University of Augsburg
Prof. Dr. Hans Ulrich Buhl
Research Center
Finance & Information Management
Department of Information Systems
Engineering & Financial Management

Universität
Augsburg
University

Discussion Paper WI-158

Complexity of Distributed Covariance Calculations

by

Gilbert Fridgen, Wolfgang Hackenbroch

July 2005

to be presented at: 7th International Conference on Optimization:
Techniques and Applications (ICOTA7), Kobe, (Japan), December 2007

Universität Augsburg, D-86135 Augsburg
Visitors: Universitätsstr. 12, 86159 Augsburg
Phone: +49 821 598-4801 (Fax: -4899)
www.fim-online.eu

Complexity of Distributed Covariance Calculations

Gilbert Fridgen
University of Augsburg, 86135 Augsburg, Germany
Gilbert.Fridgen@wiwi.uni-augsburg.de

Dr. Wolfgang Hackenbroch
University of Augsburg, 86135 Augsburg, Germany
Wolfgang.Hackenbroch@wiwi.uni-augsburg.de

Abstract

Risk/return management has evolved as one of the key success factors for enterprises especially in the financial services industry. It is highly demanding in terms of business requirements and technical resources, making it an almost ideal application for distributed computing concepts like e.g. grid computing. In this paper we focus on a specific problem—the estimation of covariance matrices that provide a powerful tool for decisions on investments. In this context we analyze different network topologies that the corresponding calculations can be performed on. We derive complexity classes for a distributed calculation scenario on these topologies. As a general result we find an upper and lower bound for the complexity of a distributed calculation in an arbitrary network structure. These results not only provide a different view on grid resource allocation but also make a contribution towards better understanding the business value of grid computing.
Keywords: Grid Computing, Risk Management, Parallel Algorithms

1 Introduction

Whereas in the beginning grid computing was restricted to large-scale scientific applications like in high-energy physics, astronomy, or biology, it has over the past years evolved to an increasingly relevant technology for the commercial sector as well. It is to some extent difficult to differentiate grid computing from related concepts of e.g. distributed computing, cluster computing, utility computing, or Service-Oriented Architectures (SOA). Yet even on a grid infrastructure resources are not unlimited thus it is necessary to consider the complexity associated with the corresponding calculations. For our analysis we focus on a specific problem in risk/return management—the calculation of covariance matrices that, on the one hand, provide a powerful tool for decisions on investments but, on the other hand, are a very complex and time-consuming assignment.

2 Modelling Approach

Distributing the calculation of covariance matrices on a grid infrastructure and computing them in parallel is not a trivial task to achieve. An algorithm must tackle questions like which grid resources are allocated to the calculation of certain parts of the covariance matrix and how the input data necessary to perform these calculations can be distributed to these grid resources over the network. Both questions are inseparably intertwined since calculating part of the matrix requires only part of the input data and unnecessary transportation of input data should be avoided.[1]

We restrict our considerations to the communication complexity, meaning the number of messages necessary to perform the distributed calculation. It is not surprising that communication complexity varies for different grid network *topologies*. Therefore we propose algorithms for a set of standard topologies to find their respective complexities. Later on we try to find a general way to determine the upper and lower bounds for the given problem on an arbitrary topology.

[1] According to the common perception of grid computing these tasks are usually done by something like a 'grid middleware'. However, as we will show in the following the proposed problem requires too much domain knowledge for a standard grid middleware to distribute it efficiently.

3 Results

Table 1: Complexity classes for different network topologies

Fully Connected	$O(n\sqrt{n})$
Star Topology	$O(n\sqrt{n})$
Line Topology	$O(n^2)$
Ring Topology	$O(n^2)$
Tree Topology	$O(n\sqrt{n})$

Looking at the standard network topologies we find very similar algorithms for the fully connected and star topologies and for the line and tree topologies respectively. While for the fully connected and star topologies the algorithms share the same complexity class they differ for the tree and line topology (see table 1).

Assuming optimality for these algorithms there are some interesting results worth to be pointed out. There seems to be no clear coherence between complexity and the number of edges relative to the number of nodes. One could presume that the complexity is lower for a higher connected topology (one with more edges) but the ring topology in fact accounts for a higher complexity than e.g. the tree topology even though it uses one more edge.

Our results rather show that from the perspective of every node, the marginal distance of other nodes is crucial for the complexity of the problem as a whole. This provides us with the means to evaluate arbitrary topologies and—by focussing on theoretically possible minimal and maximal marginal distances—to quantify the overall complexity class of the problem: We find our problem to be $\Omega(n\sqrt{n})$ and $O(n^2)$ and we can show optimality for the proposed algorithms for standard topologies.

4 Conclusion

In this paper we restricted our analysis to one well-defined problem: the grid-based calculation of covariance matrices. Although covariances are widely used in financial applications we thereby covered only a small subset of risk/return management methods and algorithms. Other approaches and applications for grid computing (like e.g. Monte-Carlo simulations which have a high parallelization potential as well) still have to be evaluated.

We considered different grid network topologies and scrutinized suitable algorithms for covariance calculation on these network structures. From there we derived the corresponding complexity classes for a distributed calculation on each topology. The basic algorithms proposed in this paper assume a very regular topology and an omniscient coordinator. Therefore none of them will be directly applicable to a real company's complex infrastructure. Nevertheless, when perceiving the nodes of the topology not as single workstations but as whole subsidiaries of a large enterprise a e.g. line topology is absolutely realistic. Furthermore the resulting complexity classes could in combination with a quantification for the benefits of the calculation be utilized for the determination of an optimal investment in covariance calculations.

For an arbitrary topology these results are as well meaningful: For example when designing a specialized algorithm for a company's network infrastructure one can apply the insights generated by the algorithms proposed (e.g. that it makes sense to send as few packages to the outmost nodes as possible). Furthermore even if in a company's network each node's effort to receive data grows in a logarithmic fashion (e.g. if each node's degree is greater than 2) one can use $O(n\sqrt{n})$ as a benchmark for the design of a specific covariance calculation method.