

# When Open Hypermedia Meets Peer-to-Peer Computing

Jing Zhou  
School of Electronics and  
Computer Science  
University of Southampton  
Southampton, Hants, United  
Kingdom  
SO17 1BJ  
jz00r@ecs.soton.ac.uk

Wendy Hall  
School of Electronics and  
Computer Science  
University of Southampton  
Southampton, Hants, United  
Kingdom  
SO17 1BJ  
wh@ecs.soton.ac.uk

David De Roure  
School of Electronics and  
Computer Science  
University of Southampton  
Southampton, Hants, United  
Kingdom  
SO17 1BJ  
dder@ecs.soton.ac.uk

## ABSTRACT

We describe the extension to our previous work on a Web-based peer-to-peer open hypermedia system, the DDLS. We enrich the peer model by introducing query history, and propose the use of the naive estimator which utilises the local knowledge of peers to estimate future information needs they would encounter. Our simulation proves that this statistical technique helps re-organise the DDLS peer network to enhance the performance of resource discovery.

## Categories and Subject Descriptors

H.5.4 [Information Interfaces and Presentation]: Hypertext/Hypermedia—Architectures; H.3.4 [Information Storage and Retrieval]: Systems and Software—Distributed systems

## General Terms

Design, Experimentation, Performance

## Keywords

Open hypermedia, Peer-to-peer, Re-organisation

## 1. INTRODUCTION

The peer-to-peer paradigm [2] which is characterised by equal capability of participating nodes and decentralisation of control can be used to support a distributed open hypermedia system for collaboration. The open hypermedia systems Microcosm [4] and Microcosm TNG [3] employed a hybrid peer-to-peer model [6] in which communication and management of computing tasks relied on central servers, whereas it is no longer feasible for an environment with ad hoc nature. Lukka et al. [5] explored the use of Freenet-like GUIDs for the location and retrieval of blocks of media content for implementing the Xadadu model. They adopted distributed hash table (DHT) techniques to facilitate the search for references to blocks on a global scale. However, the overlay created by DHTs

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

HT'04, August 9–13, 2004, Santa Cruz, California, USA.  
Copyright 2004 ACM 1-58113-848-2/04/0008 ...\$5.00.

fails to take into account the specific requirement of an open hypermedia system that the ability to locate related resources and the ability to relate information of resources to describe the overlay should be included.

We presented in [8] an approach to building a Distributed Dynamic Link Service (DDLS) based on the DLS philosophy [1]. The DDLS is an unstructured peer-to-peer system [6]. It allows resource sharing between users by retrieving and providing hyperlinks, which refer to the documents of interest, from linkbases on demand. The discovery mechanism is supported by a semantic search that satisfies the need of open hypermedia systems. A peer in the DDLS has only access to partial information about its neighbours through interaction. We therefore propose in this paper the use of the naive estimator which utilises the local knowledge of peers to estimate future information needs. The peer network is re-organised based on this estimation to facilitate resource sharing.

## 2. THE DDLS

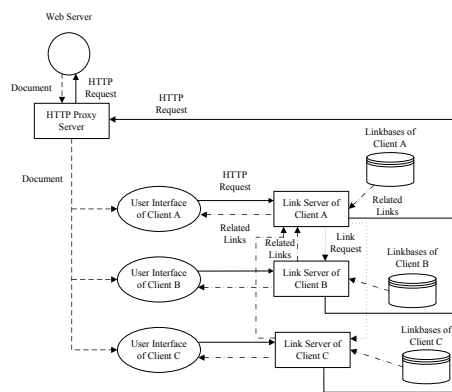


Figure 1: DDLS Architecture

Figure 1 illustrates the architecture of the DDLS. Each peer, or client, has an associated link server which serves link service requests. A link service request is captured by the user interface of a peer and then forwarded to the peer's link server. The link server queries against the peer's personal linkbases. Moreover, it forwards those involving linkbases of others to the target link servers which can be located through the resource discovery mechanism.

In the DDLS, each peer is represented by three fundamental components: the topics which describe the linkbases it maintains,

the overlap between its topics and those of its neighbours, and query history. Each linkbase maintains a list of hyperlinks related to an abstract concept. Therefore, linkbases maintained by a peer is described by a topic vector on the basis of the associated concepts. To efficiently discover the linkbases of interest, we developed a distance-based semantic search algorithm [8] in which queries are either propagated to specific neighbours or broadcast to all neighbours depending on the semantic similarity between the sender and the recipient.

Query history is a data structure on which re-organisation of the peer network (see section 3) primarily relies to predict future information needs. In our implementation, a query history is realised as a FIFO (First In First Out) queue. History information is discarded when the queue overflows. Each entry of query history embodies three fields: the query identifier, query topics and the arrival time of the query.

### 3. PEER NETWORK RE-ORGANISATION

#### 3.1 Re-organisation

The peer network comprises all the peers in the DDLS with each of them acting as either a resource provider or a resource requestor, or both. It needs to be maintained in a way that facilitates resource sharing and therefore may require constant re-organisation. We use the notion of information supply and demand to guide re-organisation of the peer network. Accordingly, we identify two criteria based on the notion. First, related peers should be incorporated into the same *cluster*. Second, peers should be situated in the vicinity of those that would satisfy their information needs. We explain the reason as follows. Having neighbours with related topics allows a peer to identify others that satisfy similar information need conveyed by queries and to propagate subsequent related queries. The local knowledge of *supply* available at neighbours is the key factor to the reduction of local broadcast. In cases where neither a peer nor its neighbours satisfy a query, the peer could locate probable information sources from its query history. In our approach, a peer analyses potential *demand* based on previous information needs and chooses appropriate peers that fulfill the requirement with high probability as neighbours.

We use the term *usefulness* to represent the relative extent to which a peer should be included as a neighbour of another peer during re-organisation. Assume that a candidate neighbour  $p_j$  of  $p_i$  publishes a list of topics  $T_j$ . Let  $\varepsilon_{i,j}$  be a metric which represents the information needs exhibited in query history of  $p_i$ . The naive estimator will be introduced to estimate the value of  $\varepsilon_{i,j}$  in the following section. Also, let  $\eta_{i,j}$  denote the extent to which that  $p_j$  would match the queries that  $p_i$  can satisfy:  $\eta_{i,j} = \frac{|T_i \cap T_j|}{|T_i|}$ . We use  $\nu_{i,j}$  to represent the usefulness of  $p_j$  with respect to  $p_i$ :  $\nu_{i,j} = \sqrt{(\kappa_1 \varepsilon_{i,j})^2 + (\kappa_2 \eta_{i,j})^2}$ , where  $\kappa_1$  and  $\kappa_2$  are constant coefficients associated with query history and the cached topic information, respectively.

#### 3.2 Naive Estimator-based Usefulness Decision

The ability of the naive estimator [7] to create a view of an information distribution can be approximated by individual observations from the information is of particular interest to us. We anticipate that, by using the naive estimator, individual peers in the DDLS can predict the future information needs based on their local knowledge derived from query history.

The information needs in the DDLS can be depicted by the distribution of query topics in query history. Suppose that the probability of the topics in query history of peer  $p_i$  is represented by

function  $f_{i,h}(t)$  of a discrete random variable  $t$ , where  $t$  denotes the minimum index of the same set of topics in history entries. Using the naive estimator, the estimate of probability of topics at  $t$  is  $\hat{f}_{i,h}(t) = \frac{1}{n} \sum_{k=1}^n \frac{1}{h} w(\frac{t-T_k}{h})$  with  $h = 0.5$ . Typically,  $\hat{f}_{i,h}(t)$  takes into account all query history entries of  $p_i$  within a specified period of time. We use  $\varepsilon_{i,j}$  to denote the estimate of the probability of  $T_j$ :  $\varepsilon_{i,j} = \sum \hat{f}_{i,h}(t)$ , where the estimate of the probability of  $T_j$  considers the estimate at all  $t$  where the entries are instances of  $T_j$ . We refer to a history entry with topic set  $T_a$  as an instance of another entry with topic set  $T_b$  if  $T_a$  is semantically subsumed by  $T_b$ . Therefore,  $\nu_{i,j}$  which represents the usefulness of  $p_j$  with regard to  $p_i$  is in the form of  $\nu_{i,j} = \sqrt{(\kappa_1 \sum \hat{f}_{i,h}(t))^2 + (\kappa_2 \frac{|T_i \cap T_j|}{|T_i|})^2}$ .

#### 3.3 Simulation

We performed a series of experiments to simulate re-organisation in the DDLS. It is observed that the impact from query history is predominant in reducing the number of average hops to achieve the maximum recall when the percentage of updating peers is relatively small, for instance 5%, 10% and 20% in our experiment. However, as the percentage of the updating peers further increases, the overlap information becomes more influential on the reduction in average hops than query history. The simulation result also indicates that, if excessive peers (40% and more) carry out updates over a time interval, a peer network without re-organisation outperforms that relying on either query history or the overlap information, or both.

### 4. CONCLUSIONS

We described in this paper the extended work on the DDLS which focuses on the use of the naive estimator in re-organisation of the peer network. We identify that query history and the overlap information between peers can guide re-organisation, with certain percentage of updating peers, to deliver better resource discovery performance. The anticipated behaviour and improvements are demonstrated by the simulation we have carried out.

### 5. REFERENCES

- [1] L. A. Carr, D. C. De Roure, W. Hall, and G. J. Hill. The Distributed Link Service: A tool for Publishers, Authors and Readers. *World Wide Web Journal*, 1(1):647–656, 1995.
- [2] D. Clark. Face-to-Face with Peer-to-Peer Networking. *Computer*, 34(1):18–21, 2001.
- [3] S. Goose. *A Framework for Distributed Open Hypermedia*. PhD thesis, University of Southampton, 1997.
- [4] W. Hall, H. Davis, and G. Hutchings. *RETHINKING HYPERMEDIA The Microcosm Approach*. Kulwer Academic Publishers, Dordrecht, Netherland, 1996.
- [5] T. J. Lukka and B. Fallenstein. Freenet-like GUIDs for Implementing Xanalogical Hypertext. In *Proceedings of HYPERTEXT'02*, pages 194–195, College Park, Maryland, USA, 2002.
- [6] C. Lv, P. Cao, E. Cohen, K. Li, and S. Shenker. Search and Replication in Unstructured Peer-to-Peer Systems. In *Proceedings of ICS'02*, 2002.
- [7] M. Rosenblatt. Remarks on Some Nonparametric Estimates of a Density Function. *Annals of Mathematical Statistics*, 27:832–835, 1956.
- [8] J. Zhou, V. K. Dialani, D. C. De Roure, and W. Hall. A Distance-based Semantic Search Algorithm for Peer-to-Peer Open Hypermedia Systems. In *Proceedings of PDCAT'03*, pages 7–11, Chengdu, China, 2003.