Peer Rewiring in Semantic Overlay Networks under Churn

[Short Paper]

Paraskevi Raftopoulou[1,2] and Euripides G.M. Petrakis[2]

[1]University of Peloponnese (UOP), Tripoli, 22100, Greece
[2]Technical University of Crete (TUC), Chania, 73100, Greece
praftop@uop.gr, petrakis@intelligence.tuc.gr

Abstract. Semantic overlay networks have been proposed as a way to organise peer-to-peer networks; peers that are semantically, thematically or socially similar are discovered and logically organised into groups. Efficient content retrieval is then performed by routing the query towards peer groups based on their likelihood to match the query. In this paper, we study the behaviour of semantic overlay networks that support full-fledged information retrieval in the presence of peer churn. We adopt a model for peer churn, and study the effect of network dynamics on peer organisation and retrieval performance. The overlay network is evaluated on a realistic peer-to-peer environment using real-world data and queries, and taking into account the dynamics of user-driven peer participation. Using this evaluation, we draw conclusions on the performance of the system in terms of clustering efficiency, communication load and retrieval accuracy in such a realistic setting.

1 Introduction

The main idea behind peer-to-peer (P2P) is that instead of relying on central components, functionality is provided through decentralised overlay architectures, where peers typically connect to a small set of other peers. Specifically in Semantic Overlay Networks (SONs), peers that are semantically, thematically or socially similar are *organised* into groups. Queries are then selectively forwarded to those groups that have the potential to provide content matching the queries. SONs, while being highly flexible, improve query performance and guarantee high degree of peer autonomy [2]. Unlike what their name imply, SONs do not necessarily use semantics in the traditional sense (e.g., ontologies), however this is the term first proposed in the literature [1].

The remainder of the paper is organised as follows. SON-like structures supporting IR functionality are reviewed in Section 2. Section 3 presents the model used to describe peer churn, while Section 4 presents a SON architecture and the related rewiring protocol. Finally, the experimental evaluation of the dynamic network is presented in Section 5.

2 Related Work and Background

Initial IR approaches implementing SON-like structures and supporting content search in a distributed collection of peers include the work of Li et al. [4], where semantic small world (SSW) is proposed. SSW aims to construct a self-organising network based on the semantics of data objects stored locally to peers. Along the same lines, Schmitz [8] assumes that peers share concepts from a common ontology and proposes strategies for organising peers into communities with similar concepts. i*Cluster* [6] extends these protocols by allowing peers with multiple and dynamic interests to form clusters.

Most of the above presented research proposals, while exploiting certain architectural or modelling aspects of peer organisation, assume for their experimental evaluation an ideal scenario where peers never leave or join the network. In this paper, we address the issues involved in the design and the evaluation of the SONs when introducing the dynamics of user-driven peer participation.

3 Churn Model

Building upon previous work [11, 3, 9], we present a model of user behaviour characterising peer arrivals and departures in a P2P system. The model takes into account heterogeneous browsing habits, formalises recurring user participation in P2P systems and explains the relationship between the various lifetime distributions observable in P2P networks.

The following assumptions are made:

1. To capture the independent nature of peers, we assume that peers behave independently of each other and processes $\{Z_i(t)\}$ and $\{Z_j(t)\}$, for any $i \neq j$, are independent.
2. For each peer, its on-line and off-line durations are independent.

4 A Semantic Overlay Network

The present work uses i*Cluster* P2P network [6], which extends the idea of peer organisation in small-world networks by allowing peers to have multiple and dynamic interests. To identify its *interests*, a peer categorises its documents by using an external reference system (i.e., an ontology as in [8] or a taxonomy such as the ACM categorisation system) or by clustering.

5 Evaluation

In this section, we evaluate the effect of peer churn on SONs, using the i*Cluster* rewiring protocol and a realistic dynamic setting. Typically, in the evaluation of P2P IR systems performance is measured in terms of *network traffic* (i.e., the number of rewiring/search messages sent over the network) and retrieval effectiveness (i.e., *recall*). In our setting precision is always 100% since only relevant documents are retrieved. Peer clustering is measured by *clustering efficiency* $\bar{\kappa}$ [5], which gives information about the underlying network structure.
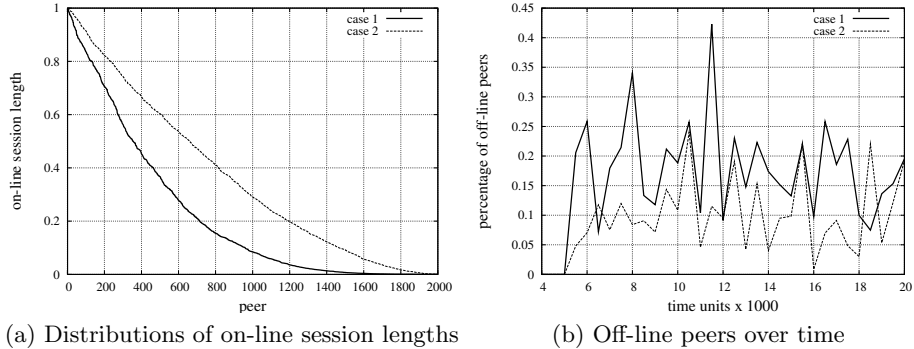
(a) Distributions of on-line session lengths   (b) Off-line peers over time

Fig. 1. Lifetime distributions

5.1 Experimental Testbed

The dataset, also used in the evaluation of [10, 7], contains over 556,000 web documents from the TREC-6 collection belonging in 100 categories. The queries employed in the evaluation of the corpus are strong representatives of document categories (i.e., the topics of the categories).

We experimented with different lifetime distributions. Figure 1(a) illustrates two different on-line session length distributions. By definition, the more skewed the distribution is, the smaller the lifetimes of the most peers are. The first case in Figure 1(a) corresponds to a difficult scenario compared to the second case, since peers are on-line for shorter time periods and leave the network more often. Figure 1(b) presents the percentage of off-line peers as a function of time. In the first scenario the percentage of the peers that are logged-off reaches up to 42% ($t = 11.5K$), while in the second scenario the percentage of the logged-off peers every moment is kept under 25%.

5.2 Experimental Evaluation

Figure 2(a) presents clustering efficiency as a measure of network organisation over time. The plots presented in the figure correspond to the two dynamic scenarios discussed in the previous section.

Notice, by collating Figures 2(a) and 1(b), that clustering efficiency is narrowly related to the percentage of peers that are off-line; when the percentage of off-line peers increases clustering efficiency decreases, and vice versa.

6 Conclusions

In this work, we studied the performance of SONs under churn, using established churn models that describe user behaviour in a realistic setting.
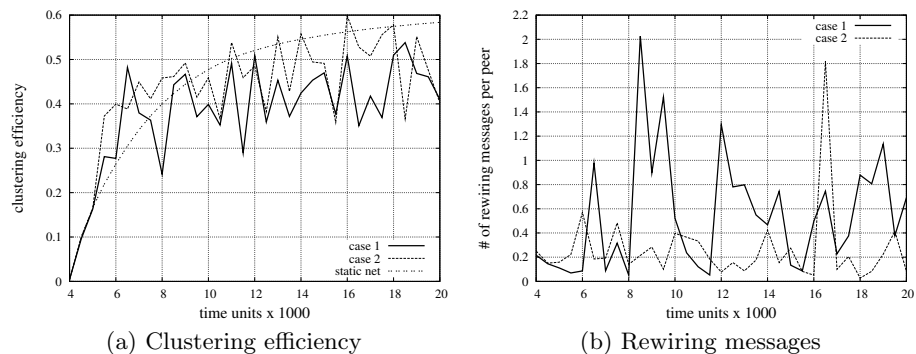
(a) Clustering efficiency   (b) Rewiring messages

Fig. 2. Rewiring effectiveness over time

References

1. A. Crespo and H. Garcia-Molina. Routing Indices for Peer-to-Peer Systems. In *ICDCS*, 2002.
2. H. Garcia-Molina and B. Yang. Efficient Search in Peer-to-Peer Networks. In *ICDCS*, 2002.
3. D. Leonard, Z. Yao, V. Rai, and D. Loguinov. On Lifetime-Based Node Failure and Stohastic Resilience of Decentralised Peer-to-Peer Networks. *IEEE/ACM Transactions on Networking*, 15(3), 2007.
4. M. Li, W.-C. Lee, and A. Sivasubramaniam. Semantic Small World: An Overlay Network for Peer-to-Peer Search. In *ICNP*, 2004.
5. P. Raftopoulou and E. Petrakis. A Measure for Cluster Cohesion in Semantic Overlay Networks. In *LSDS-IR*, 2008.
6. P. Raftopoulou and E.G.M. Petrakis. iCluster: a Self-Organising Overlay Network for P2P Information Retrieval. In *ECIR*, 2008.
7. P. Raftopoulou, E.G.M. Petrakis, and C. Tryfonopoulos. Rewiring Strategies for Semantic Overlay Networks. *Distributed and Parallel DataBases*, 2–3(26), 2009.
8. C. Schmitz. Self-Organization of a Small World by Topic. In *P2PKM*, 2004.
9. D. Stutzbach and R. Rejaie. Understanding Churn in Peer-to-Peer Networks. In *INFOCOM*, 2006.
10. J. Xu and W.B. Croft. Cluster-Based Language Models for Distributed Retrieval. In *ACM SIGIR*, 1999.
11. Z. Yao, D. Leonard, X. Wang, and D. Loguinov. Modeling Heterogeneous User Churn and Local Resilience of Unstructured P2P Networks. In *ICNP*, 2006.