Impact of workload and renewable prediction on the value of geographical workload management

Zahra Abbasi, Madhurima Pore, and Sandeep Gupta

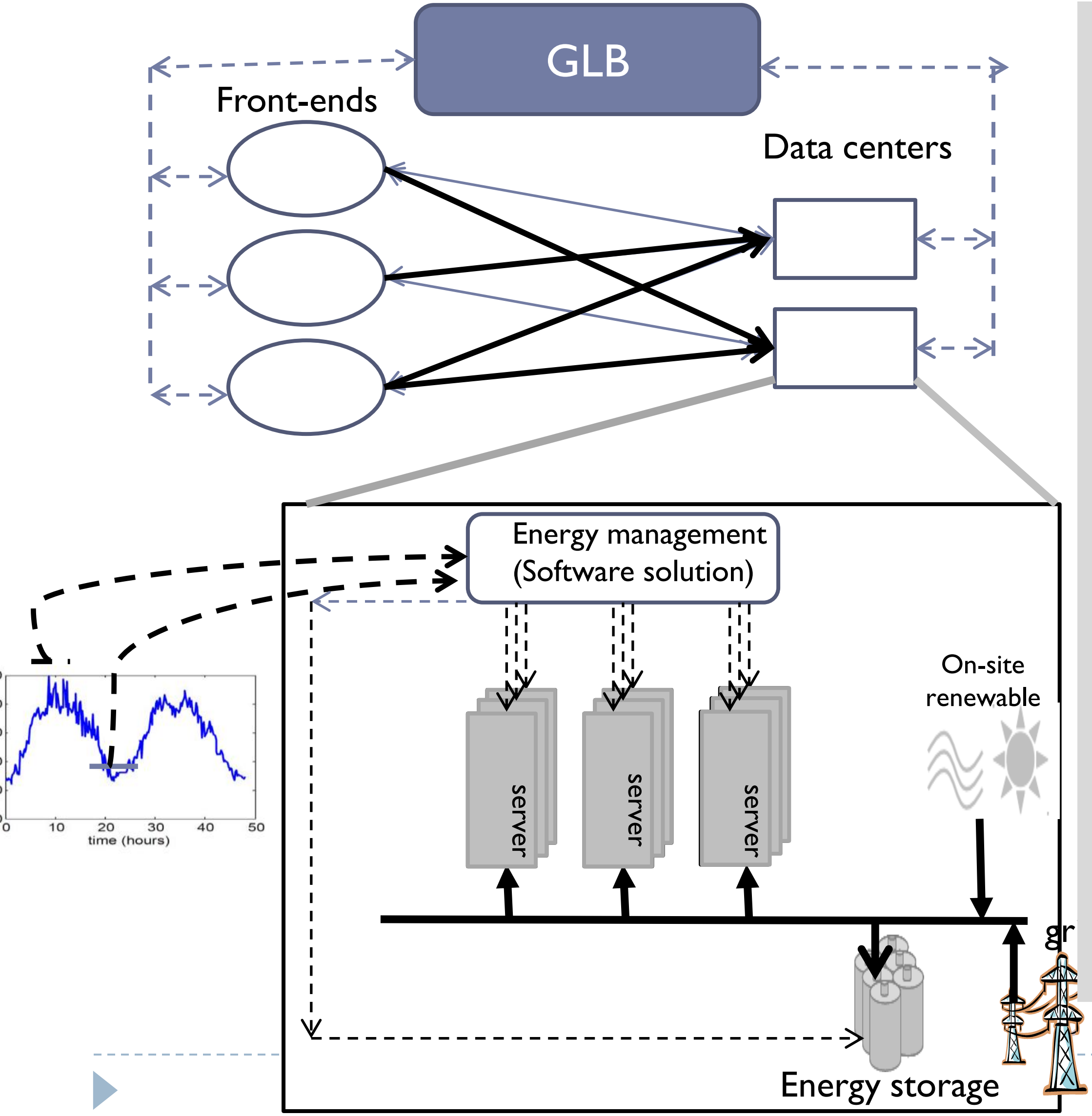
Arizona State University

Funded in parts by **NSF CNS gran** and by **Intel Corp**

Agenda

- Introduction
  - Geographical/Global Load Balancing (GLB)
  - Existing work review
- GLB Problem statement
- Trace based simulation results
  - Time-series based workload and renewable prediction
  - Impact of prediction error on the performance of GLB
- Conclusion

Geographical Load Balancing for geo-distributed Data Centers



- **Examples**
  - CDN
  - Large-scale and delay-sensitive Internet Applications (e.g., Google search)

  (Performance and fault-tolerance)

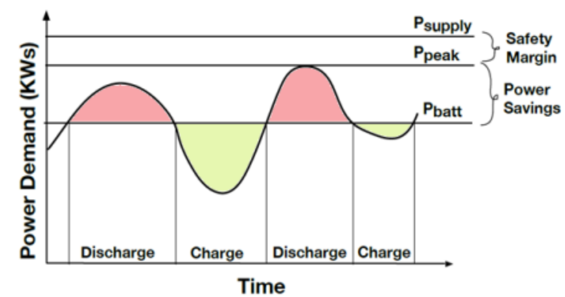
- **Energy and cost management**
  - Spatio-temporal variation of
    - **Electricity cost**
    - **Available renewable energy**
    - **Energy efficiency**
  - Peak power shaving
    - Balancing power load

- **GLB existing solutions**
  - Load balancing across data centers
  - Power demand and supply management

Overview of our study

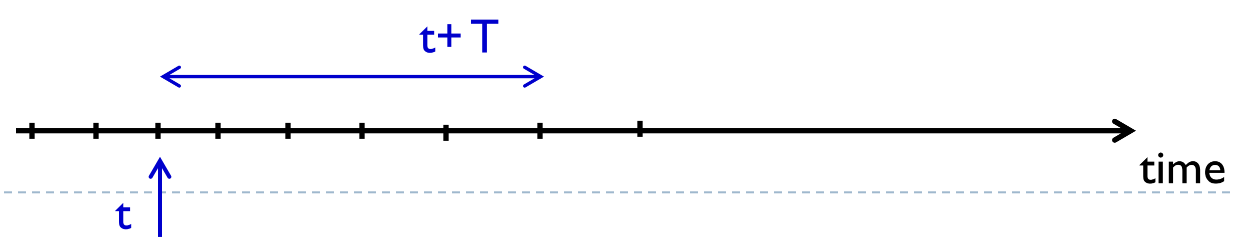
- GLB to incorporate energy buffering and the renewable energy management to
  - Maximize renewable utilization, minimize energy cost and minimize peak power cost
  - Energy buffering to balance power

  draw from grid inside data centers

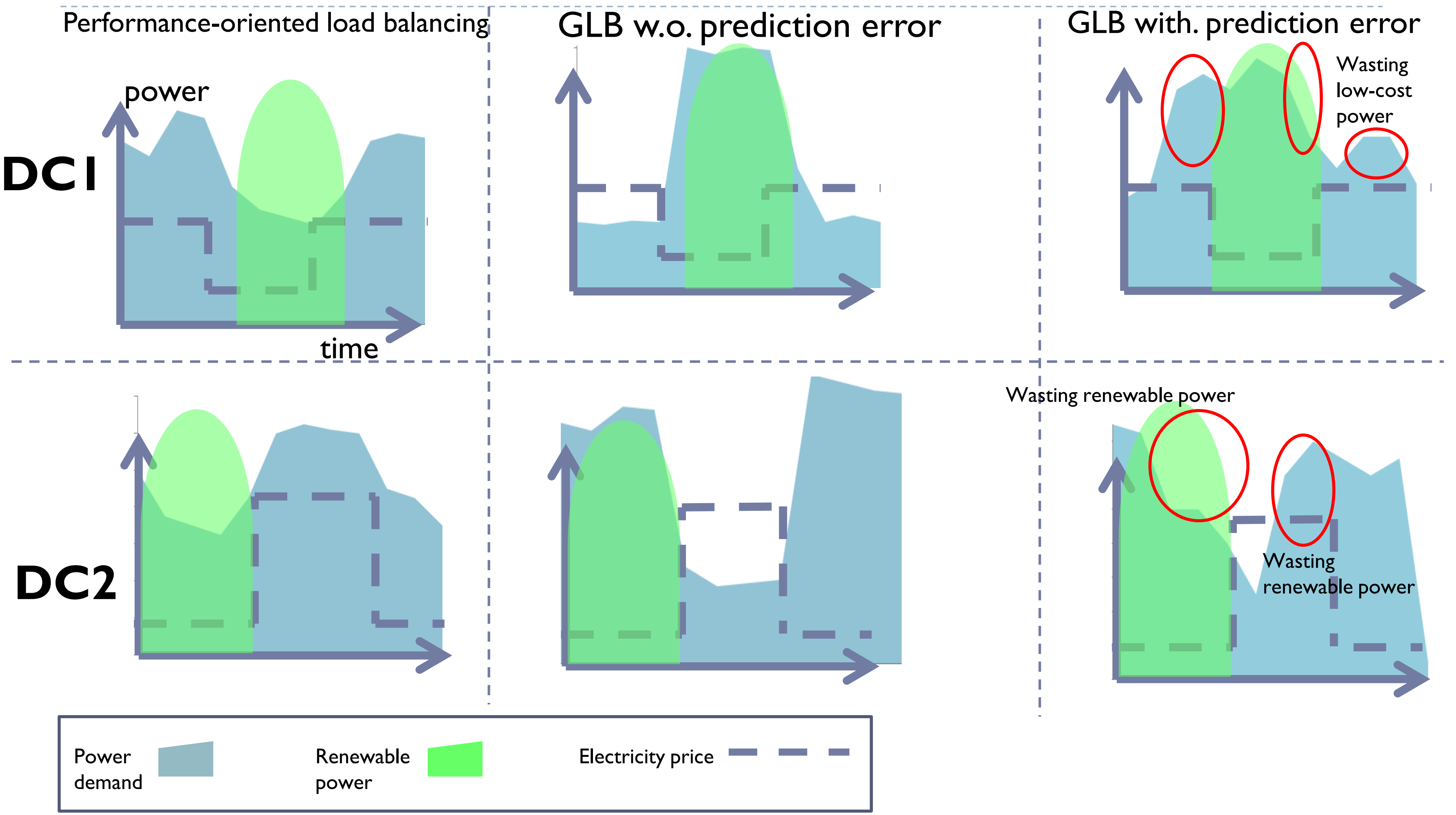


Ref: Palasamudram, Darshan S., et al. "Using batteries to reduce the power costs of internet-scale distributed networks." *Proceedings of the Third ACM Symposium on Cloud Computing*. ACM, 2012.

- Impact of prediction error on the performance of GLB
  - Efficient GLB management requires prediction over a long time horizon e.g., shaving peak power and managing server switching cost



GLB-Example



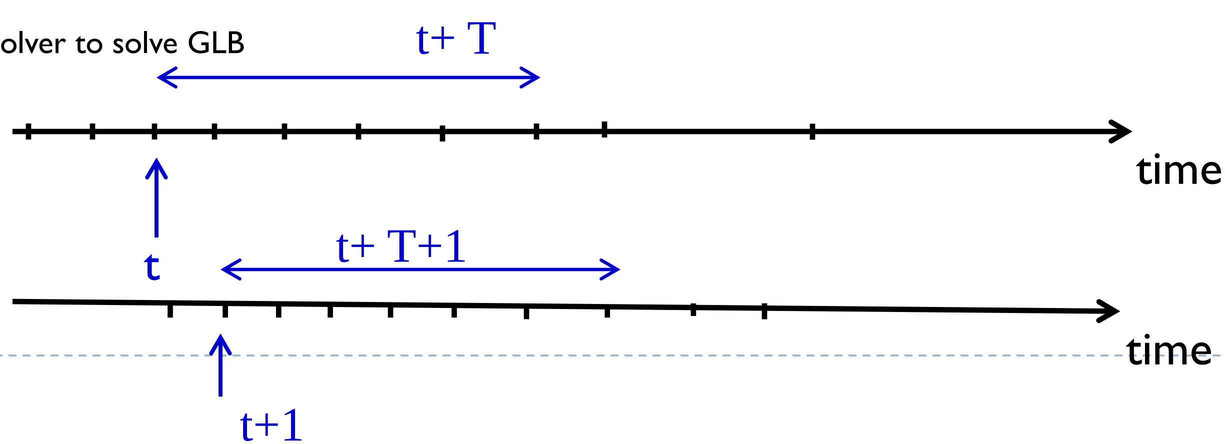
Do simple time-series based prediction techniques provide enough prediction accuracy for GLB to efficiently manage renewable energy, energy cost and peak power cost?

GLB solution

- **GLB cost model: Linear programming optimization**
  - **Find:** workload split and power source model (renewable, grid and energy storage) of data centers
  - **Minimize:** energy cost + peak power cost+ energy storage cost
  - **Subject to:**
    - Energy storage physical limitations
    - Delay constraint of applications
    - Data Centers' capacity
    - Power demand-supply matching and the available renewable energy

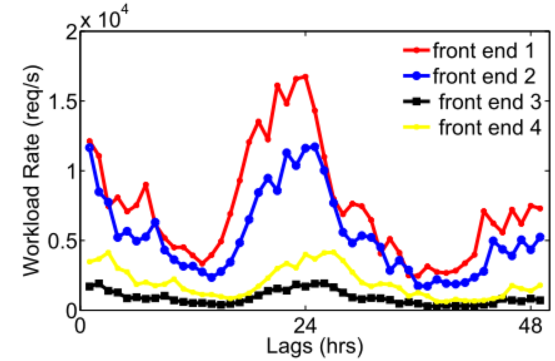
- **Using Rolling Horizon Control (RHC) to solve GLB online at each time t given inputs during t+T**

Use LP solver to solve GLB at time t



Trace based simulation study

- Four front-ends corresponding to 4 time zones in US
  - Publicly available internet traces (NASA workload) shifted in time and scaled to the number of Internet users at each time zone



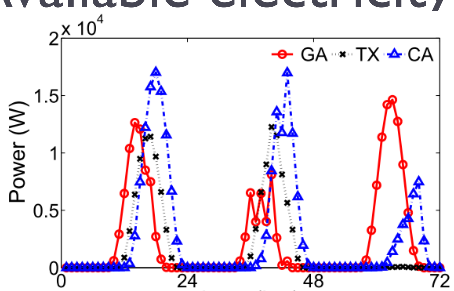
Ref: http://ita.ee.lbl.gov/html/traces.html

- Three data centers

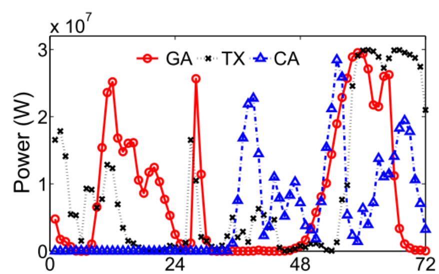
DC	Elec. price model	peak power(W)
DC1	Mountain View, CA.	395
DC2	Houston, TX	300
DC3	Atlanta, GA	450

- Available traces of solar and wind energy in nearby sites
- Available electricity price traces

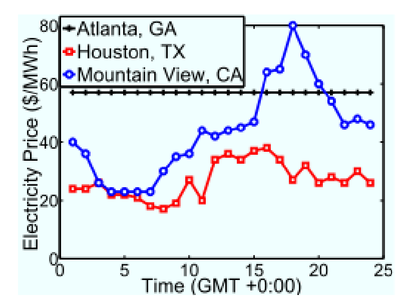
Solar power



Wind power



Ref: http://www.nrel.gov/midc/



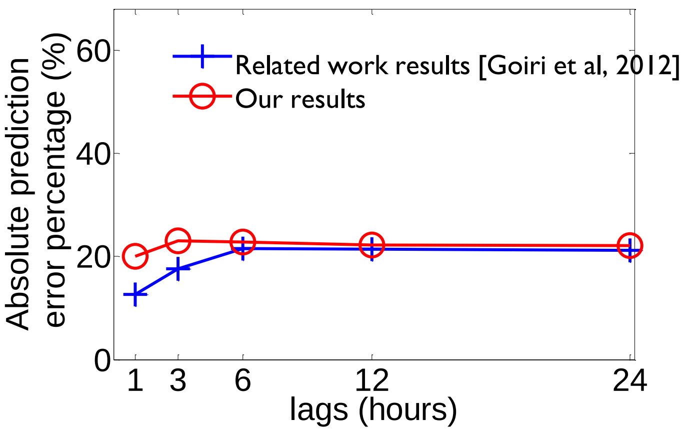
Ref: www.caiso.com, www.ercot.com

Time series based prediction models and results

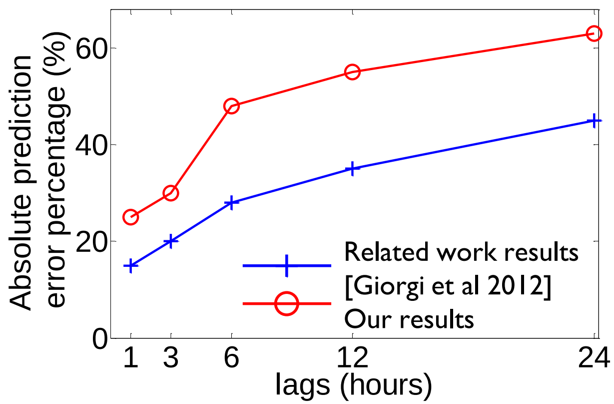
Building prediction models: One month training and one month testing

Solar energy prediction: SARIMA

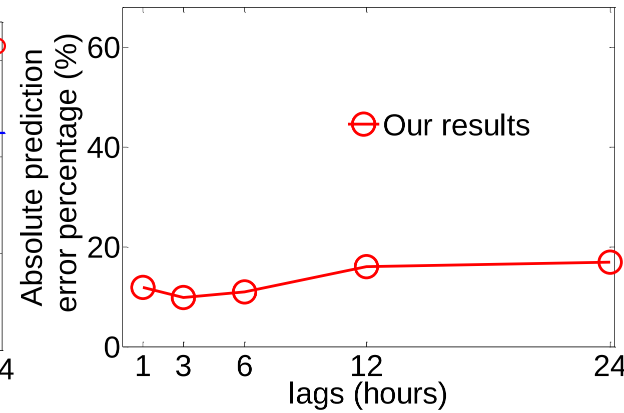
Seasonal Autoregressive Integrated moving average



Wind energy: Moving average



Workload arrival rate: SARIMA



Related work: prediction error for lag one [Zhenhua et al. 2012]: 15%

Goiri, Ínigo, Kien Le, Thu D. Nguyen, Jordi Guitart, Jordi Torres, and Ricardo Bianchini. "GreenHadoop: leveraging green energy in data-processing frameworks." In *Proceedings of the 7th ACM european conference on Computer Systems*, pp. 57-70. ACM, 2012.

De Giorgi, Maria Grazia, Antonio Ficarella, and Marco Tarantino. "Error analysis of short term wind power prediction models." *Applied Energy* 88.4 (2011): 1298-1311.

Liu, Zhenhua, et al. "Renewable and cooling aware workload management for sustainable data centers." *ACM SIGMETRICS Performance Evaluation Review*. Vol. 40. No. 1. ACM, 2012.

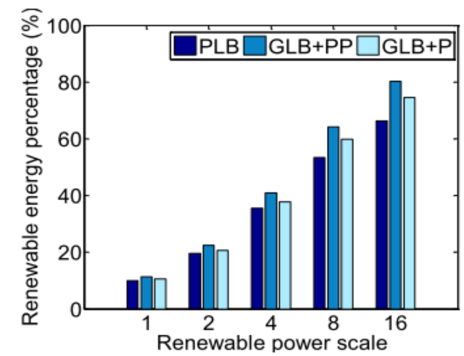
Simulation study results

- Workload management algorithms
  - Performance oriented local load balancing + perfect prediction results (zero prediction error): PLB
  - GLB + perfect prediction results (zero prediction error): GLB+PP
  - GLB + prediction data (non-zero prediction error): GLB+P

Renewable energy utilization and energy cost

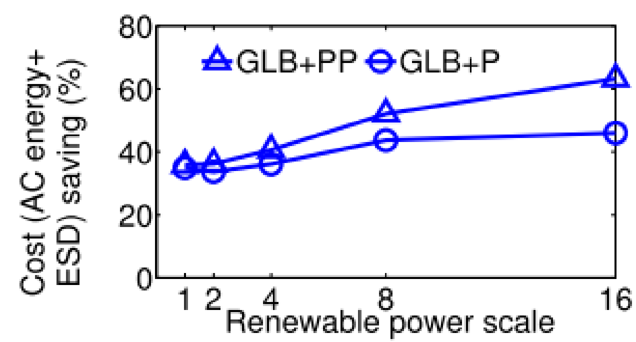
Percentage of renewable energy used out of total energy demand in data centers

- GLB w/w.o. prediction error increases the utilization of the available renewable energy
  - GLB schemes under the same magnitude of the available renewable energy and the same size of energy storage capacity increase the renewable utilization

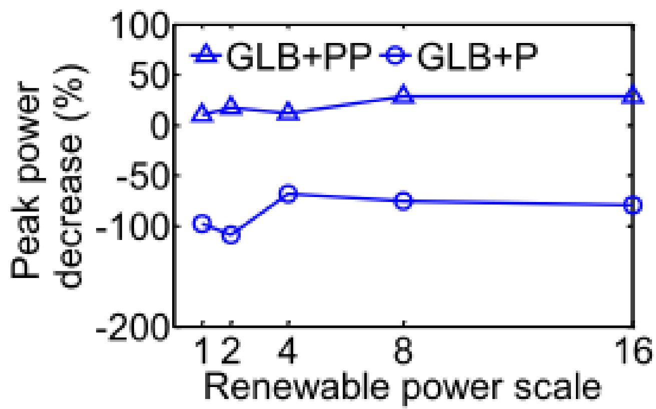


i.e., Increasing the available renewable energy

- GLB w/w.o. prediction error saves energy cost
  - Cost saving comes from increasing the utilization of the available renewable energy and low-cost electricity power



Peak power cost



Peak power peak cost saving of GLB+PP and GLB+P compared to PLB

- GLB w.o. prediction error is very effective in reducing peak power cost
- Prediction error is an adverse multiplicative factor for shaving peak power
- Pessimistic results since PLB is assumed to have zero prediction error

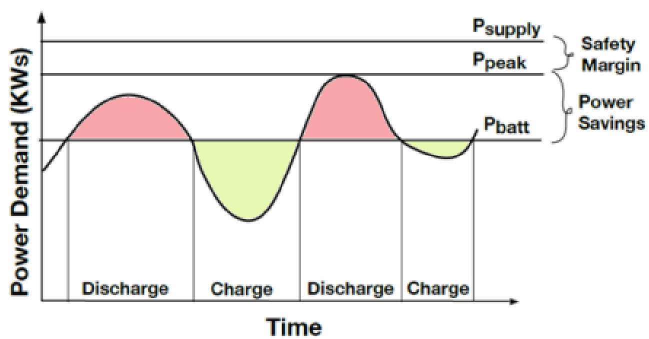
Summary of results

- Geographical (Global) workload management for geographically distributed data centers:
  - Is efficient in reducing energy cost and increasing renewable energy utilization without large scale energy storage devices given existing time-series based prediction techniques for workload and renewable energy
  - Is potentially very efficient in shaving peak power draw
  - Needs prediction techniques with low error to efficiently shave peak power

Questions?

GLB existing solutions review

- Designing efficient GLB algorithms and trace-based studies to:
  - Reduce average energy cost of DCs
  - Increase renewable energy utilization
  - Reduce energy storage size requirements
  - Potentially reduce peak power cost
    - Energy buffering to balance power draw from grid inside data centers



Ref: Palasamudram, Darshan S., et al. "Using batteries to reduce the power costs of internet-scale distributed networks." *Proceedings of the Third ACM Symposium on Cloud Computing*. ACM, 2012.

GLB Problem Statement

- Given:
  - set of data centers S with the following data over t+T
    - Electrify price:
    - Cost per excess peak power form stipulated power over a long time (e.g., month)
    - Available renewable power: $r_{i,t}^{total}$
    - State of energy storage: $B_{i,t}$
  - and set of front-ends A with workload arrival rate of $\lambda_{j,t}$
- At each slot t find:
  - Workload split of each front-ends : $\lambda_{i,j,t}$
  - Number of active servers at each DC: $y_{i,t}$
  - Power draw from grid: $p^{AC}_{i,t}$
  - Power draw from the available renewable power: $r_{i,t}$
  - Charging/Discharging from energy storage: $p^{ESD}_{i,t}$

To minimize total cost of data centers and maintain delay requirement of the applications

GLP cost model
Linear programming model

Minimize $\sum_{i=1}^{|S|}\left(\sum_{t=1}^{T}\left(p_{i,t}^{AC}\tau\alpha_{i,t}+b_{i,t}\gamma_i\right)+\max_{1\le t\le T}(p_{i,t}^{AC}-p_0)^+\beta_i'\right)$

Electricity cost (Renewable energy cost is set to zero) | Battery cost) | Peak power cost: (excess power from stipulated power)

Subject to:

Energy storage device constraints:

$$B_{i,t+1}=\min(\eta_1(B_{i,t}+\rho_1 p_{i,t}^{ESD}\tau),B_i^{max}),$$

Max charging/discharging rates
Energy storage charging/discharging cycle

Power supply Constraints:

$$r_{i,t}+p_{i,t}^{AC}=p_{i,t}^{total}+p_{i,t}^{ESD}$$

$$r_{i,t}\le r_{i,t}^{total} \qquad p_{i,t}^{total}=p_i y_i.$$

Service constraints:

$$\sum_{i=1}^{|S|}\lambda_{i,j,t}=\lambda_{j,t}$$

Performance constraints:

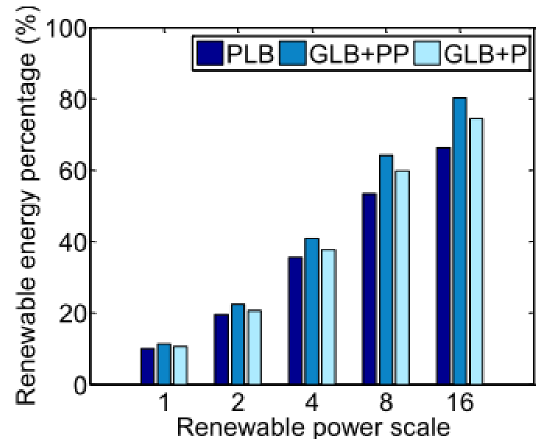
$$(d^{\text{ref}}-d_{i,j,t})\lambda_{i,j,t}\ge 0.$$

$d_{i,j,t}$: data center delay+network delay

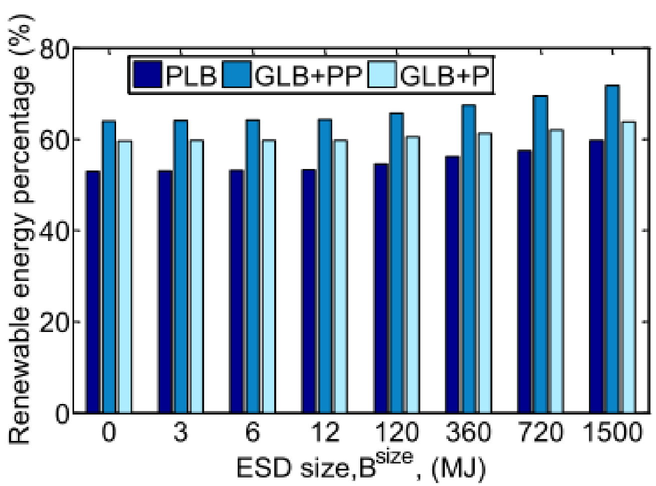
Renewable energy utilization

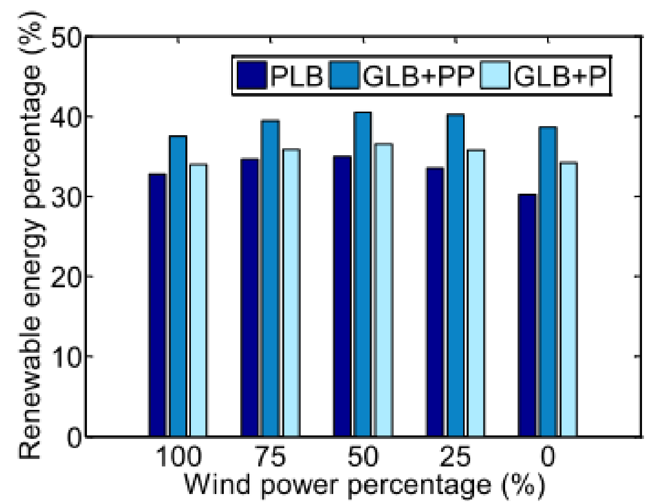
Percentage of renewable energy used out of total energy demand in data centers

- GLB w/w.o prediction error increases the utilization of the available renewable energy
- For the same renewable energy utilization GLB w/w.o. perdition error needs smaller energy storage
- GLB optimally leverages renewable energy when there is mix of wind/solar energy

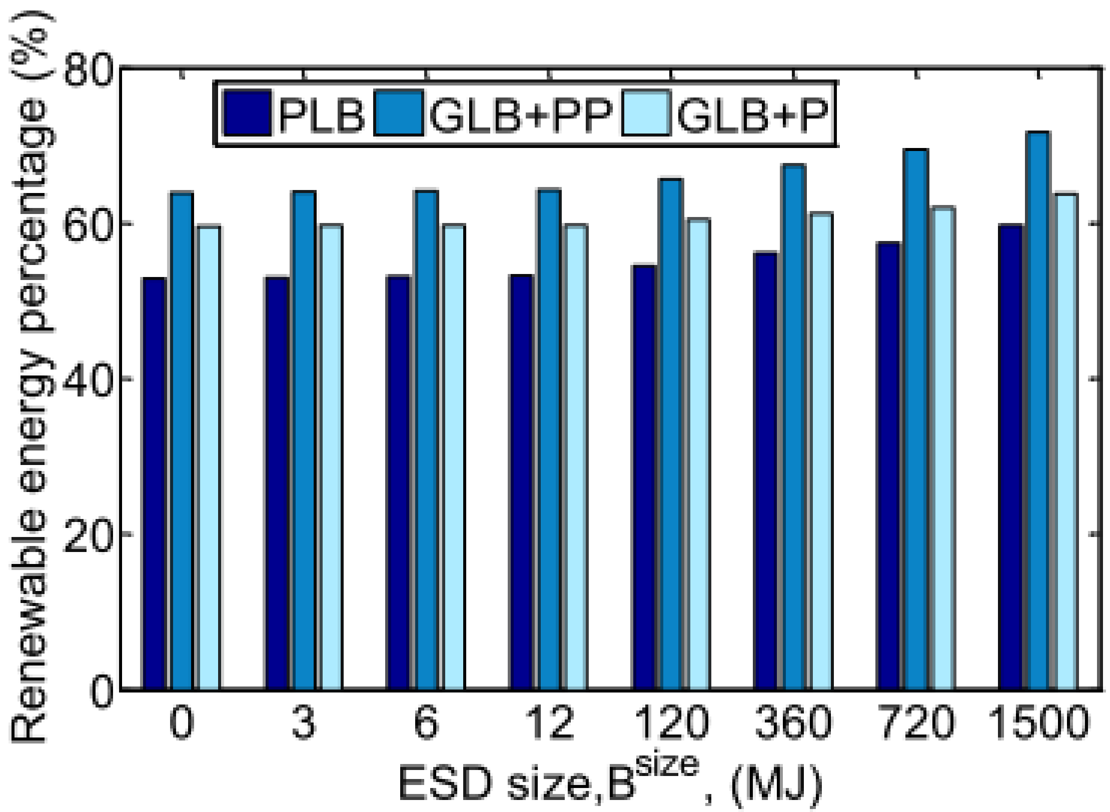


i.e., Increasing the available renewable energy

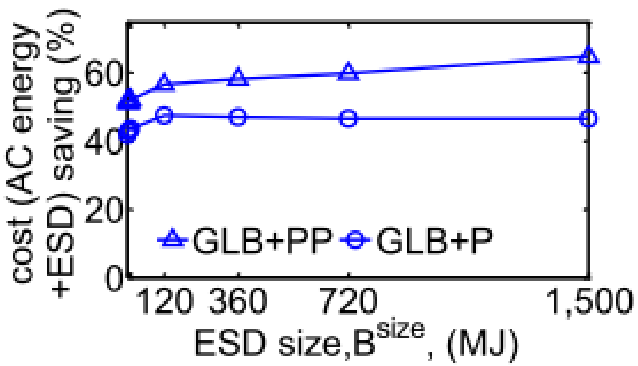


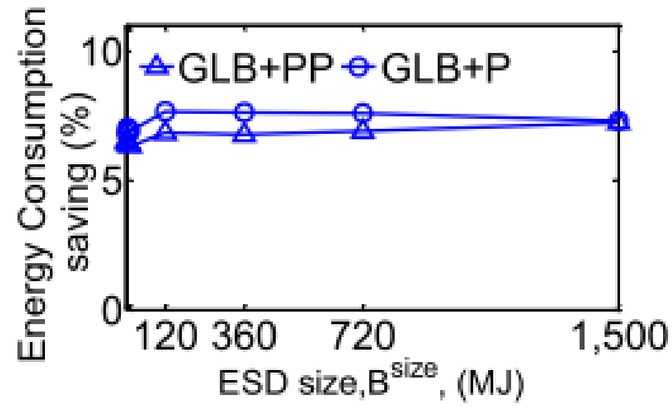


Renewable energy utilization versus ESD capacity



Energy consumption and cost





Energy saving and energy cost GLB+PP and GLB+P compared to PLB

- GLB w/w.o prediction error saves energy cost which may be along with saving total energy consumption
- With increasing energy storage size, the cost saving of GLB+P decreases