Under-provisioning the Cooling Infrastructure of Modern Datacenters.

Ioannis Manousakis
Rutgers

4/25/2016 at 01:00 pm
CoRE 305

Abstract

Cloud providers have made significant strides in reducing the cooling capital and operational costs of their datacenters, for example, by leveraging outside air (free) cooling where possible. Despite these advances, cooling costs still represent a significant expense mainly because cloud providers typically provision their cooling infrastructure for the worstcase scenario (i.e., very high load and outside temperature at the same time). Thus, in this paper, we propose to reduce cooling costs by underprovisioning the cooling infrastructure. When the cooling is underprovisioned, there might be (rare) periods when the cooling infrastructure cannot cool down the IT equipment enough. During these periods, we can either (1) reduce the processing capacity and potentially degrade the quality of service, or (2) let the IT equipment temperature increase in exchange for a controlled degradation in reliability. To determine the ideal amount of underprovisioning, we introduce CoolProvision, an optimization and simulation framework for selecting the cheapest provisioning within performance constraints defined by the provider. CoolProvision leverages an abstract trace of the expected workload, as well as cooling, performance, power, reliability, and cost models to explore the space of potential provisionings. Using data from a real small free-cooled datacenter, our results suggest that CoolProvision can reduce the cost of cooling by up to 55%. We extrapolate our experience and results to larger cloud datacenters as well.

Examination Committee: Thu D. Nguyen (chair), Abhishek Bhattacharjee, Santosh Nagarakatte, David Cash, Ricardo Bianchini (External, Microsoft Research)