
De l'intérêt du couplage de modèles pour appréhender les interactions utilisateurs - réseaux dynamiques.

Julien Siebert — Laurent Ciarletta — Vincent Chevrier

LORIA
Campus scientifique
54506 Vandoeuvre-lès-Nancy Cedex

RÉSUMÉ. Cet article s'inscrit dans le cadre de l'étude des systèmes complexes via la modélisation et la simulation informatique. Nous pensons qu'il est parfois nécessaire de faire interagir plusieurs modèles pour simuler un phénomène. Dans le cas des technologies des réseaux dynamiques (réseaux P2P, réseaux mobiles Ad Hoc), le comportement des usagers et le fonctionnement des réseaux s'influencent mutuellement. Nous proposons une approche de modélisation et un outil de simulation couplant un modèle d'utilisateurs - basé sur le paradigme multi-agents - et un modèle de réseaux dynamiques. Nous discutons, au travers d'un cas d'étude particulier, des avantages, des problématiques soulevées et des limitations d'une telle approche. Nous montrons que cette démarche de modélisation apporte un niveau de précision et une flexibilité élevés.

ABSTRACT. This article deals with modeling and computer simulations of complex systems. We think that, sometimes, in order to understand and predict some phenomena, the interaction of different models is needed. In the case of dynamic networks (P2P networks, Ad Hoc mobile networks), users' behaviour and networks' performances influence each other. We propose to couple a model of user - based upon multiagent paradigm - with a model of network, and develop a simulation tool. Within a particular case study, we will discuss advantages, limitations and new problematics raised by this approach. This method brings more precision and flexibility than classical analytic models.

MOTS-CLÉS : Modélisation, simulation, systèmes multi-agents, réseaux dynamiques, systèmes pair-à-pair

KEYWORDS: Modeling, simulation, multiagent systems, dynamic networks, peer-to-peer systems

1. Introduction

1.1. *Qu'appelle-t-on réseaux dynamiques ?*

Les réseaux dynamiques sont en théorie des réseaux dont la topologie va évoluer au fil du temps. En pratique, le terme réseaux dynamiques désigne une classe de technologies qui va, pour les plus usités, des réseaux de partage P2P¹ (comme eDonkey, KAD ou encore BitTorrent), aux technologies plus émergentes telles que les réseaux mobiles Ad Hoc (sans infrastructure pré-établie).

L'aspect dynamique peut être dû, par exemple, à la mobilité des nœuds. C'est le cas notamment dans les réseaux mobiles Ad Hoc, où les usagers du réseau se déplacent, modifiant ainsi physiquement la topologie. La dynamique du réseau peut également venir de l'ajout et/ou de la suppression de liens entre les différents nœuds. Ainsi, dans les réseaux P2P, les usagers peuvent établir de nouveaux liens directs entre eux pour, par exemple, effectuer des téléchargements. Une fois les téléchargements effectués, ces liens peuvent être détruits. Ces liens sont dits virtuels car la topologie physique n'est pas modifiée.

Enfin, les réseaux dynamiques sont des systèmes ouverts, c'est-à-dire que les utilisateurs peuvent se connecter ou se déconnecter à n'importe quel moment.

1.2. *L'impact de l'utilisateur*

Les aspects dynamiques et ouverts de ces réseaux introduisent de nouveaux défis pour l'ingénierie et l'étude de tels systèmes. Cependant, il existe un autre aspect dans ces réseaux qui nous intéresse plus directement : l'influence du comportement des utilisateurs.

En effet, le fonctionnement de ces systèmes est conditionné, en partie, par le comportement que peuvent avoir les usagers. Nous avons mentionné précédemment que le mouvement des usagers des réseaux mobiles Ad Hoc modifiait la topologie, donc le routage des messages. Dans le cas des réseaux P2P, chaque utilisateur est à la fois client et fournisseur d'un service. Si les utilisateurs partagent, ils contribuent au système en augmentant la disponibilité des ressources et rendent le réseau plus robuste et attrayant.

Inversement, le fonctionnement de ces réseaux peut influencer, en partie, le comportement des usagers. Prenons les réseaux Ad Hoc, si la connexion est mauvaise, les usagers peuvent bouger de manière à améliorer cette dernière. De la même manière, dans les réseaux P2P, si le nombre de sources pour un fichier est trop faible ou si le temps de téléchargement devient trop long, l'utilisateur peut se déconnecter et se reconnecter plus tard.

1. P2P : abréviation anglaise de pair-à-pair.

1.3. Objectifs et problématiques

Il nous semble intéressant de prendre en considération le comportement des usagers lors de l'ingénierie ou de l'étude des réseaux dynamiques. Nous souhaitons étudier les *influences mutuelles* entre comportement des usagers et le fonctionnement des réseaux. Notre objectif à long terme est, en premier lieu, de pouvoir comprendre ces interactions. Ensuite, nous souhaitons être capable d'évaluer le fonctionnement du réseau sous l'influence de différents comportements. Enfin, nous souhaitons être capable de prédire le comportement des usagers en fonction de métriques issues du réseau.

Dans la pratique, il est difficile d'effectuer des expériences sur les réseaux déployés qui contiennent un nombre d'usagers très important (de l'ordre de 10^6 utilisateurs). D'un point de vue technique, l'infrastructure nécessaire à l'observation des systèmes, à la fois réseau et usagers, est coûteuse à mettre en place. D'une part, l'observation du fonctionnement du réseau nécessite soit la collaboration d'un fournisseur d'accès, généralement peu enclin à dévoiler des informations sur son réseau, soit la mise en place de sondes et de bases de données conséquentes. D'autre part, observer le comportement des usagers nécessite d'effectuer des sondages. C'est une opération qui est aussi très coûteuse en temps et peut se révéler biaisée, notamment dans le cas des réseaux P2P souvent associés à des pratiques illégales. Enfin, du point de vue juridique, se posent à la fois le problème du respect de la vie privée des usagers et la légalité de la méthode de collecte.

De notre point de vue, nous pensons que la simulation numérique est une bonne solution pour cette étude. De plus, des travaux ont déjà été menés à la fois dans les domaines de la modélisation du comportement humain et dans la modélisation du fonctionnement des réseaux dynamiques. Cependant, à notre connaissance, aucun n'a cherché à mettre en relation ces deux domaines de manière à obtenir un niveau de précision suffisamment élevé à la fois au niveau des modèles comportementaux et des modèles de réseaux dynamiques utilisés. Afin de poursuivre nos objectifs énoncés précédemment, nous pensons qu'une étape importante dans notre travail est d'arriver à tenir compte dans nos simulations à la fois d'un modèle d'utilisateur et d'un modèle de réseau dynamique.

Dans cet article, nous décrivons un cas d'étude dans les réseaux P2P qui pose la question de l'influence mutuelle entre le fonctionnement du réseau et le comportement des utilisateurs. Après avoir montré les limites des travaux existants, nous proposons une approche de modélisation qui associe un modèle d'utilisateur basé sur le paradigme multi-agents avec un modèle de réseau P2P. Cette approche de modélisation est déjà utilisée dans le domaine de l'écologie. Cependant, elle est, à notre connaissance, originale pour le domaine des réseaux dynamiques. Ensuite, nous présentons l'implantation effectuée qui réutilise un simulateur de réseaux P2P existant. Enfin, nous illustrons les avantages de notre approche au travers de plusieurs expériences et discutons des limites et nouvelles problématiques que ces dernières ont mises à jour.

2. Description du cas d'étude : pollution et partage au sein des réseaux P2P

Dans cette partie, nous décrivons le cas d'étude qui va nous servir à expliciter notre démarche. Nous nous plaçons dans le domaine des réseaux de partage P2P. Dans la pratique, un grand problème dans ces réseaux est le risque lié au manque de partage des usagers et au partage de fichiers dits pollués. Nous proposons d'étudier ces phénomènes car ils font tous deux intervenir à la fois le comportement des usagers et le fonctionnement des réseaux P2P.

2.1. Le manque de partage dans les réseaux P2P

Les réseaux P2P de partage de ressources présentent de nombreux avantages face à l'architecture classique client-serveur. En effet, ces systèmes sont peu coûteux, tolérants aux pannes et passent aisément à l'échelle (jusqu'à plusieurs millions d'utilisateurs). Ces avantages sont dus à un concept fort : *la décentralisation des services*. Chaque utilisateur est à la fois client et fournisseur d'un service. Donc en principe, pour que le réseau fonctionne correctement, chaque utilisateur doit partager ses ressources. Or, en réalité, les mesures montrent que seulement une minorité d'utilisateurs partagent (Adar *et al.*, 2000, Hughes *et al.*, 2005, Aidouni *et al.*, 2008).

Ce phénomène, appelé *free-riding*, pose le problème suivant : le risque pour le réseau est de devenir moins robuste face aux pannes et aux déconnexions. En effet, si le nombre d'utilisateurs qui partagent des fichiers diminue, cela a pour effet de "recentraliser" le réseau sur ces derniers. C'est-à-dire que les pairs altruistes sont plus sollicités et fournissent leur service à beaucoup plus d'utilisateurs. Cela accroît le risque de panne ou de déconnexions de ces pairs. D'une manière générale, moins les utilisateurs partagent et moins le réseau est robuste.

2.2. La pollution dans les réseaux P2P

Un fichier est pollué quand son contenu ne correspond pas aux attentes initiales de l'utilisateur. La pollution peut être de plusieurs formes. Tout d'abord, les métadonnées (comme le titre, l'auteur, etc.) peuvent ne pas correspondre au contenu du fichier. C'est typiquement le cas lors du téléchargement de films lorsque l'utilisateur s'aperçoit que le titre du fichier ne correspond pas aux images visionnées. Ensuite, le contenu lui-même peut être pollué : il peut manquer une partie de l'information ou cette dernière peut être volontairement dégradée. C'est le cas lorsque des blancs ou du bruit apparaissent dans un fichier musical (type mp3 par exemple). Les mesures effectuées sur les réseaux P2P ont apporté la preuve que la pollution était très répandue pour les fichiers populaires, plus d'une copie sur deux était polluée sur le réseau KaZaA en 2005 (Liang, 2005).

Ce phénomène a pour effet de diminuer la visibilité d'un fichier sain. Ce faisant, le risque pour le système P2P est, d'une part de perdre des usagers excédés par une mau-

vaise qualité de service. La pollution fut une des principales causes de la désaffection des utilisateurs pour le réseau KaZaA. D'autre part, le risque est de voir les utilisateurs recommencer de nouvelles recherches et de nouveaux téléchargements pour cette donnée. Ceci génère du trafic qui aurait pu être évité.

2.3. Intérêt du cas d'étude

Dans la pratique, il existe des méthodes d'incitation au partage. Cependant, au vu des mesures effectuées sur les réseaux, le *free-riding* est toujours très présent. En ce qui concerne la détection de la pollution, cela dépend de l'utilisateur. Il faut que ce dernier exploite les données échangées pour se rendre compte de la présence de pollution. En effet, il est difficile de détecter automatiquement un fichier pollué. Aujourd'hui, les phénomènes de manque de partage et de pollution inhérents aux systèmes P2P posent des problèmes relativement difficiles à résoudre.

Ensuite, ces deux phénomènes s'influencent l'un et l'autre. En effet, dans le cas où les utilisateurs partagent beaucoup, si la pollution est présente, le risque est de voir la pollution se diffuser. Dans ce cas, la qualité de service va devenir mauvaise. Ainsi, le nombre d'utilisateurs qui partagent risque de diminuer. Ce faisant, le risque de diffuser de la pollution diminue (voir figure 1).

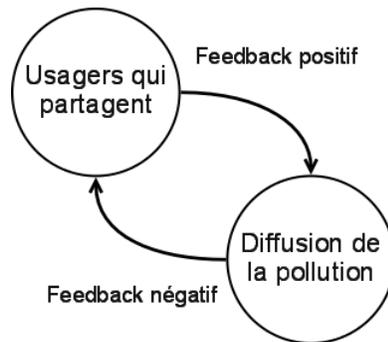


Figure 1. En plus de faire intervenir le comportement des usagers et le fonctionnement du réseau, les phénomènes de *free-riding* et de pollution peuvent s'influencer l'un et l'autre.

Nous pensons que pour comprendre et prédire ces deux phénomènes, il nous faut tenir compte à la fois du comportement des usagers et du fonctionnement du réseau. Dans la partie suivante, nous décrivons les approches de modélisation existantes et leurs limites.

3. Discussion sur les approches de modélisation existantes

Les problèmes liés au comportement des usagers des réseaux dynamiques ont déjà donné lieu à plusieurs études. Que ce soit dans les réseaux P2P ou Ad Hoc mobiles. À notre connaissance, les travaux existants se focalisent toujours sur un seul domaine (modèles d'usagers ou modèles de réseaux). Dans cette partie, nous discutons des avantages et limites des différentes approches. Nous argumentons en faveur d'une approche pluridisciplinaire liant à la fois les deux domaines.

3.1. Les modélisations globales du comportement des usagers des réseaux dynamiques

La première approche de modélisation qui est utilisée pour décrire les réseaux dynamiques est l'utilisation d'un modèle analytique (basé sur des équations) décrivant le système dans sa globalité.

Une mise en équation du phénomène de free-riding est proposée par (Feldman *et al.*, 2004). Dans ce modèle, l'utilisateur décide s'il partage en fonction de la relation entre son type (sa générosité) et le coût de sa contribution. Chaque utilisateur peut avoir un type différent. Cependant, le nombre d'usagers connectés au réseau et leurs comportements sont figés dans le temps ce qui en pratique n'est pas le cas. D'autres travaux s'intéressent au partage et à la pollution des données (Dumitriu *et al.*, 2005, Lee *et al.*, 2006, Kumar *et al.*, 2006). Leurs modèles se basent sur une approche itérative qui calcule le nombre d'utilisateurs qui partagent et le taux de pollution dans le temps. Cependant, chaque utilisateur possède le même comportement et, comme dans le modèle précédent, le nombre d'usagers et leurs comportements restent figés dans le temps. De la même manière, (Thommes *et al.*, 2006) propose un modèle épidémique de propagation de virus au sein des réseaux P2P.

La principale limite de ces approches est la suivante : il est difficile de modéliser des utilisateurs hétérogènes et dont le comportement évolue au fil du temps. En effet, ce type d'approche analytique modélise directement le fonctionnement global du système. Ce faisant, les phénomènes menant à ce fonctionnement global sont rarement explicités. Dans notre démarche, nous souhaitons, en plus d'obtenir ce genre de résultats globaux, pouvoir mesurer l'impact de certains comportements ou de paramètres du réseau sur l'ensemble du système. *Comme, par exemple, quel serait l'impact, en terme de charge du réseau (nombre de messages, temps de transmission) d'un pair qui, au début contribue beaucoup au réseau (reste connecté longtemps et partage de nombreux fichiers) et par la suite décide de ne plus rien partager ?* Pour ce type de résultats, on s'aperçoit que ces approches multi-agents ou centrées individu sont mieux adaptées (Van Dyke Parunak *et al.*, 1998, Grimm, 2005).

Les mêmes limites apparaissent dans le cas des réseaux mobiles Ad Hoc. En effet, les seules modélisations faisant intervenir le comportement des usagers sont celles des déplacements de ces derniers. Les mouvements des utilisateurs sont, là aussi, majori-

tairement décrits par des modèles analytiques globaux (Camp *et al.*, 2002). Dans ces derniers, le nombre d'usagers connectés reste identique ce qui en pratique n'est pas le cas. Comme précédemment, il est difficile de représenter au sein de ces modèles des comportements d'usagers hétérogènes et variable dans le temps.

3.2. Modélisations centrées individus (ou multi-agents)

Un modèle multi-agents (Ferber, 1997, Bousquet *et al.*, 2002) permet de décrire un système sous forme d'un ensemble d'entités (les agents) situés dans un environnement, interagissant entre elles et avec ce dernier. Les agents sont considérés comme

- autonomes : ils possèdent des buts à accomplir et sont proactifs,
- flexibles : ils réagissent aux changements de leur environnement,
- sociaux : ils sont capable d'interagir pour atteindre leur buts,
- situés et locaux : ils sont situés dans leur environnement et ne peuvent le percevoir que localement.

Dans cette approche, la dynamique globale d'un système, au niveau macroscopique, n'est pas donnée d'avance (à l'inverse des modèles précédemment cités) mais résulte de l'interaction des comportements de chaque agent au niveau microscopique. Les intérêts de cette approche sont de pouvoir représenter et simuler des systèmes ouverts ; de prendre en compte, dès la modélisation, les caractères dynamiques et hétérogènes des comportements individuels ; et de pouvoir analyser l'impact d'un comportement local sur le fonctionnement global du système (Van Dyke Parunak *et al.*, 1998).

Les réseaux informatiques sont considérés comme un champ applicatif privilégié pour les systèmes multi-agents (Rupert *et al.*, 2008b, Moro *et al.*, July, 2002, Rupert *et al.*, 2008a). Pour notre part, nous nous concentrons sur les travaux liés aux réseaux P2P.

Le paradigme multi-agents nous permet, tout d'abord, de prendre en compte des comportements hétérogènes et variables dans le temps. Prenons l'exemple des réseaux P2P, dans ce cas, le paradigme multi-agents nous permet de modéliser aisément différentes catégories d'usagers qui partagent plus ou moins leurs ressources (de l'altruiste complet jusqu'au free-rider). Des travaux dans ce sens ont été menés par (Morge *et al.*, 2007, Hales, 2004, Levointurier, 2006). Dans ces derniers, les auteurs utilisent le paradigme multi-agents et la théorie des jeux pour modéliser les usagers des réseaux P2P, créer et tester des mécanismes d'incitation au partage.

De la même manière des travaux ont été menés, dans le domaine des réseaux mobiles Ad-Hoc, pour modéliser les déplacements des usagers (Borrel *et al.*, 2006, Baumann *et al.*, 2008). Ces modèles sont ensuite utilisés pour évaluer ou construire des nouvelles technologies de réseaux mobiles. L'utilisation du paradigme multi-agents permet de générer facilement des modèles de comportements d'usagers plus réalistes que les approches classiques (voir la comparaison effectuée dans (Baumann *et al.*,

2008)). Il est possible, par exemple, de facilement décrire des usagers qui se déplacent en groupe, qui évitent les obstacles, suivent des chemins, etc.

Cependant dans ces travaux, les hypothèses faites sur le fonctionnement du réseaux sont trop restrictives. En effet, les travaux précédents se focalisent sur le comportement des usagers et négligent bien souvent les caractéristiques du réseau. Par exemple, dans la pratique, la topologie est donnée par les protocole P2P (KAD, Gnutella, etc.). Or ce sont ces derniers qui permettent à un pair de connaître l'ensemble des usagers du réseau qui offrent un service (des fichiers par exemple). Dans la littérature, les auteurs considèrent bien souvent une distribution uniforme. Cela veut dire qu'un pair à autant de chance de connaître les services offerts par ces propres voisins que ceux offerts par des usagers à l'autre bout du monde. Cette hypothèse peut se relever pertinente pour un réseau comme KAD, basé sur une table de hachage distribuée et un algorithme de routage efficace. Cependant, cela n'est plus le cas pour un réseaux de type Gnutella dans lequel les recherche sont effectuées par inondation, en passant les requêtes de proche en proche. De cette manière, un usager à plus de chance de connaître les services de ses propres voisins. De même, il n'est pratiquement jamais fait mention de latence, de nombre de messages envoyés. Cela limite les applications pratiques des travaux précédents.

3.3. *Les travaux de modélisation des réseaux dynamiques*

Nous souhaitons, d'une part, mesurer l'influence d'un type de comportement sur le fonctionnement du réseau et, d'autre part, que ce fonctionnement influence l'utilisateur. Il nous faut donc prendre en compte certaines caractéristiques de la technologie utilisée.

Nous nous sommes donc intéressés aux modélisations des technologies des réseaux dynamiques. Il existe plusieurs granularités pour représenter les échanges d'information au sein de ces réseaux. Un outil comme NS2² permet de modéliser jusqu'à l'envoi de chaque datagramme sur le réseau, alors que des outils plus spécifiques aux protocoles P2P ou Ad Hoc modélisent l'échange de messages protocolaires (Naicken *et al.*, 2006, Kurkowski *et al.*, 2005).

Les problématiques dans ce domaine sont souvent axées sur les performances des protocoles (montée en charge, passage à l'échelle). Un des reproches que la communauté des réseaux informatiques fait souvent aux simulations est leur manque de réalisme. Nous pensons, pour le cas des technologies des réseaux dynamiques, qu'il faut tenir compte d'un modèle d'utilisateur. Or, à notre connaissance, dans la plupart des outils de simulation réseau, le comportement de l'utilisateur est souvent mis de côté ou n'est pas assez réaliste.

2. <http://www.isi.edu/nsnam/ns/>

3.4. Synthèse

Nous hypothèse de travail est que, pour étudier l'interaction entre le manque de partage et la pollution dans les réseaux P2P, il faut modéliser à la fois le comportement des usagers et le fonctionnement du réseau. Les approches de modélisation disponibles nous apparaissent trop limitées si elle sont prises séparément. De plus, il nous paraît difficile et peu opportun d'unifier ces deux points de vue au sein d'un même modèle. Nous pensons qu'il nous faut faire interagir plusieurs modèles (un *modèle comportemental* et un *modèle de réseau*).

Nous proposons, dans la section suivante, une approche de modélisation qui lie à la fois un modèle d'utilisateur basé sur le paradigme multi-agents et un modèle de technologie de réseau dynamique. La première étape de notre travail a été de coupler ces différents modèles grâce au paradigme multi-agents au sein d'un même outil de simulation. La deuxième étape de notre travail a été d'explorer le cas d'étude, au travers d'expériences, afin de montrer le champ d'action de notre approche. Dans les parties suivantes, nous décrivons ce cas d'étude et présentons les expériences menées.

4. Une approche de modélisation basée sur le paradigme multi-agents

Dans cette partie, nous proposons d'utiliser le paradigme multi-agents pour faire interagir et intégrer à la fois un modèle d'utilisateur et un modèle de réseau dynamique.

4.1. Principe de l'architecture

L'idée de base est de mettre en parallèle les différents composants des réseaux dynamiques avec ceux issus du paradigme multi-agents tels qu'évoqués dans différentes approches (Demazeau, 1995, Ferber, 1997). Cette approche nous permet d'avoir à la fois les avantages de la modélisation multi-agents (hétérogénéité et la variabilité des comportements), ainsi qu'un modèle réaliste de réseau (protocoles, topologie, latences, etc).

Pour ce faire, nous avons utilisé un modèle d'agent pour décrire le comportement de l'utilisateur, un modèle d'interactions représente le réseau P2P et un modèle d'environnement correspondant au réseau physique sous jacent (Siebert, 2007). Les parties suivantes décrivent et argumentent les choix proposés. Le tableau 1 en donne les grandes lignes.

4.2. Modèle de comportement des usagers

Notre cas d'étude (le manque de partage et la diffusion de la pollution ; voir section 2) à fait l'objet de travaux (voir section 3). Nous nous sommes inspirés des modèles existants pour définir les comportements des usagers. Ainsi, nous considérons la vo-

Paradigme Agent	À quelle question répond le composant du paradigme multi-agents ?	Réseau dynamique
Agents	Quelles sont les entités autonomes et proactives du système ?	Utilisateurs
Interactions	Par quels moyens interagissent les entités du système ?	Applications et Protocoles
Environnement	Quel est le médium des interactions ? Dans quel environnement sont situées les entités ?	Réseau physique sous-jacent

Tableau 1. *Le paradigme multi-agents utilisé pour décrire les technologies des réseaux dynamiques*

lonté de partager³ et la conscience de la pollution qui correspond au fait qu'un usager détecte un fichier quand il est pollué.

Ces comportements ne font pas obligatoirement appel à des schémas comportementaux très développés. En effet, nous faisons l'hypothèse, dans une première approche, que la volonté de partager ne dépend que d'un état interne (la générosité) et de la perception du coût de contribution (la consommation de la bande passante). La conscience de la pollution quant à elle peut être vue comme un simple état interne. Ensuite, la représentation qu'a un usager de ses "voisins" et du fonctionnement du réseau peut être simplifiée. En effet, dans la pratique, il ne peut percevoir et interagir sur le réseau qu'au travers de son client. Enfin, la taille des réseaux que nous souhaitons modéliser est relativement élevée (entre 10^3 et 10^6 utilisateurs). Ces trois hypothèses nous poussent, dans un premier temps, à limiter la complexité du modèle comportemental. Nous avons donc décidé de nous baser sur une architecture d'agent réactif pour modéliser le comportement des utilisateurs (Siebert *et al.*, 2008b).

4.2.1. *Détail du modèle comportemental*

Le modèle comportemental est construit sur une boucle perception-décision-action (voir figure 2). La décision s'effectue à partir des perceptions de l'agent et des paramètres décrivant le modèle comportemental.

Les **perceptions** correspondent aux quelques informations que l'utilisateur i peut avoir sur le réseau :

- Le *débit descendant* ($download_i$), la bande passante consommée par le téléchargement de ressources.
- Le *débit montant* ($upload_i$), la bande passante consommée par l'offre de ressource.

3. Le free-riding correspond à une volonté de partager qui est nulle.

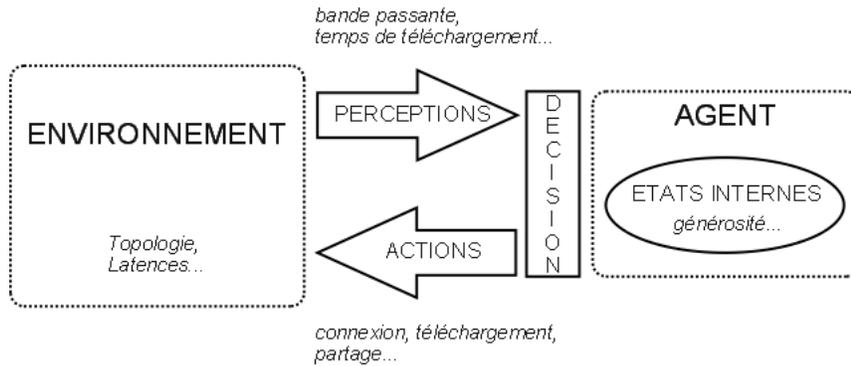


Figure 2. Boucle perception-décision-action dans le cas d'étude du manque de partage et de la pollution dans les réseaux P2P.

- Le nombre de sources disponibles n après un certain temps d'attente et qui caractérise la rareté (ou la disponibilité) d'une ressource.
- Le temps estimé (Δ) de téléchargement d'un fichier.

Les **actions** correspondent à ce que peut faire l'utilisateur lors de son activité de téléchargement :

- Connexion/déconnexion au réseau P2P.
- Recherche d'une ressource donnée.
- Lancement/Arrêt de la procédure de *téléchargement*.
- Partage ou non des ressources (ou d'une certaine quantité).
- Contrôle du taux de pollution de certaines données.

Enfin les **états internes** déterminent la manière dont sera prise une décision et donc sont en relation directe avec le type de comportement que nous souhaitons modéliser :

- La *volonté de partager* α_i (générosité). Cette notion est issue de l'économie (voir l'utilisation faite dans (Feldman *et al.*, 2004)). Elle exprime la probabilité qu'un agent i de partager ses ressources.
- La *conscience de la pollution*, évoquée dans (Lee *et al.*, 2006). β_i est la probabilité que l'agent i détecte la pollution sachant que le fichier est corrompu.

4.2.2. Décision

Les fonctions de décision de l'agent sont inspirées des modèles développés dans (Feldman *et al.*, 2004, Lee *et al.*, 2006). Nous avons adapté ces modèles à notre approche multi-agents et ajouté de nouveaux paramètres. Nous considérons qu'un agent va partager ses ressources principalement au regard de deux paramètres : sa généro-

sité α_i et le coût de contribution c_i qui est fonction des perceptions de l'agent. L'agent décide de partager si sa générosité est grande et si le coût de contribution est faible.

Un agent i détecte la pollution au sein d'un fichier en fonction de sa conscience β_i . De cette manière, un agent dont la conscience $\beta = 1$ va automatiquement détecter un fichier pollué et va agir en conséquence. Nous faisons l'hypothèse que l'utilisateur reconnaît toujours un fichier authentique.

4.3. Modélisation des aspects réseaux et implantation

Dans la pratique, pour pouvoir interagir avec les autres utilisateurs, un usager doit se connecter et utiliser un protocole P2P. C'est grâce à l'envoi de messages protocolaires que ce dernier peut connaître les services présents dans le réseau, les utiliser et en fournir.

Il nous paraît nécessaire de prendre en compte quelques caractéristiques du réseau sous-jacent (bande passante, temps de transfert). D'une part, nous souhaitons pouvoir évaluer certaines performances du réseau sous différents scénarios d'utilisation. D'autre part certaines caractéristiques, comme le temps de transfert d'un fichier par exemple, influencent le comportement de l'utilisateur.

Au niveau de la modélisation du réseau, il apparaît que celle-ci ne doit pas nécessairement être trop détaillée. En effet, du fait du nombre important d'utilisateurs et pour étudier la qualité de service du réseau P2P, il ne semble pas nécessaire de modéliser jusqu'à l'envoi de chaque datagramme, voire de modéliser l'ensemble des équipements réseaux. L'envoi de messages au niveau du protocole P2P et la prise en compte d'un protocole réaliste semblent être, en première approche, un bon compromis entre la taille du réseau et la complexité des interactions entre les pairs.

4.3.1. Choix d'un simulateur P2P existant

La volonté de modéliser les aspects réseaux nous a conduit à nous intéresser aux outils de simulation P2P. Ainsi, nous avons pu réutiliser un modèle cohérent de réseau. Après une étude bibliographique des outils disponibles (Naicken *et al.*, 2006, Siebert, 2007), nous avons décidé d'utiliser PeerfactSim.KOM⁴ (PFS). Le principal avantage de cet outil, en plus d'un nombre annoncé de nœuds simulés conséquent (de l'ordre de 10^5 (Darlagiannis, 2005)), provient de son architecture qui fait nettement la séparation entre les différents niveaux d'abstraction décrits précédemment. La figure 3 montre les différents niveaux présents dans le simulateur. Au niveau du réseau physique, le simulateur modélise, entre autres, les temps de latence pour chaque message envoyé. Au niveau du réseau overlay⁵, des protocoles P2P existants (CAN, Omicron, Chord, Kademlia) sont implantés originellement dans le simulateur. Au niveau de l'utilisateur, la partie modélisation du comportement est détachée du nœud physique correspondant.

4. <http://www.peerfact.org>

5. L'overlay est le réseau virtuel composé de tous les pairs liés deux à deux.

Ainsi, un modèle de comportement peut être implanté sans avoir à modifier les parties réseaux.

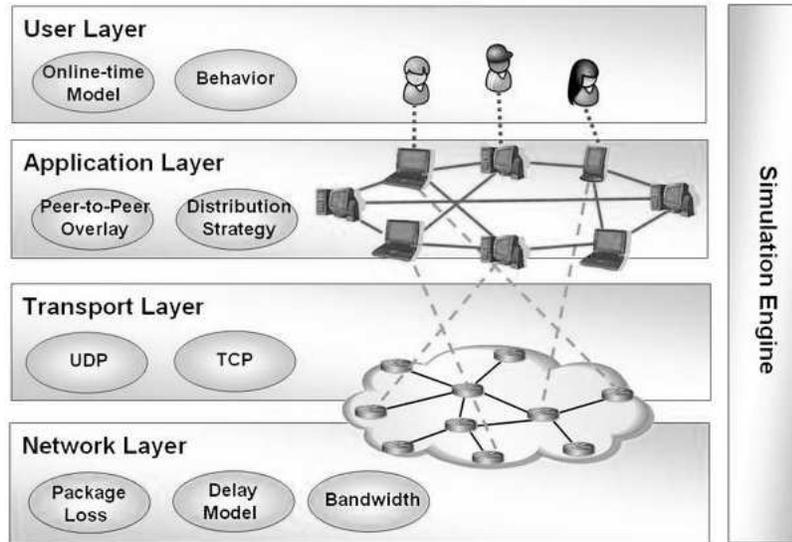


Figure 3. Architecture du simulateur Peerfactsim.KOM d'après <http://www.peerfact.org>

4.3.2. Modifications apportées

Notre problématique d'impact du comportement des utilisateurs implique que ces derniers puissent manipuler les données (les stocker, les partager, les polluer). PFS semblait, pour un simulateur P2P, adapté à nos besoins (de par son architecture). Cependant, le simulateur a été conçu pour répondre aux questions concernant les performances des protocoles P2P. Ainsi, si les protocoles d'indexation des données étaient bien présents, toute la partie transfert concret de données (modélisation, stockage des données, procédure d'échange d'un pair à un autre, etc.) n'était pas modélisée. Nous avons donc apporté quelques modifications au simulateur pour implanter ces aspects (Siebert, 2007) :

- les données sont maintenant modélisées ;
- chaque utilisateur possède désormais une liste de données qu'il partage ;
- les échanges de données entre deux pairs sont dorénavant possibles.

De même, au niveau du modèle d'utilisateur, bien que la couche utilisateur soit présente, aucun modèle de comportement n'y était implanté. Nous avons donc transcrit notre modèle d'utilisateur décrit précédemment au sein du simulateur. Nous posséd-

dons maintenant, un simulateur qui fait interagir un modèle de technologie de réseau dynamique, à savoir un réseau P2P, et un modèle de comportement d'utilisateur.

5. Discussion de l'approche et exemples d'expériences menées

Dans la partie précédente, nous avons proposé une approche de modélisation originale pour l'étude des phénomènes faisant intervenir à la fois le comportement des usagers et les technologies des réseaux dynamiques. Il nous a paru intéressant de valider l'approche conceptuelle et de se comparer aux approches existantes. Pour cela, nous avons choisi d'utiliser un cas d'école et d'illustrer, grâce aux expériences menées sur ce celui-ci, les avantages d'utiliser deux modèles complémentaires : un modèle d'usager basé sur le paradigme multi-agents et un modèle de réseau P2P.

5.1. Description du protocole expérimental

Nous nous intéressons ici au cas d'étude décrit précédemment qui concerne le phénomène de manque de partage et la diffusion de la pollution au sein des réseaux P2P. Nous nous sommes inspirés de (Lee *et al.*, 2006, Feldman *et al.*, 2004, Kumar *et al.*, 2006) pour définir ce protocole expérimental.

5.1.1. Scénario d'une simulation

Nous avons vu précédemment que notre modèle d'utilisateur nous permet de décrire les actions que peuvent effectuer les agents sur le réseau et leur raisonnement. Nous avons aussi à notre disposition plusieurs protocoles de réseaux P2P. Il nous faut ici choisir quel protocole et quel modèle de comportement utiliser lors des simulations. Nous avons choisi d'utiliser le protocole Kademia qui est implanté (sous sa forme Kad) et utilisé, entre autres, par le client eMule⁶.

Pour le comportement des usagers, nous avons défini un scénario de simulation inspiré des travaux de (Lee *et al.*, 2006). Bien que ce premier scénario ne fasse pas intervenir l'hétérogénéité et la variabilité des comportements, il est principalement utile pour comparer avec les approches plus traditionnelles, nous verrons par la suite, qu'il est facile de le modifier pour obtenir des comportements hétérogènes et dynamiques. Le scénario est décrit ci-après et résumé par la figure 4.

Notre réseau est composé de N utilisateurs. Parmi ces N utilisateurs, S sont *égoïstes*⁷. Ce sont les utilisateurs qui ne partageront aucun fichier ($\alpha_i = 0$).

Initialement (au temps $t = 0$), certains pairs publient des données, ce sont les sources initiales. Dans notre cas nous distinguons les *sources saines* (notées G_0 ⁸) qui

6. <http://www.emule-project.net/>

7. S pour l'anglais *selfish*.

8. G pour l'anglais *genuine*.

publient chacune f_{g_0} fichiers sains ; et les sources polluées (notées B_0^9) qui publient chacune f_{b_0} fichiers pollués. Il y a donc $B_0 * f_{b_0} + G_0 * f_{g_0} = F_0$ fichiers publiés initialement. Ensuite, une fois connecté et les ressources publiées, chaque pair suit le scénario suivant¹⁰ :

ÉTAPE 1 (durée maxi = e_1). Recherche des sources disponibles.

La première étape consiste en l'envoi d'une requête sur le réseau. Au temps $t_1 = t_0 + x$, avec $x \sim U(0, e_1)$. Chaque pair obtient une liste l_i de sources disponibles.

ÉTAPE 2 (durée maxi = e_2). Demande de téléchargement.

Au temps $t_2 = t_1 + x$, avec $x \sim U(0, e_2)$, l'utilisateur choisit aléatoirement (de manière uniforme) une source dans la liste construite et envoie une demande de téléchargement.

ÉTAPE 3 (durée maxi = e_3). Vérification du téléchargement.

Au temps $t_3 = t_2 + x$, avec $x \sim U(0, e_3)$, l'utilisateur vérifie que le téléchargement est terminé.

– Si le téléchargement n'est pas terminé. L'utilisateur peut alors relancer toute la procédure (Étape 5).

– Sinon : L'utilisateur passe à l'étape 4.

ÉTAPE 4 (Durée maxi = e_4). Vérifier la pollution et partager.

Au temps $t_4 = t_3 + x$, avec $x \sim U(0, e_4)$, l'utilisateur vérifie la pollution du fichier téléchargé. Sachant que le fichier est pollué, l'utilisateur i a une probabilité β_i de le détecter.

– Si celui-ci est détecté comme pollué, l'utilisateur le supprime et relance la procédure (Étape 5).

– Sinon l'utilisateur i décide de le partager ou non avec une probabilité α_i .

ÉTAPE 5 : relancer la procédure.

L'utilisateur peut relancer la procédure avec une probabilité p_r . Si c'est la cas, celle-ci recommencera après un laps de temps $\Delta_t \sim U(0, e_5)$. À noter que pour des raisons technique, r_{max} est le nombre maximal de fois où l'utilisateur peut recommencer la procédure.

5.1.2. Mesures

L'outil de simulation permettait, de base, d'effectuer des mesures sur le protocole P2P étudié (nombre de messages envoyés, reçus par un pair, nombre de sauts

9. B pour l'anglais *bogus*.

10. On notera $U(a, b)$ la loi uniforme entre a et b .

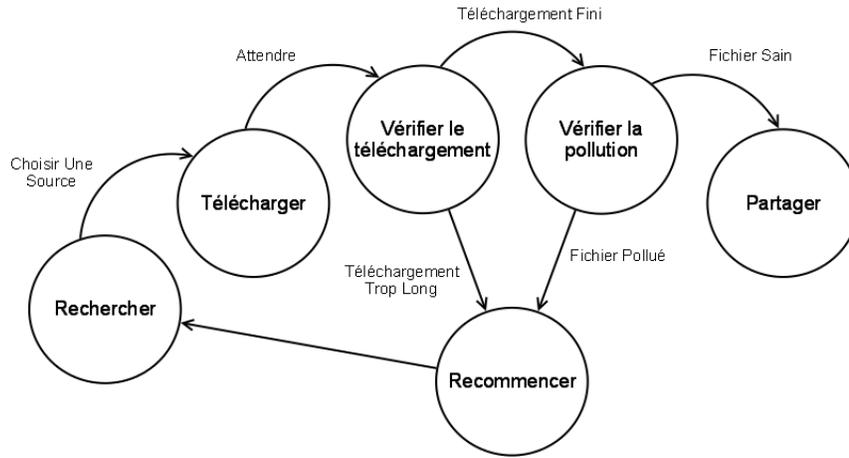


Figure 4. Représentation simplifiée du comportement d'un pair lors d'une simulation.

Utilisateurs	N
Égoïstes	S
Source saines (à $t = 0$)	G_0
Fichiers sains (à $t = 0$)	$G_0 * fg_0$
Sources polluées (à $t = 0$)	B_0
Fichiers pollués (à $t = 0$)	$B_0 * fb_0$
Nombre de fichiers initialement publiés dans le réseau (à $t = 0$)	F_0
Volonté de partager pour un utilisateur i	α_i
Conscience de la pollution pour un utilisateur i	β_i

Tableau 2. Notations des paramètres expérimentaux

avant d'atteindre la destination, nombre de pairs connectés, etc.). Ayant modifié l'outil, nous avons ajouté de nouvelles métriques. Nous décrivons ici les principales qui sont utilisées par la suite.

5.1.2.1. Métriques concernant les données échangées

Parmi les modifications apportées, figure la modélisation des données et de leur échange entre les pairs. De ce fait, nous pouvons savoir, à tout instant, quelles sont les

données partagées, si elles sont polluées ou saines et si elles ont été contrôlées par un usager ou non.

5.1.2.2. Métrique concernant le comportement des pairs

Puisque les comportements des usagers peuvent changer au cours du temps, il nous faut pouvoir connaître l'état des usagers et l'évolution de leur comportement. Pour cela, à tout instant, et pour chaque usager, il nous est possible de mesurer si l'utilisateur est un free-rider ou s'il partage des données, en quel nombre, si ces données sont polluées. Ensuite, nous pouvons connaître les perceptions des usagers, comme le nombre de sources obtenues après une requête.

5.2. Comparaison des approches de modélisation

Dans cette partie, le but recherché est de montrer que la démarche de modélisation permet d'obtenir les mêmes résultats que les autres approches appliquées à ce cas d'étude. Nous pensons que c'est une étape nécessaire avant toute discussion.

Nous comparons nos résultats à ceux du modèle proposé par (Lee *et al.*, 2006). Le but de ces travaux est de montrer l'influence de la volonté de partager et de la conscience de la pollution sur la diffusion des fichiers pollués dans le réseau. Pour ce faire, les auteurs mesurent la proportion de pollution dans le réseau (en fonction de α_i et β_i) quand le système a atteint un état d'équilibre, *i.e.* quand tous les usagers ont effectué leur téléchargement et décidé si oui ou non, ils partagent leurs fichiers.

5.2.1. Protocole expérimental

Dans ces expériences, nous avons cherché à montrer que notre approche peut reproduire les résultats du modèle de référence. Nous avons utilisé des réseaux de tailles N variant de 25 à 50000 nœuds. Comme dans le modèle de (Lee *et al.*, 2006), nous avons mesuré la proportion de fichiers pollués dans le réseau en fonction des paramètres α_i et β_i .

Prenons l'exemple de l'impact du niveau de conscience β_i de chaque utilisateur. Une expérience est décrite par (Lee *et al.*, 2006), dans laquelle, le niveau initial de pollution dans le réseau est de 20%. Les auteurs font varier la conscience qu'ont les usagers de la pollution, β_i (voir figure 5). D'une valeur de 1 (le cas où tous les usagers reconnaissent - et donc suppriment - tous les fichiers pollués) ce paramètre est décrémenté jusqu'à 0,8 (cas où les usagers ne reconnaissent que 80% des fichiers pollués). Cette expérience nous permet de confirmer que le niveau de pollution augmente quand la conscience β_i des usagers baisse et quantifier cette dynamique.

5.2.2. Résultats

Nous avons effectué des expériences du même type. La première différence majeure entre nos résultats et ceux de la littérature provient de la différence de paradigmes utilisés. En effet, le modèle initial est un modèle analytique qui peut être résolu ma-

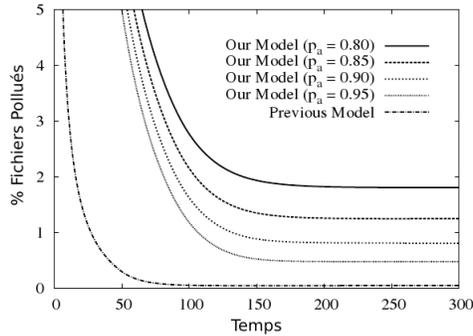


Figure 5. Expérience menée par (Lee et al., 2006). La proportion de pollution augmente quand la conscience des usagers est en baisse

thématiquement. Cela implique que les résultats obtenus sont uniques. De notre côté, le paradigme multi-agents et notre modèle probabiliste, nous obligent à utiliser la simulation numérique. C'est-à-dire que pour un même problème, nous devons faire plusieurs simulations pour obtenir des données statistiquement représentatives. Ceci explique l'aspect bruité de la courbe figure 6 par rapport à l'aspect lisse de la courbe figure 5.

Ensuite, le phénomène modélisé implique une première phase de connexion. Au début de la simulation, les usagers se connectent au réseau P2P. Ceci explique la dispersion des données lorsque le temps de simulation est faible (sur la figure 6 : $t \leq 10^5$).

Enfin, contrairement aux hypothèses faites dans le modèle initial, les usagers qui polluent le réseau initialement (B_0) ne suppriment pas leurs données. Ce qui explique que, dans la figure 6, le niveau final de pollution (pour $t \geq 4 * 10^5$) se stabilise pour $\beta_i = 1$ (la courbe claire) à la valeur B_0 .

5.3. Apport du paradigme multi-agents

Dans cette partie, nous montrons comment, grâce au paradigme multi-agents, nous pouvons facilement modéliser des comportements hétérogènes et variables dans le temps.

5.3.1. Hétérogénéité du comportement des usagers

Le but de cette partie est d'apporter des exemples où le paradigme multi-agents permet de modéliser facilement des comportements hétérogènes et variables dans le temps.

Dans cette partie, nous allons modifier le comportement d'un pair qui publie des fichiers pollués. Jusqu'à présent, nous avons suivi le modèle de (Lee et al., 2006) qui

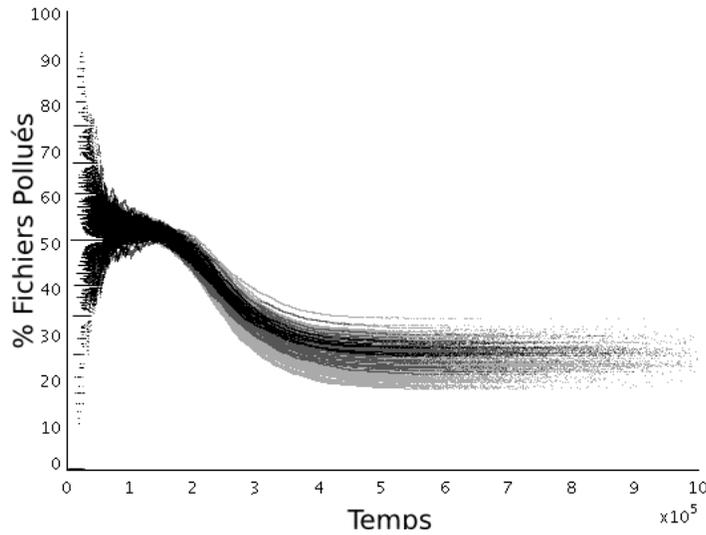


Figure 6. Évolution de la pollution sur un réseau de taille $N = 12000$ en fonction de la conscience qu'ont les usagers de la pollution (β_i). Du plus foncé au plus clair : $\beta_i = 0.8$, $\beta_i = 0.9$ et $\beta_i = 1$. Pour chaque expérience (ou couleur) la courbe est composée de l'ensemble des 100 simulations effectuées.

fait l'hypothèse que chaque utilisateur ne publie et ne télécharge qu'un seul fichier. Nous souhaitons ici faire en sorte qu'un pollueur publie plusieurs fichiers. Nous cherchons simplement à mesurer l'effet de la concentration initiale de fichiers pollués sur la diffusion de l'empoisonnement global. Autrement dit, à mesurer la différence entre une seule source qui publie beaucoup de fichiers pollués et beaucoup de sources qui publient peu de fichiers pollués.

5.3.1.1. Description du protocole expérimental

Nous nous sommes placés dans le cas où tout le monde partage et où tout le monde est conscient de la pollution :

$$\forall i \in [0, N] \alpha_i = \beta_i = 1$$

Nous avons utilisé des tailles de réseau variant de $N = 120$ à $N = 12000$. Le nombre total de fichiers initialement publiés reste constant et est égal à $N/4$. Le nombre initial de fichiers pollués est constant :

$$B_0 * fb_0 = N/10$$

Il s'en suit que le nombre de fichiers sains initialement publiés est également constant :

$$G_0 * fg_0 = N/4 - N/10$$

Puisque nous sommes dans un cas d'école dans lequel tous les utilisateurs partagent et sont totalement conscients de la pollution des données, l'état à l'équilibre est facile à prédire. En effet, tous les utilisateurs (sauf les sources initiales de pollution) vont supprimer les fichiers pollués. Un modèle macroscopique, comme celui de (Lee *et al.*, 2006), nous donnerait directement ce résultat et une seule dynamique pour l'atteindre. Cependant avec le paradigme multi-agents, il est possible de spécifier différentes configurations. Nous avons plusieurs manières d'instancier le nombre initial de fichiers pollués : en faisant varier la concentration initiale de fichiers pollués, soit B_0 et fb_0 (voir l'exemple du tableau 3). De cette manière nous faisons apparaître différentes dynamiques qui partent d'une même hypothèse "globale" (le nombre de fichiers initialement pollués) et atteignent un même état d'équilibre.

B_0	fb_0
1	12
2	6
3	4
4	3
6	2
12	1

Tableau 3. Différents types de concentrations. Exemple avec $N = 120$.

5.3.1.2. Résultats

Pour chaque série, nous avons mesuré la dynamique de la diffusion de la pollution. Le résultat principal qui apparaît est que la pollution semble se diffuser plus facilement si la concentration initiale de fichiers pollués est plus faible. C'est-à-dire que le nombre de sources B_0 est élevé mais qu'elles ne publient chacune qu'un nombre de fichiers pollués fb_0 faible.

La figure 7 illustre l'évolution temporelle de l'empoisonnement global sur un réseau de $N = 12000$ utilisateurs. Sur la courbe du haut, apparaît la quantité de fichiers pollués. Pour un réseau dans lequel 1200 utilisateurs publient chacun initialement un seul fichier pollué (pour les paramètres $B_0 = 1200$ et $fb_0 = 1$), la diffusion est plus rapide que pour un réseau où seulement 600 usagers publient chacun initialement 2 fichiers ($B_0 = 600$, $fb_0 = 2$). Le cas le plus lent, est un réseau dans lequel un seul utilisateur publie 1200 fichiers pollués ($B_0 = 1$, $fb_0 = 1200$).

Ce résultat peut s'expliquer relativement simplement par deux effets. Le premier est le *mode de diffusion*. Lorsqu'une seule source publie les fichiers pollués, elle doit le faire séquentiellement avec entre chaque publication un certain laps de temps. À l'inverse, lorsque plusieurs sources publient des fichiers pollués, elles le font en parallèle. Dans ce cas la publication et l'indexation des données sont plus rapides. Le deuxième point provient du *protocole utilisé* (Kademlia) : la recherche et l'indexation sont basées sur une table de hachage distribuée (Maymounkov *et al.*, 2002) ce qui a pour effet de donner autant de visibilité à chaque utilisateur. Ainsi, lorsque le nombre

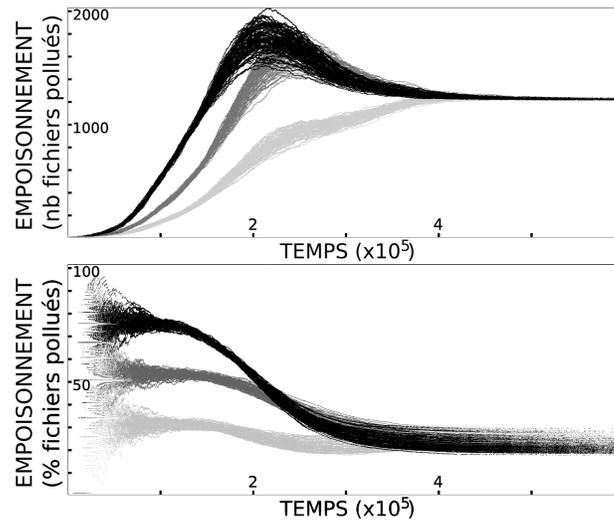


Figure 7. Évolution de la pollution sur un réseau de taille $N = 12000$. En haut : le nombre de fichiers pollués en fonction du temps. En bas : le pourcentage de pollution en fonction du temps. Du plus foncé au plus clair : $B_0 = 1200, fb_0 = 1$; $B_0 = 600, fb_0 = 2$ et $B_0 = 300, fb_0 = 4$

de sources polluantes (B_0) augmente, la probabilité de réplique de fichiers pollués est plus importante. Donc la probabilité pour un utilisateur de télécharger un fichier corrompu est plus grande et la diffusion de l'empoisonnement plus rapide.

5.3.2. Un comportement réactif à son environnement

Nous cherchons ici à illustrer comment le paradigme multi-agents nous permet de modéliser la variabilité des comportements des usagers.

Toujours dans le cas d'école défini précédemment, nous allons modéliser le fait que les usagers peuvent décider s'ils partagent ou non leurs ressources. Nous nous basons sur une modélisation du phénomène de partage effectuée par (Feldman *et al.*, 2004). Dans ces travaux, l'utilisateur, qualifié de rationnel, décide s'il partage en fonction de ce que cela lui coûte. L'hypothèse émise est que le coût du partage est inversement proportionnel à la proportion x de personnes qui partagent des ressources. Le comportement d'un usager ayant un type θ_i (le type décrit sa "générosité") sera donc le suivant :

partage si $\theta_i \geq 1/x$

free – ride sinon

Il est aisé d'introduire dans le scénario de simulation une nouvelle étape, entre l'étape 1 et 2, dans laquelle l'agent va décider si, oui ou non, il contribue au réseau. Pour re-

produire la fonction de décision précédente, il nous faut ajouter un type $\theta_i \in [0, \theta_{max}]$ à l'agent i et une perception x .

Cependant, dans le modèle macroscopique de (Feldman *et al.*, 2004), l'hypothèse émise est que l'utilisateur connaît la proportion x du nombre de personnes qui partagent des données. La contrainte de localité des agents (voir section 3.2) fait qu'une telle hypothèse est inutilisable. En effet, les agents ne peuvent pas connaître la valeur x , puisqu'il faudrait pour cela qu'ils possèdent une connaissance globale sur le système - ce qui n'est pas permis par le paradigme multi-agents. Il faut également souligner que dans la réalité, cette hypothèse est impossible à obtenir.

Il nous faut donc traduire cette variable globale x . Elle représente la proportion des personnes qui partagent des ressources. Dans notre modèle, le seul moyen d'approcher cette donnée est de compter le nombre de sources disponibles après avoir effectué une requête (étape 1 du scénario), soit¹¹ :

$$\bar{x}_i = \text{card}(l_i)$$

Cela ne nous donne pas une proportion mais bien une approximation du nombre de personnes qui partagent à un instant donné.

En tenant compte de ces contraintes et en respectant l'hypothèse de départ, la fonction de décision adaptée de celle de (Feldman *et al.*, 2004) est donc la suivante :

$$\begin{aligned} &\text{partage si } \bar{x}_i \geq \theta_i \\ &\text{free - ride sinon} \end{aligned}$$

5.3.2.1. Protocole Expérimental

Nous avons effectué des expériences en reprenant les mêmes conditions que (Feldman *et al.*, 2004) pour comparer les résultats des deux approches. Nous mesurons l'évolution du nombre d'utilisateurs qui partagent (x).

Nous avons fixé une taille de réseau $N = 15000$ nœuds, le nombre initial d'utilisateurs free-rider $S = 0$ et pas de pollution dans le réseau $B_0 = fb_0 = 0$. Dans ces conditions, nous avons fait varier le nombre de fichiers initialement publiés, *i.e.* $G_0 * fg_0$, sur des valeurs allant de $0.05 * N$ à $0.95 * N$ soit des valeurs de x initiale variant de 5% à 95%. Pour plus d'aisance, nous avons posé $fg_0 = 1$.

5.3.2.2. Résultats

Nos résultats montrent la dynamique de l'évolution du nombre d'utilisateurs qui partagent. Comme (Feldman *et al.*, 2004), le système atteint un état d'équilibre à la fin de la simulation. Cependant, alors que le modèle comparé prédit, pour un θ_{max} donné et quelle que soit la valeur x initiale, un jeu de points d'équilibre fixe, notre modèle nous donne des valeurs différentes de x à l'équilibre.

11. Voir section 5.1.1 pour la définition de l_i .

La figure 8 montre que pour un $\theta_{max} = 10$, le nombre d'usagers qui partagent à la fin de la simulation est différent si le système est initialisé avec 7500 pairs qui publient (figure 8 courbe du haut) ou avec 11250 pairs qui publient (figure 8 courbe du bas).

Ces différences viennent du fait que le modèle analytique ne prend pas en compte les aspects temporels du système. En effet, dans notre démarche, les utilisateurs se connectent, publient leurs données, celles-ci sont indexées par le réseau, puis ils effectuent leur recherche et leur téléchargement. Comme la décision des usagers se fait au fur et à mesure du temps, les premiers à décider, perçoivent moins de sources que les derniers.

5.4. Apport d'un modèle de réseau

Dans cette partie, nous cherchons à évaluer l'impact d'un type de comportement sur le réseau.

5.4.1. Hypothèse

Nous abordons ici, la question de l'impact, au niveau réseau, du nombre de pairs qui publient les données. L'intuition de départ est la suivante : en théorie, plus le nombre de pairs qui publient des ressources au départ est important, moins la charge par utilisateur est élevée (voir la section 5.3.2). En effet, ceci peut s'expliquer par le fait que tout pair qui a téléchargé une ressource devient, dans ce cas, une nouvelle source. De cette manière la diffusion des données est plus rapide et la charge est censée se distribuer sur toutes les sources.

5.4.2. Expériences et résultats

Nous avons effectué des séries d'expériences dont le but est de mesurer le nombre et le type de messages protocolaires en fonction de différents types de comportements. Nous avons aussi mesuré le nombre d'*upload* et de *download* effectués par chacun des agents. Nous avons aussi fait varier le nombre S d'utilisateurs qui ne partagent pas. Ainsi nous voulons montrer que notre intuition est valide, et que notre approche nous permet de faire le lien entre comportements des usagers et fonctionnement du réseau.

Les résultats obtenus confirment effectivement que plus les utilisateurs partagent et plus la charge du réseau se répartit équitablement. Ces résultats sont disponibles dans (Siebert *et al.*, 2008a).

5.5. Synthèse

Dans cette partie nous avons choisi d'illustrer les apports conceptuels de l'utilisation, à la fois, d'un modèle multi-agent et d'un modèle de réseau P2P. L'utilisation du paradigme multi-agents permet de modéliser facilement des comportements hétérogènes et variables dans le temps qui peuvent être influencés par le fonctionnement du

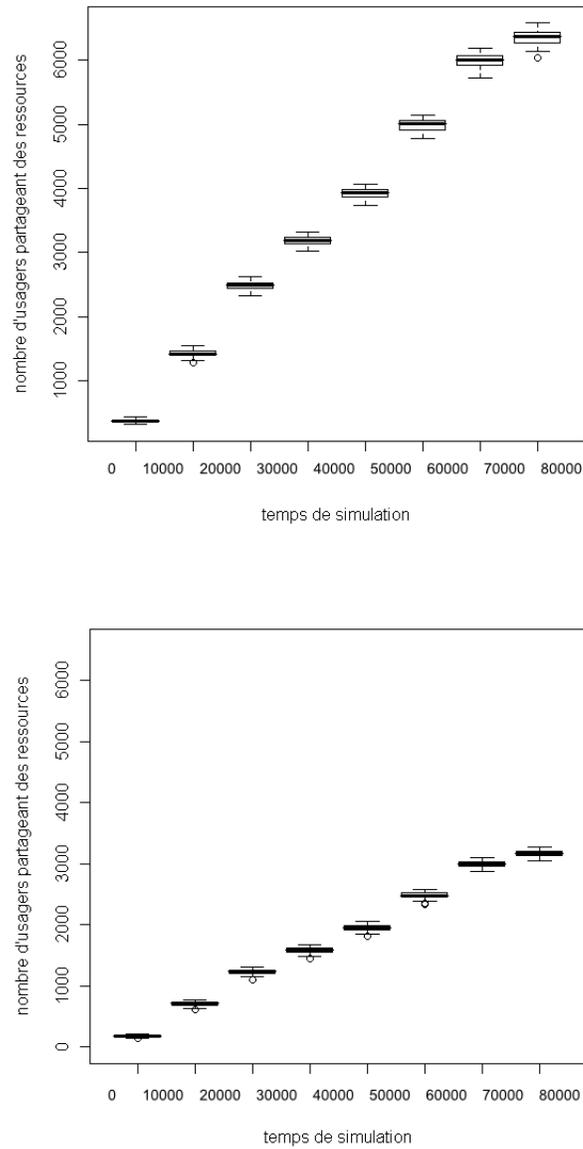


Figure 8. Évolution du nombre d'utilisateurs qui partagent, avec $\theta_{max} = 10$ et $N = 15000$. En haut : initialement le nombre de pairs publiant des données $G_0 = 0.5 * N$. En bas : initialement le nombre de pairs publiant des données $G_0 = 0.75 * N$. L'ensemble de 100 simulations est représenté ici. Pour chaque intervalle de temps de simulation sont représentés la médiane et les quartiles sous la forme d'un diagramme à moustache.

réseau. L'utilisation d'un modèle de réseau permet quant à lui, d'une part, d'influencer le comportement des usagers et, d'autre part d'évaluer le fonctionnement du réseau de manière précise.

Cette approche de modélisation est donc appropriée pour répondre à notre problématique générale qui concerne l'étude des influences mutuelles entre le fonctionnement des réseaux dynamiques et le comportement des usagers.

6. Conclusion

6.1. Bilan

Dans cet article, nous nous intéressons aux phénomènes faisant intervenir le fonctionnement des réseaux dynamiques et le comportement des usagers. Pour pouvoir étudier et prédire ceux-ci, nous pensons qu'il est utile de tenir compte de plusieurs modèles et des les faire interagir. Nous nous sommes intéressés à un cas d'étude particulier lié aux phénomènes de manque de partage et de pollution dans les réseaux P2P.

Nous avons proposé une approche de modélisation et un outil de simulation qui intègrent et font interagir à la fois un modèle d'utilisateur basé sur le paradigme multi-agents et un modèle de réseau P2P. Nous avons illustré les avantages d'une telle approche au travers de différentes expériences.

6.2. Limites

6.2.1. Cas d'école

Le cas d'école choisi permet une comparaison aisée entre les travaux issus de la littérature et notre approche. Cependant, les limites de l'existant, évoquées dans la section 3, nous ont poussés à définir un cas d'école qui, bien que didactique, peut se révéler éloigné de la réalité.

Acquérir des données sur des systèmes P2P réels pose de nombreux défis. Obtenir des données du réseau nécessite une infrastructure (bases de données, sondes) relativement conséquente, voire l'accord d'un fournisseur d'accès. Ensuite, l'obtention de données comportementales est un travail relativement long. Enfin, juridiquement se posent les questions du respect de la vie privée.

Ainsi, il nous a paru plus judicieux d'appliquer notre approche de modélisation à un cas d'école. Des expériences grandeur nature prévues à plus long terme.

6.2.2. Implantation

Nous avons implanté les modèles au sein d'un même outil de simulation. Cette approche de couplage de modèle est dite *forte*. Cette implantation nous a permis de

répondre aux questions que nous nous sommes posés sur notre cas d'étude et d'illustrer cette approche pluridisciplinaire. Nous avons utilisé ce type de couplage car il est plus simple à mettre en œuvre et efficace.

Cependant, les technologies évoluent rapidement dans les réseaux dynamiques. Ensuite les problématiques rencontrées dans les réseaux P2P sont similaires dans les réseaux mobiles Ad Hoc, voire dans certains réseaux de capteurs sans fils. Une approche qui nous semble plus adaptée est le couplage dit *faible*. C'est à dire, qu'il serait plus opportun de coupler différents simulateurs entre eux. Ainsi, il devient facile d'interchanger un modèle pour un autre. Cela permettrait une plus grande réutilisabilité et une plus grande rapidité à modéliser des phénomènes nécessitant plusieurs points de vue.

6.3. Perspectives

6.3.1. Utilisation du couplage faible

Nous souhaitons développer notre approche de modélisation en utilisant un couplage faible. Ainsi, il nous serait plus aisé d'interchanger les modèles et passer, par exemple, du réseau KAD au réseau eDonkey ou des réseaux P2P aux réseaux Mobiles Ad Hoc. Nous pensons que c'est une étape importante avant d'effectuer des expériences grandeur nature et ainsi se confronter à la réalité.

Le passage au couplage *faible* pose des questions aussi bien pratiques que théoriques. Tout d'abord, l'implantation soulève des problèmes intéressants concernant la coordination des différents simulateurs. Ensuite, des problématiques liées à l'interaction de modèles à différentes échelles apparaissent. Ces points ont déjà été évoqués par la communauté des systèmes multi-agents (Bonneaud, 2008). Nous pensons, dans un premier temps, nous diriger dans la même direction afin de voir si les problématiques propres aux réseaux dynamiques sont compatibles avec les méthodes de couplage existantes.

6.3.2. Application aux réseaux mobiles Ad Hoc

Nous voulons appliquer notre démarche aux réseaux mobiles Ad Hoc. Nous pensons que les problématiques sont relativement similaires à celles rencontrées dans les réseaux P2P. Ce champ applicatif nous permettrait, d'une part, d'éprouver notre approche de modélisation. D'autre part, la taille de ces réseaux et la mobilité des nœuds devraient à la fois nous apporter de nouveaux cas d'étude et une plus grande facilité quant à la mise en place d'expériences grandeur nature.

Remerciements

Nous tenons à remercier la Région Lorraine et le projet ANR SARAH (Services Avancés dans les Réseaux Ad Hoc) pour le soutien financier.

7. Bibliographie

- Adar E., Huberman B., Free riding on gnutella, Technical report, Technical report, Xerox PARC, 10 August, 2000.
- Aidouni F., Latapy M., Magnien C., « Ten weeks in the life of an eDonkey server », Sep, 2008.
- Baumann R., Legendre F., Sommer P., « Generic mobility simulation framework (GMSF) », *MobilityModels '08 : Proceeding of the 1st ACM SIGMOBILE workshop on Mobility models*, ACM, New York, NY, USA, p. 49-56, 2008.
- Bonneaud S., Des agents-modèle pour la simulation des systèmes complexes. Application à l'écosystème des pêches., PhD thesis, Université de Bretagne Occidentale., 2008.
- Borrel V., Legendre F., Marcelo, Fdida S., « SIMPS : Using Sociology for Personal Mobility », 2006.
- Bousquet F., Le Page C., Muller J.-P., « Modélisation et simulation multi-agent », *Assises du GDR I3*, Nancy, December, 2002.
- Camp T., Boleng J., Davies V., « A survey of mobility models for ad hoc network research », *Wireless Communications and Mobile Computing (WCMC) : Special issue on Mobile Ad Hoc Networking : Research, Trends and Applications*, vol. 2, p. 483-502, 2002.
- Darlagiannis V., Overlay Network Mechanisms for Peer-to-Peer Systems, PhD thesis, Vom Fachbereich Informatik der Technischen Universität Darmstadt, 2005.
- Demazeau Y., « From interactions to collective behaviour in agent-based systems », *European Conference on Cognitive Sciences*, 1995.
- Dumitriu D., Knightly E., Kuzmanovic A., Stoica I., Zwaenepoel W., « Denial-of-service resilience in peer-to-peer file sharing systems », *SIGMETRICS '05 : Proceedings of the 2005 ACM SIGMETRICS international conference on Measurement and modeling of computer systems*, ACM Press, New York, NY, USA, p. 38-49, 2005.
- Feldman M., Papadimitriou C., Chuang J., Stoica I., « Free-riding and whitewashing in peer-to-peer systems », *PINS '04 : Proceedings of the ACM SIGCOMM workshop on Practice and theory of incentives in networked systems*, ACM Press, New York, NY, USA, p. 228-236, 2004.
- Ferber J., « Les systèmes multi-agents : un aperçu général », *Techniques et sciences informatiques*, vol. 16, n° 8, p. 979-1012, 1997.
- Grimm V. Railsback S., *Individual-based modeling and ecology*, Princeton University Press, 2005.
- Hales D., « From Selfish Nodes to Cooperative Networks Emergent Link-Based Incentives in Peer-to-Peer Networks », *P2P '04 : Proceedings of the Fourth International Conference on Peer-to-Peer Computing (P2P'04)*, IEEE Computer Society, Washington, DC, USA, p. 151-158, 2004.
- Hughes D., Coulson G., Walkerdine J., « Free Riding on Gnutella Revisited : The Bell Tolls ? », *IEEE Distributed Systems Online*, vol. 6, n° 6, p. 1, 2005.
- Kumar R., Yao D. D., Bagchi A., Ross K. W., Rubenstein D., « Fluid modeling of pollution proliferation in P2P networks », *SIGMETRICS Perform. Eval. Rev.*, vol. 34, n° 1, p. 335-346, 2006.
- Kurkowski S., Camp T., Colagrosso M., « MANET simulation studies : the incredibles », *SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 9, n° 4, p. 50-61, 2005.

- Lee U., Choiz M., Choy J., Sanadidiy M. Y., Gerla M., « Understanding Pollution Dynamics in P2P File Sharing », *5th International Workshop on Peer-to-Peer Systems (IPTPS'06)*, Santa Barbara, CA, USA, February, 2006.
- Levointurier C., « Impact du modèle de comportement dans les réseaux pair-à-pair : une approche multi-agents », Master's thesis, Université Henri Poincaré (UHP) Nancy 1, June, 2006.
- Liang J., « Pollution in P2P File Sharing Systems », *IEEE Infocom*, Miami, FL, USA., March, 2005.
- Maymounkov P., Mazieres D., « Kademia : A peer-to-peer information system based on the XOR metric », 2002.
- Morge M., Mathieu P., « Mécanisme de rétribution pour les systèmes P2P d'échange de fichiers », *La revue d'Ingénierie des Systèmes d'Information (ISI), numéro spécial « Les systèmes d'information pair-à-pair »*, vol. 12, p. 61-84, 2007.
- Moro G., Koubarakis M., Sartori C., « Agents and Peer-to-Peer Computing », *First International Workshop, AP2PC 2002*, July, 2002.
- Naicken S., Basu A., Livingston B., Rodhetbhai S., « A Survey of Peer-to-Peer Network Simulators », *Proceedings of The Seventh Annual Postgraduate Symposium, Liverpool, UK*, 2006.
- Rupert M., Li C., Hassas S., « An Organisational Multi-agent Systems Approach for Designing Collaborative Tagging Systems », *IAT*, IEEE, p. 114-117, 2008a.
- Rupert M., Rattrout A., Hassas S., « The web from a complex adaptive systems perspective », *J. Comput. Syst. Sci.*, vol. 74, n° 2, p. 133-145, 2008b.
- Siebert J., Impact du comportement des utilisateurs sur le fonctionnement des réseaux pair-à-pair, Technical report, ESSTIN, October, 2007.
- Siebert J., Chevrier V., Ciarletta L., « Modélisation multimodèle des réseaux dynamiques : cas des réseaux pair-à-pair », *JDIR'08 - 9èmes Journées Doctorales en Informatique et Réseaux*, Villeneuve d'Ascq, France, 2008a.
- Siebert J., Ciarletta L., Chevrier V., « Impact du comportement des utilisateurs dans les réseaux pair-à-pair (P2P) : modélisation et simulation multi-agents », *Journée Francophones des Systèmes Multi-Agents*, October, 2008b.
- Thommes R., Coates M., « Epidemiological Modelling of Peer-to-Peer Viruses and Pollution », *IEEE Infocom*, Barcelona, Spain, April, 2006.
- Van Dyke Parunak H., Savit R., Riolo R. L., « Agent-Based Modeling vs. Equation-Based Modeling : A Case Study and Users' Guide », *MABS*, p. 10-25, 1998.