Combination project RUG and CIT
Technical analysis of data center:
Power consumption reduction and automated temperature control in data centers through smart scheduling decisions

Tobias Van Damme
University of Groningen
email: t.van.damme@rug.nl

Major challenges in controlling data centers lie in providing adequate cooling and preventing thermal hot spots from occurring, and optimizing the balance between the number of servers and cooling capacity. In an attempt to reduce power consumption I study and analyze thermal-aware control strategies in my PhD research.

At this time there is a large gap between theoretical and practical implementation of workload schedulers. In order to close the gap and study more realistic implementations of workload schedulers an in-depth investigation of key practical features of data centers is required. Preliminary theoretical work shows schedulers which achieve large energy reduction due to smart workload scheduling, varying from 7% up to 30% cooling cost reduction [1–6]. These values are best case scenarios and very situation dependent but indicate that it is worthwhile to investigate smart scheduling decisions. Furthermore since data center infrastructure is a rapid changing field, it is important to update the theoretical models to the latest technological advancements.

To close the gap and update the models to the latest standards an in-depth study should be done. This study should entail the identification of the current infrastructure and of the parameters which can be/are measured in a data center as well as the way a data center processes different types of workload. Questions which could be addressed in this study are the following:

- How are different types of workload characterized and how does a data center handle and differentiate between these types, e.g. OnLine Transaction Processing (OLTP) vs. batch processing.
- Which parameters can be measured, e.g. Temperature of rack/server, CPU load, fan speed, response times.
- Which parameters can be controlled, e.g. workload of servers, cooling temperature of Cooling Air Conditioning Unit (CRAC), airflow rate at the server level.
- What is the current infrastructure and power consumption of the equipment.

- What constraints are currently considered in data center control, e.g. constant temperature of servers, maximum allowable inflow temperature, response times (SLA's).
- What Virtualization workload balancing options are availiable.

This investigation can be done as a Master thesis by a student of Industrial Engineering and Management, Computer Science, Applied Physics and Applied Mathematics in cooperation with the CIT. The daily supervision of the student is handled by me and Prof Claudio De Persis and Prof Marco Aiello and the CIT provides the technical background. The student would first start a literature research and get acquainted with the field. Then after identifying key questions, contacts the appropriate experts at the CIT for short discussions concerning the questions. These discussions would take up at most 1 hour once a week, in this way the load on the CIT personnel remains low.

After the technical investigation the student will focus on extending current models with the obtained knowledge.

References

[1] T. Van Damme, C. De Persis and P. Tesi, "Optimized thermal-aware job scheduling and control of data centers", *arXiv:1611.00522*, 2016.

[2] J. Doyle, F. Knorn, D. O'Mahony and R. Shorten, "Distributed thermal aware load balancing for cooling of modular data centres", *IET Control Theory and Applications*, vol. 7, pp. 612–622, Nov 2013.

[3] J. Moore, J. Chase, R. Parthasarathy and S. Ratnesh, "Making scheduling 'cool' temperature-aware workload placement in data centers", in *USENIX Annual Technical Conference*, pp. 61–74, 2005.

[4] L. Parolini, B. Sinopoli, B. H. Krogh and Z. Wang, "A cyber-physical systems approach to data center modeling and control for energy efficiency", *Proceedings of the IEEE*, vol. 100, pp. 254–268, jan 2012.

[5] S. Li, H. Le, N. Pham, J. Heo and T. Abdelzaher, "Joint optimization of computing and cooling energy: Analytic model and a machine room case study", in *32nd Int. Conf. on Distributed Computing Systems*, pp. 396–405, IEEE, Jun 2012.

[6] Q. Tang, S. K. S. Gupta and G. Varsamopoulos, "Energy-efficient thermal-aware task scheduling for homogeneous high-performance computing data centers: a cyber-physical approach", *IEEE trans. on Parallel and distributed systems*, vol. 19, pp. 1458–1472, nov 2008.