

# PROXSEM: Interest-based Proximity Measure to Improve Search Efficiency in P2P Systems

Yann Busnel

► **To cite this version:**

Yann Busnel. PROXSEM: Interest-based Proximity Measure to Improve Search Efficiency in P2P Systems. EuroSys2006, Apr 2006, Leuven/Belgium. inria-00001249

**HAL Id: inria-00001249**

**<https://hal.inria.fr/inria-00001249>**

Submitted on 12 Apr 2006

**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.

# Interest-based Proximity to Improve Search Efficiency in Peer-to-Peer Systems

Yann Busnel

Yann.Busnel@irisa.fr

PhD Student

IRISA / ENS Cachan – Rennes, France

Peer-to-peer file sharing systems are now at the origin of most of Internet traffic. Improving the performance of such systems has generated a lot of interest both in industry and academia. More specifically, many approaches focus on the improvement of the query mechanism in such systems. In a peer-to-peer system, peers are connected to a subset of other peers with which they can communicate. Each peer maintains a cache and makes available its contents to the rest of the system. Connecting peers sharing similar interest in the context of a given application has recently been identified as a sound basis to improve the search efficiency.

However, capturing such interest-based (or semantic) proximity patterns is a difficult task. Most of current approaches measure this proximity between peers as the overlap between their cache contents. Given the well-known popularity patterns of peer-to-peer file sharing systems, the overlap between cache contents of two peers may not reflect accurately their semantic proximity. More specifically, this measure depends upon peer generosity and file popularity.

In this poster, we propose a refined proximity measure taking into account these factors. We evaluated the proposed solution by simulation against a real peer-to-peer system (eDonkey) workload. We obtain results which shown the effectiveness of the proposed approach: taking into account file popularity increases the one hop hit ratio (23 % of success against 1% for random search and 17 % for overlap size measure); taking into account peer generosity balances generous nodes load (the maximum in-degree of the network is 990 against 1702 with overlap strategy). All experiment results are presented in [1].

While peers generosity can easily be computed locally, file popularity may require a global knowledge of the system. We also propose in this study an epidemic algorithm to compute in a fully decentralised way an estimation of files popularity [1].

## References

- [1] Integrating file popularity and peer generosity in proximity measure for semantic-based overlays. Research Report RR-5731, INRIA, IRISA, Rennes, France, October 2005.