When Smart Grid Meets Geo-distributed Cloud: An Auction Approach to Datacenter Demand Response

Zhi Zhou¹, Fangming Liu¹, Zongpeng Li², Hai Jin¹

Email: fmliu@hust.edu.cn

¹Huazhong University of Science & Technology
²University of Calgary

April 30, 2015 @ IEEE INFOCOM, Hong Kong
Behind the Spotlight: What’s the Headache of Smart Grid?

- Smart grid technology: rising star in the ICT sector.

- Severe **operational stability issue**: Increase of peak power load, and penetration of intermittent distributed generation.

- Request for **demand response**.

[1] Data source: German power prices negative over weekend
Who can Response to Smart Grid?

- A variety of end users:
  - Factory
  - Building
  - Electrical Vehicle
  - Datacenter

- Datacenter power consumption: **huge yet flexible**
  - Make up 50% of the power load of a distribution grid;
  - Interactive workload, e.g., web search, online game, can be split to geo-distributed datacenters.
  - Elastic back-end batch workload: indexing, data-mining.

Datacenters are excellent candidates with great potential for smart grid demand response.

------Lawrence Berkeley National Laboratory (LBNL)
Possible to Buy Demand Response From Datacenters?

- Consumes a huge amount of energy and thus pays lots of bill\([1]\).

\[ \text{Google} \]

<table>
<thead>
<tr>
<th>Year</th>
<th>Energy Consumed (billion KWh)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>2.26</td>
</tr>
<tr>
<td>2011</td>
<td>2.68</td>
</tr>
</tbody>
</table>

- Monetary payment from DR program eases the high electricity bill.

Utilities and datacenters do not mix yet

---

\[1\] Data source: Google Green

—– The Green Grid, April 2014
Why does the Cloud Contributes Little to DR?

- Challenges and hurdles from both technical and economic aspects
  
  **Technical side:** ensuring the availability and desired performance for the risk-sensitive datacenters is being addressed.

  **Economic side:** Payment from existing demand response schemes are not enough. Designed for the stability of a single regional smart grid, price competition would incur a loss in social efficiency.
How to Solve the Inefficiency?

- Efficient market design to facilitate the cooperation between the smart grids and the geo-distributed cloud.

Energy efficiency programs need to place greater focus on marketing and outreach

—— The Green Grid, April 2014
What kind of Market?

- One seller: cloud, multi buyers: smart grids.

- Should be efficient: each smart grid really gets desired demand response without being over-charged.

- Should realize the true potential of the geo-distributed cloud, thus the smart grids should be truthful.

The answer can be: economic efficient and truthful auction.
Objective and Approach

- Economically efficient and truthful demand response auction tailored for geo-distributed datacenters to realize the true potential.

- Auction framework:
Cloud Model & Metrics

- Important performance metrics to characterize the datacenter

Front-end proxy servers:  \[ S = \{1, 2, \ldots, M\} \]

Geo-distributed datacenters:  \[ D = \{1, 2, \ldots, N\} \]

Geographical load Balancing:  \[ \sum_{j \in D} d_{ij} = D_i \]

Batch workload:  \[ \lambda_j \]

Datacenter energy demand:  \[ e_j = \alpha_j (\sum_{i \in S} d_{ij} + \lambda_j) + \beta_j \]

Electricity price:  \[ p_j \]

Capacity constraint:  \[ \sum_{i \in S} d_{ij} + \lambda_j \leq S_j \]

Propagation latency:

**Energy cost:**  \[ \sum_{j \in D} e_j p_j \]

**Dis-utility of interactive workload:**  \[ U_i(d_i) = q D_i \left( \frac{\sum_{j \in D} d_{ij} L_{ij}}{D_i} \right)^2 \]

**Revenue loss of batch workload:**  \[ V_j(\lambda_j) = \theta(S_j - \lambda_j) \]
Demand Response Bidding Model

- Key assumptions
  - One datacenter locates within the geographical span of a regional smart grid;
  
- Each smart grid submits a number of bids to express its willingness to pay for different levels of demand response.

\[
(e_j, b_j) \in \{(e_j^1, b_j^1), (e_j^2, b_j^2), \ldots, (e_j^K, b_j^K)\}
\]

Winning bid Submitted bids

Each smart grid wins one and only one bid!

Great departure from conventional auctions such spectrum auction and VM auction.
Demand Response Bidding Model

- The objective of the auction is to maximize the social welfare, i.e., the aggregated utility of both the cloud and smart grids.

- Utility of each smart grid:

\[ u_j = b_j - r_j \]

- Utility of the geo-distributed cloud participating in the demand response auction:

\[
\sum_{j \in D} \left\{ r_j - V_j(\lambda_j) - e_j p_j \right\} - \sum_{i \in S} U_i(d_i)
\]

The second difference: the auctioneer’s utility comprises not only payments from the bidders, but also the varying disutility, revenue loss and cost that depend on the allocation of the datacenter power consumption.
Winner Determination Problem

The winner determination problem is the one that determines the winning bid of each smart grid, by maximizing the social welfare.

\[
\begin{align*}
\max & \quad \sum_{j \in \mathcal{D}} \left\{ V_j(\lambda_j) - e_j p_j + b_j \right\} - \sum_{i \in \mathcal{S}} U_i(d_i), \\
\text{s.t.} & \quad \sum_{j \in \mathcal{D}} d_{ij} = D_i, \forall i \in \mathcal{S}, \\
& \quad \sum_{i \in \mathcal{S}} d_{ij} + \lambda_j \leq S_j, \forall j \in \mathcal{D}, \\
& \quad \alpha_j \left( \sum_{i \in \mathcal{S}} d_{ij} + \lambda_j \right) + \beta_j = e_j, \forall j \in \mathcal{D}, \\
& \quad (e_j, b_j) \in \left\{ (e_j^1, b_j^1), \ldots, (e_j^K, b_j^K) \right\}, \forall j \in \mathcal{D}, \\
& \quad d_{ij} \geq 0, \lambda_i \geq 0.
\end{align*}
\]

Due to the silent features, existing fruitful solutions, e.g., randomized auction, monotonic allocation are not directly applicable.
Can We Leverage the Distributed Computing Facilities?

- The cloud has enormous computing capacities distributed across its datacenters and front-end servers.

The answer is YES!
DORA: Decentralized Social welfare maximization

- Global Searching over all the bids is computational prohibitive.
- DORA: in a decentralized manner, each data-center autonomously decide the winning bid of its corresponding regional smart grid, using the Gibbs sampling technique.

1. **STEP 1**: each datacenter initials the winning bid.
2. **STEP 2**: obtain the workload management solution under the selected bids.
3. **STEP 3**: compute the parameterized transition probability
   
   \[
   Pr = \frac{\exp(T \times W)}{\exp(T \times W) + \exp(T \times W^*)}
   \]

4. **STEP 4**: with a probability or Pr to keep the current solution, and a probability of 1-Pr to explore new solution.
5. **STEP 5**: randomly select a datacenter to update its selected bid.
6. **STEP 6**: return to **STEP 2** until the stopping criteria is met.
Performance Analysis of DORA

- DORA does not always choose the current best decision.

- DORA explores new solution by introducing randomness, in order to avoid trapping into a local optimal. The degree of randomness is controlled by the tunable smoothing parameter $T > 0$.

- Parameter $T$ controls the tradeoff between economical efficiency and computational complexity.

- Performance:

  **Theorem 1:** As $T$ increases, the proposed algorithm DORA converges to the global optimal social welfare with a higher probability. When $T \to \infty$, the algorithm converges to the global optimal social welfare with probability 1.
Can We Parallelize the Computation of Workload Management?

- Convex Workload Management Problem.

\[
\begin{align*}
\min & \quad \sum_{i \in S} U_i(d_i) - \sum_{j \in D} V_j(\lambda_j), \\
\text{s.t.} & \quad \forall i : \sum_{j \in D} d_{ij} = D_i, \\
& \quad \forall j : \sum_{i \in S} d_{ij} + \lambda_j = w_j, \\
& \quad d_{ij} \geq 0, \lambda_i \geq 0,
\end{align*}
\]

where \( w_j = \frac{e_j^* - \beta_j}{\alpha_j} \).

- Need to be solved many and many times.

- Parallelizing the computation of workload management for efficient winner determination.
Distributed Workload Management — The ADMM Method

- Solves linearly constrained convex problems whose objective function is separable into two individual convex functions.

\[
\begin{align*}
\min & \quad f(x) + g(z) \\
\text{s.t.} & \quad Ax + Bz = c, \\
& \quad x \in C_1, z \in C_2,
\end{align*}
\]

- A popular approach for convex optimization is to form the augmented Lagrangian.

\[
\mathcal{L}_\rho(x, z, y) = f(x) + g(z) + y^T(Ax + Bz - c) + \frac{\rho}{2}\|Ax + Bz - c\|_2^2,
\]

- Update each block of variables in an alternating manner.

\[
\begin{align*}
x^{t+1} &= \arg\min_{x \in C_1} \mathcal{L}_\rho(x, z^t, y^t), \\
z^{t+1} &= \arg\min_{z \in C_2} \mathcal{L}_\rho(x^{t+1}, z, y^t), \\
y^{t+1} &= y^t + \rho(Ax^{t+1} + Bz^{t+1} - c).
\end{align*}
\]
Is ADMM Directly Applicable?

Directly applying the ADMM results in a centralized algorithm with high complexity, since the coupling of $d_{ij}$ happens on two orthogonal dimensions simultaneously.

\[
U_i(d_i) = qD_i \left( \frac{\sum_{j \in D} d_{ij} L_{ij}}{D_i} \right)^2
\]

\[
\sum_{j \in D} \left( \sum_{i \in S} d_{ij} + \lambda_j - w_j \right)^2
\]

For decomposition, we introduce a block of auxiliary variables $a_{ij} = d_{ij}$, and replace $\lambda_j$ with $w_j - \sum_i a_{ij}$.

\[
\text{min} \quad \sum_{i \in S} U_i(d_i) + \sum_{j \in D} V_j \left( w_j - \sum_{i \in S} a_{ij} \right)
\]

s.t. \quad \sum_{j \in D} d_{ij} = D_i, \forall i \in S,

\[
d_{ij} = a_{ij} \geq 0, \forall i \in S, \forall j \in D.
\]

Fully distributed algorithm, each facility need only to solve a small-scale convex problem.
Payment Design: VCG or Core Selection?

- To increase one's own utility, a bidder may cheat, i.e., misreporting its valuation. This would harm the social welfare.

- To maximize the real social welfare, the payment should make the dominant strategy of each bidder is to truthfully report its valuation.

- Core selection: the seller’s utility is composed of all bidder’s payment.

- VCG: demand response auction is shielded from shill bidding. Each smart grid is unable to impersonate multiple bidders in the auction.
Real-world Trace-driven Evaluation

- **Setup:**
  - Google’s 6 datacenter in US, 10 front-end servers with HP workload trace

Electricity price: the 2011 annual average day-ahead on-peak prices at the local markets

Valuation function:

\[ b_j(e_j) = \max[3000 - 120(\hat{E}_j - e_j)^2, 0] \]
Examination of social welfare and cloud utility

- Remarkable improvement
- When the interactive workload bursts, the improvement is relatively small.
Examination of smart grid utility and demand deficit

- Utility of the smart grids may not improved, since no auction no payment.
- Demand deficit is significantly decreased, and can be near to zero.
Effects of maximal bidding price and scaling ration of workload

- Demand deficit diminishes much faster as the maximal bidding price increase.
- Demand deficit increases as the scaling ratio grows.
Conclusion

- **When Smart Grid Meets Geo-distributed Cloud: An Auction Approach to Datacenter Demand Response**

  - **First effort** on market design for the demand response from the geo-distributed cloud

  - **An auction framework** with two salient features: each bidder wins one and only one bid, and the auctioneers utility compromises not only payments from the bidders, but also the dis-utility, revenue loss and energy cost

  - **Decentralized social welfare maximization** by leveraging the distributed computing facilities, based on Gibbs Sampling and ADMM method.
Q&A

Thank You!

"Cloud Datacenter & Green Computing” Research Group
Huazhong University of Science & Technology

http://grid.hust.edu.cn/fmliu/
fmliu@hust.edu.cn