Cloud Computing – Trends and Challenges

Helen Karatza, *Senior Member, IEEE*

Abstract—Advances in computing systems and networks have led to new platforms for computing, such as grids and clouds. The increasing popularity of cloud computing has offered computational services to scientists, consumers and enterprises as utilities, on a pay-per-use approach. Many business organizations migrate all or part of their data processing activities in the cloud. In addition to enterprise and scientific computing, clouds are also being used for the management of smart facilities.

In large scale distributed systems such as grids and clouds there are important issues that must be addressed, such as: performance, resource allocation, efficient scheduling, energy conservation, reliability, protection of sensitive data, security and trust, cost, availability, quality. Effective management of cloud resources is crucial to use effectively the power of these systems and achieve high system performance and lower cost.

The cloud computing paradigm can offer various types of services, such as computational resources for complex applications, web services, social networking, urban mobility, health care, environmental science, etc. Furthermore, the simultaneous usage of services from different Clouds can have additional benefits such as lower cost and high availability - however issues like interoperability, portability and service brokerage need to be addressed.

Recently, Big Data has become one of the most important research fields in science, engineering, enterprise, biology, healthcare, etc. However, in order that cloud computing will be a platform for supporting big data applications, appropriate algorithms are required for acquiring knowledge from a variety of big and not centrally collected data.

In this talk we will present recent research covering a variety of concepts on cloud computing, based on existing or simulated cloud systems, that provide insight into problems solving and we will provide future trends and challenges in the cloud computing area.

Index Terms—Cloud Computing; performance; simulation; resource allocation; scheduling, energy conservation.

BIOGRAPHY

Helen Karatza is a Professor in the Department of Informatics at the Aristotle University of Thessaloniki, Greece. Dr. Karatza's research interests include Computer Systems Modeling and Simulation, Performance Evaluation, Grid and Cloud Computing, Energy Efficiency in Large Scale Distributed Systems, Resource Allocation and Scheduling and Real-time Distributed Systems.

Professor Karatza is the Editor-in-Chief of the Elsevier Journal "Simulation Modeling Practice and Theory", Area Editor of the "Journal of Systems and Software" of Elsevier, and she has been Guest Editor of Special Issues in multiple international journals.

Helen Karatza is with the Department of Informatics, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece (e-mail: karatza@csd.auth.gr).

REFERENCES

[1] S. Dimitriadou and H.D. Karatza, "Job Scheduling in a Distributed System Using Backfilling with Inaccurate Runtime Computations," Proc. of the 4th International Conference on Complex, Intelligent and Software Intensive Systems (CISIS-2010), IEEE Computer Society, Krakow, Poland, Febr. 15-18, pp. 329-336, 2010.

[2] K. Gkoutioudi and H. D. Karatza, "Task cluster scheduling in a grid system," *Simulation Modelling Practice and Theory*, vol. 18, no. 9, pp.1242-1252, Oct. 2010.

[3] H. D. Karatza, "A simulation model of task cluster scheduling in distributed systems," Proc. of the 7th IEEE Workshop on Future Trends of Distributed Computing Systems (FTDCS'99), Cape Town, South Africa, pp. 163-168, Dec. 1999.

[4] H. D. Karatza, "Epoch task cluster scheduling in a distributed system," Proc. of the 2002 International Symposium on Performance Evaluation of Computer and Telecommunication Systems (SPECTS'02), San Diego, CA, pp. 259-265, Jul. 2002.

[5] H. D. Karatza, "Periodic task cluster scheduling in distributed systems," in *Computer System Performance Modeling in Perspective*, London, UK, World Scientific, Imperial College Press, 2006, pp. 257-276.

[6] K. I. Karaoglanoglou and H. D. Karatza, "Resource discovery in a dynamical grid based on re-routing tables," *Simulation Modelling Practice and Theory*, vol. 16, no. 6, pp. 704-720, Jul. 2008.

[7] K. Karaoglanoglou and H. Karatza, "Resource discovery in a grid system based on matchmaking-routers," Proc. of the Panhellenic Conference on Informatics (PCI 2008), Samos, Greece, IEEE Computer Society, pp. 73-77, 28-30 Aug. 2008.

[8] I. A. Moschakis and H. D. Karatza, "Evaluation of gang scheduling performance and cost in a cloud computing system," *Journal of Supercomputing*, Springer, vol. 59, no. 2, pp. 975-992, 2012.

[9] I. A. Moschakis and H. D. Karatza, "Multi-criteria scheduling of Bag-of-Tasks applications on heterogeneous interlinked Clouds with simulated annealing," *Journal of Systems and Software*, vol. 101, pp. 1-14, Mar. 2015.

[10] I. A. Moschakis and H. D. Karatza, "A meta-heuristic optimization approach to the scheduling of Bag-of-Tasks applications on heterogeneous clouds with multi-level arrivals and critical jobs," *Simulation Modelling Practice and Theory*, vol. 57, pp. 1-25, Sep. 2015.

[11] I. A. Moschakis and H. D. Karatza, "Towards scheduling for Internet-of-Things applications on Clouds: A Simulated Annealing approach," *Concurrency and Computation: Practice and Experience*, Wiley, vol. 27, no. 8, pp. 1886 – 1899, 2015.

[12] Z. Papazachos and H. D. Karatza, "Scheduling bags of tasks and gangs in a distributed system," Proc. of the 2015 International Conference on Computer, Information and Telecommunication Systems (CITS'15), Gijón, Spain, pp. 1-5, Jul. 2015.

[13] G. Stavrinides and H.D. Karatza, "Performance evaluation of gang scheduling in distributed real-time systems with possible software faults," Proc. of the 2008 International Symposium on Performance Evaluation of Computer and Telecommunication Systems (SPECTS 2008), Edinburgh, UK, pp. 1-7, June 16-18, 2008.

[14] G.L. Stavrinides and H.D. Karatza, "Fault-tolerant gang scheduling in distributed real-time systems utilizing imprecise computations," *Simulation: Transactions of the Society for Modeling and Simulation International*, Sage Publications, Vol. 85, No. 8, pp.525-536, 2009.

[15] G. L. Stavrinides and H. D. Karatza, "Scheduling multiple task graphs with end-to-end deadlines in distributed real-time systems utilizing imprecise computations," *Journal of Systems and Software*, vol. 83, no. 6, pp. 1004-1014, Jun. 2010.

[16] G. L. Stavrinides and H. D. Karatza, "Scheduling real-time DAGs in heterogeneous clusters by combining imprecise computations and bin

packing techniques for the exploitation of schedule holes," *Future Generation Computer Systems*, vol. 28, no. 7, pp. 977-988, Jul. 2012.

[17] G. L. Stavrinides and H. D. Karatza, "The impact of resource heterogeneity on the timeliness of hard real-time complex jobs," Proc. of the 7th International Conference on PErvasive Technologies Related to Assistive Environments (PETRA'14), Workshop on Distributed Sensor Systems for Assistive Environments (Di-Sensa), Rhodes, Greece, pp. 65:1-65:8, May 2014.

[18] G. L. Stavrinides and H. D. Karatza, "A cost-effective and QoS-aware approach to scheduling real-time workflow applications in PaaS and SaaS clouds," Proc. of the 3rd International Conference on Future Internet of Things and Cloud (FiCloud'15), Rome, Italy, pp. 231-239, Aug. 2015.

[19] G. L. Stavrinides and H. D. Karatza, "Scheduling different types of applications in a SaaS cloud," Proc. of the 6th International Symposium on Business Modeling and Software Design (BMSD'16), Rhodes, Greece, pp. 144-151, Jun. 2016.

[20] G. L. Stavrinides and H. D. Karatza, "Scheduling real-time parallel applications in SaaS clouds in the presence of transient software failures," Proc. of the 2016 International Symposium on Performance Evaluation of Computer and Telecommunication Systems (SPECTS'16), Montreal, Canada, pp. 1-8, Jul. 2016.

[21] G. L. Stavrinides and H. D. Karatza, "Simulation-based performance evaluation of an energy-aware heuristic for the scheduling of HPC applications in large-scale distributed systems," Proc. of the 8th ACM/SPEC International Conference on Performance Engineering (ICPE'17), 3rd International Workshop on Energy-aware Simulation (ENERGY-SIM'17), L'Aquila, Italy, pp.49-54, Apr. 2017.

[22] G. L. Stavrinides, F. R. Duro, H. D. Karatza, J. G. Blas, and J. Carretero, "Different aspects of workflow scheduling in large-scale distributed systems," *Simulation Modelling Practice and Theory*, vol. 70, pp. 120-134, Jan. 2017.

[23] G. Terzopoulos and H. D. Karatza, "Bag-of-Task scheduling on power-aware clusters using a DVFS-based mechanism," Proc. of the 10th Workshop on High-Performance, Power-Aware Computing (HPPAC'14), 28th IEEE International Parallel & Distributed Processing Symposium (IPDPS'14), Phoenix, AZ, pp. 833-84, May 2014.

[24] G. Terzopoulos and H. D. Karatza, "Bag-of-Tasks load balancing on power-aware clusters," Proc. of the 24th Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP'16), Heraklion, Crete, pp. 136-142, Feb. 2016.

[25] G. Terzopoulos and H. D. Karatza, "Power-aware Bag-of-Tasks scheduling on heterogeneous platforms," *Cluster Computing*, vol. 19, no. 2, pp. 615-631, Jun. 2016.

[26] S. Zikos and H. Karatza, "A clairvoyant site allocation policy based on service demands of jobs in a computational grid," *Simulation Modelling Practice and Theory*, Elsevier, Vol. 19, Issue 6, pp. 1465-1478, Jun. 2011.

[27] S. Zikos and H. Karatza, "Performance and energy aware cluster-level scheduling of compute-intensive jobs with unknown service times," *Simulation Modelling Practice and Theory*, Elsevier, vol. 19, pp.239–250, 2011.