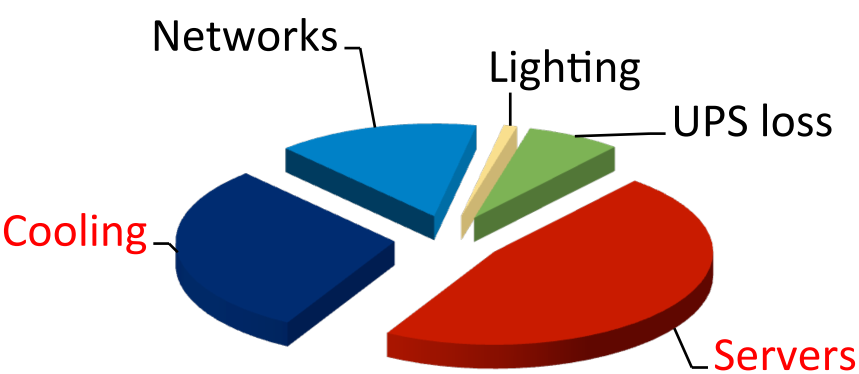
Coordinating Liquid and Free Air Cooling with Workload Allocation for Data Center Power Minimization

Li Li, Wenli Zheng, Xiaodong Wang, and Xiaorui Wang

Dept. of Electrical and Computer Engineering
The Ohio State University

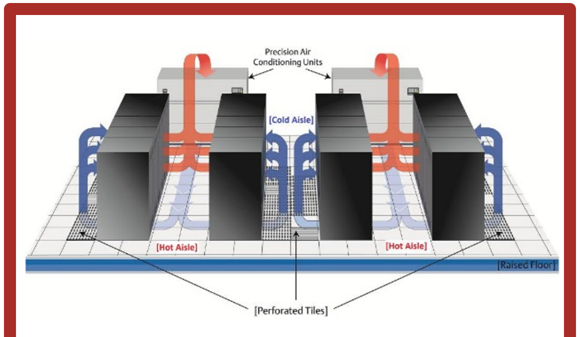
Introduction

- Data centers consume a huge amount of energy.
  - Servers and cooling are the two biggest energy consumers.
- New cooling techniques are increasingly adopted.
  - Such as *liquid cooling* and *free cooling*.
- Concern of large IT companies
  - How to minimize both server and cooling power of geo-distributed data centers equipped with heterogeneous cooling techniques.

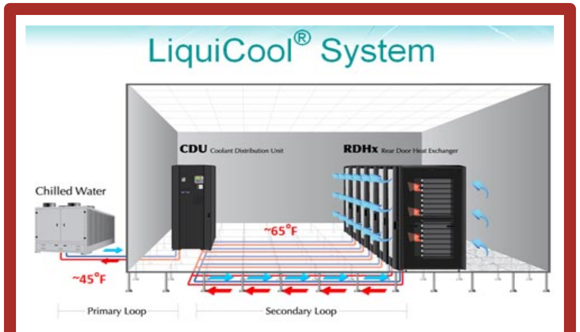


THE OHIO STATE UNIVERSITY

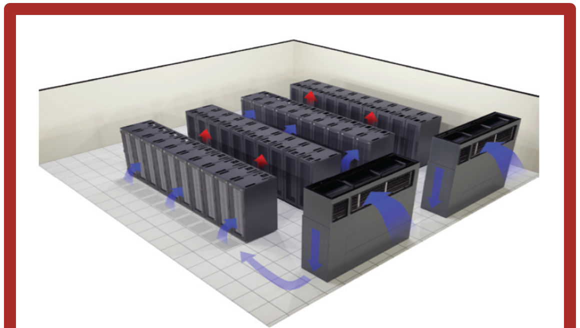
Background: Cooling System



- Air Cooling:
  - Uses CRAC (Computer Room Air Conditioning) system to supply cold air
  - Pros: Low maintenance cost
  - Cons: Low cooling efficiency & High heat recirculation



- Liquid Cooling:
  - Uses coolant to directly absorb heat from servers.
  - Pros: High cooling efficiency
  - Cons: High maintenance cost & implementation cost

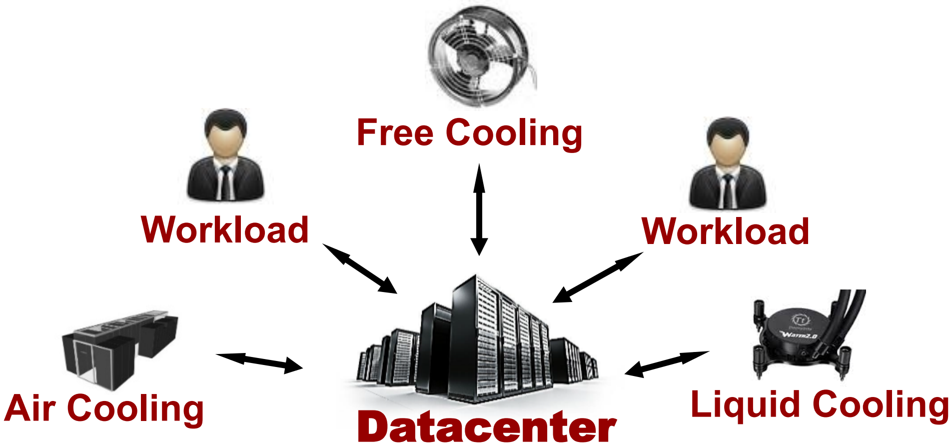


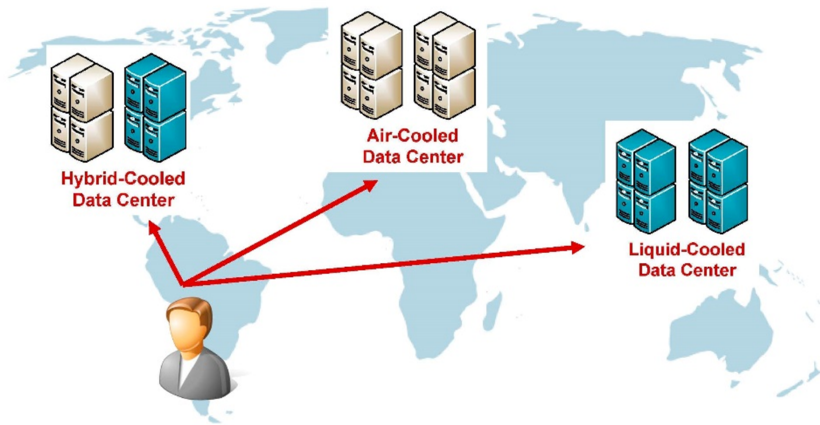
- Free Cooling:
  - Directly uses outside cold air.
  - Pros: Highest efficiency & low maintenance cost
  - Cons: Highly depends on outside temperature

Closely Related Work

- Power minimization in a single data center.
  - Server power minimization. *[ASPLOS'09] [USENIX'08] [SIGMETRICS'09]*
  - Cooling power minimization.
    - Free Cooling system. *[CLOUDCOM'11] [SEMI-THERM'12]*
    - Air Cooling system. *[CASE'10][ITHERM'06]*
    - Liquid Cooling system. *[SEMI-THERM'12][ITHERM'12]*
  - Joint minimization of server power & air cooling power. *[ASPLOS'10] [USENIX'05]*
  - Our paper minimizes total power of a data center with air, liquid and free cooling systems.
- Geographically distributed data centers.
  - Temperature aware workload management. *[SIGMETRICS'12] [ICAC'13]*
  - Our paper dispatches workload to data centers with different cooling configurations for total power minimization.

Challenges in Joint Power Minimization





- For a single hybrid-cooled data center:
  - How to efficiently coordinate air, liquid, and free cooling systems?
  - How to distribute workloads for joint power minimization?
- For geographically distributed data centers:
  - How to make online management decisions in real time?
  - Computing complexity is too high if all factors (workload distribution, cooling mode selection) are considered simultaneously.

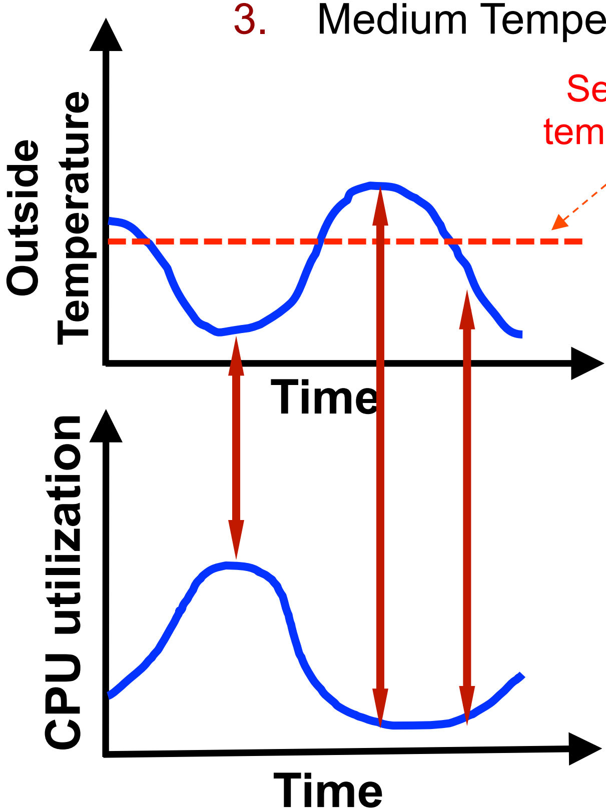
What is This Paper About?

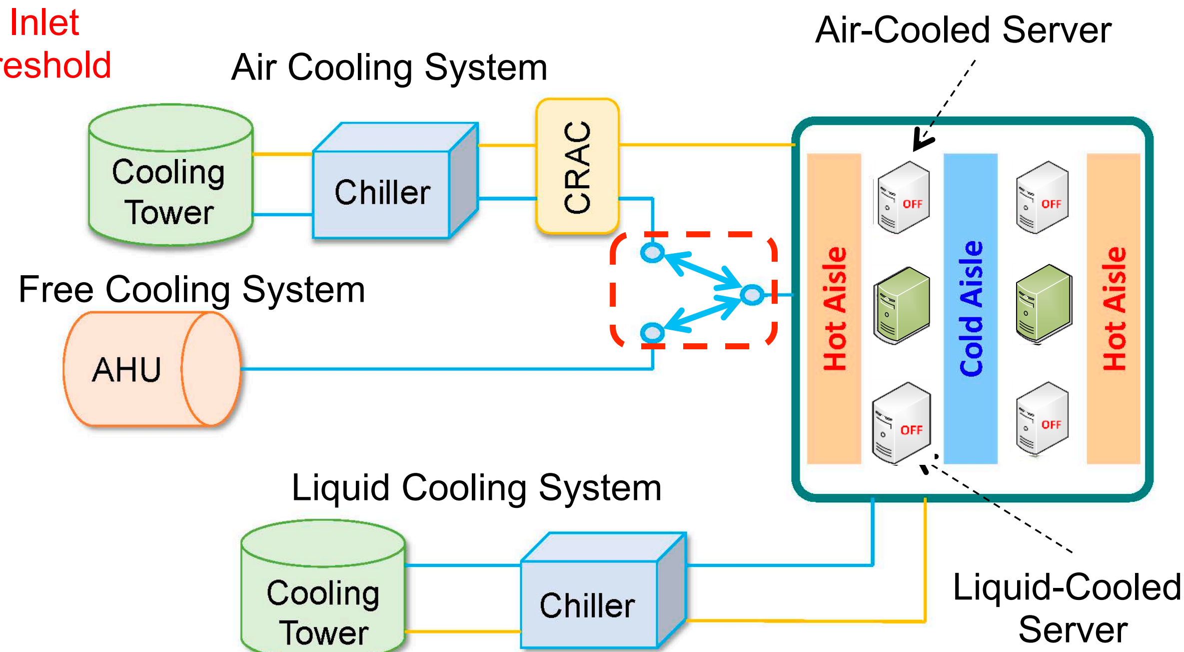
- *SmartCool:* jointly minimizes server and cooling power.
  - For reduced complexity, we decompose the problem to two levels.

- Level 1: Single hybrid-cooled data center:
  - Intelligent coordination of cooling systems to jointly optimize server and cooling power.
  - We formulate the cooling management in a hybrid-cooled data center as a constrained optimization problem.
- Level 2: Geographically distributed data centers:
  - A light-weight solution to dynamically dispatch requests among a network of data centers with heterogeneous cooling systems.

SmartCool: Single Data Center Level

- 3 example scenarios of SmartCool for cooling system coordination.
  1. Low Temperature + High Workload
  2. High Temperature + Low Workload
  3. Medium Temperature + Medium Workload





Optimization Formulation for Single Data Center

- We formulate the power minimization problem of the hybrid-cooled data center.
  - N servers are deployed in the data center.
  - M servers are liquid-cooled.
  - Total workload is W_{to}

Minimize total power consumption

W_i : Workload handled by server i.

➢ *Problem Formulation*

$$\min\{P^{server} + P^{air} + P^{liquid}\}$$

Subject to:

a. All workload is distributed
b. CPUs temperature < Threshold
c. Server Inlet temperature < Threshold

➢ *Models*

$$P_i^{server} = W_i \times P_i^{compute} + P_i^{idle}$$

$$P^{air} = \beta P_{CRAC}^{air} + (1-\beta) P_{free}^{air}$$

$$P^{liquid} = \frac{\sum_{i=1}^{M} \alpha P_i^{server}}{COP^{liquid}}$$

β: Binary variable to determine cooling mode.

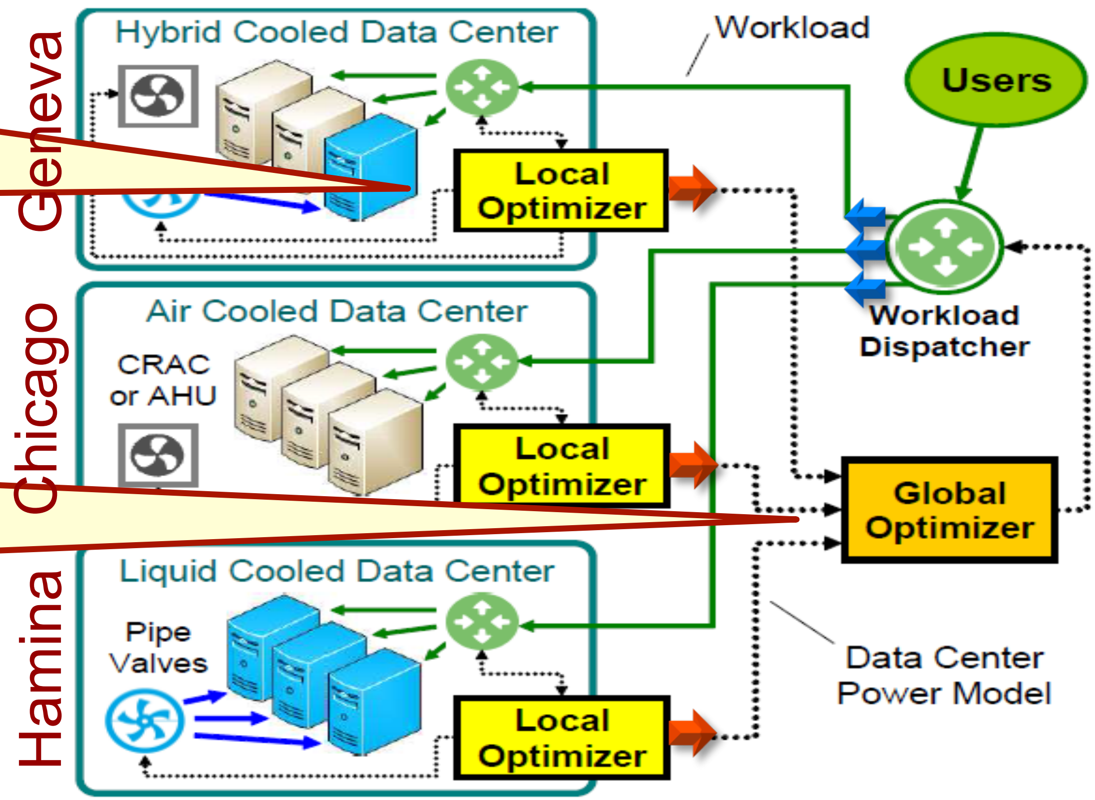
SmartCool: Distributed Data Centers Level

- A light-weight dynamic request dispatching algorithm.
  - Among different data center sites with heterogeneous cooling systems.
  - For real-time online management decisions.

Derive the cPUE model:

$$cPUE_{optimal} = f(T_{outside}, W)$$

using the single data center optimization process.



Optimize the total power by manipulating the workload assigned to each data center.

Formulation for Distributed Data Centers

- We formulate a global optimization problem for minimizing the total power consumption of geo-distributed data centers.
  - Each data center has N servers.
  - The data center system has K data centers to handle workload.

➢ *Problem Formulation*

$$\min \sum_{j=1}^{K} P_j^{DC}$$

Subject to:

a. All workload is distributed
b. CPUs temperature < Threshold
c. Server Inlet temperature < Threshold

➢ *Models*

Workload handled by certain data center.

$$P^{DC} = P^{compute}(W) * cPUE_{optimal}(T_{outside}, W)$$

$$cPUE = \frac{P^{server} + P^{cooling}}{P^{compute}(W)}$$

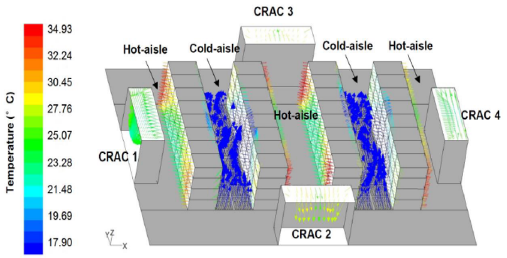
$$cPUE_{optimal} = f(T_{outside}, W)$$

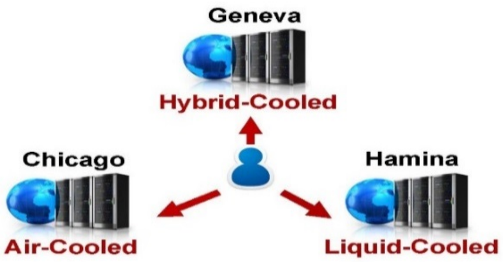
Data center outside temperature.

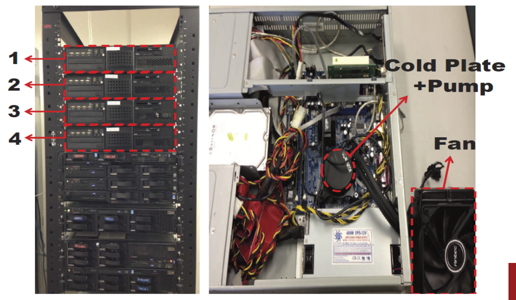
THE OHIO STATE UNIVERSITY

Experiment Setup

- Simulations for one data center
  - Standard data center configuration (alternating hot and cold aisle).
  - Four rows of servers, the first is liquid-cooled.
  - CFD is used to simulate heat recirculation.
- Simulations for multiple data centers.
  - Three geo-distributed data centers are evaluated.
  - One-week temperature traces of Geneva, Hamina and Chicago.
- Hardware experiments.
  - One liquid-cooled server and three air-cooled servers.
  - Heater is used to change the ambient temperature.







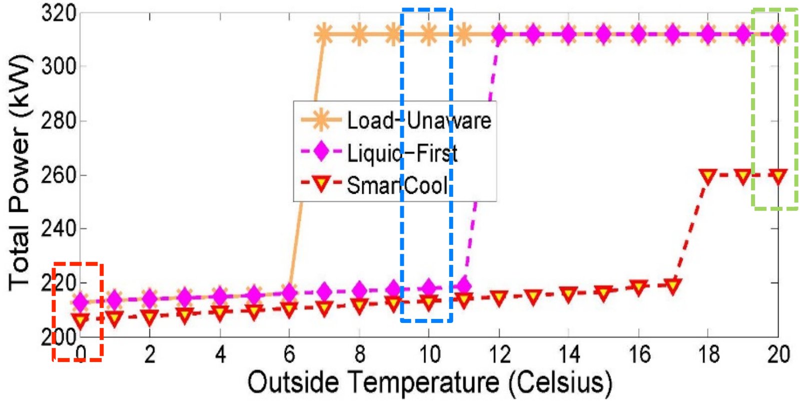
THE OHIO STATE UNIVERSITY

Baselines for Comparison

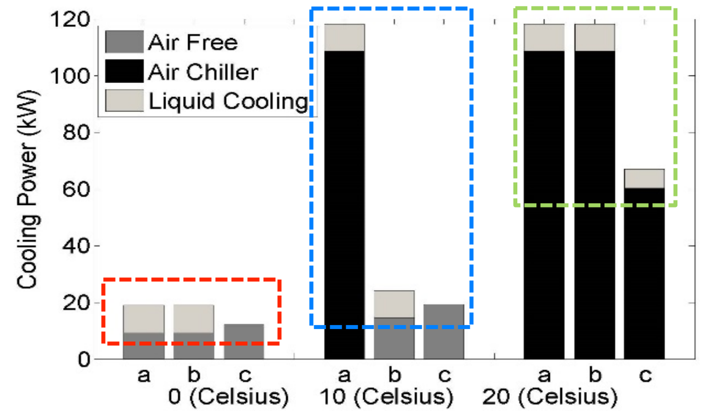
- Baselines for a single data center
  - *Load-Unaware*: (state of the practice)
    - Use a fixed temperature threshold to determine cooling mode.
    - Prefers to distribute workload to liquid-cooled servers.
  - *Liquid-First:*
    - First distribute workload in the same way as *Load-Unaware.*
    - Dynamically adjusts the temperature threshold based on real time workload.
- Baselines for geo-distributed data centers
  - *Low-Temp-First:*
    - First distribute workload to datacenter with lowest outside temperature.
  - *Liquid-First:*
    - Dispatch workload to data centers according to their cooling efficiency.

Results for A Single Data Center

- SmartCool vs. Load-Unaware and Liquid-First



Total Power at 50% Workload



a. Load-Unaware b. Liquid-First c. Smart-Cool

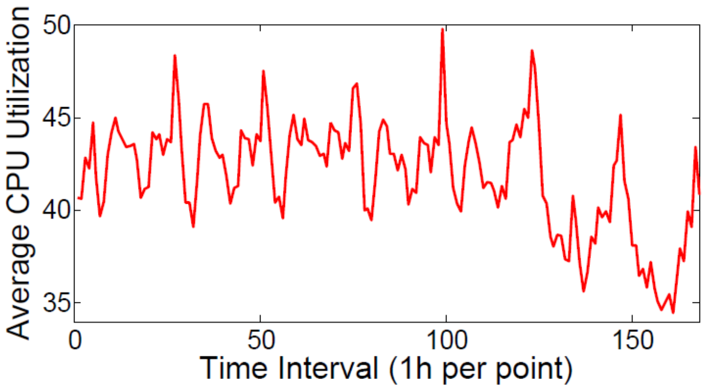
Power breakdown at 50% Workload

SmartCool:

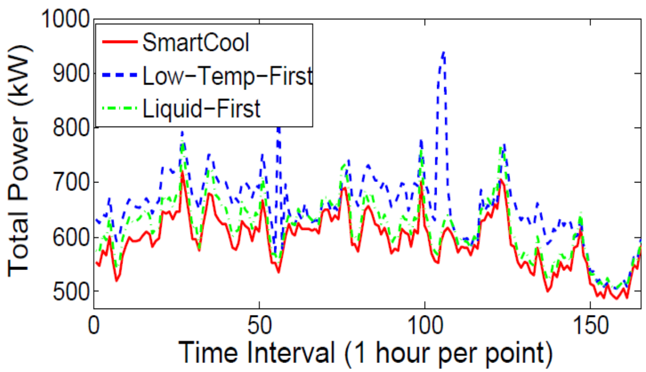
- Improves utilization of free cooling.
- Distribute workload to air-cooled servers at low temperature.
- Consumes less cooling power when traditional air cooling is used because of the consideration of heat recirculation.

Results for Geo-Distributed Data Centers

- We evaluate different power management schemes on three geo-distributed data centers with real workload and temperature traces.
- Traces:
  - One week temperature traces from Geneva, Hamina and Chicago [online].
  - Average CPU utilization from IBM data center [PACT'09].



IBM trace file



Total Power Consumption

- SmartCool jointly considers the impact of outside temperature and cooling configuration.
- More results (including hardware testbed results) are in the paper.

Conclusion

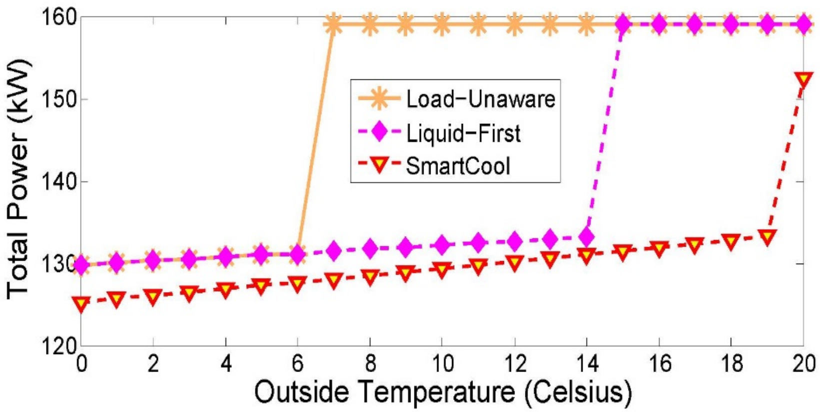
- Joint minimization of server and cooling power
  - For large IT companies with geo-distributed data centers.
  - Each data center uses heterogeneous cooling techniques.
- SmartCool features a two-level solution
  - Level 1: a single data center:
    - Effectively coordinates different cooling techniques.
    - Dynamically allocates workload for jointly optimized total power.
  - Level 2: geo-distributed data centers:
    - Light-weight dynamic request dispatching for real-time online decisions.

Thank you !

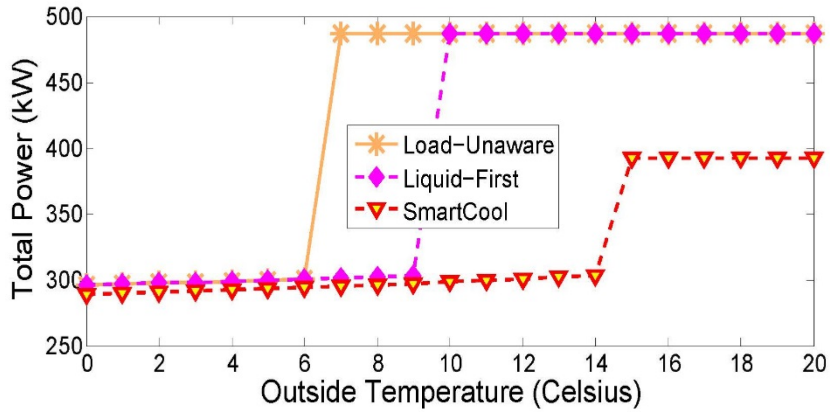
Questions?

Backup: Comparison of Cooling Schemes

- We compare *SmartCool* with *Load-Unaware and Liquid-First*.



(a) Total Power at 30% Workload

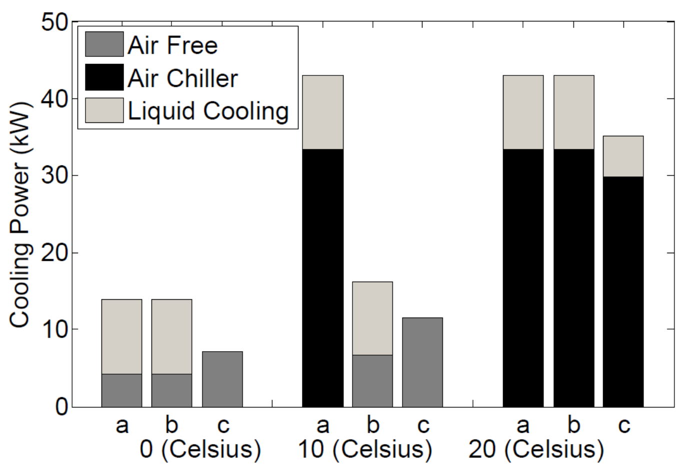


(b) Power break at 70% Workload

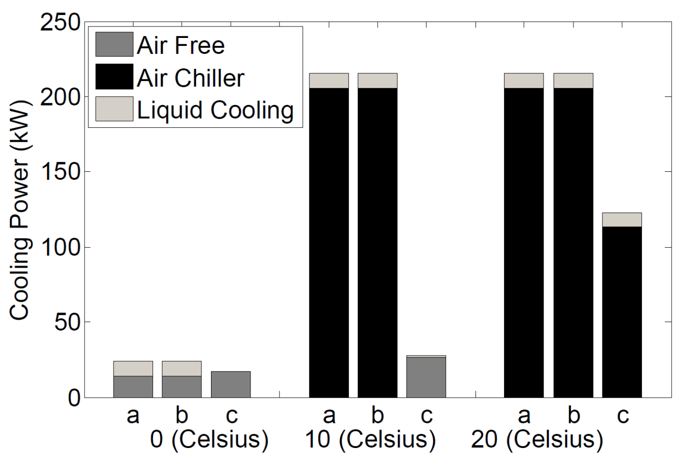
- Results have the same trend with that at 50% workload.

Cooling Power Break

- We break the cooling power consumption of three schemes:
  a. Load-Unaware b. Liquid-First c. Smart-Cool
- We evaluate:
  - *Air Free (free cooling power consumption)*
  - *Air Chiller (traditional air cooling power consumption)*
  - *Liquid Cooling (liquid cooling power consumption)*



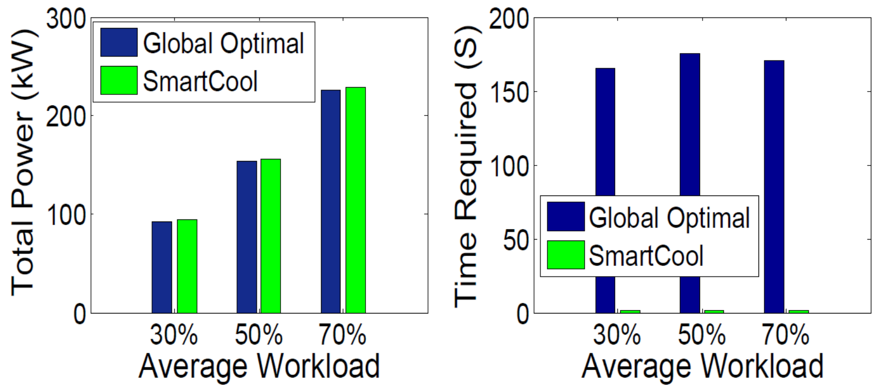
(a) Power break at 30% workload



(b) Power break at 70% workload

SmartCool vs. Global Optimal Solution

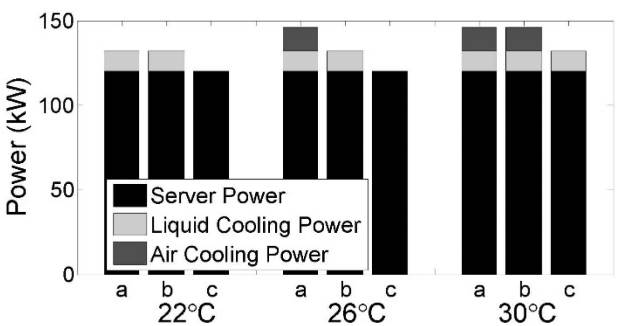
- We evaluate our two-layer power optimization algorithm.
- The two-layer solution is compared with global optimal scheme including:
  - Optimized total power consumption
  - Time overhead.
- Three smaller scale data centers are adopted.
  - Two rows of racks.
  - Each row has 4 racks and each rack has 4 blocks.



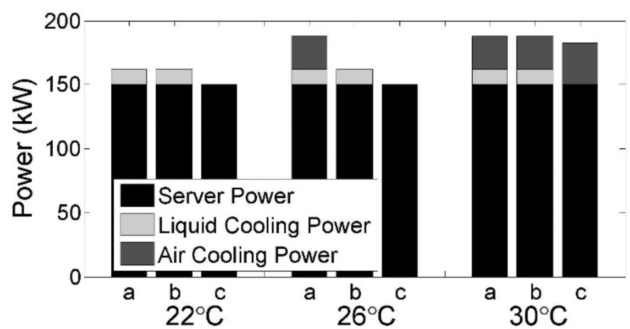
(a) Power Comparison (b) Time Comparison

Backup: Hardware Experiment

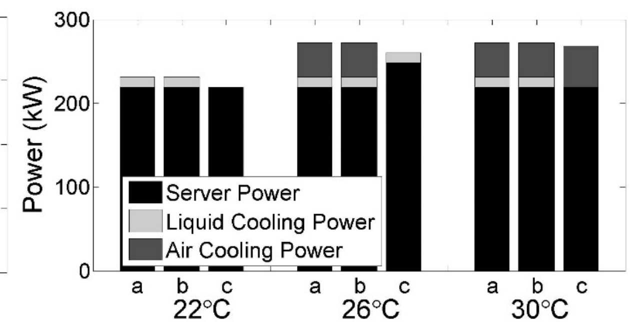
- We conduct experiments on our hardware testbed.
  - 1 liquid-cooled server and 3 air-cooled servers.
  - A heater is used to set the ambient temperatures.
  - Power meter is used to measure the power consumed by servers and cold plates.



(a) Total Power at 30% Loading



(b) Total Power at 50% Loading



(c) Total Power at 50% Loading