

# Dynamic Thermal Control and Management towards Reducing Utility Cost in Data Centers

This poster presents a collaborative work between Intel Corporation and ASU to investigate data center thermal management. This work envisions a dynamic feedback control based solution for strategically manipulating the computing resources in a data center to reduce its utility cost while maintaining the required thermal performance.

Computing clusters and server farms are increasingly housed in datacenters that are limited by power and thermal capacity. Minimizing the energy cost and improving thermal performance of such datacenters are among the key issues towards optimizing computing resources and maximally exploiting the datacenter's computation capability. Improperly designed or operated datacenters may either suffer from overheated servers (potential system failure) or from overcooled servers (leading to extra utility cost). In either case there is loss of revenue for datacenter operators.

The energy cost and thermal distribution of a datacenter depends on the multiscale properties of a large number of multimodal system variables, such as the cooling capacity, hardware thermal characteristics, and workload profiles. In this work, we propose a unique merger between the physical infrastructure and resource management functions of the cluster operating system to take a holistic view of datacenter *thermal control and management*, and make global (at the level of a datacenter), power-aware, and thermal-aware job scheduling decisions.

In the past, design and management of datacenters, and heat management of an individual computer have been treated as separate problems. In general, there is no automated *feedback control* loop between the physical environment of a datacenter and the datacenter's management software/system software scheduler for making thermal and power-efficient resource

management decisions. An integrated and system-oriented approach towards this problem, however, could provide several significant benefits towards the goal of minimizing operational cost in the datacenters. We develop an integrated thermal-aware scheduling and cooling-system control framework with the goal of optimizing the total utility cost of running a datacenter.

The main contribution of our work is the development of a new thermal control framework which resembles traditional queuing modeling framework used for performance analysis of computer systems. The thermal modeling framework will enable fast analysis of various job scheduling and cooling policies by taking into account the physical parameters of computer systems and cooling systems, their distribution in 3-D space, as well as computation parameters such as workload profiles. It also provides a theoretical foundation for analyzing energy cost of a datacenter operation.

We verify our model through extensive experimentation by prototyping the system and deploying it in several datacenters. The prototype system is based on state-of-the-art onboard and wireless/wired ambient sensors and datacenter management system.

We have developed the datacenter thermal control framework including the abstract thermal and cost models based on ASU High Performance Datacenter measurements. Currently, we are working on formalization of the different cost optimization problems and comparing their solutions. Our remaining work includes implementing thermal-aware scheduling, integrating sensor measurement with the control framework. And verification of the effectiveness and agility of the feedback control framework through real datacenter experiments.