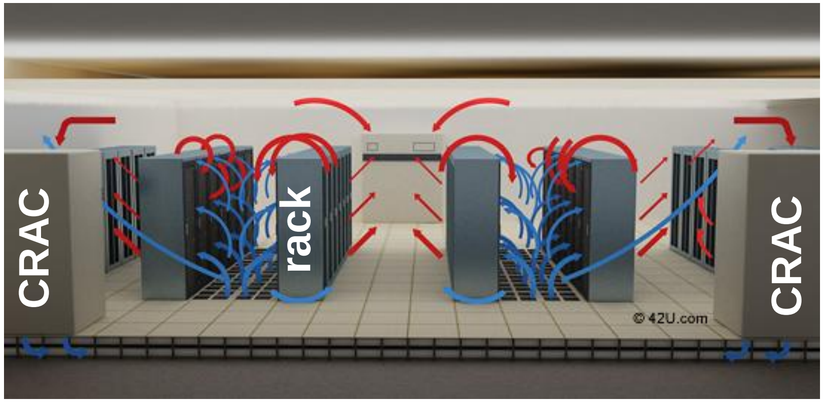
PTEC: A System for Predictive Thermal and Energy Control in Data Centers

Jinzhu Chen[1] **Rui Tan**[3] Guoliang Xing[1] Xiaorui Wang[2]

[1] Michigan State University, USA
[2] Ohio State University, USA
[3] Advanced Digital Sciences Center, Illinois at Singapore

Conservative Cooling Settings

EMC's new data center in Durham, NC



Raised-floor cooling [www.42u.com]

- **Data centers eat massive energy**
  - An industry data center = a mid-size town
- **60% non-computing energy ratio** [Uptime 2012]
  - **50% for cooling**
    24°C *in 90% data centers vs. recommended* ***27°C***
  - **10% for circulation**
    High fan speeds and simple control

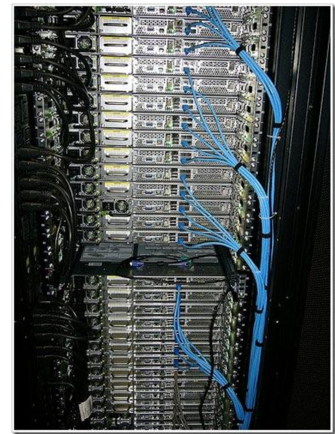
Related Work

- **New data centers: 10% for non-computing** [Google]
  - Clean slate redesign
  - Retrofitting technologies for existing data centers
- **Thermal & energy control: prevent overheat & reduce non-computing energy**
  - Single-variable (e.g., server workload)
  - Multi-variable
    Ours: AC + server fan (major ***correlated*** *energy eaters)*
- **React to detected hotspots**
  - Low temperature setpoint to resolve
  - Less energy-efficient

Predictive Thermal & Energy Control

cool air

Server inlets

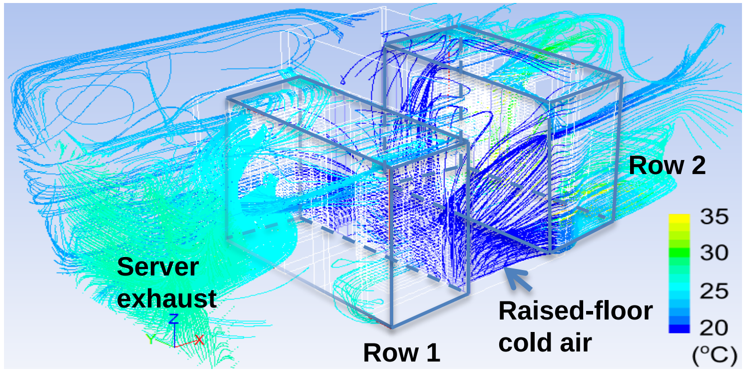


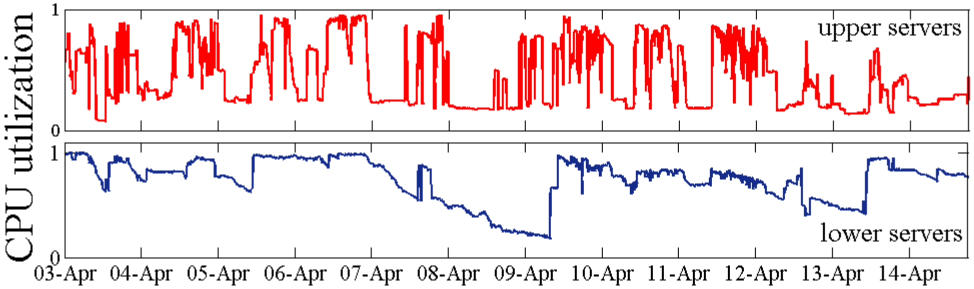
Server outlets

hot air

- **Energy-efficiently prevent hot spots**
- **Predict energy consumption & thermal conditions for each possible AC & fan control action**
  - Minimize predicted AC and fan energy
  - No predicted hot spots at server inlets

Challenges



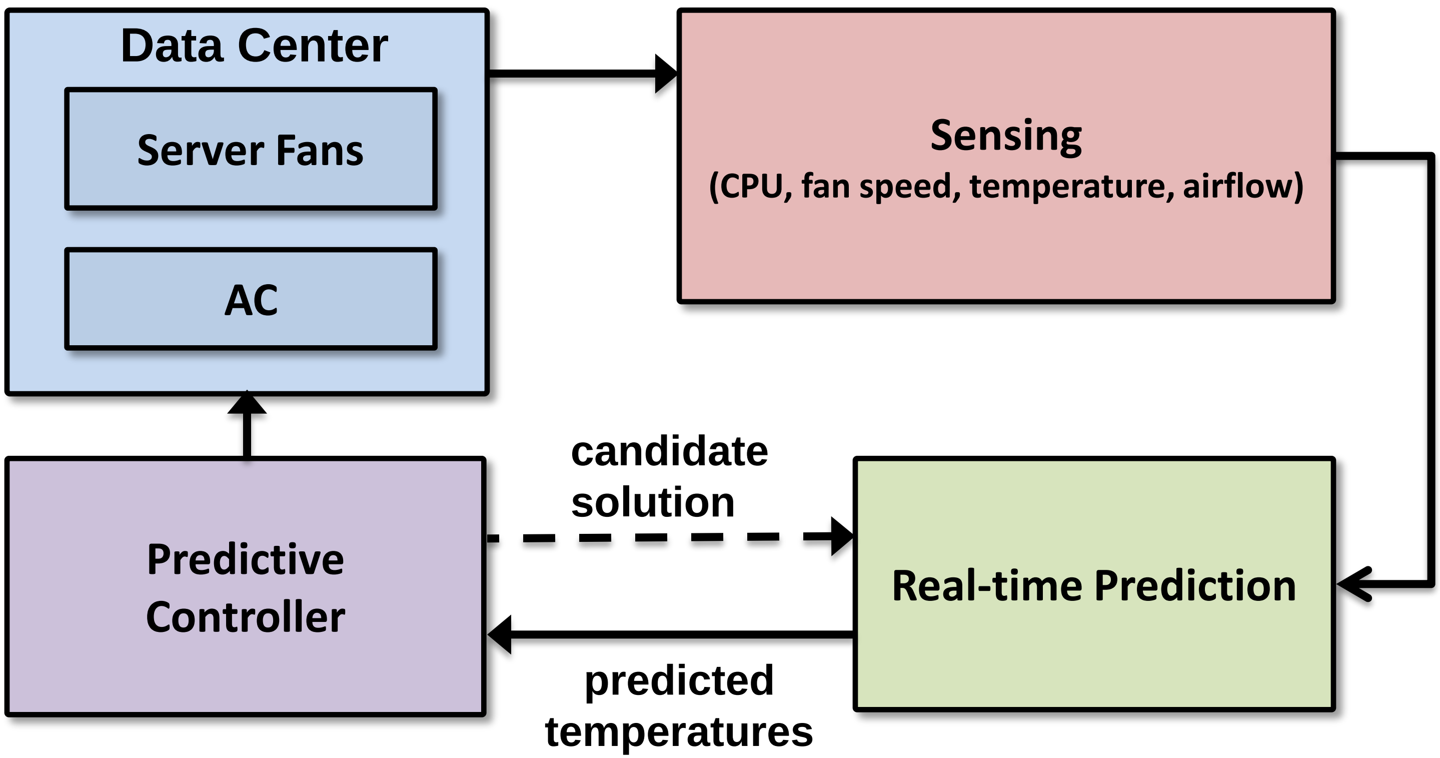


Dynamic workload in MSU HPCC over 12 days

- **Complex cyber-physical dynamics**
  – Air flow, server workload
  – Coupling btw control and thermal condition
- **Real-time and scalable**
  – No polynomial-time algorithms
  – Large # of controllable variables

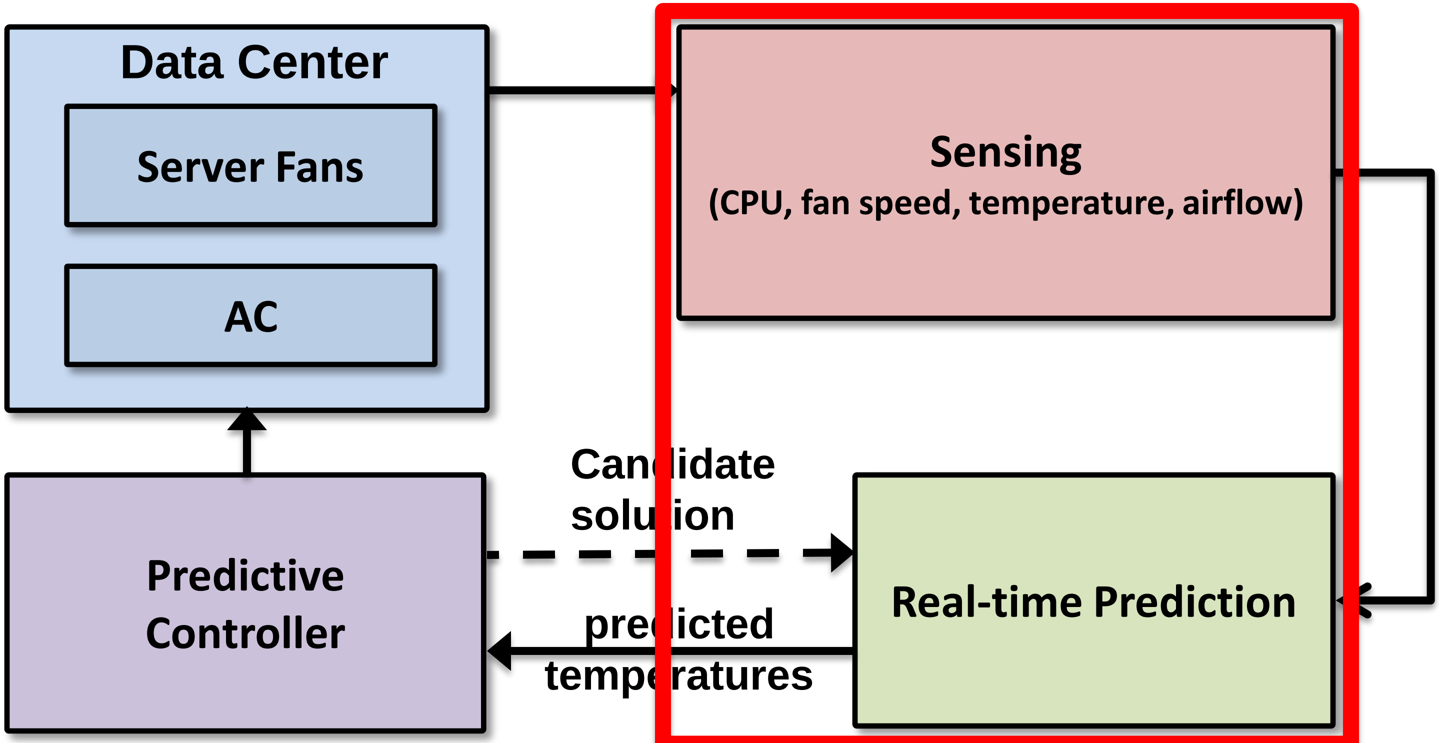
Approach Overview

- **Environment sensing**
  - Built-in sensors, external sensor network
- **Temperature & energy prediction**
  - Sensor data + energy models + candidate control action
- **Predictive controller**
  - Constrained optimization



Outline

- Motivation & Approach Overview
- **Sensing and Prediction**
- Predictive Thermal & Energy Control
- Evaluation

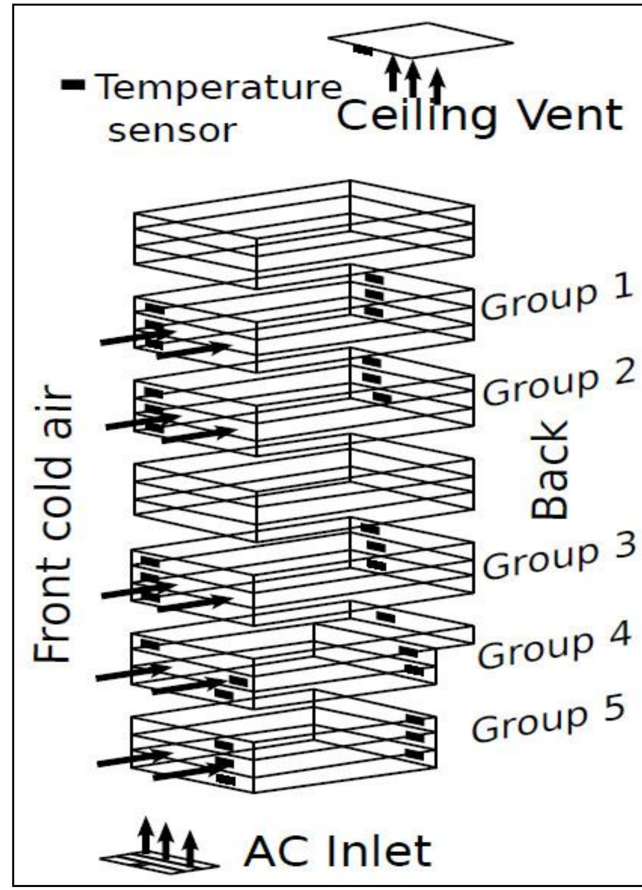


Sensing

Sensing

- Air velocity
  Air temperature
- CPU utilization
- Fan speed
- Inlet / outlet temperature

Wireless sensor / built-in sensor



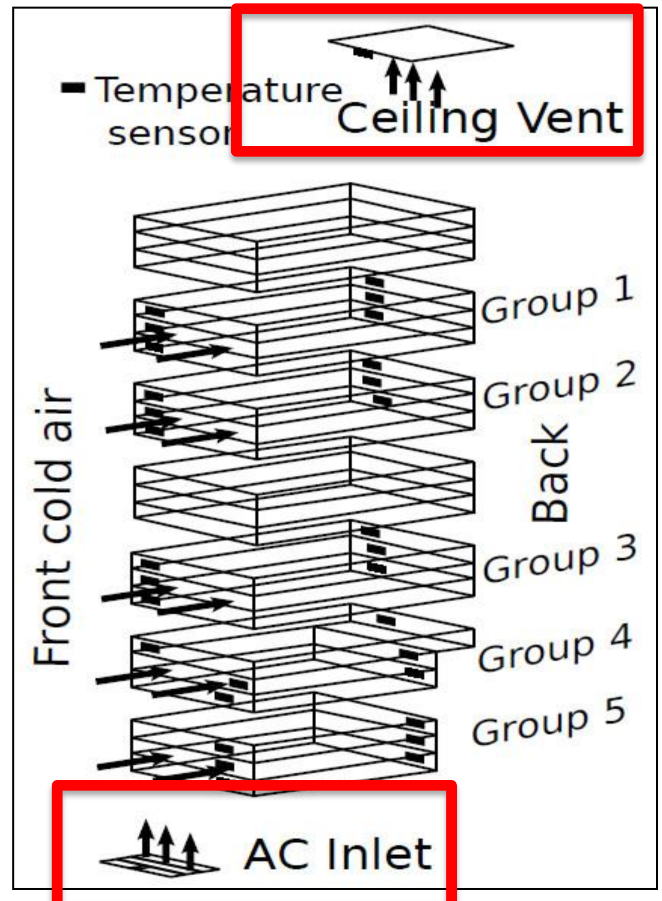
Sensor sampling rate: 30 seconds

Sensing

Sensing

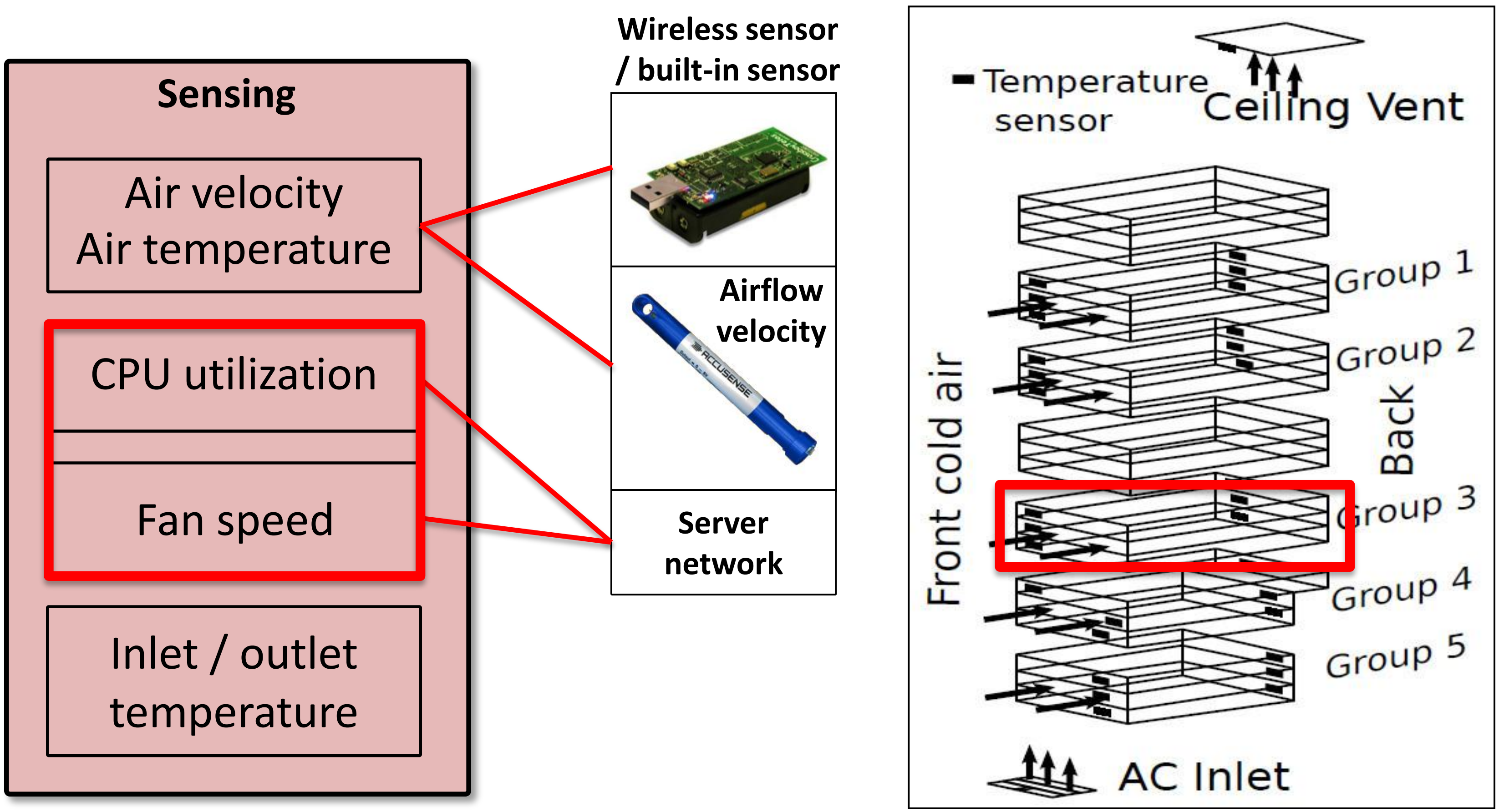
- Air velocity
  Air temperature
- CPU utilization
- Fan speed
- Inlet / outlet temperature

Wireless sensor / built-in sensor



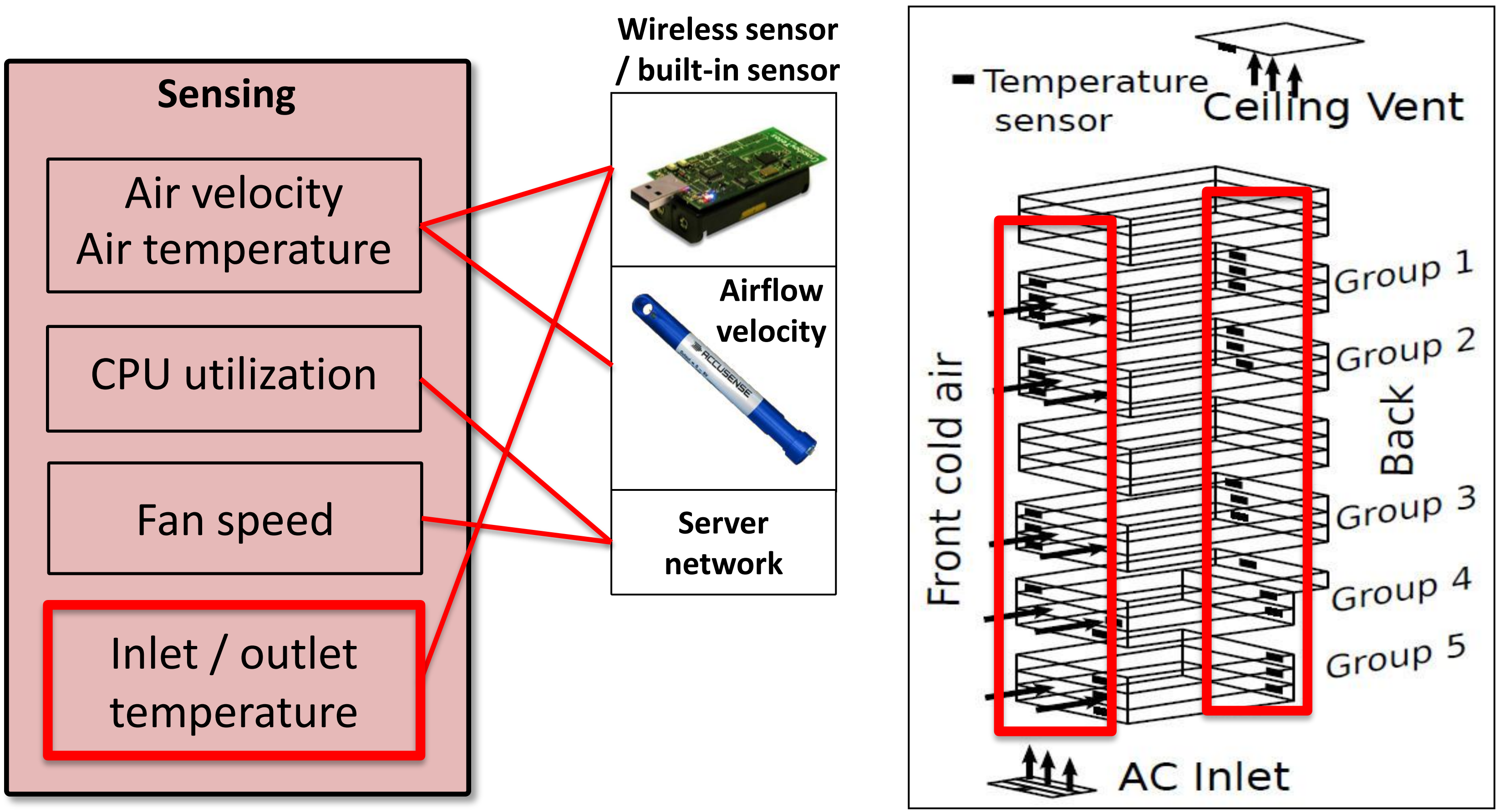
Sensor sampling rate: 30 seconds

Sensing



Sensor sampling rate: 30 seconds

Sensing



Sensor sampling rate: 30 seconds

Real-time Temperature Prediction [1]

- **Vector $\mathbf{p}_t$:** Sensor measurements at moment t
- **Prediction with a horizon of *k* sampling periods**

$$\mathbf{t}_{t+k} = \mathbf{A}_k \cdot [\mathbf{p}_t \quad \mathbf{p}_{t-1} \quad \cdots \quad \mathbf{p}_{t-h+1}]$$

[1] A High-Fidelity Temperature Distribution Forecasting System for Data Centers. RTSS 2012.

Real-time Temperature Prediction [1]

- **Vector $\mathbf{p}_t$:** Sensor measurements at moment *t*
- **Prediction with a horizon of *k* sampling periods**

$$\mathbf{t}_{t+k} = \boxed{\mathbf{A}_k} \cdot [\mathbf{p}_t \quad \mathbf{p}_{t-1} \quad \cdots \quad \mathbf{p}_{t-h+1}]$$

Regression matrix (offline trained)

[1] A High-Fidelity Temperature Distribution Forecasting System for Data Centers. RTSS 2012.

Real-time Temperature Prediction [1]

- **Vector p$_t$:** Sensor measurements at moment *t*
- **Prediction with a horizon of *k* sampling periods**

$$\mathbf{t}_{t+k} = \boxed{\mathbf{A}_k} \cdot \boxed{[\mathbf{p}_t \quad \mathbf{p}_{t-1} \quad \cdots \quad \mathbf{p}_{t-h+1}]}$$

Regression matrix (offline trained)

Historical measurements

[1] A High-Fidelity Temperature Distribution Forecasting System for Data Centers. RTSS 2012.

Real-time Temperature Prediction [1]

- **Vector $\mathbf{p}_t$:** Sensor measurements at moment *t*
- **Prediction with a horizon of *k* sampling periods**

$$\mathbf{t}_{t+k} = \boxed{\mathbf{A}_k} \cdot \boxed{[\mathbf{p}_t \quad \mathbf{p}_{t-1} \quad \cdots \quad \mathbf{p}_{t-h+1}]}$$

Regression matrix (offline trained)

Historical measurements

•Increasing *k*: temperature distribution evolution

[1] A High-Fidelity Temperature Distribution Forecasting System for Data Centers. RTSS 2012.

Real-time Temperature Prediction [1]

- **Vector $\mathbf{p}_t$:** Sensor measurements at moment *t*
- **Prediction with a horizon of *k* sampling periods**

$$\mathbf{t}_{t+k} = \boxed{\mathbf{A}_k} \cdot \boxed{[\mathbf{p}_t \quad \mathbf{p}_{t-1} \quad \cdots \quad \mathbf{p}_{t-h+1}]}$$

Regression matrix (offline trained)

Historical measurements

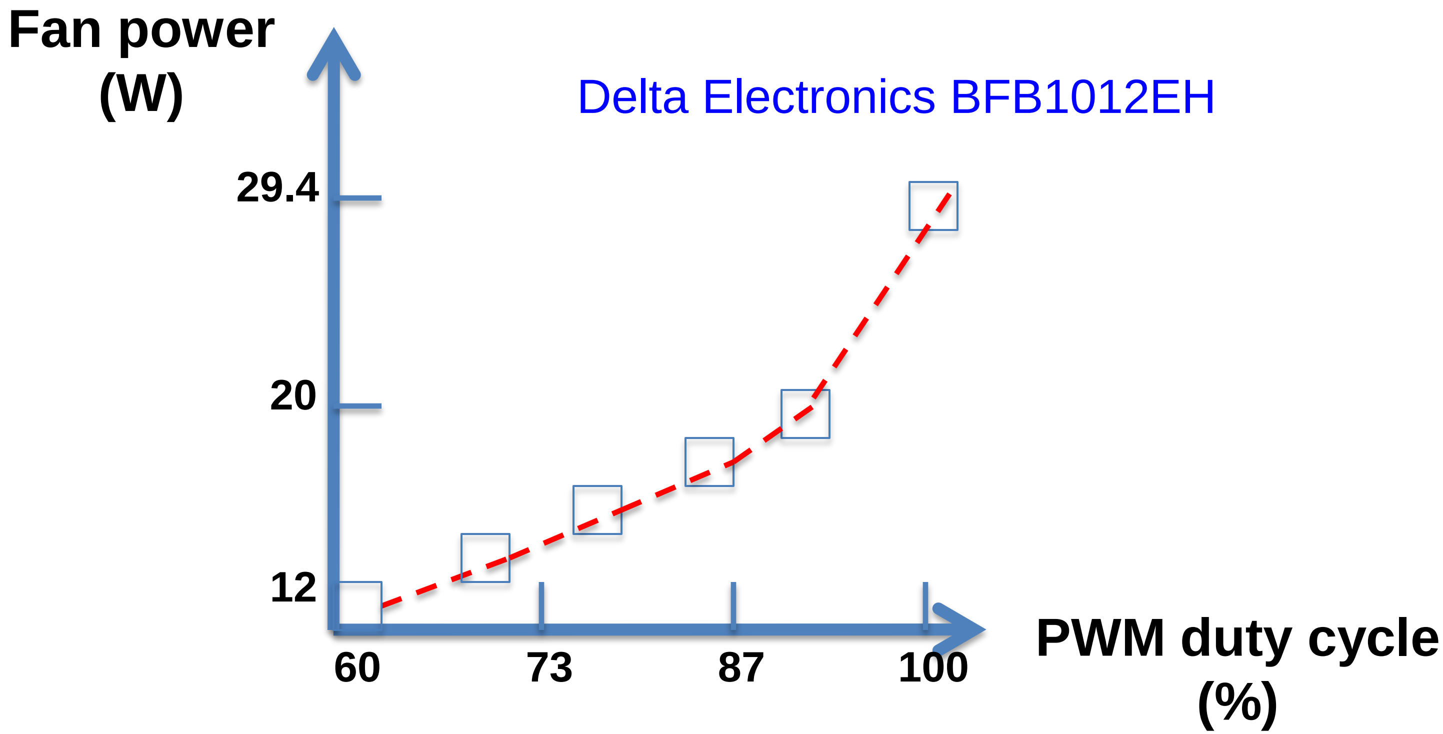
- Increasing *k*: temperature distribution evolution

- **Less than 0.5 °C error when *k* < 10 min**
  - Error increases with *k*

[1] A High-Fidelity Temperature Distribution Forecasting System for Data Centers. RTSS 2012.

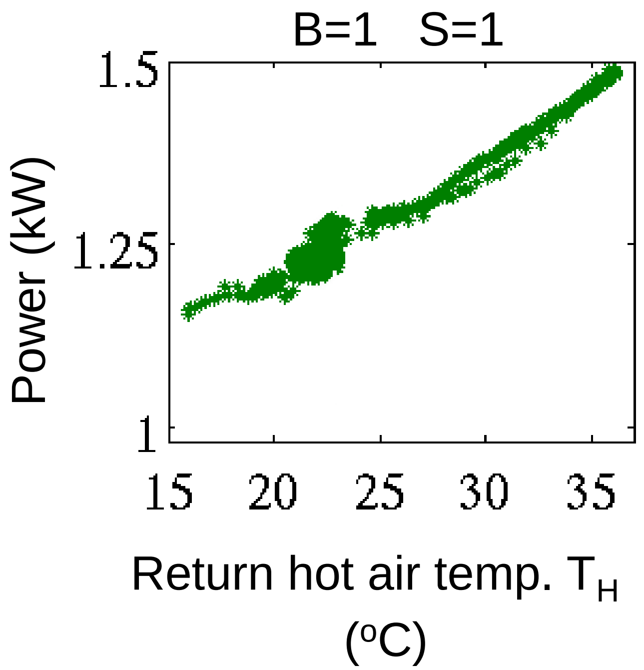
Server Fan Power Model

- **Fan regulates speed by duty cycle of PWM signal**
  - Part of control solution
- **Offline / online learning**



AC Power Model

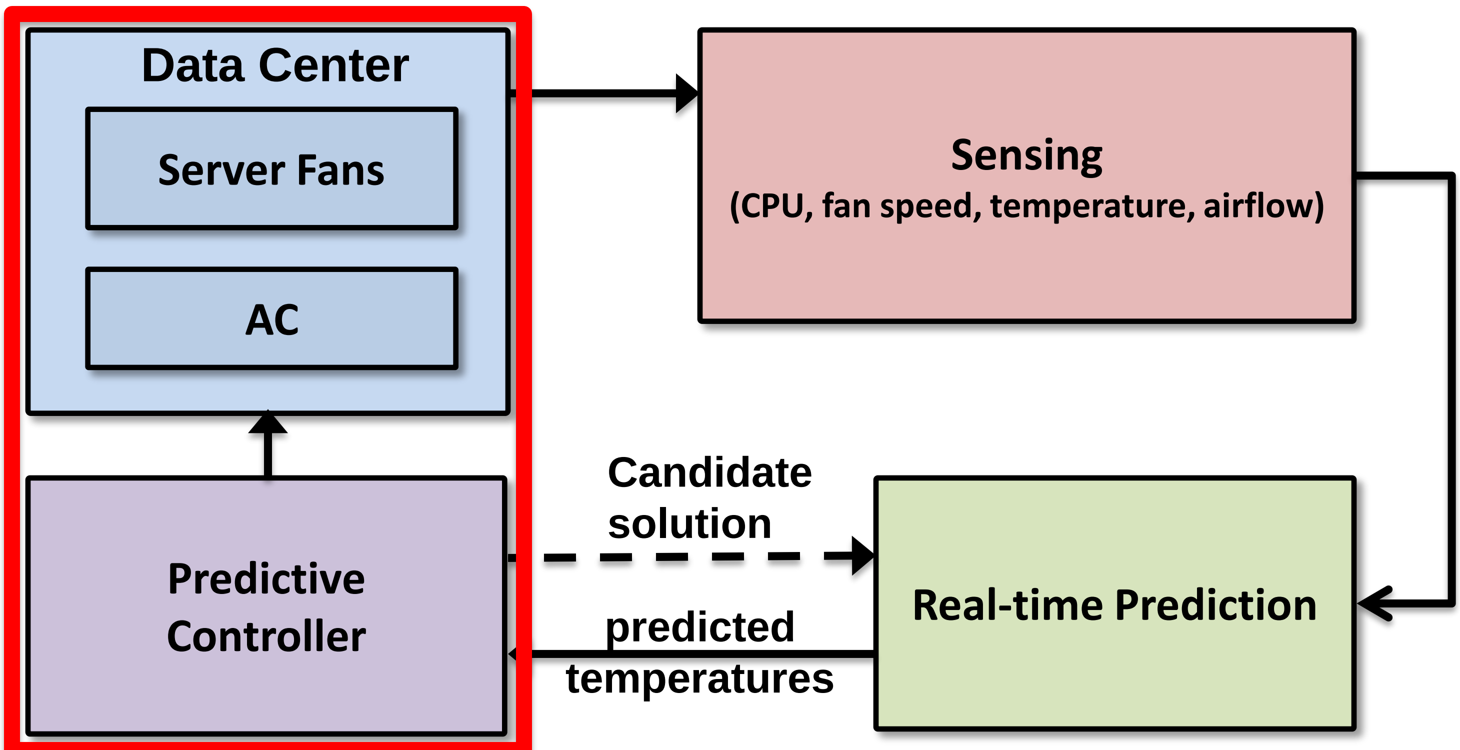
- **AC power consumption**
  - Temperature setpoint, blower speed, return hot air temperature
  - Offline learning or from spec.
- **AC in our testbed**
  - Binary blower state (B)
  - Binary compressor state (S)
  - Return hot air temp. T_H



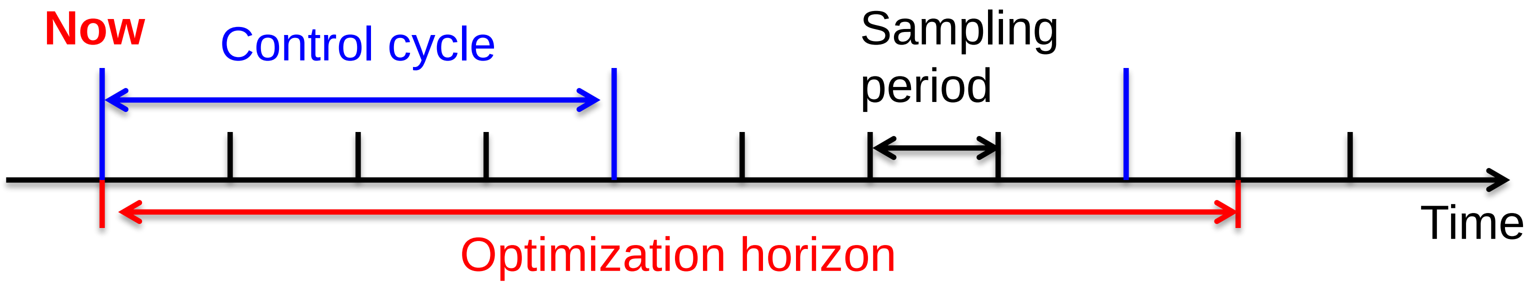
$$P_{AC} = B \cdot [S \cdot (\omega_1 \cdot T_H + \omega_0) + \omega_2]$$

Outline

- Motivation & Approach Overview
- Sensing and Prediction
- **Predictive Thermal & Energy Control**
- Evaluation

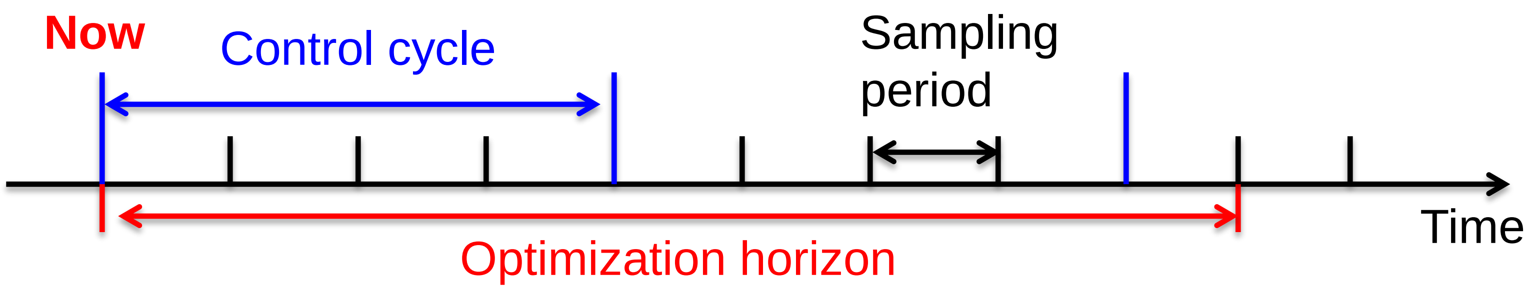


Problem Formulation



- **Find fan speeds, AC settings**
  - Minimize predicted total power in opt. horizon
  - Predicted inlet temp. upper-bounded
    Prevent overheating
  - Predicted inlet temp. variation upper-bounded
    Failure rate increases with variation [El-Sayed 2012]

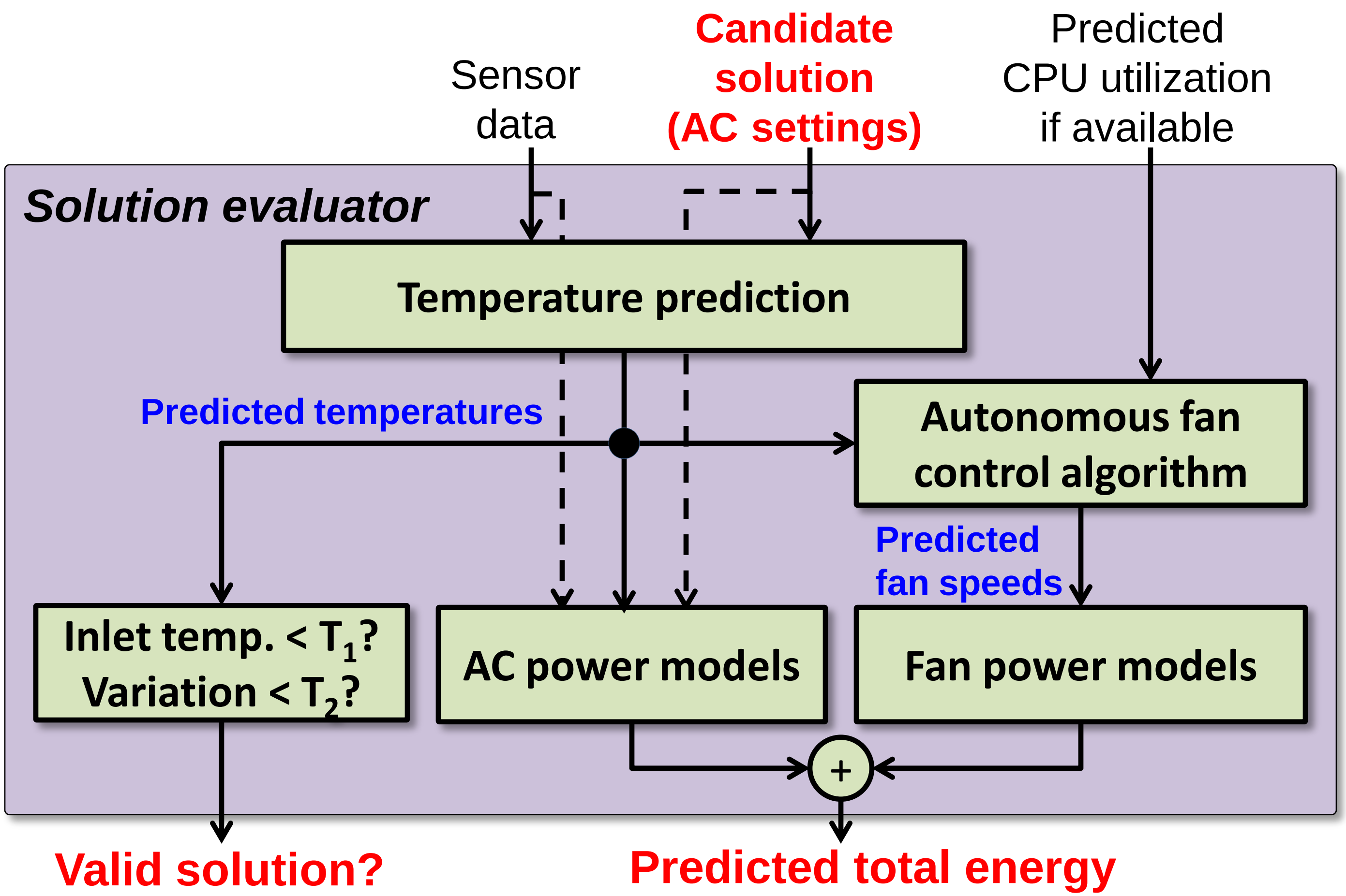
Problem Formulation



- **Find fan speeds, AC settings**
  - Minimize predicted total power in opt. horizon
  - Predicted inlet temp. upper-bounded
    Prevent overheating
  - Predicted inlet temp. variation upper-bounded
    Failure rate increases with variation [El-Sayed 2012]
- **Compute-intensive**
  - A small data center: **237** controllable variables
    229 fans + 4 blower speeds + 4 temp. setpoints

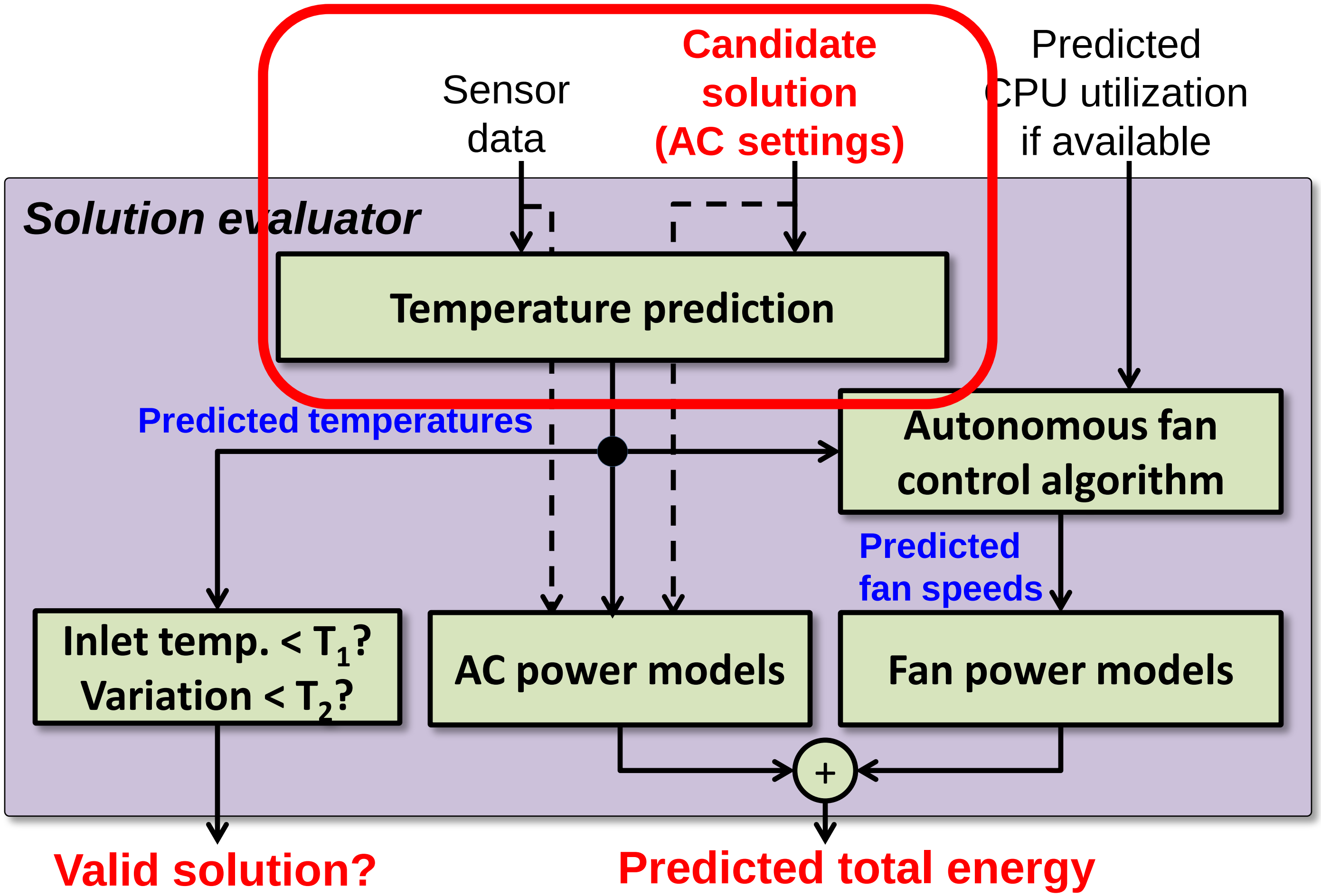
Coordinated Control

- **Autonomous fan control**
  - speed = *f* (inlet temperature, CPU utilization)



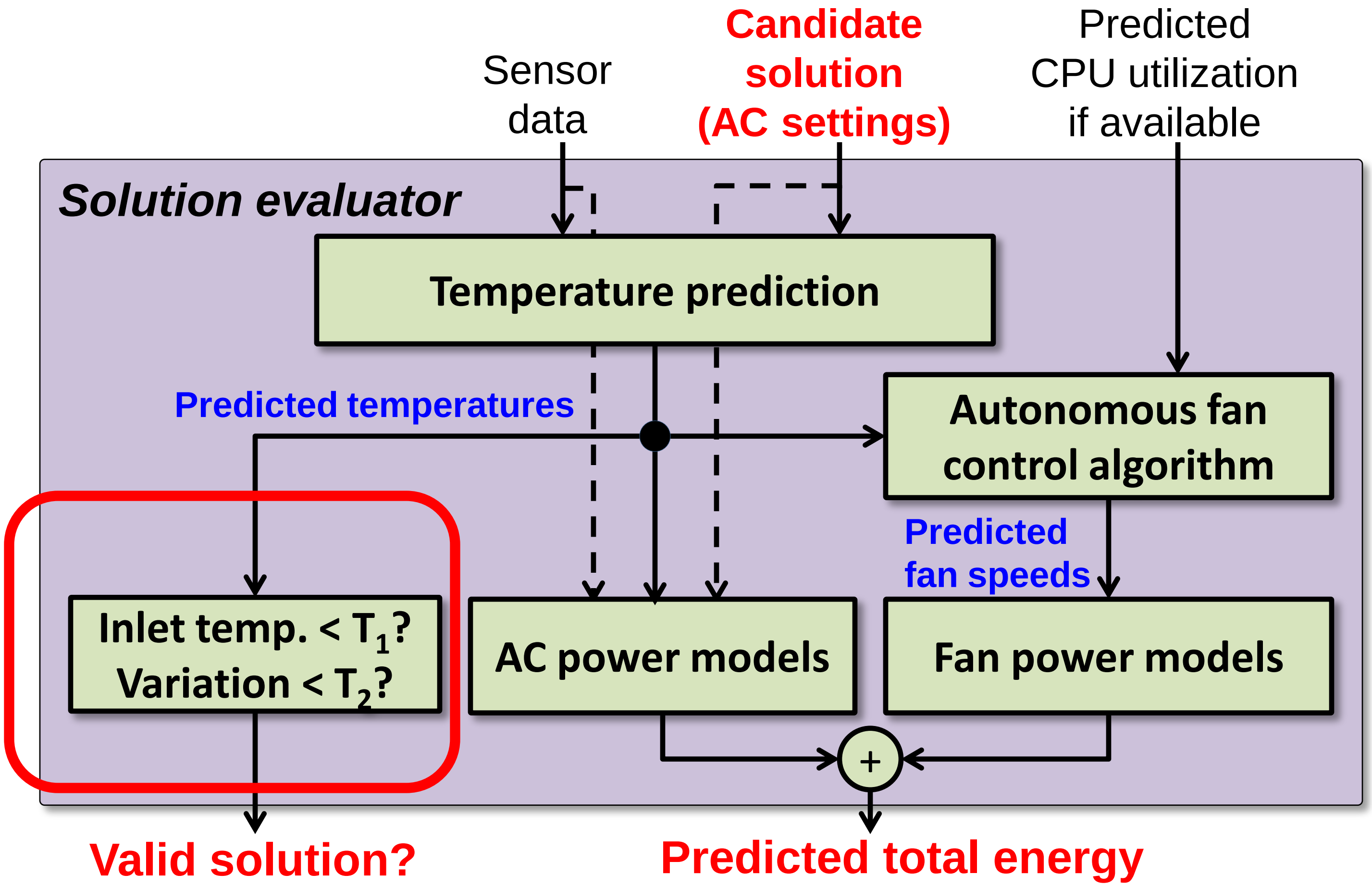
Coordinated Control

- **Autonomous fan control**
  - speed = *f* (inlet temperature, CPU utilization)



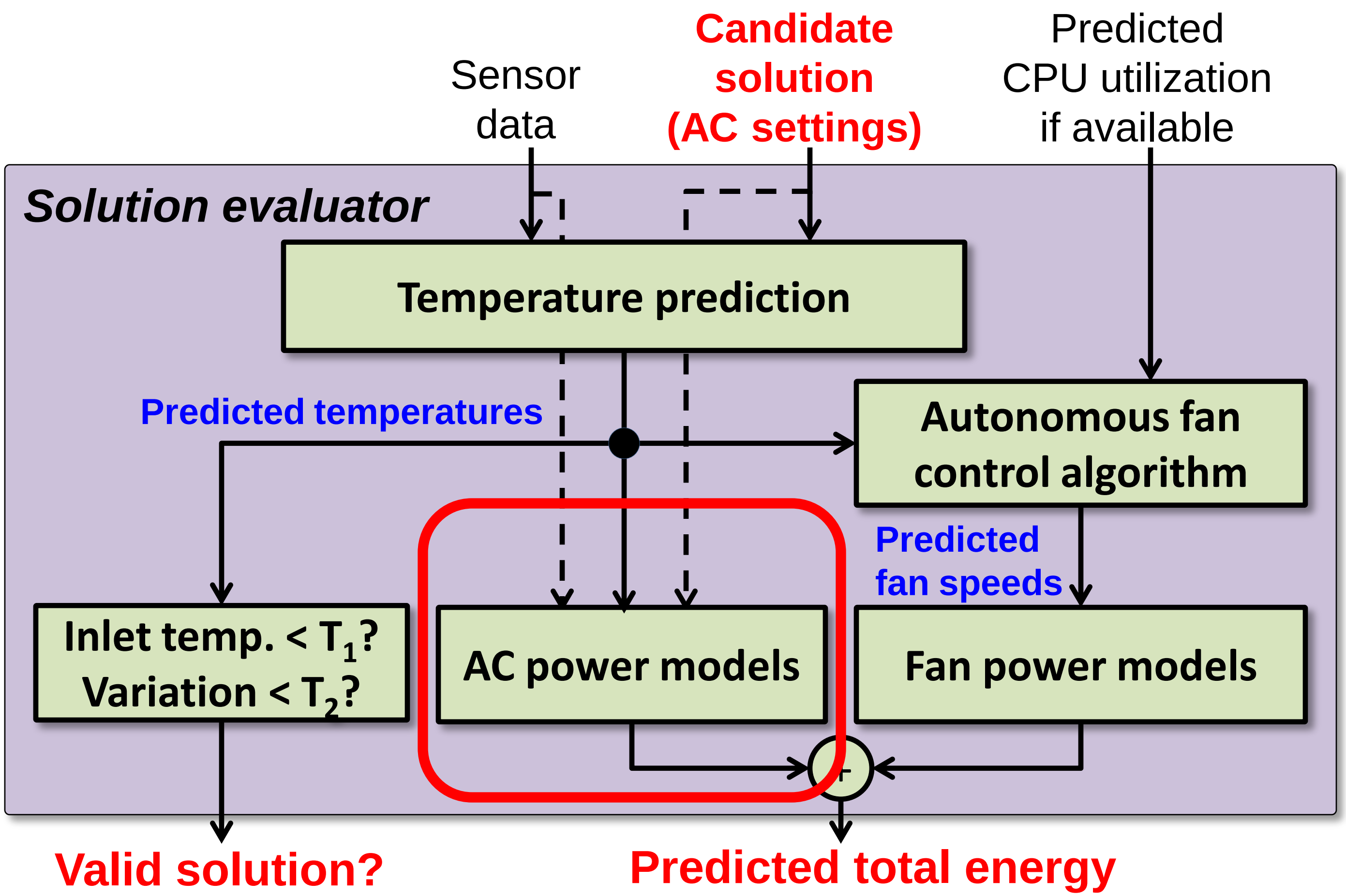
Coordinated Control

- **Autonomous fan control**
  - speed = *f* (inlet temperature, CPU utilization)



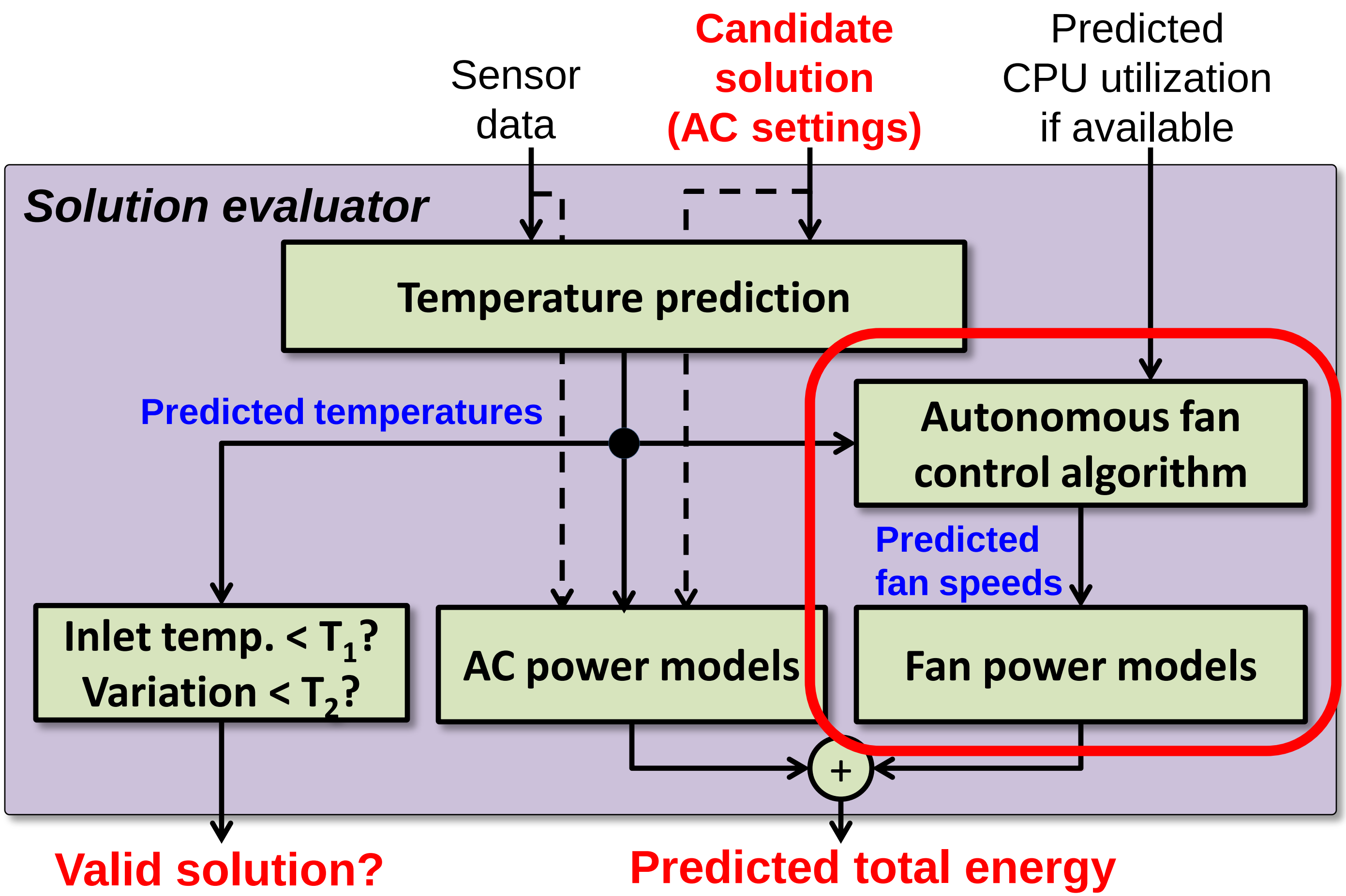
Coordinated Control

- **Autonomous fan control**
  - speed = *f* (inlet temperature, CPU utilization)



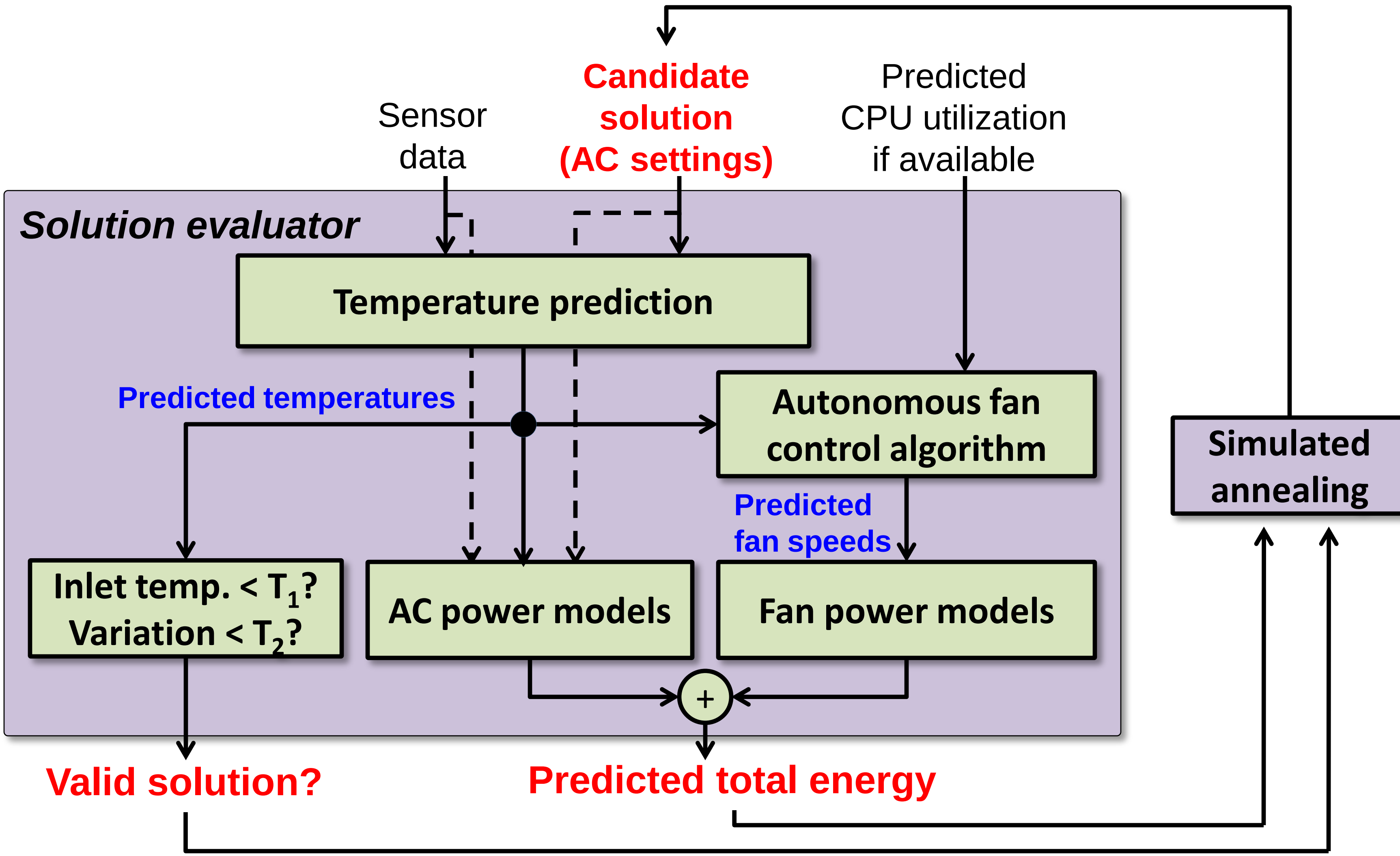
Coordinated Control

- **Autonomous fan control**
  - speed = *f* (inlet temperature, CPU utilization)



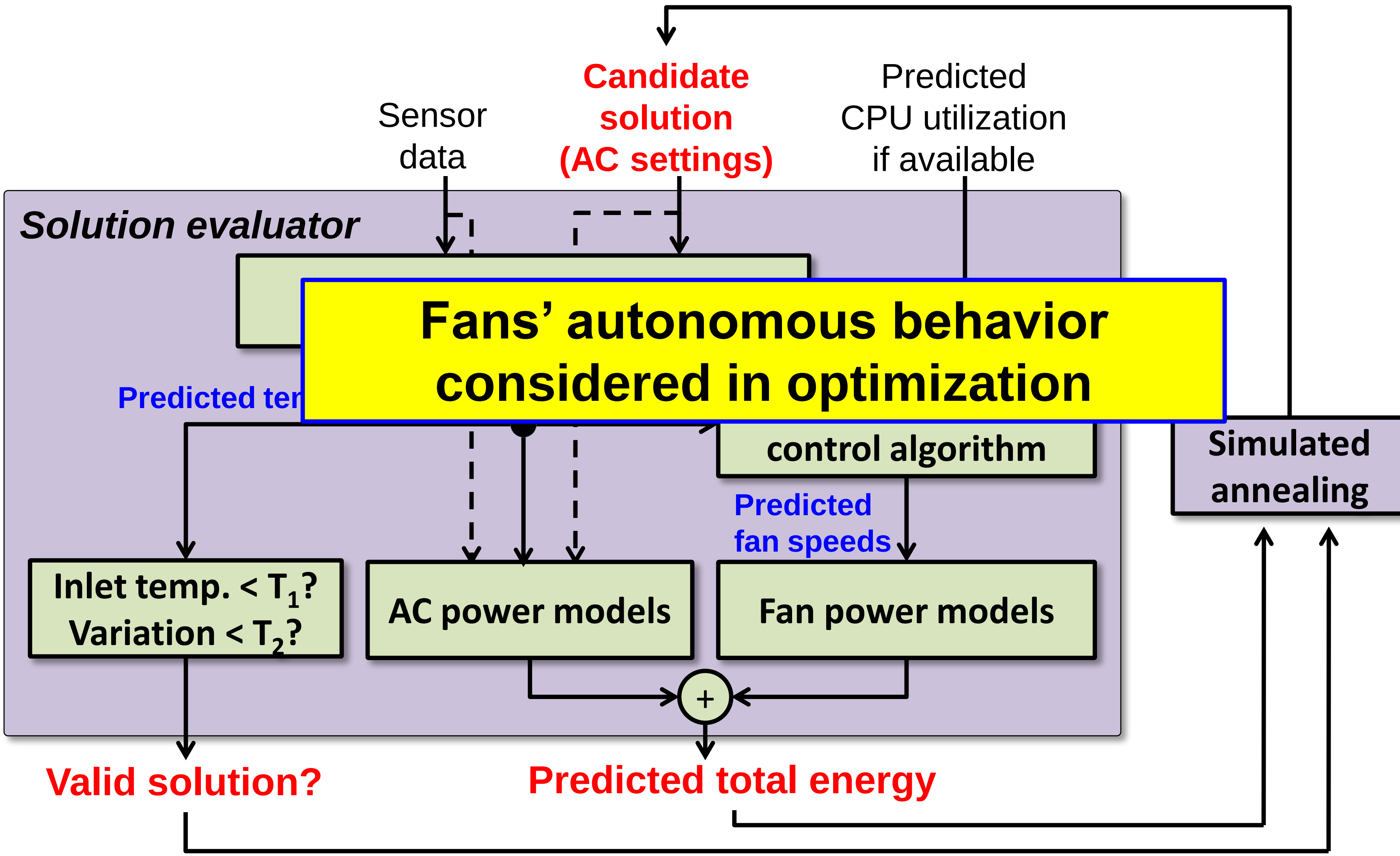
Coordinated Control

- **Autonomous fan control**
  - speed = *f* (inlet temperature, CPU utilization)



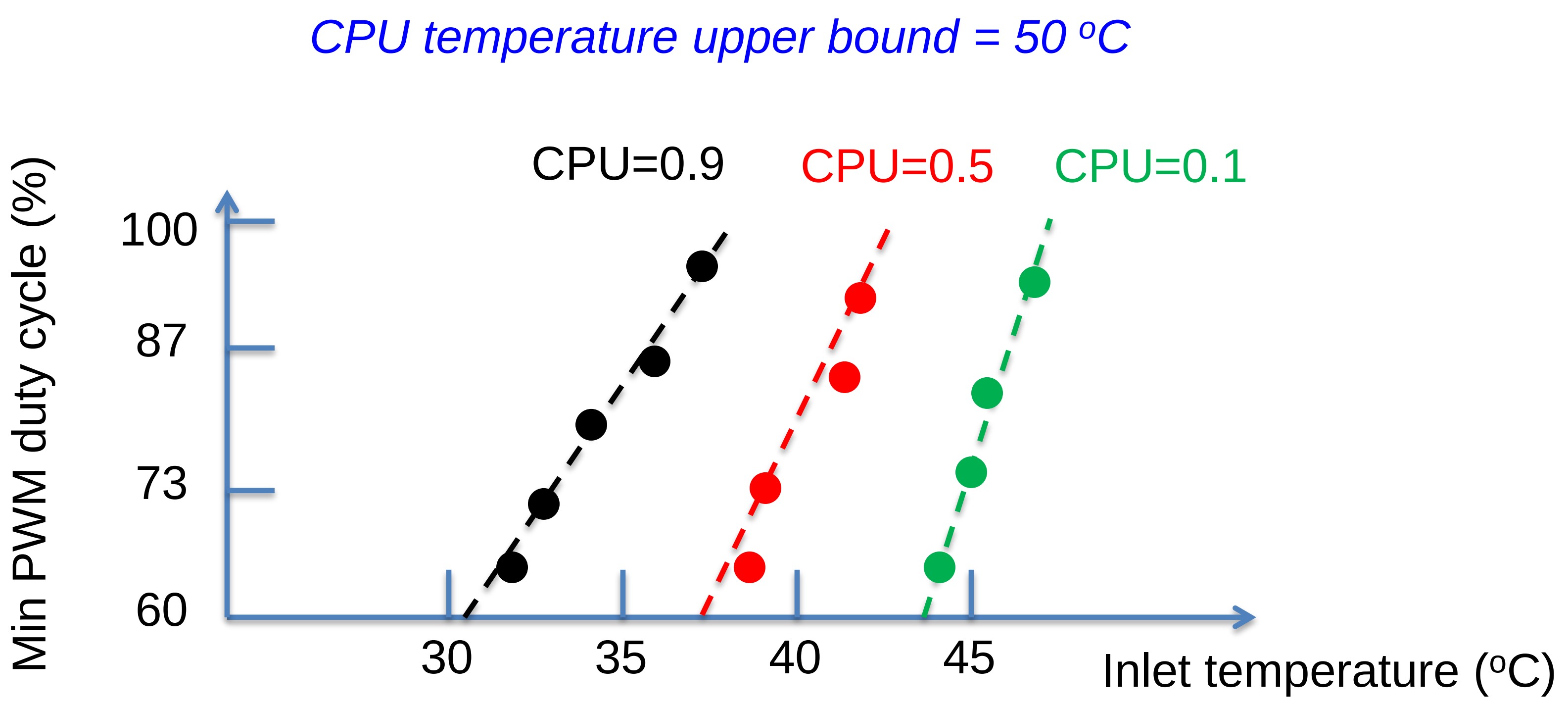
Coordinated Control

- **Autonomous fan control**
  - speed = *f* (inlet temperature, CPU utilization)



Autonomous Fan Control

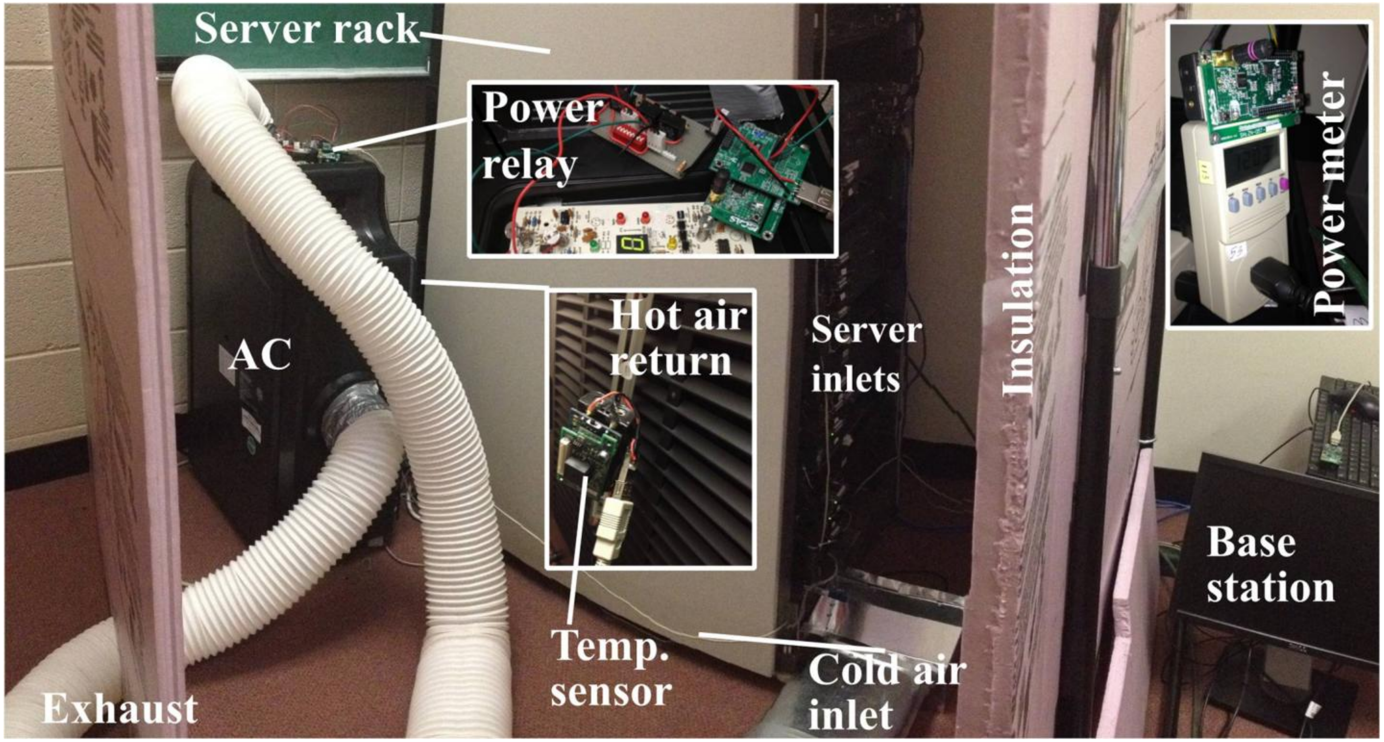
- **Ensure upper-bounded CPU temperature**
  – Measurement-based approach



Outline

- Motivation & Approach Overview
- Sensing and Prediction
- Predictive Thermal & Energy Control
- **Evaluation**

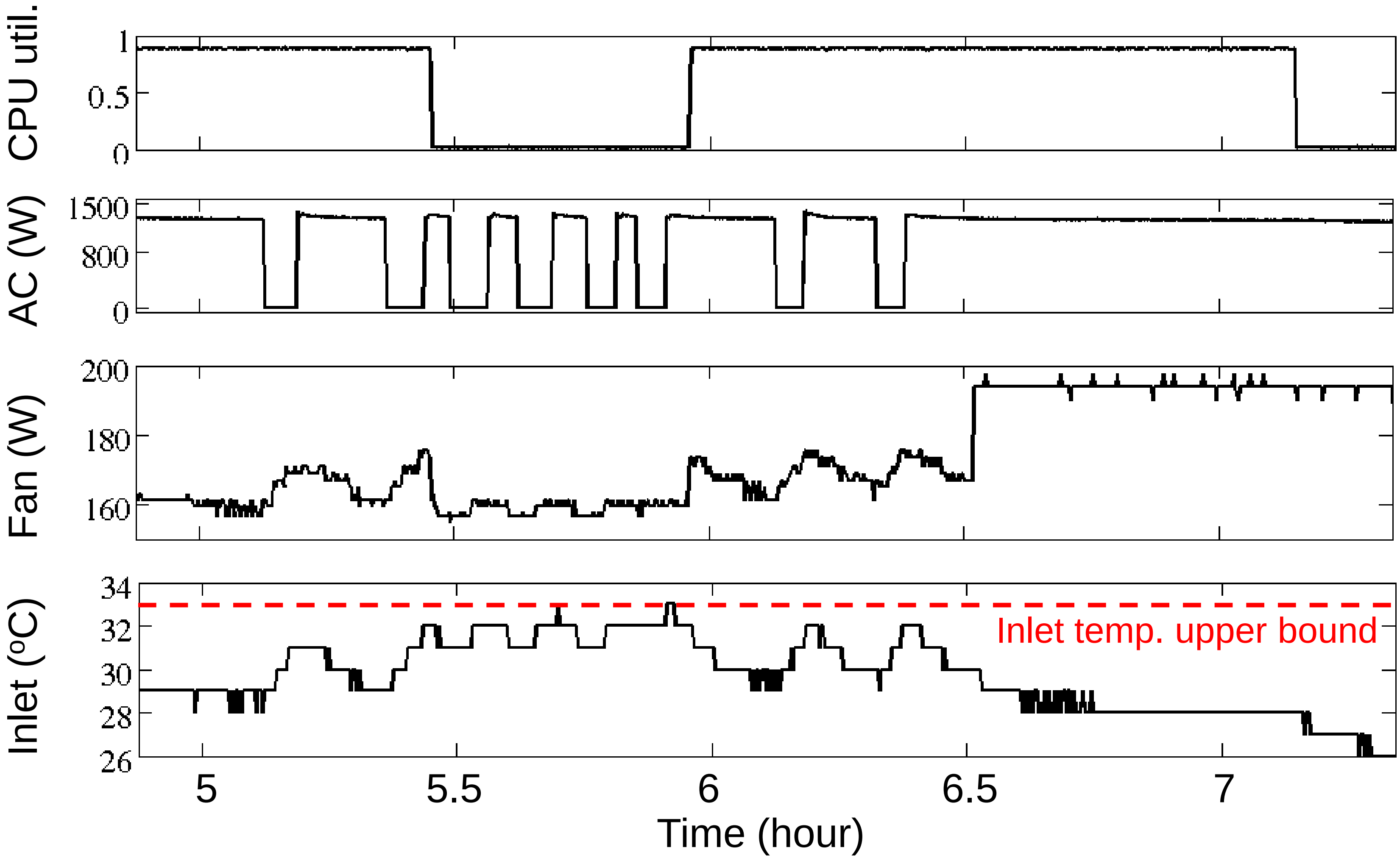
Single-Rack Experiments



- **Setup**
  - 15 servers, 32 wireless sensors, portable AC with wireless power relay
  - Controllable CPU utilization
- **System implementation**
  - Predictive controller: MATLAB on a desktop
  - Fan control: BASH on servers

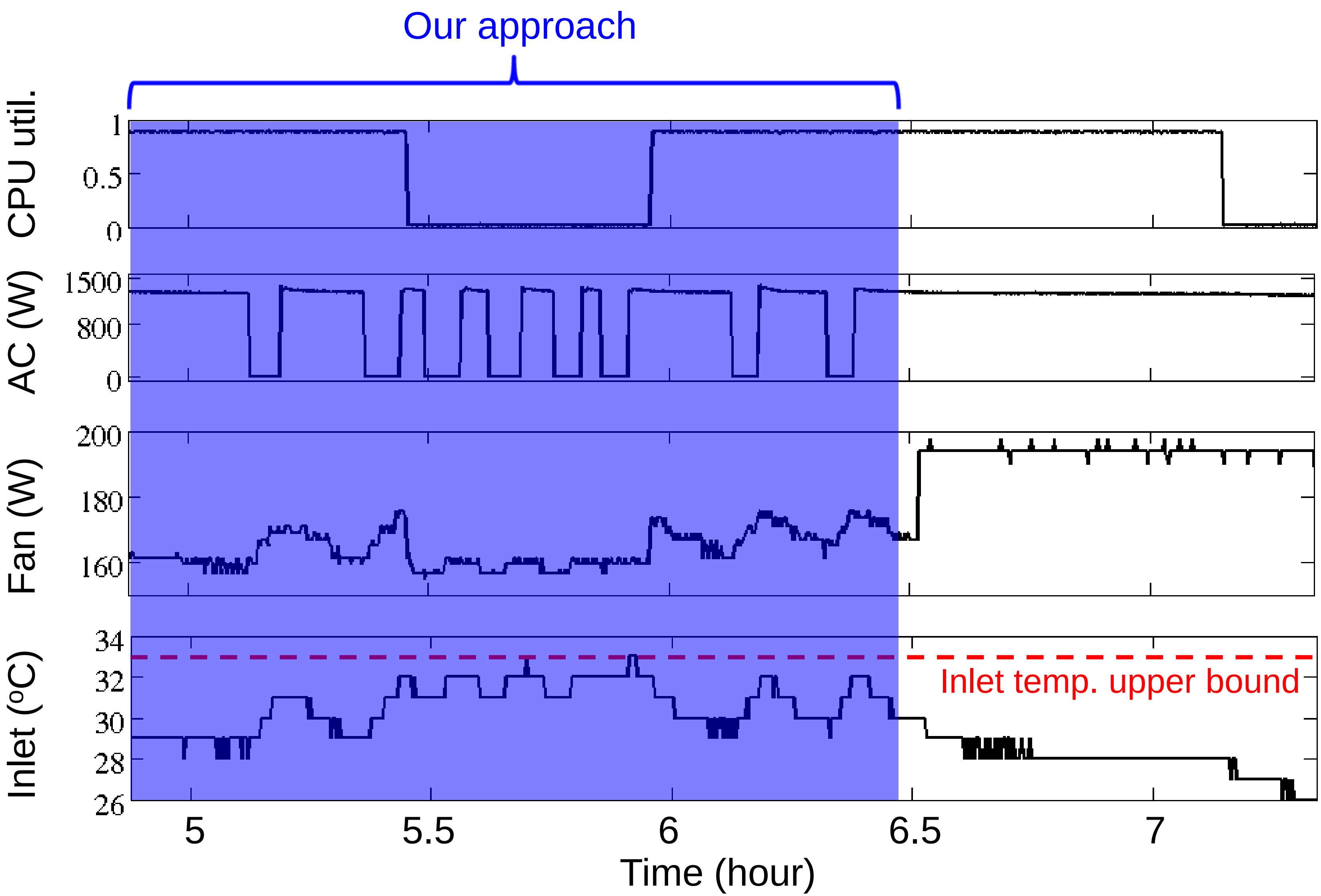
Compare with Max Cooling

- Max Cooling: fixed low AC setpoint, full server fan speed



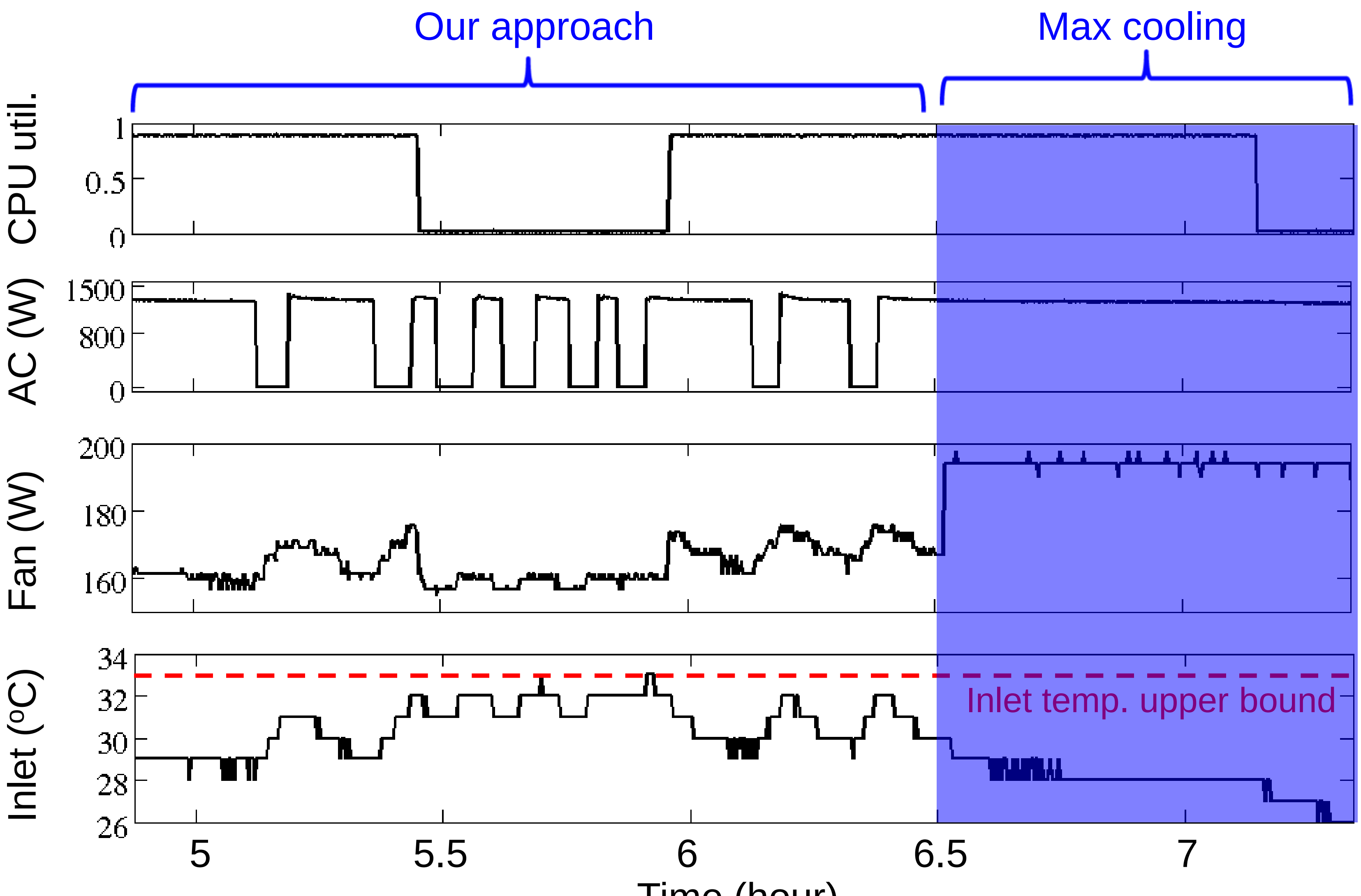
Compare with Max Cooling

- Max Cooling: fixed low AC setpoint, full server fan speed



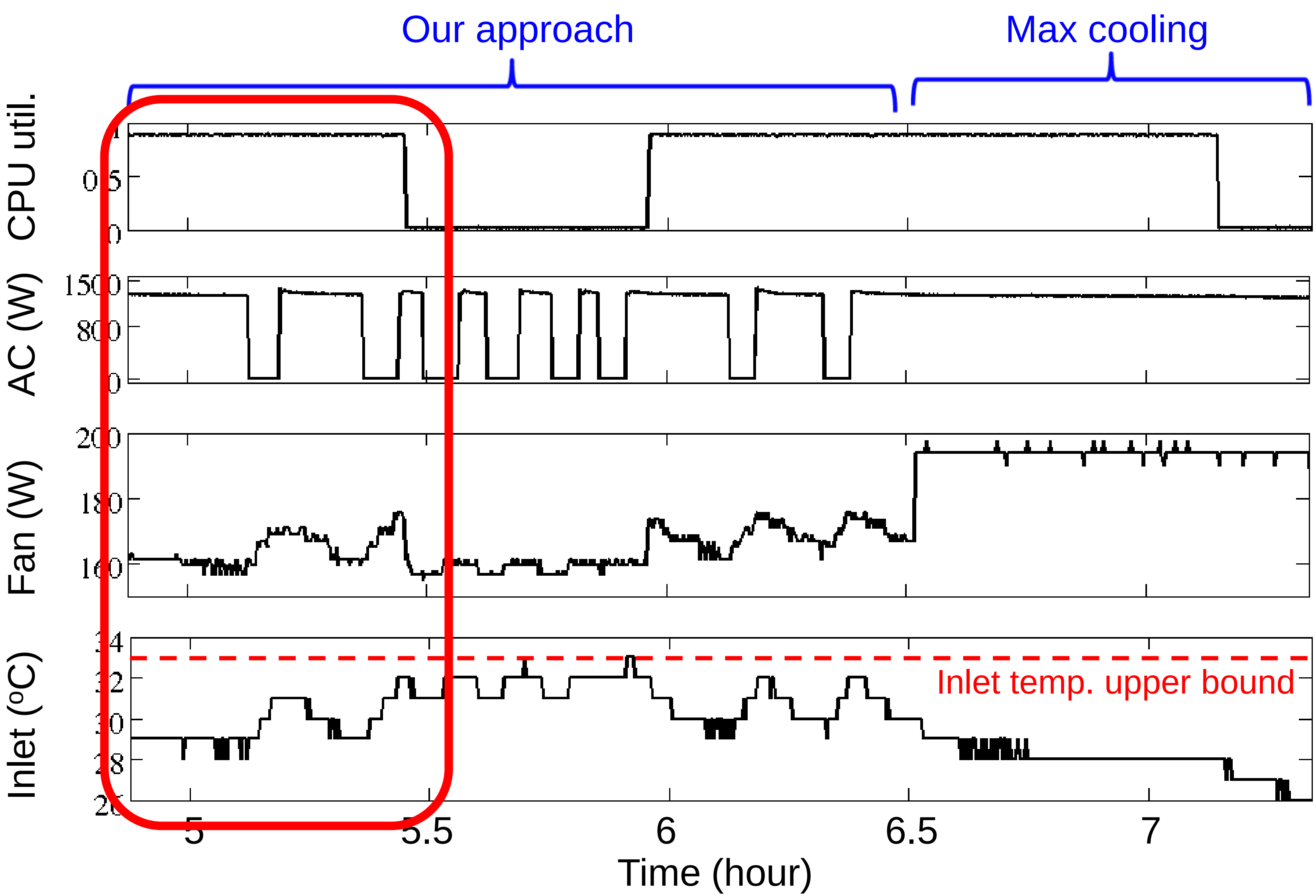
Compare with Max Cooling

- Max Cooling: fixed low AC setpoint, full server fan speed



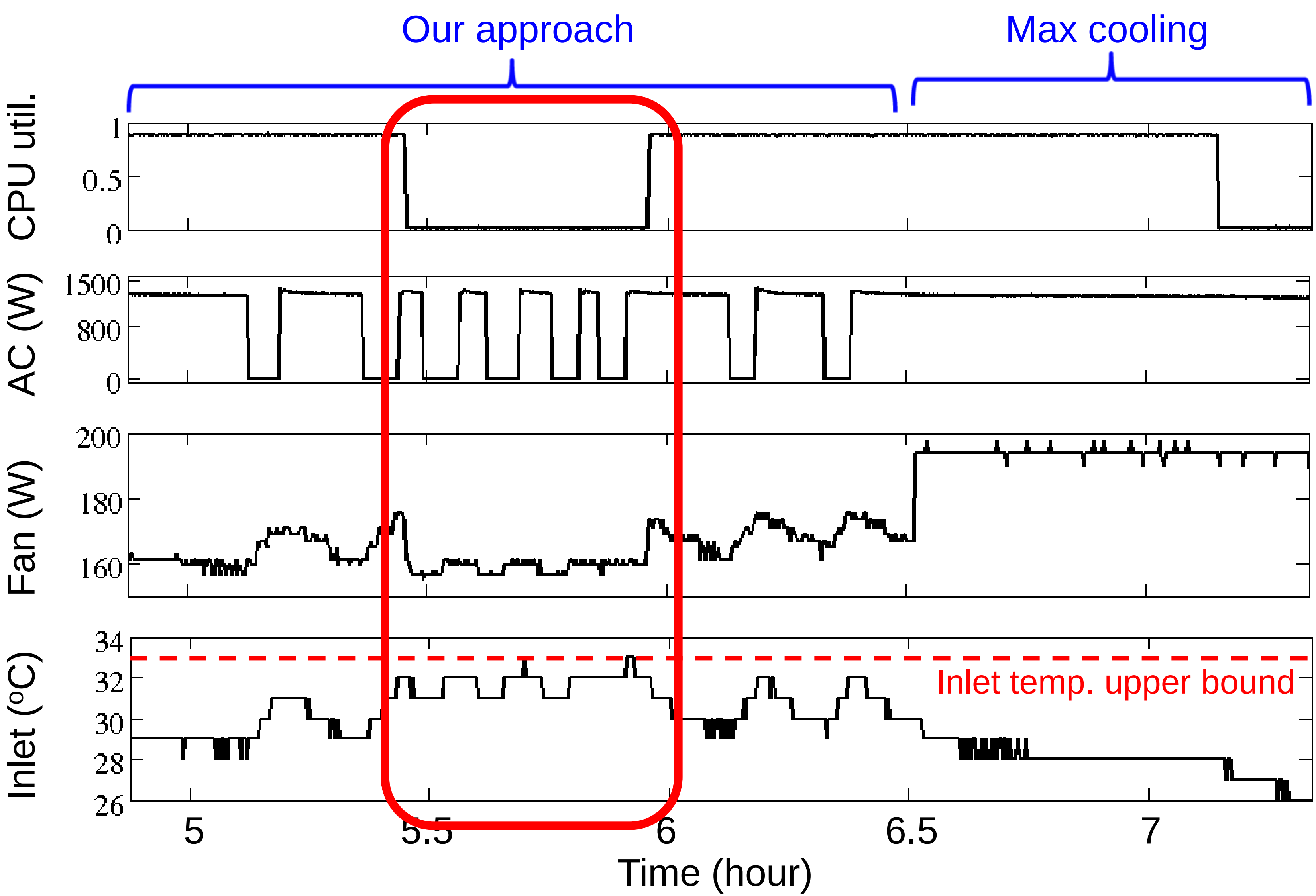
Compare with Max Cooling

- Max Cooling: fixed low AC setpoint, full server fan speed



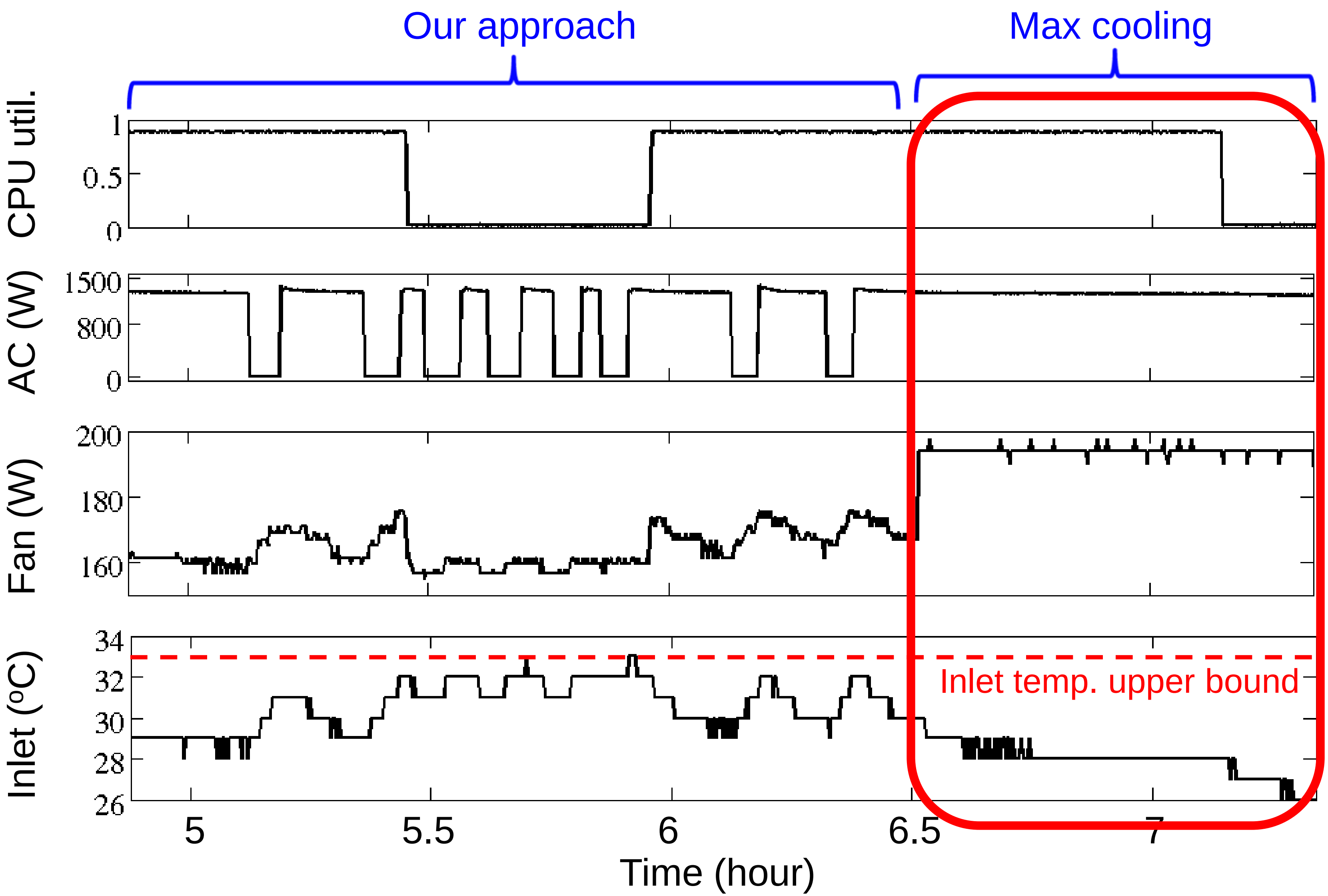
Compare with Max Cooling

- Max Cooling: fixed low AC setpoint, full server fan speed

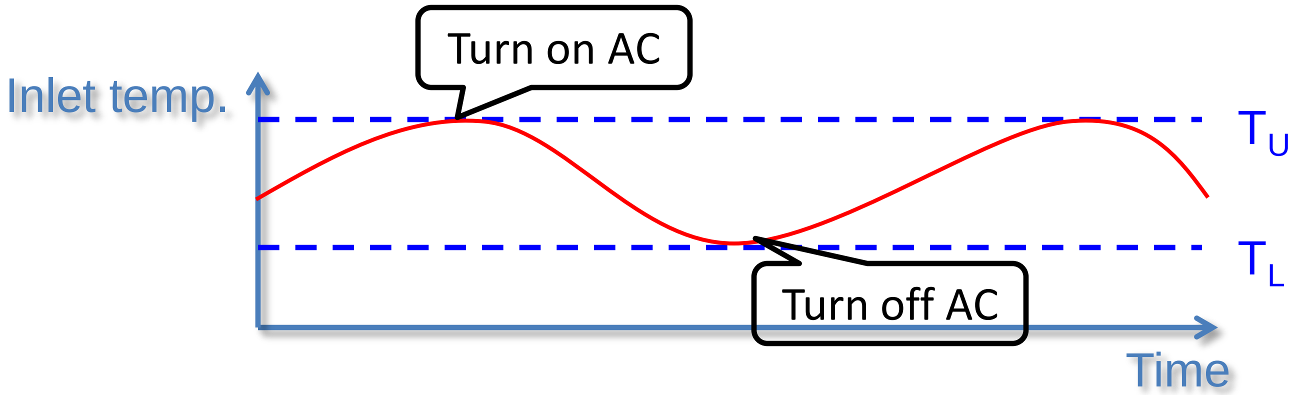


Compare with Max Cooling

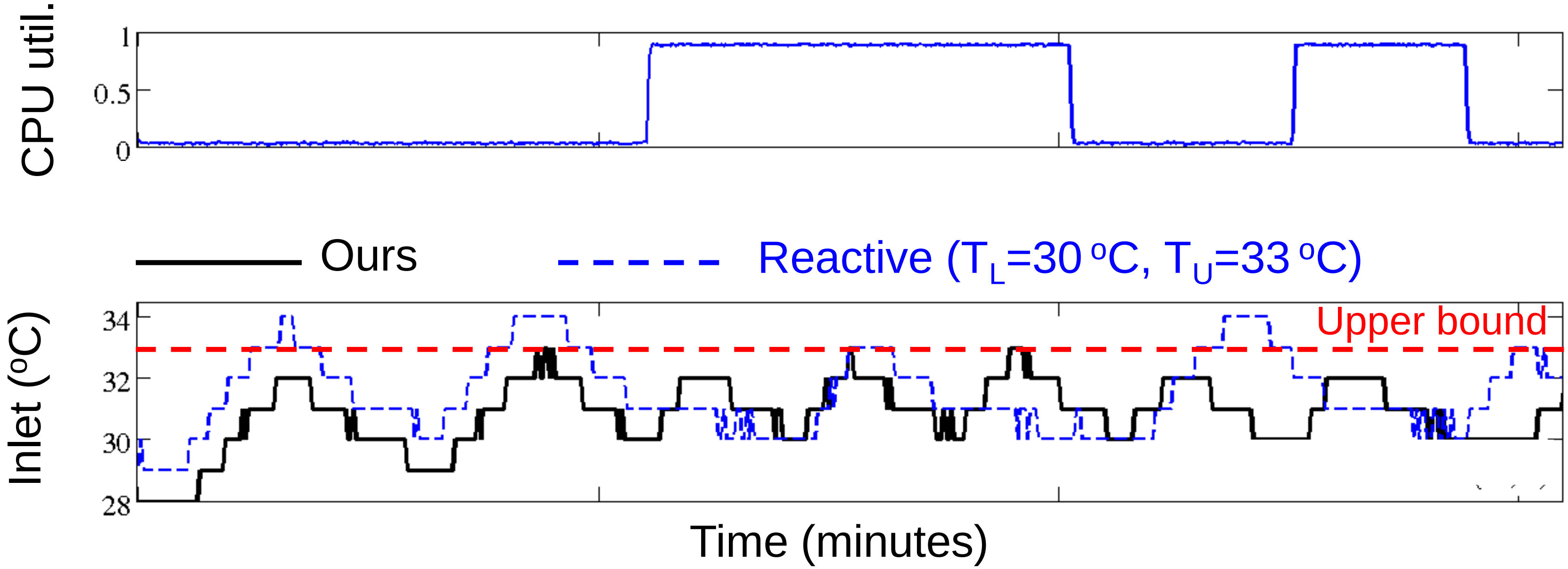
- Max Cooling: fixed low AC setpoint, full server fan speed



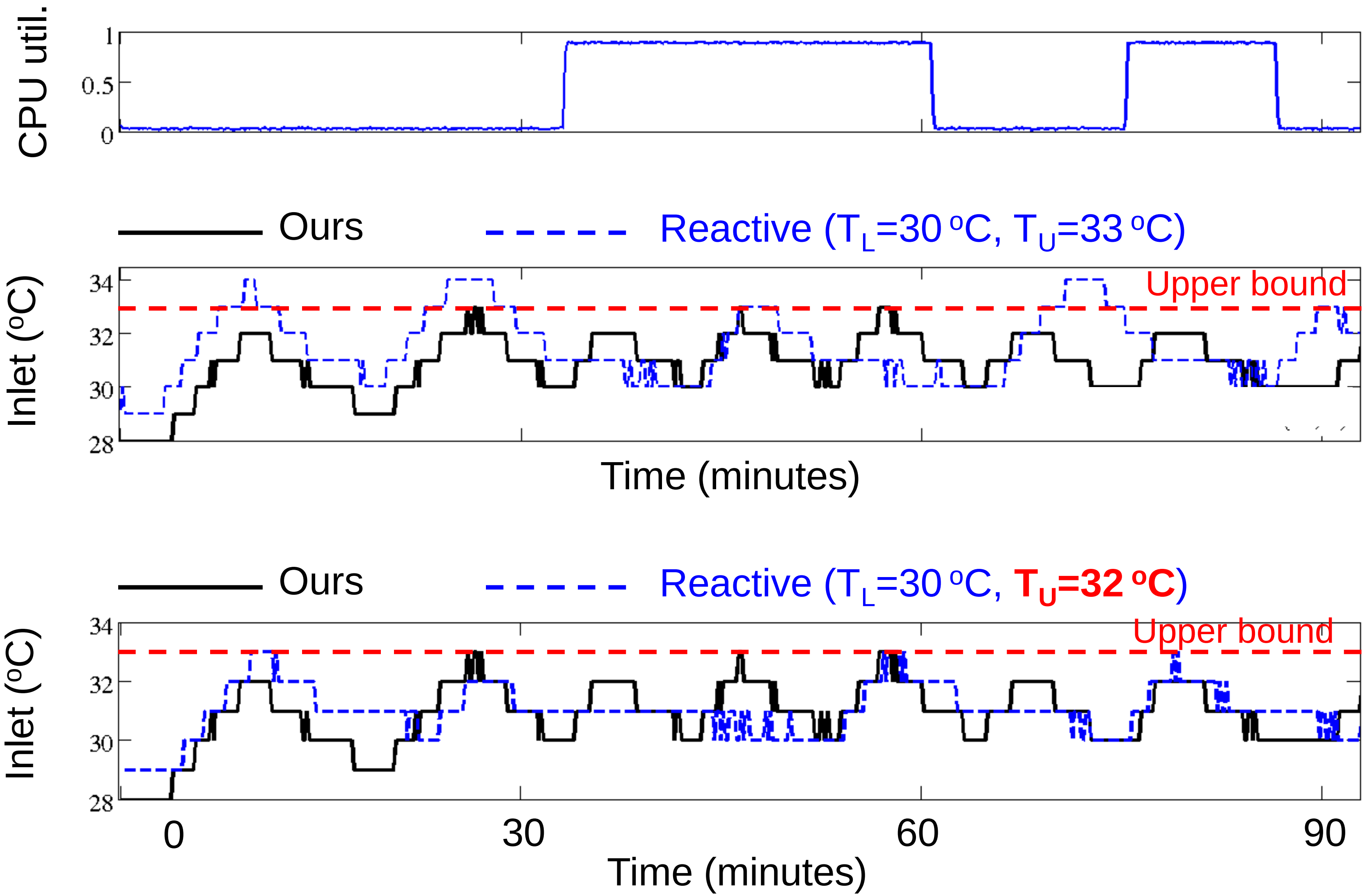
Compare with Reactive



Compare with Reactive



Compare with Reactive



Compare with Reactive (cont'd)

Inlet temp. upper bound = 33^{o}C, server idle

Reactive			Ours
T_L (^{o}C)	T_U (^{o}C)	Avg power (Watt)	Avg power (Watt)
27	30	916	638
28	30	807	
28	31	806	
29	31	817	
29	32	746	
30	32	669	
30	33	714	
31	33	640	

Compare with Reactive (cont'd)

Inlet temp. upper bound = 33°C, server idle

Reactive			Ours
T_L (°C)	T_U (°C)	Avg power (Watt)	Avg power (Watt)
27	30	916	> 638
28	30	807	
28	31	806	
29	31	817	
29	32	746	
30	32	669	
30	33	714	
31	33	640	

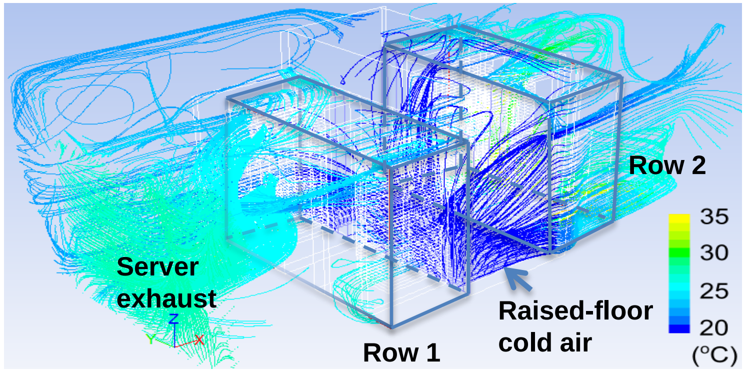
Compare with Reactive (cont'd)

Inlet temp. upper bound = 33°C, server idle

Reactive			Ours
T_L (°C)	T_U (°C)	Avg power (Watt)	Avg power (Watt)
27	30	916	> 638
28	30	807	
28	31	806	
29	31	817	
29	32	746	
30	32	669	
30	33	714	
31	33	640	

Inlet temp. > 33°C

Trace-Driven CFD Simulations

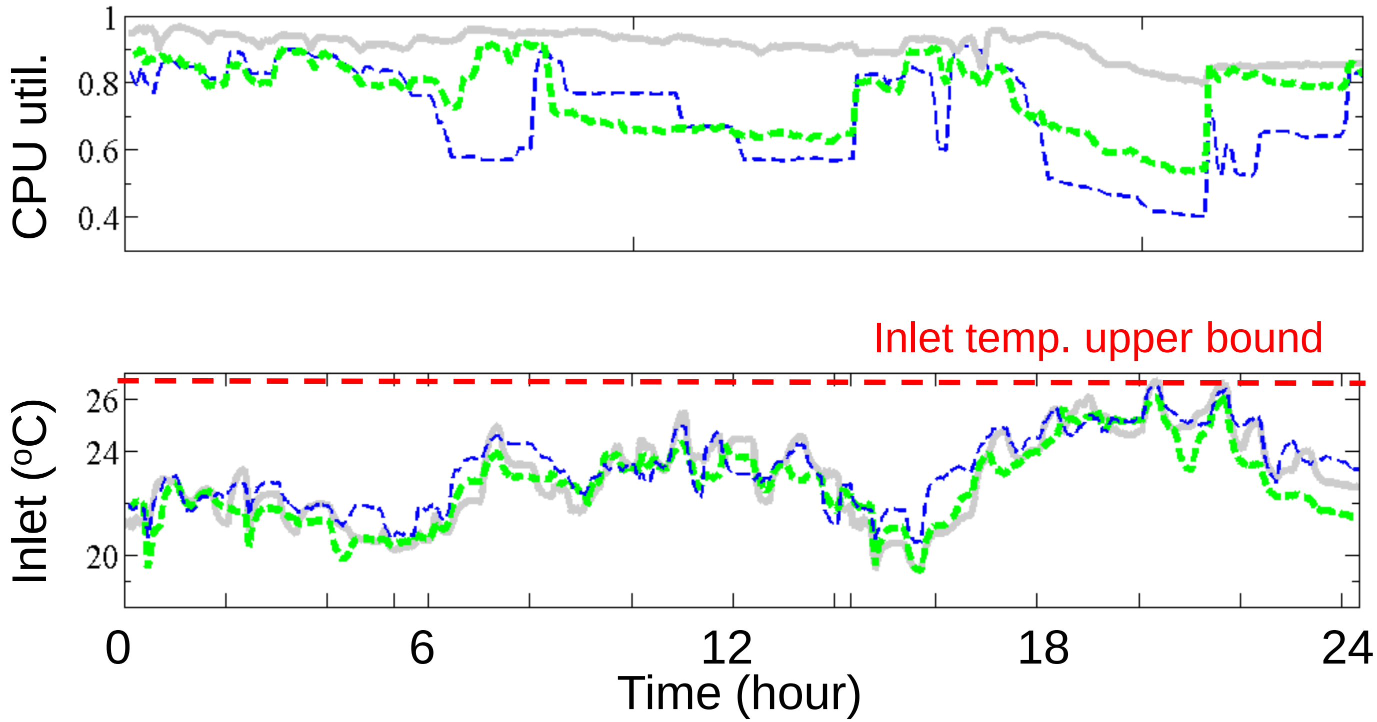


Computational fluid dynamics (CFD) model

- **MSU HPCC**
  - 5 racks, 229 servers, 4 in-row ACs
- **CPU utilization data trace**
  - 1 sample / minute
  - 12 days

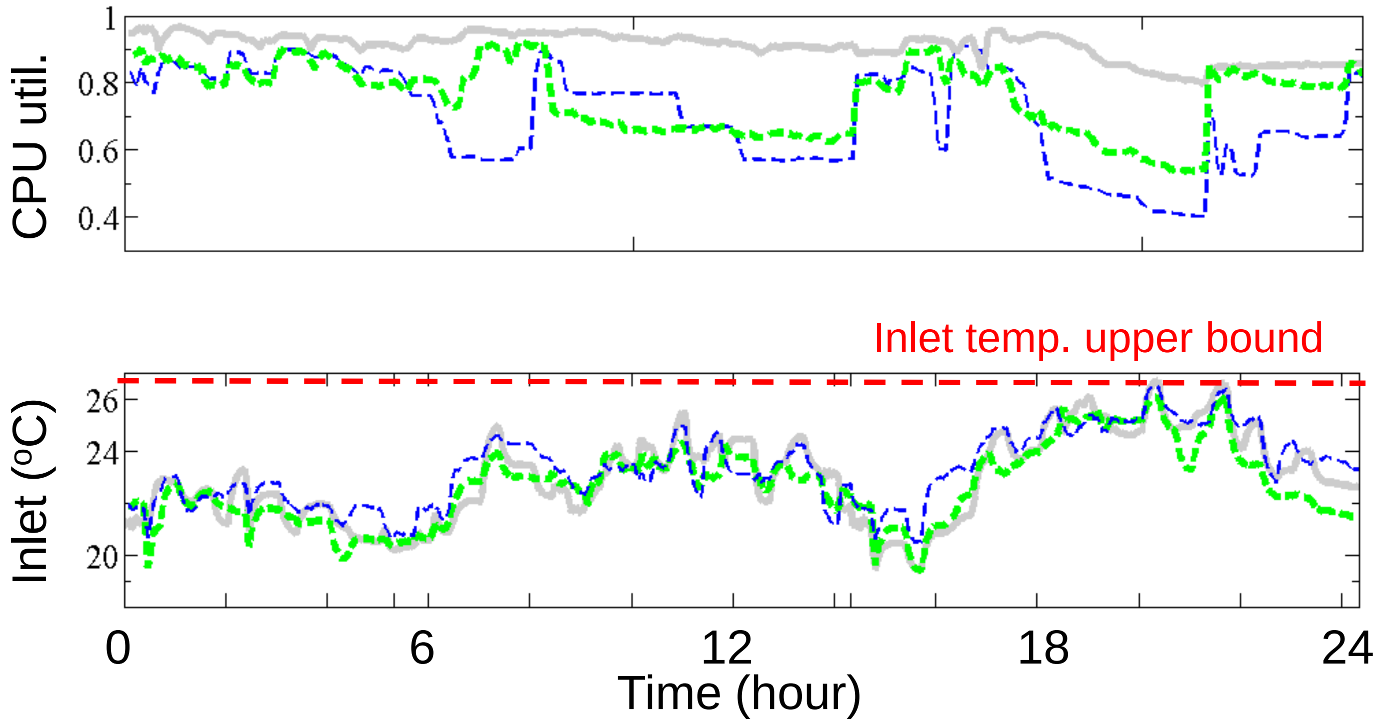
Dynamic Workload

- Results for 3 racks



Dynamic Workload

- Results for 3 racks



- Compare with TAPO [Huang et al. 2011]
  - Need fine parameter tuning
  - Poorly adapt to dynamic workload

Conclusion

- **Predictive thermal and energy control**
  - Minimize AC and fan energy
  - Upper-bound inlet temperature & variation
- **Coordinated Control**
  - Autonomous fan control to ensure CPU temp.
  - Reduce complexity
- **Testbed experiments & CFD simulations**
  - Outperform reactive approach
  - Adapt to dynamic workload