# Multidimensional Routing Indices for Efficient Distributed Query Processing

Christos Doulkeridis[1], Akrivi Vlachou[1], Kjetil Nørvåg[1], Yannis Kotidis[2], Michalis Vazirgiannis[2]
[1]Department of Computer Science, NTNU, Trondheim, Norway
[2]Department of Informatics, AUEB, Athens, Greece
{cdoulk,vlachou,noervaag}@idi.ntnu.no, {kotidis,mvazirg}@aueb.gr

## ABSTRACT

Traditional routing indices in peer-to-peer (P2P) networks are mainly designed for document retrieval applications and maintain aggregated one-dimensional values representing the number of documents that can be obtained in a certain direction in the network. In this paper, we introduce the concept of *multidimensional routing indices* (MRIs), which are suitable for handling multidimensional data represented by minimum bounding regions (MBRs). Depending on data distribution on peers, the aggregation of the MBRs may lead to MRIs that exhibit extremely poor performance, which renders them ineffective. Thus, focusing on a hybrid unstructured P2P network, we analyze the parameters for building MRIs of high selectivity. We present techniques that boost the query routing performance by detecting similar peers and grouping and reassigning these peers to other parts of the hybrid network in a distributed and scalable way. We demonstrate the advantages of our approach using large-scale simulations.

## Categories and Subject Descriptors

H.2.4 [**Database Management**]: Systems—*Query processing*

## General Terms

Algorithms, Experimentation, Performance

## Keywords

Multidimensional routing indices, P2P query processing

## 1. INTRODUCTION AND MOTIVATION

One of the most prominent applications areas of peer-to-peer (P2P) systems is content sharing and distribution. Nowadays, most of the existing systems for content distribution are P2P file-sharing networks (eMule, KaZaA) that rely on a super-peer architecture [13]. Super-peer systems are emerging as a powerful model for organizing and searching huge amounts of data, made available by independent and distributed sources. Even though there are several important applications, such as distributed image search and query processing over distributed collaborative scientific databases, which would benefit from a super-peer architecture, this has not been exploited yet. For example, consider scientists (peers) that collect massive datasets of astronomical observations or biological data and upload their descriptions to dedicated servers (super-peers), in order to share it with other scientists. Unlike plain file-sharing platforms, these applications handle multidimensional data and require advanced query types, such as nearest neighbor or range queries, in order to provide the desired functionality.

In super-peer networks, queries are typically routed first in the super-peer backbone and afterwards, if necessary, they are distributed to the peers. Routing indices [1] have been proposed for improving the performance of search in unstructured P2P networks. Traditional routing indices in P2P systems are mainly designed for document retrieval applications and maintain aggregated one-dimensional values, representing the number of documents that can be obtained in a certain direction in the network. The challenge that needs to be addressed is supporting advanced query types for applications that handle multidimensional data. Moreover, in many potential applications, it is necessary to retrieve the exact and complete result set, in contrast to typical routing indices for document retrieval, where retrieving only a fraction of the relevant data is sufficient.

Multidimensional indexing techniques have been extensively studied in centralized settings [5] for query operators like nearest neighbor, range and top-$k$ queries. Recently, there is a trend towards supporting advanced query processing in P2P networks as well [3, 8, 12], using multidimensional summary information for directing queries in the network in a deliberate way. Similar to centralized multidimensional indices, data stored on peers is described by representative (multidimensional) data descriptors, realized as minimum bounding regions (MBRs). These can be, for example, hyper-rectangles that enclose all data points on a peer, however other types of MBRs, such as spherical MBRs, can be supported.

Multidimensional routing indices (MRIs) are composed by different local multidimensional indices stored at each super-peer. We can note the distinct difference between a MRI and a distributed multidimensional index [4]. In the latter case, the nodes of a single index are themselves distributed on different servers. In contrast, the data objects stored by the

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.
*CIKM'09,* November 2–6, 2009, Hong Kong, China.
Copyright 2009 ACM 978-1-60558-512-3/09/11 ...$10.00.

MRIs on a super-peer are MBRs that summarize the data available through each neighboring super-peer. The simplest form of a MRI at a super-peer, is a list of MBRs for each neighboring super-peer. Each super-peer maintains its own MRI and utilizes it in order to efficiently route a query through the super-peer network. The MRIs are constructed in a distributed manner. A neighbor is informed about a super-peer's content by receiving a set of MBRs, representing the data that can be retrieved if a query is forwarded in that direction. This information is then aggregated with its own MBRs and forwarded to the rest of the super-peer network.

An important factor that influences the performance of multidimensional routing indices is the underlying data distribution to peers (and super-peers). Although data on peers is usually clustered into a few thematic areas that reflect the user's interests, when peers join the network by connecting to a randomly chosen super-peer, the super-peers end-up indexing MBRs spread all over the data space. This leads to what we call near-uniform data distribution at super-peer level, which degrades query processing performance, since a query may – in worst case – have to contact all super-peers. It is therefore necessary to change the data uniformity at the super-peer level, by detecting and reassigning peers with similar content to the same super-peer, thus generating an overlay network that reflects a clustered data distribution at super-peer level. Then, MRIs of better quality are constructed and queries can be directed to specific super-peers only, thus improving query processing performance. An equally important parameter that affects query processing is having neighboring super-peers maintain similar data. Again, this should occur in a self-organizing way, without explicit intervention from the P2P system designer.

The main topic of this paper is how to generate efficient multidimensional routing indices for query processing in P2P systems. Focusing on a super-peer network, we outline techniques for the identification of peers with similar content and reassignment to the same super-peer. Our approach is self-organizing, in that there is no prior assignment of space partitions or data to each super-peer, but we aim to dynamically capture the data distribution and reassign peer MBRs to different super-peers. This reassignment takes into account the super-peer topology, so that neighboring super-peers index peers with similar content. We propose a distributed and scalable approach that boosts the query routing performance since the multidimensional routing indices built on the new clustered network topology have a higher selectivity and fewer super-peers are contacted during query processing.

## 2. RELATED WORK

Routing indices [1] have been originally proposed in the context of unstructured P2P systems, as a search-enabling technique that provides a direction to routing, instead of blind forwarding. Hose et al. use distributed data summaries as routing indices [8]. They propose a tree-based structure, named QTree, for summarizing data and discuss requirements that need to be fulfilled by such routing structures. P2P data summaries have also been proposed in [7]. SIMPEER [3] is a state-of-the-art approach for exact query processing over multidimensional data distributed in a super-peer network. Sharing similar goals to this paper, such as exactness of retrieval, SIMPEER relies on the construction of routing indices to process similarity search queries.

Content addressable network (CAN) [10] was the first approach for P2P multidimensional indexing. Space partitioning, based on distributed tree structures, has been also considered. In [6], the authors identify two primary components for multidimensional query processing, namely partitioning and routing. They propose an adaptation of the kd-tree for partitioning and use of skip pointers for routing. VBI-Tree [9] is a framework for multidimensional indexing in P2P networks. Peers are organized in a balanced tree structure based on the data partitions assigned to them.

## 3. ROUTING INDICES

In the following we assume a super-peer network [3, 13], that consists of $N_{sp}$ super-peers, each connected to a limited set of at most $DEG_{sp}$ other super-peers. Super-peer networks consist of many simple peers and few enhanced super-peers, in terms of processing power, storage capacity or network connectivity. Each super-peer $SP_i$ is responsible for $DEG_p$ simple peers, which connect to $SP_i$ directly. The initial assignment of peers to super-peers is random with respect to peers' contents [3, 13]. Each peer $P_i$ holds $n_i$ $d$-dimensional points, denoted as a set $O_i$ $(1 \leq i \leq N_p)$. Assuming horizontal data distribution to the $N_p$ peers, the size of the complete set of points is $n = \sum_{i=1}^{N_p} n_i$ and the dataset $O$ is the union of all peers' datasets $O_i$ $(O = \cup O_i)$. Each peer maintains its own data objects, while the $d$-dimensional points are features extracted from the objects. In Figure 1 (left) we depict for example the two-dimensional data stored by peers $P_A$ and $P_B$, which are connected to super-peer $SP_C$.

We are interested in supporting *exact query processing* on top of the super-peer network, which means that we would like to process a query in a distributed manner, but the answer set should be the same as if the query had been executed on the dataset $O$ in a centralized setting. In the rest of the paper, we describe how range queries are processed, merely as a showcase example, although other query types can be supported as well.

**Range query** $R(q, r)$**:** Given a query object $q$ and a radius $r$, a point $p \in O$ belongs to the result set of the range query if $dist(q, p) \leq r$.

Efficient query processing over high-dimensional data, distributed in a super-peer network, is accomplished using multidimensional routing indices at super-peer level. *Local indexing* of peer MBRs at a super-peer is performed in the following way. Assume that each peer's $P_i$ data is described by a list of $k_i$ MBRs, which enclose all data objects on peer $P_i$. Each super-peer gathers the MBRs of its $DEG_p$ associated peers and stores them locally using any available centralized multidimensional indexing data structure, such as an R-Tree. In addition, each super-peer creates a list of *aggregated MBRs*[1], based on the collected MBRs of its peers.

The aggregated MBRs are exchanged between the super-peers and used for constructing *multidimensional routing indices*. The routing index $MRI$ at super-peer $SP_i$ is a set of $DEG_{sp}$ entries $MRI = \{S_1, \ldots, S_{DEG_{sp}}\}$, one for each neighboring super-peer. Each entry $S_j$ consists of $k_j$ MBRs and is associated with a super-peer $SP_j$ that is a neighbor of $SP_i$. Any peer MBR that is accessible through super-peer

[1]Any technique proposed in centralized settings for creating MBRs that enclose multidimensional data may be used.

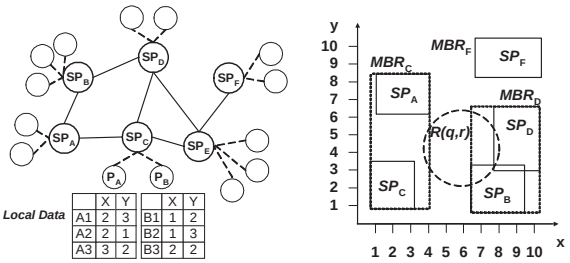



**Figure 1: Super-peer network and MRI at $SP_E$.**

$SP_j$ is enclosed by an MBR of entry $S_j$. Efficient storage and querying of the entries $S_j$ of an MRI can again be performed using an R-Tree. As an example, consider Figure 1 (right), where the MRI stored at $SP_E$ is depicted, assuming that $k_j$=1.

At query time, each super-peer is able to decide if a neighboring super-peer can contribute to the final result set, by using its multidimensional routing index. Consider again the example depicted in Figure 1 (right). Assume a range query $R(q, r)$ initiated at a super-peer $SP_E$, also called the querying super-peer. $SP_E$ decides which neighbor to forward the query to, based on the MBRs stored in its multidimensional routing index. The query is forwarded to all neighbors, for which at least one MBR overlaps with the query. This constitutes the *query routing* mechanism.

Furthermore, by storing its peers' MBRs, $SP_E$ is able to determine during query processing at least those peers that contribute to the query result set. Thus, $SP_E$ processes the query based on its peers MBRs and this process is called *local query processing* at $SP_E$. $SP_E$ contacts only peers responsible for MBRs that overlap with the query. These peers process the query based on their local data, and return their results to $SP_E$.

In our running example, after examining the MBRs of its neighbors, $SP_E$ decides to forward the query to its neighbors $SP_C$ and $SP_D$, while $SP_F$ is not contacted, since its MBRs do not overlap with the query. Notice, that $SP_C$ neither processes the query locally nor forwards it to another super-peer, because neither $SP_C$ nor $SP_A$ can contribute to the query. Actually, $SP_C$ is contacted because of the enlargement of the MBR, due to the MBRs aggregation process. On the other hand, $SP_D$ will both process the query locally and forward the query to $SP_B$.

This simple example demonstrates how the aggregation of dissimilar MBRs affects the quality of MRIs. Therefore, effective algorithms are required that reassign peers to super-peers, in such a way that similar MBRs are aggregated, thus leading to construction of qualitative MRIs. Towards this goal, we introduce two techniques for peer reassignment, which are applied prior to the actual construction of MRIs.

1. **Changing the data distribution.** The first challenge that needs to be addressed effectively is to change the distribution of peers to super-peers based on the stored data. For this purpose, each local MBR computed by a peer is reassigned to a super-peer based on similarity, in order to generate a clustered data distribution at super-peer level. Therefore, prior to the construction of the multidimensional routing indices, data aggregation is necessary, as a means to identify similar peers across the entire network. This distributed aggregation process is achieved using a content-aggregation hierarchy [2]. The super-peers are organized in a hierarchical overlay, aggregating MBRs of super-peers (or peers) of the level below. In the end, $N_R$ top-level MBRs are created at the root of the hierarchy, which describe the data available in the entire network. In the next step, the $N_R$ MBRs are dynamically decomposed to eventually form $N_{sp}$ ($N_{sp} > N_R$) groups (i.e. sets) of MBRs, essentially one group for each super-peer. The similarity between MBRs is quantified by appropriate measures and the decomposition aims to maximize the similarity of the MBRs assigned to each group, while minimizing the similarity of MBRs that belong to different groups.

2. **Mapping of MBRs to super-peers.** After the issue of discovering similar peer data has been efficiently resolved, peer MBRs are clustered into $N_{sp}$ groups based on their similarity. The remaining challenge is to assign the $N_{sp}$ groups to the super-peers, in such a way that is beneficial for query processing. A naive way is to pick for each group a super-peer randomly. However, this approach can result in a situation where neighboring super-peers index dissimilar MBRs of peers. This influences the aggregated MBRs of neighboring super-peers, and hence, the selectivity of the multidimensional routing indices. Therefore, the aim is to assign groups to super-peers, in such a way that neighboring super-peers index peers with similar content. The problem of group assignment to super-peers can be modeled in the following way. The super-peer network topology is an undirected graph that consists of $N_{sp}$ vertices and having the super-peer connections as edges. We also define an undirected fully-connected, weighted graph, mentioned also as similarity graph, that expresses the similarity between the groups. Then, the problem of group assignment to super-peers can be mapped to a *weighted graph matching problem* [11] defined based on the network topology graph and the similarity graph.

As soon as the peer reassignment is completed, the hierarchy is not maintained anymore, and MRIs are constructed at super-peer level, using the exact same construction method as in the case that no peers are reassigned.

# 4. EXPERIMENTAL STUDY

In this section we study the efficiency of the proposed approach using a simulator prototype implemented in Java. In our experiments, we used the GT-ITM topology generator[2] to create well-connected random graphs of $N_{sp}$ super-peers with average connectivity $DEG_{sp}$=4. In order to evaluate the scalability of our approach we experimented with synthetic clustered data collections. For the clustered dataset generation, we randomly pick $N_{sp}$ $d$-dimensional points and each peer obtains $k$ distinct centroids from them at random. Obviously, two peers may share the same centroid. Thereafter, the peers' objects are generated inside a radius $R$ from one of the centroids, selected in a way, such that the total volume of the data capture 10% of the data space. We conduct experiments varying the dimensionality (2-6d). We

[2]Available at: http://www.cc.gatech.edu/projects/gtitm/

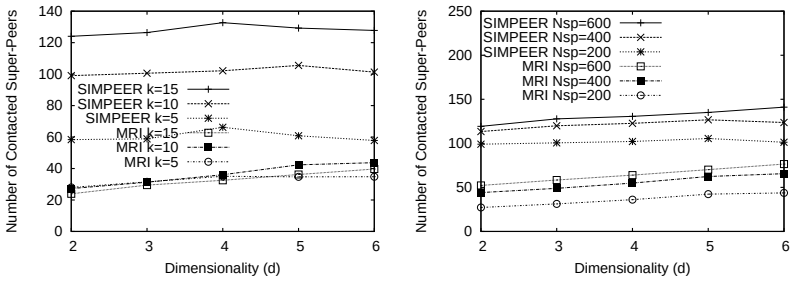



**Figure 2: Scalability study of multidimensional routing indices quality.**

keep the number of objects per peer ($n/N_p$) constant and equal to 500. Peer data are represented by spherical MBRs, and they are determined using the k-means clustering algorithm. MBR aggregation is performed using the k-means algorithm on the MBRs centroids. Our default setup consists of a network of 2000 peers, 200 super-peers, 1M data objects, $DEG_p$=10 and $k$=10. All experiments were repeated using 10 different synthetically generated datasets and the average number of the measurements is depicted in all cases. For comparative purposes, we use SIMPEER [3] as query processing mechanism, which is a state-of-the-art approach for multidimensional query processing over a super-peer architecture. We study the improvement in query processing performance achieved by the MRI approach, compared to SIMPEER without peer reassignment.

In Figure 2 (left), we evaluate the routing ability of the MBRs by simulating range queries. We select randomly 20 MBRs out of the original peer MBRs, and consider them as range queries that follow the data distribution. We measure the average number of groups that contain an overlapping MBR with the queries or, equivalently, the minimum number of super-peers that have to be contacted during query evaluation. The chart shows that MRI requires only 24 to 43 super-peers out of 200 to be contacted, in order to retrieve the exact result. However, SIMPEER requires many more super-peers, ranging from 57 to 130, depending on the selected value of $k$. Figure 2 (right) depicts the number of super-peers with results, while we increase the super-peer network size from 200 to 600 super-peers. We present the number of super-peers with results, compared to SIMPEER. MRI is always much better than SIMPEER, as fewer super-peers need to be contacted, and more importantly this gain is maintained as the number of super-peers increases, demonstrating the scalability of MRI.

# 5. CONCLUSIONS

In this paper, we introduced the concept of multidimensional routing indices in P2P networks and we addressed the challenging problem of improving their performance and selectivity. We identified an important factor that affects the performance of query routing and processing, namely the uniformity of data distribution at super-peer level. We proposed efficient and scalable techniques for changing the data distribution at super-peers in a self-organizing manner, so that peers with similar content connect to the same super-peer and neighboring super-peers index similar content. The experimental results show that our approach improves P2P similarity search.

# 6. ACKNOWLEDGMENTS

The work of Christos Doulkeridis and Akrivi Vlachou was carried out during the tenure of an ERCIM "Alain Bensoussan" Fellowship Programme.

# 7. REFERENCES

[1] A. Crespo and H. Garcia-Molina. Routing indices for peer-to-peer systems. In *Proc. of ICDCS*, page 23, 2002.

[2] C. Doulkeridis, K. Nørvåg, and M. Vazirgiannis. DESENT: Decentralized and distributed semantic overlay generation in P2P networks. *IEEE Journal on Selected Areas in Communications (J-SAC)*, 25(1):25–34, 2007.

[3] C. Doulkeridis, A. Vlachou, Y. Kotidis, and M. Vazirgiannis. Peer-to-peer similarity search in metric spaces. In *Proc. of VLDB*, pages 986–997, 2007.

[4] C. du Mouza, W. Litwin, and P. Rigaux. SD-Rtree: A scalable distributed R-tree. In *Proc. of ICDE*, pages 296–305, 2007.

[5] V. Gaede and O. Günther. Multidimensional access methods. *ACM Comput. Surv.*, 30(2):170–231, 1998.

[6] P. Ganesan, B. Yang, and H. Garcia-Molina. One torus to rule them all: Multidimensional queries in P2P systems. In *Proc. of WebDB*, pages 19–24, 2004.

[7] R. Hayek, G. Raschia, P. Valduriez, and N. Mouaddib. Summary management in P2P systems. In *Proc. of EDBT*, pages 16–25, 2008.

[8] K. Hose, C. Lemke, and K.-U. Sattler. Processing relaxed skylines in PDMS using distributed data summaries. In *Proc. of CIKM*, pages 425–434, 2006.

[9] H. V. Jagadish, B. C. Ooi, Q. H. Vu, R. Zhang, and A. Zhou. VBI-tree: A peer-to-peer framework for supporting multi-dimensional indexing schemes. In *Proc. of ICDE*, page 34, 2006.

[10] S. Ratnasamy, P. Francis, M. Handley, R. M. Karp, and S. Shenker. A scalable content-addressable network. In *Proc. of SIGCOMM*, pages 161–172, 2001.

[11] S. Umeyama. An eigendecomposition approach to weighted graph matching problems. *IEEE Trans. Pattern Anal. Mach. Intell.*, 10(5):695–703, 1988.

[12] A. Vlachou, C. Doulkeridis, K. Nørvåg, and M. Vazirgiannis. On efficient top-k query processing in highly distributed environments. In *Proc. of SIGMOD*, pages 753–764, 2008.

[13] B. Yang and H. Garcia-Molina. Designing a super-peer network. In *Proc. of ICDE*, page 49, 2003.