MAS-based, Scalable Allocation of Resources in Large-scale, Dynamic Environments

(Doctoral Consortium)

Diana Gudu
Karlsruhe Institute of Technology
Karlsruhe, Germany
diana.gudu@kit.edu

ABSTRACT

I study the resource allocation problem, with a focus on computing environments (Cloud, Internet of Things), by employing market mechanisms such as auctions that consider supply and demand, and dynamic pricing in the decision process. However, current approaches face two limitations: (i) *centralisation*, as there is typically a central entity (such as an auctioneer) that holds all the information and decides the allocation based on it, therefore either limiting the size of the auctions that can be handled, or leading to compromises in the quality of the solution, and (ii) *inflexibility*, or inability to deal with heterogeneous, dynamic environments.

In this dissertation I propose a scalable and adaptive approach to overcome these limitations: a market-based, completely decentralised multi-agent system. I investigate techniques to enhance the performance, scalability and quality of the decentralised allocation using insights from centralised approaches.

Keywords

resource allocation; auction; multi-agent system; scalable; distributed; simulation; cloud

1. INTRODUCTION

Resource allocation is a fundamental problem in many fields. Recent trends have replaced traditional resource management techniques with "market-inspired" approaches to achieve a more flexible, economically efficient, and fair allocation. In computing, the Grid has pioneered market-oriented allocation of resources: Broberg et al. [1] surveyed the state-of-the-art and showed how utility computing helped deal with unpredictable or bursty workloads, and how it provided incentives for users to be more flexible in their usage pattern. The advent of cloud computing has taken the concepts of utility computing and market-based resource allocation to the next level due to virtualization, the use of real currency, and better defined services and legal frameworks, sparking new research in already established fields such as combinatorial auctions [3, 8, 6]. A new paradigm in resource allocation is emerging, where self-interested participants, who only try to increase their utility, meet in *ad-hoc* marketplaces, negotiate and finally agree on resource exchanges [1, 2], much as in a free market.

Nevertheless, work on resource allocation so far is either (a) *centralised*: typically a central entity has a global view of user requests and available services, and single-handedly handles the allocation process [7], or (b) *one-sided*: user agents (brokers) analyse a user's various requirements to match them with the providers that can offer the needed resources [4]; at the other end, there are agents acting only on behalf of providers, e.g. holding auctions in order to sell their resources to the highest bidders and maximise provider revenue [9]. At the most, proposed decentralised marketplaces include many independent, competing brokers that mediate the resource exchanges [2], leaving the users with the difficult task of choosing the appropriate broker and optimal market mechanism. Moreover, most resource allocation frameworks only offer specific market mechanisms, exhibiting *inflexibility*, whereas heterogeneous and dynamic environments like the Cloud and the Internet of Things would benefit from higher mechanism diversity.

Appears in: *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2016), J. Thangarajah, K. Tuyls, C. Jonker, S. Marsella (eds.), May 9–13, 2016, Singapore.*

2. PROPOSED APPROACH

Motivated by the need for scalability and adaptability, I propose a complex, evolving system where independent brokers are replaced by agents that communicate and collaborate to achieve global goals based on local, incomplete information. Such a multi-agent system (MAS) would act as a mediator between providers and users, striving to find the best allocation for all participants, i.e. to maximise the *social welfare*. I aim to off-load most tasks from the client or provider side (including match-making of requirements and resource offers, selection of appropriate market mechanisms) onto the MAS. For heterogeneous, upredictable environments, a plethora of market mechanisms should be made available in the system, each specialised on a certain niche. Prior knowledge of which mechanisms are best suited for certain input patterns, together with learning from past experiences, will be employed to select appropriate mechanisms and improve the overall allocation. The MAS must therefore adapt to the dynamic environment. The inherent decentralisation of the proposed system will contribute to the scalability of the resource allocation.

2.1 Market mechanisms

Most market mechanisms, such as combinatorial auctions, are complex optimisation problems, typically NP-hard [3].

Various heuristic methods have been developed to solve these problems, inevitably causing some heuristics to give good results for certain problem instances (or use cases) and underperform in others. In a related work, Leyton-Brown et al. [5] propose an approach to identify features that characterise the empirical hardness of combinatorial auctions, by using domain knowledge and machine learning.

I state that a *comprehensive analysis* of centralised market mechanisms aimed at gathering in-depth knowledge for each mechanism is essential in making the multi-agent system more scalable and efficient.

In a paper currently under review, I extensively analysed one such centralised market mechanism, namely a greedy heuristic approach for solving double combinatorial auctions, based on the work of Samimi et al. [6]. I evaluated how metrics such as time, success rate of client requests, resource utilization, average agent utility, and average price per unit of goods sold behave for different inputs, as well as with respect to demand and supply fluctuations. An important finding was that the results for sparse providers (providers specialised on offering disjoint subsets of the possible resource types) were unpredictable and highly sensitive to input, limiting the algorithm to be used only on dense providers and dense client requests. I then analysed the scalability of the same algorithm by distributing it over multiple agents, and found promising speedup results, as well as an expected efficiency loss due to decision making with incomplete information. There is certainly more research to be done in this area, more specifically a thorough analysis and classification of the most suited mechanisms for each use case brought forth, and designing new algorithms where necessary.

2.2 Multi-agent system

I propose a MAS design consisting of three types of agents: *Client Agents*, which deal with customer requests and submit them to an appropriate Matching Agent after a basic analysis; *Resource Agents*, which expose the available providers and their specifications to the appropriate Matching Agents; and finally *Matching Agents*, which use market mechanisms (e.g. auctions, bargaining) to find the best resource allocation for the client requests and provider offers they receive. Each Matching Agent essentially solves an optimisation problem by maximising the sum of all participants' utilities, i.e. the social welfare. This underlines the distributed character of the proposed approach. A participant's utility can include, but is not limited to, the price of resources and associated QoS metrics. However, local optimisation is not sufficient for a globally optimal allocation. Therein lies the intelligent, adaptive behaviour: in the communication between Matching Agents, aimed at finding a trade-off between the amount of information exchanged and the improvement in solution quality (performance vs. efficiency). The communication is based on prior knowledge of market mechanisms (cf. Sec. 2.1) for filtering and forwarding requests based on use cases, as well as learning techniques for a better distribution of requests.

The design expresses the generality of our approach by separating the clients and resource providers on the one hand, and the resource allocation on the other hand. The MAS also provides fair treatment to clients and providers (through social welfare optimisation), as well as incentive for them to participate in a truthful manner (through a careful auction design, more specifically the pricing algorithm).

An ensuing challenge concerns the evaluation of the decentralised system, as well as its comparison with an equivalent centralised one.

2.3 Research question

Essentially, through the proposed thesis, I resolve to address the following research question:

How can one improve the scalability and adaptivity of resource allocation in large-scale, dynamic environments by using a decentralised, market-driven approach?

To that end, the MAS-based approach I proposed encompasses the following: (1) *decentralised decision-making* for increased scalability, (2) *market-driven* allocation and pricing, that is *flexible* and *customised* to the needs of the participants, (3) *fine-grained allocation* of resources, aimed at improving resource utilization for providers, thereby decreasing costs for clients, and (4) *social welfare optimisation*, as opposed to one-sided optimisation which only benefits either the resource providers or the users.

REFERENCES

[1] J. Broberg, S. Venugopal, and R. Buyya. Market-oriented grids and utility computing: The state-of-the-art and future directions. *Journal of Grid Computing*, 6(3):255–276, 2008. doi:10.1007/s10723-007-9095-3.

[2] R. Buyya, C. S. Yeo, and S. Venugopal. Market-oriented cloud computing: Vision, hype, and reality for delivering it services as computing utilities. In *High Performance Computing and Communications, 2008. HPCC'08. 10th IEEE International Conference on*, pages 5–13. IEEE, 2008. doi:10.1109/HPCC.2008.172.

[3] S. De Vries and R. V. Vohra. Combinatorial auctions: A survey. *INFORMS Journal on computing*, 15(3):284–309, 2003. doi:10.1287/ijoc.15.3.284.16077.

[4] F. Jrad. *A service broker for Intercloud computing*. PhD thesis, Karlsruher Institut für Technologie, 2014. urn:nbn:de:swb:90-423243.

[5] K. Leyton-Brown, E. Nudelman, and Y. Shoham. Learning the empirical hardness of optimization problems: The case of combinatorial auctions. In *Principles and Practice of Constraint Programming-CP 2002*, pages 556–572. Springer, 2002. doi:10.1007/3-540-46135-3_37.

[6] P. Samimi, Y. Teimouri, and M. Mukhtar. A combinatorial double auction resource allocation model in cloud computing. *Information Sciences*, 2014. doi:016/j.ins.2014.02.008.

[7] J. Watzl. *A framework for exchange-based trading of cloud computing commodities*. PhD thesis, Ludwig Maximilian University of Munich (LMU), 2014. urn:nbn:de:bvb:19-168702.

[8] S. Zaman and D. Grosu. Combinatorial auction-based allocation of virtual machine instances in clouds. *Journal of Parallel and Distributed Computing*, 73(4):495–508, 2013. doi:10.1016/j.jpdc.2012.12.006.

[9] Q. Zhang, Q. Zhu, and R. Boutaba. Dynamic resource allocation for spot markets in cloud computing environments. In *Utility and Cloud Computing (UCC), 2011 Fourth IEEE International Conference on*, pages 178–185. IEEE, 2011. doi:10.1109/UCC.2011.33.