

Définition et diffusion de signatures sémantiques dans les systèmes pair-à-pair

Raja Chiky*, Bruno Defude**, Georges Hébrail*

* GET-ENST Paris

Laboratoire LTCI - UMR 5141 CNRS
Département Informatique et Réseaux
46 rue Barrault, 75634 Paris Cedex 13
Email: chiky@enst.fr, hebrail@enst.fr

**GET-INT

Département Informatique
9 rue Charles Fourier, 91011 Évry cedex
Email: bruno.defude@int-evry.fr

Résumé. Les systèmes pair-à-pair (peer-to-peer, P2P, égal-à-égal) se sont popularisés ces dernières années avec les systèmes de partage de fichiers sur Internet. De nombreuses recherches concernant l'optimisation de la localisation des données ont émergé et constituent un axe de recherche très actif. La prise en compte de la sémantique du contenu des pairs dans le routage des requêtes permet d'améliorer considérablement la localisation des données. Nous nous concentrons sur l'approche PlanetP, faisant usage de la notion de filtre de Bloom, qui consiste à propager une signature sémantique des pairs (filtres de Bloom) à travers le réseau. Nous présentons cette approche et en proposons une amélioration : la création de filtres de Bloom dynamiques, dans le sens où leur taille dépend de la charge des pairs (nombre de documents partagés).

1 Introduction

Pour la recherche, le partage et l'échange de ressources (données, programmes, services), le modèle pair-à-pair constitue une alternative au modèle client/serveur. Les pairs peuvent à la fois offrir (rôle serveur) et demander (rôle client) des ressources. Il existe de nombreuses architectures des systèmes pair-à-pair, se basant sur des techniques différentes de localisation des données, qui se traduisent par des méthodes différentes de routage des requêtes. Pour améliorer la localisation d'une ressource recherchée par un pair, on ajoute de l'information aux tables de routage des requêtes : il peut s'agir du contenu des pairs, de l'historique de leurs requêtes, ou des concepts qu'ils traitent...

La difficulté rencontrée lors de l'intégration de la sémantique du contenu des pairs, est de déterminer un espace de représentation commun à tous les pairs du réseau. Quelques systèmes tels que SON (*Semantic Overlay Network*)(Crespo et al., 2002) utilisent des concepts définis à priori pour résoudre ce problème. Mais cette solution ne s'applique qu'à un domaine précis.