrôle les manipulations des données sensibles de sa sphère privé.*

La sphère privée donne donc la possibilité aux utilisateurs de contrôler l'utilisation et la diffusion de leurs données sensibles lors des traitements automatiques liés à l'utilisation massive des technologies informatiques. Nous considérons également que seule l'entité concernée par les données sensibles détient les droits de propriété sur ces données.

Les problèmes relatifs au respect de la sphère privée nous conduisent également à porter notre attention sur ce que peuvent être les violations de la sphère privée. L'ensemble des travaux présentés, notamment [Baase (2002)], nous apportent plusieurs éléments de réponse que nous présentons à travers la définition d'une entité ayant un comportement *suspicieux*.

Définition 2.3 (*Comportement suspicieux*)

Une entité au comportement suspicieux est une entité violant une sphère privée en accédant, en altérant, en endommageant ou en détruisant des données sensibles de la sphère privée d'une autre entité que cette dernière ne lui ai donné les autorisations pour manipuler ses données sensibles.

Maintenant que nous avons fourni une première définition de la sphère privée et des violations possibles envers cette sphère, nous pouvons étudier ce concept d'un point de vue informatique avec des technologies où l'utilisateur est au centre du fait du traitement automatique de ses données sensibles. Nous commençons par présenter la préservation de la sphère privée sur Internet car son évolution actuelle a renforcé l'importance de la notion de données sensibles. Ensuite, nous présentons ce respect dans les systèmes de gestion de bases de données, qui détiennent généralement les données sensibles que les utilisateurs fournissent aux services, que ce soit pour les applications de gestion ou les applications Internet par exemple. Pour finir, nous abordons ce respect dans le cadre de recherche de cette thèse, les systèmes multi-agents.

2.2 Respect de la sphère privée sur Internet

Nous présentons dans cette section le standard du W3C concernant le respect de la sphère privée, la plate-forme pour les préférences de confidentialité, suivi de l'approche de ce problème dans les réseaux pair-à-pair (P2P) afin d'en étudier les apports possibles pour les systèmes multi-agents.

2.2.1 Plate-forme pour les préférences de confidentialité

La plate-forme pour les préférences de confidentialité (P3P pour *Platform for Privacy Preferences*) [W3C (2002b); Cranor (2002)] est une initiative du consortium W3C visant à développer un standard pour gérer les informations de la sphère privée, principalement au sein des sites Internet de commerce électronique, en se limitant aux côtés client et serveur lors des communications mettant en jeu des données sensibles. Dans la suite de ce manuscrit, de telles communications sont appelées *transactions de données sensibles*.

Grâce à la plate-forme pour les préférences de confidentialité, un utilisateur définit ses *préférences*, formalisées grâce à un fichier XML, pour la gestion de ses données personnelles sensibles. Une préférence définit les contraintes que l'utilisateur souhaite imposer sur la gestion et la manipulation de ses données personnelles. Elle permet à l'utilisateur d'indiquer les données qu'il estime être sensibles et qui doivent donc être protégées en définissant les objectifs pour lesquels elles peuvent être utilisées et diffusées ainsi que le temps de rétention autorisé pour la conservation de ses données sensibles par un tiers.

De son côté, le serveur ayant à gérer ces données définit une *politique* dans un fichier XML, et s'engage à s'y tenir. Une politique spécifie les objectifs de

```
<POLICIES xmlns="http://www.w3.org/2002/01/P3Pv1">
 <POLICY name="pourNavigateur"
      discuri="http://www.exemple.com/confidentialite_navigation.html"
      xml:lang="en">
  <ENTITY>
   <DATA-GROUP>
    <DATA ref="#business.name">CatalogueExemple</DATA>
    <DATA ref="#business.contact-info.postal.street">4000 Lincoln Ave.</DATA>
    <DATA ref="#business.contact-info.postal.city">Birmingham</DATA>
    <DATA ref="#business.contact-info.postal.stateprov">MI</DATA>
    <DATA ref="#business.contact-info.postal.postalcode">48009</DATA>
    <DATA ref="#business.contact-info.postal.country">USA</DATA>
    <DATA ref="#business.contact-info.online.email">catalogue@example.com
    </DATA>
    <DATA ref="#business.contact-info.telecom.telephone.intcode">1</DATA>
    <DATA ref="#business.contact-info.telecom.telephone.loccode">248</DATA>
    <DATA ref="#business.contact-info.telecom.telephone.number">3926753</DATA>
   </DATA-GROUP>
  </ENTITY>
  <ACCESS><nonident/></ACCESS>
  <DISPUTES-GROUP>
   <DISPUTES resolution-type="independent"
     service="http://www.sceauconfidentiel.example.org"
     short-description="sceauconfidentiel.example.org">
    <IMG src="http://www.sceauconfidentiel.example.org/Logo.gif"
    alt="Le logo de SceauConfidentiel"/>
    <REMEDIES><correct/></REMEDIES>
   </DISPUTES>
  </DISPUTES-GROUP>
  <STATEMENT>
   <PURPOSE><admin/><develop/></PURPOSE>
   <RECIPIENT><ours/></RECIPIENT>
   <RETENTION><stated-purpose/></RETENTION>
                        <!-- Remarquez également que la politique
                        de confidentialité du site lisible
                        par un humain DOIT mentionner la purge
                        des données toutes les deux semaines,
                        ou fournir un lien vers cette information. -->
   <DATA-GROUP>
    <DATA ref="#dynamic.clickstream"/>
    <DATA ref="#dynamic.http"/>
   </DATA-GROUP>
  </STATEMENT>
 </POLICY>
</POLICIES>
```

FIG. 2.1 – Exemple de politique spécifiée pour la plate-forme des préférences de confidentialité.

la collecte de données qui va être réalisée, ainsi que la durée de conservation des données privées recueillies.

Exemple 2.2

> *La figure 2.1 (extraite de http ://www.yoyodesign.org/doc/w3c/p3p1/) représente la politique du site www.exemple.com. Cette politique indique que les données collectées sont les coordonnées des internautes. Ces données sont anonymes et ne sont utilisées que pour l'amélioration du site. Ce site est également certifié par une autorité de contrôle du respect de la sphère privée, SceauConfidentielExemple. La dernière information contenue dans cette politique concerne la rétention des données qui ne sont gardées en mémoire que deux semaines.*

Afin d'établir une correspondance automatique entre politique et préférence, le W3C propose l'utilisation du moteur d'inférence fondé sur le langage APPEL (*A P3P Preference Exchange Language.*) qui permet aux utilisateurs de définir leurs préférences [W3C (2002a)].

Exemple 2.3

> *La préférence exprimée dans la figure 2.2 (extraite de http ://www.yoyodesign.org/doc/w3c/p3p1/) déclare que toutes les données peuvent être collectées quel que soit l'objectif de la transaction de données sensibles.*

Ce moteur établit une correspondance entre une politique et une préférence permettant ainsi d'un point de vue théorique le respect de la préférence par le site qui recueille les données sensibles. Cependant cette mise en correspondance automatique pose quelques problèmes au niveau de la sémantique de la spécification des objectifs. En effet, aucune contrainte sémantique n'est définie pour l'expression des objectifs au sein de la plate-forme pour les préférences de confidentialité. De ce fait, ce moteur d'inférence n'est pas toujours en mesure de garantir une analyse exacte des objectifs de la transaction de données sensibles et n'aboutit pas toujours à un résultat satisfaisant pour les utilisateurs et les serveurs.

2.2.2 Réseaux pair-à-pair

Les réseaux pair-à-pair permettant le partage de données et donc notamment des données sensibles, ils sont au cœur des problèmes liés au respect de la sphère privée.

D'un point de vue général, les réseaux pair-à-pair permettent de gérer la diffusion des données sensibles des utilisateurs grâce à une topologie ami-à-ami (F2F pour *Friend-to-Friend.*). Cette topologie permet aux utilisateurs d'échanger leurs données uniquement avec d'autres utilisateurs connus et dignes de

```
<APPEL:RULESET xmlns:APPEL="http://www.w3.org/2000/APPEL"
               crtdby="W3C" crtdon="15-March-2000 16:41:21 GMT">

  <APPEL:RULE behavior="inform"
     description="Service collects data for marketing, profiling,
                  or "other" purposes.">
    <POLICY APPEL:connective="and">
      <STATEMENT APPEL:connective="and">
        <PURPOSE APPEL:connective="or">
             <contact/><profiling/><other/>
        </PURPOSE>
      </STATEMENT>
    </POLICY>
  </APPEL:RULE>

  <APPEL:RULE behavior="inform"
     description="Service shares information with legal entities
                  following different practices, public fora, or
                  unrelated third parties.">
    <POLICY APPEL:connective="and">
      <STATEMENT APPEL:connective="and">
        <RECIPIENT APPEL:connective="or">
          <other/><public/><unrelated/>
        <RECIPIENT/>
      </STATEMENT>
    </POLICY>
  </APPEL:RULE>

  <APPEL:RULE behavior="warn"
     description="Site collects healthcare information.">
    <POLICY APPEL:connective="and">
      <STATEMENT APPEL:connective="and">
       <DATA-GROUP APPEL:connective="or">
            <DATA category="health"/>
        </DATA-GROUP>
      </STATEMENT>
    </POLICY>
  </APPEL:RULE>

  <APPEL:RULE behavior = "accept"
    description = "privacy policy matches Information Only preferences">
      <APPEL:OTHERWISE/>
  </APPEL:RULE>

</APPEL:RULESET>
```

FIG. 2.2 – Exemple de préférence explicitée avec le langage APPEL.

confiance tout en assurant leur anonymat. La mise en place des réseaux ami-à-ami a été développée dans des réseaux pair-à-pair tels que RetroShare[1], Freenet[2], NodeZilla[3], GNUnet[4] et OneSwarm [Isdal *et al.* (2009)] par exemple.

Une autre approche de la sphère privée est proposée dans [Damiani *et al.* (2004)] qui considère les identités virtuelles des utilisateurs comme les données sensibles dans le contexte précis de la lutte contre le pourriel (*spam* en anglais). Il propose une approche décentralisée pour filtrer les pourriels par le biais d'une architecture de type pair-à-pair. Les serveurs partagent ainsi leur connaissance sur les envoyeurs de pourriels grâce à cette architecture. Cette approche permet de réduire le niveau de pourriel en détectant de plus en plus de pourriels en utilisant des techniques collaboratives de filtrage.

Citons un dernier travail qui exploite les réseaux pair-à-pair et les agents mobiles pour le respect de la sphère privée lors du stockage des données sensibles. [Pommier et Bourdon (2008)] propose de modéliser une donnée par un système multi-agent où chaque agent représente un fragment de cette donnée. Comme le propose [Deswarte *et al.* (1991)], chaque agent est ensuite dupliqué pour la pérennité de la donnée, les agents sont par la suite disséminés d'un point de vue spatial et temporel, et ce de manière aléatoire, dans un réseau pair-à-pair. Le déplacement des agents dans le réseau n'est pas prévisionnel, ce qui empêche toute attaque. Les agents déposent des phéromones sur chaque nœud visité, ce qui permet de déterminer un niveau de confiance pour chaque nœud. Les agents se déplacent ensuite en vol de nuée d'oiseaux en choisissant les nœuds qui ont le moins de phéromones et le plus haut niveau de confiance afin que la donnée ne puisse pas être reconstruite en récupérant l'ensemble des fragments. Notons que les systèmes basés sur ces principes demandent un temps d'exécution important et ne peuvent pas convenir aux attentes des utilisateurs en termes de performance malgré un niveau de sécurité performant.

2.2.3 Discussion

La plate-forme pour les préférences de confidentialité permet de spécifier des contraintes sur la gestion de données sensibles d'une manière générique pour les utilisateurs et les serveurs. La transaction de données sensibles requiert un accord entre l'utilisateur et l'entité qui recueille les données sensibles sur le but de cette transaction ainsi que la durée de rétention de ces données. Cet accord se traduit par une mise en correspondance réussie entre la politique du site et la préférence de l'utilisateur.

[1] http ://retroshare.sourceforge.net/
[2] http ://freenetproject.org/
[3] http ://www.nodezilla.net/
[4] http ://gnunet.org/

La plate-forme pour les préférences de confidentialité soulève toutefois un problème de sémantique pour ce type de transaction : les entités inscrites dans une transaction de données sensibles doivent être en mesure de se comprendre sur l'expression des objectifs de cette transaction afin de trouver un accord. Or aucune contrainte sémantique n'étant spécifiée, ces objectifs sont spécifiés généralement en langage naturel, comme le montrent les deux précédents exemples (figures 2.1 et 2.2), ce qui peut provoquer une barrière de la langue, notamment au niveau des utilisateurs qui doivent pouvoir comprendre les politiques des sites qui collectent leurs données sensibles.

Pour conclure sur cette technologie, nous rejoignons certaines critiques émises depuis la proposition de la plate-forme pour les préférences de confidentialité [Thibadeau (2000)]. En effet, ce standard fournit seulement une description des accords passés entre un client et un serveur sur une transaction mettant en œuvre des données sensibles. Aucune vérification n'est réalisée sur l'engagement du serveur. De plus, l'intégration de l'utilisateur est assez pauvre au niveau de la personnalisation de la sphère privée car les préférences sont définies de manière générique et ne permettent pas à l'utilisateur de les définir contextuellement en fonction du site qui recueille les données.

Au bilan, nous pouvons remarquer que le respect de la sphère privée des utilisateurs sur Internet demande à être considéré lors du stockage des données sensibles, lors de la diffusion de ces données (avec un engagement de l'entité qui les demande) et également lors du traitement des données sensibles diffusées.

2.3 Respect de la sphère privée dans les systèmes de gestion de bases de données

La préservation de la sphère privée des utilisateurs est un problème primordial dans le cadre des systèmes de gestion de bases de données car ces dernières gèrent principalement les données sensibles des utilisateurs. Nous proposons de donner une vision de cette problématique par l'étude de deux propositions principales de ce domaine : le contrôle d'accès à base de rôles [Ferraiolo et Kuhn (1992); Sandhu *et al.* (1996)] et les bases de données hippocratiques [Agrawal *et al.* (2002)] .

2.3.1 Contrôle d'accès à base de rôles

Le contrôle d'accès à base de rôles (RBAC pour *Role Based Access Control*) [Ferraiolo et Kuhn (1992); Sandhu *et al.* (1996)] a été développé dans le but de permettre le contrôle dynamique des accès aux données dans les organisations dynamiques et les systèmes d'information complexes.

Le contrôle d'accès à base de rôles s'appuie sur une politique de contrôle d'accès aux données par le biais de la notion de *rôle* qui représente la fonction des utilisateurs. Les rôles définissent un ensemble de permissions permettant d'accéder ou non aux données contenues dans la base de données. Une relation d'affectation entre les rôles et les utilisateurs permet d'attribuer l'ensemble de permissions du rôle aux utilisateurs. Une permission est représentée par un ensemble d'opérations possibles sur un objet donné.

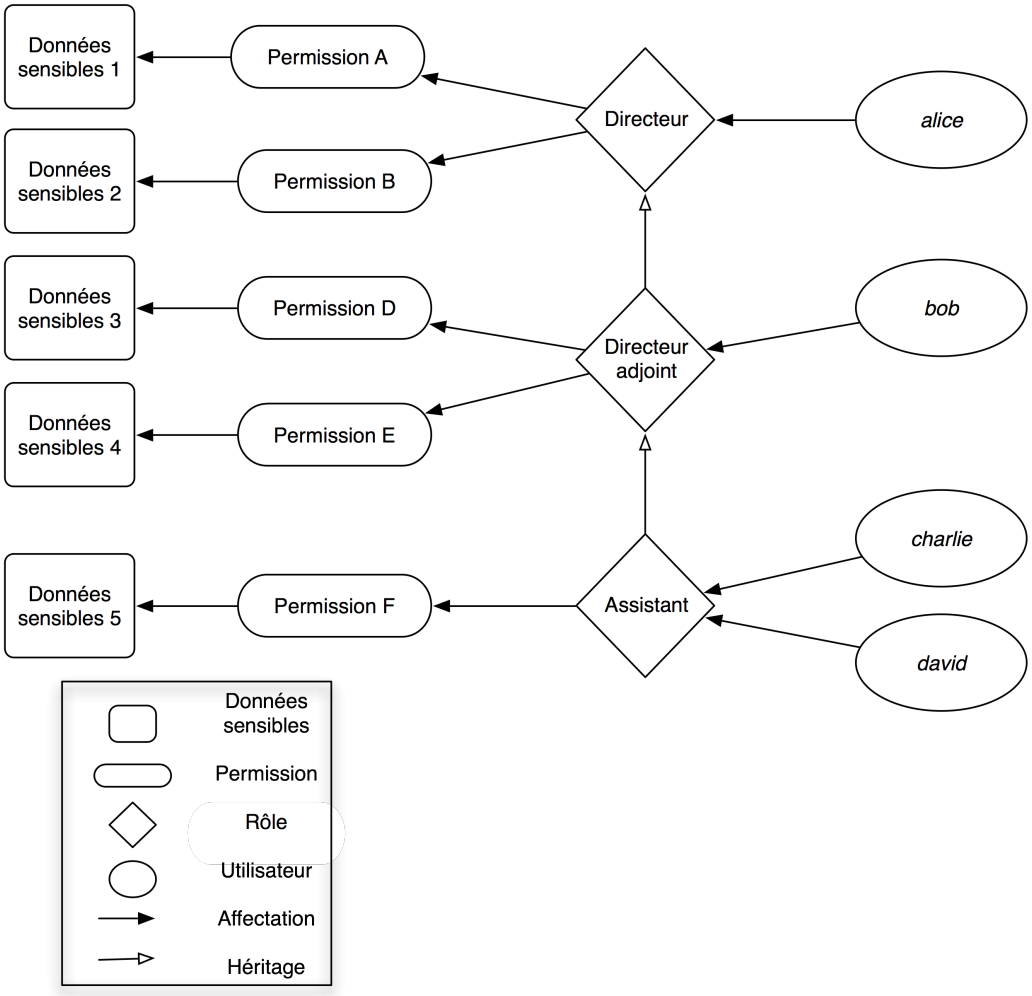

FIG. 2.3 – Exemple de contrôle d'accès à base de rôles.

Pour assurer une gestion dynamique et flexible de l'accès aux données, le contrôle d'accès à base de rôles utilise des sessions pour représenter les connexions des utilisateurs. Chaque session représente une relation entre un utilisateur et un rôle et permet ainsi de lui attribuer dynamiquement un ensemble de permissions.

Notons que les rôles dans le contrôle d'accès à base de rôles peuvent définir un lien hiérarchique entre les rôles selon une organisation donnée. Cette fonctionnalité permet de ne pas redéfinir des permissions pour rôle qui hérite des permissions d'un autre et simplifie donc les sessions attribuant ces permissions aux rôles, et donc aux utilisateurs.

Exemple 2.4

La figure 2.3 représente un exemple de contrôle d'accès à base de rôles. Les utilisateurs (alice, bob, charlie et david) sont affectés à un rôle donné (directeur, directeur adjoint ou assitant), ce qui leur permet d'obtenir un ensemble de permissions sur les données sensibles de la base de données.

Les assistants, charlie et david, n'ont que la permission F sur les données sensibles 5.

Bob, le directeur adjoint, possède la même permission que les assistants du fait de l'héritage entre ces deux rôles. Cet utilisateur possède également deux autres permissions, les permissions D et E qui s'appliquent respectivement sur les données sensibles 3 et 4.

La directrice, alice, possède les mêmes permissions que le directeur adjoint bob. En plus, cet utilisateur détient deux autres permissions, A et B, sur les données sensibles 1 et 2. Dans cet exemple, il s'agit du seul utilisateur qui a accès à l'ensemble des données sensibles de la base de données, les autres ayant seulement une vision partielle de cet ensemble.

2.3.2 Bases de données hippocratiques

Englobant les principes de la plate-forme pour les préférences de confidentialité et du contrôle d'accès à base de rôles, le modèle de bases de données hippocratiques [Agrawal *et al.* (2002)] renforce ces deux derniers travaux dans le domaine des bases de données en définissant dix principes pour le respect de la sphère privée.

Pour cela, Agrawal *et al.* s'inspirent du serment d'Hippocrate[5] en l'adaptant à la proposition d'Allan Westin [Westin (1967)] qui définit le caractère personnel de la sphère privée. Ainsi un système de gestion de base de données pour être hippocratique doit se plier à un ensemble de principes moraux et éthiques (tel le médecin envers ses patients). Au nombre de dix, ces principes définissent un modèle de système de gestion de bases de données prenant en compte la gestion et la protection de la sphère privée des utilisateurs :

1. **Consentement de l'utilisateur** : un système de gestion de bases de données hippocratiques s'engage à ce qu'un utilisateur puisse donner son accord pour chacune des données propres qui sont recueillies.

[5]Dont une version est disponible sur http ://www.chu-rouen.fr/documed/serment.html.

2. **Connaissance des différents objectifs** : un système de gestion de bases de données hippocratiques s'engage à ce qu'un utilisateur puisse être au courant des objectifs dans lesquels cette collecte est réalisée afin de donner son consentement sur la collecte de données (cf. principe 1).
3. **Limitation de la collecte des données** : un système de gestion de bases de données hippocratiques s'engage à limiter les données qu'il recueille à celles nécessaires aux objectifs transmis au donneur (cf. principe 2).
4. **Limitation de l'utilisation des données** : un système de gestion de bases de données hippocratiques s'engage à restreindre l'utilisation des données recueillies aux objectifs transmis au donneur.
5. **Limitation de la diffusion des données** : un système de gestion de bases de données hippocratiques s'engage à diffuser les données sensibles des utilisateurs uniquement si cela s'avère nécessaire pour réaliser un objectif donné.
6. **Limitation de la rétention des données** : un système de gestion de bases de données hippocratiques s'engage à ne conserver les données que pendant le laps de temps correspondant au délai nécessaire à leur utilisation.
7. **Sécurité** : un système de gestion de bases de données hippocratiques assure un niveau élevé de sécurité afin d'éviter toute manipulation frauduleuse. La mise en place d'un tel niveau de sécurité intervient lors des accès aux bases et lors du stockage des données sensibles.
8. **Transparence des données** : un système de gestion de bases de données hippocratiques s'engage à ce qu'un utilisateur puisse avoir accès aux données qui lui sont propres afin de connaître celles qui sont encore stockées. L'utilisateur doit pouvoir effectuer des mises à jour de ses données.
9. **Exactitude des données** : un système de gestion de bases de données hippocratiques s'engage sur l'exactitude et la cohérence des données stockées pendant leur durée de rétention.
10. **Conformité** : un système de gestion de bases de données hippocratiques s'engage à ce que tout utilisateur puisse avoir la possibilité de vérifier le respect de chacun des principes précédents.

[Agrawal *et al.* (2002)] présente une ligne directrice pour la création d'un système de gestion de bases de données hippocratiques en définissant les problèmes-clé du respect de la sphère privée grâce aux dix principes présentés précédemment. Certains de ces problèmes ont fait l'objet de travaux ultérieures :

- Le principe 5, la limitation de la diffusion des données sensibles, a été étudié dans [LeFevre *et al.* (2004)].

- L'utilisation d'un tatouage des données sensibles qui permet d'appliquer le principe 6, la limitation de la rétention des données sensibles a été présenté dans [Agrawal *et al.* (2003)]. Ce tatouage altère les données si elles viennent à être manipulées après un certain laps de temps représentant le temps de rétention accordé, rendant ainsi les données inutilisables.
- L'intégration de l'utilisateur par la création d'un langage de préférences utilisateur impliquant des restrictions sur les colonnes, les lignes et/ou sur les cellules des tables d'une base de données afin d'expliciter aux utilisateurs les principes 1 à 6 a été étudié dans [Agrawal *et al.* (2005)]. Ce langage permet également une traduction des politiques des serveurs définies par le biais de la plate-forme pour les préférences de confidentialité et ainsi de faire une mise en correspondance valide avec les préférences des utilisateurs.

2.3.3 Discussion

Le contrôle d'accès à base de rôles ne s'intéresse pas à la phase de collecte des données sensibles mais uniquement à la consultation de ces dernières après leur collecte. Même si le contrôle d'accès à base de rôles impose plus de contraintes sur l'utilisation de données sensibles que la plate-forme pour les préférences de confidentialité, nous pouvons regretter l'absence de contrôle sur le devenir de ces données après leur accès, même si des travaux comme [Ahmed et Tripathi (2003)] instaure un système de vérification des permissions selon les utilisateurs. Le contrôle d'accès à base de rôles ne permet donc pas de garantir le respect de la sphère privée après la collecte des données sensibles.

Les bases de données hippocratiques prennent en compte les principaux aspects moraux du respect de la sphère privée au travers des dix principes qu'elles imposent. Les données sensibles sont préservées lors de leur stockage, lors des communications et leur devenir est également pris en considération. De plus, les principes imposant une limitation permettent de diminuer les risques de comportement suspicieux en limitant les manipulations (collecte, utilisation, diffusion et rétention) des données sensibles collectées.

Cependant nous pouvons noter qu'aucun mécanisme de sanction envers les entités du système ne respectant pas un, plusieurs ou les dix principes n'est abordé dans ces travaux, ce qui n'assure pas totalement le respect de la sphère privée. En effet, dans le contexte des systèmes de gestion de bases de données, la prévention des utilisateurs envers les comportements suspicieux par le dixième principe ne peut être assurée si aucune entité suspicieuse découverte n'est dénoncée aux autres utilisateurs. De plus, les bases de données hippocratiques utilisant principalement sur des technologies cryptographiques, le coût

en performance d'un tel système de gestion reste encore très élevé pour une mise en œuvre efficace.

2.4 Respect de la sphère privée dans les systèmes multi-agents

Le respect de la sphère privée devient un aspect important dans le domaine des systèmes multi-agents du fait qu'il repose sur les interactions entre agents logiciels autonomes pouvant posséder chacun des données sensibles [Wooldridge et Jennings (1995)]. Nous présentons dans cette section quelques uns des travaux portant sur les bases de données hippocratiques, sur les approches cryptographiques et les agents garants dans le contexte des systèmes multi-agents.

Notons que nous n'abordons pas dans cette section les systèmes multi-agents normatifs, où le comportement des agents est régi par un ensemble de normes, car, à la suite de nos recherches, nous n'avons pas trouvé de référence traitant du respect de la sphère privée dans ce type de systèmes multi-agents. Nous les aborderons cependant dans le chapitre suivant lorsque nous traiterons de la réputation dans les systèmes multi-agents.

2.4.1 Bases de données hippocratiques et systèmes multi-agents

[Massacci *et al.* (2007)] réponde à quelques unes des limites des travaux portant sur les systèmes de gestion de bases de données hippocratiques, dont l'absence de mécanismes de sanction. TROPOS est une méthodologie de développement de logiciel orientée agent prenant en compte les besoins non fonctionnels lors de la conception d'un système multi-agent.

Pour ce faire, Massacci *et al.* utilisent les relations de confiance afin de diminuer le nombre de transactions de données sensibles lorsqu'une de ces relations n'est pas satisfaisante, c'est-à-dire que la base de données a enfreint trop de principes de trop nombreuses fois.

2.4.2 Approches cryptographiques

Dans les problèmes de satisfaction de contraintes distribuées, le respect de la sphère privée est assimilé à la protection de données lors du partage de connaissance entre agents. Ce respect est assuré par la diminution du partage et donc par l'augmentation des secrets (états courants cachés des agents) entre les agents. Avec cette approche, les algorithmes des solveurs sont de moins en

moins efficaces, ne donnant pas de solution dans la majorité des cas, pour un respect de plus en plus complet, comme le remarque [Freuder *et al.* (2001)].

[Freuder *et al.* (2001)] reprend également les principes de limitation des bases de données hippocratiques (cf. principes 3.-4.-5.-6.) afin de diminuer les risques encourus en diminuant le plus possible les manipulations des données sensibles par les agents. Chaque problème est dans un premier temps abordé avec le moins de données sensibles possibles. Si aucune solution n'est trouvée, les limitations sont relâchées en intégrant plus de données sensibles et ce jusqu'à ce que le solveur trouve une solution. Nous pouvons déjà constater que cette proposition induit un coût assez onéreux pour les performances, le solveur devant être relancé à chaque échec avec un ensemble de données croissant.

Exemple 2.5

Prenons comme exemple illustratif le coloriage d'une carte. Chaque zone de la carte doit être assimilée à une couleur (bleu, jaune, rouge, vert) qui est une donnée sensible pour certaines zones représentées par des agents. Aucune zone voisine ne doit avoir la même couleur.

La première exécution du solveur va considérer toutes les couleurs estimées comme sensibles et ne va pas permettre leur diffusion. Le manque de communication entre les agents va donc entraîner un problème, car ne connaissant pas la couleur de son voisin, un agent ne peut pas savoir s'il possède la même ou non.

Ainsi jusqu'à ce qu'une solution soit trouvée, la taille de l'ensemble des données sensibles diminue à chaque exécution, ce qui permet aux agents une meilleure connaissance de la carte et donc de s'approprier la couleur adéquate.

Ce type d'approche pour la préservation de la sphère privée est également repris dans les travaux de Yokoo *et al.* [Yokoo *et al.* (2005)] même si les algorithmes développés se révèlent être trop coûteux.

Pour diminuer ce coût, de nombreuses approches proposent d'utiliser des algorithmes de permutation aléatoire entre les agents pour assurer le respect de la sphère privée comme par exemple [Silaghi et Rajeshirke (2004); Greenstadt *et al.* (2006)].

2.4.3 Utilisation d'agents garants

Dans le cadre des systèmes multi-agents, en plus d'un niveau de sécurité élevé durant le stockage et les transactions de données sensibles, [Bergenti (2005)] propose l'intervention d'un agent garant pour la transmission d'une donnée sensible entre deux agents afin d'assurer le respect de leur sphère privée. Cet agent représente en fait un intermédiaire entre les deux autres agents de la transaction des données sensibles, il garantit les données transmises ainsi que

le respect des volontés des deux parties. L'avantage majeur de cette technique réside dans le fait que les agents peuvent concentrer leur confiance en une seule entité, l'agent garant.

Une autre approche est présentée dans [Rezgui *et al.* (2002)] qui propose un modèle permettant d'assurer le respect de la sphère privée au sein des services Web. Pour ce faire, il définit plusieurs éléments essentiels à la transaction de données sensibles entre un utilisateur et un Service Web et entre services Web. L'utilisateur définit un profil dynamique permettant de qualifier les données en termes de protection. Ensuite, le service définit une politique pour les données qu'il souhaite recueillir. La transaction de données sensibles entre l'utilisateur et le service Web passe par un agent intermédiaire. Cet agent crée des filtres à partir des profils de l'utilisateur afin de garantir la protection de ses données sensibles lors de chaque transaction en vérifiant si la politique du service correspond bien aux souhaits de l'utilisateur. Ce type d'approche est également utilisé dans [Cissée et Albayrak (2007)] qui propose d'utiliser un agent garant avec des techniques de filtrage et des profils utilisateurs pour préserver la sphère privée lors des interactions.

Une dernière utilisation d'agents garants est proposée dans [Piolle (2009)] où chaque agent assiste un utilisateur qui lui délègue ses données sensibles. Pour le respect de la sphère privée, Piolle développe un modèle d'agent spécifique, les agents PAw (pour *Privacy-Aware*), utilisant la logique DLP (pour *Deontic Logic for Privacy*) afin de prendre en considération les différentes réglementations concernant la sphère privée et provenant de différentes autorités. Cette proposition permet une gestion des données sensibles prenant en compte les conflits possibles entre les normes et les réglementations. Ces agents sont implémentés dans une architecture d'informatique de confiance[6] [Chen *et al.* (2009)] pour assurer la sécurité des transactions.

2.4.4 Discussion

Les approches cryptographiques soulèvent des problèmes de sécurité lors du respect de la sphère privée. Des méthodes cryptographiques sont nécessaires à ce respect mais, étant trop coûteuses, elle doivent être assistées par d'autres méthodes de plus haut niveau. Ces approches s'intéressent principalement au traitement des données sensibles sans attacher d'importance aux problématiques liées à la collecte et au devenir de ces données ce qui est pourtant primordiale dans le contexte du respect la sphère privée. Cette vision de la préservation de la sphère privée ne permet pas de prendre en considération l'existence de l'utilisateur. En effet, ces technologies n'intègrent pas les préfé-

[6]En anglais, Trusted Computing.

rences de l'utilisateur en prenant entièrement le contrôle de la gestion et du traitement de ses données sensibles.

Ces approches présentent toutefois le même principal point faible que celui mis en avant dans le contexte de la plate-forme pour les préférences de confidentialité : aucune vérification n'est effectuée une fois la donnée sensible transmise.

2.5 Synthèse : nécessité liée au respect de la sphère privée

Une première constatation est que la sphère privée n'est pas uniquement constituée des données sensibles de l'utilisateur. Du fait de la délégation de données sensibles à une entité informatique, cette sphère doit également regrouper les éléments de gestion nécessaires à celles-ci tout en permettant une personnalisation [Westin (1967); Thomson (1975); Warren et Brandeis (1985); Lessig (2000); Baase (2002); Demeulenaere (2002); Müller (2006)] de la part de l'utilisateur.

Ces études nous indiquent également que la manière de recueillir des données sensibles et leur traitement ne constituent pas les principales étapes à prendre en considération dans le contexte de la préservation de la vie privée. En effet, le respect de la sphère privée demande à être étudié lors de trois phases critiques comme le souligne [Spiekermann et Cranor (2009)] :

1. le **stockage** des données sensibles,
2. la **transaction** des données sensibles,
3. le **devenir** des données sensibles qui est relatif au comportement de l'entité qui reçoit de telles données.

Lors de ces trois phases, nous pouvons constater que plusieurs agents entrent en jeu. En effet, comme le montre la figure 2.4, toute la société d'agents est concernée par le respect de la sphère privée. Lors du *stockage* seul un agent doit assurer cette phase. La phase *transaction* demande à ce que deux agents préservent la sphère privée et la phase *devenir* fait appel à toute la société d'agents pour garantir la non présence de comportement suspicieux.

2.5.1 Stockage des données sensibles

D'un point de vue informatique, les besoins en respect de la sphère privée lors du stockage des données sensibles sont de l'ordre de la sécurité car aucune entité ne doit avoir accès à une donnée sensible sans le consentement de son sujet.

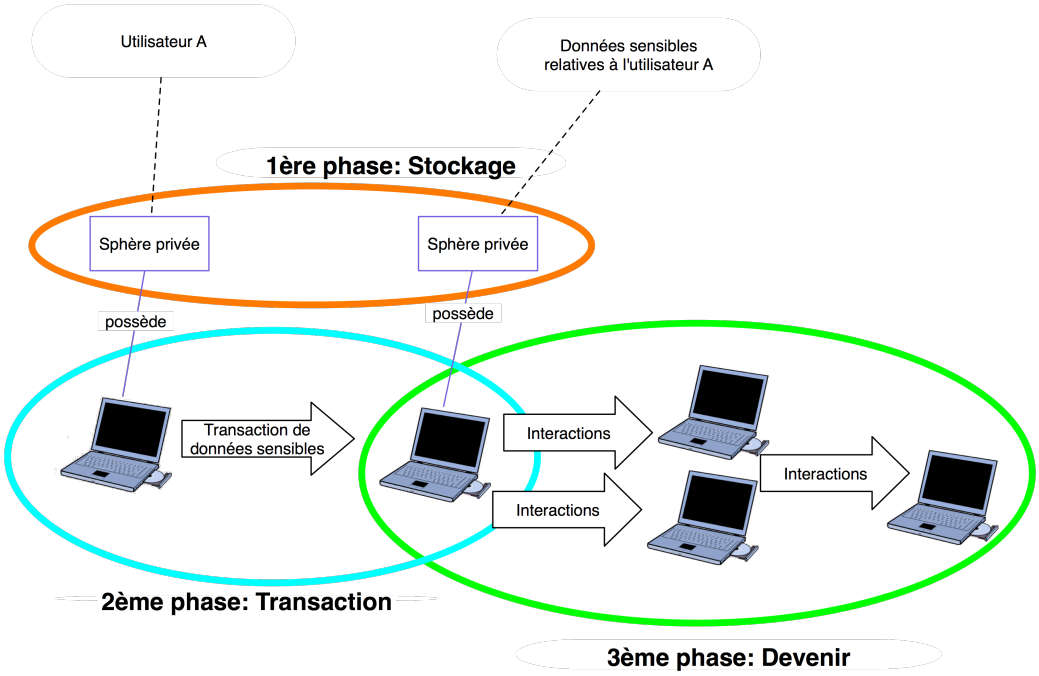

FIG. 2.4 – Phases critiques du respect de la sphère privée.

Cet aspect de la préservation des données sensibles des utilisateurs est également explicité d'un point de vue légal. En effet, la législation française, avec notamment la loi [République Française (2004)], réglemente la transmission des données sensibles recueillies en l'interdisant par défaut. Notons également que la directive européenne [The European Parliament and the Council (2002)] définit que, dans le cas où l'utilisateur accepte la diffusion de ses données sensibles à un tiers, cette diffusion doit être *minimale* et l'utilisateur doit en être prévenu. En France, le contrôle de l'application de ces lois s'effectue grâce à la Commission Nationale de l'Informatique et des Libertés[7] (CNIL) qui a pour rôle d'informer les personnes de leurs droits et de leurs obligations en termes de protection de la sphère privée, de sanctionner les comportements suspicieux par des sanctions pécuniaires par exemple et également de réglementer le respect de la vie privée en proposant des normes simplifiées. Pour finir, citons également la Fondation Internet Nouvelle Génération (FING) qui, dans le projet *identités actives* [Fondation Internet Nouvelle Génération (FING) (2008)], définit comme essentiel le fait de donner aux utilisateurs des droits sur la diffusion de leurs données sensibles afin que leur image virtuelle et réelle ne soit pas entachée.

[7]http ://www.cnil.fr

Le stockage des données sensibles contenues dans la sphère privée impose donc un niveau de sécurité élevé afin de prévenir les attaques possibles et d'en contrôler les accès comme le souligne par exemple [Agrawal *et al.* (2002)] avec le septième principe[8] des bases de données hippocratiques ou encore [Bergenti (2005)].

2.5.2 Transaction des données sensibles

La problématique du respect de la sphère privée relative à la transaction de données sensibles requiert de s'intéresser à deux axes de recherche. Le premier se rapporte à la *sécurité de la transaction* : lors de la transaction de données sensibles, les agents doivent utiliser un support de communication sécurisé, comme par exemple avec l'utilisation de clés de chiffrement asymétriques, afin que ces données ne soient pas interceptées [Sandhu *et al.* (1996); Agrawal *et al.* (2002); Bergenti (2005); Cissée et Albayrak (2007); Rezgui *et al.* (2002)]. Les besoins de la sphère privée se focalisent ici sur le domaine de la sécurité en termes de réseaux et de cryptographie.

Le deuxième problème soulevé par la transaction de données sensibles concerne le *consentement* de l'entité qui fournit de telles données. Afin de donner son accord pour une transaction, cette entité doit être en mesure de déterminer les impacts futurs d'une telle transaction en termes de collecte, diffusion, utilisation et rétention [W3C (2002b); Cranor (2002); Agrawal *et al.* (2002); Piolle (2009)].

2.5.3 Devenir des données sensibles

Les deux précédentes sous-sections décrivent les besoins premiers du respect de la sphère privée tandis que celle-ci décrit une phase souvent oubliée : le devenir des données sensibles après une transaction, ou plus simplement l'étude du comportement de l'entité qui recueille les données après une transaction de données sensibles.

La plate-forme pour les préférences de confidentialité et les bases de données hippocratiques prennent en considération cette phase par la description du comportement de l'entité qui recueille les données sensibles (les serveurs, les bases de données par exemple) vis-à-vis des différentes manipulations des données (utilisation, diffusion...) qu'elle aspire à exécuter.

Afin que l'entité (l'utilisateur ou le serveur) qui fournit les données sensibles puisse déterminer les impacts de la diffusion de celles-ci, l'entité qui les

[8]Un système de gestion de bases de données hippocratiques assure un niveau élevé de sécurité afin d'éviter toute collecte frauduleuse. La mise en place d'un tel niveau de sécurité intervient lors des accès aux bases et lors du stockage des données sensibles.

Travaux	Stockage	Transaction	Devenir
[W3C (2002b)]		Politique de manipulation	
Réseau ami-à-ami		Diffusion limitée	
[Damiani *et al.* (2004)]			Prévention des pourriels
[Pommier et Bourdon (2008)]	Sécurité		
[Ferraiolo et Kuhn (1992); Sandhu *et al.* (1996)]		Contrôle d'accès	
[Agrawal *et al.* (2002)]	Sécurité	Sécurité et politique de manipulation minimale	Vérification de la politique par l'utilisateur
[Massacci *et al.* (2007)]	Sécurité	Sécurité et politique de manipulation minimale	Vérification de la politique par l'utilisateur et sanction des comportements suspicieux
[Freuder *et al.* (2001); Yokoo *et al.* (2005); Silaghi et Rajeshirke (2004); Greenstadt *et al.* (2006)]		Diffusion limitée	
[Rezgui *et al.* (2002); Bergenti (2005); Cissée et Albayrak (2007); Piolle (2009)]	Sécurité	Sécurité et politique de manipulation	

TAB. 2.1 – Récapitulatif des travaux présentés pour le respect de la sphère privée.

reçoit doit fournir une description des manipulations qu'elle envisage de réaliser sur celles-ci. De plus, cette entité doit s'engager à ne pas exécuter d'autres manipulations hormis celles spécifiées en termes de collecte, utilisation, rétention et diffusion [W3C (2002b); Cranor (2002); Agrawal *et al.* (2002)]. Une fois terminée la transaction de données sensibles, les entités en jeu dans ce type d'interaction doivent être capables de vérifier que l'engagement de l'entité qui reçoit les données a bien été respecté [Agrawal *et al.* (2002)]. Pour finir, le système doit garantir son comportement en détectant les violations de cet engagement, en les sanctionnant et en prévenant les utilisateurs des comportements suspicieux.

Le problème qui se pose ici est la vérification des contraintes imposées à l'entité qui recueille les données sensibles : comment s'assurer qu'elle respecte son engagement vis-à-vis des données qu'elle recueille ?

2.5.4 Discussion

Comme le montre le tableau récapitulatif des approches que nous avons présentées dans ce chapitre (Tableau 2.1), nous pouvons noter que les principales recherches sur le respect de la sphère privée se focalisent sur la transaction et le stockage, sans prendre en compte la vérification de l'engagement que prend l'entité qui reçoit les données sensibles, ce qui représente le risque le plus important de nos jours en informatique.

La dernière phase critique, le devenir, est un des problèmes les moins pris en considération dans les travaux traitant de la préservation de la sphère privée. Nous proposons donc de consacrer le prochain chapitre à l'étude des mécanismes de régulation adaptables aux agents dans le but de mettre en œuvre cette dernière phase critique et de répondre ainsi à la dernière question que nous soulevons dans ce chapitre.

Chapitre 3

Régulation des systèmes multi-agents pour la sphère privée

Sommaire

Le devenir des données sensibles, phase critique du respect de la sphère privée, nous impose de considérer le comportement des agents vis-à-vis des données sensibles qu'ils recueillent après une transaction de données sensibles. Dans ce contexte, il est notamment essentiel que les agents soient capables de juger les autres agents avant de leur transmettre des données sensibles.

Cela permet ainsi de déterminer la fiabilité d'un agent vis-à-vis du respect de l'engagement qu'il a accepté sur les manipulations des données sensibles qu'il recueille. Du fait des caractéristiques de la sphère privée (cf. chapitre 2, section 2.1), cette fiabilité doit être contextuelle et personnelle tout en évoluant en fonction du comportement des agents.

Pour prendre en considération les besoins de cette phase critique, nous présentons dans ce chapitre les deux principales catégories de régulations (en termes de contrôle) qui peuvent être adaptées aux systèmes multi-agents : la régulation de comportement par une autorité tiers n'appartenant pas au système multi-agent et la régulation effectuée par les agents du système. Nous concluons ce chapitre par une mise en relation de ces mécanismes avec le respect de la sphère privée.

3.1 Régulation par des autorités tiers

3.1.1 Agent tiers et normes sociales

Afin de réguler le comportement des agents, un premier type de mécanismes possibles consiste à introduire une autorité tiers (entité ou agent) dans la société d'agents. Ce tiers a pour rôle de vérifier le comportement de chaque agent, de détecter les violations et d'appliquer des sanctions si elles existent. Ce type de régulation est généralement employé dans les systèmes de sécurité comme PONDER [Lupu *et al.* (2000)] par exemple. PONDER a pour but de développer et d'instaurer un langage de politiques de sécurité définissant les normes de bon comportement, à l'aide d'autorisations et d'obligations, et de ce fait, les critères représentant un comportement suspicieux.

Exemple 3.1

Dans le système PONDER, la déclaration d'une politique de sécurité telle que :
inst auth $-$ $/negativeAuth/testRouters\{$
subject $/testEngineers/trainee$;
action $performance\_test()$;
target $< routerT >$ $/routers$;
when $time.between$ ("0900","1700")}
spécifie qu'il n'est pas autorisé d'effectuer des tests de performance sur les routeurs durant l'intervalle de temps compris entre l'heure 0900 et l'heure 1700.

L'instauration de ces politiques est réalisée par une entité tiers, indépendante des agents, qui vérifie l'état des agents ainsi que l'ensemble de leurs

interactions dans le but de détecter toutes les violations des politiques de sécurité.

Cette technologie est adaptable aux systèmes multi-agents en y intégrant une entité de confiance sanctionnant les agents présentant un comportement suspicieux et une plate-forme de confiance garantissant le bon déroulement des interactions. La détection de l'infraction aux politiques peut se faire par exemple par écoute flottante [Busetta *et al.* (2002)] au niveau des interactions entre agents ou au travers de la plate-forme de confiance [Chen *et al.* (2009)] au niveau du raisonnement et de la mémoire des agents.

Une seconde utilisation des autorités tiers adaptables aux systèmes multiagents concerne les institutions électroniques et les systèmes normatifs employant des normes sociales de bon comportement. Notons que lors de nos recherches nous n'avons trouvé aucune référence dans ce domaine appliquée au respect de la sphère privée. En effet, les systèmes normatifs tels que [Hannoun *et al.* (2000)] ou AGREEN [Báez-Barranco *et al.* (2007)] par exemple, les normes décrivent de quelle manière les agents doivent se comporter afin de réaliser leurs tâches individuelles et leurs tâches communes sans pour autant prendre ne considération la sphère privée des agents.

Les travaux, comme [Artikis *et al.* (2002); Esteva *et al.* (2002, 2004); García-Camino *et al.* (2005); Fornara et Colombetti (2008)], reposent tous sur les mêmes fondements. Un bon comportement est défini en fonction d'un ensemble d'obligations, de permissions et d'interdictions d'exécuter certaines actions. Chaque agent de la société est contraint, au niveau de ses interactions, par une ou plusieurs autorités tiers de contrôle. Chaque interaction, incluant donc les transactions de données sensibles, transite ainsi par une autre entité, agent ou non, qui l'analyse et peut donc consulter les données sensibles échangées. Ce tiers est certifié en ce qui concerne le respect de la sphère privée, mais rien ne garantit son bon comportement une fois le certificat obtenu.

3.1.2 Discussion

Les mécanismes de régulation externe aux agents imposent donc que les transactions de données sensibles soient réalisées par l'intermédiaire d'une ou plusieurs autres entités. Ces autorités possèdent alors une vision globale de la société, au niveau de la sphère privée des agents, ce qui constituent une violation de la préservation de la sphère privée. En effet, afin de respecter cette sphère, comme nous l'avons remarqué dans le chapitre 2, les données sensibles des agents ne doivent pas être accessibles à un tiers sans l'autorisation de l'agent. Or la régulation externe de comportement requiert la connaissance de ce type de données par les autorités tiers qui peuvent conserver les données sensibles diffusées. De plus, ce type de mécanismes impose que les transactions de données sensibles s'effectuent par le même intermédiaire. Les communications

entre agents sont donc régies par un ou plusieurs tiers pouvant récupérer les données sensibles échangées ce qui représente aussi une violation de la sphère privée.

Pour ce faire, une possibilité serait de certifier cette entité tiers mais nous pouvons remarquer qu'une fois le certificat obtenu, aucune autorité ne vérifie si cette entité en respecte les clauses.

Pour conclure sur ce type de régulation, le respect de la sphère privée impose une vision locale de cette sphère aux agents afin de ne pas commettre de violations de cette dernière, du fait de son caractère personnel et de sa confidentialité, ce qui est incompatible avec la vision globale que les tiers possèdent.

Etudions maintenant le contrôle social, un mécanisme de régulation interne adaptable aux agents, afin d'en déterminer la compatibilité avec le respect de la sphère privée.

3.2 Régulation par contrôle social

Le second type de régulation auquel nous nous intéressons est fondé sur une gestion interne aux agents des comportements. Dans ce contexte, on suppose que la société d'agents se régule par elle-même grâce aux agents qui la composent. Dans le domaine des systèmes multi-agents, le contrôle social est une technique de régulation interne largement étudiée et employée dans les travaux relatifs à la confiance et à la réputation. Le contrôle social tel que l'aborde [Castelfranchi (2000)] par exemple correspond à une collaboration entre agents s'appuyant sur la confiance et la réputation au vu d'établir des relations de confiance.

Nous proposons dans cette section d'analyser ce mécanisme afin d'en déterminer les avantages et les incompatibilités pour le respect de la sphère privée. Nous présentons une synthèse des différents travaux portant sur la confiance puis sur la réputation pour détailler la construction et la gestion de la confiance interpersonnelle. Nous concluons en établissant les liens qui existent entre le respect de la sphère privée et la confiance.

3.2.1 Confiance

La confiance et la réputation n'ont pas de définition communément acceptée, que ce soit en informatique ou dans le domaine des sciences sociales, même si en général, avoir confiance en quelqu'un est décrit comme "*croire qu'il est honnête, sincère, et qu'il n'agit pas délibérément dans l'intention de nous nuire*" (dictionnaire Collins). La confiance est donc un sentiment subjectif qui

permet d'espérer en une personne pour une tâche donnée, ce qui apporte une assurance et un sentiment de sécurité.

La phase *devenir des données sensibles* impose une évaluation du comportement futur des autres agents tout en permettant de calculer les risques encourus en cas d'un comportement suspicieux. Nous nous intéressons à la confiance définie comme un **processus de décision** (faire confiance ou non) [Fukuyama (1995); McKnight et Chervany (2001); Quéré (2001); Demolombe (2004)]. Cette vision nous permet d'assimiler la confiance à une prise de risque pour une action donnée. En effet, lorsqu'un agent décide de faire confiance à un autre agent, il estime que les risques encourus sont négligeables, voire nuls, ou trop grands pour exécuter une interaction donnée.

Cette vision de la confiance permet de réduire l'ignorance ou l'incertitude de l'agent sur le comportement futur d'un autre agent [Luhmann (1988)]. Elle réduit les risques encourus à l'exécution d'une action donnée par un agent cible.

A partir de nombreux travaux sur la confiance, nous pouvons présenter une synthèse des caractéristiques admises de celle-ci. La confiance est **contextuelle** [Fukuyama (1995); McKnight et Chervany (2001); Quéré (2001); Demolombe (2004)] : faire confiance signifie émettre l'hypothèse qu'une action donnée soit bien réalisée et ce dans un cadre précis (des bénéfices sont alors apportés à l'agent). Elle est également définie comme **subjective** et **personnalisable** [Abdul-Rahman (2004); Lacomme *et al.* (2009)] : chaque agent possède son propre niveau de confiance envers un autre agent selon ses critères personnels. La confiance est **relative à différentes sources d'information** car elle ne peut s'établir sans l'étude de l'historique des interactions [Sabater (2002); Abdul-Rahman (2004)]. Elle est donc longue à établir et se nourrit (i) des expériences directes, (ii) des informations émanant d'un témoin et (iii) des aspects sociologiques[1]. La confiance est **dynamique** car elle est fondée sur les historiques des interactions des agents. De ce fait, son évolution dynamique est fortement liée à l'évolution des interactions [Sabater (2002); Abdul-Rahman (2004)]. Pour finir, la confiance est **réciproque** [Ostrom (1998); Conte et Paolucci (2002); Falcone *et al.* (2002); Mui *et al.* (2002)] ce qui incite les agents à bien se comporter entre eux [Quéré (2001)] et permet également d'instaurer un contrôle social.

La confiance est également décomposée selon quatre classes [Dasgupta (1990); McKnight et Chervany (2001); Quéré (2001)]. La **confiance dispositionnelle** représente les dispositions d'un agent à faire confiance plus ou moins facilement à un agent sans qu'un historique des interactions ait pu être établi. La confiance interpersonnelle fait référence aux relations de confiance unà-un et représente la confiance qu'accorde un agent à un autre agent donné. La

[1]Informations tirées de l'analyse du réseau social [Melaye *et al.* (2006)].

confiance de groupe représente la confiance qu'un agent accorde à un groupe particulier d'agents (agents ayant au moins une caractéristique commune). La **confiance institutionnelle** (ou systémique) fait référence à la confiance mise en jeu entre un agent et une institution (ou système).

3.2.2 Réputation

La réputation est considérée comme un mécanisme de contrôle social [Castelfranchi (2000)], du fait de la réciprocité de la relation de confiance [Abdul-Rahman (2004)].

La réputation est relative aux informations sur lesquelles un agent s'appuie pour prendre la décision d'instaurer une relation de confiance avec un autre agent. Les réputations sont en quelques sortes les entrées du processus de construction et de gestion de la confiance interpersonnelle [Muller (2006)] : un agent décide de faire confiance à un autre après avoir raisonné sur les différentes réputations qu'il a pu acquérir.

De ce fait, la réputation possède les mêmes caractéristiques que la confiance (contextualité, subjectivité, personnalisation et fondées sur les expériences passées). La réputation n'étant pas une simple modélisation de la confiance, elle possède deux autres caractéristiques afin de les différencier au niveau de leur conception. La première définit le caractère **communicable** de la réputation. La communication de réputation se fait par le concept de recommandations [Conte et Paolucci (2002)], déterminées par les réputations propagées [Casare et Sichman (2005)]. Une réputation étant subjective à chacun, un agent ne communique pas la réputation qu'il accorde à un autre mais il informe les autres agents sur la cible en calculant une recommandation à partir de ses propres réputations. La seconde caractéristique de la réputation concerne le caractère **multi-facette** de celle-ci [Zacharia *et al.* (1999); Rubiera *et al.* (2003); Muller (2006)] : un agent peut décider de faire confiance à un autre agent selon plusieurs critères, on a alors une réputation par critère.

Exemple 3.2

> *Une personne peut faire confiance à une autre sur le plan professionnel mais elle ne lui accorde aucune confiance pour ses confidences personnelles.*

Selon [Casare et Sichman (2005)], il existe quatre classes de réputations. La **réputation directe** résulte des expériences directes entre deux agents. La **réputation observée** est construite par les observations qu'un agent peut faire sur les interactions entre les autre agents. La **réputation propagée** représente la réputation fournie par les autres agent. La **réputation stéréotypée** est la réputation qu'un agent accorde à un autre selon ses caractéristiques sans prendre en considération les interactions passées.

Maintenant qu'une vue globale de la confiance et de la réputation a été présentée, nous allons étudier le processus visant à construire et gérer la confiance interpersonnelle.

3.2.3 Construction et gestion de la confiance interpersonnelle

Une relation de confiance interpersonnelle entre deux agents se construit sur l'historique de leurs interactions et sur leur confiance dispositionnelle. Cette relation évolue au cours du temps en fonction des interactions exécutées. A partir des diverses sources d'informations, un agent peut émettre un jugement de confiance sur un agent cible (agent qui est visé par la relation de confiance qu'un autre agent essaie d'établir envers lui). Une fois ce jugement émis, il peut alors le diffuser aux autres agents et surtout décider d'installer une relation de confiance ou pas pour une tâche donnée [McKnight et Chervany (2001); Demolombe (2004)].

Exemple 3.3

> *L'agent alice fait confiance à l'agent bob. Après plusieurs interactions entre bob et l'agent charlie, charlie décide de ne plus faire confiance à bob. Il indique alors cette décision à alice qui l'intègre dans son processus de confiance. Avec l'ajout de ces nouvelles informations, alice décide alors de ne plus faire confiance à bob.*

La confiance permet d'établir une relation de fiabilité entre deux agents. Cette relation se construit à partir d'un processus décomposé en quatre étapes [McKnight et Chervany (2001); Demolombe (2004)] résumées sur la figure 3.1 qui intègre les différentes réputations mises en jeu :

1. **Initialisation** : Dans le cas où aucune relation de confiance n'a pu être établie entre l'agent et l'agent cible, l'agent évalue l'agent cible par rapport à sa confiance dispositionnelle (déterminée à partir des réputations stéréotypées) si aucune interaction n'a pu avoir lieu. A partir des différentes sources d'information (réputations directes et propagées) que l'agent a pu recueillir, il raisonne sur l'agent cible afin de pouvoir émettre un jugement de confiance (l'agent cible est-il un agent dans lequel on peut avoir confiance ou non ?).
2. **Révision** : La confiance étant dynamique, le jugement émis doit être révisé avec la mise à jour des différentes sources d'informations (réputations directes et propagées).
3. **Propagation** : Les sources d'informations pouvant émaner d'un témoin (compilées sous forme de recommandations, obtenues à partir des réputations propagées reçues), une fois qu'un agent a établi un jugement de

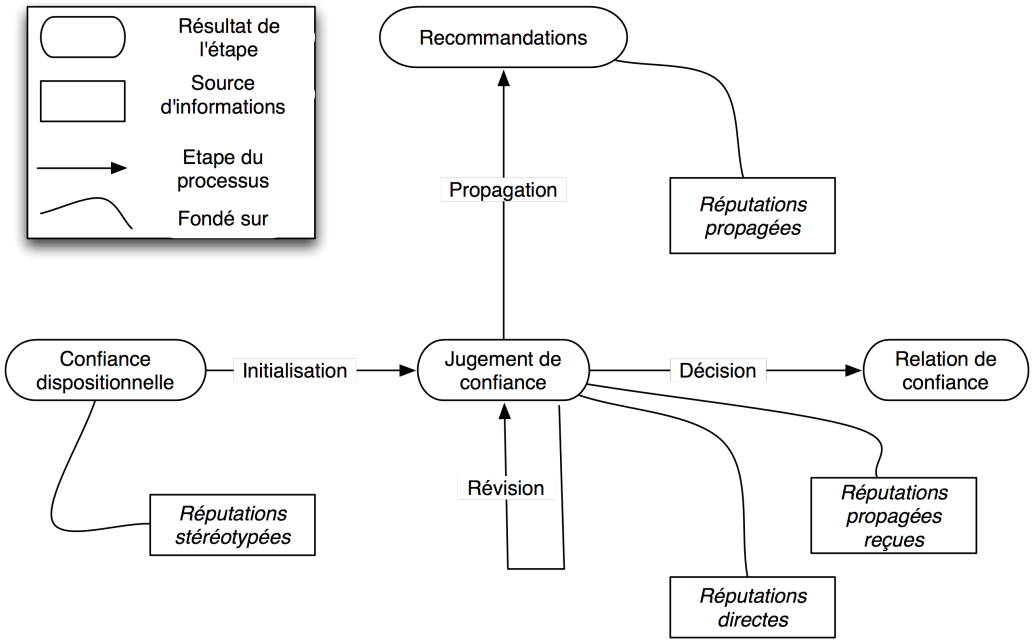

Fig. 3.1 – Décomposition du processus de construction et de gestion de la confiance.

confiance, il est susceptible de prévenir les autres agents de son jugement ainsi que la révision de ce dernier.

4. **Décision** : Une fois que l'agent a émis son jugement de confiance, il décide alors d'instaurer ou non une relation de confiance entre lui et l'agent cible.

3.2.4 Discussion

Pour instaurer un contrôle social visant à réguler les agents sur le respect de la sphère privée, nous nous intéressons à la confiance interpersonnelle étendue à la confiance dispositionnelle pour l'intégration des préférences utilisateurs du fait du caractère personnel de la sphère privée.

La réputation est relative aux croyances sur lesquelles une entité s'appuie pour établir une relation de confiance avec une autre entité. Le contrôle social que nous proposons porte sur des contraintes de relations un-à-un avec une vision locale aux agents, ce qui nous conduit à focaliser nos travaux sur les réputations directes, propagées et stéréotypées.

3.3 Régulation de comportement et respect de la sphère privée

3.3.1 Exigences et critères

Afin d'instaurer un contrôle social, nous devons choisir un modèle opérationnel de confiance respectant nos besoins en termes de préservation de la sphère privée. Pour cela, nous nous focalisons sur les modèles qui correspondent à la confiance et à la réputation présentées précédemment.

La gestion de la réputation doit comprendre plusieurs facettes. Le modèle employé doit également permettre la gestion des réputations directes afin que les agents puissent émettre un jugement sur leur fiabilité les uns envers les autres selon l'historique de leurs interactions et de leurs observations. De plus, afin de modéliser les préférences des utilisateurs, le modèle doit intégrer les réputations stéréotypées pour permettre aux utilisateurs de prendre part à l'initialisation du processus de confiance tout en respectant leurs souhaits. Pour finir, instaurer un contrôle social impose que les réputations puissent être propagées afin qu'une certaine prévention des comportements suspicieux soit mise en place. Notons que nous en préoccupons des réputations observées car, les interactions ayant un rapport avec la sphère privée ne devant pas être interceptées, les agents n'ont donc pas la possibilité d'observer ces interactions.

	Init.	Révision	Raisonnement	Décision	Propagation
[Castelfranchi et Falcone (1998)]	X	X	X	X	X
[Sabater (2002)]	X	X	X	X	X
[Abdul-Rahman (2004)]	X	X*		X	X
[Muller (2006)]	X	X	X	X	X

TAB. 3.1 – Etapes du processus de construction et de gestion d'une relation de confiance (** avec l'intervention de l'utilisateur*).

Travaillant sur des systèmes multi-agents centrés utilisateur, nous devons également sélectionner un modèle de confiance en adéquation avec la personnalisation de l'agent utilisateur. Dans notre contexte de recherche, un utilisateur délègue à un agent artificiel (ou logiciel) sa sphère privée et donc la protection

de ses données sensibles. Cette délégation sous-entend également une délégation complète des étapes du processus de construction et de gestion de la confiance, résumées dans le tableau 3.1.

Nous focalisons donc notre choix sur trois modèles de confiance multiagents : [Castelfranchi et Falcone (1998)], ReGReT [Sabater (2002)] et LIAR [Muller (2006)] qui répondent à ces exigences, [Abdul-Rahman (2004)] ne proposant pas une entière délégation du processus de construction et de gestion de la confiance.

3.3.2 Modèles existants

Afin de finaliser notre choix sur le modèle de confiance approprié, il nous faut différencier les trois modèles [Castelfranchi et Falcone (1998); Sabater (2002); Muller (2006)]. Commençons par présenter plus en détail chacun de ces modèles.

3.3.2.1 Travaux de Castelfranchi

[Castelfranchi et Falcone (1998)] aborde la confiance comme un processus cognitif dans le cadre des systèmes multi-agents. Ils se basent sur la théorie définissant la confiance comme nécessaire à la délégation de tâches.

La confiance est ici abordée comme un ensemble d'attitudes mentales représentant plusieurs types de croyances : sur les compétences, sur la nécessité de la tâche et sur les dispositions de l'agent à exécuter cette tâche. Ce modèle intègre donc la confiance interpersonnelle et dispositionnelle.

Ces travaux décrivent l'instauration de relations de confiance entre des agents de type BDI [Bratman (1987)]. Les agents sont décrits comme ayant trois états possibles :

- Belief (noté $B_i\alpha$) : l'agent i croit α même si α peut être erronée. Les croyances peuvent porter sur l'état du monde et sur l'état des autres agents.
- Desire (noté $D_i\alpha$) : l'agent i veut réaliser α (il souhaite que l'état du monde devienne α).
- Intention (noté $I_i\alpha$) : l'agent i va effectuer l'action α afin de réaliser ses désirs.

Les états mentaux des agents sont fondés sur trois sources d'informations pour établir une relation de confiance selon un contexte donné (ou une action donnée) Ω :

- Les expériences directes entre deux agents (relation un-à-un) correspondant aux réputations directes ;
- Les recommandations relatives aux réputations propagées ;

– La confiance systémique avec les réputations stéréotypées.

A partir de ces différentes sources d'informations, les agents peuvent déterminer deux types de croyances intermédiaires afin d'obtenir un niveau de confiance *DoT* (*Degree of Trust*) :

– Le niveau de compétence *DoA* (*Degree of Ability*) ;

– Le niveau de bonne volonté *DoW* (*Degree of Willingness*).

Le niveau de confiance $DoT_{j,\Omega}$ est obtenu par une fonction monotone exploitant les croyances $DoA_{j,\Omega}$ et $DoW_{j,\Omega}$. Afin d'établir une relation de confiance avec l'agent j, l'agent i compare le résultat de la fonction précédente selon un seuil où la valeur 0 représente un manque absolu de confiance et la valeur 1 une confiance absolue dans l'agent i selon le contexte Ω.

Remarquons que ce modèle a été étendu de nombreuses fois, comme par exemple, dans [Herzig *et al.* (2008)] qui ajoute dans le raisonnement des agents des croyances collectives aux croyances individuelles ou encore [Lacomme *et al.* (2009) pour la personnalisation de réseaux de confiance multi-agents. Notre thèse ne traitant pas des réseaux d'agents et ne prenant pas en compte d'un point de vue formel les croyances collectives afin de préserver la sphère privée, nous ne nous préoccupons pas de ces travaux.

3.3.2.2 Modèle ReGReT

ReGReT [Sabater (2002)] aborde la confiance selon trois dimensions pour un contexte de commerce électronique :

1. Dimension individuelle : cette dimension se fonde sur les interactions directes entre les agents. Elle est calculée à partir des opinions directes construites sur les évaluations des expériences directes entre deux agents et elle dépend du temps (les évaluations les plus anciennes ont moins de poids que les plus récentes).
2. Dimension sociale : cette dimension se construit sur les sources d'informations extérieures telles que les recommandations de témoignage (réputation individuelle communiquée par un agent tiers), de voisinage (moyenne pondérée des réputations individuelles de l'agent cible) ou systémiques (dépend des caractéristiques sociales communes comme l'appartenance à un groupe donné par exemple).
3. Dimension ontologique : cette dimension permet de prendre en compte le contexte dans le calcul des réputations.

La relation de confiance s'instaure par le biais d'un processus de décision où une valeur de fiabilité est attribuée à chacune de ses dimensions. Ce processus consiste à utiliser la confiance individuelle si elle est jugée comme fiable. Dans le cas contraire, la relation de faire confiance est établie à partir d'une combinaison linéaires des autres dimensions.

Ce modèle permet de fusionner les différentes réputations, avec une pondération variable, en une réputation globale des agents grâce à l'introduction d'une ontologie de domaine. Les processus liés à la construction et à la gestion de la confiance sont automatisés.

3.3.2.3 Modèle LIAR

Muller *et al.* définissent la confiance à l'aide de réputations qui sont considérées comme des croyances dans le contexte des réseaux pair-à-pair [Muller et Vercouter (2004); Muller (2006)].

La réputation dans ce modèle a pour valeur une probabilité subjective pouvant être égale à unknow lorsqu'un agent ne dispose pas d'assez d'informations sur l'agent cible pour émettre un jugement de confiance. Le processus automatisé de construction et de gestion de la confiance prend en considération dans un premier temps les réputations calculées à partir des interactions directes. Si elles ne sont pas suffisantes pour prendre la décision d'instaurer une relation de confiance, les réputations propagées, décomposées en réputation de témoignage et réputation de voisinage, ainsi que la réputation relative à la confiance institutionnelle sont alors à la base du jugement de confiance.

LIAR introduit également une nouvelle caractéristique à la réputation : le fait d'être multi-dimension ce qui permet de classer les réputations en fonction de leur provenance. Ce modèle ne permet pas de combiner les différentes réputations calculées afin d'avoir un jugement général de l'agent cible, ce qui peut entraîner un blocage dans certaines situations, comme par exemple lorsqu'un agent n'a pas encore interagit avec les autres agents et qui ne possède donc pas de réputations le conernant.

3.3.3 Discussion

Pour le respect de la sphère privée, nous optons pour un modèle générique, adaptable à tout contexte de recherche de système multi-agent : le modèle de Castelfranchi et Falcone [Castelfranchi et Falcone (1998)], ReGreT étant un modèle conçu pour le commerce électronique et L.I.A.R. pour les réseaux pairàpair. De plus, le modèle proposé par Castelfranchi et Falcone est un modèle calculatoire fondé sur les sciences cognitives ce qui correspond à une vision centrée utilisateur [Lacomme *et al.* (2009)] qui est celle que nous souhaitons avoir dans cette thèse.

3.4 Synthèse

Du fait de la personnalisation de la sphère privée, une vision locale à l'agent est nécessaire pour son respect, les agents, le système et/ou une entité tiers ne peuvent pas disposer d'une vision globale des engagements des agents sur les transactions de données sensibles. Regrouper les mécanismes de régulation en une seule et même entité représente en effet un danger pour la sphère privée car cela impose une vision globale des engagements passés entre les agents sur les données de leur sphère privée. Nous devons ainsi opter pour des mécanismes de régulation internes tels que le contrôle social [Castelfranchi (2000)] fondé sur la confiance et la réputation où les agents régulent eux-même leur comportement en fonction de celui des autres agents.

Nous proposons de considérer un modèle de confiance existant pour en faire un modèle de confiance préservant la sphère privée. Il s'agit plus d'une extension compatible aux modèles déjà existants comme le propose par exemple [Rehák et Pechoucek (2007)] pour la représentation du contexte des relations de confiance.

La confiance et la réputation apportent donc plusieurs avantages pour le respect de la sphère privée en répondant à ses différents besoins non liés à la sécurité. Elles permettent premièrement d'implémenter la fonction de jugement nécessaire aux agents et ce dans sa globalité. Un système de prévention des comportements suspicieux peut également être instauré grâce à la présence des recommandations. Ces dernières permettent également d'intégrer plus aisément l'utilisateur dans le processus de construction et de gestion de la confiance. Pour finir, le processus relatif aux relations de confiance permet de garantir grâce au contrôle social le comportement futur des consommateurs sans introduire une régulation externe aux agents, ce qui constituerait une violation de la sphère privée.

Cependant, pour assurer un respect de la sphère privée le plus complet possible, nous devons également considérer que les informations relatives à la confiance et à la réputation peuvent être elles-mêmes jugées sensibles. Nous devons ainsi intégrer ce type de données dans la sphère privée des agents et leur appliquer l'ensemble des besoins des trois phases critiques présentées dans le deuxième chapitre, au sein du processus de construction et de gestion de la confiance afin de ne pas instaurer un contrôle social suspicieux.

La prochaine partie de cette thèse présente notre proposition de modèle multi-agent préservant la sphère privée, les Systèmes Multi-Agents Hippocratiques (HiMAS), prenant en compte les trois phases critiques du respect de la sphère privée. Le chapitre 4 est consacré aux fondements de notre modèle. Les chapitre 5 et 6 présentent quant à eux respectivement la modélisation des deux phases critiques (*transaction* et *devenir des données sensibles*) en décri-

vant un protocole de transaction de données sensibles et un contrôle social hippocratique.

Chapitre 4

Fondements du modèle HiMAS

Sommaire

Notre première contribution, dans le cadre de cette thèse, consiste à définir la problématique attachée à la sphère privée et à présenter un modèle multi-agent, que nous appelons Systèmes Multi-Agents Hippocratiques, ou HiMAS

[0]Les travaux proposés dans ce chapitre ont été présentés dans [Crépin *et al.* (2007, 2008a,b, à paraître 2009)].

pour *Hippocratic Multi-Agent Systems*. Ce modèle permet aux agents d'un tel système de représenter le concept de sphère privée et d'assurer sa gestion et sa protection, en prenant en compte les trois phases critiques (stockage, transaction et devenir) avec l'ensemble des exigences qui en découlent. Le terme *hippocratique* implique que ce modèle traite des aspects moraux et éthiques du respect de la sphère privée en imposant aux agents du système multi-agent de se plier à un ensemble de neuf principes normatifs, tel le médecin envers son patient lorsqu'il cherche à respecter le serment d'Hippocrate.

L'existence des trois phases critiques dans le cadre du respect de la sphère privée nous amène à constater que les bases de données hippocratiques, présentées dans la section 2.3.2, prennent en compte une grande partie de nos besoins dans le cadre des systèmes multi-agents, hormis l'instauration de mécanismes de sanction pour les comportements suspicieux. De ce fait, nous proposons d'étendre les travaux d'Agrawal *et al.* au domaine des systèmes multi-agents en considérant les agents comme autonomes et ayant pour but d'assister les utilisateurs dans la gestion et la protection de leur sphère privée. Pour ce faire, nous proposons d'intégrer neuf des dix principes hippocratiques définis dans [Agrawal *et al.* (2002)] au sein des systèmes multi-agents en y introduisant des mécanismes de sanction pour les comportements suspicieux.

Afin d'illustrer le modèle de systèmes multi-agents hippocratiques, nous considérons tout au long de cette thèse une application concrète de gestion d'agendas distribués nommée AGENDA [Demazeau *et al.* (2006)] comme contexte applicatif. La migration de cette application en un système multi-agent hippocratique sera présentée dans le chapitre 8. Au sein d'AGENDA, un rendez-vous est caractérisé par un identifiant, les participants, une date de début, une date de fin, un niveau d'*importance* et un niveau d'*urgence*. Ces deux derniers paramètres sont subjectifs à chaque utilisateur et permettent de donner un niveau de priorité pour l'agencement du calendrier. Chaque utilisateur est assisté par un agent ayant en charge son emploi du temps. Cet agenda peut être partagé avec les autres agents. Au niveau des interactions possibles entre les agents d'AGENDA pour fixer un rendez-vous entre plusieurs utilisateurs, notre proposition se focalise sur le partage des agendas : les agents disposent d'un accès direct aux données de l'agenda d'un autre agent sur simple requête. Cette tâche implique de nombreuses communications de données sensibles relatives aux calendriers des utilisateurs entre deux agents ce qui nous permet d'illustrer notre proposition.

Pour présenter le modèle de systèmes multi-agents hippocratiques que nous proposons, nous définissons d'abord notre modèle de la sphère privée puis nous développons les neuf grands principes normatifs des HiMAS. Pour finir, nous détaillons le centrage utilisateur des HiMAS en développant les mécanismes de délégation d'informations sensibles entre les utilisateurs et les agents. Les deux prochains chapitres présentent la formalisation des principes

relatifs aux deux phases critiques de la transaction et du devenir des données sensibles avec l'introduction d'un protocole de transaction de données sensibles et un contrôle social hippocratique.

4.1 Sphère privée

Comme nous l'avons vu dans différentes études présentées dans le deuxième chapitre, nous considérons que la sphère privée d'un agent concerne toutes les données qu'il désire protéger des autres agents, associées à leurs éléments de gestion. Seul l'agent concerné par les données sensibles détient les droits de propriété sur ces données. Nous considérons la sphère privée comme personnelle, personnalisable et contextuelle.

Par convention, nous notons les ensembles avec une majuscule et les éléments appartenant à ces ensembles par une minuscule.

Les agents d'un système multi-agent hippocratique construisent, font évoluer et protègent la représentation de leur sphère privée selon les souhaits des utilisateurs. En accord avec la définition 2.2. de la sphère privée, nous proposons la définition suivante.

Définition 4.1 (*Sphère privée*)

La sphère privée d'un agent, notée SP est un quadruplet de quatre ensembles non vides :

$$SP :=< Eléments,\ Autorisations,\ Règles,\ Normes >$$

La suite de cette section est consacrée à la présentation de chacun de ces ensembles.

4.1.1 Eléments de la sphère privée

Afin d'intégrer les données sensibles des utilisateurs dans la sphère privée des agents, nous proposons de les représenter grâce à des éléments. Ces éléments attachent aux données sensibles les renseignements nécessaires au respect de leur protection.

Définition 4.2 (*élément de la sphère privée*)

Soit $\mathcal{A}$ l'ensemble des agents du système multi-agent, Id l'ensemble des identifiants possibles pour la sphère privée d'un agent, $Données$ l'ensemble des données sensibles d'une sphère privée, $Contextes$ l'ensemble des contextes des données sensibles, relatifs au domaine du HiMAS. Nous

définissons un élément de la sphère privée, élément, comme un sextuplet :

$$élément \in Eléments$$

$$élément :=< id, donnée, Propriétaires,$$

$$contexte, Sujets, Références > avec$$

- $id \in Id$: *identifiant de l'élément,*
- $donnée \in Données$: *donnée sensible de la sphère privée,*
- $Propriétaires \subset \mathcal{A}$: *ensemble des identifiants des agents connus possédant la donnée sensible à un instant donné,*
- $contexte \in Contextes$: *contexte des éléments de la donnée sensible propres au domaine des données,*
- $Sujets \subset \mathcal{A}$: *ensemble des agents concernés par la donnée sensible,*
- $Références \subset Données$: *ensemble des références vers les données sensibles qui peuvent être déduites à partir de la connaissance de donnée.*

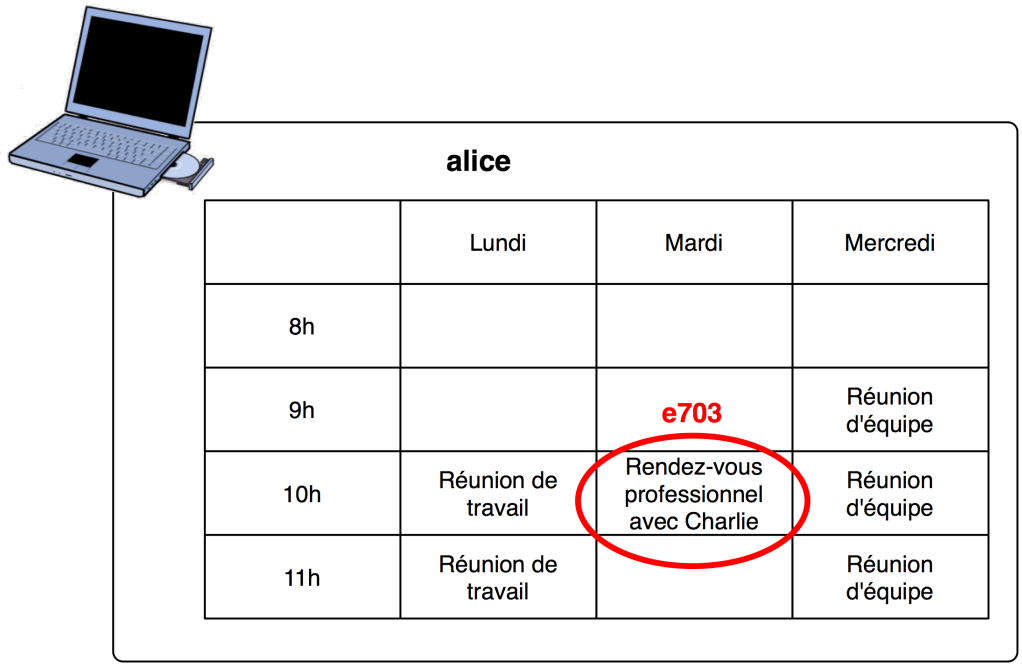

FIG. 4.1 – Agenda de l'agent *alice*.

Exemple 4.1

Dans le cas de la gestion d'agendas, soit e703 l'identifiant de l'élément concernant la donnée sensible rdv représentant le rendez-vous du Mardi à 10h dans l'agenda de l'agent alice, figure 4.1.

$$< e703,\ rdv,\ \{alice,\ charlie\},\ professionnel,$$

$$\{alice,\ charlie\}, \{mardi - 10h,\ \{alice,\ charlie\}, important, urgent\} >$$

Ce rendez-vous a lieu à une date précise, $mardi - 10h$. Les agents alice et charlie sont les participants et sont les seuls à connaître cette donnée sensible.

Afin qu'un agent puisse raisonner sur la diffusion des données sensibles, un élément est associé à un ensemble de propriétaires. Dans le cas de l'élément $e703$, les propriétaires de la donnée sensible rdv sont les agents alice et charlie.

Un élément appartient également à un certain contexte : par exemple, un rendez-vous de travail est associé au milieu professionnel, donc le contexte de $e703$ est $professionnel$.

Un élément de la sphère privée contient également l'ensemble des sujets de la donnée sensible, c'est-à-dire les agents concernés par celles-ci. Pour $e703$, les sujets sont les agents participant à la réunion, soit alice et charlie.

Une donnée sensible peut faire référence à d'autres données sensibles, ici rdv fait référence à la date $mardi - 10h$, au niveau d'importance et d'urgence de la réunion et aux participants $\{alice,\ charlie\}$.

Afin de mieux illustrer les éléments de la sphère privée, prenons un autre exemple relatif au domaine médical.

Exemple 4.2

Lorsqu'un médecin $médecin47$ effectue des analyses de sang $analyses465$ pour un patient $patient993$, ce médecin intègre les résultats de ces analyses dans sa sphère privée afin de respecter le serment d'Hippocrate. Soit $e465993$ l'élément de la sphère privée du médecin qui représente ces analyses.

$$< e465993,\ analyses465,\ \{médecin47, patient993\},\ médical,$$

$$patient993, \{taux_1 ... taux_n\} >$$

L'élément $e465993$ représente la donnée sensible relative aux analyses de sang $analyses465$ du patient $patient993$.

Les propriétaires de cette donnée sensible sont le médecin qui les a prescrites et le patient concerné, soit $médecin47$ et $patient993$. Notons que le seul sujet de cette donnée sensible est le patient pour qui les analyses ont été prescrites.

Ces analyses sont relatives au contexte médical et sont composées par les ensembles des taux sanguins demandés, $\{taux_1 ... taux_n\}$.

4.1.2 Autorisations d'un élément de la sphère privée

Afin que les agents manipulent les données en respectant la sphère privée, nous attachons un ensemble d'autorisations aux éléments. Les autorisations d'un élément de la sphère privée permettent à un agent de définir les opérations qu'il autorise sur une donnée sensible. Ces opérations portent sur les différentes manipulations des données sensibles qui pourraient porter atteinte à leur protection, comme leur diffusion, leur altération ou leur utilisation. Selon le domaine applicatif du HiMAS considéré, l'ensemble de ces autorisations peut être modifié par l'ajout ou la suppression d'une ou de plusieurs manipulations possibles.

Définition 4.3 (*Autorisation d'un élément de la sphère privée*)

Nous définissons une autorisation par :

$$autorisation \in Autorisations = \{utiliser, supprimer, diffuser, changer, mentirsur\}$$

Exemple 4.3

Soit e703 l'élément défini précédemment, nous pouvons définir les autorisations suivantes :

- *utiliser*(*e*703) *: La donnée sensible contenue dans l'élément e*703 *peut être utilisée par l'agent dont la sphère privée contient e*703*.*
- *supprimer*(*e*703) *: L'agent dont la sphère privée contient l'élément e*703 *a le droit de supprimer cet élément de sa sphère privée.*
- *diffuser*(*e*703) *: L'agent possédant l'élément e*703 *dans sa sphère privée a le droit de diffuser la donnée sensible de e*703*.*
- *changer*(*e*703) *: L'agent qui possède l'élément e*703 *dans sa sphère privée peut modifier cet élément, afin de mettre à jour la donnée sensible qu'il contient par exemple.*
- *mentirsur*(*e*703) *: L'agent possédant l'élément e*703 *dans sa sphère privée a le droit de mentir sur la donnée sensible contenue dans cet élément pour en assurer sa protection.*

4.1.3 Règles de la sphère privée d'un agent

Du fait que la sphère privée est définie de manière contextuelle, qu'elle évolue dynamiquement au cours du temps et qu'elle est intrinsèquement personnelle, nous lui associons un ensemble de règles permettant de spécifier les conditions d'activation des autorisations présentées précédemment.

Définition 4.4 (*Règle de la sphère privée*)

Soit Dates l'ensemble des date et Conditions l'ensemble des condition du HiMAS. Une règle de la sphère privée est définie de la façon suivante :

$$règle \in Règles$$

$$règle = autorisation(id) \leftarrow condition(contexte, date_{début}, date_{fin}) où$$

- $autorisation \in Autorisations$ *: l'autorisation possible sur l'élément d'identifiant* id*,*
- $id \in Id$ *: l'identifiant de l'élément,*
- $condition \in Conditions$ *: la condition activant l'autorisation* $autorisation$ *sur l'élément* id*,*
- $contexte \in Contextes$ *: contexte relatif à la condition* $condition$*,*
- $\{date_{début}, date_{fin}\} \in Dates$ *: la durée de validité de la règle.*

Les conditions de type *condition* sont relatives au contexte courant de l'agent et à celui de l'élément de la sphère privée. De plus, lorsqu'un agent envoie des données sensibles de sa sphère privée à un autre agent, cette transaction a pour but de satisfaire un objectif donné. De ce fait, les conditions sont également relatives aux objectifs de la transaction, qui sont développés dans le chapitre 5.

Exemple 4.4

Soit $e703$ *l'élément défini précédemment,* $professionnel$ *le contexte relatif aux rendez-vous de travail,* $10-10-2009, 31-12-2009$ *deux dates et* $fixerRdv(professionnel)$ *la condition représentant l'objectif d'une transaction de données sensibles pour le fait de fixer un rendez-vous de travail.*

$$utiliser(e703) \leftarrow$$

$$fixerRdv(professionnel, 10-10-2009, 31-12-2009)$$

est une règle permettant à l'agent possédant dans sa sphère privée l'élément $e703$ *d'utiliser la donnée sensible de* $e703$ *afin de prendre un autre rendez-vous seulement dans le contexte professionnel du début à la fin de la session courante.*

Les règles d'une sphère privée permettent à un agent de définir la dynamique interne de celle-ci selon ses propres souhaits et donc de prendre en compte le caractère personnel de la sphère privée.

L'ensemble des règles est dynamique, son contenu peut varier en fonction des différents événements qui se produisent dans le système multi-agent. Par exemple, une règle peut permettre à un agent de diffuser un rendez-vous professionnel dans un premier temps, et, suite à l'ajout d'un nouveau participant

à cette réunion, imposer que cette donnée ne puisse plus être diffusée afin de respecter les souhaits de ce dernier participant. En plus des changements dus à l'évolution temporelle, les règles de la sphère privée conduisent également à une évolution personnelle, selon les souhaits de l'agent propriétaire de celle-ci.

Cette dynamique peut être entraînée par deux types de mécanismes. Le premier, comme le montre l'exemple précédent, représente une évolution automatique des règles par l'agent suite à un nouvel événement dans la société. Le second représente une évolution dite manuelle où l'utilisateur modifie une ou plusieurs règles en fonction de ses besoin.

4.1.4 Normes de la sphère privée

Le dernier point abordé au niveau de la représentation d'une sphère privée concerne l'impact que peut avoir la société d'agents dans laquelle se trouve l'agent.

Nous intégrons les normes à la sphère privée des agents afin que, lorsqu'un agent appartient à différentes sociétés et qu'il est donc soumis à différents ensembles de normes, le partage de connaissances entre les sociétés sur les normes qu'elles imposent se soit pas possible.

Nous définissons les normes relatives à la sphère privée de manière similaire aux règles associées à celle-ci, à la façon de [Boella *et al.* (2007)]. Nous considérons que les normes sont des règles de bon comportement communes à l'ensemble des agents du système multi-agent.

Définition 4.5 (*Normes de la sphère privée*)

Soit Normes l'ensemble des normes communes à l'ensemble des agents. Nous nous inspirons de la logique déontique [von Wright (1951)] pour définir l'opérateur O représentant une obligation (notre thèse ne se focalisant sur cet aspect, nous ne développerons pas en détails cet opérateur). Une norme de la sphère privée est définie comme suit :

$$norme \in Normes$$

$norme = O(autorisation(id)) \leftarrow condition(contexte, date_{début}, date_{fin})$ *où*

- $autorisation \in Autorisations$ *: l'autorisation possible sur l'élément d'identifiant* id*,*
- $id \in Id$ *: l'identifiant de l'élément,*
- $condition \in Conditions$ *: la condition activant l'autorisation* $autorisation$ *sur l'élément* id*,*
- $\{date_{début}, date_{fin}\} \in Dates$ *: la durée de validité de la norme.*

Toutefois, contrairement aux règles, les normes sont connues de tous les agents de la société qui les impose et elles doivent être respectées par l'ensemble des agents. Demeulenaere remarque dans [Demeulenaere (2002)] que le périmètre de la sphère privée peut être influencé par les conventions de la société, même si l'agent reste entièrement maître de son comportement vis-à-vis de ces règles. Cela se traduit par le fait que certaines normes de la société peuvent imposer aux agents d'inclure ou non des éléments dans leur sphère privée, tout en laissant la possibilité à l'agent de violer ces normes. Les conséquences de ces violations sont à étudier afin d'en déterminer les différents impacts sur les agents ainsi que sur la société d'agents, de même que l'impact des contradictions possibles entre les normes et les règles de la sphère privée d'un agent. Cette étude dépasse toutefois les objectifs de cette thèse et nous ne l'évoquerons pas dans la suite de ce manuscrit. Le lecteur intéressé par cet axe de recherche peut se référer à [Piolle et Demazeau (2008)].

De plus, notons, comme le constate Demeulenaere dans [Demeulenaere (2002)], que ces normes peuvent évoluer au cours de l'exécution du système multi-agent : certains comportements des agents peuvent conduire à rendre une norme caduque ou encore à en instaurer une nouvelle. Par exemple, une norme interdisant de mentir sur un rendez-vous personnel dans un contexte professionnel, devient caduque si l'ensemble des agents ne la respecte pas. Pour de plus amples informations sur cet axe de recherche, nous renvoyons le lecteur à la thèse de Guillaume Piolle étudiant le respect de la sphère privée à travers le raisonnement normatif des agents [Piolle (2009)].

4.1.5 Relations internes de la sphère privée

Un agent personnalise sa sphère privée en définissant l'ensemble des éléments qui la constituent et à partir de l'ensemble des données qu'il estime être sensibles (cf. figure 4.2). Il définit également l'ensemble des règles qui se rapportent aux autorisations sur les éléments de sa sphère privée selon ses souhaits, comme le montre la figure 4.2. Cette personnalisation permet de prendre en considération l'ensemble des souhaits de l'utilisateur sur la gestion de ses données sensibles.

Au niveau du raisonnement de l'agent, les autorisations représentent les manipulations possibles des éléments de la sphère privée par les agents et sont activées selon les conditions des règles (cf. figure 4.2). Les normes peuvent imposer aux agents de nouvelles autorisations de la sphère privée devant être respectées et les règles peuvent induire de nouvelles normes de la société comme le montre la figure 4.2. En effet, lorsque l'ensemble des agents respecte les mêmes règles, la société d'agents peut alors décider de définir de nouvelles normes correspondant à ces règles. Les différentes influences entre les normes et les règles sont un aspect important de la gestion de la sphère privée qui

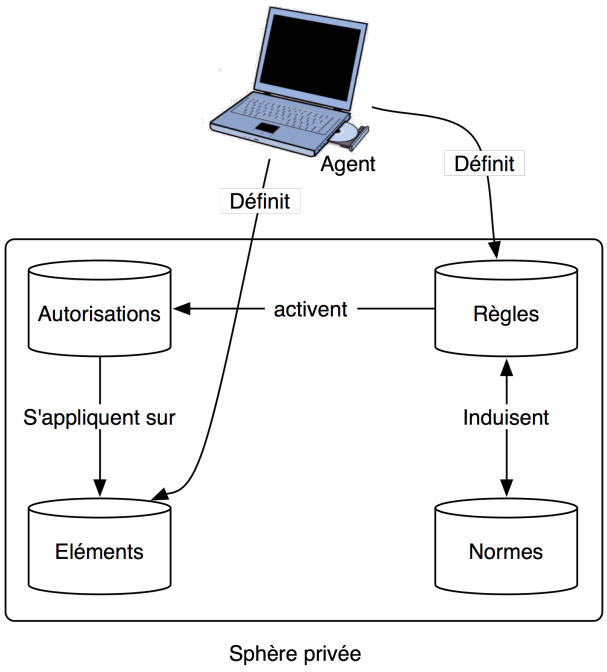

FIG. 4.2 – Sphère privée d'un agent.

représente une perspective intéressante de nos travaux de recherche mais que nous n'évoquerons également pas dans la suite de cette thèse.

Concluons cette section par un exemple illustrant la sphère privée d'un agent d'un HiMAS.

Exemple 4.5

Représentons la sphère privée de l'agent alice concernant l'agenda de l'utilisateur qu'il représente sur une période de trois matinées (représentée dans la figure 4.3). Nous considérons que les trois rendez-vous (lundi 10h-12h, mardi 10h-11h et mercredi 9h-10h) sont des données sensibles de la sphère privée de cet agent.

Ainsi l'ensemble Eléments de la sphère privée d'alice est constitué des trois éléments suivants :

$$< e702,\ rdv,\ \{alice,\ bob\},\ professionnel,$$
$$\{alice,\ bob\}, \{lundi - 10h,\ \{alice,\ bob\}, important, urgent\} >$$
$$< e703,\ rdv,\ \{alice,\ charlie\},\ professionnel,$$
$$\{alice,\ charlie\}, \{mardi - 10h,\ \{alice,\ charlie\}, important, urgent\} >$$
$$< e704,\ rdv,\ \{alice,\ équipe\},\ professionnel,$$
$$\{alice,\ équip\}, \{mercredi - 9h,\ \{alice,\ charlie\}, important\} >$$

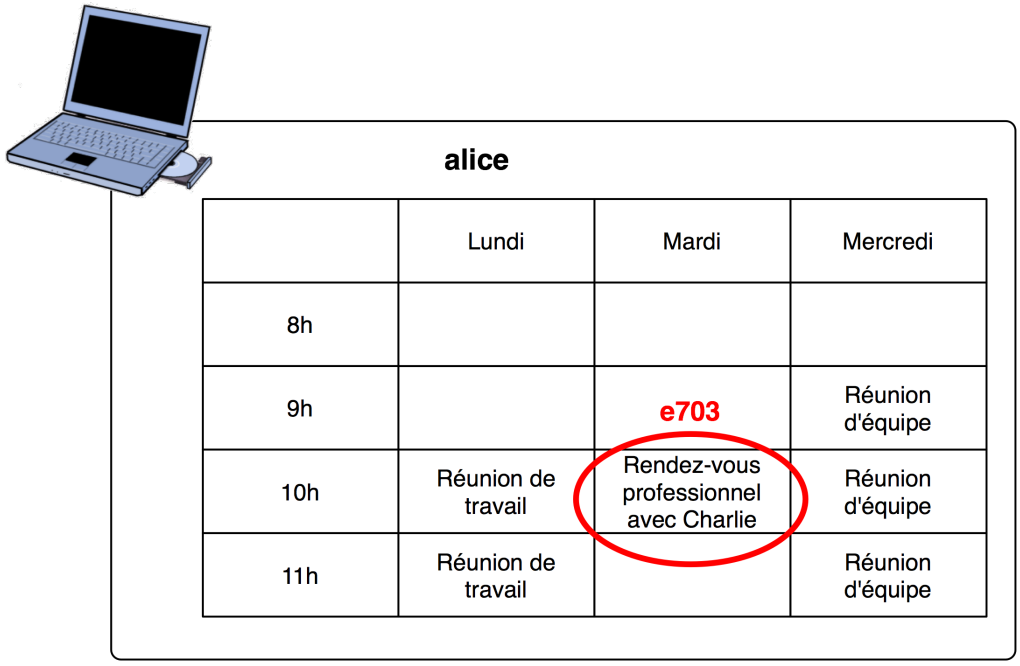

FIG. 4.3 – Agenda de l'agent *alice*.

L'ensemble des autorisations possibles, Autorisations, est constitué des autorisations décrites dans la sous-section 4.1.2 :

$$Autorisations = \{utiliser, supprimer, diffuser, changer, mentirsur\}$$

Pour illustrer l'ensemble des règles, Règles, de la sphère privée d'alice, nous instaurons une règle pour chaque donnée sensible :

- *La donnée sensible de e*702 *ne doit pas être diffusée afin de prendre un autre rendez-vous quel que que soit le domaine du début à la fin de la session courante.*

$$\neg diffuser(e704) \leftarrow fixerRdv(null, débutSession, FinSession)$$

- *La donnée sensible de e*703 *peut être utilisée afin de prendre un autre rendez-vous seulement dans le contexte professionnel du début à la fin de la session courante.*

$$utiliser(e703) \leftarrow fixerRdv(professionnel, débutSession, FinSession)$$

- *La donnée sensible de l'élément e*704 *peut être diffusée afin de prendre un autre rendez-vous quelque que soit le domaine du début à la fin de la session courante.*

$$diffuser(e704) \leftarrow fixerRdv(null, débutSession, FinSession)$$

L'ensemble des normes de la sphère privée, Normes, étant commun à la société d'agents, nous ne le développons pas dans cet exemple.

La sphère privée étant définie, nous pouvons maintenant présenter les neuf principes normatifs qui permettent aux agents d'assurer son respect.

4.2 Principes hippocratiques

Le modèle de systèmes multi-agents hippocratiques, HiMAS, est inspiré de celui proposé par Agrawal *et al.* dans le cadre des bases de données hippocratiques [Agrawal *et al.* (2002)].

Dès lors qu'un agent est capable de représenter sa sphère privée, il doit prendre en considération sa préservation. Pour ce faire, nous proposons d'instaurer des principes normatifs pour les agents d'un HiMAS. Cependant l'ensemble de ces principes ne suffit pas totalement à protéger la sphère privée. En effet, si un agent vient à divulguer une donnée sensible à un autre agent, il doit être en mesure de juger de la fiabilité de ce dernier agent afin de déterminer les risques qu'il encoure et ainsi permettre la mise en place d'un mécanisme de sanction envers les agents suspicieux (ce mécanisme sera détaillé dans le chapitre 6). Nous imposons donc aux agents d'un HiMAS d'émettre un jugement sur la fiabilité des autres agents de la société en plus du respect des neufs principes normatifs.

Avant de préciser ces principes normatifs, nous présentons, à travers la notion de rôle, les différentes positions que les agents d'un HiMAS peuvent prendre par rapport à la sphère privée avec la définition de plusieurs rôles.

4.2.1 Rôles des agents

Pour représenter les positionnements possibles d'un agent par rapport à la sphère privée, nous définissons trois rôles (cf. figure 4.4). Notons que cette vision centrée utilisateur est à l'opposé de celle centrée service en termes de consommateur et de fournisseur.

Définition 4.6 (*Consommateur*)

Le rôle de consommateur caractérise l'agent qui demande une donnée sensible.

Définition 4.7 (*Fournisseur*)

Le rôle de fournisseur caractérise l'agent qui reçoit une demande de donnée sensible.

Définition 4.8 (*Sujet*)

Le rôle de sujet caractérise l'agent concerné par une donnée sensible. Lors d'une transaction de données, le fournisseur et le sujet ne sont pas forcé-

ment le même agent comme, par exemple, lorsque le fournisseur transmet une donnée sensible d'un tiers.

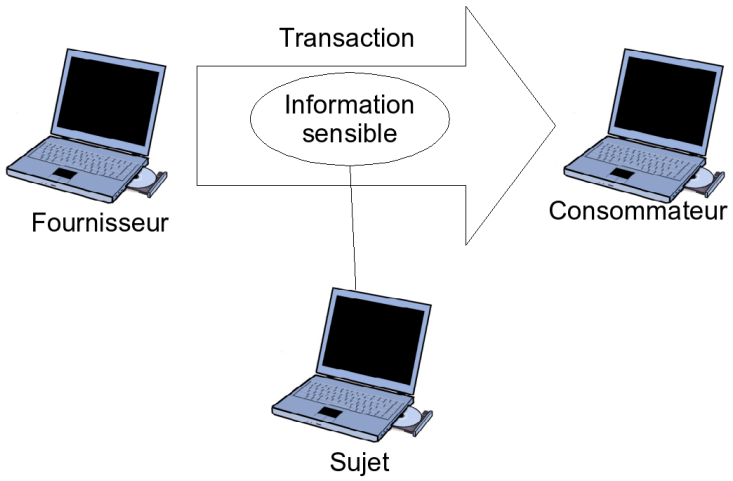

FIG. 4.4 – Rôles des agents d'un HiMAS.

4.2.2 Principes normatifs

Selon le modèle que nous proposons, un système multi-agent hippocratique doit respecter les neuf principes normatifs suivants. Afin de prendre en compte les besoins des trois phases du respect de la sphère privée (stockage, transaction et devenir), nous proposons de les représenter sur la figure 4.5 (adaptation de la figure 2.2. du chapitre 2).

1. Consentement (*Phase transaction de données sensibles*) : Chaque transaction de données sensibles doit requérir le consentement du fournisseur.

Exemple 4.6

Lorsqu'un consommateur demande à un fournisseur son planning à une date précise, le fournisseur doit donner son accord. Dans le cas où le fournisseur et le sujet ne sont pas le même agent, ce consentement est également demandé au sujet et doit respecter ses souhaits en termes de protection. Par exemple, si le consommateur demande à un fournisseur le planning d'un troisième agent (représentant le sujet), le fournisseur transmettra cette donnée seulement à condition que le sujet et lui-même soient consentants.

2. Connaissance des objectifs (*Phase transaction de données sensibles*) : Le fournisseur doit connaître les objectifs motivant la collecte des données

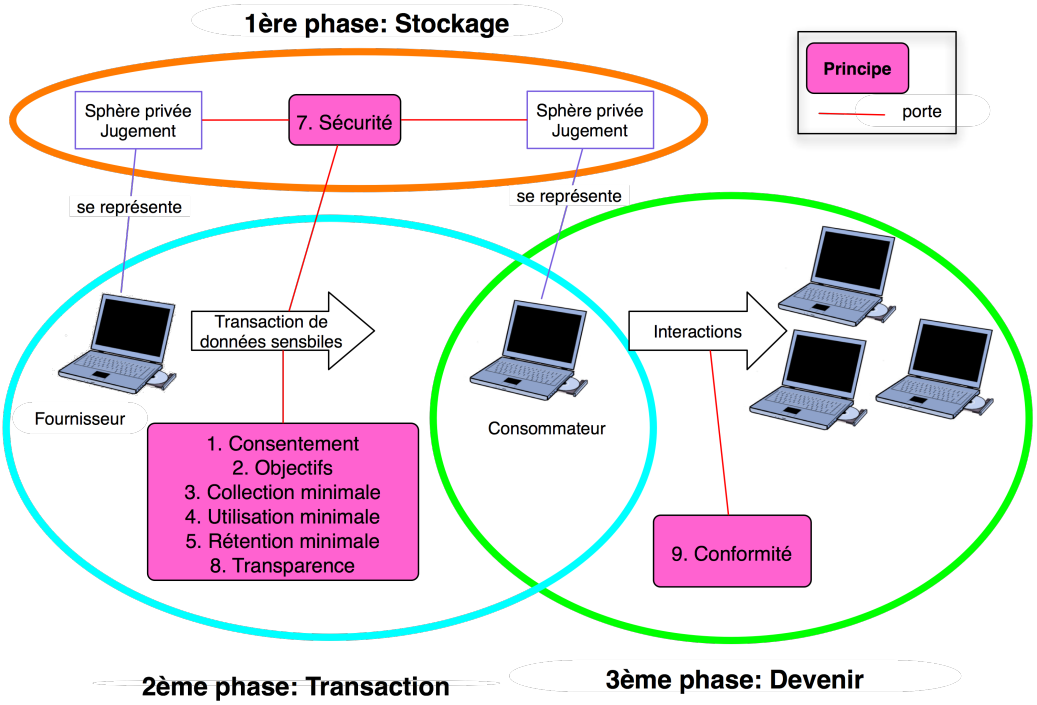

FIG. 4.5 – Systèmes Multi-Agents Hippocratiques (HiMAS).

sensibles. Ainsi, s'il le souhaite, il a la possibilité de calculer les conséquences de cet échange.

Exemple 4.7

Le consommateur indique que s'il désire avoir le planning du fournisseur, c'est dans l'objectif de prendre un rendez-vous avec lui. Le fournisseur peut ainsi décider dans de meilleures conditions s'il transmet ou non la donnée sensible.

3. Collecte minimale (*Phase transaction de données sensibles*) : Tout consommateur doit s'engager à collecter une quantité suffisante de données pour la réalisation d'un même ensemble d'objectifs.

Exemple 4.8

Lorsque le consommateur demande son planning au fournisseur pour fixer un nouveau rendez-vous, le consommateur a uniquement besoin de connaître les plages horaires libres et occupées du fournisseur. Il ne doit pas chercher à obtenir plus de données comme par exemple le sujet des rendez-vous ou les participants.

4. Utilisation minimale (*Phase transaction de données sensibles*) : Tout consommateur doit s'engager à n'utiliser une donnée sensible demandée à un fournisseur que pour satisfaire les objectifs qu'il a spécifié et pour rien de plus.

Exemple 4.9

Dans le contexte de la gestion d'agendas, le consommateur doit utiliser par exemple le planning demandé uniquement pour fixer un nouveau rendez-vous entre lui et le fournisseur. Le consommateur ne peut pas demander à utiliser cette donnée sensible pour la communiquer à un tiers car cela ne fait pas partie de ses objectifs.

5. Diffusion minimale (*Phase transaction de données sensibles*) : Tout consommateur doit s'engager à ne diffuser une donnée sensible que si ses objectifs l'exigent et ce, le moins de fois possible et aux moins d'agents possibles. L'ampleur de cette limitation dépend des objectifs de la transaction.

Exemple 4.10

Dans le cas de la prise de rendez-vous, le consommateur n'a nullement besoin de diffuser le planning du fournisseur pour prendre un rendez-vous uniquement avec le fournisseur.

6. Rétention minimale (*Phase transaction de données sensibles*) : Tout consommateur doit s'engager à ne conserver une donnée sensible que pendant un certain laps de temps. Celui-ci doit être fixé au plus petit délai requis pour la réalisation des objectifs du consommateur.

Exemple 4.11

Pour la prise de rendez-vous, le consommateur s'engage à effacer la partie du planning reçue du fournisseur une fois le rendez-vous pris ou, le cas échéant, la date du rendez-vous dépassée.

7. Sécurité (*Phases stockage et transaction de données sensibles*) : La sécurité des données sensibles doit être garantie pendant leur stockage et les transactions.

8. Transparence (*Phase transaction de données sensibles*) : Les données sensibles transmises doivent rester accessibles au sujet et/ou fournisseur pendant la durée de rétention. La transparence permet au fournisseur de mettre à jour les données sensibles qu'il a pu transmettre.

Exemple 4.12

Par exemple, si le planning du fournisseur change et que le nouveau rendez-vous n'a pas encore été pris, il doit avoir la possibilité de mettre

à jour le planning connu par le consommateur, et ce afin que la prise de rendez-vous se fonde sur des données exactes.

9. Conformité (*Phase devenir des données sensibles*) : Tout agent appartenant à un système multi-agent hippocratique doit être capable de vérifier le respect des principes précédents et de prévenir et de dénoncer tout comportement suspicieux.

Notons qu'un principe des bases de données hippocratiques n'a pas été retenu pour les HiMAS : celui qui impose l'exactitude des données sensibles. En effet, dans le contexte des systèmes multi-agents, un agent doit avoir la possibilité de mentir pour protéger sa sphère privée. Par exemple, le fait de refuser de donner une donnée sensible à un agent suspicieux peut souvent révéler cette donnée ou du moins une partie. Nous pouvons illustrer cette remarque avec un exemple de question fermée. Par exemple, lorsqu'une personne demande à une autre "Pouvez-vous me dire si vous votez Parti Socialiste ?", l'absence de réponse peut être interprétée comme une réponse affirmative.

Lorsqu'un fournisseur juge un consommateur comme suspicieux, deux possibilités s'offrent à lui. La première consiste à ignorer la requête du consommateur. La deuxième consiste à mentir sur la donnée sensible demandée. L'utilisation du mensonge permet de ne pas dévoiler au consommateur qu'il est jugé comme suspicieux (cf. définition 2.3). Cette solution permet de ce fait de discréditer le consommateur suspicieux auprès des autres agents lorsqu'il dévoilera une donnée sensible erronée (mais cela risque aussi de discréditer le fournisseur qui aura menti).

4.3 Centrage utilisateur

Dans les systèmes multi-agents centrés utilisateur, les agents autonomes sont en charge de certaines tâches attribuées par l'utilisateur afin de lui porter assistance. Dans les HiMAS, ces tâches concernent le respect de la sphère privée en termes de gestion et de protection des données sensibles déléguées.

Avant d'étudier la modélisation des principes à un niveau social dans les deux prochains chapitres, nous nous intéressons d'abord dans cette section au niveau individuel (ou agent) : de quelle manière des agents autonomes peuvent-ils construire leur sphère privée en fonction d'un utilisateur tout en respectant les souhaits de celui-ci sur les données sensibles qu'il leur délègue (nous faisons référence ici au caractère personnel et personnalisable de la sphère privée) ?

Cette délégation nécessite de représenter les différents liens entre un utilisateur et son agent assistant tout en respectant les caractéristiques de la sphère privée, notamment sa personnalisation et son aspect personnel.

La personnalisation étant un vaste domaine de recherche [Montaner *et al.* (2003); Gauch *et al.* (2007)], nous proposons d'adapter les techniques employées pour la création et la maintenance des profils utilisateur dans l'esprit des travaux de [Maes (1994)] qui définissent des canevas internes aux agents pour représenter les données et leurs paramètres de manière rationnelle afin d'en assurer une gestion particulière.

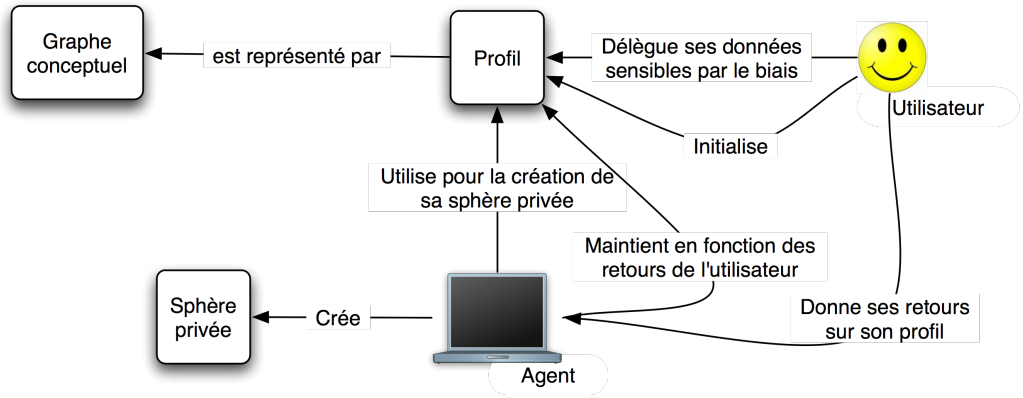

FIG. 4.6 – Mise en place du centrage utilisateur dans les HiMAS.

Afin d'intégrer l'utilisateur dans les HiMAS, nous introduisons dans le raisonnement des agents d'un HiMAS les quatre étapes nécessaires à la création d'un profil utilisateur [Montaner *et al.* (2003)] représentées dans la figure 4.6 en les adaptant au respect de la sphère privée :

- *La représentation du profil de l'utilisateur* : comment créer les différents liens entre un agent et un utilisateur dans le but de la délégation des données sensibles selon les souhaits exprimés par l'utilisateur sur leur respect (compréhension mutuelle) ?
- *L'initialisation du profil* : de quelle manière un agent peut-il initialiser un profil en limitant l'intervention de l'utilisateur tout en respectant la sphère privée de celui-ci ?
- *Les retours de l'utilisateur* : comment assurer la maintenance de la sphère privée des agents, dont la collecte des données sensibles, en fonction des souhaits de l'utilisateur ?
- *L'adaptation du profil* : de quelle manière assurer les mises à jour du profil en fonction de l'utilisateur et des adaptations requises par l'application ?

4.3.1 Représentation des profils utilisateurs

Un profil utilisateur est généralement représenté selon trois grandes catégories [Gauch *et al.* (2007)] : (i) par mots-clés [Moukas *et al.* (1997); Sakagami

et Kamba (1997)], (ii) par réseaux sémantiques [Gentili *et al.* (2003); Micarelli et Sciarrone (2004)] et (iii) par concepts [Guarino *et al.* (1999); Pretschner et Gauch (1999)]. Les deux premières approches sont fondées sur l'utilisation de mots-clés (en réseau pour la deuxième) pondérés selon les intérêts de l'utilisateur. Elles ne permettent pas une abstraction relative aux données sensibles déléguées à l'agent ce qui empêche un classement générique des données, technique facilitant pourtant la délégation par l'utilisateur (coût moins élevé en temps utilisateur) et la gestion des données par l'agent.

Cependant ce dernier point est spécifique au profil utilisateur représenté par des concepts. En effet, selon cette technique, les données sont caractérisées selon une méta-donnée pouvant posséder plusieurs paramètres, ordonnées selon une certaine hiérarchie.

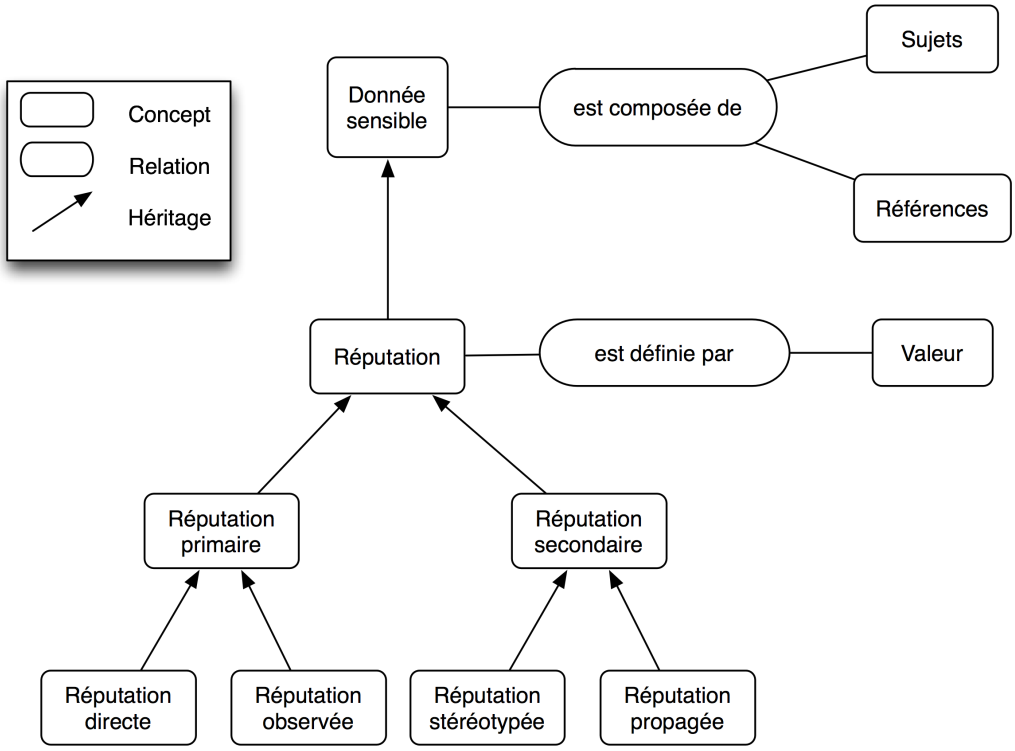

FIG. 4.7 – Graphe conceptuel représentant les types, indépendants du domaine, des données sensibles relatives aux utilisateurs.

Nous proposons de fonder la délégation des données sensibles d'un utilisateur à un agent d'un HiMAS par l'instanciation d'un graphe conceptuel [Sowa (1984)] représentant les méta-données [Pitrat (1990)] relatives aux données sensibles du domaine ainsi qu'aux données utilisées dans le processus de construction et de gestion de la confiance, ou plus précisément les réputations.

L'ensemble des concepts définis dans ce graphe est relatif à *Données* de la sphère privée des agents.

La représentation du profil utilisateur se fait au travers de deux graphes conceptuels représentant les données sensibles non relatives au domaine et celles relatives au domaine du HiMAS.

Dans le graphe conceptuel proposé dans la figure 4.7, chaque concept représente un type de donnée sensible indépendamment du domaine, ainsi que les liens en termes de hiérarchie et de sémantique qui peuvent exister entre les concepts. Chaque donnée sensible est composée d'un ensemble de sujets et de références afin que les agents puissent les intégrer dans leurs éléments de la sphère privée (cf. définition 3.2)[1]. Afin de faciliter la création des règles de gestion de la sphère privée, le contexte de la donnée sensible est directement intégré au type de donnée lors de la création du profil relatif au domaine. Comme le montre la figure 4.7, les données sensibles non relatives au domaine des HiMAS représentent les données utilisées dans la construction et la gestion de la confiance : les réputations [Casare et Sichman (2005)] qui permettent aux agents d'un HiMAS d'appliquer le principe de conformité avec un contrôle social basé sur la confiance et la réputation (cf. chapitre 3).

Pour les données sensibles relatives au domaine du HiMAS, nous proposons à nouveau un exemple relatif à l'application AGENDA dont l'implémentation est présentée dans le chapitre 8. Cet exemple se focalise sur les données sensibles relatives au domaine et n'aborde donc pas les réputations, cet aspect étant développé dans le chapitre 6.

Exemple 4.13

Dans le domaine de la gestion d'agenda, les données sensibles déléguées par l'utilisateur à un agent sont ses rendez-vous. Dans AGENDA*, les rendez-vous dépendent d'un certain contexte qui peut être personnel (associatif et familial) ou professionnel (équipe, laboratoire et projet européen). Les données sensibles sont représentées dans la figure 4.8 qui intègre la hiérarchie entre ces types de rendez-vous.*

Ayant défini les types de données sensibles, définissons maintenant les concepts relatifs aux règles de gestion de la sphère privée (cf. définition 3.4.) afin de représenter parfaitement l'utilisateur dans notre approche. Ces règles définissent les conditions d'activation des autorisations (utiliser, supprimer, diffuser, changer et mentir) sur les données sensibles que l'agent possède. Ce graphe conceptuel permet aux agents d'un HiMAS de définir les ensembles *Autorisations* et *Règles* de leur sphère privée. Ces concepts sont également rattachés aux types de données sensibles afin que les agents d'un HiMAS

[1]Seules ces informations doivent être fournies par l'utilisateur pour la création d'un élément de la sphère privée d'un agent.

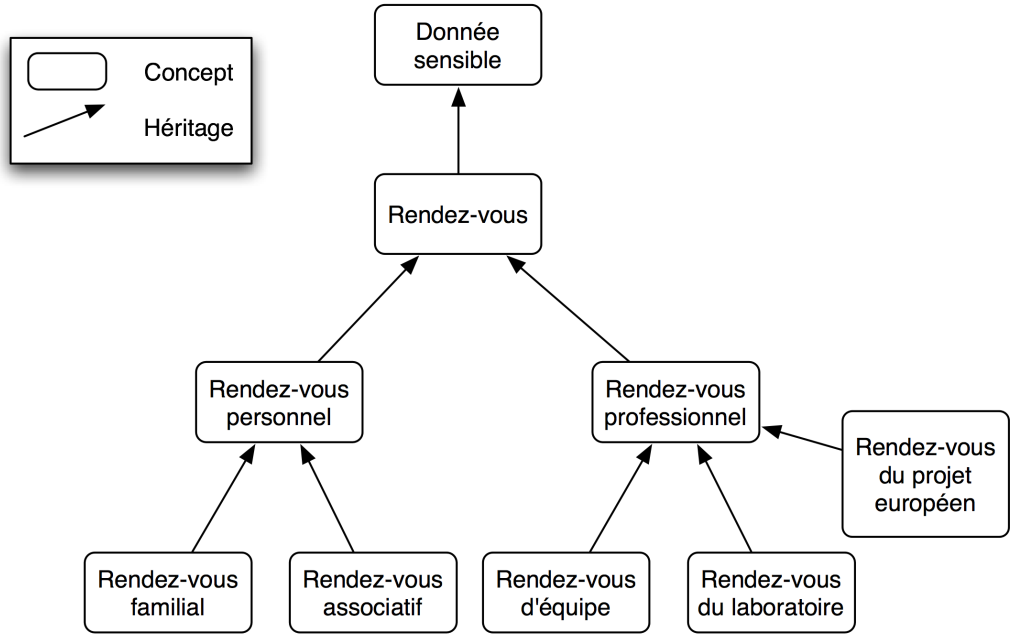

FIG. 4.8 – Graphe conceptuel représentant les types dépendants du domaine des données sensibles relatives aux utilisateurs dans le domaine de la gestion d'agenda.

puissent les appliquer selon les souhaits de l'utilisateur comme le montre la figure 4.9.

4.3.2 Initialisation des profils

Une fois que les agents d'un HiMAS possèdent la capacité de communiquer de manière compréhensible avec l'utilisateur par le biais des graphes conceptuels présentés précédemment (représentation des profils), ils vont devoir créer à partir de ces graphes et sans l'intervention de l'utilisateur (de manière automatique) leur sphère privée tout en respectant les souhaits des utilisateurs qu'ils représentent.

Un agent d'un HiMAS crée sa sphère privée grâce à l'initialisation du profil de l'utilisateur et selon les concepts représentant les types de données sensibles et les règles de gestion qui y sont attachées. Cette initialisation est possible selon deux approches principales :

- une initialisation manuelle où l'utilisateur instancie les données de son profil à partir d'un profil vide [Balabanovic et Shoham (1997); Chen et Sycara (1998)],

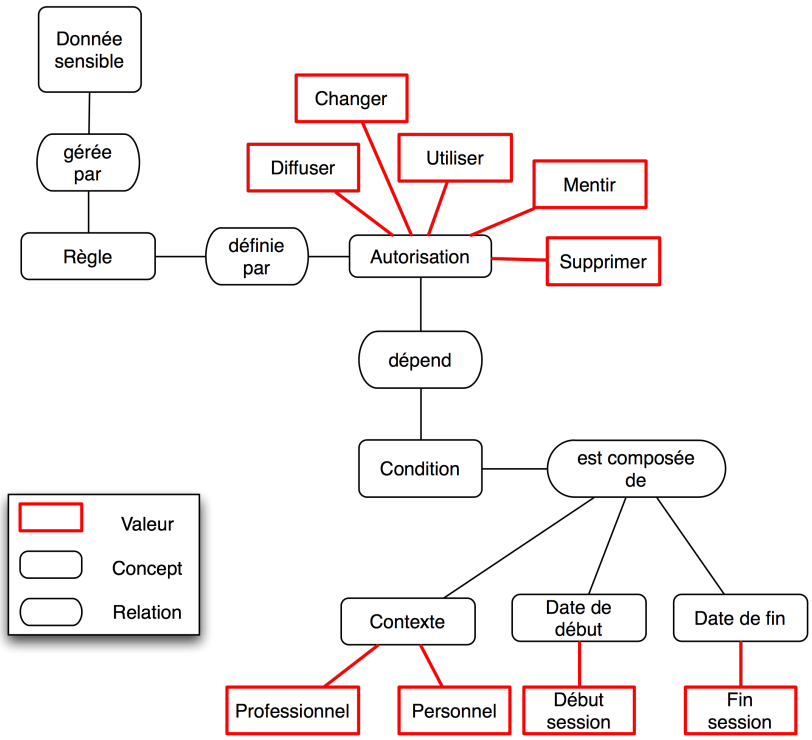

FIG. 4.9 – Graphe conceptuel représentant les règles de gestion de la sphère privée pour l'intégration de l'utilisateur.

– une initialisation stéréotypée prenant en compte les caractéristiques de l'utilisateur [Rich (1998)].

4.3.2.1 Initialisation manuelle

Dans le cas d'une initialisation manuelle, l'utilisateur choisit les types de données sensibles relatifs au domaine que l'agent doit intégrer dans sa sphère privée dans un profil lui étant proposé vide.

Exemple 4.14

Dans AGENDA*, un utilisateur peut décider que seuls les rendez-vous familiaux sont sensibles et laisser les autres types de rendez-vous publics.*

Une fois les types de données sensibles sélectionnés, l'agent propose à l'utilisateur, toujours à l'aide des concepts définis, de créer des règles de gestion pour les types choisis. Pour ce faire, il propose à l'utilisateur toutes les autorisations possibles en fonction du domaine ainsi que l'ensemble maximal des conditions qui peuvent exister dans ce domaine.

Exemple 4.15

Dans notre exemple, l'agent propose à l'utilisateur d'associer aux rendez-vous familiaux les autorisations portant sur la diffusion et le partage de ces rendez-vous selon que le consommateur fasse parti de sa famille ou non.

4.3.2.2 Initialisation stéréotypée

Une initialisation stéréotypée entraîne la création de la sphère privée de l'agent sans l'intervention de l'utilisateur. Afin de protéger tout type de données sensibles, l'agent déclare que tous les concepts du profil sont à intégrer dans sa sphère privée.

Exemple 4.16

Dans AGENDA*, les agents déclarent que tous les types de données représentées dans le graphe de la figure 4.8 (rendez-vous personnel et professionnel) sont des données sensibles de leur sphère privée.*

Au niveau des règles de gestion, l'instanciation du graphe conceptuel comprend les ensembles maximaux des autorisations et des conditions attachées permettant de définir un comportement non suspicieux en fonction du domaine. L'agent attache donc à chaque type de données sensibles l'ensemble des règles décrites par l'association des autorisations et des conditions.

Exemple 4.17

Dans notre exemple, les agents associent à chaque type de rendez-vous toutes les règles décrites dans l'instanciation du graphe représentant les règles de la sphère privée (cf. figure 4.9).

Afin que l'utilisateur puisse personnaliser la sphère privée de son agent, un raffinement personnel du stéréotype est possible en appliquant les mécanismes liés à une initialisation manuelle.

Exemple 4.18

Dans AGENDA*, une fois que les agents ont créé leur sphère privée avec une initialisation stéréotypée, l'utilisateur peut décider par exemple que ses rendez-vous professionnels ne sont pas des données sensibles.*

4.3.2.3 Discussion

L'initialisation manuelle semble être la plus adaptée pour le respect de la sphère privée car elle permet de prendre en compte le caractère personnel de la sphère privée. En effet, un tel mécanisme permet à l'utilisateur de personnaliser les données sensibles de l'agent mais également les règles de gestion des éléments de la sphère privée des agents, ce qui est en parfaite concordance avec notre définition de cette sphère et de son respect.

Cependant, nous pouvons remarquer que ce type d'initialisation est coûteux en temps pour l'utilisateur, ce qui peut être pallié par l'utilisation de profils stéréotypés. L'initialisation par un profil vide constitue une violation de la sphère privée car toutes les données sensibles que l'agent possède deviennent publiques et leur protection n'est donc pas assurée. De ce fait, nous ne pouvons pas opter pour une telle technique d'initialisation.

L'initialisation selon des stéréotypes, tout en diminuant le coût du temps utilisateur, ne constitue pas une violation de la sphère privée car elle définit un ensemble maximal des données sensibles ainsi qu'un ensemble maximal de leurs règles de gestion. Pour ce faire, les stéréotypes employés (et définis selon le domaine) doivent comprendre les ensembles maximaux de données pouvant être estimées comme sensibles ainsi que l'ensemble des règles définissant un comportement non suspicieux selon le domaine[2].

4.3.3 Retours des utilisateurs

Les retours de l'utilisateur sur le profil que l'agent lui attribue permettent de collecter les données sensibles de l'utilisateur. Ils peuvent être de trois types : (i) implicite, (ii) explicite ou (iii) hybride (i.e. une combinaison des deux types précédents) [Gauch *et al.* (2007)]. Comme le remarque [Quiroga et Mostafa (1999)], l'approche implicite et celle explicite apportent le même niveau de satisfaction de la part de l'utilisateur qui en majorité préfère un mécanisme

[2]Le développement de ces ensembles est abordé plus en détails dans le prochain chapitre.

hybride par souci de gain de temps et de respect de ses souhaits en termes de personnalisation.

L'approche implicite ne demande pas l'intervention de l'utilisateur car l'agent infère les données lui étant relatives par les traces laissées par le comportement de l'utilisateur comme par exemple dans [Kelly et Teevan (2003)]. Cette étude de l'activité de l'utilisateur et de son historique est généralement utilisée pour la personnalisation des navigateurs [Lieberman (1995); Marais et Bharat (1997)].

Les retours explicites nécessitent une bonne volonté de la part de l'utilisateur car ils requièrent un coût important au niveau de son temps. Ce type de retours s'exprime généralement par des cases à cocher, par des valeurs à choisir sur une échelle.

Les approches hybrides emploient des retours implicites complétés, voire corrigés, par des retours explicites des utilisateurs. Ce type d'approches possède les mêmes avantages et les mêmes inconvénients que les deux précédentes.

Le respect de la sphère privée impose que celle-ci reste entièrement personnelle en ce qui concerne les données sensibles qu'elle contient. Du fait de cette caractéristique, nous devons opter pour des retours explicites de l'utilisateur afin que les agents puissent gérer leurs données sensibles en respectant au mieux les souhaits de l'utilisateur qui leur délègue ses informations. En effet, un agent ne peut pas choisir délibérément qu'une donnée soit sensible ou non sans avoir eu une interaction avec l'utilisateur.

Cependant les retours implicites sont également employés dans les HiMAS pour les données sensibles concernant un autre agent, donc un autre utilisateur. Par exemple, lorsque deux agents effectuent une transaction de données sensibles. En effet, le consommateur intègre directement les données sensibles échangées, avec les souhaits de l'utilisateur concerné attachés, sans demander de retour explicite aux entités concernées par ces données. Cet exemple nous amène à présenter la maintenance des profils.

4.3.4 Maintenance des profils

Une fois qu'un agent a créé sa sphère privée, il lui faut la maintenir en respectant les souhaits de l'utilisateur. Cette maintenance est activée par deux interactions : (i) la délégation d'une nouvelle donnée sensible de la part de l'utilisateur et (ii) l'inclusion de nouvelles données suite à une transaction de données sensibles.

4.3.4.1 Délégation de nouvelles données sensibles

A chaque nouvelle donnée sensible déléguée par l'utilisateur à un agent, ce dernier doit intégrer cette donnée dans sa sphère privée. Cette délégation s'effectue grâce aux graphes conceptuels donnant une représentation des profils des utilisateurs. Ainsi l'agent associe également à cette donnée les règles de gestion attachées au concept correspondant. Cette délégation est possible grâce aux retours explicites de la part de l'utilisateur.

Notons que le rafraîchissement et la suppression d'une donnée sensible se font par retour explicite de la part de l'utilisateur et sont gérés de la même manière que la délégation d'une nouvelle donnée sensible.

4.3.4.2 Inclusion de nouvelles données sensibles suite à une transaction

Suite à une transaction de données sensibles, un consommateur a pour devoir d'intégrer les données sensibles reçues dans sa sphère privée selon les souhaits du fournisseur engagé dans une telle communication afin de respecter les principes normatifs des HiMAS.

L'inclusion de nouvelles données sensibles dans la sphère privée d'un agent suite à une transaction se fait donc par des mécanismes de retours implicites, car aucun retour n'est demandé par l'agent à l'utilisateur concerné, ses souhaits étant attachés à la donnée sensible transmisse par le fournisseur.

Comme le montre la figure 4.10, un fournisseur transmet des données sensibles, déléguées par son utilisateur, à un consommateur. Une fois la transaction de données sensibles terminée (étape 1 de la figure 4.10), le consommateur intègre les données sensibles dans sa sphère privée (étape 2 de la figure 4.10), même si elles ne font pas partie des types de données sensibles définis par l'utilisateur. En effet, ces données ne concernent pas l'utilisateur qui délègue ses informations sensibles au consommateur mais à un autre utilisateur. Ensuite il attache à ces données sensibles les règles de gestion découlant des souhaits du fournisseur.

4.4 Synthèse

Le modèle de Système Multi-Agent Hippocratique permet aux agents d'un tel système de se représenter leur sphère privée et de la gérer en fonction des souhaits des utilisateurs qu'ils représentent.

Les HiMAS imposent également aux agents de respecter neuf principes normatifs afin de préserver la sphère privée lors des trois phases critiques du

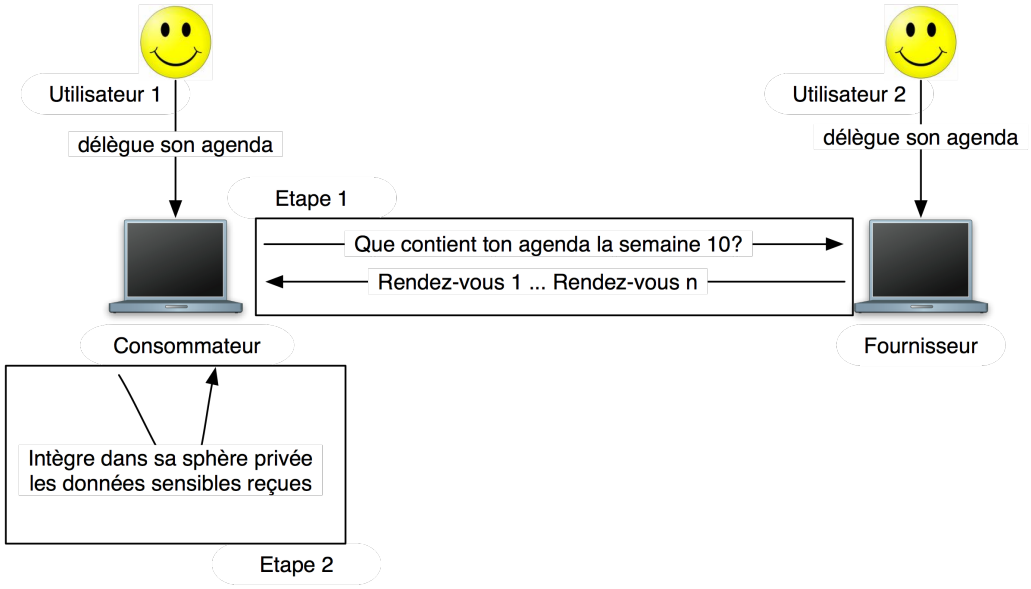

FIG. 4.10 – Inclusion de nouvelles données sensibles suite à une transaction de données sensibles.

respect de cette dernière (*stockage*, *transaction* et *devenir* des données sensibles) en prenant en considération l'ensemble de leurs besoins.

La suite de cette thèse s'intéresse à deux problèmes résultant du respect de la sphère privée des utilisateurs par des agents autonomes après la délégation des données sensibles des utilisateurs : la transaction de données sensibles et le devenir de celles-ci par la suite. C'est ce que nous abordons dans les deux prochains chapitres.

Chapitre 5

Transaction de données sensibles

Sommaire

Dès lors que les agents d'un HiMAS ont la capacité de se représenter leur sphère privée, les besoins des deuxième et troisième phases critiques du respect de la sphère privée (la *transaction de données sensibles* et le *devenir des données sensibles*) doivent être pris en compte afin de compléter le modèle que nous proposons.

[0]Une partie des travaux présentés dans ce chapitre a été publiée dans [Crépin *et al.* (à paraître 2009, 2008, à paraître 2009)].

Ce chapitre se consacre à la transaction de données sensibles et le chapitre suivant au devenir des données sensibles diffusées afin de fournir une formalisation complète des HiMAS. Nous développons dans ce chapitre un protocole de transaction de données sensibles entre deux agents autonomes d'un HiMAS afin de répondre à l'ensemble des besoins que nous avons pu identifier en conclusion du deuxième chapitre. De plus, notre proposition nous permet également de poser des bases solides pour la dernière phase critique du respect de la sphère privée, le devenir des informations sensibles.

Ce chapitre concerne les principes de consentement, de connaissance des objectifs, de transparence et de limitation en termes de diffusion, collecte, rétention et utilisation. Notre proposition est fondée sur des méta-politiques et la plate-forme pour les préférences de confidentialité [Cranor (2002); W3C (2002b)]. Il s'agit de permettre aux agents d'intégrer dans leur raisonnement les principes énoncés précédemment et de vérifier les principes de consentement, de connaissance des objectifs et de collecte minimale.

Nous commençons par présenter les principes intervenant dans la transaction de données sensibles grâce aux méta-politiques, ce qui nous permet dans la suite de ce chapitre de proposer un protocole dédié aux transactions de données sensibles à un niveau méta puis à un niveau protocole.

5.1 Expression sémantique des principes et méta-politique

Afin de déterminer précisément les liens entre les principes normatifs des HiMAS, nous proposons de les définir sous un aspect sémantique en fonction du raisonnement des agents des HiMAS. Cette étude nous amène à constater que ces principes forment en fait une ligne directrice pour la création des politiques des agents, ce qui nous conduit à présenter par la suite les méta-politiques pour finir par conclure sur la formalisation des principes mis en œuvre lors des transactions de données sensibles.

5.1.1 Expression sémantique des principes

Nous regroupons les principes des HiMAS en trois classes selon qu'ils interviennent lors de la transaction de données sensibles, lors des interactions entre agents, ou qu'ils sont en lien avec le système.

5.1.1.1 Transaction de données sensibles

Lors d'une transaction de données sensibles, le fournisseur définit une politique et le consommateur une préférence afin que chaque agent du HiMAS puisse définir ses autorisations sur les manipulations des données sensibles.

Sept des neuf principes des HiMAS doivent être respectés pour qu'une transaction de données sensibles puisse se dérouler dans un tel système :

2. Connaissance des objectifs : Le consommateur demande des données sensibles à un fournisseur afin de réaliser certaines tâches. Celles-ci permettent de définir les objectifs de la politique du consommateur qui peut alors les transmettre au fournisseur.

Exemple 5.1

Par exemple, lorsqu'un consommateur a pour tâche de trouver un créneau horaire libre avec un fournisseur afin de fixer ultérieurement un rendez-vous avec ce dernier, il peut déterminer que l'objectif de sa politique est de fixer un rendez-vous.

3. Collecte minimale : Une fois que le consommateur a défini les objectifs de sa politique, il peut alors sélectionner les données sensibles dont il a besoin pour les réaliser et uniquement celles-là.

Exemple 5.2

Dans l'exemple précédent, l'objectif $fixerRdv$ implique que le consommateur a juste besoin de recevoir les créneaux horaires libres du fournisseur.

4. Utilisation minimale : Les objectifs de sa politique étant définis, le consommateur peut déterminer les utilisations minimales possibles des données recueillies.

Exemple 5.3

Si l'objectif de la politique est de fixer un rendez-vous, le consommateur doit uniquement stocker les données sensibles reçues afin de négocier la prise de rendez-vous.

5. Diffusion minimale : La spécification des objectifs de la politique permet au consommateur de déterminer l'ensemble minimal des agents pouvant recevoir les données sensibles recueillies.

Exemple 5.4

Dans le cas où l'objectif de la politique du consommateur est de fixer un rendez-vous avec le fournisseur, il n'a pas besoin de diffuser les données sensibles reçues.

6. Rétention minimale : La spécification des objectifs de la politique définit combien de temps au maximum le consommateur va pouvoir garder en mémoire les données sensibles.

Exemple 5.5

Par exemple, pour fixer un rendez-vous, le consommateur n'a pas besoin de garder le planning du fournisseur une fois le rendez-vous pris.

8. Transparence : La transparence implique que le fournisseur et/ou le sujet appartiennent à la liste de diffusion.

1. Consentement : La mise en correspondance entre une politique et une préférence représente le principe du consentement, établi après le respect des précédents principes.

Les principes intervenant lors d'une transaction de données sensibles déterminent donc la construction de la politique du consommateur et celle de la préférence du fournisseur.

5.1.1.2 Interactions

Le principe de **conformité** (principe 9) est quant à lui relié aux interactions entre agents. En effet, pour appliquer ce principe, des transactions de données sensibles doivent avoir été exécutées et, respectant la sécurité lors du stockage des données sensibles, la seule possibilité pour les agents d'un HiMAS de raisonner sur le devenir des données sensibles diffusées est d'étudier les interactions se déroulant après les transactions.

5.1.1.3 Sécurité système

Le principe de **Sécurité** (principe 7) ne concerne pas directement le raisonnement des agents d'un HiMAS. Ce principe intervient lors de la conception du système multi-agent et ne fait donc pas partie de la formalisation présentée dans cette thèse car il est indépendant du raisonnement des agents d'un HiMAS.

5.1.2 Méta-politique

Nous présentons ici brièvement les principaux travaux relavant du le domaine des méta-politiques pour la sécurité.

Les méta-politiques ont été introduites par Hosmer dans [Hosmer (1991, 1992)]. Ces articles décrivent des politiques qui portent sur des politiques. Ces méta-politiques permettent de définir des règles de coordination sur les politiques de sécurité d'un système. Les travaux de Kühnhauser [Kühnhauser

(1995)] les utilisent pour l'interface de politiques complexes coexistantes et pour la coopération et la résolution de conflits entre politiques. Dans le système PONDER [Lupu *et al.* (2000); Twidle et Lupu (2007)], elles sont utilisées pour décrire les politiques de sécurité et résoudre les conflits entre ces politiques.

Les méta-politiques ont en général pour objectif de gérer l'ensemble des politiques de sécurité d'un système en garantissant leur définition et la détection de conflits.

5.1.3 Discussion

Les principes des HiMAS définissent des lignes directrices pour le raisonnement des agents et donc pour leur politique et préférence. Ainsi, ces principes représentent des méta-politiques pour le comportement des agents en relation avec la transmission et la manipulation des données sensibles.

Cependant, pour la préservation de la sphère privée au sein de systèmes multi-agents, la notion de politique n'est pas la même que dans les travaux portant sur la sécurité. Les principes d'un HiMAS permettent aux agents de raisonner sur un ensemble de contraintes portant sur leur comportement et non de gérer l'ensemble des politiques des agents. Nous proposons donc d'étudier les méta-politiques comme une spécialisation des méta-connaissances introduites par Pitrat [Pitrat (1990)] : elles portent sur des connaissances représentant uniquement les politiques des consommateurs d'un HiMAS.

Cette différence nous amène également à représenter différemment les principes d'un HiMAS. Sachant que la sphère privée est contextuelle, ces principes doivent proposer une définition formelle et générique des lignes directrices du comportement que les agents doivent adopter lors d'une transaction de données sensibles. Afin de permettre aux agents de raisonner sur ces principes, nous les définissons dans un dictionnaire sous forme de graphe conceptuel [Sowa (1984)] où chaque concept représente l'élément majeur d'un principe relié sémantiquement à un autre par une relation binaire.

5.2 Interprétation des principes au niveau méta

Nous proposons de décrire les principes d'un HiMAS par le biais de deux dictionnaires : un générique représentant les liens entre les principes et un deuxième représentant l'intégration du domaine.

Le dictionnaire générique définit les principes à respecter lors d'une transaction de données sensibles en guidant la conception d'une politique et d'une préférence.

La sphère privée étant dépendante du contexte et donc du domaine, les politiques et les préférences sont donc également contextuelles. Etant le résultat du raisonnement des agents, elles sont aussi dépendantes des principes impliqués dans la transaction de données sensibles. Une représentation du domaine, reliée aux principes d'un HiMAS, est donc nécessaire. Un dictionnaire du domaine est donc considéré afin d'inclure les éléments du domaine du HiMAS de l'application pour le raisonnement des agents. Ces derniers construisent alors leur préférence et leur politique, et donc une transaction de données sensibles, en se référant au dictionnaire du domaine. Ce protocole de transaction de données sensibles est résumé dans la figure 5.1. Il est ainsi possible de vérifier les limitations exprimées dans les principes d'un HiMAS grâce au dictionnaire du domaine.

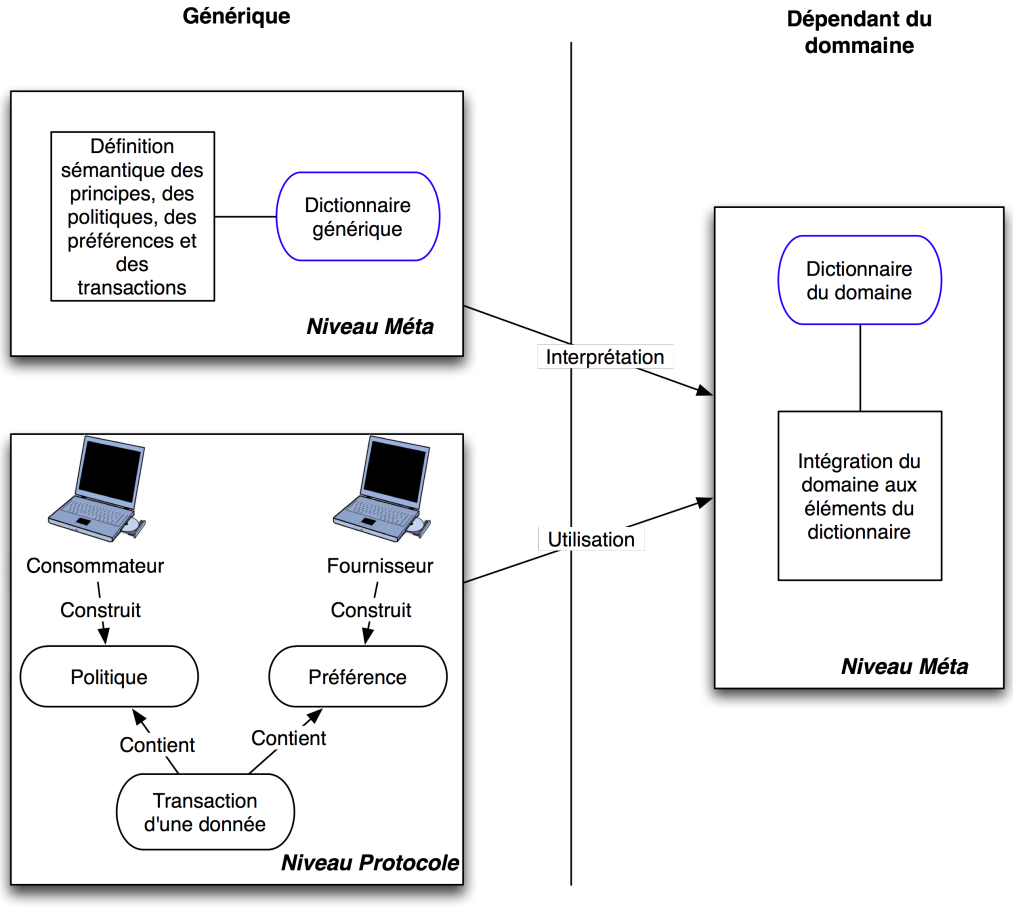

FIG. 5.1 – Modélisation du protocole de transaction de données sensibles entre les agents d'un HiMAS.

Ces deux dictionnaires sont communs à tous les agents d'un HiMAS afin que chaque agent fonde son raisonnement sur un même vocabulaire et une

même sémantique. Ils sont extérieurs aux agents et consultables par l'ensemble de ces derniers. De cette manière, les modifications apportées aux dictionnaires ne posent pas de problème de propagation et requièrent uniquement l'intervention d'une seule entité de contrôle. De plus, avec cette approche, plusieurs HiMAS relatifs à un domaine commun peuvent se référer aux mêmes dictionnaires, ce qui permet de considérer l'ouverture entre plusieurs HiMAS ayant le même domaine.

5.2.1 Dictionnaire générique

Le dictionnaire générique s'intéresse aux liens sémantiques qui existent entre les principes pris en compte dans une transaction de données sensibles afin de les formaliser pour les intégrer au raisonnement des agents d'un HiMAS.

Le deuxième principe des HiMAS (connaissances des objectifs) est au cœur du raisonnement relatif à la transaction de données sensibles (cf. figure 5.2). Ce principe permet aux agents de définir la durée de rétention, la collecte de données, la liste de diffusion incluant le principe de transparence, les différentes utilisations possibles ainsi que le format de la donnée demandée (liste des références requises). A partir de la connaissance des objectifs, le fournisseur est également apte à donner son consentement ou à refuser de transmettre une donnée sensible.

Principe	**Concept associé**
Connaissance des objectifs	*Motif* composé d'un ensemble d' *Objectif*
Collecte minimale	*Collecte* composé d'un ensemble de *Donnée*
Utilisation minimale	*UtilisationsPossibles* composé d'un ensemble d' *Utilisation*
Diffusion minimale	*ListeDiffusion* composé d'un ensemble d' *Agent*
Rétention minimale	*DuréeRétention*
Transparence	*Sujet* et *Fournisseur* inclus dans *Agent*
Consentement	*Consentement*

Tab. 5.1 – Concepts représentant les principes des HiMAS.

Afin de permettre aux agents de raisonner sur ces principes, nous les définissons dans un dictionnaire sous forme de graphe conceptuel. La liste des concepts utilisés pour représenter les principes sur lesquels nous fondons notre protocole est donnée dans le tableau 5.1 et les relations sémantiques entre ces concepts dans la figure 5.2.

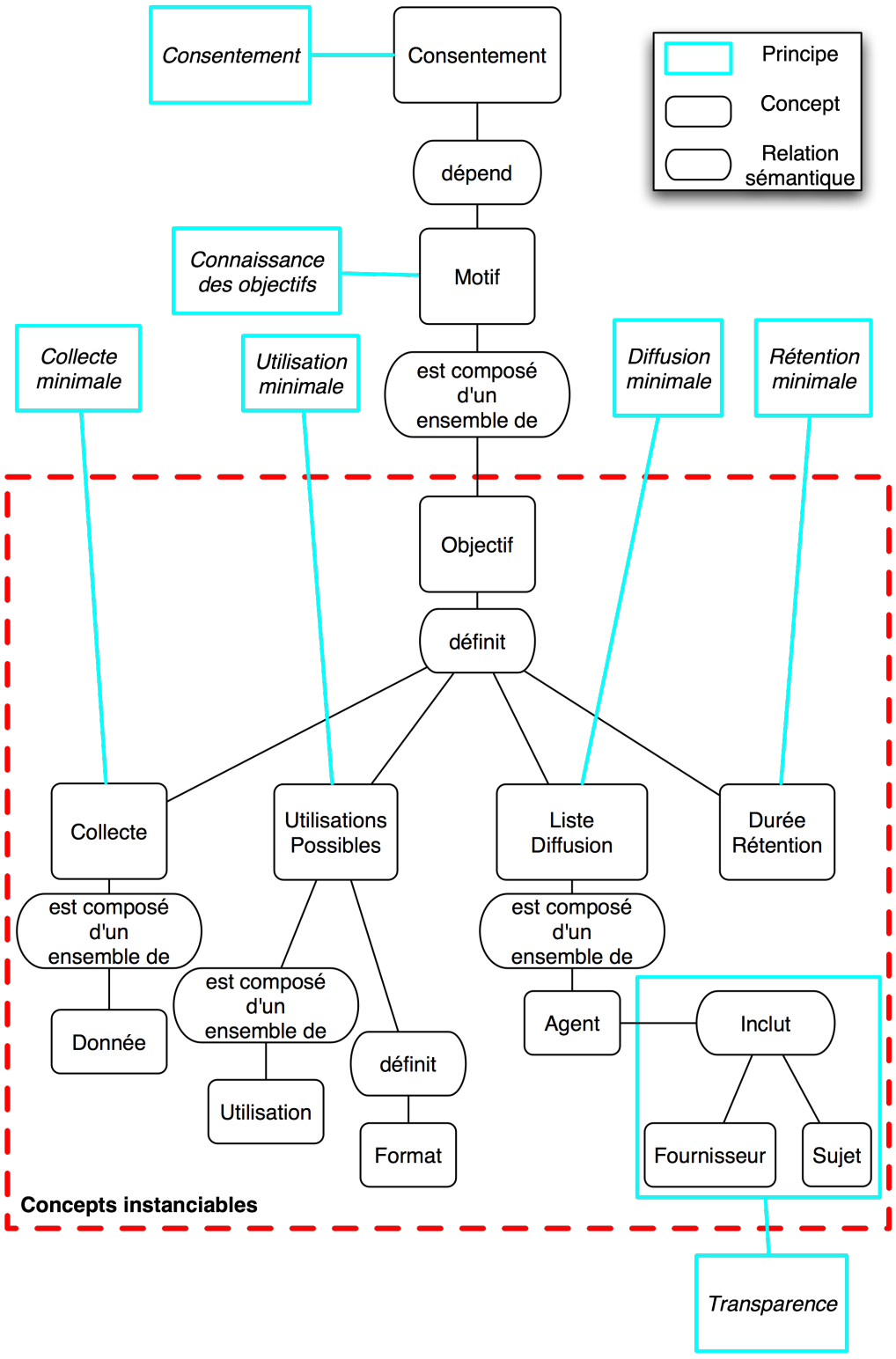

FIG. 5.2 – Graphe conceptuel des principes liés au raisonnement lors de la transaction de données sensibles dans un HiMAS.

La description formelle du graphe conceptuel présenté dans la figure 5.2 est décrite dans le tableau 5.2. Chaque principe et la notion de format, élément nécessaire au respect de la sphère privée dans les politiques et les préférences des agents d'un HiMAS, représentent un concept relié aux autres par une relation *définit*, *dépend* ou *inclut* selon le lien sémantique qui existe entre deux concepts. Afin d'obtenir une définition formelle de ce graphe conceptuel, nous utilisons un fragment de la logique existentielle, positive et conjonctive du premier ordre pour ne pas obtenir d'information logique contradictoire au sein du graphe conceptuel. Nous représentons ainsi chacun de ces concepts par un prédicat atomique et chaque relation par un prédicat binaire.

	Principe : 1. Connaissance des objectifs
$\forall p\ Motif(p)$	$\rightarrow\ \exists x\ composéDe(p,x) \wedge Objectif(x)$
	Principe : 3. Collecte minimale
$\forall x\ Objectif(x)$	$\rightarrow\ \exists y\ définit(x,y) \wedge Collecte(y)$
$\forall y\ Collecte(y)$	$\rightarrow\ \exists z\ composéDe(y,z) \wedge Donnée(z)$
	Principe : 4. Utilisation minimale
$\forall x\ Objectif(x)$	$\rightarrow\ \exists y\ composéDe(x,y) \wedge UtilisationsPossibles(y)$
$\forall y\ UtilisationsPossibles(y)$	$\rightarrow\ \exists z\ composéDe(y,z) \wedge Utilisation(z)$
$\forall y\ UtilisationsPossibles(y)$	$\rightarrow\ \exists z\ définit(y,z) \wedge Format(z)$
	Principe : 5. Diffusion minimale
$\forall x\ Objectif(x)$	$\rightarrow\ \exists y\ définit(x,y) \wedge ListeDiffusion(y)$
$\forall y\ ListeDiffusion(y)$	$\rightarrow\ \exists z\ composéDe(y,z) \wedge Agent(z)$
	Principe : 8. Transparence
$\forall z\ Agent(z)$	$\rightarrow\ \exists w\ inclut(z,w) \wedge Sujet(w)$
$\forall z\ Agent(z)$	$\rightarrow\ \exists w\ inclut(z,w) \wedge Fournisseur(w)$
	Principe : 6. Rétention minimale
$\forall x\ Objectif(x)$	$\rightarrow\ \exists y\ définit(x,y) \wedge DuréeRétention(y)$
	Principe : 2. Consentement
$\forall c\ Consentement(c)$	$\rightarrow\ \exists p\ dépends(c,p) \wedge Motif(p)$

TAB. 5.2 – Formalisation des principes mis en œuvre lors d'une transaction de données sensibles en logique du premier ordre.

Afin que les agents d'un HiMAS puissent définir leurs politiques et leurs préférences, le dictionnaire générique doit être associé au domaine du HiMAS grâce à un dictionnaire du domaine.

5.2.2 Dictionnaire du domaine

Le dictionnaire générique définit un vocabulaire pour le dictionnaire du domaine. Ce dernier instancie le dictionnaire générique en donnant toutes les

valeurs possibles aux concepts selon le domaine et en mettant en relation ses valeurs.

Afin d'illustrer notre dictionnaire du domaine, nous avons choisi un cas de transaction de données sensibles : fixer un rendez-vous de groupe. Nous considérons les créneaux horaires libres et occupés des agendas des membres du groupe comme étant les données sensibles à protéger.

Exemple 5.6

Dans cet exemple, un consommateur veut fixer un rendez-vous avec un fournisseur et d'autres agents (groupe G) dans une période donnée (un intervalle de temps borné par deux créneaux de temps). La figure 8.12 illustre ce cas d'étude. Nous considérons ici que le fournisseur incarne également le rôle de sujet. Pour fixer un tel rendez-vous, nous définissons les contraintes suivantes pour les politique des agents :

- *Les données sensibles que le consommateur peut collecter sont les créneaux libres pour une période donnée.*
- *Les données sensibles peuvent être fournies avec toutes les références que le fournisseur permet de diffuser.*
- *Les données recueillies par le consommateur ne peuvent pas être conservées en mémoire au-delà d'une date fixée.*
- *Le consommateur peut diffuser ces données sensibles aux agents du groupe G et il doit en garantir l'accès au fournisseur.*
- *Les utilisations possibles des données sensibles dans le contexte de la détermination d'un rendez-vous de groupe sont de stocker les données recueillies, de les utiliser pour négocier un rendez-vous avec le fournisseur et de partager ces données avec les agents du groupe G.*

5.2.3 Discussion

Le dictionnaire générique et le dictionnaire du domaine sont définis à partir des principes d'un HiMAS liés au raisonnement des agents lors d'une transaction. Ils représentent le vocabulaire contextuel nécessaire aux agents pour se comprendre et envisager les différentes conséquences d'une transaction de données sensibles.

A partir des dictionnaires, le consommateur peut établir sa politique en respectant les limitations imposées par les principes d'un HiMAS. Ce respect est assuré par le biais de la définition des liens sémantiques qui existent entre les concepts des principes d'un HiMAS. Il en va de même pour un fournisseur lorsque celui-ci construit sa préférence.

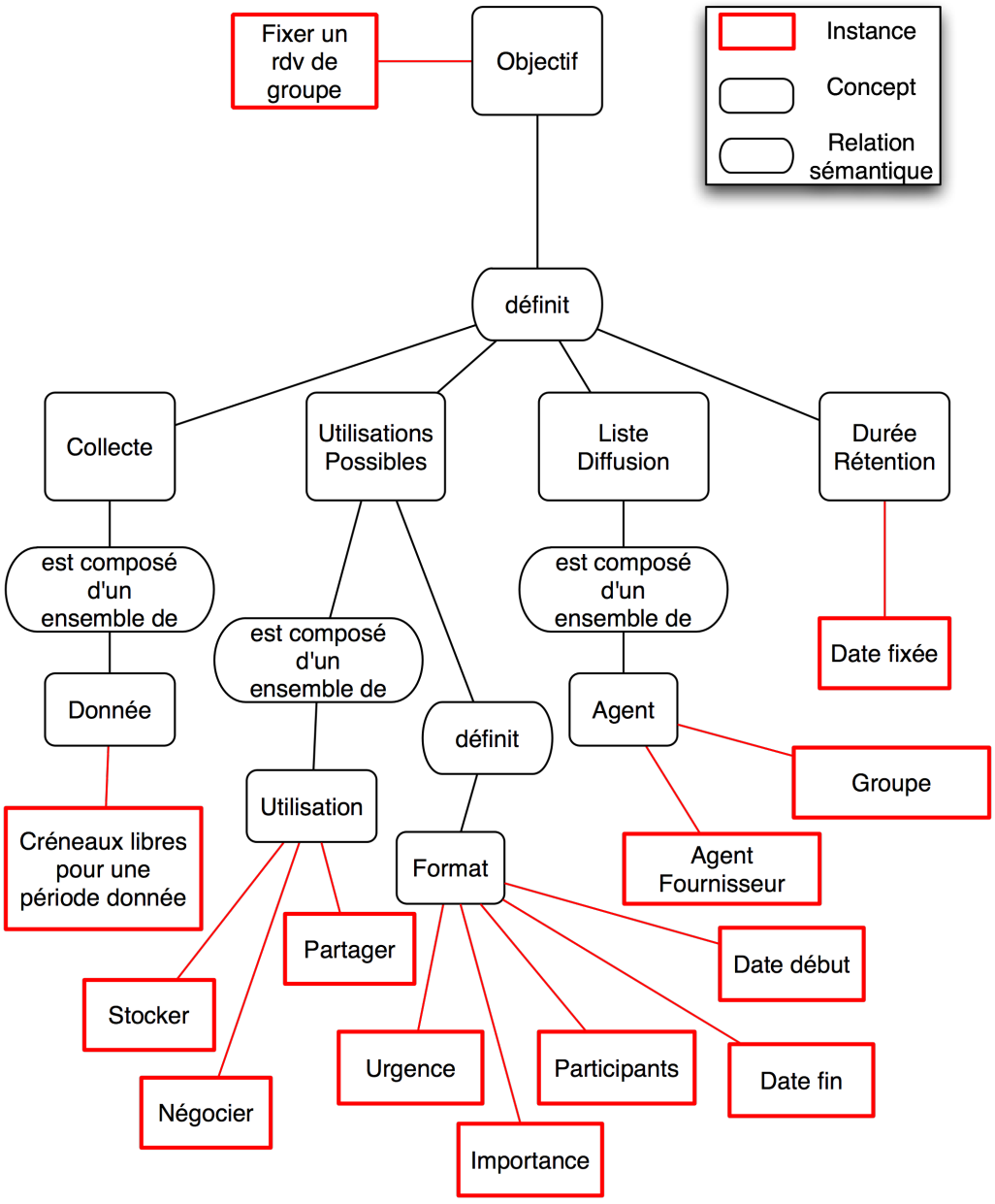

FIG. 5.3 – Dictionnaire du domaine pour l'objectif fixer un rendez-vous de groupe.

5.3 Interprétation des principes au niveau protocole

Notre protocole de transaction de données sensibles se fonde sur ces deux dictionnaires et définit les transactions de données au sein d'un HiMAS.

La figure 5.4 présente le protocole de transaction de données sensibles sous un formalisme UML qui est exploité tout au long de cette section.

Ce protocole de transaction est centré utilisateur et se place à l'opposé des protocoles de communication rencontrés dans la littérature qui sont essentiellement centrés service tout en reprenant les principes de [W3C (2002b)]. Pour que le respect de la sphère privée soit complet, ce protocole doit être intégré à un média de communication sécurisé (principe de sécurité des HiMAS).

Afin que les agents d'un HiMAS puissent représenter leurs souhaits en termes de respect de la sphère privée, nous utilisons les concepts de politique et de préférence que nous choisissons d'exprimer de façon similaire à [W3C (2002b); Cranor (2002)].

Définition 5.1 (*Politique et Préférence*)

Soit $\mathcal{A}$ l'ensemble des agents du HiMAS, Dates l'ensemble des dates et Données l'ensemble des données sensibles. Une politique est un quintuplet :

$$politique := < Objectifs, Utilisations, format,$$
$$ListeDiffusion, duréeRétention >$$

Une préférence est un quintuplet :

$$préférence < Objectifs, Utilisations, format,$$
$$ListeDiffusion, duréeRétention >$$

- $Objectifs \subset Motif$ *: un ensemble d'objectifs (les objectifs se rapprochent de la notion de désir, comme par exemple dans le modèle BDI [Bratman (1987)]) ,*
- $Utilisations \subset UtilisationsPossibles$ *: les futures utilisations possibles de la donnée sensible,*
- $format \in Données$ *: le format de la donnée (liste des références requises),*
- $ListeDiffusion \subset \mathcal{A}$ *: liste des agents pouvant recevoir les données sensibles,*
- $duréeRétention \in Dates$ *: une date de suppression des données sensibles.*

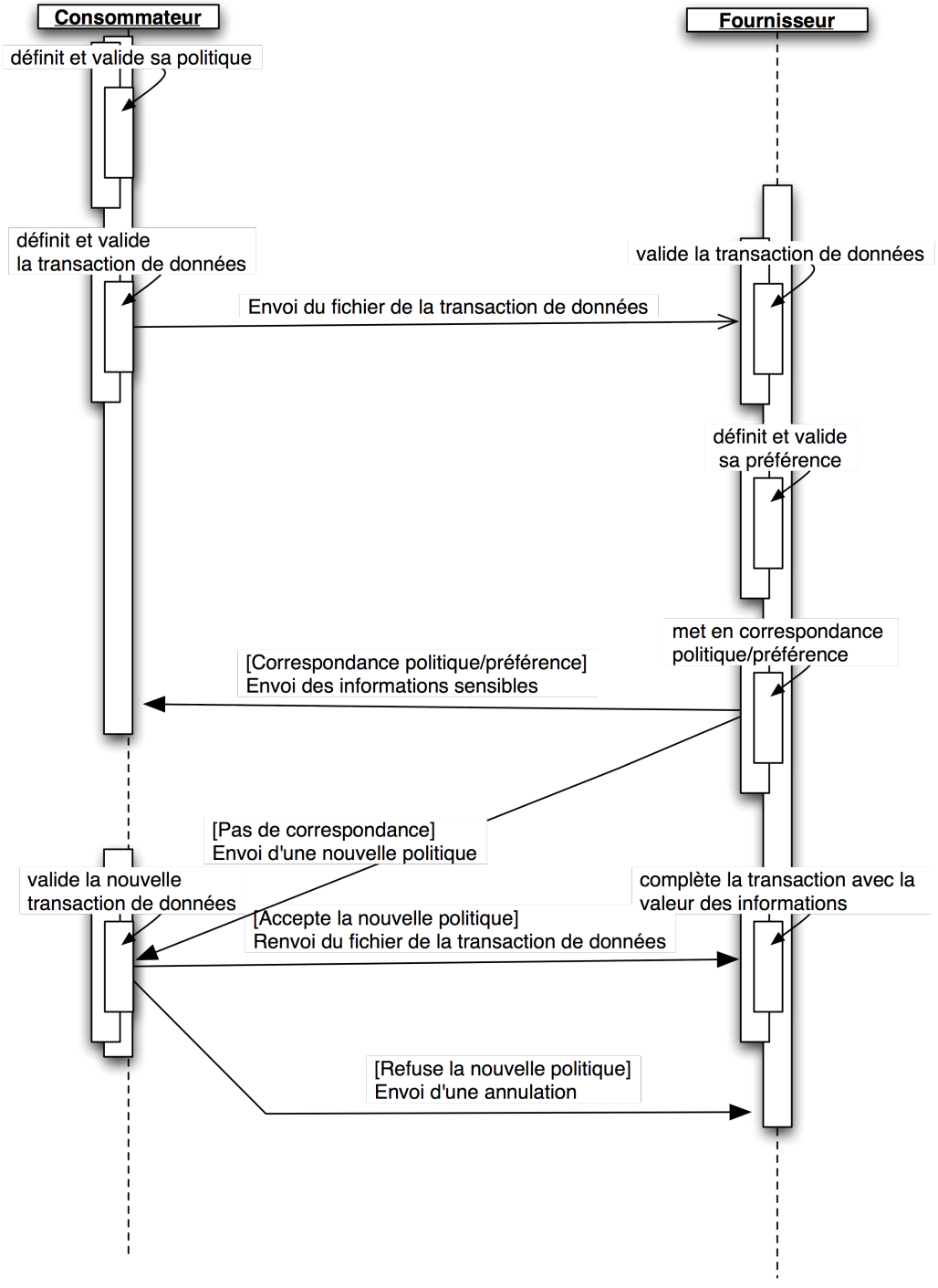

FIG. 5.4 – Protocole de transaction de données sensibles au sein d'un HiMAS.

Lors d'une transaction de données sensibles, le fournisseur (resp. le consommateur) construit sa préférence (resp. sa politique) en fonction de ses besoins et selon le dictionnaire du domaine du HiMAS.

Nous allons maintenant présenter la construction de chaque composant de ce type de communication : la politique, la préférence et la transaction de données sensibles.

5.3.1 Politique

Un consommateur construit sa politique en fonction des objectifs qu'il doit réaliser. A partir de la connaissance de ses objectifs, cet agent est capable de construire sa politique en utilisant le dictionnaire du domaine afin qu'il puisse être compris des autres agents et que son comportement soit correct envers le respect de la sphère privée.

Dans notre dictionnaire du domaine, les objectifs du consommateur sont reliés sémantiquement à tous les autres concepts utilisés dans une transaction de données sensibles. Le dictionnaire du domaine contient, pour chaque objectif, toutes les valeurs possibles respectant la sphère privée. Ainsi, un consommateur peut savoir s'il viole la sphère privée d'un fournisseur ou non en vérifiant que les éléments de sa politique sont contenus dans le dictionnaire du domaine et qu'ils sont liés correctement d'un point de vue sémantique.

A partir de ses objectifs et du dictionnaire du domaine, un consommateur peut construire sa politique en respectant la sphère privée au niveau sémantique.

Afin de définir entièrement notre protocole, il faut que cette politique soit correcte d'un point de vue syntaxique. Pour ce faire, une politique doit contenir, en accord avec la définition 5.1, la spécification des objectifs, la date de suppression des données sensibles collectées, la liste de diffusion de ces informations et l'ensemble des références demandées pour chaque information (figure 5.5 et tableau 5.3).

$\forall y$	$politique(y)$	$\rightarrow \exists z\ objectifs(z) \wedge Représente(y, z)$
$\forall y$	$politique(y)$	$\rightarrow \exists z\ utilisations(z) \wedge Représente(y, z)$
$\forall y$	$politique(y)$	$\rightarrow \exists z\ format(z) \wedge Représente(y, z)$
$\forall y$	$politique(y)$	$\rightarrow \exists z\ listeDiffusion(z) \wedge Représente(y, z)$
$\forall y$	$politique(y)$	$\rightarrow \exists z\ duréeRétention(z) \wedge Représente(y, z)$

TAB. 5.3 – Formalisation des politiques.

Exemple 5.7

Le consommateur bob désire demander son planning, incluant les rendez-vous importants et urgents, au fournisseur alice afin de fixer un rendez-

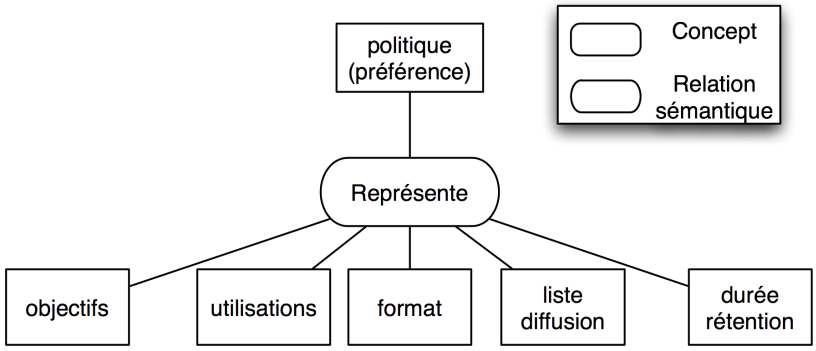

FIG. 5.5 – Politiques et préférences.

vous de groupe avec le groupe G. Soit $politque_{983}$ *la politique de bob :*

$$politique_{983} :=< fixerRdvGroupe,$$

$$\{stocker, négocier, partager\}, \{important, urgent\},$$

$$\{alice, G\}, 31/12/2009 >$$

Pour résumer, un consommateur effectue quatre opérations pour commencer une transaction de données sensibles :

1. Définir les éléments de sa politique.
2. Valider sa politique avec le dictionnaire du domaine.
3. Valider sa politique d'un point de vue syntaxique.
4. Envoyer sa politique au fournisseur par une transaction de données sensibles.

5.3.2 Préférence

A partir des éléments de gestion de sa sphère privée, un fournisseur établit les règles d'utilisation, de diffusion et de rétention des informations sensibles qu'il détient. Une fois qu'il a reçu une demande de transaction de données sensibles, ses règles lui permettent de confirmer ou d'infirmer la politique du consommateur.

Exemple 5.8

Suite à la demande du consommateur bob, alice crée à partir de son profil utilisateur sa préférence, notée $préférence_{1365}$ *:*

$$préférence_{1365} :=< fixerRdvGroupe,$$

$$\{stocker, négocier, partager\}, \{important, urgent\},$$
$$\{alice, G\}, 31/12/2009 >$$

Avant de s'intéresser à la politique du consommateur, le fournisseur doit en premier lieu vérifier la validité de la transaction d'un point de vue syntaxique (cf. définition 4.6 et figure 5.5) et d'un point de vue sémantique (vérification du dictionnaire du domaine). Ces deux validations permettent de déterminer si un consommateur présente un comportement malicieux vis-à-vis des limitations imposées par les principes d'un HiMAS.

5.3.3 Transaction de données sensibles

Une fois que le consommateur a défini et validé sa politique, la transaction de données sensibles peut être effectuée à l'initiative du consommateur.

Chacune des valeurs possibles pour les éléments d'une transaction de données sensibles est définie dans le dictionnaire du domaine afin que le consommateur construise une transaction valide pour le respect de la sphère privée d'un point de vue sémantique.

Définition 5.2 (*Transaction de données sensibles*)

Une transaction de données sensibles est définie par :

$transaction :=< consentement, Donnéessensibles, politique, contexte >$

- $consentement \in \{vrai, faux, null\}$ *: le consentement du fournisseur, initialisé à null,*
- $Donnéessensibles \subset Données$ *: les données sensibles demandées,*
- $politique \in Politiques$ *: la politique du consommateur,*
- $contexte \in Contextes$ *: contexte de la transaction de données sensibles relatif au contexte des données sensibles.*

Pour construire une transaction de données sensibles syntaxiquement correcte, nous adoptons le même procédé que pour une politique. Nous définissons une transaction de données sensibles d'un point de vue formel (tableau 5.4 et figure 5.6).

$\forall y\, transaction(y)$	$\rightarrow$	$\exists z\ consentement(z) \wedge représente(y, z)$
$\forall y\, transaction(y)$	$\rightarrow$	$\exists z\ Donnéessensibles(z) \wedge représente(y, z)$
$\forall y\, transaction(y)$	$\rightarrow$	$\exists z\ politique(z) \wedge représente(y, z)$
$\forall y\, transaction(y)$	$\rightarrow$	$\exists z\ contexte(z) \wedge dépend(y, z)$

TAB. 5.4 – Formalisation des transactions de données sensibles.

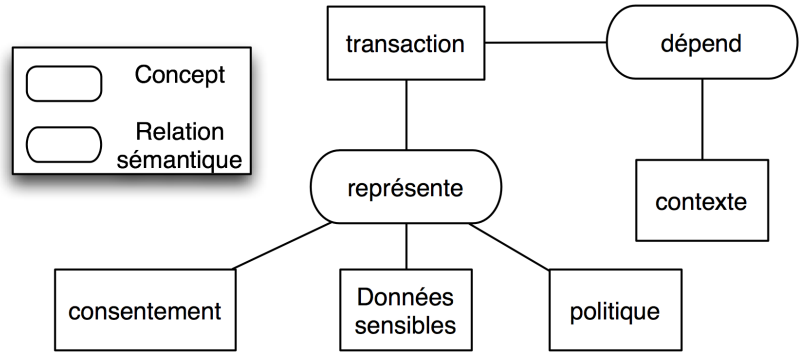

FIG. 5.6 – Transaction de données sensibles.

Une fois le fichier de la transaction de données sensibles créé et validé, le consommateur peut alors l'envoyer au fournisseur afin que ce dernier en prenne connaissance.

Notons que ce formalisme ne fait pas référence à la préférence du fournisseur. En effet, une préférence et une politique s'appuyant sur les mêmes concepts, nous modélisons la préférence du fournisseur par les modifications qu'il impute à la politique du consommateur si elle ne lui convient pas lors de la réception de la transaction de données sensibles.

Si la transaction de données sensibles est validée syntaxiquement et sémantiquement par le fournisseur, ce dernier peut alors chercher à mettre en correspondance sa préférence avec la politique du consommateur (figure 5.4). Dans le cas où cette mise en correspondance ne réussit pas, le fournisseur peut modifier la politique du consommateur selon ses souhaits et lui renvoyer le nouveau fichier correspondant à la nouvelle transaction de données sensibles afin qu'il en accepte (ou non) les nouveaux termes, comme le montre la figure 5.4.

Une fois le consommateur et le fournisseur en accord sur les termes de la politique, le fournisseur complète la transaction de données sensibles avec les valeurs des données sensibles demandées. Si aucun accord n'est trouvé, la transaction n'aboutit pas et le fournisseur ne peut pas satisfaire la requête du consommateur.

Exemple 5.9

Lorsque bob demande son planning à alice avec la politique $politique_{983}$*, alice accepte cette transaction car sa préférence,* $préférence_{1365}$*, concorde avec* $politique_{983}$*.*

Dans le cas où alice a comme préférence $préférence_{297}$*, définie comme suit, ce fournisseur propose à bob d'accepter sa préférence comme*

politique.

$$préférence_{297} :=< fixerRdvGroupe,$$
$$\{stocker, négocier, partager\}, \{important\},$$
$$\{alice, G\}, 31/12/2009 >$$

Si bob refuse, la transaction de données sensibles échoue.

Si bob accepte la préférence d'alice, il modifie sa politique comme suit :

$$politique_{983} :=< fixerRdvGroupe,$$
$$\{stocker, négocier, partager\}, \{important\},$$
$$\{alice, G\}, 31/12/2009 >$$

Ainsi la transaction de données sensibles est formalisée comme suit :

$$< vrai, agenda, politique_{983}, professionnel >$$

5.4 Synthèse

Notre protocole de transaction de données sensibles permet le **respect de sept des neuf principes** d'un HiMAS : connaissance des objectifs, collecte limitée, utilisation limitée, diffusion limitée, rétention limitée, transparence et consentement. En effet, ce protocole centré utilisateur offre la capacité aux agents d'un HiMAS de définir des politiques et des préférences en accord avec le respect de la privée et de ce fait, d'effectuer des transactions de données sensibles ne faisant pas intervenir de comportement suspicieux.

Le respect de ces principes s'effectue par la conception d'un protocole en deux niveaux :

1. **Niveau méta** : les principes sont formalisés dans un dictionnaire générique qui définit les liens entre les concepts des principes puis dans un dictionnaire du domaine qui permet de donner une valeur contextuelle aux concepts des principes.
2. **Niveau protocole** : les agents utilisent le dictionnaire du domaine pour construire une transaction de données sensibles respectant la sphère privée.

En liant les concepts des principes, nous déterminons dans un dictionnaire du domaine l'ensemble maximal des manipulations de données sensibles qu'un consommateur peut effectuer sur les informations sensibles qu'il recueille. Un fournisseur peut alors vérifier qu'un consommateur respecte les principes limitatifs en se référant à ce dictionnaire. Afin qu'aucun principe ne soit omis

dans une transaction de données sensibles, nous formalisons également cette transaction. Cette formalisation contribue également à la détection des agents suspicieux, c'est-à-dire qui ne respectent pas les principes imposés lors d'une transaction de données sensibles.

Le fait d'inclure un dictionnaire du domaine dans notre protocole permet de ne pas être confronté au même problème que la plate-forme pour les préférences de confidentialité [Thibadeau (2000)]. En effet, la mise en correspondance entre une politique et une préférence se fait en fonction du dictionnaire du domaine, ce qui permet aux agents de comprendre les intentions des consommateurs. Le deuxième avantage de l'introduction d'un dictionnaire du domaine réside en la possibilité de définir les limitations imposées par les principes d'un HiMAS.

Ayant défini les principes liés au raisonnement des agents lors d'une transaction de données sensibles, il nous faut maintenant définir le principe de **conformité** lié aux autres interactions dans un HiMAS.

Chapitre 6

Contrôle social hippocratique

Sommaire

Les agents d'un HiMAS étant capables de construire leurs préférences et leurs politiques en termes de manipulations des données sensibles échangées en y intégrant les principes des HiMAS, nous nous intéressons maintenant à la manière dont le principe de conformité peut être appliqué au vu des possibles manipulations des données sensibles stockées après les transactions de données sensibles. Notre proposition se focalise sur le respect des principes d'utilisation, de rétention, de diffusion et de transparence, principes qui ne peuvent être vérifiés qu'après ces transactions spécifiques.

Pour ce faire, dans un premier temps, nous proposons d'attacher à chaque donnée sensible la politique résultante du protocole de transaction de données sensibles proposée dans le chapitre précédent. Ainsi, à chaque donnée sensible échangée sont liées les contraintes sur les futures manipulations des données sensibles à respecter pour ne pas violer la sphère privée. Ces contraintes portent sur la diffusion, la rétention, l'utilisation et la transparence des données sensibles.

Les HiMAS doivent également donner aux agents la capacité de vérifier le respect ou la violation des politiques résultantes des transactions de données sensibles. Nous proposons de modéliser cette tâche sous la forme d'un contrôle social installant une régulation interne du comportement des agents afin de ne pas porter atteinte au respect de la sphère privée. Ce contrôle social est fondé sur un processus de construction et de gestion de la confiance dont les différents paramètres d'entrée sont les réputations relatives aux autres agents et donc aux utilisateurs. Pour ce faire, afin de préserver la sphère privée des utilisateurs, nous incluons ces réputations dans la sphère privée des agents. Cette tâche s'effectue en appliquant les neuf principes d'un HiMAS au processus de construction et de gestion de la confiance utilisé afin d'instaurer un contrôle social hippocratique. Pour cela, nous définissons le modèle de confiance que les agents utilisent pour juger de leur fiabilité et établir par la suite une relation de confiance en adéquation avec le modèle HiMAS. Nous proposons de modéliser la base de ce contrôle social par l'engagement social obtenu suite à une transaction de données sensibles entre un consommateur et un fournisseur. Ces deux éléments mettent en place notre proposition de contrôle social hippocratique. Nous montrons les répercussions qu'il engendre sur les interactions entre les agents d'un HiMAS.

6.1 Modèle de confiance pour un contrôle social hippocratique

Les différents travaux présentés sur la confiance et la réputation au chapitre 3 ont permis de choisir un modèle de confiance multi-agent pour le respect de la sphère privée : le modèle de Castelfranchi et Falcone [Castelfranchi et Falcone (1998)]. Nous proposons dans cette sous-section d'étendre ce modèle pour que les agents d'un HiMAS puissent utiliser les fondements de ce modèle pour la préservation de la sphère privée.

Nous fondons cette extension sur un modèle de confiance punitif : les relations de confiance ne peuvent que se détruire au fur et à mesure du temps, il n'existe aucune possibilité pour reconstruire une relation de confiance. Ce choix permet d'augmenter la volonté des agents d'éviter tout comportement suspicieux afin de pouvoir interagir le plus longtemps possible avec les autres agents.

Afin d'utiliser un contrôle social hippocratique pour le respect de la sphère privée, nous agrémentons le modèle de confiance proposé par [Castelfranchi et Falcone (1998)] de la prise en considération du caractère multi-facette des réputations ainsi que de l'intégration de nouvelles croyances de confiance.

6.1.1 Contexte et facette

Comme la sphère privée est contextuelle, chaque croyance de confiance doit également dépendre du contexte. Pour ce faire, nous associons à chaque croyance relative à la confiance, un contexte relatif à celui de la transaction de données sensibles concernée.

Définition 6.1 (*Contexte d'une croyance relative à la confiance*)

Le contexte d'une croyance de confiance, noté Ω*, représente le contexte de la transaction de données sensibles concernée :*

$$\Omega \in Contexte$$

Exemple 6.1

Dans AGENDA*, le contexte peut être relatif à la famille, au milieu professionnel ou encore au laboratoire de l'utilisateur.*

Nous utilisons les relations de confiance afin de prévenir les comportements suspicieux des agents d'un HiMAS. Un comportement suspicieux représente une violation de la politique attachée aux données sensibles après une de leurs transactions. Cette violation peut se référer à l'utilisation, la diffusion, la rétention des données sensibles et/ou au principe de transparence. Afin que les agents d'un HiMAS puissent avoir la capacité de juger les autres agents du HiMAS en fonction de ces types de violations, nous attachons une facette aux croyances relatives à la confiance.

Définition 6.2 (*Facette d'une croyance relative à la confiance*)

La facette, notée f*, d'une croyance relative à la confiance représente la violation possible de la politique attachée aux données sensibles.*

$$f \in Facettes = \{utilisation, diffusion, rétention, transparence\}$$

6.1.2 Croyances relatives à la confiance

Le modèle établi par [Castelfranchi et Falcone (1998)] introduit deux croyances de confiance (la croyance de compétence, DoA pour *degree of ability*, et la croyance de bonne volonté, DoW pour *degree of willingness*) qui sont déterminées à partir de trois sources d'informations relatives à la confiance : les expériences directes, les réputations propagées et la confiance systémique. Afin de prendre en considération ces trois sources d'informations, et donc d'obtenir les deux croyances de confiance permettant d'établir une relation de confiance (DoA et DoW), nous introduisons trois nouvelles croyances de confiance.

Chaque croyance relative à la confiance est représentée par un réel appartenant à l'intervalle [0,1].

Définition 6.3 (*Croyance de réputation directe*)

Soit $\mathcal{A}$ l'ensemble des agents, Contextes l'ensemble des contextes et Facettes l'ensembles des facettes.

La croyance de réputation directe qu'un fournisseur accorde à un consommateur est liée à l'ensemble des expériences directes relatives au respect des politiques par le consommateur $c \in \mathcal{A}$ pour la facette $f \in Facettes$ pour le contexte $\Omega \in Contextes$. Cette croyance est notée $DoDR_{c,\Omega,f}$ pour degree of direct reputation.

Définition 6.4 (*Croyance de réputation propagée*)

Soit $\mathcal{A}$ l'ensemble des agents, Contextes l'ensemble des contextes et Facettes l'ensembles des facettes.

La croyance de réputation propagée qu'un fournisseur accorde à un consommateur est relative aux réputations propagées sur le respect et les violations de politiques par le consommateur $c \in \mathcal{A}$ pour la facette $f \in Facettes$ qu'il a recueillie pour le contexte $\Omega \in Contextes$. Cette croyance est notée $DoPR_{c,\Omega,f}$ pour degree of propagated reputation.

Définition 6.5 (*Croyance de réputation stéréotypée*)

Soit $\mathcal{A}$ l'ensemble des agents, Contextes l'ensemble des contextes et Facettes l'ensembles des facettes.

La croyance de réputation stéréotypée qu'un fournisseur accorde à un consommateur est relative aux caractéristiques (groupe par exemple) du consommateur $c \in \mathcal{A}$ pour la facette $f \in Facettes$ pour le contexte $\Omega \in Contextes$ en relation avec les politiques que c s'est engagé à tenir. Cette croyance est notée $DoSR_{c,\Omega,f}$ pour degree of stereotyped reputation.

A partir de ces trois croyances, les fournisseurs d'un HiMAS sont aptes à calculer les deux croyances de confiances nécessaires à l'établissement ou non d'une relation de confiance, comme le montre la fonction de confiance représentée dans la figure 6.1.

Définition 6.6 (*Croyance de compétence*)

Soit $\mathcal{A}$ l'ensemble des agents, Contextes l'ensemble des contextes et Facettes l'ensembles des facettes.

La croyance de compétence qu'un fournisseur accorde à un consommateur est relative aux croyances $DoDR$ et $DoPR$. La croyance de compétence que le fournisseur accorde au consommateur $c \in \mathcal{A}$ pour la facette $f \in Facettes$ pour le contexte $\Omega \in Contextes$ est calculée par la fonction

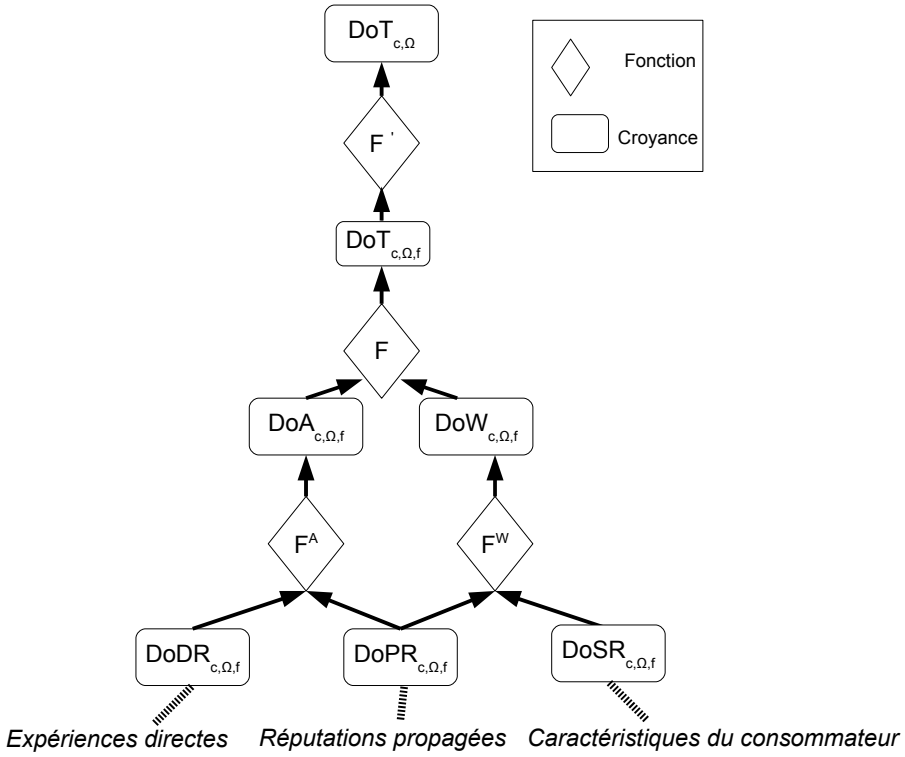

FIG. 6.1 – Fonction de confiance.

monotone décroissante suivante :

$$DoA_{c,\Omega,f} = F^A(DoDR_{c,\Omega,f}, DoPR_{c,\Omega,f})$$

Exemple 6.2

Soit l'agent alice qui reçoit une demande de transaction de données sensibles de la part du consommateur bob dans un contexte $professionnel$. Prenons pour illustrer cet exemple, les croyances relatives à la facette $diffusion$.

alice possède comme croyances de confiance relatives à bob :

- $DoDR_{bob,professionnel,diffusion} = 0,7$
- $DoPR_{bob,professionnel,diffusion} = 0,85$.

Pour calculer la croyance de compétence que cet agent accorde à bob, prenons comme fonction de confiance la moyenne des valeurs des croyances de confiance (notons que, selon les besoins de l'application, cette fonction peut différer en pondérant les croyances ou peut différer en utilisant une fonction plus complexe et mieux élaborée en fonction de ces besoins par exemple) :

$$DoA_{bob,professionnel,diffusion} = (0,7 + 0,85)/2 = 0,775$$

Définition 6.7 (*Croyance de bonne volonté*)

Soit $\mathcal{A}$ l'ensemble des agents, $Contextes$ l'ensemble des contextes et $Facettes$ l'ensembles des facettes.

La croyance de bonne volonté qu'un fournisseur accorde à un consommateur est relative aux croyances $DoPR$ et $DoSR$. La croyance de bonne volonté que le fournisseur accorde au consommateur $\in \mathcal{A}$ pour la facette $f \in Facettes$ pour le contexte $\Omega \in Contextes$ est calculée avec la fonction monotone décroissante suivante :

$$DoW_{c,\Omega,f} = F^W(DoPR_{c,\Omega,f}, DoSR_{c,\Omega,f})$$

Exemple 6.3

alice accorde comme valeur 0,6 pour sa croyance de réputation stéréotypée $DoSR_{bob,professionnel,diffusion}$.

Pour calculer la croyance de bonne volonté pour la facette $diffusion$ dans le contexte $professionnel$ qu'alice accorde à bob, alice applique la fonction suivante :

$$DoW_{bob,professionnel,diffusion} = (0,85 + 0,6)/2 = 0,725$$

Une fois ces deux croyances déterminées, un agent est capable de déterminer la croyance de confiance pour une facette donnée qu'il accorde à un autre agent (figure 6.1).

Définition 6.8 (*Croyance de confiance pour une facette donnée*)

Soit $\mathcal{A}$ l'ensemble des agents, Contextes l'ensemble des contextes et Facettes l'ensembles des facettes.

La croyance de confiance pour une facette donnée qu'un fournisseur accorde à un consommateur est relative aux croyances $DoA_{c,\Omega,f}$ et $DoW_{c,\Omega,f}$. La croyance de confiance pour une facette donnée que le fournisseur accorde au consommateur $c \in \mathcal{A}$ pour la facette $f \in Facettes$ pour le contexte $\Omega \in Contextes$ est calculée avec la fonction monotone décroissante suivante :

$$DoT_{c,\Omega,f} = F(DoA_{c,\Omega,f}, DoW_{c,\Omega,f})$$

Exemple 6.4

Pour calculer la croyance de confiance pour la facette $diffusion$ dans le contexte $professionnel$ qu'alice accorde à bob, cet agent applique la fonction suivante :

Pour calculer la croyance de bonne volonté pour la facette $diffusion$ dans le contexte $professionnel$ que cet agent accorde à bob, alice applique la fonction suivante :

$$DoT_{bob,professionnel,diffusion} = (0,775 + 0,725)/2 = 0,75$$

Définition 6.9 (*Croyance de confiance*)

Soit $\mathcal{A}$ l'ensemble des agents et Contextes l'ensemble des contextes.

La croyance de confiance qu'un fournisseur accorde à un consommateur est relative aux croyances de confiance pour une facette donnée. La croyance de confiance que le fournisseur accorde au consommateur $c \in \mathcal{A}$ pour le contexte $\Omega \in Contextes$ est calculée avec la fonction monotone décroissante suivante :

$$DoT_{c,\Omega} = F'(DoA_{c,\Omega,f_1}..DoA_{c,\Omega,f_n}, DoW_{c,\Omega,f_1}..DoW_{c,\Omega,f_n})$$

Exemple 6.5

Soit les croyances de confiance relatives à l'agent bob pour le contexte $professionnel$ que l'agent alice possède :

- $DoA_{bob,professionnel,utilisation} = 0,7$
- $DoA_{bob,professionnel,diffusion} = 0,775$

– $DoA_{bob,professionnel,rétention} = 0,6$
– $DoA_{bob,professionnel,transparence} = 0,95$
– $DoW_{bob,professionnel,utilisation} = 0,5$
– $DoW_{bob,professionnel,diffusion} = 0,725$
– $DoW_{bob,professionnel,rétention} = 0,65$
– $DoW_{bob,professionnel,transparence} = 0,825$

Pour calculer la croyance de confiance que cet agent accorde à bob dans le contexte professionnel, alice applique la fonction suivante :

$$DoT_{bob,professionnel} =$$

$$(((0,7 + 0,775 + 0,6 + 0,95)/4) + (0,5 + 0,725 + 0,65 + 0,825)/4)/2$$

$$= 0,7125$$

Afin d'instaurer ou non une relation de confiance avec un autre agent, un agent doit calculer toutes les croyances de confiance qu'il possède pour chaque facette comme le montre la figure 6.1. Si cette croyance est supérieure à un seuil donné[1], une relation de confiance existe entre ces deux agents. Dans le cas contraire, aucune relation de confiance entre ces deux agents n'est instaurée.

Nous pouvons remarquer que le raisonnement appliqué à la fonction de confiance développée dans notre modèle de confiance hippocratique est similaire au fonctionnement de la fonction de seuil dans les réseaux de Hopfield [Hopfield (2007)] lors de l'évaluation séquentielle des coefficients synaptiques.

Afin d'établir un contrôle social hippocratique, nous devons définir l'objet des relations de confiance obtenues à partir du modèle que nous venons de présenter. Pour ce faire, nous proposons un engagement social hippocratique représentant le respect de la politique par un consommateur après une transaction de données sensibles.

6.2 Engagement social hippocratique

Afin de pouvoir étudier la dynamique des interactions relatives à un contrôle social, nous devons modéliser l'objet du contrôle social hippocratique. Pour ce faire, nous proposons d'utiliser la notion d'engagement social introduite par [Singh (1991, 2000)] et développée entre autres par [Bentahar *et al.* (2003); Pasquier *et al.* (2005)] en l'adaptant au respect de la sphère privée.

Le contenu de l'engagement social concerne la politique attachée aux données transmises après une transaction de données sensibles. Afin de ne pas

[1]Nous laissons le soin à l'utilisateur de définir ce seuil afin de lui laisser la possibilité d'exprimer sa disposition à faire confiance.

rendre ces données visibles, le contenu représente seulement l'intitulé de ces données dans la politique qui y est relative.

Afin de modéliser l'état de l'engagement social hippocratique, nous rejoignons [Bentahar *et al.* (2003)] qui le définissent comme suit :

- inactif : l'engagement a été rejeté par un des deux agents,
- actif : les deux agents ont explicitement accepté l'engagement social,
- violé : l'engagement social a été violé par un des deux agents qui subit donc la sanction prévue,
- accompli : les deux agents ont respecté et terminé l'engagement social,
- annulé : les deux agents ont annulé leur engagement social sans qu'une sanction ne soit appliquée ou l'engagement social est annulé après avoir été violé et après que l'agent suspicieux ait reçu sa sanction.

Définition 6.10 (*Engagement social hippocratique*)

Soit $\mathcal{A}$ l'ensemble des agents, Politiques l'ensemble des politiques, Dates l'ensemble des dates, Etats l'ensemble des états et Sanctions l'ensemble des sanctions.

Un engagement social hippocratique est modélisé comme suit :

$$HiC(consommateur, fournisseur, politique, duréeRétention, état, sanction)$$

- $consommateur \in \mathcal{A}$ *: le consommateur,*
- $fournisseur \in \mathcal{A}$ *: le fournisseur,*
- $politique \in Politiques$ *: la politique obtenue après une transaction de données sensibles,*
- $duréeRétention \in Dates$ *: la durée de l'engagement est égale à la durée de rétention des données sensibles exprimée dans la politique,*
- $état \in Etats$ *: l'état de l'engagement où* $Etats = \{inactif, actif, violé, accompli, annulé\}$,
- $sanction \in Sanctions$ *: la sanction qui sera appliquée au consommateur en cas de non respect de sa politique. Cette sanction représente une baisse du niveau de la réputation directe selon la ou les facettes qui ont été violées dans le contexte.*

Dans un HiMAS, cet engagement est mutuel mais les sanctions ne s'appliquent donc qu'au consommateur si l'engagement social est violé. En effet, seul le consommateur s'engage à respecter les contraintes du fournisseur sur ses données sensibles.

Les sanctions qui sont appliquées sont relatives au contrôle social que nous voulons instaurer. De ce fait, les sanctions encourues par un consommateur représentent une mauvaise réputation propagée pour une ou plusieurs facettes citées précédemment (utilisation, diffusion, rétention et transparence).

Définition 6.11 (*Sanction*)

Soit $\mathcal{A}$ l'ensemble des agents, Contextes l'ensemble des contextes et Facettes l'ensembles des facettes.

Lorsqu'un engagement social hippocratique est violé par un consommateur c, le fournisseur baisse sa réputation en fonction des facettes qui n'ont pas été respectées dans la politique exprimée dans la transaction de données sensibles :

$$Decrease(DoDR_{c,\Omega,f}, DoPR_{c,\Omega,f}, punition)$$

- $c \in \mathcal{A}$ *: le consommateur,*
- $\Omega \in Contexte$*,*
- $f \in Facettes$ *: la politique obtenue après une transaction de données sensibles,*
- $punition \in [0,1]$ *: la valeur de la baisse de réputations encourues. Cette valeur est fixée par l'utilisateur.*

Exemple 6.6

Lors de la transaction de données sensibles présentée en exemple dans le chapitre précédent (un consommateur désire obtenir le planning d'un fournisseur pour fixer un rendez-vous de groupe), le consommateur alice et le fournisseur bob créent l'engagement social hippocratique suivant sur la politique $politique_{983}$:

$$HiC(alice, bob, politique_{983}, 31/12/2009, actif,$$

$$Decrease(DoDR_{alice,professionnel,f}, DoPR_{alice,professionnel,f}, 0,15)$$

Nous pouvons maintenant présenter la manière dont les agents d'un HiMAS utilisent les relations de confiance et l'engagement social pour s'assurer du principe de conformité.

6.3 Formalisation du principe de conformité

Instaurer un contrôle social hippocratique permet aux agents d'un HiMAS de juger de la fiabilité des autres agents du système grâce aux relations de confiance et donc d'appliquer le principe de conformité. Le premier impact de cet apport s'applique lors du raisonnement des agents d'un HiMAS : ils doivent déterminer l'existence ou non d'une relation de confiance avec tout consommateur leur demandant des données sensibles. Le deuxième impact de ce contrôle social hippocratique s'effectue sur le protocole de transactions de données sensibles qui prend maintenant en considération ces relations de confiance.

6.3.1 Relation de confiance hippocratique

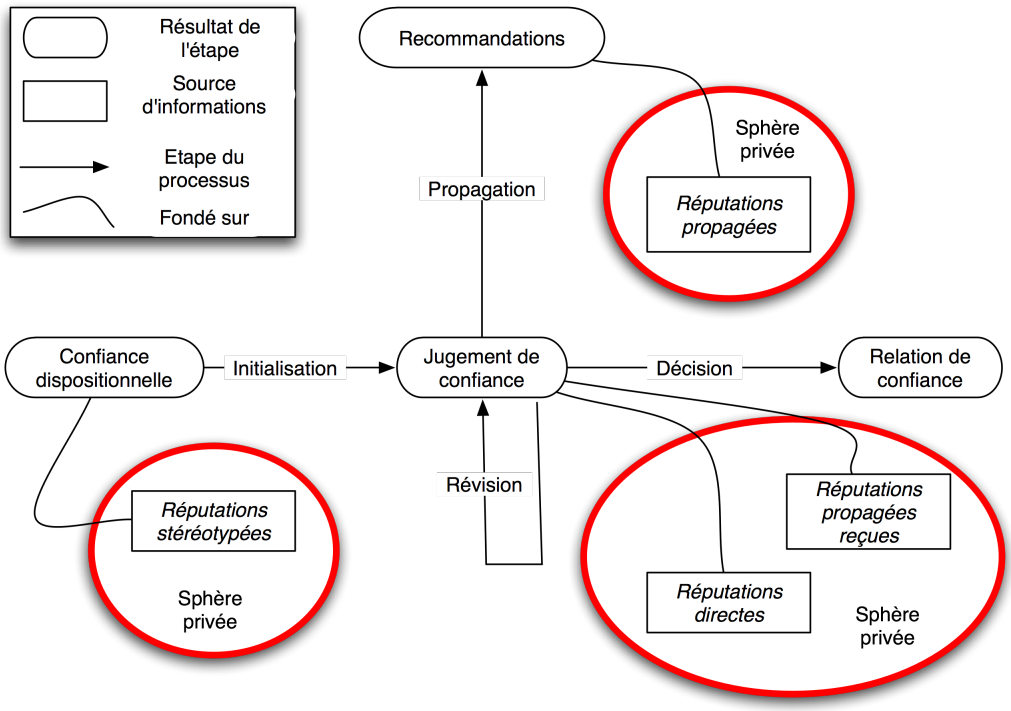

FIG. 6.2 – Processus de construction et de gestion de la confiance hippocratique.

Du fait de la propagation des réputations, nous nous devons d'inclure le traitement des informations liées à la confiance dans la sphère privée des agents d'un HiMAS. Etablir une relation de confiance dans nos recherches est donc étroitement lié à la préservation de la sphère privée. En effet, cette relation permet aux agents d'un HiMAS de porter un jugement sur la fiabilité des consommateurs vis-à-vis du respect de leur politique portant sur les données sensibles qu'ils recueillent. Ainsi, le processus qui mène à l'instauration d'une telle relation doit pouvoir être préservé au même titre que la sphère privée.

Pour assurer le respect de la sphère privée, nous proposons d'appliquer au processus de construction et de gestion de la confiance présenté précédemment les neufs principes des HiMAS en incluant, dans la sphère privée des agents, les réputations qu'ils possèdent, comme le montre la figure 6.2.

Comme pour le respect de la sphère privée, nous nous préoccupons uniquement des transactions de données sensibles, ici les réputations propagées, et de leur devenir après leur communication. En effet, pendant leur stockage, nous supposons leur sécurité assurée en termes d'attaques et d'intrusions.

De plus, lors d'une transaction de réputations propagées, les agents doivent respecter le protocole présenté dans la section précédente. Pour ce faire, nous développons dans notre dictionnaire du domaine un objectif spécifique à l'échange de réputations propagées, le contrôle social. Cet objectif définit le fait que les données recueillies sont les réputations propagées complètes et qu'elles ne doivent être utilisées que pour juger de la fiabilité d'un autre agent ou réviser un jugement antérieur s'il existe. Ces données ne doivent pas être diffusées et ne doivent rester accessibles que pendant le processus de construction et de gestion de la confiance. Ces contraintes sont représentées par la figure 6.3.

Afin de pouvoir instaurer un contrôle social, les agents d'un HiMAS sont amenés à échanger des réputations propagées. Avec notre protocole, nous imposons aux agents de fournir leurs propres données relatives à la confiance, qui peuvent être compilées grâce à d'autres réputations propagées, et non de faire suivre celles d'un autre agent.

La dernière étape pour l'élaboration d'une relation de confiance hippocratique concerne le principe de conformité. Les agents doivent être capables de juger de la fiabilité des consommateurs sur leurs politiques relatives aux recommandations (envois des réputations propagées). Pour ce faire, nous traitons ces réputations comme des données sensibles à part entière. Du fait que notre étude se focalise sur le respect de la sphère privée, le contrôle social que nous instaurons ne prend pas en considération la véracité des réputations propagées, seulement le respect des politiques des consommateurs et des préférences des fournisseurs. Nous émettons effectivement l'hypothèse selon laquelle les agents sont sincères sur les réputations qu'ils propagent. Cependant, nous pouvons considérer cette facette dans nos perspectives de travail, notamment pour étudier la coopération d'agents suspicieux par le biais des réseaux de confiance [Melaye et Demazeau (2005); Lacomme *et al.* (2009)] par exemple.

En résumé, une confiance hippocratique intègre dans son processus les neuf principes normatifs d'un HiMAS en assurant la sécurité des valeurs de confiance, en respectant le protocole de transaction de données sensibles et en permettant aux agents de vérifier la véracité et le respect des réputations propagées.

6.3.2 Transaction de données sensibles et principe de conformité

Maintenant que les agents d'un HiMAS possèdent la capacité à juger de la fiabilité des consommateurs, il faut l'intégrer dans notre protocole de transaction de données sensibles afin de prévenir les comportements suspicieux.

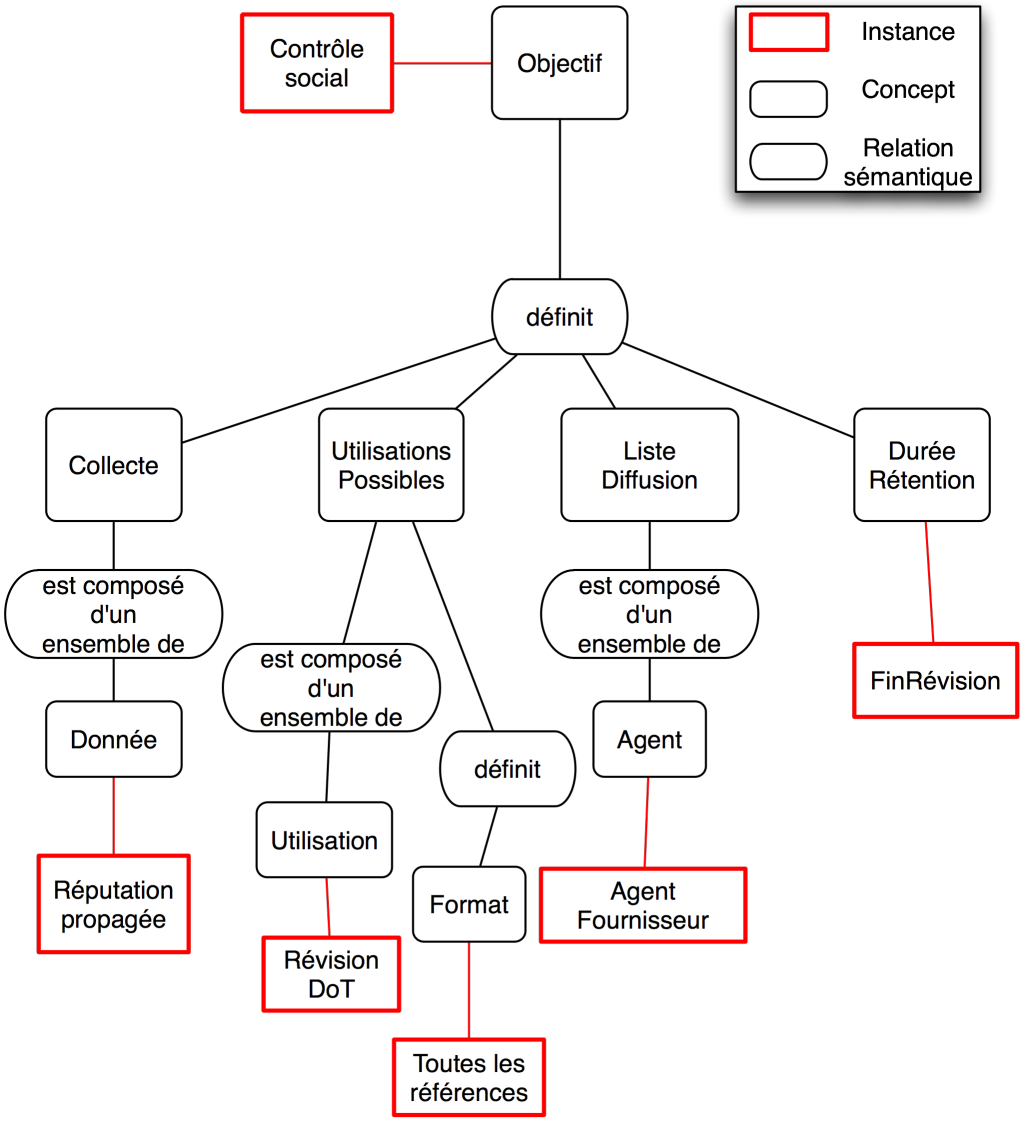

FIG. 6.3 – Dictionnaire du domaine pour l'objectif "contrôle social".

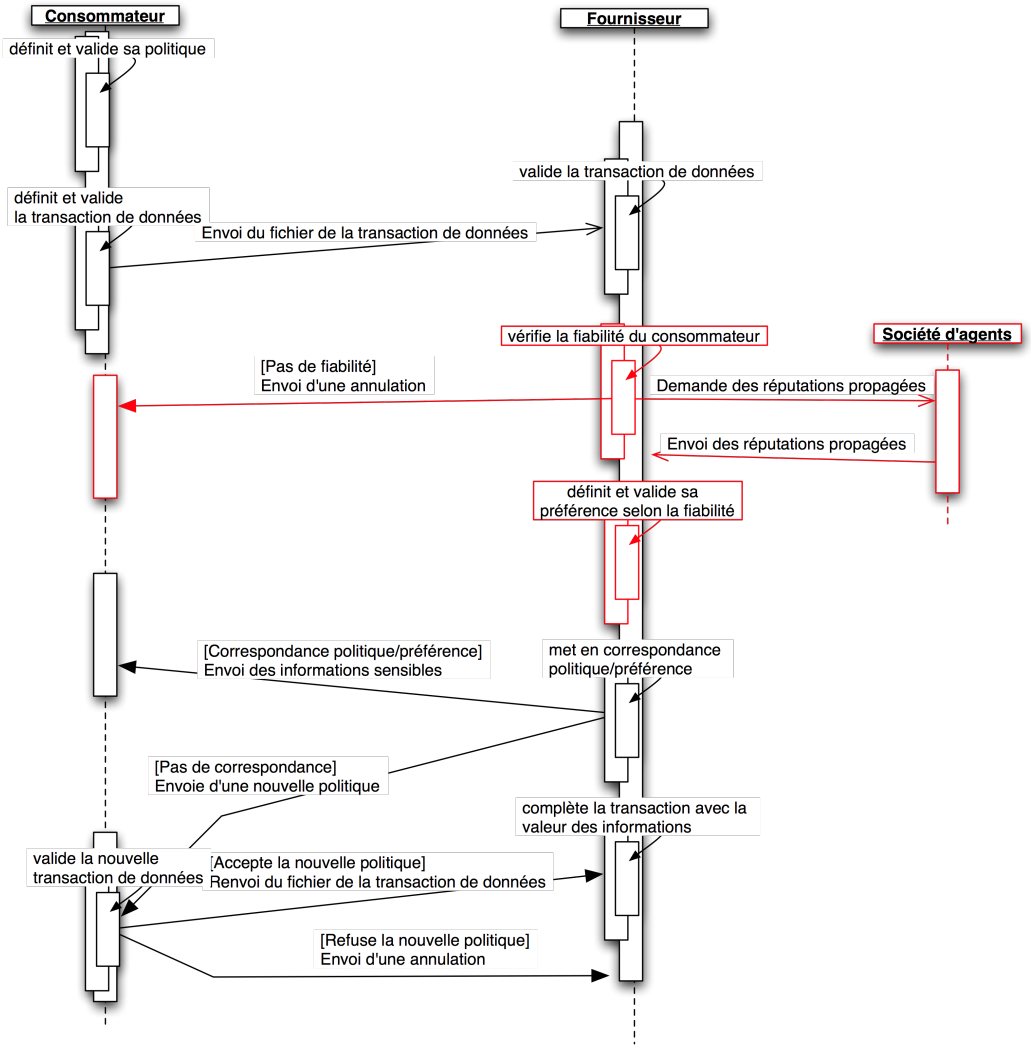

FIG. 6.4 – Protocole de transaction de données sensibles avec contrôle social hippocratique.

Les impacts du contrôle social hippocratique sur les transactions de données se focalisent sur le raisonnement du fournisseur une fois que le consommateur lui ait fait parvenir sa politique. Ces impacts sont résumés sur la figure 6.4.

Le premier s'exprime par une nouvelle étape du raisonnement du fournisseur : cet agent détermine la fiabilité du consommateur. Pour ce faire, il utilise les croyances relatives à la confiance envers le consommateur en fonction de ses propres expériences, avec ses réputations directes, mais également des expériences des autres agents par le biais des réputations propagées. Pour déterminer la fiabilité du consommateur, il applique le fonction de confiance (Définition 6.9) et change donc la valeur de son $DoT_{c,\Omega}$. Une fois que le fournisseur a déterminé la fiabilité du consommateur, il décide d'annuler la transaction si cette fiabilité est jugée insuffisante (le seuil de fiabilité n'est pas atteint).

Dans le cas contraire, le fournisseur poursuit la transaction de données sensibles selon le protocole avec notamment quelques adaptations dans la définition de sa préférence. En effet, la fiabilité du consommateur porte sur les quatre facettes de la réputation que nous avons présentées précédemment. Dans le cas où le consommateur n'est pas jugé assez fiable sur une ou plusieurs de ces facettes (valeur de fiabilité proche du seuil mais acceptable), le fournisseur peut décider de restreindre les contraintes de ces facettes dans la préférence du consommateur selon les choix des utilisateurs.

Pour conclure, présentons un exemple de transaction de données sensibles incluant le contrôle social hippocratique.

Exemple 6.7

Reprenons l'exemple de la transaction de données sensibles présentée dans le chapitre précédent.

Le consommateur bob, à partir de ses objectifs et du dictionnaire du domaine, crée et valide sa politique $politique_{983}$. *Il l'envoie au fournisseur alice afin d'obtenir son agenda dans le but de fixer un rendez-vous professionnel avec le groupe G.*

alice commence par vérifier la validation sémantique et syntaxique de $politique_{983}$. *Ensuite, elle compile la fonction F, en demandant les réputations propagées à la société d'agents, puis la fonction F′ pour calculer sa croyance de confiance envers bob dans le cadre professionnel,* $DoT_{bob,professionnel}$.

Dans le cas où $DoT_{bob,professionnel}$ *est supérieure au seuil fixé par l'utilisateur d'alice, alice crée et valide sa préférence* $préférence_{1365}$. *Cette préférence concorde avec la politique de bob. alice envoie donc les données sensibles requises à bob et l'engagement social hippocratique suivant est instauré entre ces deux agents :* $HiC(alice, bob, politique_{983}, 31/12/2009, actif, Decrease(DoDR_{alice,professionnel,F}, DoSR_{alice,professionnel,f}, punition)$.

> *Si $DoT_{bob,professionnel}$ est inférieure au seuil donné, alors alice annule la transaction de données sensibles.*

6.4 Synthèse

Les principes relatifs au raisonnement des agents d'un HiMAS sont formalisés dans nos travaux grâce à un protocole de communication spécifique aux *transactions de données sensibles* (répondant aux besoins de la deuxième phase critique du respect de la sphère privée) et grâce à un contrôle social hippocratique représentant la troisième phase critique qui concerne le *devenir des données sensibles*.

Au cours des échanges de données sensibles, les agents d'un HiMAS construisent leurs politiques et leurs préférences, selon le rôle qu'ils endossent, et intègrent ces éléments à la transaction de données sensibles afin de s'engager socialement sur le respect de ces politiques et de ces préférences. Une fois la transaction aboutie, le contrôle social hippocratique entre en jeu permettant ainsi de prévenir des violations des politiques que les consommateurs se sont engagés à respecter. Ces deux mécanismes permettent donc de modéliser pour les agents les principes de consentement, connaissance des objectifs, collecte minimale, utilisation minimale, diffusion minimale, rétention minimale, transparence et conformité tout en respectant la sphère privée.

Le prochain chapitre de notre thèse se focalise sur l'évaluation de notre proposition en présentant les expérimentations effectuées permettant de valider le contrôle social hippocratique.

Chapitre 7

Evaluation des HiMAS

Sommaire

Afin de valider notre modèle de systèmes multi-agents hippocratiques, nous présentons maintenant une évaluation de celui-ci en nous focalisant sur le contrôle social. Cette évaluation est composée de deux phases : l'évaluation des paramètres du modèle de confiance proposé et celle du contrôle social hippocratique. Commençons par présenter le scénario développé pour les expérimentations.

7.1 Scénario d'expérimentation

Cette section décrit le protocole que nous utilisons afin d'exécuter les expérimentations sur le contrôle social hippocratique que nous avons développé

sur la plate-forme agent JACK [AOS (1997 2009)]. Nous décrivons d'abord le domaine d'expérimentation, puis l'initialisation des agents pour finir par la présentation des paramètres d'expérimentation.

7.1.1 Domaine d'expérimentation

Nous choisissons comme domaine d'expérimentation la gestion décentralisée d'agendas [Demazeau *et al.* (2006)]. Dans ce contexte, chaque agent possède l'agenda d'un utilisateur et les agents vont partager et/ou diffuser les rendez-vous de leur agenda afin de prendre de nouveaux rendez-vous selon le protocole de transaction de données sensibles proposé au chapitre 6.

Après chaque transaction de données sensibles, le consommateur inclut dans sa sphère privée les données sensibles reçues, avec leur politique attachée. Lorsque ce consommateur devient fournisseur de cette donnée, il la transmet avec la politique qui est lui attachée. Ainsi, le nouveau consommateur est en mesure de vérifier l'engagement social hippocratique passé entre le premier consommateur et le premier fournisseur. Si cet engagement a été violé, le dernier consommateur prévient le premier fournisseur qui dès lors applique la punition à sa croyance de réputation directe, $DoDR_{c,\Omega,f}$, de la fonction de confiance présentée dans le chapitre 6 avec un palier donné pour la ou les facettes qui n'ont pas été respectées

Pour la gestion des réputations propagées, nous proposons d'intégrer un agent *référent* à notre HiMAS expérimental comme le montre la figure 7.1. Nous introduisons cet agent spécifique car nos agents sont en relation directe avec les utilisateurs ce qui nous permet de ne pas nuire aux utilisateurs (toute réputation propagée n'est pas dénonciatrice d'un utilisateur par un autre). Cet agent a pour rôle de recueillir les réputations propagées de chaque agent du HiMAS, de les compiler afin de ne diffuser qu'une seule valeur à chaque agent demandant une réputation propagée sur un consommateur donné (réception des réputations propagées des agents, compilation de ces dernières et envoie du résultat aux agents comme indiqué sur la figure 7.1). Les agents du HiMAS doivent donc communiquer les réputations qu'ils propagent à l'agent référent dès que leur valeur diminue. Ils doivent aussi demander une réputation propagée sur le consommateur engagé dans la transaction de données sensibles à l'agent référent avant chaque transaction de données sensibles afin de pouvoir compiler leur propre valeur de confiance. Cet agent référent permet ainsi une optimisation du nombre de messages envoyés pendant les expérimentations même s'il ne représente pas la meilleure solution dans l'absolu pour le respect de la sphère privée. Il permet donc d'optimiser les simulations au niveau de la mémoire et ainsi d'augmenter le nombre d'agents de la société.

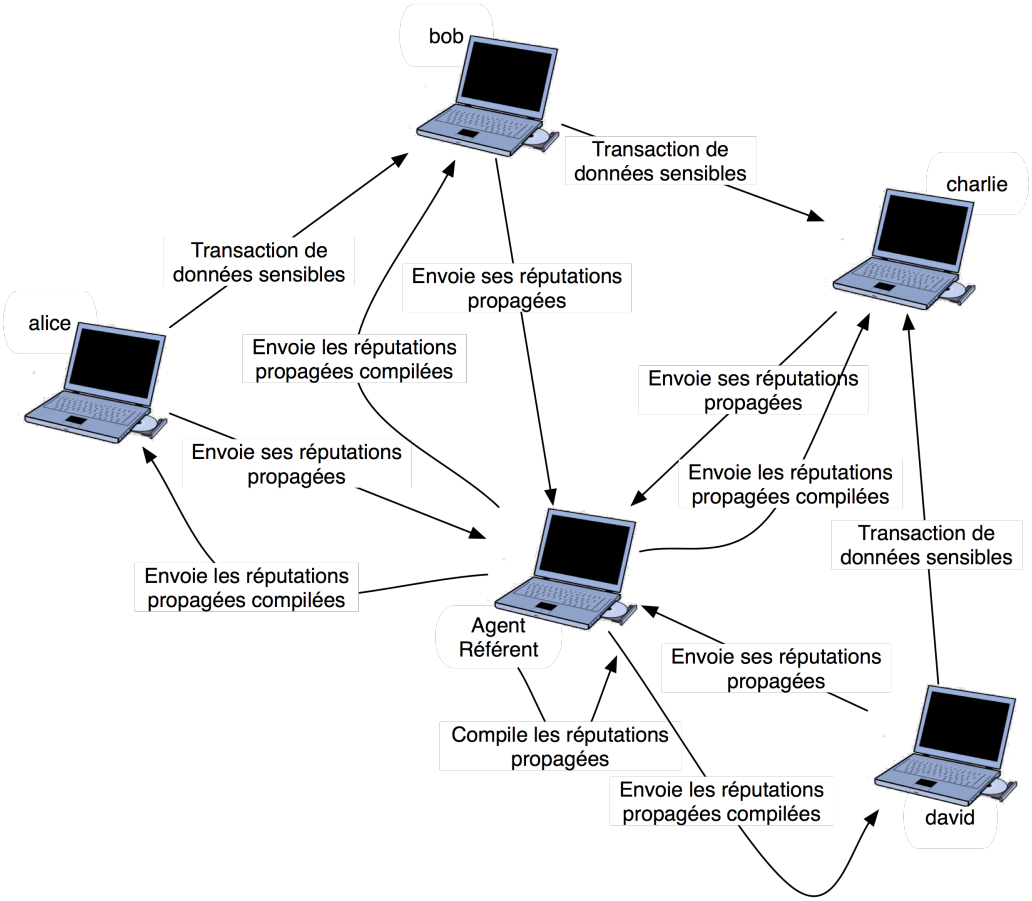

FIG. 7.1 – Scénario d'expérimentation.

7.1.2 Initialisation des agents

Pour chaque agent, son agenda est initialisé aléatoirement : le nombre de rendez-vous que contient son agenda ainsi que leurs paramètres (importance, urgence, date de début, date de fin, participants et objet) sont générés pour chaque expérimentation.

Lors de chaque transaction de données sensibles, les agents construisent leur politique et leur préférence également de manière aléatoire. Pour cela, ils choisissent dans le dictionnaire du domaine de la gestion d'agenda[1] soit toutes les valeurs possibles pour chaque élément de sa politique, soit un sous-ensemble de ces valeurs.

Nous initialisons aussi les croyances de réputation directe, $DoDR_{c,\Omega,f}$, et de réputation propagée, $DoPR_{c,\Omega,f}$, à leur valeur maximale (égale à 1) car le

[1]Ce dictionnaire du domaine est présenté plus en détails dans le chapitre suivant.

modèle de confiance que nous proposons est un modèle punitif visant à *forcer* les agents à bien se comporter vis-à-vis du respect de la sphère privée.

Les agents suspicieux ne respectent aucune facette (utilisation, diffusion, rétention, transparence) afin de faciliter leur détection. Dès qu'un agent détecte un agent suspicieux (valeur de la confiance en-dessous du seuil fixé), il ne communique plus de données sensibles avec lui.

Les partages de rendez-vous sont également engendrés de manière aléatoire : chaque agent demande son agenda ou l'agenda d'un tiers à un agent de sa liste de contact.

7.1.3 Paramètres d'expérimentation

Nous proposons de fonder notre évaluation sur [Joumaa *et al.* (2008, 2009)] qui définit l'évaluation de la performance des systèmes multi-agents comme relative aux interactions entre agents, et notamment celles représentant un envoi de messages. Dans un HiMAS, les envois de messages qui nous préoccupent sont les transactions de données sensibles, les résultats des expérimentions représentent une évaluation du contrôle social hippocratique relative au nombre de transactions de données sensibles terminées au sein du HiMAS.

Afin d'obtenir un nombre moyen de transactions de données sensibles par agent pour évaluer notre contrôle social hippocratique, chaque expérimentation a été exécutée 20 fois afin de s'assurer de la validité des résultats des simulations initialisées aléatoirement. Nous avons ensuite calculé la moyenne des résultats obtenus pour vérifier l'influence des différents paramètres que nous expérimentons.

Chaque simulation s'arrête par l'exclusion d'un agent suspicieux. Pour ce faire, nous appliquons le principe démocratique du tiers d'exclusion : dès qu'un tiers de la société d'agents considère un agent donné comme suspicieux, ce dernier est exclu du système car dès lors aucun agent n'interagit avec lui.

Nous avons expérimenté dans un premier temps les valeurs possibles des paramètres du modèle de confiance fixées par l'utilisateur afin de pouvoir raffiner notre modèle de confiance hippocratique. Ces paramètres sont la valeur du seuil de la fonction de confiance, la valeur de la punition encourue en cas de violation d'un engagement social hippocratique et la valeur des réputations stéréotypées ($DoSR_{c,\Omega,f}$).

Nous évaluons ensuite l'efficacité et les limites du contrôle social hippocratique qui formalise le principe de conformité au sein des HiMAS selon la typologie du réseau d'agents et le nombre d'agents suspicieux de la société d'agents.

7.2 Evaluation des paramètres du modèle de confiance

Afin de proposer un modèle de confiance pertinent pour un contrôle social hippocratique, nous réalisons une étude empirique afin de déterminer les meilleures valeurs des paramètres de notre système. Les paramètres étudiés sont :

- la valeur du seuil au-delà duquel les agents n'établissent plus de relation de confiance avec un autre agent suspicieux,
- la valeur de la punition appliquée aux facettes à chaque détection d'un comportement suspicieux,
- la valeur des réputations stéréotypées.

Afin d'obtenir un temps d'exécution convenable, le HiMAS mis en œuvre dans chaque expérimentation comporte 50 agents dont un seul est suspicieux. Le réseau d'agents utilisé est de type réseau social où chaque agent possède une liste de contacts construite sans hiérarchie et sans contrainte de dépendance entre les agents.

7.2.1 Seuil et punition de la fonction de confiance

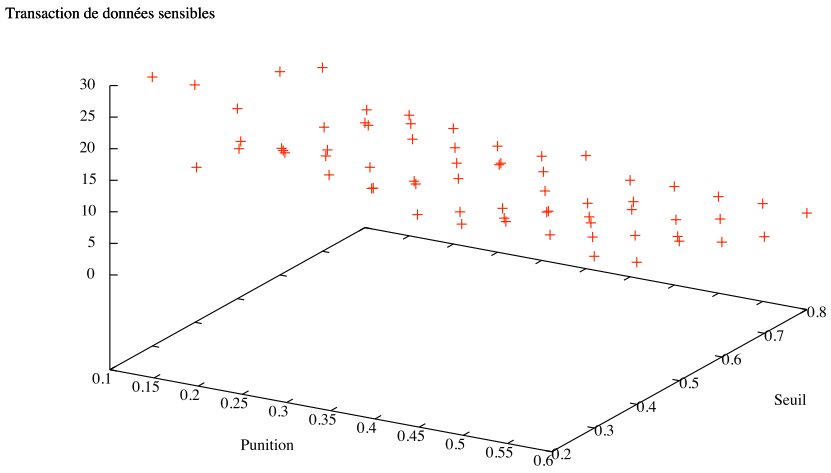

FIG. 7.2 – Nombre moyen de transactions de données sensibles par agent pour détecter un comportement suspicieux en fonction de la valeur de la punition appliquée et de la valeur du seuil de la fonction de confiance.

Les valeurs des réputations appartenant à l'intervalle [0,1], le seuil et la punition de la fonction de confiance doivent donc appartenir à l'intervalle]0,1[. Nous testons des valeurs du seuil comprises entre 0,3 et 0,8, variant par palier de 0,1. En ce qui concerne la punition, les valeurs expérimentées vont de 0,1 à 0,6 par palier de 0,05.

La figure 7.2 représente la variation du nombre moyen de transactions de données sensibles par agent en fonction de la valeur de la punition et du seuil de la fonction de confiance. La figure 7.3, quant à elle, représente la variation du nombre moyen de transactions de données sensibles par agent suspicieux en fonction des mêmes paramètres.

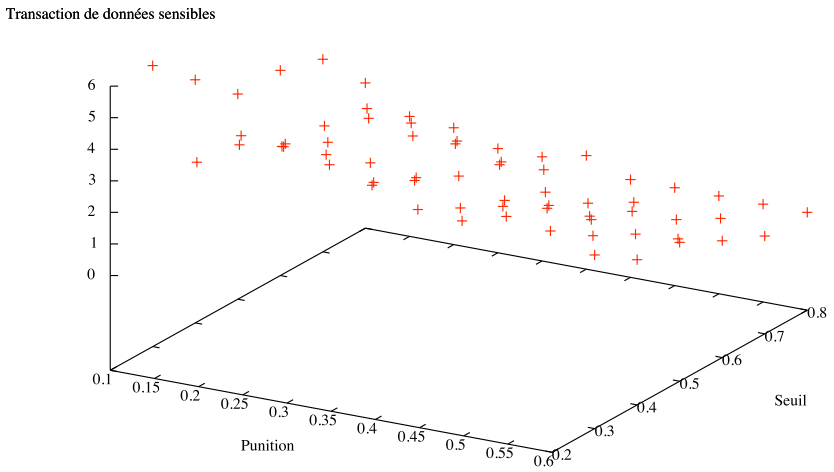

FIG. 7.3 – Nombre moyen de transactions de données sensibles par agent suspicieux pour détecter un comportement suspicieux en fonction de la valeur de la punition appliquée et de la valeur du seuil de la fonction de confiance.

Sachant que les agents peuvent violer un engagement social hippocratique sans pourtant être suspicieux, comme par exemple pour débloquer une situation importante pour la réalisation de leurs tâches, nous ne pouvons pas opter pour des valeurs permettant d'exclure un agent estimé comme suspicieux seulement après 1 transaction de données sensibles. De ce fait, comme le montre la figure 7.3, la valeur du seuil ne doit pas excéder 0,6 pour une valeur de la punition comprise entre 0,4 et 0,6.

En effet, par exemple, lorsque le seuil est fixé à 0,8 (ou 0,6) et la punition à 0,6 (ou 0,7), les agents effectuent en moyenne 0,3 transactions de données sensibles dont 0,1 avec l'agent suspicieux, ce qui indique que l'ensemble des

agents n'a pas pu effectuer au moins une transaction de données sensibles et qu'à la première violation détectée, l'agent dit suspicieux est exclu. A l'inverse, si on choisit un seuil et une punition peu élevés (respectivement 0,1 et 0,3), les agents effectuent en moyenne 27,6 transactions de données sensibles dont 5,9 avec l'agent suspicieux ce qui permet à l'agent suspicieux de récolter trop de données sensibles avant son exclusion. Cependant, avec un seuil fixé à 0,5 et une punition fixée à 0,15, les résultats obtenus nous fournissent un compromis acceptable avec une moyenne de 10,1 transactions de données par agent dont 2,1 avec l'agent suspicieux.

En dehors de ces valeurs, nous constatons qu'une moyenne de 3 transactions de données sensibles est nécessaire pour détecter un agent suspicieux par l'ensemble des agents. Cette moyenne diminue avec l'augmentation de la valeur du seuil et de la punition de la fonction de confiance.

Afin de ne pas exclure trop vite un agent suspicieux, car aucune relation de confiance ne peut être instaurée de nouveau, nous choisissons, pour la suite de nos expérimentations, de définir un seuil égal à 0,5 et une punition égale à 0,15.

7.2.2 Influence des réputations stéréotypées

Le troisième paramètre, dont nous souhaitons évaluer l'influence sur le contrôle social hippocratique, concerne les initialisations des réputations stéréotypées. En effet, ce type de réputation est le seul à pouvoir être initialisé à une valeur autre que 1 et qui n'est pas directement influencé par les transactions de données sensibles entre agents. Nous proposons donc d'effectuer plusieurs expérimentations où l'ensemble des 50 agents auront une valeur prédéfinie de réputation stéréotypée comprise entre 0,1 et 0,9, évoluant par palier de 0,1.

Nous rappelons que suite aux résultats obtenus dans la sous-section précédente, nous choisissons de définir un seuil de la fonction de confiance égal à 0,5 et une valeur de punition de 0,15.

La figure 7.4 présente la variation du nombre moyen de transactions de données sensibles par agent, et par agent suspicieux, en fonction de la valeur initiale de la réputation stéréotypée, pour détecter un comportement suspicieux.

Comme le montre la figure 7.4, la valeur de la réputation stéréotypée n'influence pas le nombre moyen de transactions de données sensibles par agent et par agent suspicieux. En effet, si la valeur de la réputation stéréotypée est fixée à 0,1, les agents effectuent en moyenne 15,8 transactions de données sensibles dont 3 avec l'agent suspicieux et avec une réputation stéréotypée fixée à 0,9, les agents effectuent en moyenne 15,2 transactions de données sensibles

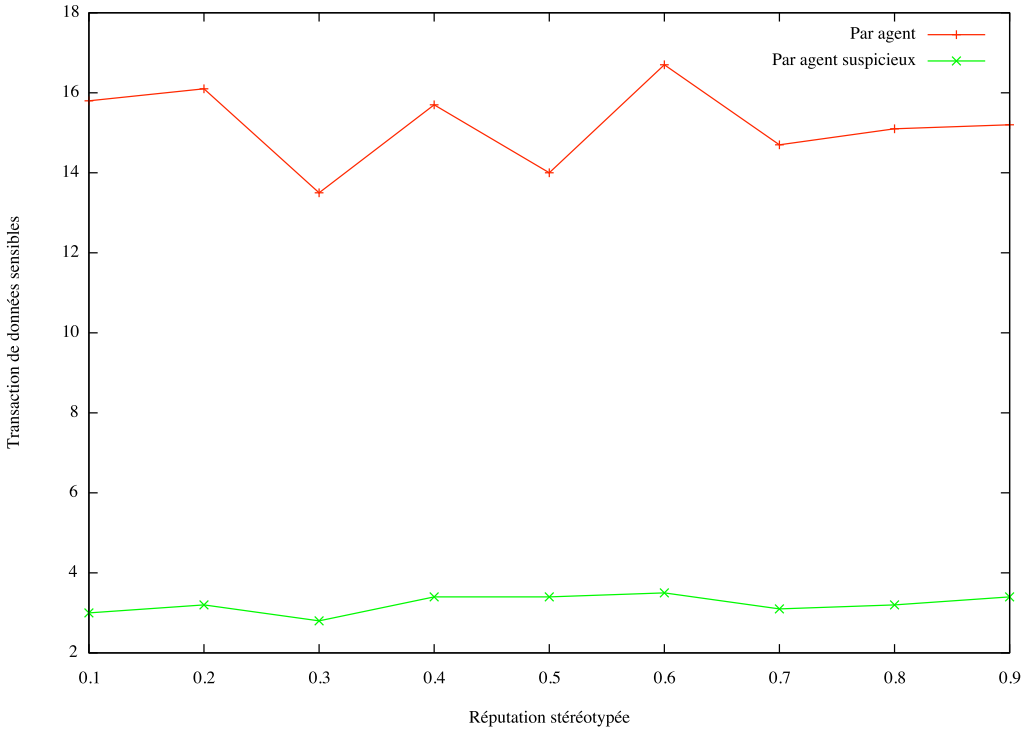

FIG. 7.4 – Nombre moyen de transactions de données sensibles pour détecter un comportement suspicieux en fonction de la valeur de la réputation stéréotypée.

dont 3,4 avec l'agent suspicieux, ce qui est équivalent en termes de transactions de données sensibles.

7.3 Evaluation du contrôle social hippocratique

Nous jugeons dans cette section de l'efficacité des modèles de confiance pour le respect de la sphère privée sans prendre en compte les performances du modèle en lui-même. Pour évaluer le contrôle social hippocratique que nous proposons, nos expérimentations sont paramétrées selon : le nombre d'agents, le nombre d'agents suspicieux et le type du réseau d'agents.

7.3.1 Selon la typologie du réseau d'agents

La première série d'expérimentations se focalise sur le type de réseau d'agents. Nous expérimentons notre contrôle social hippocratique avec un ensemble de 10, 50, 100 ou 150 agents, dont un suspicieux, selon trois types de réseaux d'agents : un réseau dit social, un réseau en arbre et un réseau en couche. Ces trois types de réseaux sont illustrés par la figure 7.5.

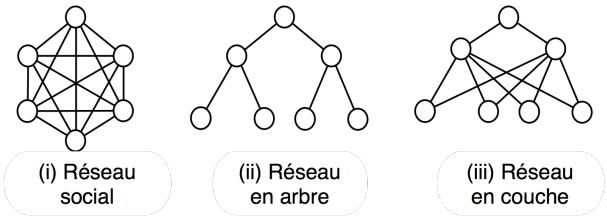

FIG. 7.5 – Typologie des réseaux testés dans les expérimentations.

Dans un réseau social ((i) sur la figure 7.5), les agents peuvent interagir avec les autres sans aucune contrainte ou dépendance entre eux. Dans un réseau en arbre ((ii) sur la figure 7.5), les interactions entre agents sont régies par une contrainte de position dans le réseau. En effet, un agent ne peut interagir qu'avec l'agent se trouvant au-dessus ou au-dessous de lui dans l'arbre. Dans un réseau en couche ((iii) sur la figure 7.5), les agents ne peuvent interagir qu'avec les agents de la couche supérieure ou de la couche inférieure de l'arbre.

Les résultats de ces expérimentations sont présentés sur les figures 7.6 et 7.7.

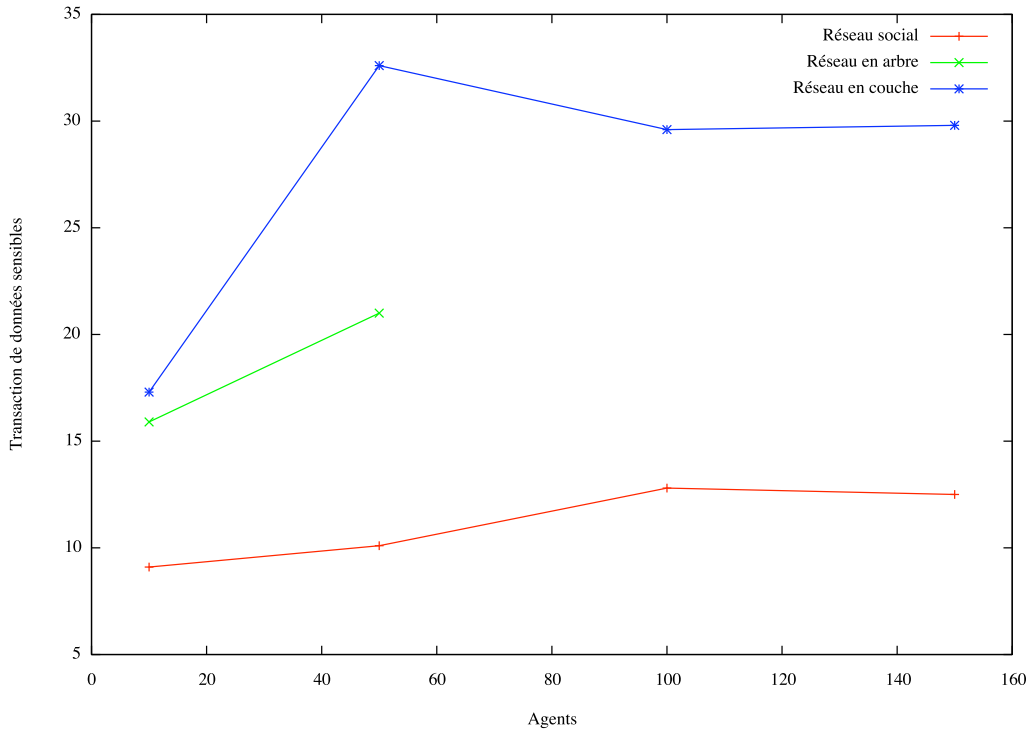

FIG. 7.6 – Nombre moyen de transactions de données sensibles requises avec l'agent suspicieux pour détecter un comportement suspicieux en fonction du type du réseau et du nombre d'agents.

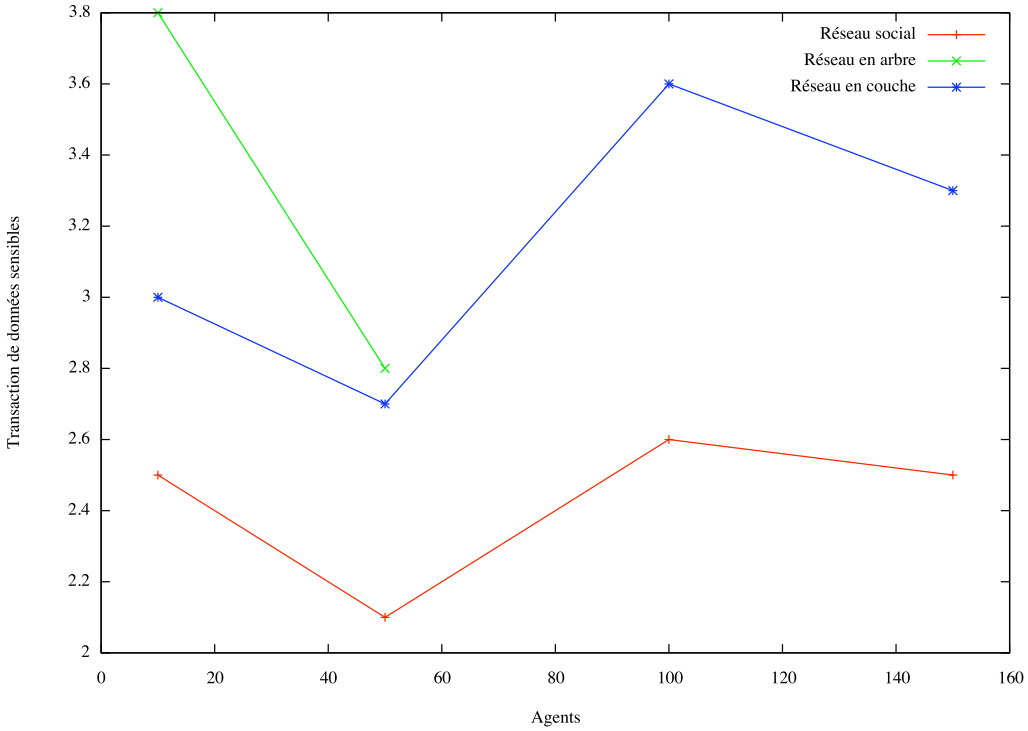

FIG. 7.7 – Nombre moyen de transactions de données sensibles requises avec l'agent suspicieux pour détecter un comportement suspicieux en fonction du type du réseau et du nombre d'agents.

7.3.1.1 Réseau social

Pour les réseaux de type social, la détection d'un agent suspicieux est très peu influencée par le nombre d'agents de la société. En effet, même si le nombre de transactions de données croît avec le nombre d'agents (figures 7.6 et 7.7), ceux-ci ont besoin d'effectuer en moyenne entre 9 et 13 transactions de données sensibles chacun, dont environ 2 avec l'agent suspicieux, avant de le bannir. Par exemple, dans une société de 10 agents, les agents effectuent en moyenne 9,1 transactions de données sensibles dont 2,5 avec l'agent suspicieux et dans une société de 100 agents, les agents effectuent en moyenne 12,8 transactions de données sensibles dont 2,6 avec l'agent suspicieux.

7.3.1.2 Réseau en arbre

Les expérimentations avec un réseau de type arbre n'ont pas pu excéder la simulation d'une société de 50 agents du fait de la taille mémoire requise (grand nombre de messages et leurs traces).

Comme le montrent les figures figures 7.6 et 7.7, dans une société de 50 agents, les agents effectuent en moyenne 15,9 transactions de données sensibles dont 3,8 avec l'agent suspicieux et dans une société de 100 agents, les agents effectuent en moyenne 21 transactions de données sensibles dont 2,8 avec l'agent suspicieux.

Cependant, même si le nombre de transactions de données sensibles augmente (voir figures 7.6 et 7.7), les agents doivent effectuer en moyenne 3 transactions de données sensibles avec l'agent suspicieux pour le détecter et le bannir, soit 1 transaction de données sensibles de plus que pour les réseaux sociaux, ce qui est une différence insignifiante.

Nous pouvons aussi remarquer que, dans ce type de réseau, l'exclusion de l'agent suspicieux scinde le réseau en plusieurs sous-réseaux indépendants du point vue de leurs communications comme l'illustre la figure 7.8.

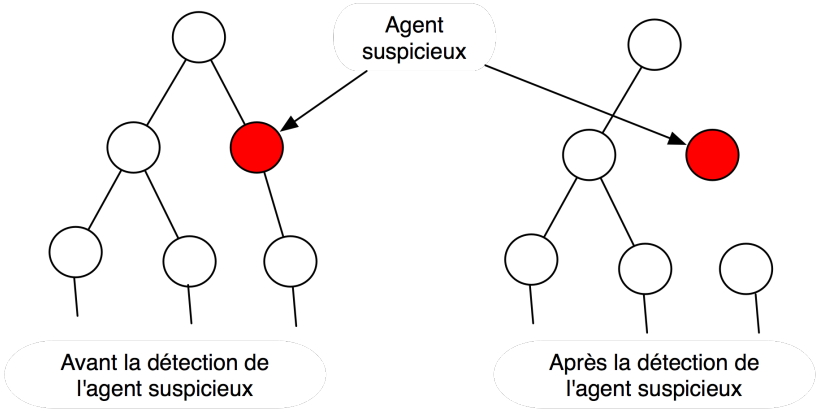

FIG. 7.8 – Exclusion d'un agent suspicieux dans un réseau en arbre.

7.3.1.3 Réseau en couche

Dans les réseaux en couche, le nombre de transactions de données sensibles est plus élevé que dans les réseaux sociaux (figures 7.6 et 7.7). Ceci est principalement dû au type du réseau qui impose plus souvent aux agents, pour connaître l'agenda d'un autre, de passer par un tiers. Cependant, nous pouvons constater que les agents ont besoin d'effectuer en moyenne 3 transactions de données sensibles avec l'agent suspicieux pour le détecter et le bannir. Le contrôle social hippocratique est donc aussi performant dans les réseaux en couche que dans les réseaux sociaux.

Par exemple, comme le montrent les figures 7.6 et 7.7, dans une société de 50 agents, les agents effectuent en moyenne 32,6 transactions de données sensibles dont 2,7 avec l'agent suspicieux alors que dans une société de 100

agents, les agents effectuent en moyenne 29,6 transactions de données sensibles dont 3,6 avec l'agent suspicieux.

7.3.2 Selon le nombre d'agents suspicieux

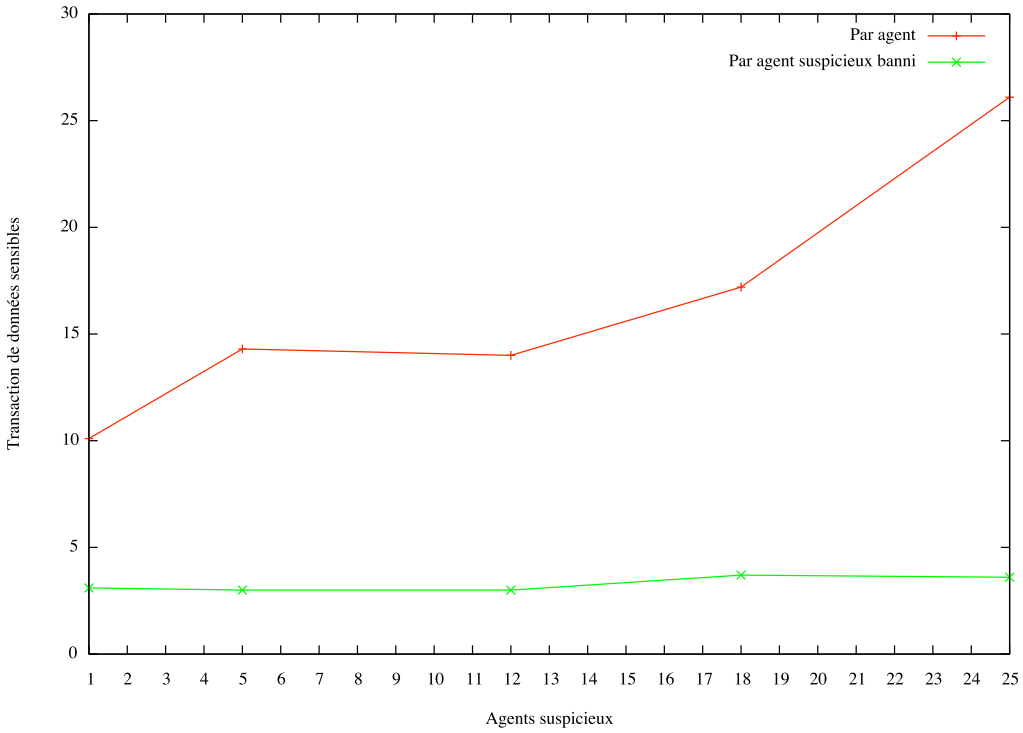

FIG. 7.9 – Nombre de transactions de données sensibles requises pour détecter un comportement suspicieux en fonction du nombre d'agents suspicieux dans un HiMAS de 50 agents.

La deuxième phase de notre validation expérimentale concerne l'influence du nombre d'agents suspicieux sur le nombre de transactions de données sensibles dans un système multi-agent hippocratique. Dans un système de 50 agents, en réseau social, nous introduisons 1, 5, 12, 18 et 25 agents au comportement suspicieux. Ensuite, dans un HiMAS de 150 agents, nous introduisons 1, 5, 10, 25, 50 et 75 agents suspicieux. Nous étudions ainsi les limites du contrôle social hippocratique en fonction du nombre d'agents suspicieux introduit dans la société d'agents.

Comme le montre les figures 7.9 et 7.10, l'augmentation du nombre d'agents suspicieux dans la société d'agents influence très légèrement leur détection. En effet, le nombre de transactions de données sensibles avec le premier agent suspicieux exclu ne varie que légèrement dans une limite de 33% d'agents suspicieux (soit 18 agents suspicieux pour un HiMAS de 50 agents et 50 agents

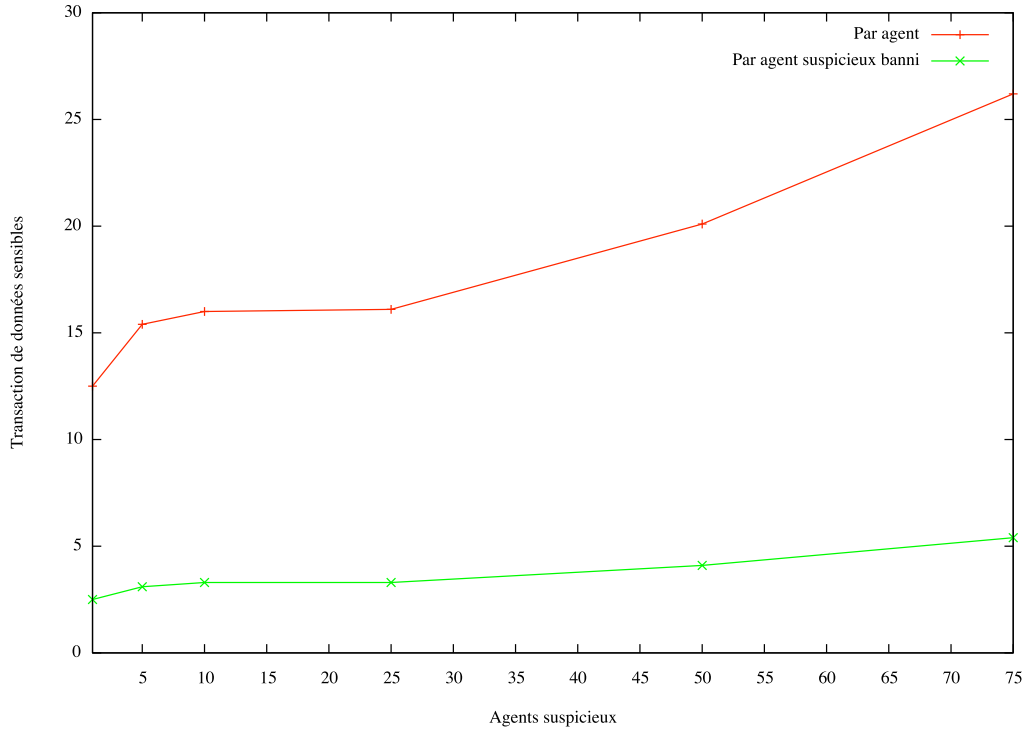

FIG. 7.10 – Nombre de transactions de données sensibles requises pour détecter un comportement suspicieux en fonction du nombre d'agents suspicieux dans un HiMAS de 150 agents.

pour un HiMAS de 150 agents), malgré une forte augmentation du nombre total de transactions de données sensibles. Au-delà de ce pourcentage, nous pouvons remarquer que le contrôle social hippocratique devient de moins en moins performant avec l'ajout d'autres agents suspicieux. Notons que ces résultats nous permettent également d'avancer l'hypothèse que le nombre d'agents du HiMAS influence peu le contrôle social hippocratique.

En effet, dans une société de 50 agents (figure 7.9), lorsqu'un seul agent suspicieux est introduit, les agents effectuent en moyenne 10,1 transactions de données sensibles dont 3,1 avec l'agent suspicieux et lorsque 75 agents suspicieux sont introduits, la moyenne des transactions de données sensibles est de 26,1 par agent dont 3,6 avec l'agent suspicieux exclu.

Dans une société de 150 agents (figure 7.10), avec l'introduction d'un seul agent suspicieux, les agents effectuent en moyenne 12,5 transaction de données sensibles dont 2,5 avec l'agent suspicieux alors que lorsque 25 agents suspicieux sont introduits dans la société, la moyenne des transactions de données sensibles par agent monte à 26,2 dont 5,4 avec l'agent suspicieux banni.

7.4 Synthèse

Nous pouvons avancer l'hypothèse que le nombre d'agents et le type du réseau ont très peu d'influence sur la détection d'un agent suspicieux. En effet, selon nos différentes expérimentations, les agents ont besoin d'effectuer en moyenne 3 transactions de données sensibles pour détecter et bannir un agent suspicieux avec un seuil de 0,5 et une punition de 0,15 pour la fonction de confiance.

Nous avons pu aussi constater que le pourcentage d'agents suspicieux dans la société d'agents n'influence que légèrement l'efficacité de la détection d'un agent suspicieux.

Maintenant que nous avons évalué notre modèle d'un point de vue expérimental, nous présentons dans le prochain chapitre la mise en œuvre d'un HiMAS par la migration d'une application concrète de gestion décentralisée d'agenda en un système multi-agent hippocratique.

Chapitre 8

Migration d'un système multi-agent vers un HiMAS

Sommaire

Dans le but d'illustrer d'un point de vue opérationnel notre modèle de HiMAS, nous avons étendu une application multi-agent de gestion d'agendas décentralisés en un système multi-agent hippocratique. Dans cette application, chaque utilisateur est représenté par un agent logiciel qui prend en charge la gestion de son agenda. La spécificité de cette application fait que les agents

[0]Une version initiale des travaux réalisés dans ce chapitre a été publiée dans [Crépin *et al.* (2009)].

échangent fréquemment des messages contenant des données sensibles des utilisateurs ce qui nous offre un terrain propice à notre modèle.

Nous commençons par décrire l'application agenda dans la première section. La deuxième section est consacrée au niveau individuel des HiMAS, ou plus précisément à la création de la sphère privée des agents en fonction des utilisateurs. Pour finir, nous présentons l'implémentation du niveau social des HiMAS : le protocole de transaction de données sensibles et le contrôle social hippocratique.

8.1 Application AGENDA

[Demazeau *et al.* (2006)] propose une application multi-agent de gestion d'agendas décentralisés. L'architecture de cette application permet à chaque agent de gérer le calendrier d'un utilisateur (figure 8.1). Chaque utilisateur délègue en effet l'ensemble de ses rendez-vous à un agent autonome qui a pour objectif de l'assister dans la gestion de son agenda en termes de prise de rendez-vous et d'organisation de ces derniers.

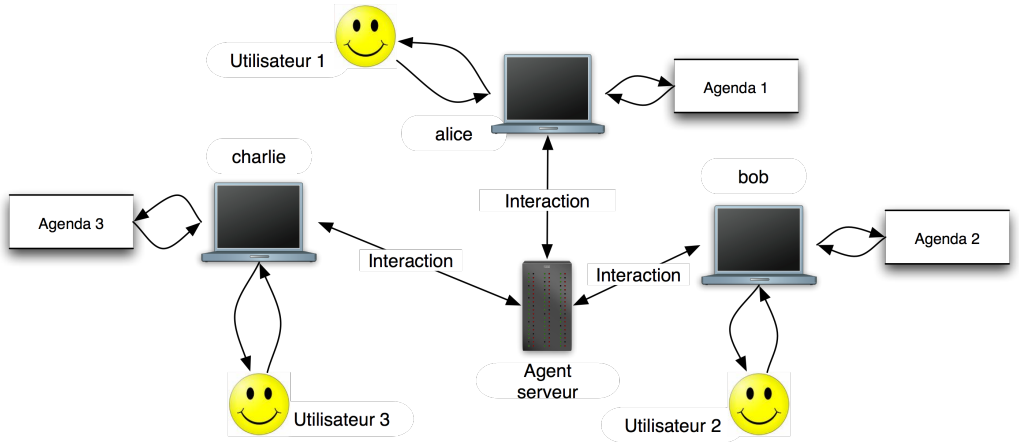

FIG. 8.1 – Architecture de l'application de gestion d'agendas distribués.

L'application utilise un agent spécifique, l'agent serveur (cf. figure 8.1), qui a pour rôle d'inscrire chaque nouvel agent et de mettre en relation l'ensemble des agents du système en gérant les envois de messages.

L'envoi de messages utilise le protocole de communication Jabber [Jabber Software Foundation (2001)]. Ce protocole permet d'assurer la sécurité et la confidentialité des communications ainsi que la décentralisation de l'applica-

tion. L'interface homme-machine est développée grâce à MiG Calendar[1] qui permet la visualisation des calendriers des utilisateurs.

8.1.1 Prise de rendez-vous

Pour prendre un nouveau rendez-vous, AGENDA propose deux fonctionnalités à l'utilisateur : (i) une négociation semi-automatique et (ii) le partage d'agenda.

La première fonctionnalité, la négociation de rendez-vous, est assurée par le système GeNCA [Mathieu et Verrons (2005)] qui permet une négociation en temps fini de contrats représentant le rendez-vous à prendre. L'utilisateur choisit ses préférences en termes de négociation selon les caractéristiques des rendez-vous et selon ses besoins en termes de gestion d'agenda. Il indique ensuite à son agent assistant la nouvelle réunion à fixer avec l'ensemble de ses paramètres (participants, sujet, date de début, date de fin, urgence et importance).

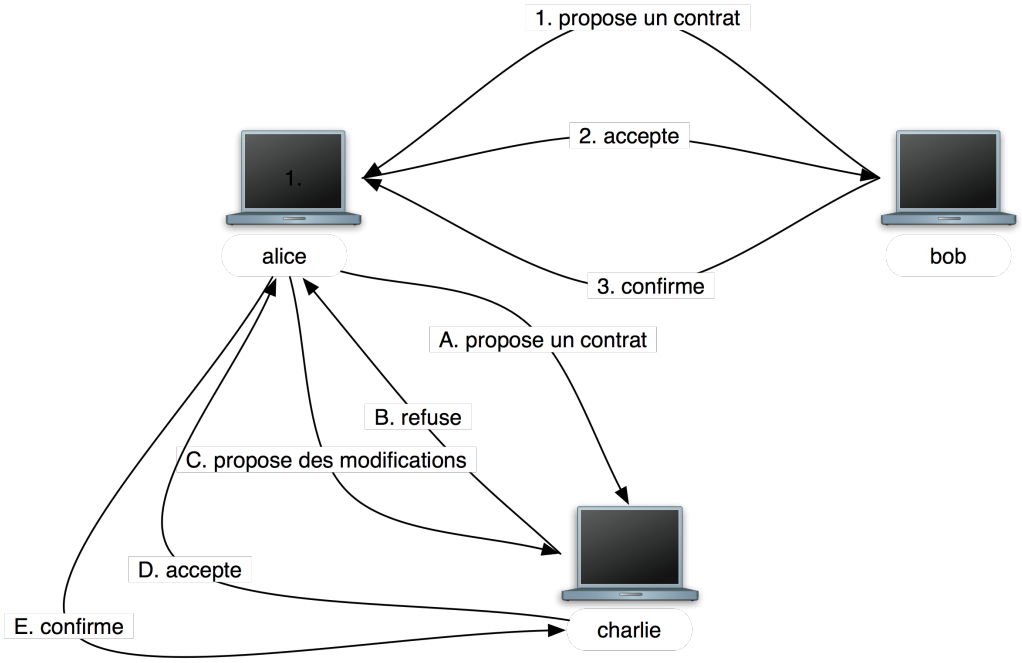

FIG. 8.2 – Illustration de la négociation sur un exemple.

Commence alors la négociation dont les actions possibles sont représentées dans la figure 8.2. L'agent prend en charge la négociation de cette réunion en envoyant une proposition de contrat aux autres agents concernés. Ces agents

[1]http ://migcalendar.com/index.php

décident alors d'une date commune si elle existe, proposent des modifications sur le contrat, s'ils en ont la possibilité, ou annulent la prise de rendez-vous le cas échéant, comme le montre la figure 8.2.

La deuxième possibilité pour prendre un rendez-vous est de consulter les agendas des autres utilisateurs afin de connaître leurs disponibilités avant de lancer une négociation. Cette tâche est possible grâce au partage d'agendas. Lorsqu'une relation de confiance est instaurée entre agents[2], les agents s'adressent leur agenda sur simple demande. Le cas échéant, si aucune relation de confiance n'est déclarée, le partage de l'agenda échoue.

L'extension d'AGENDA que nous proposons dans ce chapitre se focalise sur ce partage d'agendas en y introduisant le protocole de transaction de données sensibles et le contrôle social hippocratique.

8.1.2 Modélisation des rendez-vous

Chaque agenda est constitué de créneaux horaires qui peuvent être libres ou occupés par un rendez-vous lui-même caractérisé par un nom, une catégorie, les participants, une date de début, une date de fin, un niveau d'urgence et un niveau d'importance, comme le montre la figure 8.3.

Un utilisateur gérant plusieurs facettes de sa personnalité pour son emploi du temps, un rendez-vous appartient également à une catégorie[3] dont les deux principales, dans notre exemple (figure 8.3), sont *professionnel*, décomposée en *équipe*, *laboratoire* et *projet européen*, et *personnel*, décomposée en *famille* et *association*.

Afin d'organiser les rendez-vous, AGENDA propose la possibilité d'affecter une caractéristique d'*importance* et une caractéristique d'*urgence* pour chaque rendez-vous, ce qui permet de donner un ordre de priorité entre les rendez-vous au sein de l'agenda. L'importance est généralement rattachée à un degré de force ou d'intérêt, et l'urgence à une situation qui ne doit pas être traitée avec du retard (dictionnaire Larousse).

Ces deux caractéristiques sont subjectives : les utilisateurs choisissent l'ordre de priorité de leur événement en fonction de leur propre définition de l'importance et de l'urgence comme le montre le tableau 8.1. Par exemple, un utilisateur peut décider que l'urgence prime sur l'importance alors qu'un autre peut décider de l'inverse. Seuls les rendez-vous urgents et importants restent les plus prioritaires et ceux non importants et non urgents les moins prioritaires quel que soit l'utilisateur, comme le résume le tableau 8.1. Lors de

[2]Initialement, cette relation de confiance est binaire et l'utilisateur décide entièrement de sa valeur (vrai ou faux).

[3]Une catégorie peut être assimilée à la notion de groupe.

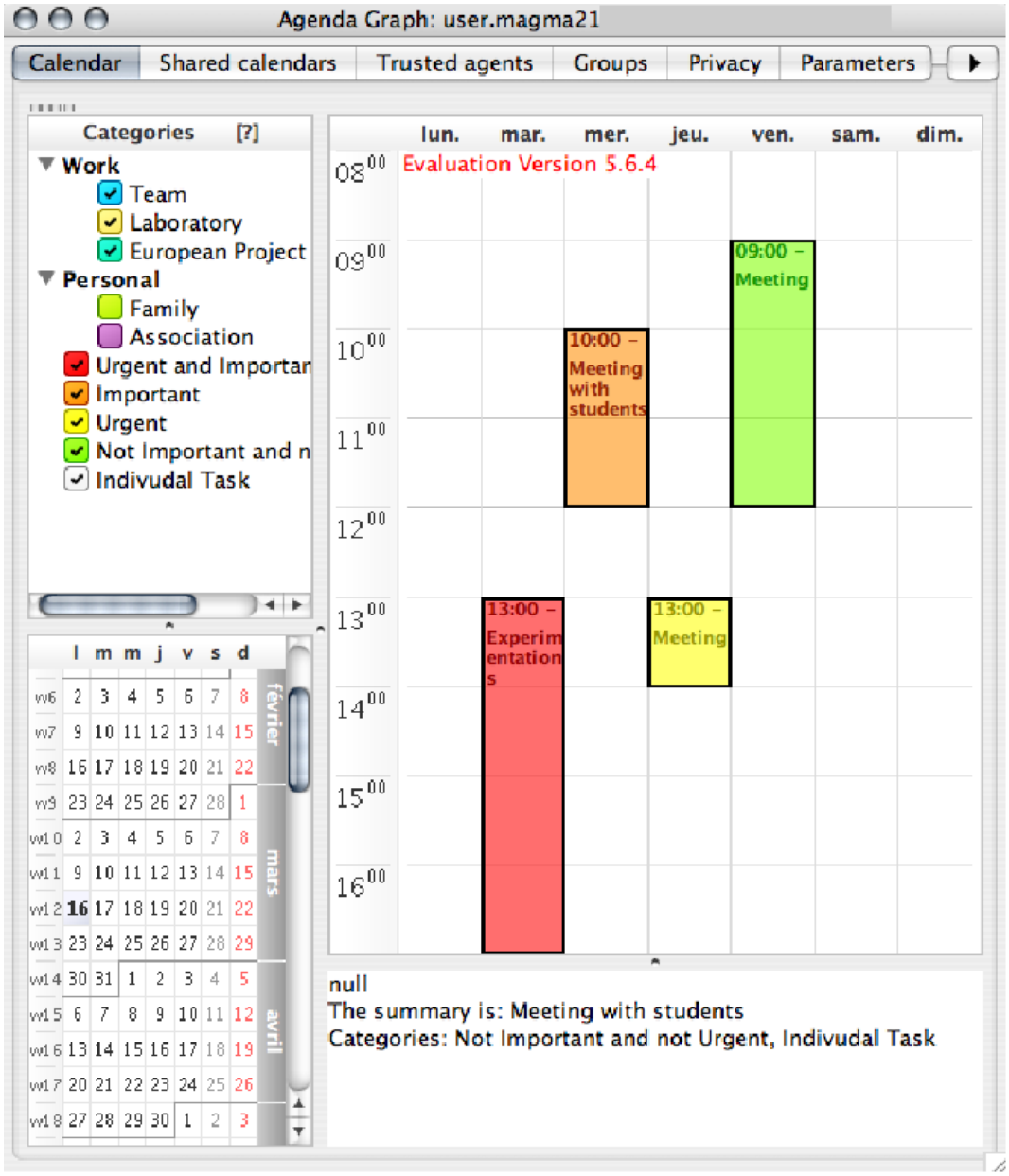

FIG. 8.3 – Interface homme-machine de l'application AGENDA.

la négociation, les agents se fondent sur ces deux paramètres pour définir leur stratégie[4].

Ordre de priorité	Important	Non Important
Urgent	1	2 ou 3
Non Urgent	2 ou 3	4

TAB. 8.1 – Priorité des rendez-vous selon l'importance et l'urgence.

Nous proposons d'étendre AGENDA pour y intégrer le protocole de transaction de données sensibles et le contrôle social hippocratique détaillés respectivement dans le chapitres 5 et le chapitre 6. Pour cela, nous considérons les rendez-vous des agendas des utilisateurs ainsi que l'ensemble de leurs paramètres comme étant des données sensibles. Commençons par présenter l'intégration de la sphère privée des agents au sein d'AGENDA.

8.2 Sphère privée des agents et utilisateurs

Pour qu'un agent puisse se représenter, construire et gérer sa sphère privée en fonction des souhaits de l'utilisateur qu'il représente, la première étape de la migration d'AGENDA vers un HiMAS concerne la délégation de l'agenda de l'utilisateur à son agent assistant par la création d'un profil spécifique à la préservation de la sphère privée, comme proposé dans le chapitre 4.

Rappelons que la création du profil des utilisateurs est constituée de trois phases. Dans un premier temps, l'utilisateur indique à son agent logiciel assistant quelles sont les données qu'il estime être sensibles. Ainsi chaque nouveau rendez-vous considéré comme une donnée sensible par un utilisateur est directement intégré dans la sphère privée de son agent. L'utilisateur personnalise ensuite la fonction de confiance pour le contrôle social hippocratique. Pour finir, afin que les agents respectent les souhaits des utilisateurs, ces derniers indiquent les règles de gestion de la sphère privée des agents.

8.2.1 Agendas et rendez-vous

Afin que chaque utilisateur délègue à un agent logiciel autonome ses données sensibles relatives à son agenda, nous avons choisi de représenter ces données par des concepts hiérarchisés (cf. figure 4.7 du chapitre 4, section 4.3.1). La création du dictionnaire du domaine pour l'application AGENDA a été effectuée en fonction des différentes catégories de rendez-vous que l'application propose : les rendez-vous personnels, dont familiaux et associatifs, et les

[4]Pour de plus amples informations sur cette partie de l'application, nous renvoyons le lecteur à [Demazeau *et al.* (2006)]

rendez-vous professionnels dont ceux avec l'équipe, avec le laboratoire et pour les projets européens.

Les agents gèrent les données sensibles des utilisateurs grâce à un ensemble de règles définissant des autorisations selon un ensemble de conditions données. Les conditions portent sur les objectifs de la transaction de données sensibles, ceux-ci étant dépendants des catégories des autres agents. Nous proposons donc aux utilisateurs de choisir quels types de rendez-vous ils estiment être sensibles en fonction de la catégorie, puis de définir leurs contraintes de partage d'agendas en fonction de l'objectif de la transaction de données sensibles.

Exemple 8.1

Par exemple, un utilisateur peut estimer que ses rendez-vous d'ordre personnel sont sensibles si l'objectif de la demande est en rapport avec son milieu professionnel alors que si l'agent appartient à son domaine familial, ce type de rendez-vous n'est pas sensible.

Chaque donnée sensible est composée d'un ensemble de références (cf. définition 4.2, chapitre 4, section 4.1.1). Dans le cas d'AGENDA, les références relatives aux rendez-vous sont les paramètres des rendez-vous : l'objet, les participants, le niveau d'urgence et le niveau d'importance. Le choix de la transmission ou non de ces références par l'utilisateur est géré par la création des règles de gestion de la sphère privée.

FIG. 8.4 – Données sensibles.

Le profil de chaque utilisateur pour la création de la sphère privée des agents relative aux agendas est donc composé de sept données sensibles pour chaque catégorie. Par exemple, si un utilisateur coche *Family* sur la ligne *Work* (cf. figure 8.4) l'agent intègre à sa sphère privée tous les rendez-vous familiaux lorsque que le contexte de la transaction de données sensibles a comme contexte le milieu professionel. Les objectifs pour les transactions de données sont pris en compte lors de la délégation des éléments des règles de gestion de la sphère privée pour la création des politiques et des préférences.

8.2.2 Paramètres de la fonction de confiance

Comme le montre le graphe conceptuel de la figure 4.6 (chapitre 4, section 4.3.1), chaque réputation est considérée comme une donnée sensible qui doit être transmise en respectant le protocole spécifique de transaction de données sensibles présenté dans le chapitre 5.

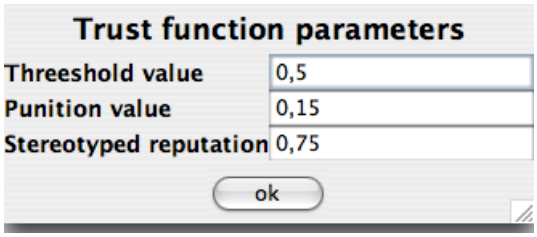

FIG. 8.5 – Personnalisation des paramètres de la fonction de confiance.

Les seules actions que l'utilisateur peut réaliser sur le contrôle social hippocratique, et donc sur les réputations, concernent le choix de la valeur de la punition et du seuil de la fonction pour établir une relation de confiance. De plus, afin de personnaliser cette fonction, nous avons introduit les réputations stéréotypées dans le calcul de la valeur de confiance. Ces valeurs sont saisies par l'intermédiaire d'une fenêtre d'interaction comme le montre la figure 8.5.

Remarquons que la valeur des croyances de confiance vont évoluer au cours de l'exécution de l'application de deux manières possibles. La première est manuelle : l'utilisateur peut modifier les paramètres de la fonction de confiance au cours de l'exécution de l'application. La seconde est automatique. En effet, en appliquant le contrôle social hippocratique, les agents s'échangent des réputations propagées ce qui influence la valeur de la croyance de confiance qu'un agent possède pour un autre agent.

8.2.3 Règles, politiques et préférences

Les règles de gestion de la sphère privée permettent aux agents de définir leurs politiques et leurs préférences selon les souhaits des utilisateurs qu'ils assistent lors des transactions de données sensibles.

La diffusion d'une donnée sensible dépend des objectifs de la transaction. Afin d'obtenir un comportement non suspicieux, les agents sont soumis, au niveau des règles de gestion de leur sphère privée, aux limites imposées par le dictionnaire du domaine. L'agent propose donc à son utilisateur d'imposer plus de contraintes en termes de collection, d'utilisations possibles, de liste de diffusion et de durée de rétention en fonction des éléments contenus dans le dictionnaire du domaine, comme le montre la figure 8.6.

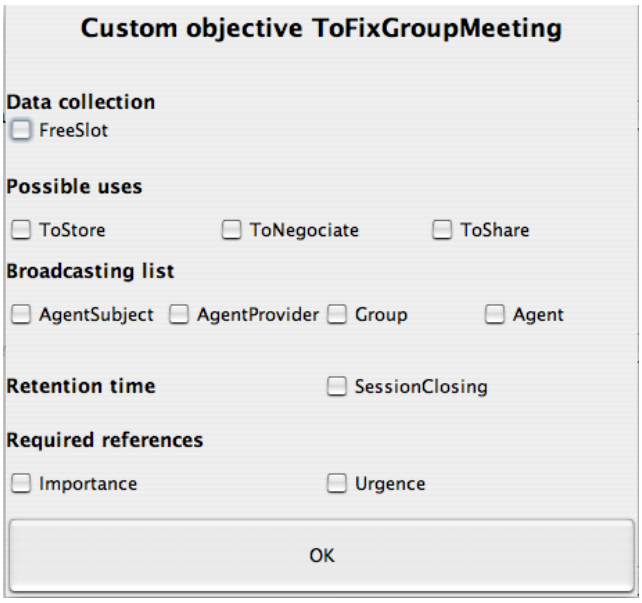

FIG. 8.6 – Interface pour la personnalisation des préférences.

Une fois que les agents peuvent se représenter leur sphère privée en fonction des utilisateurs, ils sont en mesure de communiquer des données sensibles sans violer leur sphère privée et en respectant les souhaits des utilisateurs.

8.3 Transaction et devenir des données sensibles

Afin que les agents puissent échanger les données sensibles de leur sphère privée, il nous faut d'abord interpréter le dictionnaire générique en fonction du domaine de la gestion de calendrier. Ensuite nous intégrons aux agents la possibilité de créer leurs politiques et préférences en fonction du profil utilisateur dans le but d'effectuer des transactions de données sensibles sans violation de la sphère privée.

8.3.1 Implémentation du dictionnaire générique

L'implémentation du dictionnaire générique se fait par le biais d'un Schéma RDF (RDFS) [W3C (2002d)] dans un document OWL [W3C (2004)]. Ce document est généré grâce à l'application Protégé [Stanford Center for Biomedical Informatics Research (1997 2009)]. RDFS est un langage de représentation de connaissance recommandé par le W3C permettant de structurer des documents RDF [W3C (2002c)]. Dans notre protocole, ces documents représentent le dictionnaire du domaine. Le langage RDFS permet de représenter

un ensemble de classes, de sous-classes et de propriétés entre classes (ou sous-classes), ce qui correspond au dictionnaire générique.

```
<RDF:RDF>
  ...
  <RDFS:Class RDF:ID="Liste-diffusion"/>
  <RDFS:Class RDF:ID="Objectif"/>
  <RDFS:Class RDF:ID="Donnee-collectee"/>
  <RDFS:Class RDF:ID="Duree-retention"/>
  <RDFS:Class RDF:ID="Utilisation-possible"/>
 <owl:TransitiveProperty RDF:ID="definit">
    <RDFS:domain RDF:resource="#Objectif"/>
    <RDF:type RDF:resource="owl#ObjectProperty"/>
    <RDFS:range>
      <owl:Class>
        <owl:unionOf RDF:parseType="Collecte">
          <RDFS:Class RDF:about="#Duree-retention"/>
          <RDFS:Class RDF:about="#Utilisation-possible"/>
          <RDFS:Class RDF:about="#Liste-diffusion"/>
        </owl:unionOf>
      </owl:Class>
    </RDFS:range>
  </owl:TransitiveProperty>
  ...
</RDF:RDF>
```

FIG. 8.7 – Exemple d'implémentation du dictionnaire générique avec RDFS.

La figure 8.7 présente un extrait de l'implantation de notre dictionnaire générique. Cette partie de fichier correspond à la définition des concepts liés aux principes de connaissance des objectifs, de diffusion limitée, de collection limitée, de rétention limitée et des liens sémantiques qui existent entre eux.

8.3.2 Interprétation et implémentation du dictionnaire du domaine

Afin de transformer AGENDA en un HiMAS, nous avons choisi de développer trois objectifs possibles pour une transaction de données sensibles dans le domaine de la gestion d'agenda :

1. *S'informer* : prendre connaissance de l'agenda d'un autre agent sans traitement des données sensibles ;
2. *Fixer un rendez-vous* : prendre connaissance de l'agenda d'un autre agent afin de prendre un rendez-vous avec ce dernier ;
3. *Fixer un rendez-vous de groupe* : prendre connaissance de l'agenda d'un autre agent afin de prendre rendez-vous avec ce dernier et un groupe d'agents.

Pour intégrer ces trois objectifs et donc permettre aux agents d'effectuer une transaction de données sensibles, nous interprétons et implémentons le

dictionnaire du domaine. Cette étape consiste en la création des ensembles maximaux des valeurs possibles pour chaque concept du dictionnaire tout en appliquant le respect de la sphère privée. A partir du dictionnaire du domaine obtenu et de l'intégration des préférences des utilisateurs, les agents sont capables de communiquer les données sensibles qu'ils connaissent sans violer la sphère privée.

8.3.2.1 S'informer

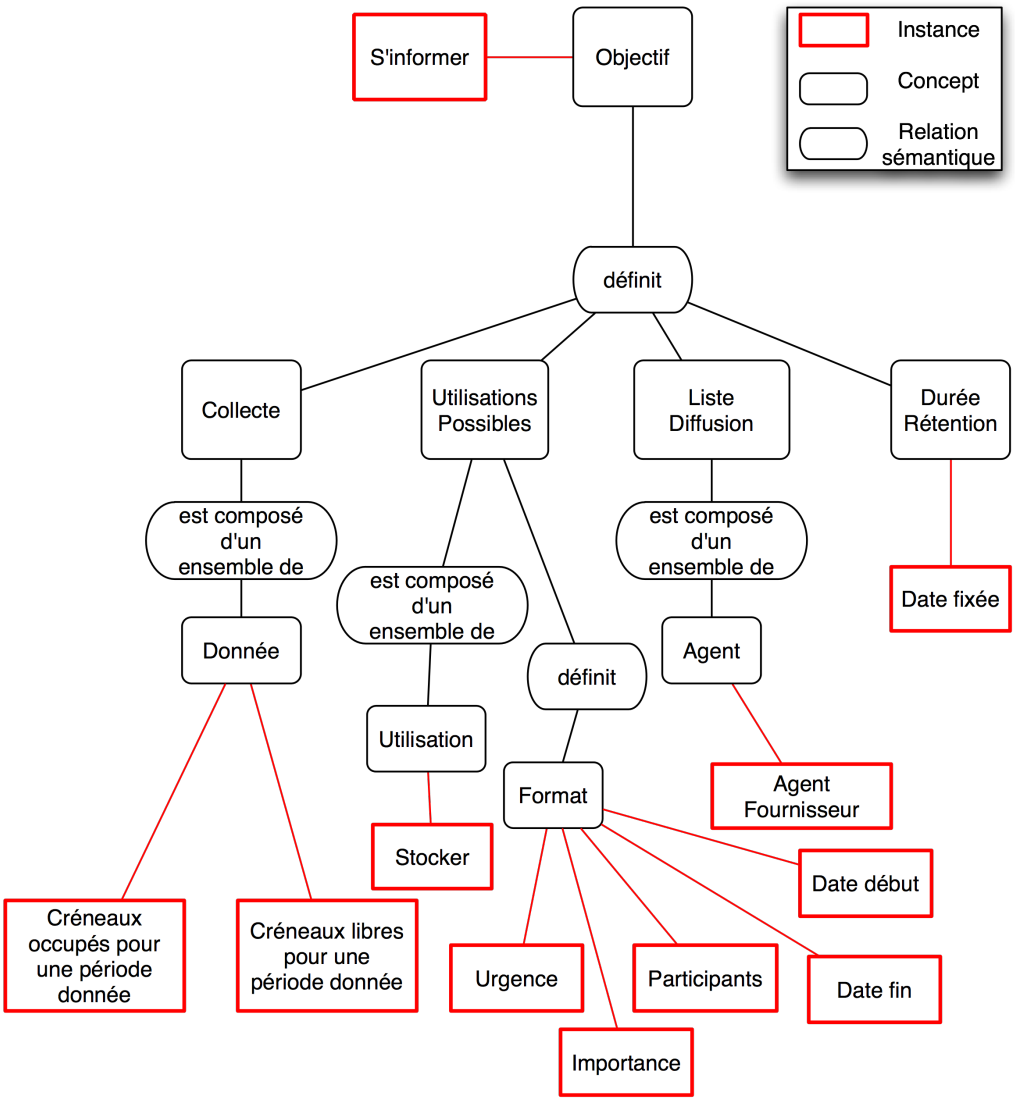

FIG. 8.8 – Instanciation du dictionnaire du domaine pour l'objectif "s'informer".

Cet objectif définit un consommateur qui veut s'informer sur les disponibilités d'un fournisseur pour une période donnée, figure 8.8. Dans le contexte de cet objectif, nous posons les restrictions suivantes :

- Les données sensibles que le consommateur peut collecter sont les créneaux horaires libres et les créneaux occupés pour une période donnée.
- Les données sensibles peuvent être fournies avec toutes les références que le fournisseur permet de diffuser.
- Les données recueillies par le consommateur ne peuvent pas être conservées en mémoire au-delà d'une date fixée.
- Le consommateur ne peut pas diffuser les données sensibles transmises par un fournisseur mais il doit en garantir l'accès à celui-ci qui est également le sujet de ces informations.
- La seule utilisation possible des données sensibles pour cet objectif est de stocker les informations recueillies.

L'implémentation de ce cas d'étude se fait par l'insertion dans notre dictionnaire du domaine des valeurs définies dans la figure 8.8. Cette insertion se traduit par la balise présentée dans la figure 8.9.

```
<RDF:RDF>
  <Liste-diffusion RDF:ID="null">
    <inclut RDF:resource="#agent-fournisseur"/>
  </Liste-diffusion>
  <Objectif RDF:ID="informer">
    <definit RDF:resource="#creneaux-occupes"/>
    <definit RDF:resource="#creneaux-libres"/>
    <definit RDF:resource="#stocker"/>
    <definit RDF:resource="#date-fixee"/>
    ...
  </Objectif>
</RDF:RDF>
```

FIG. 8.9 – Exemple d'implémentation du dictionnaire du domaine pour l'objectif "s'informer".

8.3.2.2 Fixer un rendez-vous

Ce deuxième objectif définit un consommateur qui veut fixer une rendez-vous avec un fournisseur dans une période donnée (cf. figure 8.10). Le consommateur précise donc dans sa politique que son objectif est de fixer un rendez-vous entre lui et le fournisseur. A partir de cet objectif, nous posons les restrictions suivantes :

- Les données sensibles que le consommateur peut collecter sont les créneaux horaires libres pour une période donnée.
- Les données sensibles peuvent être transmises avec toutes les références que le fournisseur permet de diffuser.

– Les données recueillies par le consommateur ne peuvent pas être gardées en mémoire au-delà d'une date fixée.
– Le consommateur ne peut pas diffuser les données sensibles transmises mais il doit en garantir l'accès au fournisseur qui est également le sujet de ces données.
– Les utilisations possibles des données sensibles pour l'objectif *fixer un rendez-vous* sont de stocker les informations recueillies et de les utiliser pour négocier un rendez-vous avec le fournisseur.

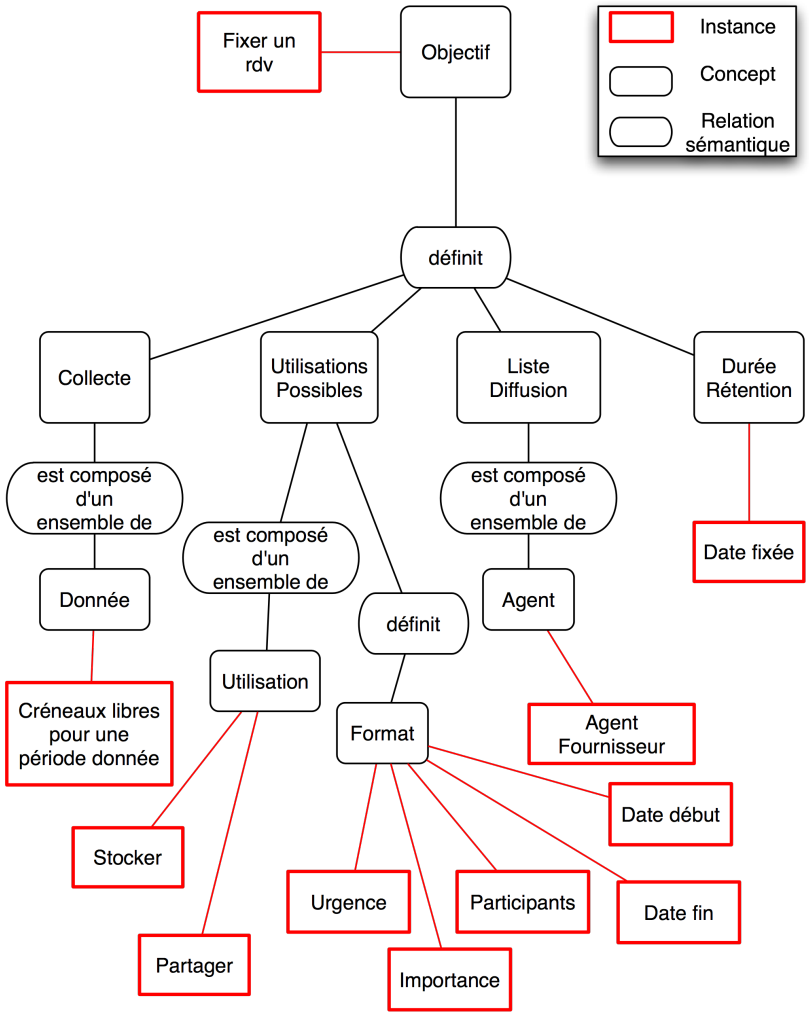

FIG. 8.10 – Instanciation du dictionnaire du domaine pour l'objectif "fixer un rendez-vous".

L'implémentation de cet objectif se fait en rajoutant dans le dictionnaire du domaine la balise décrite dans la figure 8.11, traduction en un fichier OWL de la figure 8.10.

```
<RDF:RDF>
  <Liste-diffusion RDF:ID="null">
    <inclut RDF:resource="#agent-fournisseur"/>
  </Liste-diffusion>
  <Objectif RDF:ID="fixer-rdv">
    <definit RDF:resource="#creneaux-libres"/>
    <definit RDF:resource="#stocker"/>
    <definit RDF:resource="#negocier"/>
    <definit RDF:resource="#date-fixee"/>
    ...
  </Objectif>
</RDF:RDF>
```

FIG. 8.11 – Exemple d'implémentation du dictionnaire du domaine : "fixer un rendez-vous".

8.3.2.3 Fixer un rendez-vous de groupe

Ce dernier objectif définit un consommateur qui veut fixer un rendez-vous avec un fournisseur et d'autres agents (groupe G) dans une période donnée (un intervalle de temps borné par deux créneaux de temps). Nous considérons ici que le fournisseur incarne également le rôle de sujet.

Pour fixer un tel rendez-vous, nous définissons les contraintes suivantes pour les politiques des agents :

- Les données sensibles que le consommateur peut collecter sont les créneaux libres pour une période donnée.
- Les données sensibles peuvent être fournies avec toutes les références que le fournisseur permet de diffuser.
- Les données recueillies par le consommateur ne peuvent pas être conservées en mémoire au-delà d'une date fixée.
- Le consommateur peut diffuser ces données sensibles aux agents du groupe G et il doit en garantir l'accès au fournisseur.
- Les utilisations possibles des données sensibles dans le contexte de la détermination d'un rendez-vous de groupe sont de stocker les données recueillies, de les utiliser pour négocier un rendez-vous avec le fournisseur et de partager ces données avec les agents du groupe G.

Le dictionnaire du domaine s'implémente en instanciant les classes de la figure 8.7 avec les valeurs fournies dans la figure 8.12. Pour notre exemple, la classe *Objectifs* s'instancie par la valeur "fixer-rdv-groupe" et cette valeur définit les valeurs "créneaux-libres", "négocier, stocker, partager", "date-fixée" et "agent-fournisseur, groupe G" pour les classes *Données*, *Utilisations*, *Durée-Rétention* et *ListeDiffusion*.

L'implémentation de cet exemple se fait en rajoutant dans le dictionnaire du domaine la balise décrite dans la figure 8.13.

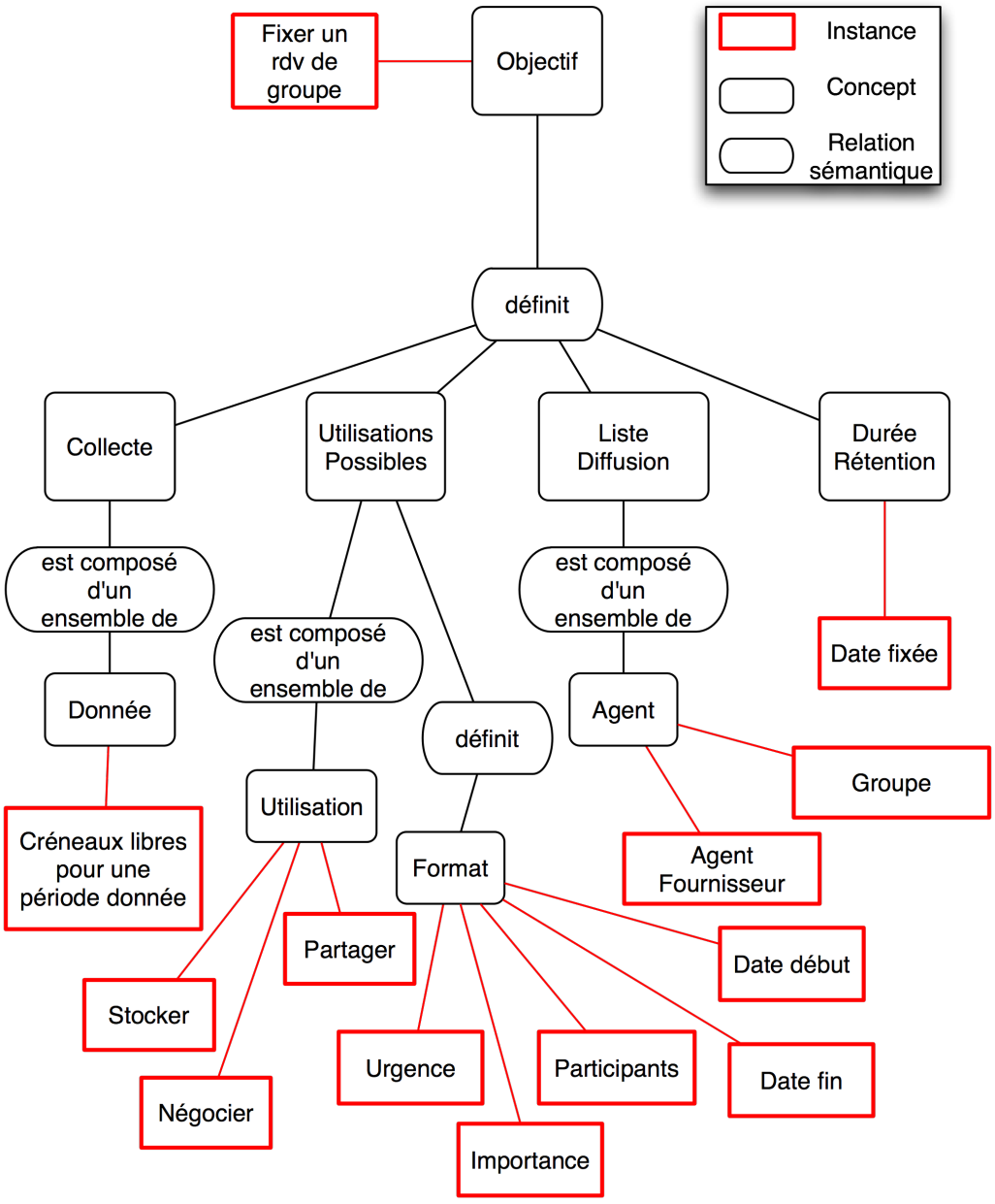

FIG. 8.12 – Dictionnaire du domaine pour l'objectif "fixer un rendez-vous de groupe".

```
<RDF:RDF>
  <Liste-diffusion RDF:ID="groupe">
    <inclut RDF:resource="#agent-fournisseur"/>
    <inclut RDF:resource="#agent-sujet"/>
  </Liste-diffusion>
  <Objectif RDF:ID="fixer-rdv-groupe">
    <definit RDF:resource="#groupe"/>
    <definit RDF:resource="#creneaux-libres"/>
    <definit RDF:resource="#partager"/>
    <definit RDF:resource="#stocker"/>
    <definit RDF:resource="#negocier"/>
    <definit RDF:resource="#date-fixee"/>
    ...
  </Objectif>
</RDF:RDF>
```

FIG. 8.13 – Exemple d'implémentation du dictionnaire du domaine pour l'objectif "fixer un rendez-vous de groupe".

Le dictionnaire du domaine étant implémenté dans un fichier OWL accessible à l'ensemble des agents, ils peuvent maintenant appliquer le protocole de transaction de données sensibles et le contrôle social hippocratique pour respecter la sphère privée dont ils ont la charge.

8.3.3 Raisonnement des agents

A chaque partage d'agenda, le consommateur construit sa politique en fonction des souhaits de l'utilisateur qu'il représente. La création de la politique se fait en appliquant les règles spécifiées par l'utilisateur lors de la création de la sphère privée de l'agent.

L'utilisateur choisit les objectifs de la transaction et l'agent crée automatiquement sa politique en vérifiant sa validité avec le dictionnaire du domaine pour vérifier la sémantique.

Il vérifie également la syntaxe du fichier de la transaction de données sensibles grâce à un schéma XSD prédéfini qui spécifie ce que doit contenir les fichiers des transactions de données sensibles.

Si ces deux validations sont correctes, la transaction de données sensibles commence par l'envoi du fichier XML au fournisseur. Dans le cas contraire, le consommateur prévient l'utilisateur de l'erreur engendrée et la transaction ne débute pas. Prenons comme exemple, le consommateur *user.magma*1 qui demande à l'agent *user.magma*2 son agenda pour fixer un rendez-vous. Le fichier de transaction de données sensibles envoyé est représenté sur la figure 8.14. L'agent demande à *user.magma*2 de lui envoyer les créneaux horaires

```
<TransactionDonneesSensibles>
        <ID value="8"/>
        <consent value="false"/>
                <objectif value="fixer-rdv-groupe">
                <collection>
                        <donneeSensible id="creneaux-libres" value="null"/>
                </collection>
                <listeDiffusion>
                        <Sujet ID="user.magma2"/>
                        <Fournissuer ID="user.magma2"/>
                        <Groupe ID="travail"/>
                </listeDiffusion>                <utilisations>
                        <utilisation value="stocker"/>
                        <utilisation value="négocier"/>
                        <utilisation value="partager"/>
                </utilisations>
                <dureeRetention value="fin-session"/>
                <format>
                        <reference value="importance"/>
                        <reference value="urgence"/>
                </format>
        </objective>
</TransactionDonnéesSensibles>
```

FIG. 8.14 – Fichier de transaction de données sensibles envoyé à l'agent fournisseur *user.magma*2.

libres de son agenda dans le but de fixer un rendez-vous avec le groupe *travail*. Il propose d'utiliser ces données pour les stocker, les partager avec les agents du groupe et de les utiliser pour négocier le rendez-vous avec le groupe *travail*. Les références demandées sont l'importance et l'urgence des rendez-vous. Ces données seront supprimées en fin de session.

Lorsque l'agent *user.magma*2 reçoit le fichier de transaction de données sensibles, il commence par raisonner sur la relation de confiance qu'il entretient avec *user.magma*1. Pour ce faire, il commence par demander à la société d'agents, via l'agent serveur, de lui communiquer la réputation propagée qu'elle accorde à l'agent *user.magma*1. Ensuite il intègre les réputations propagées reçues au calcul de la valeur de confiance qu'il accorde à *user.magma*1. Si celle-ci est supérieure au seuil fixé par l'utilisateur, la transaction de données sensibles peut continuer. Dans le cas contraire, *user.magma*2 annule la transaction de données sensibles.

Par exemple, *user.magma*2 possède comme réputation de confiance pour *user.magma*1 $DoT_{user.magma1,professionnel} = 0,7125$. *user.magma*2 demande aux agents de la société de lui communiquer leur réputation propagée sur *user.magma*1 dans le contexte $professionnel$. Les seules réputations propagées reçues par *user.magma*2 portent sur la facette $diffusion$. Se basant sur le fait qu'aucun nouvel agent a intégré la société, *user.magma*2 compile la valeur des réputations propagées en une moyenne, ce qui lui donne comme valeur 0,75. *user.magma*2 détermine ensuite la moyenne de cette valeur avec

la valeur de sa croyance de réputation propagées (0,85). Ainsi *user.magma*2 possède maintenant comme $DoPR_{user.magma1,professionnel}$ la valeur 0,8 et peut donc calculer[5] son nouveau niveau de croyance envers *user.magma*2 qui prend comme valeur 0,7031. Cette valeur étant supérieure au seuil fixé par l'utilisateur qui est de 0,5 (cf. figure 8.5), *user.magma*2 établit une relation de confiance avec *user.magma*1.

Une fois la relation de confiance établie entre *user.magma*2 et *user.magma*1, *user.magma*2 valide le fichier reçu sémantiquement d'une part avec le dictionnaire du domaine, et syntaxiquement d'autre part avec le schéma XSD afin de vérifier que les intentions de l'agent *user.magma*1 ne sont pas suspicieuses. Dans le cas d'un comportement suspicieux, *user.magma*2 diminue la réputation qu'il accorde à *user.magma*1 sur les facettes qui ont été violées et la transaction de données sensibles ne peut avoir lieu.

```
<TransactionDonneesSensibles>
        <ID value="8"/>
        <consent value="false"/>
                <objectif value="fixer-rdv-groupe">
                <collection>
                        <donneeSensible id="creneaux-libres" value="null"/>
                </collection>
                <listeDiffusion>
                        <Sujet ID="user.magma2"/>
                        <Fournissuer ID="user.magma2"/>
                        <Groupe ID="travail"/>
                </listeDiffusion>               <utilisations>
                        <utilisation value="stocker"/>
                        <utilisation value="négocier"/>
                        <utilisation value="partager"/>
                </utilisations>
                <dureeRetention value="fin-session"/>
                <format>
                        <reference value="importance"/>
                </format>
        </objective>
</TransactionDonnéesSensibles>
```

FIG. 8.15 – Fichier de transaction de données sensibles modifié par *user.magma*2.

Si le fichier est validé, la transaction de données sensibles peut continuer par la création de la préférence de l'agent *user.magma*2 selon les souhaits des utilisateurs intégrés dans son profil. Si la préférence et la politique concordent, *user.magma*2 envoie son consentement et partage alors son agenda. Dans le cas contraire, *user.magma*2 modifie le fichier de transaction selon les souhaits que l'utilisateur lui a indiqués et l'envoie à *user.magma*1. Dans notre exemple, l'utilisateur ne veut pas transmettre ses rendez-vous urgents, comme le montre la figure 8.15 qui représente le même fichier que pour la politique mais où la balise *format* ne contient qu'une balise *reference*. Dès lors *user.magma*1 accepte

[5]Ce calcul se fait par le biais de la fonction présentée dans le chapitre 6.

ou non ses changements selon les indications de l'utilisateur et la transaction aboutit ou bien est annulée.

Remarquons que lorsqu'un consommateur reçoit l'agenda d'un tiers par un fournisseur, il vérifie la politique associée par le premier fournisseur et s'il détecte une infraction à cette politique, il baisse la réputation du fournisseur comme le veut le contrôle social hippocratique. Les utilisateurs ont également la possibilité de détecter des comportements suspicieux. Dans ce cas, l'utilisateur informe son agent qui intègre ces nouvelles réputations dans son calcul de la valeur de confiance par le calcul des nouvelles valeurs des croyances de confiance.

8.4 Synthèse

La migration de l'application AGENDA vers un HiMAS illustre une implémentation de notre modèle. Ce travail nous permet de mettre en œuvre concrètement notre protocole de transaction de données sensibles ainsi que le contrôle social hippocratique pour préserver la sphère privée des agents relative aux agendas des utilisateurs tout en respectant leurs souhaits grâce aux profils.

Afin de compléter les travaux proposés dans ce chapitre, une perspective intéressante consiste à évaluer qualitativement les méthodes employées pour respecter la sphère privée en soumettant aux utilisateurs une grille de critères et ce, principalement au niveau de la délégation des données sensibles et au niveau de la création des politiques et des préférences. Ainsi nous pourrions poursuivre cette implémentation en termes de qualité de service utilisateur.

Chapitre 9

Conclusions et perspectives

Sommaire

Nous réalisons dans ce chapitre une analyse critique de nos contributions pour le respect de la sphère privée dans les systèmes multi-agents centrés utilisateur. Cette analyse nous permet d'émettre des conclusions sur notre travail ainsi que de dégager les perspectives que notre proposition peut engendrer.

9.1 Problématique

Cette thèse aborde les problèmes engendrés par la préservation de la *privacy* selon un point de vue moral en informatique, que nous étudions à travers le respect de la sphère privée. Deux axes de recherches sont généralement mis en relation dans ce contexte [Deswarte et Melchor (2006); Spiekermann et Cranor (2009)] : (i) les travaux en cryptographie et (ii) les travaux portant sur la gestion des données sensibles en termes de diffusion, d'utilisations et d'abus. Notre cadre de recherche se focalise sur un type particulier de systèmes multi-agents, les systèmes-multi agents centrés utilisateur [Demazeau (2003)], où les agents artificiels et autonomes prennent en charge la gestion et la protection des données sensibles des utilisateurs.

Le respect de la sphère privée requiert une approche se décomposant en trois phases critiques : (i) le stockage des données sensibles, (ii) la transaction de données sensibles et (iii) le devenir des données sensibles diffusées. Afin de ne pas violer la sphère privée des agents représentant les utilisateurs, le système multi-agent doit assurer un niveau de sécurité performant lors du stockage et

des transactions des données sensibles afin de prévenir les possibles attaques et intrusions. De plus, les agents doivent prendre en considération, dans leur raisonnement, des mécanismes leur permettant de vérifier les intentions des autres agents lors des transactions de données sensibles et détecter tout comportement suspicieux lors des traitements des données sensibles diffusées.

Ainsi, pour que le respect de la sphère privée soit complet, trois propositions inter-dépendantes doivent être abordées : la sécurité, un protocole de communication dédié aux transactions de données sensibles et un mécanisme de régulation pour les comportements des agents.

9.2 Contributions

Afin d'assurer la gestion des données sensibles des utilisateurs au sein des systèmes multi-agents centrés utilisateur, nous avons intégré une sphère privée à chaque agent artificiel représentant l'ensemble des données sensibles et les règles de gestion associées respectant les souhaits exprimés par les utilisateurs lors de la création de leur profil.

Cette thèse propose également un modèle, les systèmes multi-agents hippocratiques (HiMAS) inspiré des bases de données hippocratiques [Agrawal *et al.* (2002)], intégrant le concept de sphère privée et assurant son respect lors des phases critiques présentées précédemment grâce à un ensemble de neuf principes normatifs.

Le principe de sécurité garantit le stockage et les transactions de données au niveau des attaques et des intrusions suspicieuses possibles.

La phase critique concernant la transaction de données sensibles est étudiée au travers de sept principes : le consentement, la connaissance des objectifs, la collection minimale, l'utilisation minimale, la diffusion minimale, la rétention minimale et la transparence. Ces principes sont formalisés et intégrés au raisonnement des agents par la spécification d'un protocole de transaction de données sensibles que les agents d'un système multi-agent hippocratique doivent exécuter pour communiquer les données sensibles que leur sphère privée contient. Ce protocole s'inspire de la plate-forme pour les préférences de confidentialité [W3C (2002b)] pour la création des politiques des consommateurs et des préférences des fournisseurs. L'introduction d'un dictionnaire de domaine commun à l'ensemble des agents sous forme d'un graphe conceptuel interprété permet une compréhension mutuelle pour la mise en correspondance des politiques et des préférences. De plus, cette technique donne la possibilité aux agents de vérifier les intentions des autres agents en définissant une limite maximale des valeurs des éléments des politiques et des préférences en fonction du domaine, ce qui permet d'éliminer les comportements suspicieux explicites.

La dernière phase critique, le devenir des données sensibles, est assurée par le principe de conformité. Chaque agent d'un système multi-agent hippocratique doit être en mesure de vérifier le respect des principes normatifs imposés par notre modèle. Nous proposons de formaliser ce principe par un mécanisme de régulation interne, le contrôle social hippocratique fondé sur des relations de confiance, afin de ne pas provoquer de violations de la sphère privée lors de la régulation de comportement. Le contrôle social hippocratique permet également l'inclusion des données relatives aux relations de confiance, donc relatives aux utilisateurs, à la sphère privée des agents par le respect du protocole de transaction de données sensibles spécifique aux envois de réputations propagées. Ce contrôle social hippocratique donne la capacité aux agents de vérifier la fiabilité des autres agents en dénonçant chaque violation observée à la communauté d'agents qui intègre les sanctions encourues dans leur raisonnement.

La dernière contribution de nos travaux consiste en l'implémentation du modèle proposé dans une application multi-agent concrète de gestion décentralisée d'agendas. Nous avons introduit au sein de cette application la représentation de la sphère privée pour les agents en charge des données sensibles des utilisateurs, le protocole de transaction de données sensibles lors de l'envoi des agendas et le contrôle social hippocratique a ainsi pu être validé expérimentalement.

9.3 Limites

Notre étude des systèmes multi-agents hippocratiques se focalise essentiellement sur les principes normatifs relatifs au raisonnement des agents, n'incluant donc pas le principe de sécurité qui est strictement d'ordre cryptographique. Afin de compléter notre proposition, il serait intéressant d'étudier les possibilités offertes par ce domaine de recherche afin de l'intégrer aux systèmes multi-agents hippocratiques. Cette étude doit prendre en considération les performances en termes de sécurité mais également en termes de temps utilisateurs. En effet, au vu de la complexité des techniques cryptographiques et travaillant dans des systèmes multi-agents centrés utilisateur, nos principales préoccupations portent sur la satisfaction des utilisateurs en termes de protection et de temps d'exécution.

La modélisation de la sphère privée et des mécanismes de protection dédiés que nous proposons intègrent implicitement l'aspect temporel dû à l'évolution de la sphère privée, notamment par la datation des règles et des normes de la sphère privée. L'expression explicite du temps apporterait un modèle complet, permettant ainsi de mieux considérer le respect de la sphère privée au fil de l'exécution d'un système multi-agent hippocratique. En effet, il serait

intéressant d'étendre notre modélisation de la sphère privée à l'aide de la logique temporelle, comme par exemple [Pnueli (1977)], afin de pouvoir étudier plus précisément la dynamique de la sphère et les impacts engendrés par cette caractéristique sur la protection des données sensibles.

Le protocole de transaction de données sensibles intégré au sein des systèmes multi-agents hippocratiques permet de pallier les critiques émises pour la plate-forme pour les préférences de confidentialité [Thibadeau (2000)]. Cependant la création du dictionnaire du domaine est une tâche très coûteuse car elle requiert une parfaite connaissance du domaine afin de définir les ensembles maximaux des instances des concepts. Cela implique donc un temps d'analyse et de développement assez long. De plus, rien n'assure le fait que le dictionnaire du domaine soit complet et permette donc toutes les transactions de données sensibles souhaitées par les utilisateurs.

Le mécanisme de régulation choisi, le contrôle social, ne permet qu'une vérification a posteriori de certaines violations de la sphère privée. En effet, ces violations ne peuvent être détectées que par la connaissance d'une manipulation frauduleuse. Dans le cas où un agent manipule d'une manière suspicieuse ses données sensibles en dehors du système par exemple, aucune vérification n'est possible. De plus ce type de mécanisme ne permet pas la réparation des dommages subis par les violations de la sphère privée.

9.4 Perspectives

Une première prolongation importante de nos travaux consisterait à étudier l'introduction des normes et des conflits engendrés avec les règles personnelles de la sphère privée. Pour ce faire, une première piste de travail est d'intégrer l'agent PAw [Piolle (2009)] dans les systèmes multi-agents hippocratiques afin de prendre en considération les problèmes relatifs aux normes.

Une autre perspective en lien avec la gestion de la sphère privée concernerait la création et la maintenance du profil utilisateur. A l'heure actuelle, les agents d'un système multi-agent hippocratique gèrent le profil de l'utilisateur, et donc leur sphère privée, par des retours explicites de l'utilisateur afin de respecter leurs souhaits en termes de protection de données sensibles. Cependant, afin d'alléger la charge de l'utilisateur, il serait intéressant d'appliquer des mécanismes implicites de personnalisation, comme par exemple [Krulwich (1997)], au sein de ces agents qui ne violeraient pas la sphère privée et qui correspondraient en tous points aux souhaits des utilisateurs.

Pour finir, une perspective très intéressante consisterait à étudier le mensonge comme moyen de protection de la sphère privée. En effet, lorsqu'un agent doit fournir ses données sensibles à un agent suspicieux afin d'accéder à un service nécessaire à la réalisation de ses objectifs, mentir représente une solution

où aucune violation de la sphère privée n'est réalisée et qui permet à cet agent d'effectuer les tâches qui sont lui assignées. L'utilisation du mensonge n'est pas aisée [Grasland (2009)] car elle demande aux agents de raisonner sur l'ensemble des diffusions des données erronées et d'inférer leurs possibles diffusions par les autres agents afin de ne pas créer d'ambiguïté au sein de la société et ce afin de ne pas perdre en niveau de fiabilité pour les autres agents qui risquent de considérer cet acte comme suspicieux. L'intégration d'un tel concept donnerait sans aucun doute une envergure supplémentaire aux systèmes multi-agents hippocratiques.

Bibliographie

Alfarez Abdul-Rahman. *A Framework for Decentralised Trust Reasoning*. PhD thesis, Department of Computer Science, University College London, 2004.

Rakesh Agrawal, Jerry Kiernan, Ramakrishnan Srikant, et Yirong Xu. Hippocratic databases. In *Proceedings of the International Conference on Very Large Data Bases*, pages 143–154. Morgan Kaufmann, 2002.

Rakesh Agrawal, Peter J. Haas, et Jerry Kiernan. Watermarking relational data: framework, algorithms and analysis. *The International Journal on Very Large Data Bases*, 12(2) :157–169, 2003.

Rakesh Agrawal, Paul Bird, Tyrone Grandison, Jerry Kiernan, Scott Logan, et Walid Rjaibi. Extending relational database systems to automatically enforce privacy policies. In *Proceedings of the International Conference on Data Engineering*, pages 1013–1022. IEEE Computer Society, 2005.

Tanvir Ahmed et Anand R. Tripathi. Static verification of security requirements in role based cscw systems. In *Proceedings of the ACM symposium on Access control models and technologies*, pages 196–203. ACM, 2003.

AOS. Jack development environment, 1997-2009. www.agent-software.com.

Alexander Artikis, Jeremy Pitt, et Marek J. Sergot. Animated specifications of computational societies. In *Proceedings of the International Joint Conference on Autonomous Agents & Multiagent Systems*, pages 1053–1061. ACM, 2002.

Sara Baase. *A Gift of Fire: Social, Legal, and Ethical Issues in Computing*. Prentice Hall PTR, Upper Saddle River, NJ, USA, 2002.

José-Antonio Báez-Barranco, Tiberiu Stratulat, et Jacques Ferber. A unified model for physical and social environments. In *Proceedings of the Environments for Multi-Agent Systems III, Third International Workshop*, Lecture Notes in Computer Science, pages 41–50. Springer, 2007.

Marko Balabanovic et Yoav Shoham. Content-based, collaborative recommendation. 40(3) :66–72, 1997.

Victoria Bellotti et Abigail Sellen. Design for privacy in ubiquitous computing environments. In *Proceedings of the European Conference on Computer Supported Cooperative Work*, pages 77–92. Kluwer Academic Publishers, 1993.

Jamal Bentahar, Bernard Moulin, et Brahim Chaib-draa. Towards a formal framework for conversational agents. In *Proceedings of the Agent Communication Languages and Conversation Policies Workshop at AAMAS 2003*, éditeurs Marc-Philippe Huget et Frank Dignum, 2003.

Federico Bergenti. Secure, trusted and privacy-aware interactions in large-scale multiagent systems. In *Proceedings of the Workshop From Objects to Agents*, pages 144–150. Pitagora Editrice Bologna, 2005.

Guido Boella, Leendert W. N. van der Torre, et Harko Verhagen. Introduction to normative multiagent systems. In *Normative Multi-agefnt Systems*, volume 07122 de *Dagstuhl Seminar Proceedings*. Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany, 2007.

Michael E. Bratman. Intention, plans, and practical reason. O'Reilly, Harvard University Press: Cambridge MA, 1987.

Paolo Busetta, Antonia Donà, et Michele Nori. Channeled multicast for group communications. In *Proceedings of the International Joint Conference on Autonomous agents and multiagent systems*, pages 1280–1287. ACM, 2002.

Sara J. Casare et Jaime Simão Sichman. Towards a functional ontology of reputation. In *Proceedins of the International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 505–511. ACM, 2005.

Cristiano Castelfranchi et Rino Falcone. Principles of trust for MAS: Cognitive anatomy, social importance, and quantification. In *Proceedings of the International Conference on Multiagent Systems*, pages 72–79. IEEE Computer Society, 1998.

Cristiano Castelfranchi. Engineering social order. In *Proceedings of the International Workshop Engineering Societies in the Agent World*, volume 1972 de *Lecture Notes in Computer Science*, pages 1–18. Springer, 2000.

éditeurs Liqun Chen, Chris J. Mitchell, et Andrew Martin (éditeurs). *Trusted Computing, Second International Conference, Trust 2009, Proceedings*, volume 5471 de *Lecture Notes in Computer Science*. Springer, 2009.

Liren Chen et Katia P. Sycara. WebMate: A personal agent for browsing and searching. In *Proceedings of Agents*, pages 132–139. ACM, 1998.

Richard Cissée et Sahin Albayrak. Experimental analysis of privacy loss in dcop algorithms. In *Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 1424–1426. ACM, 2007.

Rosiria Conte et Mario Paolucci. *Reputation in Artificial Societies. Social Beliefs for Social Order*. Boston : Kluwer, 2002.

Lorrie Faith Cranor. *Web Privacy with P3P*. O'Reilly, 2002.

Ludivine Crépin, Laurent Vercouter, François Jacquenet, Yves Demazeau, et Olivier Boissier. Hippocratic Multi-Agent Systems. In *Proceedings of the International Conference of Entreprise Information Systems*, éditeurs José Cordeiro et Joaquim Filipe, pages 301–308, 2008.

Ludivine Crépin, Laurent Vercouter, François Jacquenet, Yves Demazeau, et Olivier Boissier. Towards HiMAS: A model for privacy preserving multi-agent systems. In *Proceedings of the International Workshop Trust in agent societes during the International Joint Conference on Autonomous Agents and Multiagent Systems*, 2008.

Ludivine Crépin, Yves Demazeau, François Jacquenet, et Olivier Boissier. Privacy preservation in a decentralized calendar system. In *Proceedings of the International Conference on Practical Applications of Agents and Multi-Agent Systems*, pages 529–537. Springer, 2009.

Ludivine Crépin, Yves Demazeau, Olivier Boissier, et François Jacquenet. Transaction de données sensibles au sein de Systèmes Multi-Agents Hippocratiques. *Revue d'Intelligence Artificielle*, à paraître, 2009.

Ludivine Crépin, Laurent Vercouter, Yves Demazeau, François Jacquenet, et Olivier Boissier. Systèmes Mutli-Agents Hippocratiques. In *Intelligence Artificielle et Web Intelligence (atelier IAWI de la plate-forme AFIA 2007)*, 2007.

Ludivine Crépin, Yves Demazeau, Olivier Boissier, et Francois Jacquenet. Transaction de données sensibles au sein d'un Système Multi-Agent Hippocratique. In *Proceedings of the Journées Francophones sur les Systèmes Multi-Agents*, éditeurs Mandiau et Chevallier, pages 223–232, 2008.

Ludivine Crépin, Yves Demazeau, Olivier Boissier, et Francois Jacquenet. Sensitive data transaction in Hippocratic Multi-Agent Systems. In *Proceedings of the International Workshop on Engineering Societies in the Agents World*. Springer, à paraître, 2009.

Ernesto Damiani, Sabrina De Capitani di Vimercati, Stefano Paraboschi, et Pierangela Samarati. P2P-based collaborative spam detection and filtering.

In *Proceedings of the International Conference on Peer-to-Peer Computing*, pages 176–183. IEEE Computer Society, 2004.

Partha Dasgupta. Trust: Making and breaking cooperative relations. In *Trust as a commodity*, pages 49–72. New York, NY, USA : B. Blackwel, 1990.

Yves Demazeau, Dimitri Melaye, et Marie-Hélène Verrons. A decentralized calendar system featuring sharing, trusting and negotiating. In *Proceedings of the International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, volume 4031 de *Lecture Notes in Computer Science*, pages 731–740. Springer, 2006.

Yves Demazeau. Créativité Emergente Centrée Utilisateur. In *Proceedings of the Journées Francophones sur les Systèmes Multi-Agents*, pages 31–36. Hermès, 2003.

Pierre Demeulenaere. Les difficultés de la caractérisation de la notion de la vie privée d'un point de vue sociologique. In *La protection de la vie privée dans la société d'information*, volume 11, 2002. Groupe d'études Société d'information et vie privée.

Robert Demolombe. Reasoning about trust: A formal logical framework. In *Proceedings of the International Conference Trust Management, iTrust*, volume 2995 de *Lecture Notes in Computer Science*, pages 291–303. Springer, 2004.

Yves Deswarte, Laurent Blain, et Jean charles Fabre. Intrusion tolerance in distributed computing systems. In *Proceedings of the IEEE Symposium on Research in Security and Privacy*, pages 110–121, 1991.

Yves Deswarte et Carlos Aguilar Melchor. Current and future privacy enhancing technologies for the internet. *Annales des Télécommunications*, 61 :399–417, 2006.

Marc Esteva, David de la Cruz, et Carles Sierra. ISLANDER: an electronic institutions editor. In *Proceedings of the International Joint Conference on Autonomous Agents & Multiagent Systems*, pages 1045–1052. ACM, 2002.

Marc Esteva, Bruno Rosell, Juan A. Rodríguez-Aguilar, et Josep Lluís Arcos. AMELI: An agent-based middleware for electronic institutions. In *Proceedings of the International Joint Conference on Autonomous Agents & Multiagent Systems*, pages 236–243. IEEE Computer Society, 2004.

Rino Falcone, Giovanni Pezzulo, et Cristiano Castelfranchi. A fuzzy approach to a belief-based trust computation. In *Proceedings of the International*

Workshop Trust, Reputation, and Security during the International Conference Autonomous Agent and Multiagent Systems, volume 2631 de *Lecture Notes in Computer Science*, pages 73–86. Springer, 2002.

David F. Ferraiolo et D. Richard Kuhn. Role-based access controls. In *Proceedings of the National Computer Security Conference*, pages 554–563, 1992.

Fondation Internet Nouvelle Génération (FING). Identités actives. www.fing.org/ - http ://www.identitesactives.net/, 2008.

Stanford Center for Biomedical Informatics Research. Protégé, 1997-2009. http ://protege.stanford.edu/.

Nicoletta Fornara et Marco Colombetti. Specifying and enforcing norms in artificial institutions. In *Proceedings of the International Joint Conference on Autonomous Agents & Multiagent Systems*, pages 1481–1484. IFAAMAS, 2008.

Jabber Software Foundation. Jabber. http ://www.jabber.org/, 2001.

République Française. Loi numéro 2004-801du 6 août 2004 relative à la protection des personnes physiques à l'égard des traitements de données à caractère personnel. In *Journal Officiel de la République Française*, 2004.

Eugene C. Freuder, Marius Minca, et Richard J. Wallace. Privacy/efficiency tradeoffs in distributed meeting scheduling by constraint-based agents. In *Notes of the International Joint Conference on Artificial Intelligence Workshop on Distributed Constraint Reasoning*, pages 63–71, 2001.

Francis Fukuyama. *Trust. The Social Virtues and the Creation of Prosperity.* Free Press, 1995.

Andrés García-Camino, Pablo Noriega, et Juan A. Rodríguez-Aguilar. Implementing norms in electronic institutions. In *Proceedings of the International Joint Conference on Autonomous Agents & Multiagent Systems*, pages 667–673. ACM, 2005.

Susan Gauch, Mirco Speretta, Aravind Chandramouli, et Alessandro Micarelli. User profiles for personalized information access. *The Adaptive Web*, 4321 :54–89, 2007.

Gianluigi Gentili, Alessandro Micarelli, et Filippo Sciarrone. Infoweb: An adaptive information filtering system for the cultural heritage domain. *Applied Artificial Intelligence*, 17 :715–744, 2003.

Yves Grasland. Identités multiples et gestion du mensonge. Rapport de stage M2R - Université Joseph Fournier, 2009.

Rachel Greenstadt, Jonathan P. Pearce, Emma Bowring, et Milind Tambe. Experimental analysis of privacy loss in dcop algorithms. In *Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 1424–1426. ACM, 2006.

Nicola Guarino, Claudio Masolo, et Guido Vetere. Ontoseek: Content-based access to the web. In *IEEE Intelligent Systems*, volume 14 (3), pages 70–80. IEEE Computer Society, 1999.

Mahdi Hannoun, Olivier Boissier, Jaime Simão Sichman, et Claudette Sayettat. MOISE: An organizational model for multi-agent systems. In *Proceedings of the Advances in Artificial Intelligence, International Joint Conference, 7th Ibero-American Conference on AI*, volume 1952 de *Lecture Notes in Computer Science*, pages 156–165. Springer, 2000.

Andreas Herzig, Emiliano Lorini, Jomi F. Hübner, Jonathan Ben-Naim, Olivier Boissier, Cristiano Castelfranchi, Robert Demolombe, Dominique Longin, Laurent Perrussel, et Laurent Vercouter. Prolegomena for a logic of trust and reputation. In *Proceedings of the International Workshop on Normative Multiagent Systems (NorMAS 2008)*, 2008.

John J. Hopfield. Hopfield network. *Scholarpedia*, 2(5) :1977, 2007.

Hilary H. Hosmer. Metapolicies I. *ACM SIGSAC Data Management Workshop*, 10(2-3) :18–43, 1991.

Hilary H. Hosmer. Metapolicies II. In *Proceedings of the National Computer Security Conference*, pages 369–378. Elsevier Advanced Technology Publications, 1992.

Tomas Isdal, Michael Piatek, Arvind Krishnamurthy, et Thomas Anderson. Friend-to-friend data sharing with oneswarm. Rapport technique, http ://oneswarm.cs.washington.edu/, 2009.

Hussein Joumaa, Yves Demazeau, et Jean-Marc Vincent. Evaluation of Multi-Agent Systems: The case of Interaction. In *Proceedings of the International Conference on Information & Communication Technologies: from Theory to Applications*, Lecture Notes in Computer Science, pages 703–709. IEEE Computer Society, 2008.

Hussein Joumaa, Yves Demazeau, et Jean-Marc Vincent. Performance visualization of a multi-agent system application. In *Proceedings of the International Conference on Practical Applications of Agents and Multi-Agent Systems*, pages 188–196. Springer, 2009.

Diane Kelly et Jaime Teevan. Implicit feedback for inferring user preference: a bibliography. In *SIGIR Forum*, volume 37 (2), pages 18–28. ACM, 2003.

Bruce Krulwich. LIFESTYLE FINDER: Intelligent user profiling using large-scale demographic data. *AI Magazine*, 18(2) :37–45, 1997.

Winfried E. Kühnhauser. A paradigm for user-defined security policies. In *Symposium on Reliable Distributed Systems*, pages 135–144, 1995.

Laurent Lacomme, Yves Demazeau, et Valérie Camps. Personalization of a trust network. In *Proceedings of the International Conference on Agents and Artificial Intelligence*, pages 408–415. IEEE/ACM, 2009.

Kristen LeFevre, Rakesh Agrawal, Vuk Ercegovac, Raghu Ramakrishnan, Yirong Xu, et David J. DeWitt. Limiting disclosure in hippocratic databases. In *Proceedings of the International Conference on Very Large Data Bases*, pages 108–119. Morgan Kaufmann, 2004.

Lawrence Lessig. *Code and Other Laws of Cyberspace*. Basic Books, New York, 2000.

Henry Lieberman. Letizia: An agent that assists web browsing. In *Proceedings of the International Joint Conferences on Artificial Intelligence*, pages 924–929. Morgan Kaufmann, 1995.

Niklas Luhmann. Familiarity, confidence, trust: Problems and alternatives. In *Proceedings of Trust: Making and Breaking of Cooperative Relations*, pages 94–107. Basil Blackwell, 1988.

Emil Lupu, Morris Sloman, Naranker Dulay, et Nicodemos Damianou. PONDER: Realising enterprise viewpoint concepts. In *Proceedings of the International Enterprise Distributed Object Computing Conference*, pages 66–75. IEEE Computer Society, 2000.

Pattie Maes. Agents that reduce work and information overload. *Communications of the ACM*, 37(7) :30–40, 1994.

Hannes Marais et Krishna Bharat. Supporting cooperative and personal surfing with a desktop assistant. In *Proceedings of the annual ACM symposium on User interface software and technology*, pages 129–138. ACM Press, 1997.

Fabio Massacci, John Mylopoulos, et Nicola Zannone. From Hippocratic Databases to secure TROPOS: a computer-aided re-engineering approach. *International Journal of Software Engineering and Knowledge Engineering*, 17(2) :265–284, 2007.

Philippe Mathieu et Marie-Hélène Verrons. A general negotiation model using XML. *Artificial Intelligence and Simulation of Behaviour Journal*, 1(6) :523–542, 2005.

D. Harrison McKnight et Norman L. Chervany. Trust and distrust definitions: One bite at a time. In *Proceedings of the Workshop on Deception, Fraud, and Trust in Agent Societies held during the Autonomous Agents Conference*, pages 27–54. Springer-Verlag, 2001.

Dimitri Melaye, Yves Demazeau, et Thierry Bouron. Which adequate trust model for trust networks ? In *Proceedings of the IFIP Conference on Artificial Intelligence Applications and Innovations*, pages 236–244, 2006.

Dimitri Melaye et Yves Demazeau. Bayesian dynamic trust model. In *Proceedings of the International Central and Eastern European Conference on Multi-Agent Systems*, volume 3690 de *Lecture Notes in Computer Science*, pages 480–489. Springer, 2005.

Alessandro Micarelli et Filippo Sciarrone. Anatomy and empirical evaluation of an adaptive web-based information filtering system. *User Modeling and User-Adapted Interaction*, 14(2-3) :159–200, 2004.

Miquel Montaner, Beatriz López, et Josep Lluís de la Rosa. A taxonomy of recommender agents on the internet. *Artificial Intelligence Review*, 19(4) :285–330, 2003.

Alexandros G. Moukas, Alexandros G. Moukas, et Alexandros G. Moukas. Amalthaea: Information filtering and discovery using a multiagent evolving system. 11(5) :437–457, 1997.

Lik Mui, Mojdeh Mohtashemi, et Ari Halberstadt. Notions of reputation in multi-agents systems: a review. In *Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 280–287. ACM, 2002.

Guillaume Muller et Laurent Vercouter. Decentralized monitoring of agent communications with a reputation model. In *Proceedings of Trusting Agents for Trusting Electronic Societies*, volume 3577 de *Lecture Notes in Computer Science*, pages 144–161. Springer, 2004.

Guillaume Muller. *Utilisation de normes et de réputations pour détecter et sanctionner les contradictions.* PhD thesis, Ecole Nationale Supérieure des Mines, Saint-Etienne, 2006.

Günter Müller. Introduction of privacy and security in highly dynamic systems. *Communications of the ACM*, 49(9) :1013–1022, September 2006.

Elinor Ostrom. A behavioral approach to the rational choice theory of collective action: Presidential address, american political science association. *The American Political Science Review*, 92(1) :1–22, 1998.

Leysia Palen et Paul Dourish. Unpacking "privacy" for a networked world. In *Proceedings of the Conference on Human Factors in Computing Systems*, pages 129–136. ACM, 2003.

Philippe Pasquier, Roberto Flores, et Brahim Chaib-draa. Modeling flexible social commitments and their enforcement. In *Proceedings of the Fifth International Workshop Engineering Societies in the Agents World*, pages 139–151. Springer-Verlag, 2005.

Guillaume Piolle et Yves Demazeau. Obligations with deadlines and maintained interdictions in privacy regulation frameworks. In *Proceedings of the 2008 IEEE/WIC/ACM International Conference on Intelligent Agent Technology*, pages 162–168. IEEE, 2008.

Guillaume Piolle. *Agents utilisateurs pour la protection des données personnelles : modélisation logique et outils informatiques*. PhD thesis, Université Joseph Fourier - Grenoble I, Grenoble, France, 2009.

Jacques Pitrat. *Métaconnaissance, Futur de l'Intelligence Artificielle*. Hermés, 1990.

Amir Pnueli. The temporal logic of programs. In *Proceedings of the IEEE Symposium on the Foundations of Computer Science (FOCS-77)*, pages 46–57, Providence, Rhode Island, USA, 1977. IEEE Computer Society Press.

Hugo Pommier et François Bourdon. Agents mobiles et réseaux pair-à-pair : vers une gestion sécurisée de l'information répartie. In *Proceedings of the Journées Francophones sur les Systèmes Multi-Agents*, éditeurs Mandiau et Chevallier, pages 171–180, 2008.

Alexander Pretschner et Susan Gauch. Ontology based personalized search. In *Proceedings of the IEEE International Conference on Tools with Artificial Intelligence*, pages 391–398. IEEE Computer Society, 1999.

Louis Quéré. La structure cognitive et normative de la confiance. In *Réseaux*, volume 19 (108), pages 125–152, 2001.

Luz Marina Quiroga et Javed Mostafa. Empirical evaluation of explicit versus implicit acquisition of user profiles in information filtering systems. In *Proceedings of the ACM Conference on Digital libraries*, pages 238–239. ACM, 1999.

Martin Rehák et Michal Pechoucek. Trust modeling with context representation and generalized identities. In *Proceedings of the International Workshop Cooperative Information Agents*, volume 4676 de *Lecture Notes in Computer Science*, pages 298–312. Springer, 2007.

Abdelmounaam Rezgui, Mourad Ouzzani, Athman Bouguettaya, et Brahim Medjahed. Preserving privacy in web services. In *Proceedings of the Workshop on Web Information and Data Management*, pages 56–62. ACM, 2002.

Elaine Rich. User modeling via stereotypes. In *Readings in intelligent user interfaces*, pages 329–342. Morgan Kaufmann Publishers, 1998.

Javier Carbo Rubiera, José M. Molina López, et Jorge Dávila Muro. Reaching agreements through fuzzy counter-offers. In *Proceedings of the International Conference on Web Engineering*, volume 2722 de *Lecture Notes in Computer Science*, pages 90–93. Springer, 2003.

Jordi Sabater. *Trust and Reputation for Agent Societies.* PhD thesis, Universitat Autònoma de Barcelona, Spain, 2002.

Hidekazu Sakagami et Tomonari Kamba. Learning personal preferences on online newspaper articles from user behaviors. *Computer Networks and ISDN Systems*, 29(8-13) :1447–1455, 1997.

Ravi S. Sandhu, Edward J. Coyne, Hal L. Feinstein, et Charles E. Youman. Role-based access control models. *IEEE Computer*, 29(2) :38–47, 1996.

Marius-Calin Silaghi et Vaibhav Rajeshirke. The effect of policies for selecting the solution of a discsp on privacy loss. In *Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 1396–1397. IEEE Computer Society, 2004.

Munindar P. Singh. Social and psychological commitments in multiagent systems. In *AAAI Fall Symposium on Knowledge and Action at Social and Organizational Levels*, pages 104–106. AAAI, 1991.

Munindar P. Singh. A social semantics for agent communication languages. In *Issues in Agent Communication*, pages 31–45. Springer-Verlag, 2000.

Daniel J. Solove. A taxonomy of privacy. In *University of Pennsylvania Law Review*, volume 154 (3), pages 477–561. GWU Law School Public Law Research, 2006.

John F. Sowa. *Conceptual Structures: Information Processing in Mind and Machine.* Addison-Wesley, 1984.

Sarah Spiekermann et Lorrie Faith Cranor. Engineering privacy. *IEEE Transactions on Software Engineering*, 35(1) :67–82, 2009.

The European Parliament and the Council. Directive 2002/58/ec of the european parliament and of the council of 12 july 2002 concerning the processins of personal data and the protection of privacy in the electronic communications sector. In *Official Journal of the European Communities*, 2002.

Robert Thibadeau. A critique of P3P: Privacy on web. dollar.ecom.cmu.edu/p3pcritique/, 2000.

Judith J. Thomson. The right of privacy. Philosophy and Public Affairs 4 : 295-314, 1975.

Kevin P. Twidle et Emil Lupu. Ponder2 - policy-based self managed cells. In *Proceedings of the International Conference on Autonomous Infrastructure, Management and Security*, volume 4543 de *Lecture Notes in Computer Science*, page 230. Springer, 2007.

Georg Henrik von Wright. Deontic logic. In *Mind*, pages 1–15, 1951.

W3C. A P3P preference exchange language 1.0. http ://www.w3.org/TR/P3P-preferences/, 2002.

W3C. Plateform for privacy preferences. http ://www.w3.org/P3P/, 2002.

W3C. Ressource description framework, http ://www.w3.org/rdf/. 2002.

W3C. Ressource description framework schema, www.w3.org/tr/rdf-schema/. 2002.

W3C. Owl web ontology language. http ://www.w3.org/TR/owl-features/, 2004.

Samuel D. Warren et Louis D. Brandeis. *The right to privacy.* Wadsworth Publ. Co., Belmont, CA, USA, 1985.

Alan F. Westin. Special report: legal safeguards to insure privacy in a computer society. *Communications of the ACM*, 10(9) :533–537, 1967.

Michael Wooldridge et Nicholas R. Jennings. Intelligent agents: Theory and practice. In *Proceedings of European Conference on Artificial Intelligence Workshop on Agent Theories, Architectures, and Languages*, volume 890 de *Lecture Notes in Computer Science*, pages 115–152. Springer, 1995.

Makoto Yokoo, Koutarou Suzuki, et Katsutoshi Hirayama. Secure distributed constraint satisfaction: reaching agreement without revealing private information. *Artificial Intelligence*, 161(1-2) :229–245, 2005.

Giorgos Zacharia, Alexandros Moukas, et Pattie Maes. Collaborative reputation mechanisms in electronic marketplaces. In *Proceedings of the Hawaii International Conference on System Sciences*, pages 371 – 388, 1999.

Résumé

Avec l'explosion des technologies liées à Internet et aux systèmes multi-agents, l'évolution actuelle des systèmes d'information induit un traitement automatique massif des données des utilisateurs. Le développement des systèmes multi-agents centrés utilisateur amène un nouveau centre d'intérêt : la gestion et la protection des données sensibles des utilisateurs. Les recherches présentées dans cette thèse se focalisent donc sur le respect de la *privacy* des utilisateurs lorsqu'ils décident de céder une partie du contrôle de leurs données sensibles à un agent autonome interagissant avec d'autres agents. La *privacy* n'ayant pas de définition communément acceptée en français (vie privée, confidentialité...), nous proposons d'utiliser la notion de *sphère privée* pour faire référence aux problématiques soulevées. Du fait du grand nombre des interactions au sein des systèmes multi-agents, les risques encourus par les données sensibles deviennent de plus en plus importants en termes de divulgation, d'altération etc... Nos préoccupations portent essentiellement sur les communications de données sensibles et sur leur devenir après leur diffusion. Afin de prendre en considération les questions primordiales pour la préservation de la sphère privée, nous proposons le modèle de Systèmes Multi-Agents Hippocratiques (HiMAS) qui définit un cadre dans lequel les agents ont la capacité de gérer les données sensibles des utilisateurs et de protéger ces données contre des comportements suspicieux à l'aide d'un protocole d'interaction spécifique et de mécanismes de régulation de comportement tels que la confiance et la réputation. Notre approche se fonde sur un aspect moral et éthique afin de venir compléter les nombreuses propositions sur ces problématiques dans les domaines de la sécurité et des réseaux.

Mots clé Systèmes Multi-Agents, Sphère Privée, Interactions, Protocole de Communication, Contrôle Social, Confiance, Réputation.

Abstract

With the explosion of the Web and multi-agent technologies, the current evolution of information systems leads to an automatic processing of users' data. The development of user centered multi-agent systems brings a new research topic : the management and the protection of users' sensitive data in order to preserve privacy. This thesis focuses on privacy management coming from the user's delegation to an agent of his sensitive data. Interaction between agents being one of the main feature of a multi-agent system, the possible risks for the sensitive data become more and more important in terms of disclosure, alteration etc... This thesis primarily focuses on sensitive data communications and on the sensitive data becoming after being sent. In order to consider these important questions about the privacy preservation, we propose the model of Hippocratic Multi-Agent System (HiMAS). This model gives to agent the capacity to manage the sensitive users' data thanks to the notion of *private sphere*, and to protect this kind of data against suspicious behavior thanks to a specific interaction protocol and some mechanisms for the regulation of the agent behavior as trust and reputation. Our approach is based on a moral and ethic focus in order to assist the many propositions on these problems in security and network research.

Key words Multi-Agent System, Privacy, Interaction, Communication Protocol, Social Order, Trust, Reputation.