CoolProvision: Underprovisioning Datacenter Cooling

RUTGERS
THE STATE UNIVERSITY OF NEW JERSEY

Ioannis Manousakis, Thu D. Nguyen
{ioannis.manousakis, tdnguyen}@cs.rutgers.edu

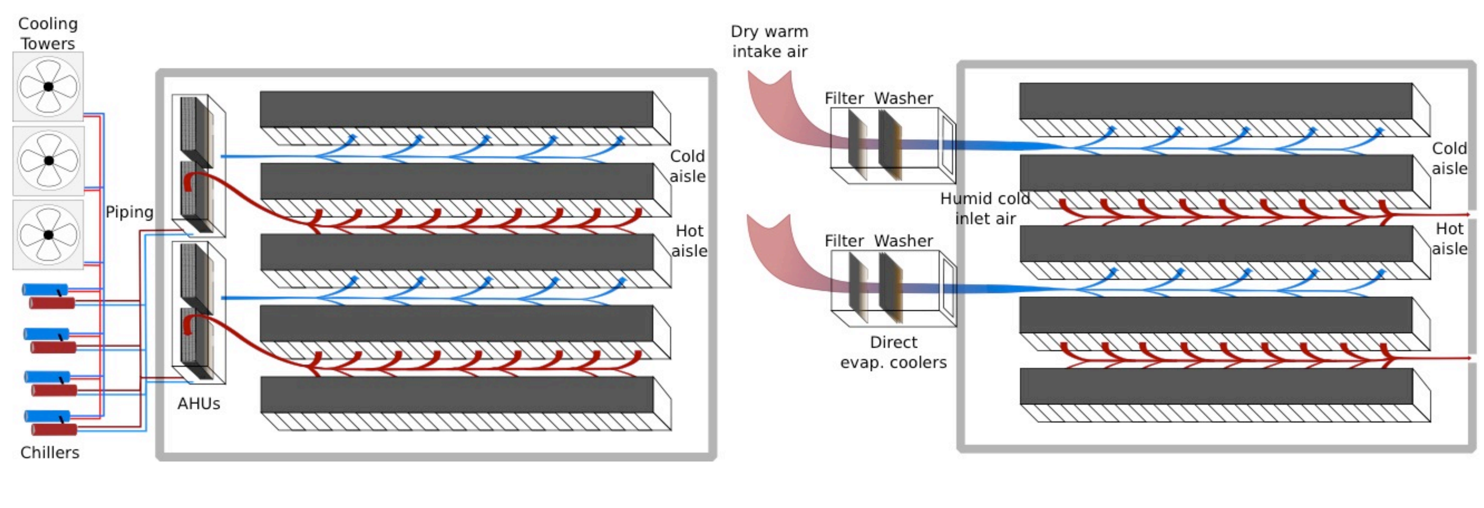
Sriram Sankar
sriram@godaddy.com

Inigo Goiri and Ricardo Bianchini
{goiri, ricardob}@microsoft.com

Overview

Datacenter Cooling

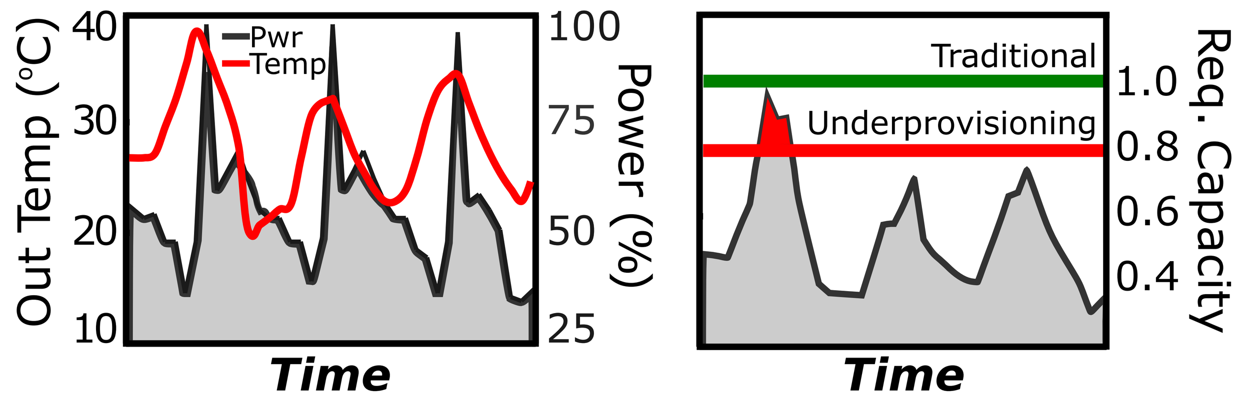
- Traditional cooling accounts for 40% of construction costs
- High energy costs (location-dependent)



a) Traditional Cooling (Chillers) b) Direct Evaporative Cooling

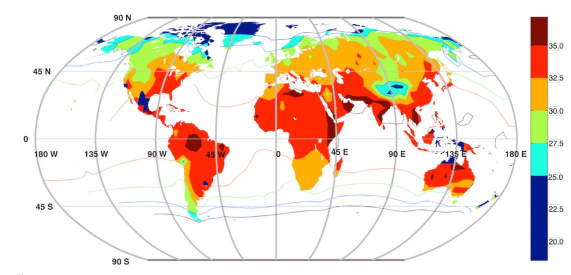
Design goals:

- Systematic way to reduce datacenter cooling within perf. constraints
- Trades capital costs, energy costs, hardware reliability and performance



Benefits:

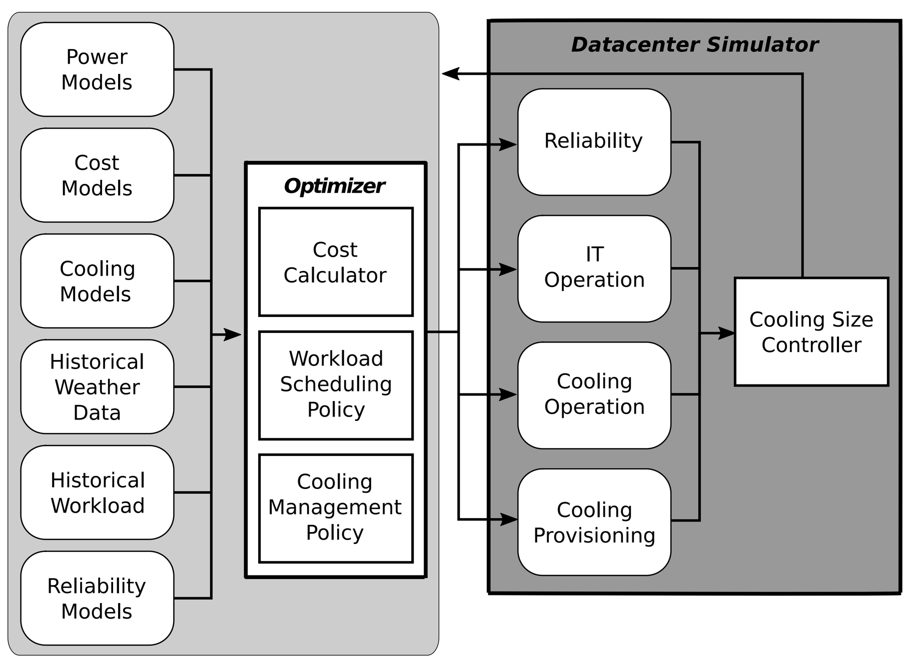
- Reduced capital costs and drastically improved NVPs
- Workload- and location-aware cooling deployments



Design

Architecture

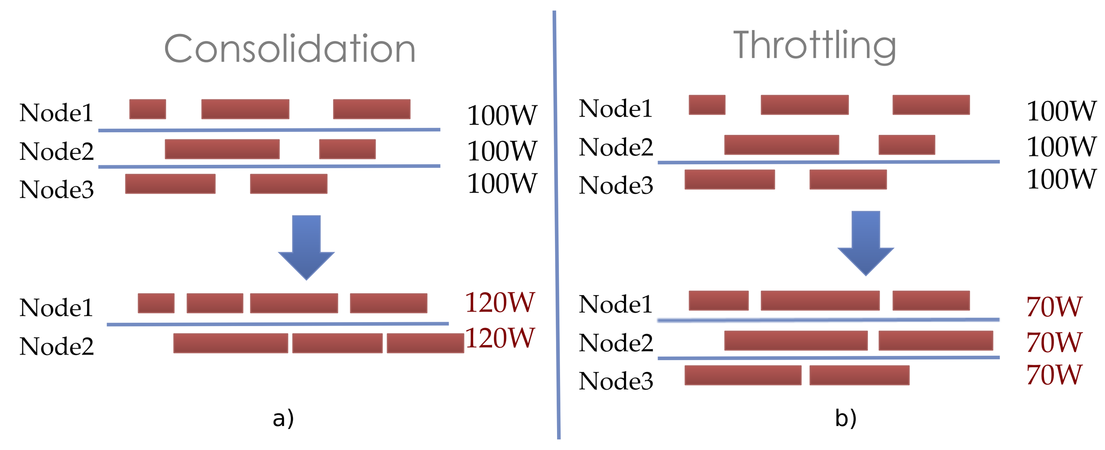
- Modular design

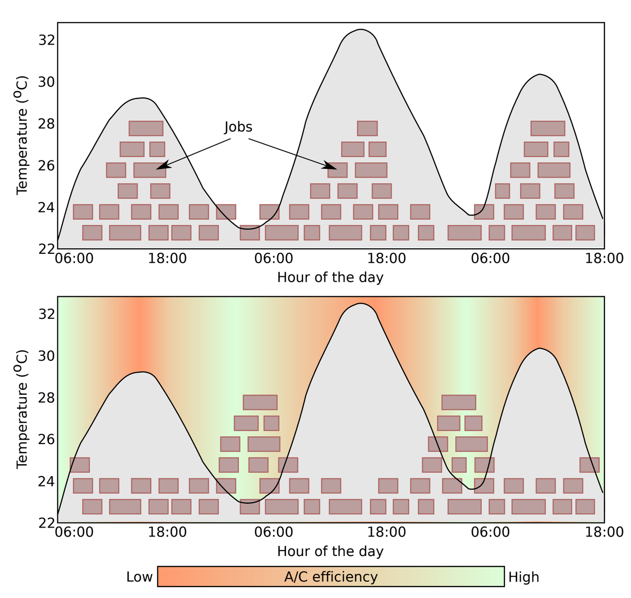


- Inputs
  1. High-level thermal models
  2. Historical weather (temp, RH%)
  3. Expected workload profile
  4. Server power models
  5. Power – performance trade-offs
- Outputs
  1. Cheapest cooling technology
  2. Cheapest cooling size
- Constraints
  - Upper inlet temperature and RH%
  - Maximum workload degradation

Workload Management Policies

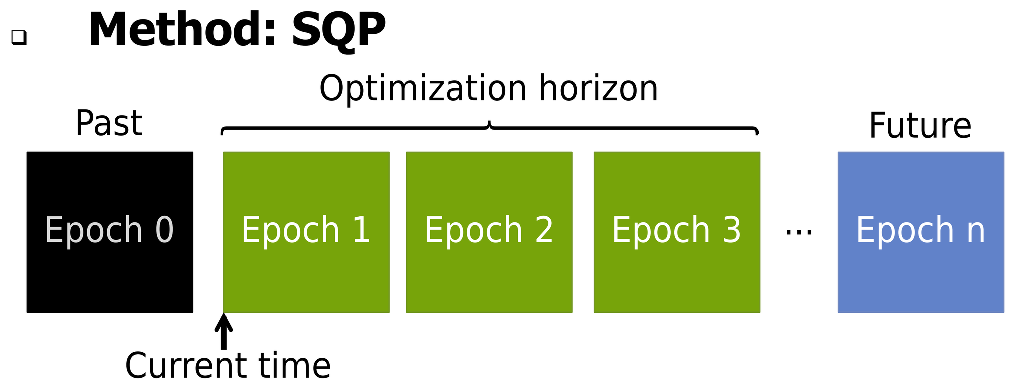
- Reduce server power and heat during harsh conditions
  1. High outside temperature
  2. High relative humidity
  3. High load
- Generic workloads → DVFS
- Interactive and VMs → Consolidation
- Deferrable analytics → Job deferring





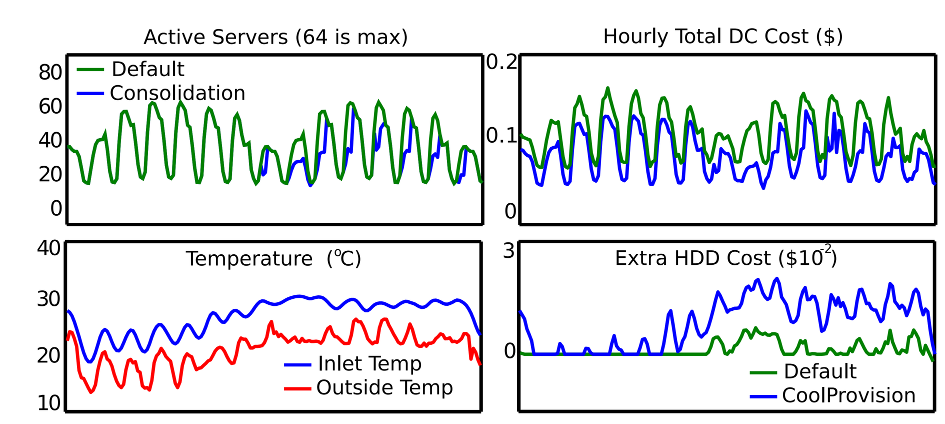
Optimization

- Epoch – based optimization
- Tunable horizon
- Method: SQP



- Implementation in MATLAB (Simulator)
- Trained models with real small factor A/C and evaporative cooler (Parasol)

Operational Example



Results

Four Representative Locations