

Modélisation multimodèle des réseaux dynamiques : cas des réseaux pair-à-pair

Julien Siebert, Vincent Chevrier, Laurent Ciarletta

► **To cite this version:**

Julien Siebert, Vincent Chevrier, Laurent Ciarletta. Modélisation multimodèle des réseaux dynamiques : cas des réseaux pair-à-pair. 9èmes Journées Doctorales en Informatique et Réseaux - JDIR'08, Jan 2008, Villeneuve d'Ascq, France. 2008. <hal-00202453>

HAL Id: hal-00202453

<https://hal.archives-ouvertes.fr/hal-00202453>

Submitted on 7 Jan 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Modélisation multimodèle des réseaux dynamiques : cas des réseaux pair-à-pair

Julien Siebert, Vincent Chevrier et Laurent Ciarletta

LORIA - Campus Scientifique - BP 239 - 54506 Vandœuvre-lès-Nancy Cedex
{julien.siebert, vincent.chevrier, laurent.ciarletta}@loria.fr

Résumé Dans cet article, nous présentons une démarche conceptuelle liant des modélisations issues des mondes pair-à-pair et multi-agents afin de prendre en compte le comportement de l'utilisateur dans la modélisation et la simulation des réseaux pair-à-pair. Nous présentons aussi un outil de simulation que nous avons adapté ainsi que nos premiers résultats qui indiquent que cette démarche est cohérente.

Mots-clés: Simulation Pair-à-Pair, Simulation Multi-Agents, Multi-modélisation, Coopération.

1 Introduction

Dans les réseaux pair-à-pair (P2P¹), les utilisateurs sont à la fois clients et fournisseurs d'un service (mémoire, bande passante, puissance de calcul). De ce fait, la qualité de ce service est directement liée au comportement des usagers du réseau. On le voit notamment avec le problème du manque de coopération² (voir [1,2]), ou encore avec le problème de la diffusion de ressources erronées (voir [3]). Nous nous sommes intéressés à la prise en compte du comportement des utilisateurs dans la modélisation et la simulation des réseaux P2P (voir [4]). Les questions que l'on souhaite aborder sont les suivantes : dans quelle mesure un comportement donné agit sur les performances du réseau et, inversement, quelles sont les caractéristiques du réseau qui influencent le comportement des usagers. Le point de vue que nous souhaitons défendre est qu'il faut pouvoir, dans une même simulation, prendre en compte plusieurs niveaux de représentation (utilisateur, réseau overlay, réseau physique).

Nous proposons d'utiliser le paradigme multi-agents afin de pouvoir coupler les modèles de ces différentes échelles d'abstraction. Dans cet article, nous présentons d'abord les travaux existants en modélisation du comportement des utilisateurs. Ensuite nous exposons le paradigme et la simulation multi-agents. Puis nous présentons notre approche conceptuelle. Enfin, nous exposons nos premières expérimentations et nous concluons.

¹ De l'anglais peer-to-peer.

² *free-riding*.

2 État de l'art

2.1 Modélisation et simulation du comportement des utilisateurs dans les réseaux P2P

Les études faisant intervenir la modélisation du comportement des utilisateurs au sein des réseaux P2P traitent principalement des problématiques de la coopération (environ 70% des utilisateurs ne partagent pas leurs ressources [1,2]) et de l'empoisonnement des données (près de 50% des ressources les plus populaires sont polluées³ sur le réseau FastTrack [3]). La modélisation de l'utilisateur s'appuie sur des travaux réalisés en économie ou directement d'observations.

Feldman *et al.* [5] proposent une mise en équation du phénomène de coopération, chaque utilisateur décidant s'il partage en fonction de la relation entre son *type* (sa générosité) et le coût de sa contribution. Dumitriu et Lee *et al.* [6,7] s'intéressent à la coopération et à l'empoisonnement des données et se basent sur une approche itérative en définissant dès le départ l'ensemble des règles déterminant leur système.

On peut voir poindre ici plusieurs critiques concernant ces approches purement analytiques. Tout d'abord, les aspects dynamiques des comportements et leur hétérogénéité ne sont pas toujours pris en compte dans les modèles. Ensuite, l'aspect réseau est souvent absent des modélisations. Or, nous estimons que certaines caractéristiques influencent le comportement des utilisateurs (comme, par exemple, la disponibilité de la bande passante). Pour pallier ces contraintes, les travaux dans ce domaine commencent à s'intéresser à la simulation multi-agents (voir [8,9]). Nous montrerons par la suite qu'une telle approche apporte effectivement son lot de solutions mais qu'il existe aussi un réel manque en matière de modélisation réseau au sein de la communauté des systèmes multi-agents (SMA).

2.2 Modélisation et simulation multi-agents

Un système multi-agents peut être défini comme un système composé d'un environnement où peut évoluer un ensemble organisé d'agents en interaction (voir la définition donnée par Ferber [10]). L'utilisation qui en est faite dans les réseaux est détaillée dans [11]. La simulation multi-agents se propose de modéliser explicitement les comportements des entités et considère que la dynamique globale d'un système, au niveau macroscopique, résulte directement de l'interaction des comportements au niveau microscopique (Parunak *et al.* [12]). La simulation multi-agents est utilisée dans l'étude des phénomènes complexes dans divers domaines comme l'éthologie [13], les sciences de l'Homme et de la Société [14] (écologie, sociologie, les simulations urbaines) mais encore relativement peu dans la modélisation des réseaux (voir [15]).

Les intérêts de cette approche sont de pouvoir représenter et simuler des systèmes faisant intervenir un grand nombre d'individus ; de prendre en compte,

³ Soit le contenu est erroné (bruit, données manquantes...) soit les métadonnées (titre, auteur...) différent du contenu.

dès la modélisation, les caractères dynamiques et hétérogènes des comportements individuels ; et d'analyser l'impact d'un comportement local sur le fonctionnement global du système. Ainsi, le paradigme multi-agents semble tout indiqué pour étudier l'impact du comportement des utilisateurs dans les réseaux P2P (voir [8,9]). Cependant, au vu de ces travaux, il est apparu que les aspects réseaux (protocole P2P, bande passante, délai) n'étaient, pour ainsi dire, jamais pris en compte [4].

Ce constat, la thèse que nous défendons (prendre en compte les trois niveaux d'abstraction) et les contraintes de temps nous ont conduit à nous intéresser aux outils de simulation P2P afin de pouvoir y implanter notre approche et de réutiliser un modèle cohérent de réseau.

2.3 Modélisation et simulation P2P

Les problématiques dans ce domaine sont souvent axées « performances au niveau overlay » (montées en charge, passage à l'échelle). Ainsi, les outils développés intègrent rarement une couche modélisation de l'utilisateur. Après une étude bibliographique (voir le détail dans [4]) nous avons décidé d'utiliser PeerfactSim.KOM (PFS)⁴. Le principal avantage de cet outil, en plus d'un nombre annoncé de nœuds simulés conséquent (10^5 sur un réseau Omicron [16]), provient de son architecture qui fait nettement la séparation entre les différents niveaux d'abstraction décrits précédemment.

Au niveau réseau physique, le simulateur modélise, entre autres, les temps de latence pour chaque message envoyé.

Au niveau réseau overlay, des protocoles P2P existants (CAN, Omicron, Chord, Kademlia) sont implantés originellement dans le simulateur.

Au niveau utilisateur, la partie comportement de l'utilisateur est totalement détachée du nœud physique correspondant. Ainsi, on peut implanter un modèle de comportement sans avoir à modifier les parties réseaux.

3 Vers une approche multimodèle liant P2P et SMA au travers d'un cas d'étude

3.1 Approche générale

Les modélisations du comportement des utilisateurs intègrent rarement les aspects du réseau utilisé (protocole, délai, bande passante) et, à l'inverse, les modélisations des réseaux ne tiennent guère compte des aspects liés aux comportements des usagers (hétérogénéité, dynamique). La problématique sur laquelle nous souhaitons travailler, l'influence réciproque réseau/usagers, nous incite à prendre en compte les différentes échelles de représentation des systèmes P2P (utilisateur, réseau overlay, réseau physique). Ainsi, nous pensons que le paradigme multi-agents peut nous permettre de coupler ces différents niveaux d'abstraction.

⁴ <http://www.peerfactsim.com>

L'idée de base est de mettre en parallèle les composants d'un SMA (Agents, Environnement, Interactions et Organisation [17]) avec les différentes échelles de représentation dans les systèmes P2P (voir la figure 1). C'est cette démarche

P2P	SMA
Utilisateur	Agent
Protocole	Interactions - Organisation
Réseau physique	Environnement

Fig. 1. Mise en relation des paradigmes SMA et P2P [4].

qui va nous guider tout au long de notre travail. Afin de la valider, nous nous sommes intéressés aux problématiques de diffusion de pollution et du manque de coopération au sein des systèmes P2P. Nous avons commencé par utiliser une approche relativement simple où l'utilisateur est décrit par un modèle d'agent réactif (voir ci-après). Le protocole P2P (Kademlia) et le modèle de réseau physique sont tous deux issus du simulateur PFS. Nous y avons apporté quelques modifications pour tenir compte des échanges de fichiers et de la pollution (voir [18]). Les sections suivantes décrivent ce cas d'étude.

3.2 Un premier modèle comportemental : l'agent réactif

Notre hypothèse de travail est la suivante : les comportements que nous avons souhaités étudier (coopération, diffusion de la pollution) ne font pas obligatoirement appel à des schémas comportementaux très développés. On peut considérer, par exemple, 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). De même, la représentation qu'a un usager de ses « voisins » et du fonctionnement du réseau peut être simplifiée. En effet, on peut considérer qu'il ne peut percevoir et interagir sur le réseau qu'au travers de son client. De plus, la taille des réseaux que nous souhaitons modéliser est relativement élevée (entre 10^3 et 10^6 utilisateurs), ce qui nous a contraint sur la complexité du modèle comportemental. Nous avons donc décidé de nous baser sur une architecture d'agent réactif pour modéliser le comportement des utilisateurs [4].

Les agents réactifs possèdent peu d'états internes et ont des capacités de raisonnement rudimentaires. Ils n'ont pas de représentation très poussée d'eux même et de leur environnement et ne peuvent pas, par exemple, établir des conjectures sur le fonctionnement de leur entourage. Leur comportement est proche du réflexe. L'agent perçoit certains états de l'environnement via ses capteurs et décide des actions à mener en fonction de ses percepts et de ses états internes. Dans notre cas, nous considérons les paramètres suivants :

Perceptions

- Les débits montant (upload) et descendant (download), la bande passante consommée par l’offre et la consommation de ressources. On peut considérer le rapport entre ces deux débits comme le coût de la contribution de l’agent comme évoqué dans [5].
- Nombre de sources disponibles après un certain temps d’attente. Cette donnée peut être vue comme la rareté d’un fichier sur le réseau.
- Temps de téléchargement estimé. Ceci peut caractériser la patience de l’usager face au téléchargement.

Actions

- Connexion/déconnexion au réseau P2P.
- Envoi d’une requête, 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.

États internes

- La volonté de partager (générosité α). Cette notion est issue de l’économie et des travaux de Andreoni sur le modèle *warm-glow model* (voir l’utilisation faite dans [5]). On peut considérer cet état comme la quantité (en pourcentage) de ressource que l’utilisateur est prêt à partager.
- La conscience de la pollution p_a , évoquée dans [7]. On peut voir cet état comme la probabilité que l’utilisateur détecte la pollution sachant que le fichier est corrompu. De cette manière, on peut dire qu’un agent dont la conscience $p_a = 1$ va automatiquement détecter un fichier pollué et va agir en conséquence. De plus, on fait l’hypothèse que l’utilisateur reconnaît toujours un fichier authentique.

4 Premières Expérimentations et Résultats**4.1 Scénario de base**

Initialement On fait l’hypothèse qu’il y a N utilisateurs au total dans le réseau. Chaque utilisateur ne garde qu’une seule version d’un fichier. Au début de la simulation, il y a G_0 utilisateurs qui mettent à disposition une version correcte (un fichier sans pollution) et B_0 utilisateurs qui mettent à disposition une version polluée. $0 \leq G_0 + B_0 = P \leq N$.

Étape 1 (durée maxi = e_1). Recherche des sources disponibles La première étape consiste en l’envoi d’une requête. Au temps $t_1 = t_0 + x$, avec $x \sim \mathcal{U}[0, e_1]$. Chaque pair obtient une liste de sources disponibles.

Étape 2 (durée maxi = e_2). Demande de téléchargement Au temps $t_2 = t_1 + x$, avec $x \sim \mathcal{U}[0, e_2]$, 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 $x \sim \mathcal{U}[0, e_3]$, 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). Vérification de la pollution Au temps $t_4 = t_3 + x$, avec $x \sim \mathcal{U}[0, e_4]$, l'utilisateur vérifie la pollution du fichier téléchargé. Sachant que le fichier est pollué, l'utilisateur a une probabilité p_a de le détecter.

- Si celui-ci est détecté, l'utilisateur le supprime et relance la procédure.
- Sinon l'utilisateur décide de le partager ou non avec une probabilité p_c .

Étape 5 : recommencer 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 $\Delta_t \sim \mathcal{U}[0, e_5]$. À noter que pour des raisons technique on peut définir un nombre maximal (r_{max}) de fois où l'utilisateur peut recommencer la procédure.

4.2 Expérimentations

Ce premier modèle, relativement simple, nous a permis d'étudier la faisabilité et la validité de notre approche et de prendre en main l'outil de simulation. Pour ce faire, nous nous sommes placés dans un cas relativement idéal : tous les pairs partagent ($p_c = 1$) et aucun fichier n'est pollué ($B_0 = 0$). On se propose d'aborder la question de l'impact du nombre de pairs qui publient les données. L'intuition de départ est la suivante : dans ce type de réseau, plus le nombre de pairs qui publient des ressources est important, moins la charge par utilisateur est élevée. C'est ce type de comportement que l'on souhaite mettre en évidence avec notre simulateur.

Les premiers résultats montrent que l'on peut faire le lien entre un certain comportement des utilisateurs et les performances du réseaux. Dans l'exemple de la figure 2, on observe, de manière qualitative, que, pour un faible nombre P de pairs qui publient des données (fig. 2, en haut), un nombre restreint d'utilisateurs supporte la majorité de la charge du réseau. Ensuite, lorsque P augmente, la charge se répartit de manière homogène sur l'ensemble des pairs (fig. 2, au centre et en bas). Cet exemple, somme toute classique, nous montre, d'une part, que notre démarche conceptuelle et l'outil qui en résulte nous permet effectivement de traiter notre problématique d'influence comportement des usagers/performances réseau (et inversement), et d'autre part, que le comportement de notre modèle dans un cas idéal est relativement réaliste.

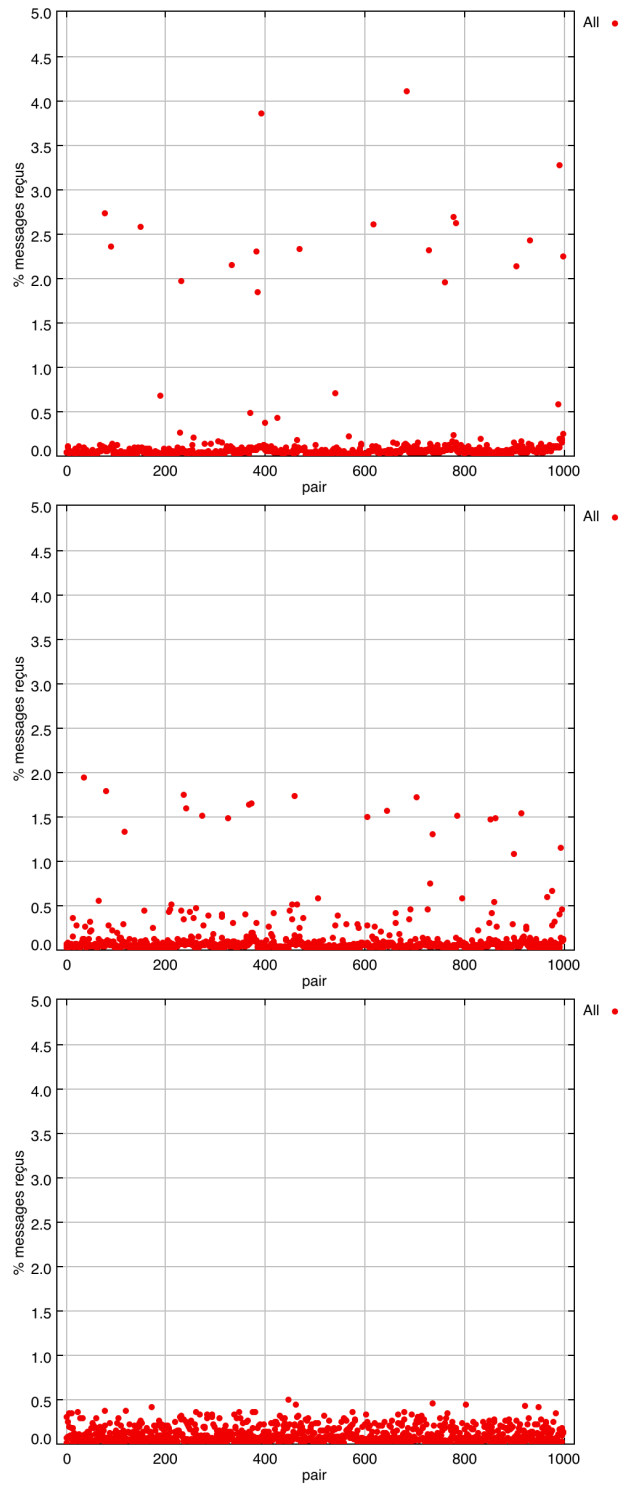


Fig. 2. Un exemple de résultat : la charge par utilisateur. En abscisse le numéro des pairs, en ordonnée le % du nombre total de messages reçus. $N = 1000$. De haut en bas : $P = 10$, $P = 100$ et $P = 500$.

Nous avons simulé plusieurs séries d'expérimentations avec, pour chacune, des jeux de paramètres différents sur des tailles de réseau (N) allant de 25 à 50000 nœuds. Nous avons profité de ces expériences pour étudier le passage à l'échelle de l'outil. Les simulations en cours (un réseau Kademia de 50000 nœuds) prennent en moyenne 5 minutes sur une machine de type Intel(R) Core(TM)2 E6600 cadencée à 2.40GHz, 4Go de mémoire RAM et avec la machine virtuelle java 1.6.0.03-b05. Cependant, par manque de place et de temps, nous ne pouvons les présenter dans cet article.

Ces premières expériences menées [18] sont encourageantes, dans le sens où notre outil et la modélisation implantée semblent avoir un comportement réaliste. De plus, elles corroborent l'idée que notre démarche conceptuelle de multimodélisation (P2P & SMA) peut porter ses fruits.

5 Conclusion et perspectives

Dans cet article, nous avons proposé une approche multimodèle pour la simulation des réseaux P2P. La thèse défendue ici est que, dans le contexte des réseaux P2P, il faut pouvoir prendre en compte plusieurs niveaux d'abstraction (utilisateur, protocole/service, réseau physique) pour avoir une compréhension globale. Des modèles existent déjà pour chaque échelle de représentation. Cependant, le lien est rarement établi entre ces différents niveaux. Nous proposons donc une démarche conceptuelle basée sur le paradigme multi-agents qui permet de coupler ces modèles.

Nous avons pris le parti d'adapter un simulateur P2P existant afin d'y implanter notre démarche conceptuelle. Les expériences menées, certes très simples, sont un premier pas qui nous conforte dans l'idée que notre méthode de modélisation peut porter ses fruits.

Nous avons maintenant la maîtrise d'un outil de simulation (PeerFactSim.KOM) que nous avons modifié pour pouvoir étudier la question de l'impact du comportement des utilisateurs sur le fonctionnement des réseaux P2P. Les modifications apportées sur cet outil sont résumées par la figure 3.

Niveau d'abstraction	Modifications apportées
Utilisateur	Modèle d'agent réactif
Client	Liste des données partagées
Données	Taux de pollution
Réseau Physique	Procédure d'échange de données entre deux utilisateurs

Fig. 3. Modifications apportées sur le simulateur PeerFactSim.KOM [18].

Pour la suite des travaux, nous souhaitons travailler à la fois sur la conception de l'outil de simulation et les aspects conceptuels inhérents (le multi-agents pour

la multimodélisation) et sur les problématiques de l'impact du comportement des utilisateurs sur le fonctionnement des réseaux P2P (diffusion de pollution, manque de partage). Sur ce point, nous souhaitons confronter notre approche avec les approches existantes afin de démontrer les apports du paradigme multi-agents.

À plus long terme, nous pensons pouvoir appliquer cette approche aux autres types de réseaux dynamiques. On pourrait alors analyser plusieurs types de comportements (issus des sciences cognitives, de l'économie...) à la fois dans le domaine du P2P mais aussi dans d'autres domaines tels que les réseaux adhoc et maillés (mesh).

Références

1. Adar, E., Huberman, B. : Free riding on gnutella. Technical report, Xerox PARC (2000)
2. Hughes, D., Coulson, G., Walkerdine, J. : Free riding on gnutella revisited : The bell tolls ? IEEE Distributed Systems Online **6**(6) (2005) 1
3. Liang, J. : Pollution in p2p file sharing systems. In : In IEEE Infocom 2005. (2005)
4. Siebert, J. : Impact du comportement des utilisateurs sur le fonctionnement de réseaux pair-à-pair : modélisation et simulation multi-agents. Master's thesis, Université Henri Poincaré (UHP) Nancy 1 (juin 2007)
5. Feldman, M., Papadimitriou, C., Chuang, J., Stoica, I. : Free-riding and whitewashing in peer-to-peer systems. In : PINS '04 : Proceedings of the ACM SIGCOMM workshop on Practice and theory of incentives in networked systems, New York, NY, USA, ACM Press (2004) 228–236
6. Dumitriu, D., Knightly, E., Kuzmanovic, A., Stoica, I., Zwaenepoel, W. : Denial-of-service resilience in peer-to-peer file sharing systems. In : SIGMETRICS '05 : Proceedings of the 2005 ACM SIGMETRICS international conference on Measurement and modeling of computer systems, New York, NY, USA, ACM Press (2005) 38–49
7. Lee, U., Choiz, M., Choy, J., Sanadidiy, M.Y., Gerla, M. : Understanding pollution dynamics in p2p file sharing. In : 5th International Workshop on Peer-to-Peer Systems (IPTPS'06), Santa Babara, CA, USA (February 2006)
8. 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 (juin 2006)
9. Hales, D. : From selfish nodes to cooperative networks ” emergent link-based incentives in peer-to-peer networks. In : P2P '04 : Proceedings of the Fourth International Conference on Peer-to-Peer Computing (P2P'04), Washington, DC, USA, IEEE Computer Society (2004) 151–158
10. Ferber, J. : Les systèmes multi-agents : un aperçu général. Techniques et sciences informatiques **16**(8) (1997) 979–1012
11. Pujolle, G. : 39. In : Les Réseaux, Édition 2005. Eyrolles (2005) 1003 – 1020
12. Parunak, H.V.D., Savit, R., Riolo, R.L. : Agent-based modeling vs. equation-based modeling : A case study and users' guide. In : Workshop on Modeling Agent Based Systems (MABS'98). (1998) 10–25

13. Bonabeau, E., Dorigo, M., Theraulaz, G. : *Swarm Intelligence From Natural to Artificial Systems*. Oxford University Press (1999)
14. Amblard, F., Phan, D. : *Modélisation et Simulation Multi-Agent Application pour les Sciences de l'Homme et de la Société*. Hermes Lavoisier (2006)
15. Gaïti, D., Merghem-Boulahia, L. : 5. In : *L'autonomie dans les réseaux*. Hermes Lavoisier (2006) 133–165
16. Darlagiannis, V. : *Overlay Network Mechanisms for Peer-to-Peer Systems*. PhD thesis, Vom Fachbereich Informatik der Technischen Universität Darmstadt (2005)
17. Demazeau, Y. : *From interactions to collective behaviour in agent-based systems*. In : *ECCS'95 : First European conference on cognitive science*. (1995) 117–132
18. Siebert, J. : *Impact du comportement des utilisateurs sur le fonctionnement de réseaux pair-à-pair*. rapport technique. Technical report, École Supérieure des Sciences et Technologies de l'Ingénieur de Nancy (ESSTIN), Université Henri Poincaré (UHP) (Septembre 2007)