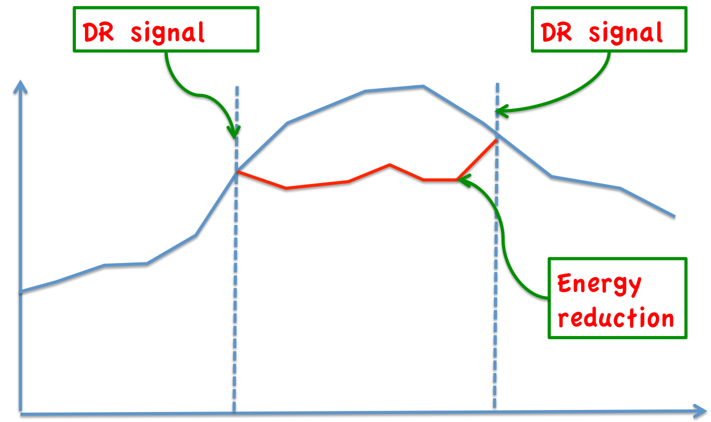
Energy Demand Response Modeling for High Performance Computing Systems

Kishwar Ahmed and Jason Liu
Florida International University

Workshop on Modeling and Simulation of Systems and Applications
August 15-17, 2018◆ University of Washington, Seattle, Washington

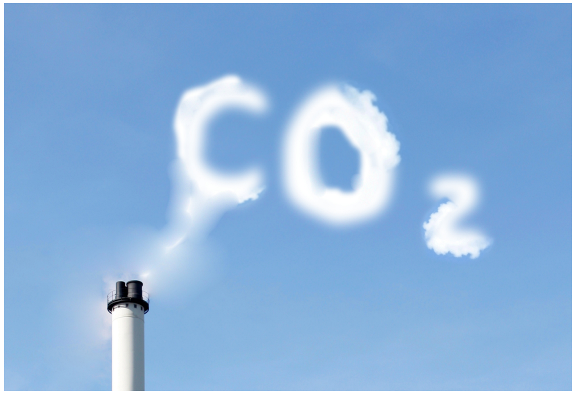
Demand Response

- Participants reduce energy consumption during
  - Emergency events
  - High electricity price period
- Emergency demand response
  - Mandatory energy reduction to target level
- Economic demand response
  - Voluntary participation based on economic incentives



Why Demand Response?

Financial benefits



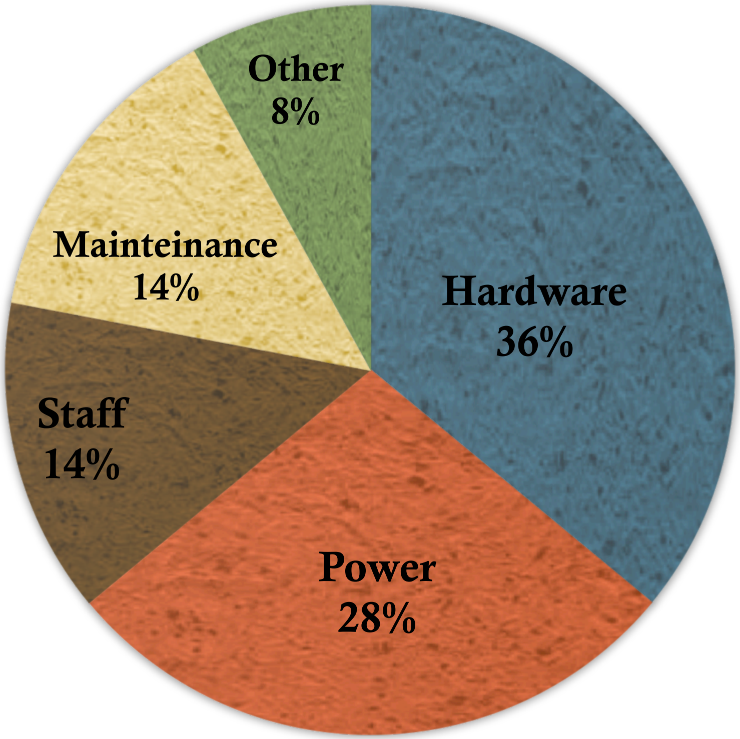
Environmental benefits

Power system stability

- Increase in demand response participation
  - Many well-known companies, such as Google, Apple, etc.
  - Participation in demand response to double in 2020

HPC is Energy-Costly!

- Worldwide investment on supercomputers
  - In 2016: $38 billion
- Supercomputer's lifelong energy cost almost equals investment cost
- Advent of Exascale
  - 20MW → $20 million/year for electricity



Source: "Total Cost of Ownership in High Performance Computing. HPC data center cost considerations: investment, operation and maintenance." in SoSE 2014

HPC in Demand Response

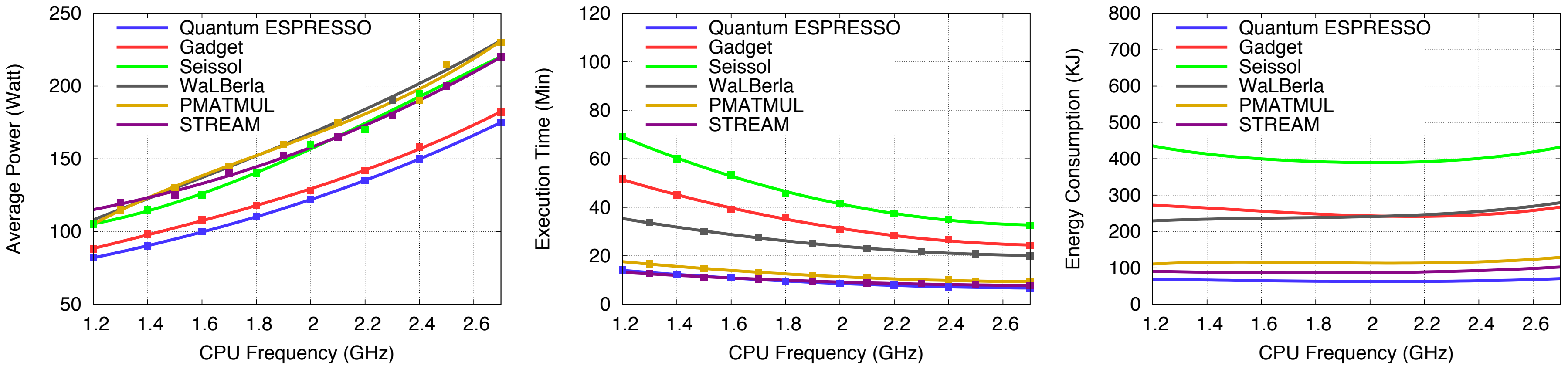
- Can HPC systems reduce the energy consumption and energy cost through emergenc*y* and economic demand response participation?
  - Supercomputers are willing to participate [Patki et al., 2016; Bates et al. 2015]
- Our solutions:
  - Emergency demand-response model
    - Application performance loss vs. energy reduction gain?
  - Economic demand-response model
    - How to incentivize HPC users for demand response participation?

Emergency DR Model

- Power/performance prediction model
  - Empirical data
  - Polynomial regression
- Demand response job scheduling
  - FCFS with possible job eviction (to ensure power limit)
- Resource provisioning
  - DVFS, power capping, node scaling

Power/Performance Prediction Model

Apply regression (quite a few alternatives) on power and execution time



$p(j,f) = a_j + b_j \cdot f + c_j \cdot f^2 + d_j \cdot f^3$

$t(j,f) = \alpha_j + \beta_j \cdot f + \gamma_j \cdot f^2$

$e(j,f) = n_j \cdot p(j,f) \cdot t(j,f)$

Job Scheduling

- During normal operation:
  - Traditional job scheduling
  - Optimized for best performance (max frequency)
- During demand response period:
  - Minimize energy for resource allocation
    - DVFS, power-capping, node scaling
  - Reduce power limit
    - May have to evict some jobs

Resource Allocation

- During normal operation
  - Run applications at maximum frequency for best performance
- During demand response: energy conservation

$$\text{Minimize: } \sum_{j \in R} e_R(j, f_j)$$

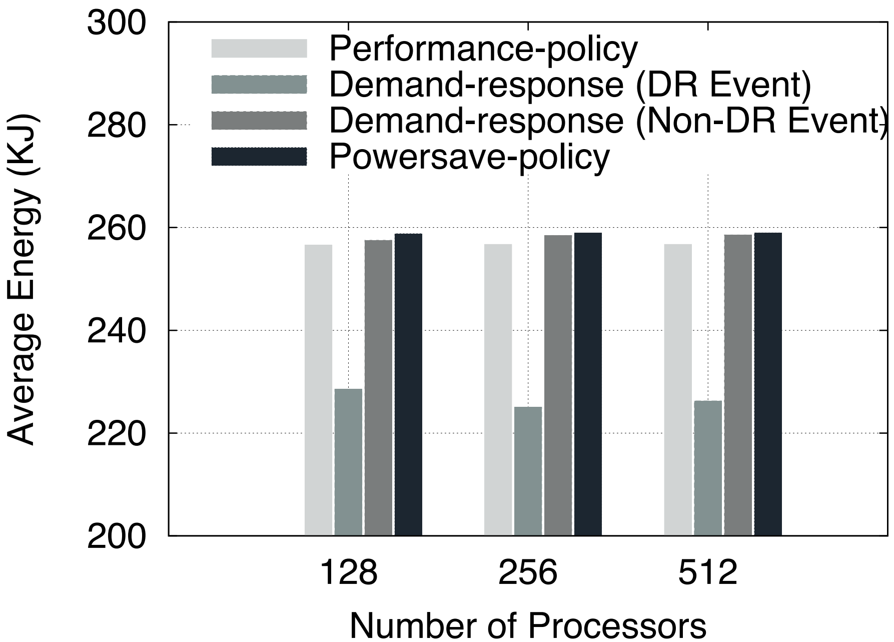
$$\text{subject to } \quad f_{min} \leq f_j \leq f_{max}$$

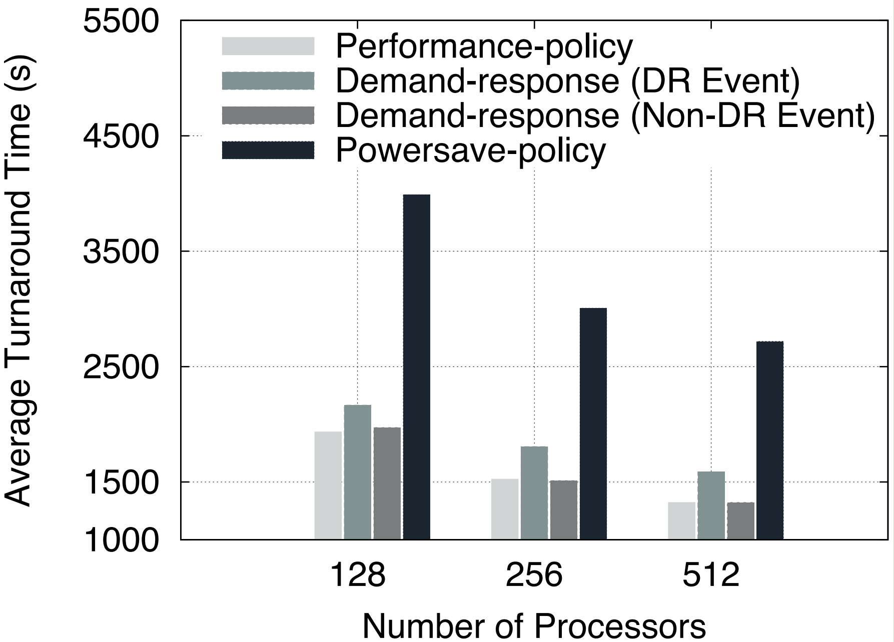
$$p_{run} = \sum_{j \in R} p(j, f_j) \leq \hat{p}$$

$$\text{where, } \quad e_R(j, f_j) = (1 - \alpha_j) \cdot n_j \cdot p(j, f_j) \cdot t(j, f_j)$$

Model Evaluation

Vary system size: 128, 256, and 512 processors





Reduced energy consumption at moderate increase in turnaround time

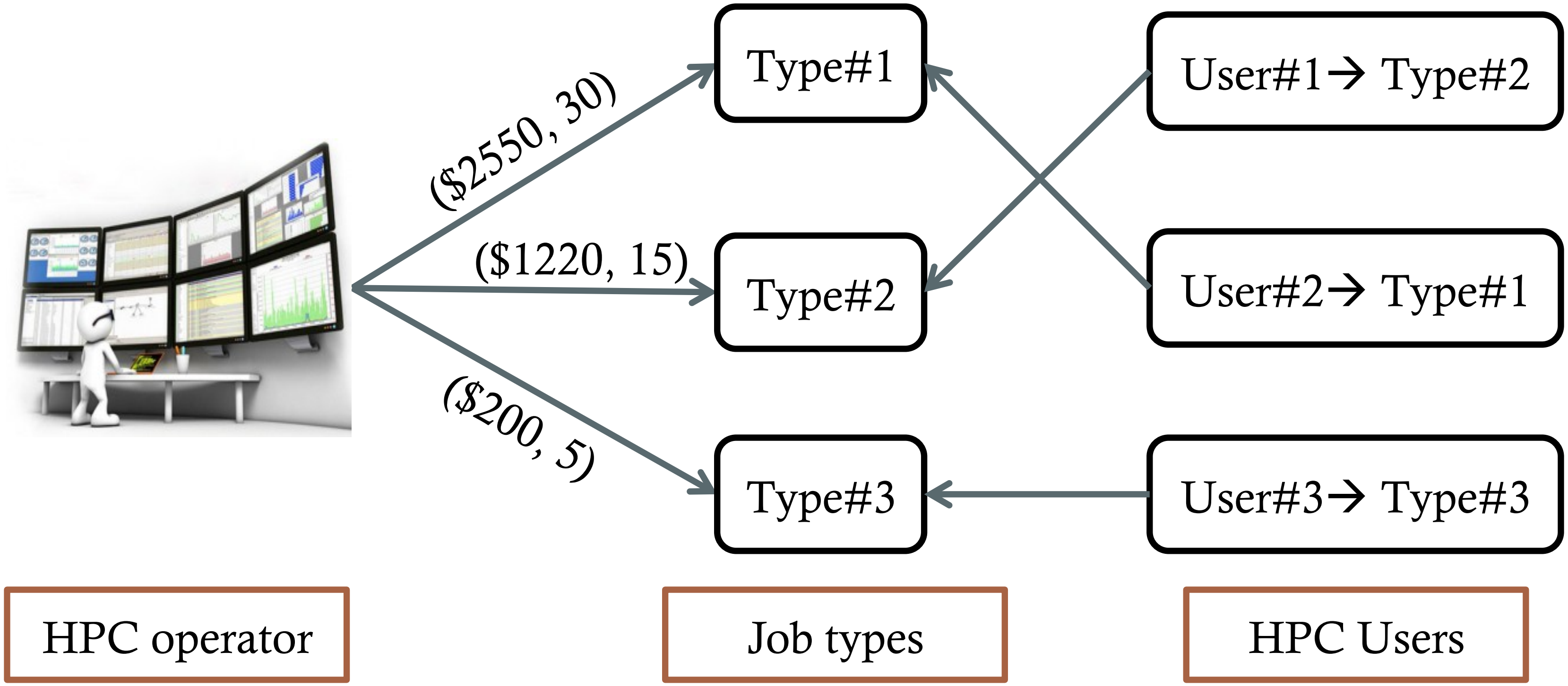
Economic DR Model

- Economic demand response
  - Voluntary participation based on economic incentives
- How to incentivize HPC users for participation?
  - Participation may introduce execution delays
  - Need a proper rewarding mechanism
- HPC economic DR model
  - A contract-based rewarding mechanism to incentivize HPC users' participations

Contract Theory

- A formal (economic) study to develop contracts between parties
  - Principal: who offers the contracts (HPC operator)
  - Agents: who are offered the contracts and can accept/reject (HPC users)
- Widely used in theory and practice
  - Economics (e.g., managerial compensation)
  - Communication (e.g., cellular network)

An High-Level Example



User’s utility maximized when selects own type’s contract

Resource Allocation

Maximize: $\sum_{i=1}^{n} m_i \cdot (\phi \cdot \gamma \cdot \Delta e_i - r_i)$

subject to, $f_{min} \leq f_i' \leq f_{max}$, IR, and IC constraints

Definition (Individual Rationality (IR) Constraint)

Participants in contract mechanism achieve non-negative pay-off or utility

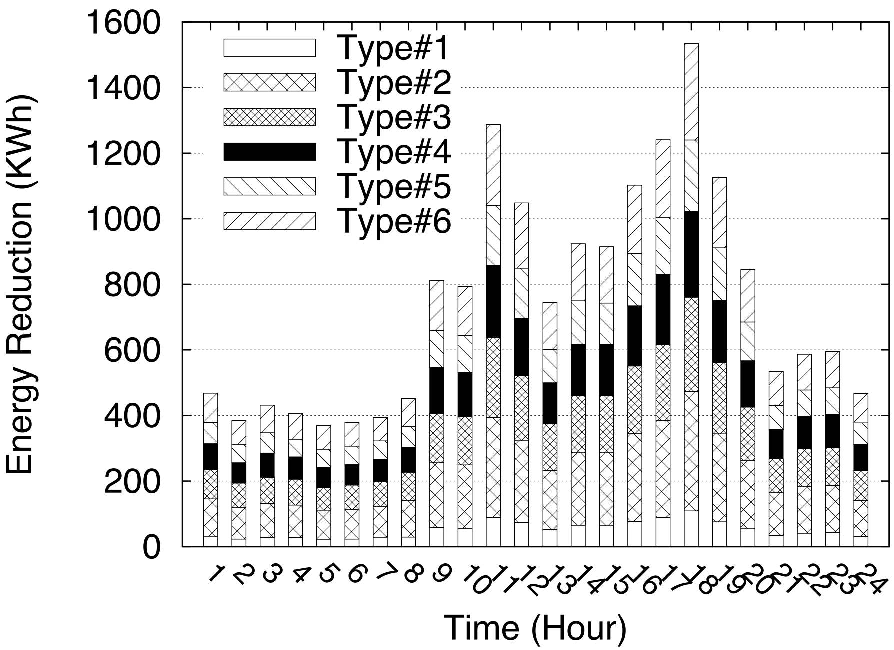
$$u_i = r_i - \theta_i \cdot c(\Delta t_i) \geq 0$$

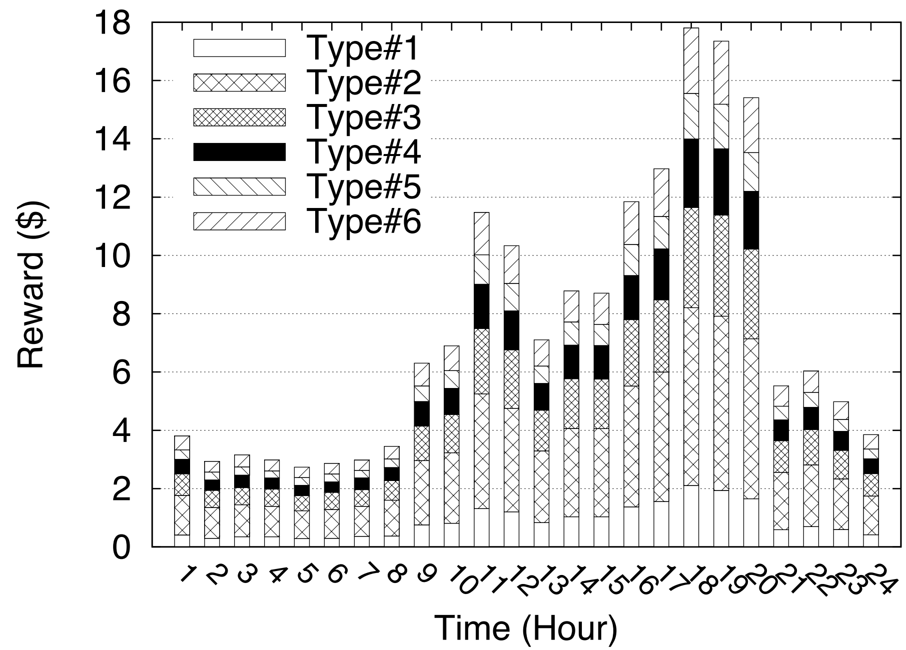
Definition (Incentive Compatibility (IC) Constraint)

Utility is maximized when participant chooses own contract type

$$r_i - \theta_i \cdot c(\Delta t_i) \geq r_{i'} - \theta_i \cdot c(\Delta t_{i'})$$

Energy and Reward





Energy reduction and rewards throughout entire time periods

Conclusions

- HPC demand-response models
  - Emergency demand response participation
  - Economic demand response participation
- A *win-win* situation to all:
  - HPC systems reduce energy cost
  - HPC users earn rewards
  - Power grid achieves energy reduction and power stability

Thanks!

Acknowledgements:

Kishwar Ahmed, Jason Liu, ModSim Workshop, August 2018