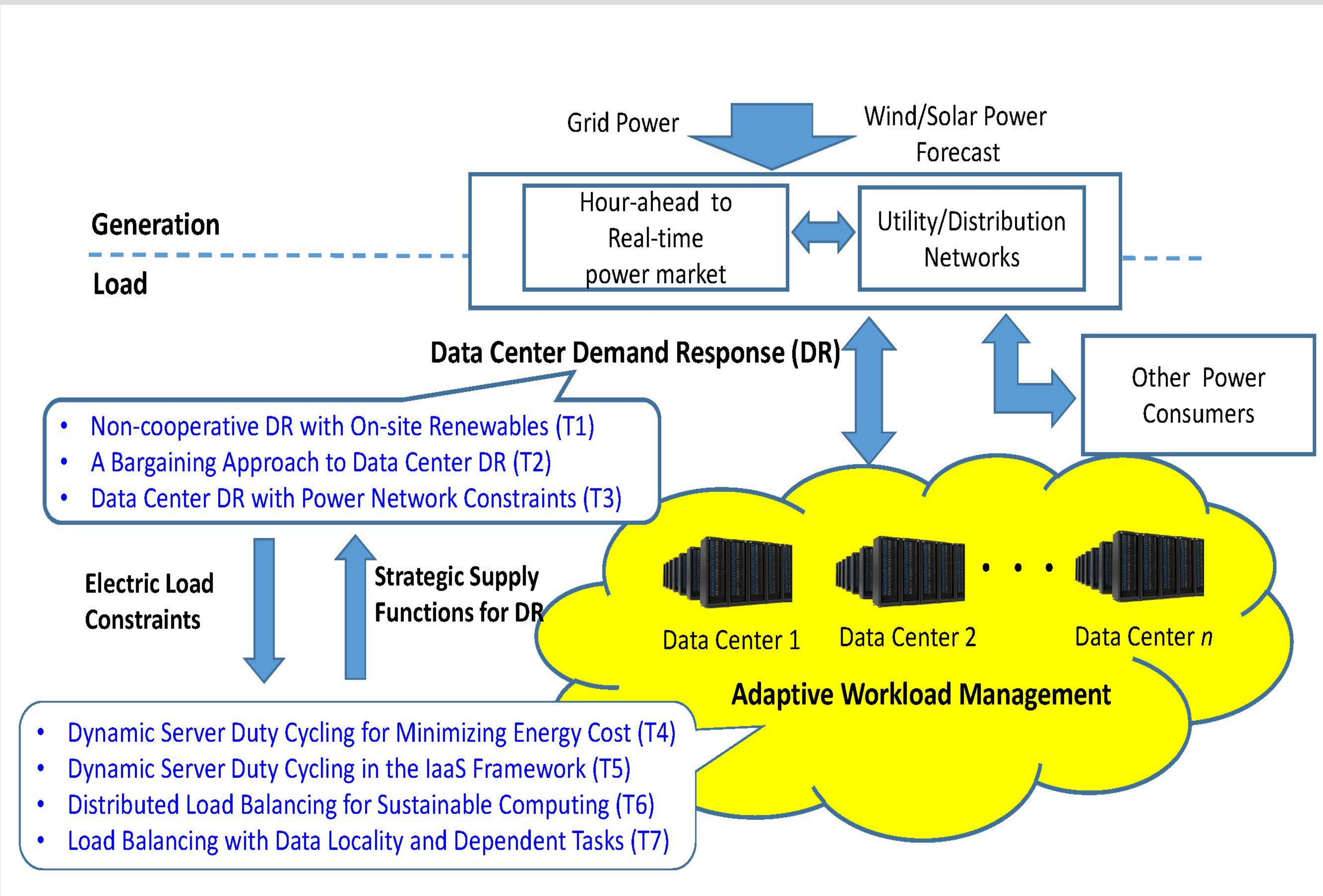
Demand Response and Workload Management for Data Centers with Increased Renewable Penetration

Junshan Zhang & Lei Ying, ASU; R. Srikant, UIUC; Steven Low, Caltech

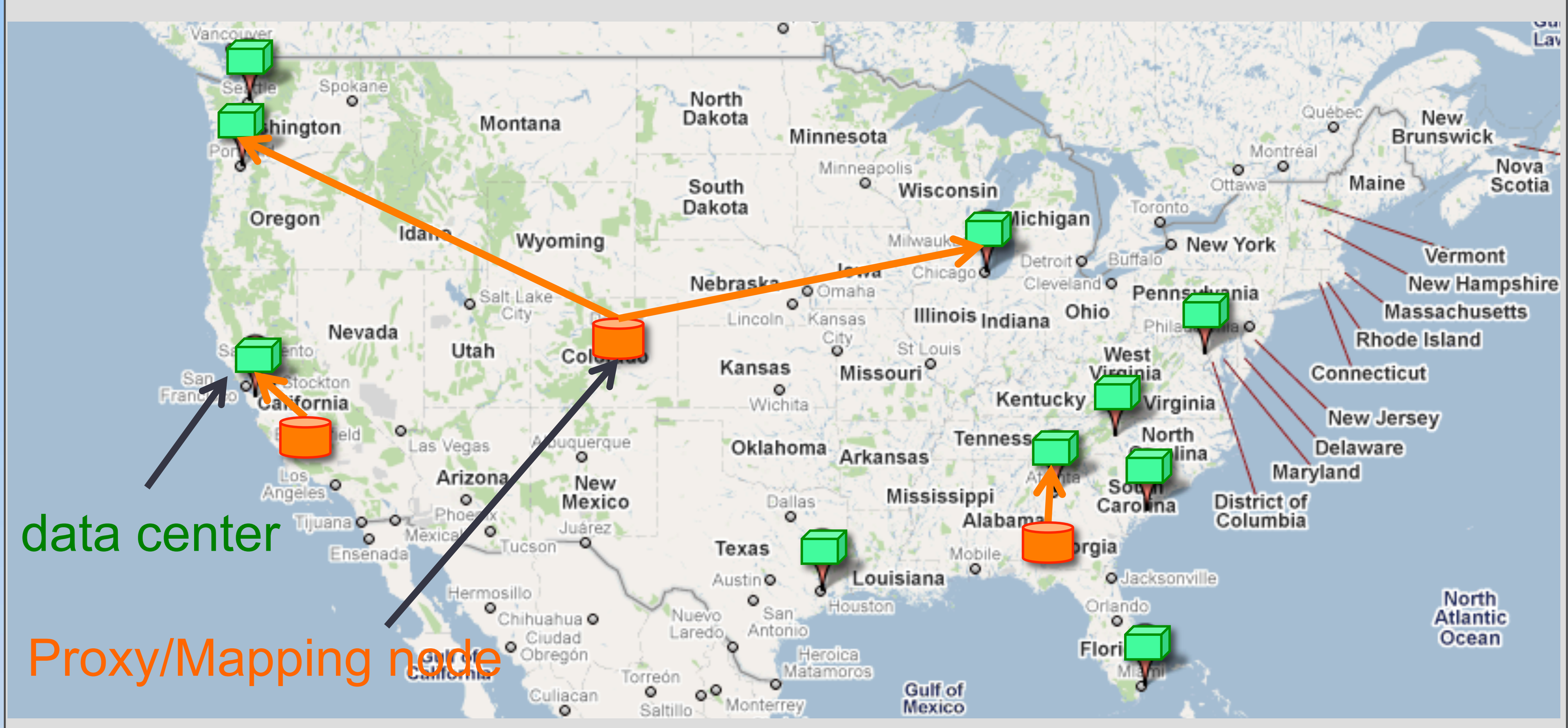
Strategic Demand Response and Workload Management towards Sustainable Data Center



Data Center DR with Power Network Constraints (Low)

Opportunities and challenges

1. Geographically distributed data centers can provide DR services by real-time load balancing across these data centers
2. Power flow constraints on the distribution grid constrain optimal DR decisions and introduce computational challenges



Problem formulation (optimal power flow)

$$\min_{V\in\mathbb{C}^n} \ \mathrm{tr}\left(CVV^H\right) \quad \text{min consumption cost}$$

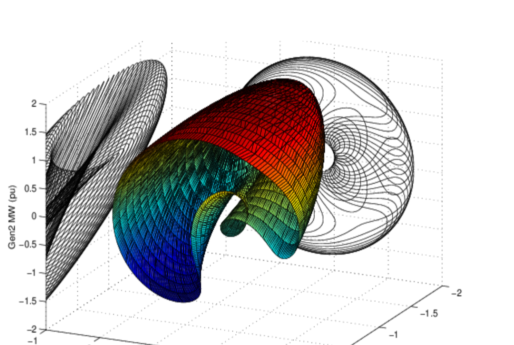
$$\text{s.t.} \quad s_j = \mathrm{tr}\left(Y_j^H VV^H\right) \quad \text{power flow equations}$$

$$\underline{v}_j \le |V_j|^2 \le \overline{v}_j, \quad \underline{s}_j \le s_j \le \overline{s}_j \quad \text{operational constraints}$$

nonconvex feasible set

- Y_j^H not Hermitian (nor positive semidefinite)
- C is positive semidefinite (and Hermitian)

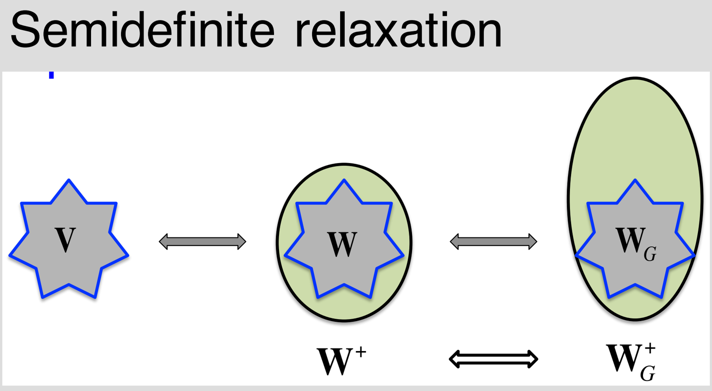
nonconvex QCQP



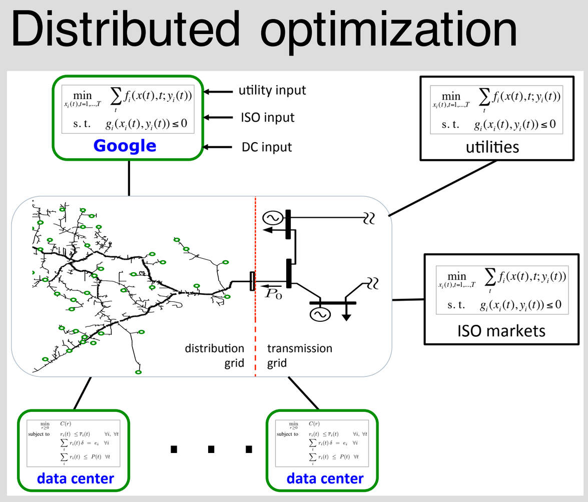
[Ian Hiskens]

Solution approach

Semidefinite relaxation



Distributed optimization



Strategic Participation of Data Centers (Zhang-Ying)

1) Non-cooperative Data Center Demand Response (DR); a Stackelberg game approach

data center n strategy: $\max_{d_n\ge 0} pd_n - C_n(d_n - R_n, W_n, p)$

utility n strategy: $\min_{p\ge 0} \mathbb{E}_{\{W_n,R_n\}}\{h(D - \sum_n d_n^*(W_n,p)) + p\sum_n d_n^*(W_n,p)\}$

2) Data Center Demand Response (DR); A Bargaining Approach

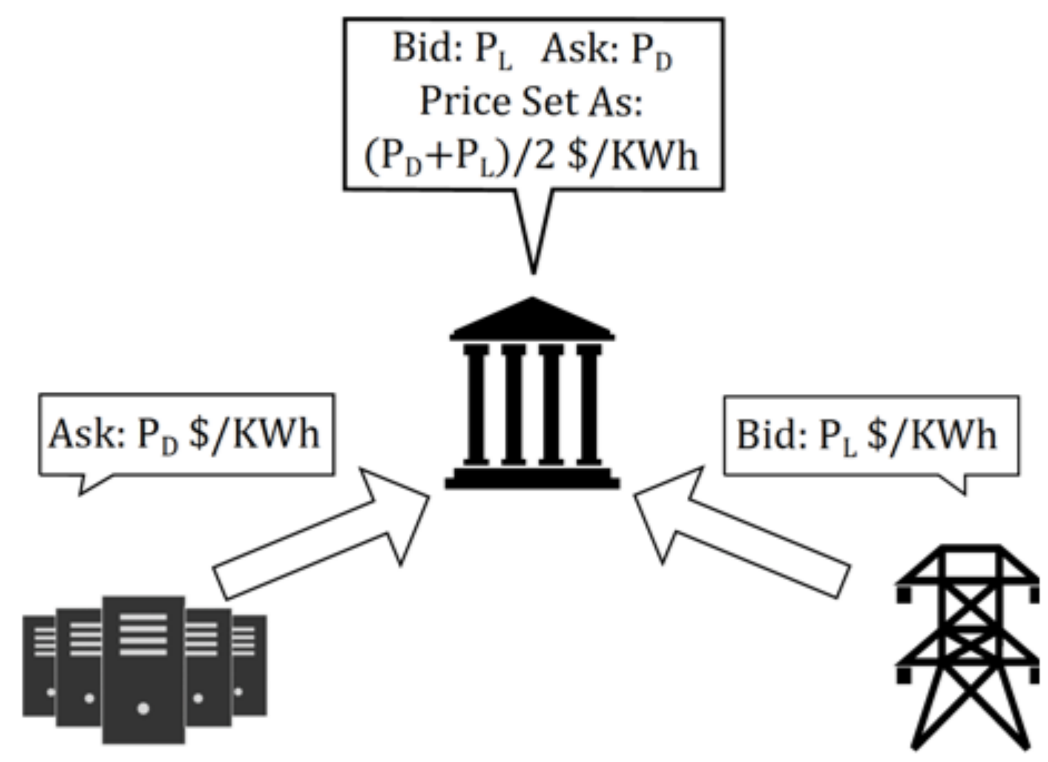


Figure: An illustration of the bargaining process

demand: $d(p) = \arg\max_q h(D) - h(D-q) - pq$

supply: $s(p) = \arg\max_q pq - C(q)$

The bargaining process

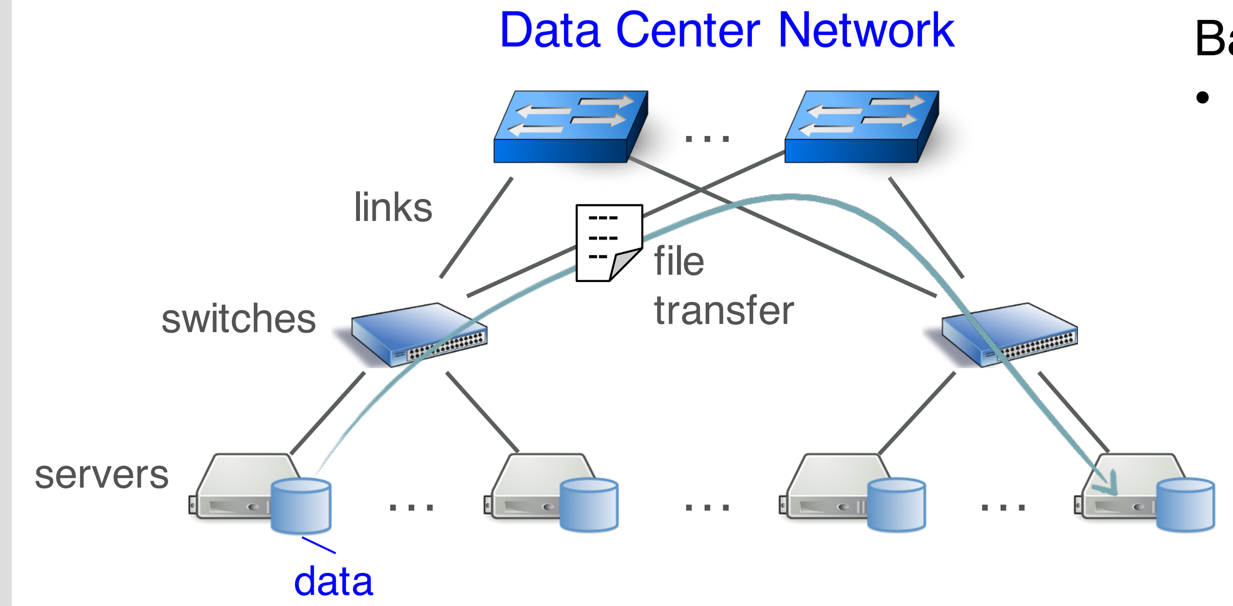
Algorithm 1 The bargaining process

Input: arbitrator gives an initial value $\epsilon \sim exp(\mu)$
Output: decide: bargaining succeed/fail
iteration: $k = 0$
  while not stable **do**
    if k=odd **then**
      utilities change bids p_L, in response to data-centers' asks
    else if k=even **then**
      data-centers change asks p_D, in response to utilities' bids
    $k++$;
  final spread: $\Delta p = p_L - p_D$
  if $\Delta p \ge \epsilon$ **then**
    return bargaining succeed
  else if $\Delta p < \epsilon$ **then**
    return bargaining fail

Delay–Power Tradeoffs in Data Centers (Srikant-Ying)

- Data centers need to deliver service quality guarantees on delays to end users
- Delay performance is a function of the utilization of a data center, which greatly affects the power consumption
- Delay analysis facilitates the planning of power generation based on the delay demand

Data Center Network

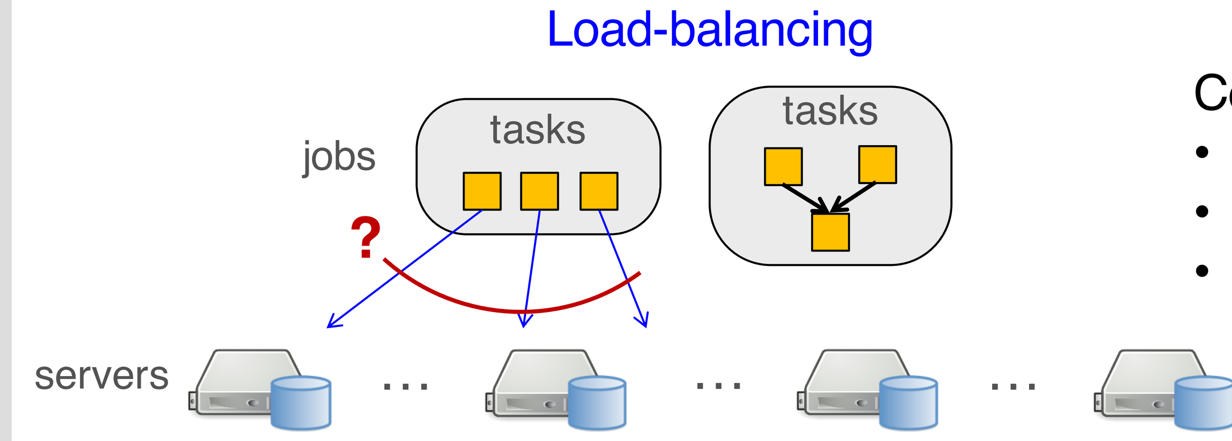


Bandwidth Allocation Policy

- The delay of data transfers is determined by the bandwidths allocated to them

$$\max_{(x_1,\ldots,x_R)} \sum_r n_r \log x_r$$
$$\text{subject to} \sum_{r:\ell\in r} n_r x_r \le C_\ell, \forall \ell,$$
$$x_r \ge 0, \forall r,$$

Load-balancing



Concerns

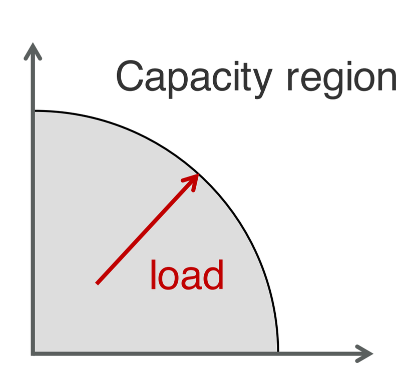
- Data locality
- Precedence relation among tasks
- Delay analysis on job level

Delay Analysis Approaches

- Exact analysis of delay is usually intractable

Heavy-traffic analysis

- System load approaches capacity limit
- Simplifications arise



Mean-field analysis

- Large number of servers in a system

Stochastic System → Deterministic Dynamical System