

Impact du comportement des utilisateurs dans les réseaux pair-à-pair (P2P) : modélisation et simulation multi-agents

Julien Siebert, Laurent Ciarletta, Vincent Chevrier

► **To cite this version:**

Julien Siebert, Laurent Ciarletta, Vincent Chevrier. Impact du comportement des utilisateurs dans les réseaux pair-à-pair (P2P) : modélisation et simulation multi-agents. 16es Journées Francophones des Systèmes Multi-Agents - JFSMA'08, Oct 2008, Brest, France. pp.129-138, 2008. <hal-00331919>

HAL Id: hal-00331919

<https://hal.archives-ouvertes.fr/hal-00331919>

Submitted on 20 Oct 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.

Impact du comportement des utilisateurs dans les réseaux pair-à-pair (P2P) : modélisation et simulation multi-agents

Julien Siebert, Laurent Ciarletta, Vincent Chevrier
{julien.siebert, laurent.ciarletta, vincent.chevrier}@loria.fr

LORIA - Campus Scientifique
BP 239
54506 Vandoeuvre-lès-Nancy Cedex

Résumé

Dans cet article, nous présentons les interconnexions qui peuvent se faire entre, d'une part, la modélisation et simulation multi-agents, et le domaine des réseaux pair-à-pair (P2P) d'autre part. Plus particulièrement, dans ce domaine, nous traitons de la prise en compte, dans les modèles de simulation, des comportements de l'utilisateur sur l'une des caractéristiques du fonctionnement de ces réseaux : la qualité de service.

Dans un premier temps, nous montrons les caractéristiques qui nous semblent souhaitables pour modéliser de manière adéquate de tels réseaux, montrons les limites actuelles que nous constatons dans le domaine multi-agents et proposons une approche multi-modèles. Dans un second temps, nous détaillons la mise en œuvre de cette proposition puis validons son bien-fondé en détaillant quelques expérimentations réalisées qui montrent que notre outil correspond bien à nos attentes de modélisation.

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

Abstract

In this article, we present the interrelations between, on the one hand, agent-based modelling and simulation, and, on the other hand, peer-to-peer networks realm. More precisely, in this field, we talk about taking into account users' models when simulating quality of services.

At first, we highlight the attributes we think useful to model such networks. Then, we present the limits we have found in the multi-agents area and we propose a multi-model approach. Next, we detail our implementation and validate it by presenting some experiments which show that our tool corresponds to our modelling needs.

Keywords: *Peer-to-Peer Simulation, Agent-based Simulation, Multi-model, Free-riding, Poisoning*

1 Introduction

Les réseaux pair-à-pair (P2P¹) sont un moyen avantageux d'offrir et de partager des services. Ils ont remporté ces dernières années un succès rapide auprès du grand public et des professionnels. En effet, ces systèmes sont peu coûteux, tolérants aux pannes et aux déconnexions, faciles d'utilisation et passent aisément à l'échelle (jusqu'à plusieurs millions d'utilisateurs²). Ces avantages sont dus à un concept fort : *la décentralisation des services*. Chaque utilisateur est à la fois client et fournisseur d'un service.

Dans cet article, on s'intéresse aux applications d'échange de contenu (fichiers) qui sont les plus connues et utilisées (comme eMule ou bitTorrent, par exemple). Prenons un exemple (idéal) : *Imaginons qu'Alice se soit connectée, via un client³ à un réseau P2P. Elle souhaite y partager et diffuser un service (un logiciel qu'elle aurait développé par exemple). Alice propose un service en déposant simplement son logiciel dans le dossier partagé de son client. Le reste lui est totalement transparent.*

Dans la deuxième étape, le système **indexe** les nouvelles ressources (le logiciel d'Alice). Ici, plusieurs méthodes, plus ou moins centralisées, existent selon le protocole utilisé (voir le détail dans [ATS04]).

Imaginons maintenant que Bob souhaite, via ce réseau, récupérer un service du même type que celui qu'offre Alice. Cependant Bob n'a aucune idée de qui partage ce dernier. Il doit donc le rechercher. Ici aussi, plusieurs méthodes plus ou moins efficaces ont été développées, allant des méthodes d'inondations (on transfère une requête à nos voisins, chaque voisin la transfère à ses voisins, et ainsi de suite...) jusqu'à l'utilisation de tables de hachage distribuées (voir le

¹De l'anglais *peer-to-peer*.

²http://www.ipoque.com/news_&_events/internet_studies/

³Nous utiliserons ce terme dans cet article pour désigner le logiciel qui permet le téléchargement.

détail dans [ATS04]).

Bob connaît maintenant Alice et son logiciel. Il ne lui reste plus qu'à le télécharger directement depuis la machine d'Alice. Il peut désormais lui aussi partager ce logiciel. La boucle est ainsi bouclée.

On s'aperçoit, au travers de cet exemple, qu'idéalement, plus le système contient d'utilisateurs et plus le nombre de ressources (uniques ou copiées) est important. Cependant que se passe-t-il si Bob décide d'utiliser le système sans partager quoi que ce soit ? C'est ce qu'on appelle le *free-riding*. On estime qu'environ 70% des utilisateurs ne partagent pas (voir [AH00, HCW05]). Dans ce cas, la répllication des données qui assure la robustesse du système est mise en péril.

En outre, que se passe-t-il si Bob décide de partager ses ressources mais en mentant sur leur contenu (nom différent, contenu dégradé, présence de virus...)? C'est ce que l'on appelle l'empoisonnement⁴ (voir [Lia05]).

On voit ici que la qualité du service rendu est sensible au comportement des usagers. Les phénomènes décrits ci-dessus font perdre aux systèmes P2P leurs principaux avantages. Bien que des méthodes existent pour tenter d'y remédier, ils restent aujourd'hui une question ouverte. En effet, la difficulté tient du fait que, d'une part, il est relativement complexe de *contrôler* et de *prévoir* ce que chacun fait dans un système aussi décentralisé, et d'autre part, que les phénomènes décrits interviennent à *différents niveaux d'abstraction* du système.

Dans cet article, nous nous concentrons sur un cas d'étude particulier. Nous prenons le parti d'étudier, par la modélisation et simulation, les relations qui existent entre le comportement des utilisateurs et la qualité de service⁵ des réseaux P2P. La modélisation et la simulation multi-agents nous apparaissent comme les voies les plus adaptées. Au vu des interrelations qui existent entre la QoS et les comportements des usagers, nous pensons qu'il est nécessaire de modéliser ces différents niveaux d'abstractions : *utilisateur, application P2P et réseau*. Cette démarche nous permettra de détecter les comportements dangereux/bénéfiques pour le réseau, mais aussi de comparer différents services dans différentes conditions d'utilisation. Bien que cet article traite principalement des réseaux P2P

⁴Aussi appelé pollution.

⁵En anglais : Quality of service (QoS)

d'échange de contenus, nous souhaitons notre démarche plus générale et applicable aux autres types d'applications qui suivent les concepts inhérents au P2P.

Cet article est structuré comme suit : la section suivante présente un état de l'art de la question. Ensuite, après avoir montré les caractéristiques qui nous semblent souhaitables de modéliser, nous exposons notre proposition. L'avant-dernière section rend compte de nos expérimentations et résultats. Enfin, en guise de conclusion, nous discutons du travail effectué et des travaux futurs.

2 Etat de l'art

Nous nous focalisons ici sur la prise en compte du comportement de l'utilisateur dans la modélisation et la simulation de phénomènes dans les réseaux P2P. Dans ce domaine, nous constatons d'abord un manque de modélisation du comportement des utilisateurs. Nous montrons ensuite que le domaine des systèmes multi-agents (SMA) quant à lui est apte à modéliser ces aspects mais manque cruellement de modèles pour représenter de manière appropriée le monde réseau.

Enfin, une troisième partie détaille l'outil PeerfactSim.KOM qui nous a servi de base pour l'implantation de notre proposition.

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

Nous nous sommes intéressés aux études faisant intervenir la modélisation du comportement des utilisateurs au sein des réseaux P2P ; plus spécifiquement, celles traitant principalement des problématiques de la coopération et de l'empoisonnement des données. En général, la modélisation de l'utilisateur s'appuie sur des travaux réalisés en économie ou directement d'observations.

Feldman et al. [FPCS04] 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, Lee et al. [DKK⁺05, LCC⁺06] 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. Thommes et al. [TC06] définissent un modèle de diffusion de virus au

sein des systèmes P2P, basé sur des équations différentielles.

Tous ces travaux apportent des résultats sur un comportement moyen du système. Dans notre démarche, nous souhaitons, en plus de pouvoir obtenir ce genre de résultats globaux, pouvoir mesurer l'impact de comportements ou de paramètres réseaux locaux sur l'ensemble du système. *Par exemple, quel est l'impact de la concentration de fichiers pollués sur la diffusion de l'empoisonnement ?* Pour ce type de résultats, on s'aperçoit que ces approches moyennes (equation based model, voir [VDPSR98]) ne sont pas les mieux adaptées. De plus, il semble difficile de prendre en compte la dynamique et l'hétérogénéité des comportements des utilisateurs au travers de ces approches. Enfin, nous souhaitons mesurer les relations entre comportement et qualité de service. Cette dernière dépend, entre autres, du type de réseau P2P utilisé. Or, les aspects réseaux (bandes passantes, délais, protocoles réalistes) sont rarement présents et relativement difficiles à exprimer au sein de ces modélisations.

Pour pallier ces contraintes, les travaux dans ce domaine semblent converger vers la simulation multi-agents. 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 du monde réseau.

2.2 Modélisation et simulation multi-agents

La simulation multi-agents [Fer97, BLPM02] 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 (dixit [Rej05]). Les intérêts de cette approche sont de pouvoir représenter et simuler des systèmes faisant intervenir un grand nombre d'individus ; de prendre en compte, 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 [VDPSR98].

Ainsi, le paradigme multi-agents semble tout indiqué pour étudier notre problématique d'impact du comportement des utilisateurs dans les réseaux P2P. Il permettrait par exemple de prendre en compte le fait que certains usagers

sont plus ou moins enclins à partager leurs ressources (fichiers, bande passante...) en fonction de leur perception de l'environnement (le réseau) ; ou encore de mesurer l'impact de paramètres locaux sur le comportement global du système, comme par exemple l'effet du nombre de fichiers publiés par chaque usager sur la vitesse de diffusion des ressources dans le réseau.

Cependant, au vu des travaux existants (voir notamment [Lev06, Hal04]) et aussi de peer-sim⁶, le seul outil de simulation multi-agents traitant de P2P à notre connaissance, il est apparu que les aspects réseaux (protocole P2P, bande passante, délais...) n'étaient, pour ainsi dire, jamais pris en compte. Il nous semble donc qu'il existe un réel manque au sein de la communauté multi-agents en termes d'outils de modélisation et de simulation des réseaux P2P.

Il existe des utilisations du paradigme multi-agent dans les réseaux, détaillées dans [RRH08, Puj05], mais celles-ci ne correspondent pas à notre besoin.

2.3 Modélisation et simulation P2P

Des contraintes de temps et la volonté de modéliser à la fois les aspects réseaux et comportementaux nous ont conduit à nous intéresser aux outils de simulation P2P. Ainsi, nous pourrions y implanter notre approche et réutiliser un modèle cohérent de réseau.

Les problématiques dans ce domaine sont souvent axées sur les performances des protocoles P2P (montée en charge, passage à l'échelle). Ainsi, les outils développés intègrent rarement une couche modélisation de l'utilisateur [NBLR06, Sie07].

Nous avons décidé d'utiliser Peerfact-Sim.KOM⁷ (PFS). Le principal avantage de cet outil, en plus d'un nombre annoncé de noeuds simulés conséquent (de l'ordre de 10^5 [Dar05]), provient de son architecture qui fait nettement la séparation entre les différents niveaux d'abstraction décrits précédemment. Au niveau du réseau physique, le simulateur modélise, entre autres, les temps de latence pour chaque message envoyé. Au niveau du réseau overlay⁸, des protocoles P2P existants

⁶<http://peersim.sourceforge.net>

⁷<http://www.peerfact.org>

⁸On appelle overlay le réseau virtuel composé de tous les pairs liés deux à deux.

(CAN, Omicron, Chord, Kademia) sont implantés originellement dans le simulateur. Au niveau de l'utilisateur, la partie modélisation du comportement est détachée du noeud physique correspondant. Ainsi, on peut implanter un modèle de comportement sans avoir à modifier les parties réseaux.

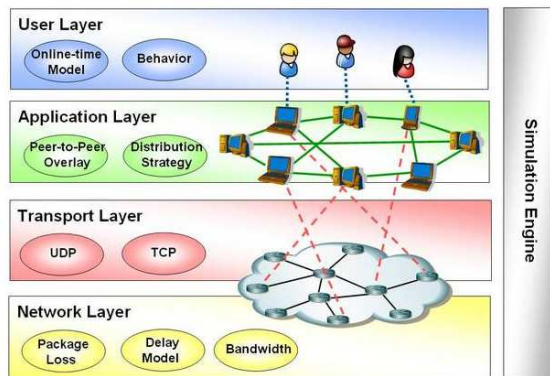


FIG. 1 – Architecture du simulateur PFS.
Source <http://www.peerfact.org>

Comme nous le montrerons par la suite, ces caractéristiques font que PFS offre une base de départ tout à fait intéressante qui peut être étendue à moindre coût pour intégrer des considérations plus en rapport avec la modélisation multi-agents. La section suivante développe notre approche de modélisation.

3 Vers une approche multi-modèles liant P2P et SMA

Nous proposons ici une approche pour l'étude des interactions entre la qualité de service des réseaux P2P et le comportement des utilisateurs. Nous avons vu précédemment qu'utiliser un seul point de vue semblait insuffisant pour répondre à notre problématique. Il nous paraît nécessaire de faire coopérer au sein d'une même modélisation P2P et SMA. Nous avons donc proposé une démarche conceptuelle qui permet de faire interagir ces deux domaines.

L'idée de base est de mettre en parallèle les différents composants des systèmes P2P avec ceux issus du paradigme multi-agents tels qu'évoqués dans différentes approches [Dem95, Fer97] (voir la figure 2).

Paradigme Agent	Monde P2P
Agent	Utilisateur
Interactions	Applications et Protocoles P2P
Environnement	Réseau physique sous-jacent

FIG. 2 – Mise en relation du paradigme multi-agents et du monde P2P [SCC08]

3.1 Un modèle comportemental multi-agents

Choix d'un modèle réactif.

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, 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). 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 étant relativement élevée (entre 10^3 et 10^6 utilisateurs), elle nous contraint 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 [SCC08]. Dans notre cas, les perceptions et les actions sont données directement par le client P2P.

Détail du modèle comportemental.

Nous avons choisi ici de nous concentrer sur un cas d'étude particulier : le partage et la diffusion de la pollution. Le modèle comportemental est construit sur une boucle perception-décision-action (voir fig. 3). La décision s'effectue à partir des perceptions de l'agent et des paramètres décrivant le modèle comportemental.

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

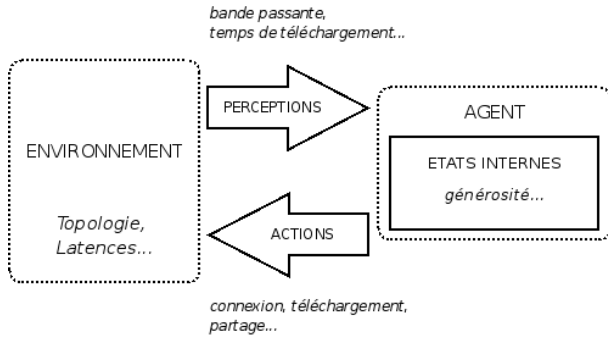


FIG. 3 – Boucle perception-décision-action.

- 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. On peut considérer le rapport entre ces deux débits (noté $c_i = download_i/upload_i$) comme le coût de la contribution de l’agent i .
- Le *nombre de sources* disponibles n après un certain temps d’attente et qui caractérise la rareté 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 [FPCS04]). Elle exprime la probabilité qu’à un agent i de partager ses ressources.
- La *conscience de la pollution*, évoquée dans [LCC⁺06]. On notera β_i la probabilité que l’agent i détecte la pollution sachant que le fichier est corrompu.

Décision.

Les fonctions de décision de l’agent sont inspirées des modèles développés dans [FPCS04, LCC⁺06]. Nous avons adapté ces modèles à notre approche multi-agents et ajouté de nouveaux paramètres.

On peut considérer qu’un agent va partager ses ressources principalement au regard de deux paramètres : sa générosité α_i et le coût de contribution c_i . L’agent décidera de partager si la fonction de type

$$\alpha_i * c_i^{-a} > \text{seuil partage} \quad (a > 0)$$

Ainsi, un agent avec une grande générosité aura tendance à partager plus ; et plus le coût de contribution sera élevé et moins les agents partageront.

Un agent i détecte la pollution au sein d’un fichier (on note $\tau_f \in [0, 1]$ le taux de pollution d’un fichier f) et supprime ce dernier si et seulement si :

$$\beta_i > 1 - \tau_f$$

De cette manière, on peut dire qu’un agent dont la conscience $\beta = 1$ va automatiquement détecter un fichier pollué et va agir en conséquence. On fait l’hypothèse que l’utilisateur reconnaît toujours un fichier authentique ($\tau_f = 0$).

On pourra aussi considérer que les agents tolèrent un temps d’attente ω plus long (entre le lancement et l’arrêt du téléchargement) quand le nombre de sources disponibles pour un fichier est petit : $\omega \propto 1/n$.

3.2 Modélisation des aspects réseaux et implantation

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 QoS du réseau P2P, il ne semble pas nécessaire de modéliser jusqu’à l’envoi de chaque datagramme sur le réseau, voire de modéliser l’ensemble des équipements réseaux.

L’envoi de messages au niveau du protocole P2P (overlay) 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. Cependant, il nous paraît nécessaire de prendre en compte quelques caractéristiques du

réseau sous-jacent (bande passante, temps de transfert, modélisation des données transférées) en vue d'évaluer les performances réseaux et parce qu'elles impactent aussi le comportement de l'utilisateur.

Notre problématique d'impact du comportement des utilisateurs implique que ces derniers puissent manipuler les données (les stocker, les partager, les empoisonner...). On a vu précédemment que 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, en revanche toute la partie transfert « concret » de données (modélisation, stockage des données, procédure d'échange d'un pair à un autre...) n'était pas modélisée. Nous avons donc apporté quelques modifications au simulateur pour implanter ces aspects ([Sie07]) :

- 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.

3.3 Synthèse

Dans cette section, nous avons exposé une approche conceptuelle de modélisation pour la prise en compte du comportement des usagers dans les réseaux P2P. Cette dernière fait appel au paradigme multi-agents tout en intégrant des modèles issus du monde réseau. Nous avons aussi décrit un modèle comportemental d'utilisateur et les modifications de l'outil Peerfact-Sim.KOM qui ont été nécessaires afin de procéder à l'implantation du tout.

4 Expérimentations

Nous présentons ici les expériences que nous avons effectuées. Les premières expérimentations ont servi à valider notre implantation. La deuxième série quant à elle s'intéresse à démontrer les intérêts de l'approche multi-agents par rapport aux méthodes existantes.

4.1 Scénario

Nous avons établi un scénario de simulation en cinq étapes. Celui-ci est détaillé dans [SCC08], nous en rappelons ici les caractéristiques principales.

Notre réseau est composé de N utilisateurs. Parmi ces N utilisateurs, on considère S *égoïstes*⁹. Ce sont les utilisateurs qui ne partageront aucun fichier ($\alpha_i = 0$).

Initialement, certains pairs publient des données, ce sont les sources initiales. Dans notre cas nous distinguons les *sources saines* (notées G_0 ¹⁰) qui publient chacune $f g_0$ *fichiers sains* ; et les *sources polluées* (notées B_0 ¹¹) qui publient chacune $f b_0$ *fichiers pollués*. Il y a donc $B_0 * f b_0 + G_0 * f g_0$ fichiers publiés initialement.

Ensuite, une fois connecté et les ressources publiées, chaque pair suit le scénario suivant :

ÉTAPE 1 : Rechercher un fichier

- Envoyer une requête sur le réseau.
- Attendre un laps de temps.

ÉTAPE 2 : Télécharger

- Parmi les sources disponibles, en choisir une.
- Initier la procédure de téléchargement.

ÉTAPE 3 : Vérifier le téléchargement

- Si le téléchargement est trop long, passer à l'étape 5.
- Sinon, passer à l'étape 4.

ÉTAPE 4 : Vérifier la pollution et partager

- Si le fichier est pollué, le supprimer et passer à l'étape 5.
 - Sinon, choisir de partager ou non le fichier.
- Fin.*

ÉTAPE 5 : relancer la procédure

- Repartir à l'étape 1 après un certain laps de temps.

Utilisateurs	N
Egoïstes	S
Source saines	G_0
Fichiers sains	$f g_0$
Sources polluées	B_0
Fichiers pollués	$f b_0$

FIG. 4 – Notations des paramètres expérimentaux

⁹S pour l'anglais *selfish*.

¹⁰G pour l'anglais *genuine*.

¹¹B pour l'anglais *bogus*.

4.2 Première série d'expérimentations

Avant d'étudier la problématique de l'impact du comportement des utilisateurs dans les réseaux P2P, nous souhaitons valider notre approche conceptuelle et le comportement de l'outil. Pour cela, nous nous sommes intéressés au passage à l'échelle et au réalisme de notre outil modifié. Ces résultats sont détaillés dans [SCC08]. Nous en reprenons les caractéristiques principales ici.

Nous nous sommes placés dans un cas idéal sans pollution ni égoïsme ($B_0 = 0$, $fb_0 = 0$ et $S = 0$).

Passage à l'échelle : les expérimentations se sont déroulées sur des réseaux Kademia [MM02] de tailles variant de $N = 25$ à $N = 50000$. Les simulations les plus importantes en nombre de nœuds prennent en moyenne 15 minutes (Java 6, 2Gb RAM, core duo 2.4GHz).

Réalisme : nous avons étudié l'impact du nombre de sources initiales ($B_0 + G_0$) sur la charge supportée par chaque utilisateur (ici le nombre de messages reçus par pair). L'idée de base est la suivante : idéalement, dans un réseau P2P, plus les utilisateurs publient et plus la charge se répartit équitablement. Nous avons effectivement montré ce comportement dans notre outil modifié (voir le détail des résultats dans [SCC08]).

Ces premières expérimentations nous ont apporté la preuve que le coût d'implantation d'une telle démarche n'est pas prohibitif. Ensuite, elles nous ont permis de prendre en main le simulateur et ont montré que notre outil modifié a un comportement réaliste.

4.3 Deuxième série

Cette deuxième série d'expérimentations a pour but de démontrer les avantages de l'approche multi-agents ; c'est à dire pouvoir mesurer l'impact de facteurs locaux sur le système global.

On cherche ici à mesurer l'effet de la concentration initiale de fichiers pollués sur la diffusion de l'empoisonnement global. Autrement dit, on cherche à voir la différence entre une seule source qui publie beaucoup de fichiers pollués et beaucoup de sources qui publient peu de fichiers pollués. Nous nous sommes placés dans le cas où tout le monde partage et où tout

le monde est conscient de la pollution :

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

Description des expériences.

Nos expériences se déroulent sur des réseaux de tailles $N = 120$, $N = 1200$ et $N = 12000$. Le nombre total de fichiers initialement publiés 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 :

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

Sous ces contraintes, nous avons fait varier la concentration initiale de fichiers pollués, soit B_0 et fb_0 (voir l'exemple en fig. 5). Pour chaque série, nous avons joué 100 fois le scénario avec, pour chacun, une graine aléatoire différente.

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

FIG. 5 – Paramètres variants. Exemple avec $N = 120$, $G_0 = 9$ et $fg_0 = 2$

Résultats.

Nous avons mesuré le nombre de fichiers sains et pollués présents dans le réseau en fonction du temps. On peut ainsi voir l'évolution temporelle de la pollution au travers du réseau. 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 ; et ce, pour les différentes tailles N de réseaux testés.

On voit, sur la figure 6, l'évolution temporelle de l'empoisonnement global sur un réseau de $N = 12000$ utilisateurs. Sur la courbe du haut, on a fait apparaître la quantité de fichiers pollués. Pour les paramètres $B_0 = 1200$ et $fb_0 = 1$

(courbe foncée), la diffusion est plus rapide que pour les paramètres $B_0 = 600$, $fb_0 = 2$ (courbe grise médium) et $B_0 = 300$, $fb_0 = 4$ (courbe claire).

De la même manière, la courbe du bas affiche l'évolution du pourcentage de fichiers pollués. On peut tirer ici les mêmes conclusions.

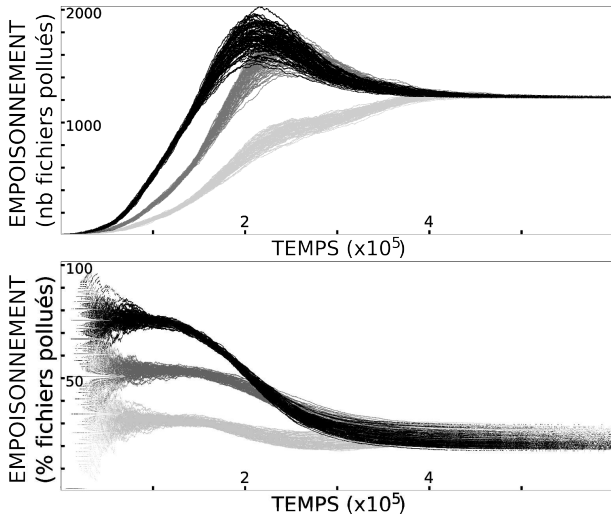


FIG. 6 – Evolution de la pollution sur un réseau de taille $N = 12000$. En haut : le nombre de fichiers pollués en fonction du temps. En bas : le pourcentage de pollution en fonction du temps. Du plus foncé au plus clair : $B_0 = 1200$, $fb_0 = 1$; $B_0 = 600$, $fb_0 = 2$ et $B_0 = 300$, $fb_0 = 4$

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 ([MM02]) ce qui a pour effet de donner autant de visibilité à chaque utilisateur. Ainsi, lorsque le nombre de sources polluantes (B_0) augmente, la probabilité de réplcation 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.

On peut voir apparaître le même type de comportement de diffusion de la pollution dans des modèles plus classiques (comme dans

[LCC⁺06]). Simplement, notre approche multi-agents nous permet de pousser l'analyse plus loin, en intégrant des paramètres locaux (ici le nombre de fichiers publiés initialement fb_0 et fg_0) et en faisant interagir un modèle comportemental avec un modèle de système P2P plus précis.

4.4 Synthèse

Nous avons montré au travers de ces deux séries d'expérimentations qu'effectivement, notre approche multi-modèles liant P2P et SMA, nous permet d'étudier les relations entre la qualité de service des réseaux P2P et le comportements des usagers.

5 Conclusion et perspectives

Dans cet article, nous nous sommes intéressés à la problématique de la modélisation du comportement des utilisateurs dans les réseaux P2P. Nous avons montré pourquoi les approches existantes ne nous semblent pas satisfaisantes et avons proposé une approche multi-modèles faisant interagir les mondes multi-agents et P2P. Nous avons implanté cette approche en utilisant un simulateur réseau existant (Peerfact-Sim.KOM) que nous avons modifié. Les expériences réalisées montrent que cette approche permet de prendre en compte à la fois les aspects réseaux et la dynamique et l'hétérogénéité des comportements au sein d'une même modélisation.

À court terme, nous souhaitons confronter notre approche avec les modèles existants. Nous planifions également d'exécuter d'autres expériences pour continuer notre étude. Nous avons d'ailleurs récemment entrepris de mesurer l'impact d'autres caractéristiques sur la vitesse de diffusion de l'empoisonnement tels que la popularité des fichiers ou encore des paramètres directement issus du protocole Kademlia. Pour le moment, notre implantation ne nous permet que difficilement (en terme de temps de développement) de tirer parti des avantages du paradigme multi-agents. En effet, l'outil de base est certes modulable d'un point de vue réseau P2P, on peut facilement implanter un nouveau protocole ou changer les paramètres du réseau sous-jacent, mais cela devient difficile au niveau comportement des usagers. À moyen terme, nous souhaitons travailler sur cet aspect et proposer une implantation modulable qui permette facilement de définir et d'implanter les notions de types de

comportements, d'actions et de groupes d'utilisateurs.

Bien que cet article traite d'un cas d'étude particulier dans le P2P, les travaux réalisés ont mis en lumière des questions scientifiques plus générales. En effet, en suivant cette démarche, il devient nécessaire de faire interagir différents modèles à des niveaux d'abstraction et des échelles spatiales et temporelles différents. Nous souhaitons, à plus long terme, étudier cet aspect "multi-modèles" et proposer un outil permettant de faire interagir différents modèles de comportements et/ou différents protocoles et/ou différents réseaux sous-jacents afin d'analyser les relations entre le comportement des utilisateurs et les réseaux dynamiques (P2P, adhoc, mesh) d'une manière générale.

Références

- [AH00] E. Adar and B. Huberman. Free riding on gnutella. Technical report, Technical report, Xerox PARC, 10 August 2000.
- [ATS04] Stephanos Androutsellis-Theotokis and Diomidis Spinellis. A survey of peer-to-peer content distribution technologies. *ACM Comput. Surv.*, 36(4) :335–371, 2004.
- [BLPM02] F. Bousquet, C. Le Page, and J-P. Muller. Modélisation et simulation multi-agent. In *Assises du GDR 13*, Nancy, December 2002.
- [Dar05] Vasilios Darlagiannis. *Overlay Network Mechanisms for Peer-to-Peer Systems*. PhD thesis, Vom Fachbereich Informatik der Technischen Universität Darmstadt, 2005.
- [Dem95] Y. Demazeau. From interactions to collective behaviour in agent-based systems. In *European Conference on Cognitive Sciences*, 1995.
- [DKK⁺05] D. Dumitriu, E. Knightly, A. Kuzmanovic, I. Stoica, and W. Zwaenepoel. 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*, pages 38–49, New York, NY, USA, 2005. ACM Press.
- [Fer97] Jacques Ferber. Les systèmes multi-agents : un aperçu général. *Techniques et sciences informatiques*, 16(8) :979–1012, 1997.
- [FPCS04] Michal Feldman, Christos Papadimitriou, John Chuang, and Ion Stoica. Free-riding and white-washing in peer-to-peer systems. In *PINS '04 : Proceedings of the ACM SIGCOMM workshop on Practice and theory of incentives in networked systems*, pages 228–236, New York, NY, USA, 2004. ACM Press.
- [Hal04] David Hales. 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)*, pages 151–158, Washington, DC, USA, 2004. IEEE Computer Society.
- [HCW05] Daniel Hughes, Geoff Coulson, and James Walkerdine. Free riding on gnutella revisited : The bell tolls? *IEEE Distributed Systems Online*, 6(6) :1, 2005.
- [LCC⁺06] Uichin Lee, Min Choiz, Junghoo Choy, M. Y. Sanadidiy, and Mario Gerla. Understanding pollution dynamics in p2p file sharing. In *5th International Workshop on Peer-to-Peer Systems (IPTPS'06)*, Santa Babara, CA, USA, February 2006.
- [Lev06] Christophe. Levointurier. 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.
- [Lia05] Jian Liang. Pollution in p2p file sharing systems. In *IEEE Infocom*, Miami, FL, USA,, March 2005.
- [MM02] P. Maymounkov and D. Mazieres. Kademia : A peer-to-peer information system based on the xor metric, 2002.
- [NBLR06] S. Naicken, A. Basu, B. Livingston, and S. Rodhetbhai. A survey of peer-to-peer network simulators. *Proceedings of The Seventh*

Annual Postgraduate Symposium, Liverpool, UK, 2006.

- [Puj05] Guy Pujol. *Les Réseaux, Edition 2005*, chapter 39, pages 1003–1020. Eyrolles, 2005.
- [Rej05] Lilia Rejeb. *Simulation multi-agents de modèles économiques*. PhD thesis, Université de Reims Champagne-Ardennes, 15 December 2005.
- [RRH08] Maya Rupert, Amjad Rattout, and Salima Hassas. The web from a complex adaptive systems perspective. *J. Comput. Syst. Sci.*, 74(2) :133–145, 2008.
- [SCC08] Julien Siebert, Vincent Chevrier, and Laurent Ciarletta. Modélisation multimodèle des réseaux dynamiques : cas des réseaux pair-à-pair (in french). In *JDIR'08 - 9èmes Journées Doctorales en Informatique et Réseaux*, Villeneuve d'Ascq, France, 2008.
- [Sie07] Julien Siebert. Impact du comportement des utilisateurs sur le fonctionnement des réseaux pair-à-pair. Technical report, ESSTIN, October 2007.
- [TC06] Richard. Thommes and Mark. Coates. Epidemiological Modelling of Peer-to-Peer Viruses and Pollution. In *IEEE Infocom*, Barcelona, Spain, April 2006.
- [VDPSR98] H. Van Dyke Parunak, R. Savit, and R L. Riolo. Agent-based modeling vs. equation-based modeling : A case study and users' guide. In *MABS*, pages 10–25, 1998.