Heterogeneity-Aware Distributed Access Structure

Alejandra González Beltrán Peter Milligan Paul Sage
School of Computer Science, Queen's University Belfast
18 Malone Road, Belfast, United Kingdom, BT7 1NN
{a.gonzalez-beltran, p.milligan, p.sage}@qub.ac.uk

Abstract

Efficient access to distributed and dynamic multidimensional data is vital for applications in large, heterogeneous, decentralised, resource-sharing environments such as Grids and Peer-to-Peer systems. Most systems providing this functionality assume homogeneous participants. This paper proposes HADAS, an access structure exploiting heterogeneity to build a self-aware adaptive information system.

1. Introduction

Accessing distributed and dynamic high-dimensional data in a timely fashion is an important function for large, decentralised, resource-sharing environments such as Grids and Peer-to-Peer (P2P) systems. A system providing such functionality should allow for complex queries such as range and similarity searches.

The use of P2P systems, both unstructured [5] and structured [7], has been proposed for information discovery because they provide a scalable, self-organising substrate for fully decentralised applications. In this work, structured networks are used as they support bounded searches with guaranteed results. Among these, there have been two approaches: providing distributed hash table (DHT) and distributed search tree (DST) functionality. Some systems supporting flexible information discovery are based on DHTs [7]. However, the use of hashing in DHTs destroys the ordering of keys, making range and similarity queries inefficient. DST systems such as [2, 4] overcome this limitation, allowing for range queries.

Although P2P and Grid environments are characterised by the heterogeneity of the resources (locality, storage capacity, connectivity, etc.), systems for information services [5, 7] have assumed that nodes have equal responsibilities.

This paper proposes a Heterogeneity-Aware Distributed Access Structure (HADAS), which is a multidimensional index built on top of a structured P2P network. The main contribution is a method to exploit nodes' heterogeneity by storing reflective information in the structure itself, used to build and maintain the overlay network.

2. Routing Substrate

The heterogeneity-aware overlay could be implemented using any structured P2P network as routing substrate. This paper uses SkipGraph/SkipNet [2, 4], which define basically the same randomized overlay network based on skip lists with logarithmic (in the number of nodes) degree and lookup performance. The nodes are connected as a collection of circular skip lists or rings. Each node is assigned two identifiers: a *numeric ID* randomly and uniformly chosen, which encodes ring membership (at level h ring, nodes' *numeric IDs* share a common prefix of h bits), and an arbitrary *name ID*, whose order is preserved in each ring.

3. Data and Query Management

Data Representation. A distributed entity (resource, service or application) is characterised by a set of features, which can be seen as a point in the *multidimensional space* $[0, 1]^d$, by transforming each feature into a value in $[0, 1]$.

Space Partitioning. Database [3] and P2P systems [7] have used space-filling curves (SFCs) for multidimensional indexing. SFCs are recursively defined continuous mappings from a multidimensional space to a linear ordering, which preserve data locality. HADAS uses the Z-SFC ($Z : [0, 1]^d \to [0, 1]$). The ordered space $[0, 1]$ constitutes the *name ID* space for SkipNet's nodes. Each node is in charge of the region in $[0, 1]^d$ whose Z-image is the interval determined by the *name ID* of its predecessor in the lowest level ring and its own *name ID*.

Instead of using bit-interleaving for obtaining $Z(x)$ and compare nodes' *name IDs*, HADAS uses a comparison routine[1] that determines the relative position of two points in the Z-order without computing their Z-images. This rou-

Proceedings of the Fifth IEEE International Conference on Peer-to-Peer Computing (P2P'05)
0-7695-2376-5/05 $20.00 © 2005 IEEE

tine has better performance and is independent of the data distribution.

Query Management. HADAS supports several types of queries (range, k-nearest-neighbour, prefix search), combining searching methods in the high-dimensional space [3] and query distribution in the routing substrate, as in [7].

4. Heterogeneity-Aware Overlay Network

HADAS allows the storage and efficient search of high-dimensional data objects from different data sets. Its nodes are dynamic, heterogeneous and geographically distributed. The key idea proposed to enable exploitation of heterogeneity is to use HADAS to store information about its own nodes' state, such as storage capacity, current load, locality, processing power and network bandwidth. This information composes a *reflection* data set that makes the access structure *self-aware*. In order to build this data set, nodes periodically store their state in HADAS and independently determine when this data should be updated.

The reflection data set is used in the overlay network construction and maintenance to provide desirable properties that can be customised for each application.

In this paper, attention is focused on the properties: storage load balance (each node's storage load has to be proportional to its storage capacity) and physical network proximity (consecutive nodes in the SkipNet's lower-level ring are physically close in the underlying IP network) while considering nodes' availability.

Then, each tuple representing a node's state contains: the load percentage (ratio between current number of objects stored and its storage capacity), a binary number encoding the ordering of a set of relatively stable landmark machines in order of increasing round-trip times from the node [6] and a number representing availability (the greater the value, the longer the node has been connected in HADAS).

The node joining process starts by determining reflection information for the joining node. Then, through a known member of HADAS or gateway, it queries the structure for the objects in the reflection data set which satisfy the condition that their load percentage is greater than or equals to that of the gateway[1] and have the same ordering of landmark nodes. This results in a set of nodes that are probably overloaded and physically close to joining node. Further selection locally determines the node with highest availability. This node is chosen to proportionally partition its region according to storage capacities. Thus, the new node's *name ID* is determined, which encodes heterogeneity information, and the SkipNet joining procedure can start. Online repair mechanisms dedicated to preserve the desired properties, adapting HADAS as changes are produced, also take advantage of the reflection data set. Thus, instead of using gossip-based mechanisms to contact its neighbours, a node whose load has recently increased, queries HADAS to find close nodes whose load percentage is less than its own. From the resulting nodes, it carefully choses the most appropriate one to share its load, so that it causes the minimum disruption in terms of space partitioning and data movement.

5. Discussion

In summary, HADAS is a distributed index allowing multi-attribute complex queries on arbitrary data sets. During network construction and maintenance, HADAS holds information about its own nodes to exploit heterogeneity and simultaneously satisfy properties such as storage load balance and network proximity. HADAS can support other properties by extending the reflection data set and node selection criteria can impose priorities on them.

HADAS development is in progress and follows the Bamboo DHT[2] model, which enables system evaluation by using its simulator, cluster emulation[3] and execution on PlanetLab[4]. Thus, the use of Bamboo as routing substrate will be straightforward and will allow the performance comparison of DHTs and DSTs. Another outcome will be an analysis of the trade-off between effectiveness, with respect to the target properties, and bandwidth costs incurred by making the structure self-aware.

Acknowledgments. The authors are grateful to APG QUB for funding the work of Alejandra González Beltrán.

References

[1] S. Aluru and F. Sevilgen. Efficient Methods for Database Storage and Retrieval Using Space-Filling Curves. In *ISCIS*, 2004.

[2] J. Aspnes and G. Shah. Skip Graphs. In *SODA*, 2003.

[3] C. Böhm, S. Berchtold, and D. Keim. Searching in High-Dimensional Spaces. *ACM Computing Surveys*, 33(3):322–373, 2001.

[4] N. Harvey, M. Jones, S. Saroiu, M. Theimer, and A. Wolman. SkipNet: A Scalable Overlay Network with Practical Locality Properties. In *USITS*, 2003.

[5] A. Ianmitchi and I. Foster. *Grid Resource Management*, chapter 25, pages 413–429. Kluwer, 2004.

[6] S. Ratnasamy, M. Handley, R. Karp, and S. Shenker. Topologically-aware Overlay Construction and Server Selection. In *INFOCOM*, 2002.

[7] C. Schmidt and M. Parashar. Enabling Flexible Queries with Guarantees in P2P Systems. *IEEE Internet Computing*, 3(8):19–26, 2004.

[1] Assuming the storage load balance property is true, all nodes should have roughly the same load percentage. Thus, the nodes whose load percentage exceeds that of the gateway are likely to be overloaded.

[2] Bamboo DHT http://www.bamboo-dht.org

[3] ModelNet http://issg.cs.duke.edu/modelnet.html

[4] PlanetLab http://www.planet-lab.org