Proceedings of the *International Conference on Theory and Application of Mathematics and Informatics ICTAMI 2005* - Alba Iulia, Romania

GAME THEORY IN DATA ALLOCATION FOR DISTRIBUTED DATABASES

Horea Adrian Grebla

Abstract. This paper studies distributed database performance optimization by mapping data fragmentation and allocation to load balancing games. The system is modelled as a set of network locations (agents) who share several resources (data fragments). The aim of the agents is to perform a strategy that maximizes their gain but optimizes systems performance too. The system optimizes and becomes balanced if it is constructed a Nash equilibrium, so the problem is to find sequences of utility-improving moves that lead to a Nash equilibrium, starting from some given assignment of resources to agents. The contribution of this paper represents a parallel comparison of 2 types of methods in resource allocation games that we consider to be suitable to our problem: non-cooperative versus cooperative strategies.

2000 *Mathematics Subject Classification*: 91A80, 68U99.

1. Introduction

Distributed databases (DDBs) still represent a research field as networks expand, organizations store critical information on geographically distributed locations and web technologies incorporate remote administration for databases, allow distributed queries and supply native transparency levels for users. Stored data at different sites of a computer network can have a variety of forms, ranging from flat files, to hierarchical, relational or object-oriented databases. Distributed databases architectures still present some management and optimization challenges. The system needs to be highly scalable with no critical points of failure. Also the latency must not affect the performance of applications. The aim is to provide uniform access to physically distributed data to all users, no mater what the distance between the access location and places data resides, no matter what network access user poses.

2. Distributed Database Design Issues

Distributed database management system has to ensure [1] local applications for each computational component as well as global applications on more computational machines; it also has to provide a high-level query language with distributed query power, for distributed applications development. Must be ensured transparency levels that confer the image of a unique database. To improve the performance data can be partitioned and stored on the systems components. A distributed database system allows data fragmentation if a relation stored within can be divided in pieces called fragments that can be stored on different sites residing on the same or different machines. The purpose of data fragmentation is to store the fragments closer to where they are more frequently used to achieve best performance. The partitions can be created horizontal, vertical or mixed [2](the combination of horizontal and vertical fragmentation). Let $R[A_1,A_2,\ldots,A_n]$ be a relation where A_i, $i = 1,\ldots,n$ are attributes. A horizontal fragment ca

$$R = R_1 \cup R_2 \cup \ldots \cup R_k.$$

A vertical fragment is obtained by a projection operation:

$$R_i = \prod_{\{A_{x1},A_{x2},\ldots,A_{xp}\}} (R),$$

where A_{xi}, $i = 1,\ldots,p$ are attributes. The initial relation can be reconstructed by join of the fragments:

$$R = R_1 \otimes R_2 \otimes \ldots \otimes R_l.$$

From the perspective of information exchange we consider the DBMS independent approach, which allows construction of open systems, with good scaling capabilities in heterogeneous environments. A DDB system can be represented [3] as a graph where the sites are given by (V), the set of vertices, and the edges (E) given by the direct connections between sites. An example of a distributed database system is depicted in Fig.1

The system must preserve distributed data independence [1], such that any change of physical location of data must not disturb application functionality. The way to achieve this independence is to build a Global Directory that registers data placed on each site.

3. Problem description

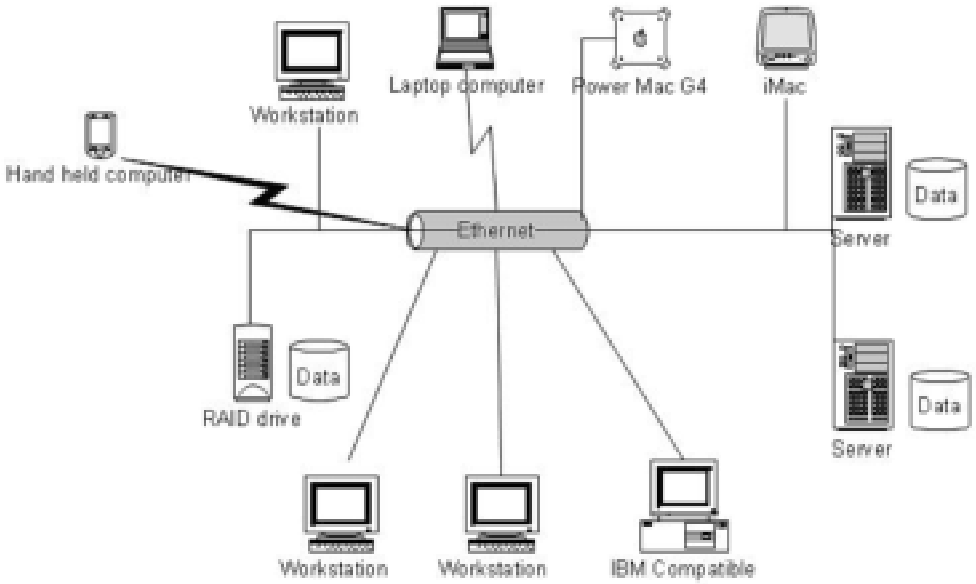

Figure 1: Distributed Database System

The allocation of resources across is an old problem that has been studied extensively.

Assume that there are a set fragments $F = \{F_1, F_2, .., F_n\}$ and a network consisting of sites $S = \{S_1, S_2, \dots, S_m\}$ on which a set of applications $Q = \{Q_1, Q_2, \dots, Q_p\}$ is running.

The allocation problem is to find the optimal distribution of F to$S$. The optimality can be defined with respect to :

• Minimal cost , the cost of storing Fi at Sj, the cost of querying Fi at Sj, the cost of updating Fi where it is stored, and the cost of data communication.

• Performance , to minimize the response time and to maximize the system throughput at each site

The allocation problem can be specified as a cost-minimization problem:

$$\min[\sum_{i=1}^{m}(\sum_{j|S_j\in I} x_j u_j c_{ij} + t_j \min_{j|S_j\in I} c_{ij}) + \sum_{j|S_j\in I} x_j d_j]$$

This formulation [4] has proven to be NP complete.

4. Game Theory and Resource Sharing

Game theory studies humanan behavior from the perspectives of the disciplines involved in: mathematics, economics and the other social and behavioral sciences. Neoclassical economics [5] assumes that humans are rational in their choices such that each person maximizes its profits in all circumstances. This represents the basics in the study of the allocation of resources. It evaluates the efficiency of a system such that if the costs are greater than benefits then another strategy must be chosen in order to maximize benefits. In multiplayer games the choice of a player means the strategy it addopts, and the outcome of the interactions depend on all choices of the participants, that means on the combined strategies. One can identifie two possible solutions: a ”non-cooperative” solution in which each person maximizes its own benefits , and a ”cooperative” solution in which the strategies of the players are chosen so as to obtain the best result for the group.

4.1. Cooperative games

Games in which the participants can make commitments to coordinate their strategies are ”cooperative games” and have ”cooperative solutions”. In a cooperative game, the problem is to choose a strategy that leads to the best outcome for all players. It has to be defined a criterion [5] to rank outcomes from the point of view of the group of players as a whole. It can be said that one outcome is better than another if at least one player is better off and no-one is worse off. This is called the Pareto criterion, after the Italian economist and mechanical engineer, Vilfredo Pareto. If an outcome can not be improved upon, if no-one can be made better off without making somebody else worse off, then itis said that the outcome is Pareto Optimal. If there were a unique Pareto optimal outcome for a cooperative game, that would seem to be a good solution concept, but, in general, there are infinitely many Pareto Optima for most complicated game. All the same, this was the solution criterion that von Neumann and Morgenstern used, and the set of all Pareto-Optimal outcomes is called the ”solution set”. One economic problem that still present good research activity is allocation of resources in a distributed environment. ”Allocation” is an economic term, and economists are often concerned with the efficiency of allocations. The standard definition of efficient allocation in economics is ”Pareto optimality”. In defining an efficient allocation, it is best to proceed by a double-negative. An allocation is inefficient if there is at least one player who can do better, while no other player is worse off. Conversely, the allocation is efficient in the Paretian sense if no-one can be made better off without making someone else worse off.

4.2. Non-cooperative games

Non-cooperative game theory provides [5] a normative framework for analyzing strategic interactions of agents. There has been interest in the computational complexity of natural questions in game theory. Starting in the 1970s, theoreticians have focused on the complexity of playing particular highly structured games (usually board games, such as chess or Go). These games tend to be alternating-move zero-sum games with enormous state spaces, which can nevertheless be concisely represented due to the simple rules of the transition between states. As a result, effort on finding results for general classes of games has often focused on complex languages in which such structured games can be concisely represented.

5. Suitable solutions to the problem:

Cooperative approach: The problem can be modelled as the market game; in a merket game we deal with a notion called core; The core of a cooperative game consists of all undominated allocations in the game. In other words, the core [5] consists of all allocations with the property that no subgroup within the coalition can do better by deserting the coalition. An allocation in the core of a game will always be an efficient allocation. it is illustrated by the market game, a game of exchange. Consider a market [9] comprising n buyers and m items. Each buyer has an endowment given by a portfolio of items. A finite quantity of each item is available and is assumed to be divisible. Further, each (buyer, item) pair has an associated utility function. This function is assumed to be non-negative and linear. The market equilibrium problem is to compute a price vector and a feasible assignment of goods to buyers such that no buyer is induced to change his assignments with respect to the given set of prices and market clearing is achieved, i.e. there is no surplus or deficit of the goods.

Non-cooperative approach: The allocation problem can be viewed as Santa Fe Bar Problem; in Arthur.s original simulations [8], agents attempt to predict how many others will attend the Santa Fe bar each time using a simple kind of deterministic inductive reasoning. If they predict attendance will be less than cthe bar; if they predict attendance will be greater than cat home. Each agent uses a number of .rules of thumb. such as simple averages, moving averages, and linear or nonlinear filters to formulate predictions and then acts on the prediction that was correct most frequently in recent past. When Arthur simulated a bar-going society of 100 inductively rational agents, he found that the attendance of the bar tended to hover near 60 though population varied greatly, often exceeding 70 or dropping below 50. The time series of aggregate attendance appeared random, despite the deterministic rules of the underlying agents.

6. Conclusions

In practice, servers are configured to act selfishly in order to maximize their own benefit. For example, in web servicies, each administrative domain utilizes local servers to better support clients in their own domains. They have obvious incentives to cache objects that maximize the benefit in their domains, possibly at the expense of globally optimum behavior [5]. It has been an open question whether these caching scenarios and protocols maintain their desirable global properties (low total social cost, for example) in the face of selfish behavior.

The techniques proposed can bring improvements to a distributed database system. The redesign phase, meaning re-fragmentation and re-allocation, can be done by the mentioned technique such as to obtain an equilibrium in the system from each user.s perspective. The use of non-cooperative game theory in the agency logic can automate the system and easy its management.

References

[1] Oszu, M. T., Valduriez, P., *Principles of Distributed Database Systems*, Prentice Hall, Englewood Cliffs, NJ, 1999

[2] Piattini, M. and Diaz, O., *Advanced Database Technology and Design*, Artech House, Inc. 685 Canton Street Norwood, MA 02062, 2000

[3] Kasa, Z., Varga, V., *A Graph Model of Distributed Database System*, Studia Universitatis "Babes-Bolyai", Informatica, vol. XLI, no. 2, 1996, p. 3-8.

[4] M. R. Garey and D. S. Johnson., *Computers and Intractability:A Guide to the Theory of NP-Completeness,* W. H.Freeman and Company, San Francisco. 1979.

[5] McCain, Roger, *Game Theory: A Non-Technical Introduction to the Analysis of Strategy*, Barnes and Noble, September 2003

[6] K. Jain, M. Mahdian, and A. Saberi. *Approximating market equilibrium.* In Workshop on Approximation Algorithms for Combinatorial Optimization (APPROX 2003), 2003.

[7] Greenwald,A., Mishra,B. and Parikh, R. *The Santa Fe Bar problem revisited:Theoretical and practical implications*, In Mimeo. New York:New York Univ. Press, 1998.

[8] Arthur, W.B. *Inductive reasoning and bounded rationality: The Santa Fe bar problem*, In American Economic Revue, vol. 84, pp. 406-411, May 1994.

[9] Garg R., Kapoor S., *Auction Algorithms for Market Equilibrium*, STOC 2004, June 13.15, 2004, Chicago, Illinois, USA.

Horea Adrian Grebla
Department of Computer Science
”Babes-Bolyai” University of Cluj-Napoca
400084-Cluj-Napoca, 1 M. Kogalniceanu Street
email:*horea@cs.ubbcluj.ro*