# Optimizing Peak Power-Related Costs in Cloud Data Centers

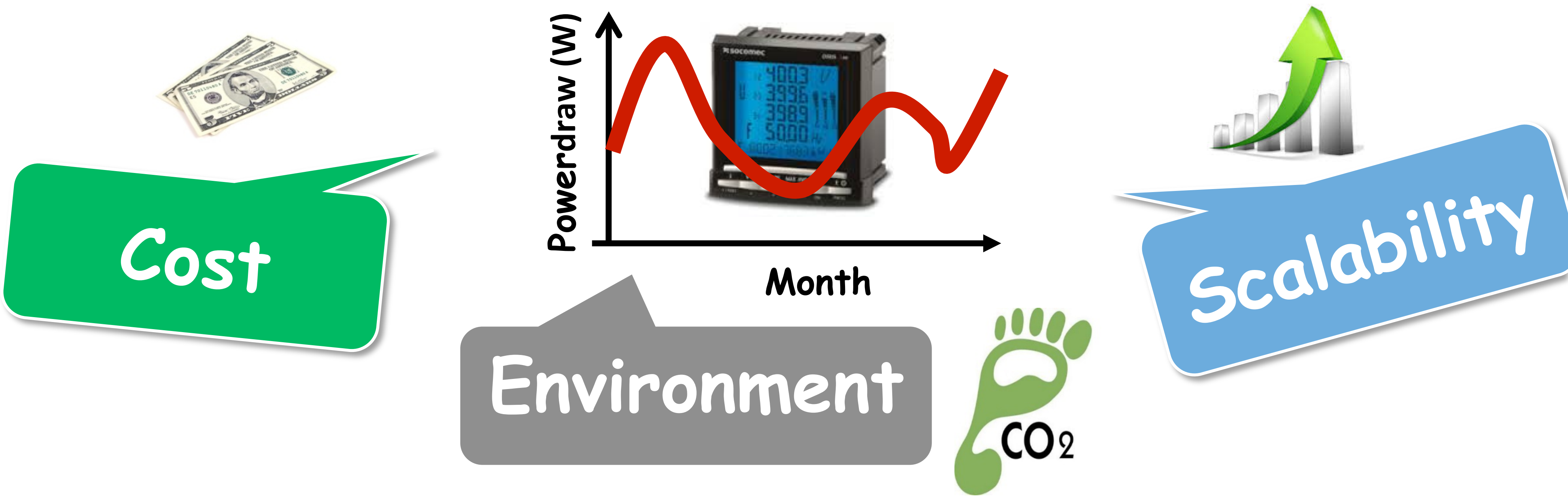
**Bhuvan Urgaonkar**

*Department of Computer Science and Engineering*

*Penn State University*

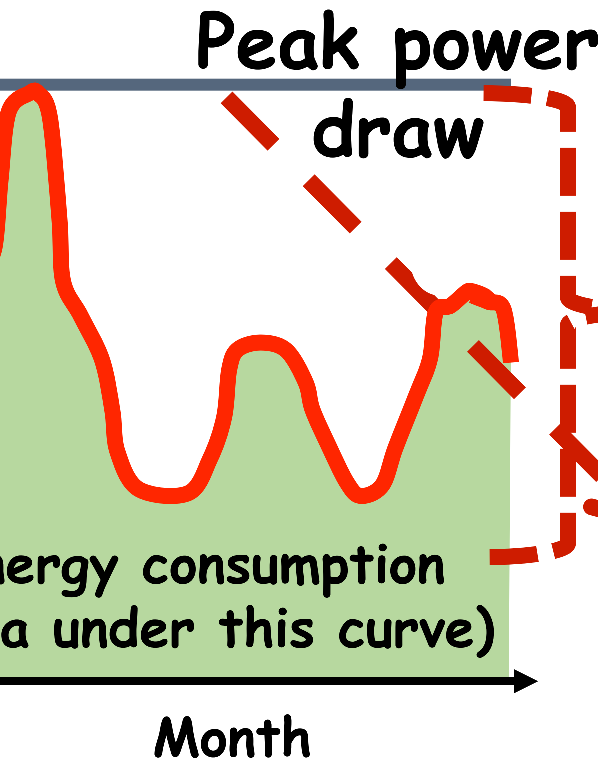
Collaborators: C. Wang, G. Kesidis

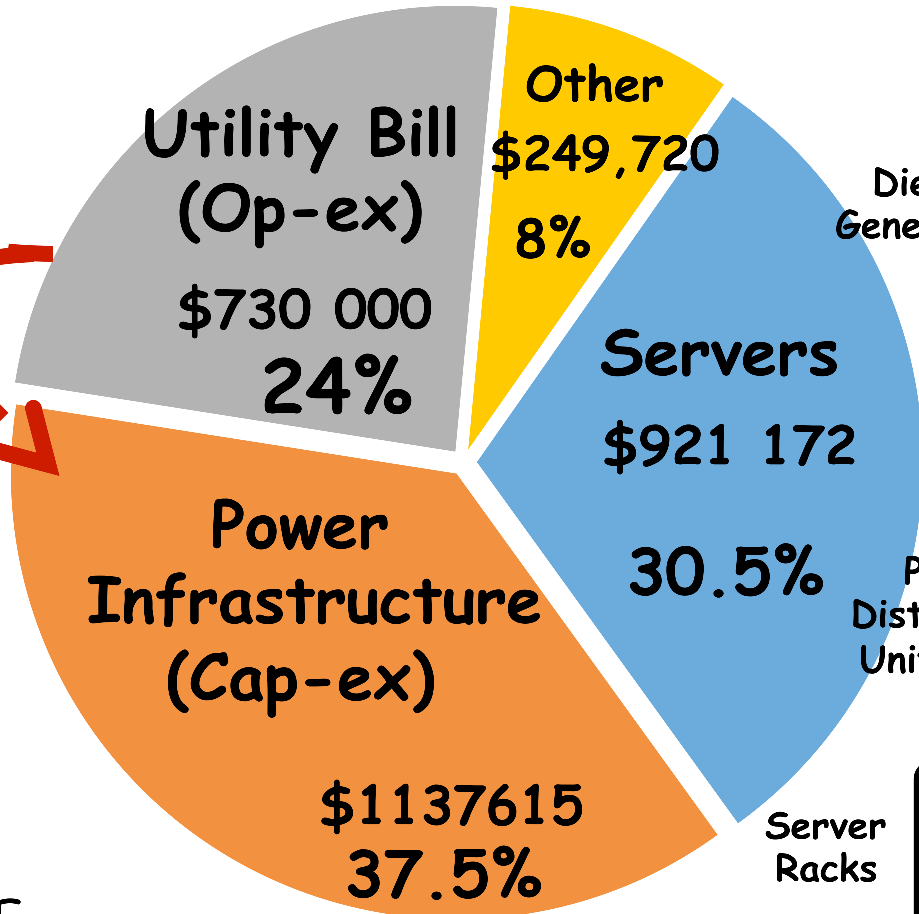
# Data Centers Consume Lots of Power!




f treated as a country, *fifth* in the world for electricity use
ouble in next 5 years, imposing a peak load of over 20 GW on the grid

# Monthly Cost of 10MW Data Center



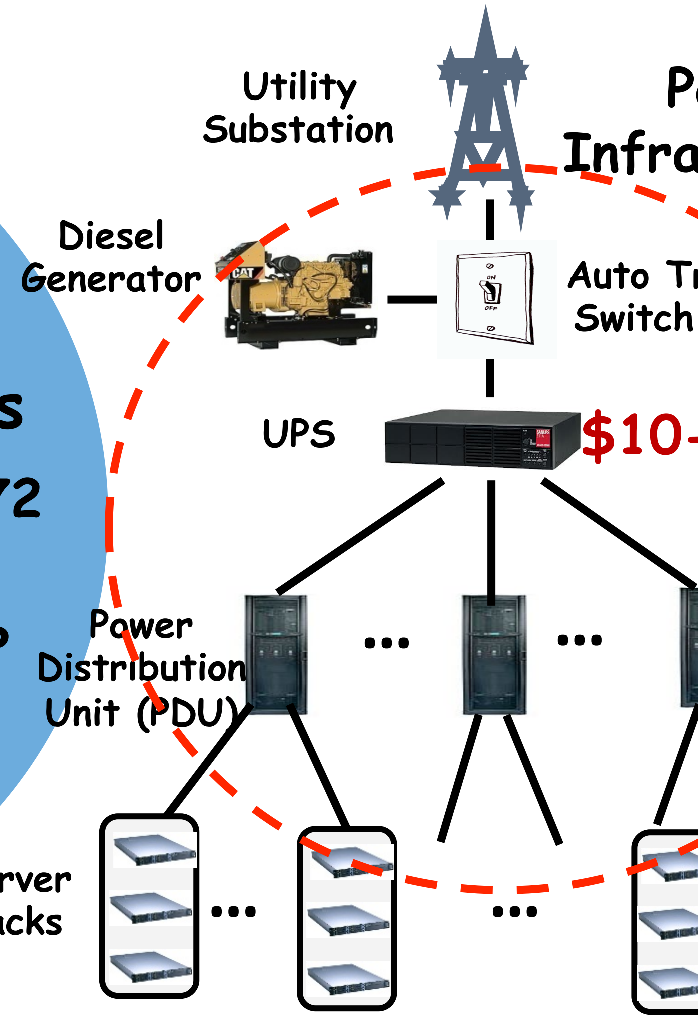
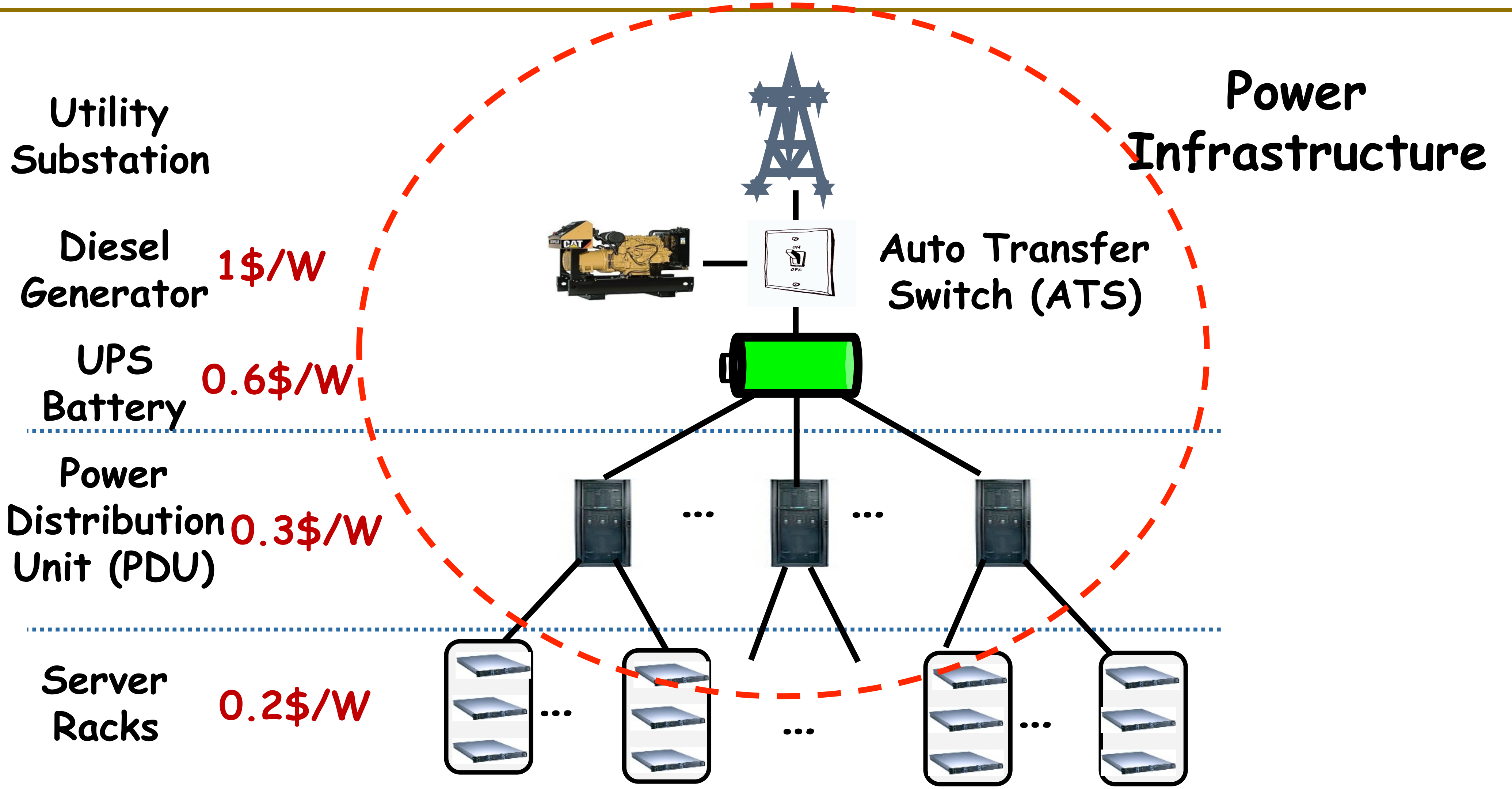



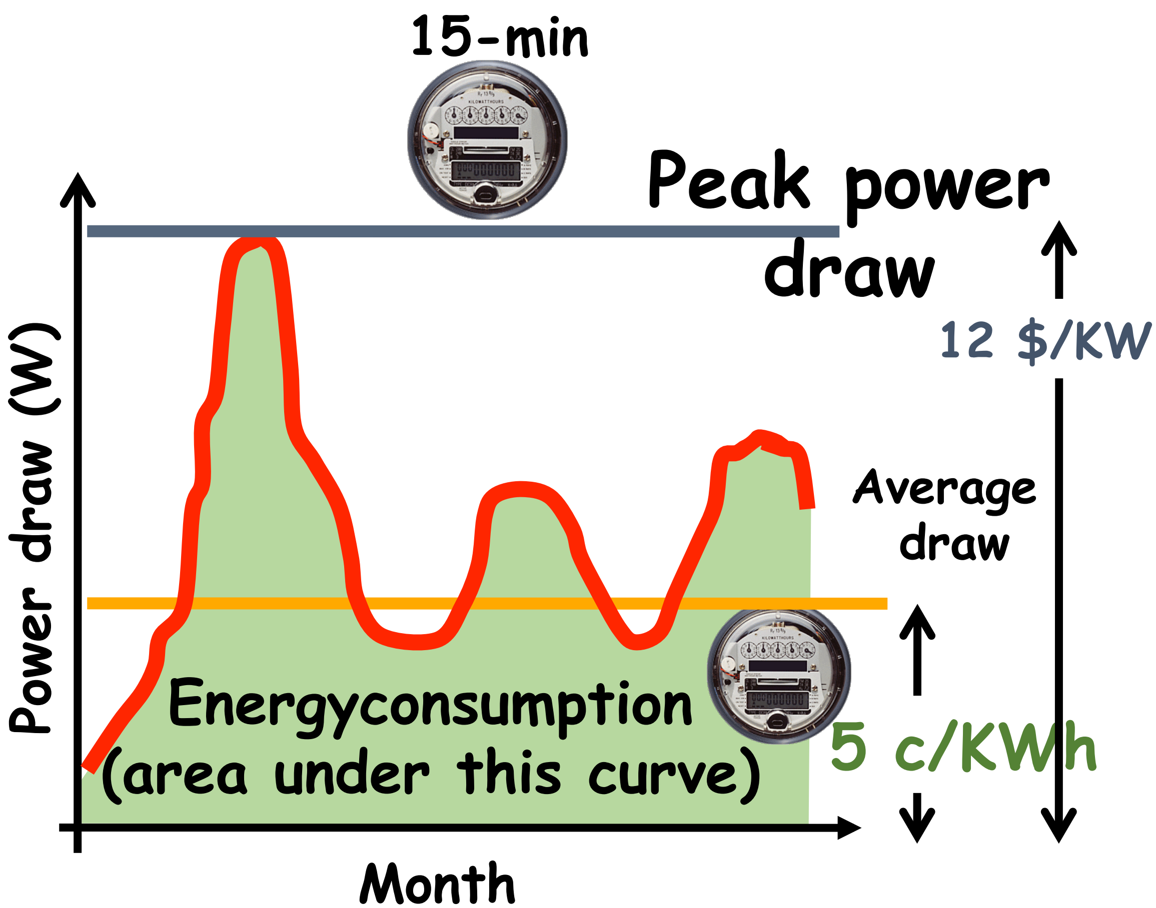

**Chart:**
*ce:* Book by Barroso et al.,
*ion:* 20,000 servers, 1.5 PUE,
Cap-ex, Duke Energy Op-ex,
*ver & 12 yr infrastructure*
*mortization (Tier-2)*

***All cost are amortized at a monthly granularity***

# Provisioned Peak Power Impact on Cap-ex



# Consumed Peak Draw Contribution to Op-Ex (Explicit Peak-based Tariff)





Peak to Average ratio: 3:1

**Duke Utility Tariffs**
(12 $/KW, 5 c/KWh)

***Note:*** **Tariff rates collected from Duke Energy Utility.**

# onsumed Peak Draw Contribution to Op-Ex (Implicit)

eal-time pricing with high "coincident" peak charges


eak draw affects both Cap-Ex and Op-E

# Optimizing Cap-Ex and Op-Ex

Cap-Ex optimization: How much capacity to provision for the next several years?

- An offline problem

# Optimizing Cap-Ex and Op-Ex

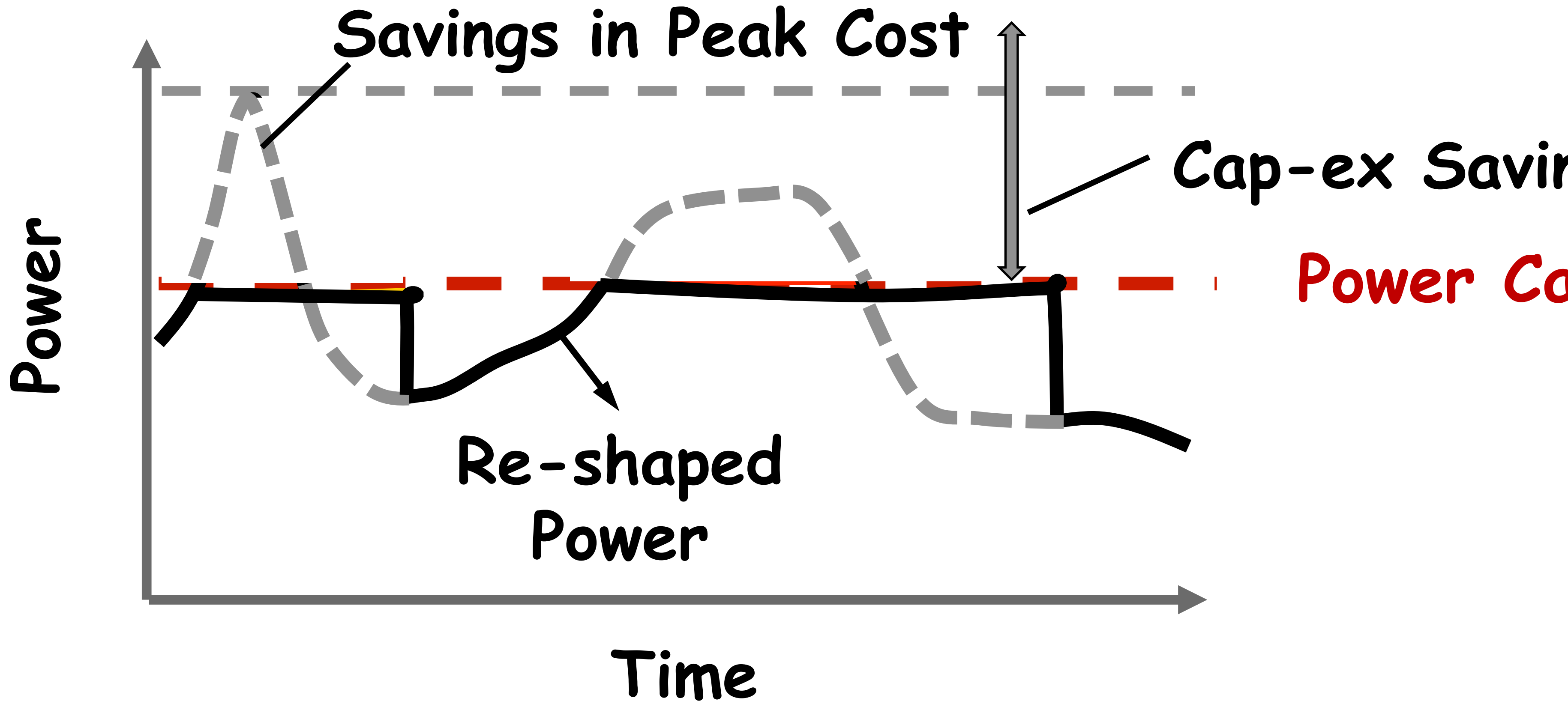
Cap-Ex optimization: How much capacity to provision for the next several years?

- An offline problem

Op-ex: How much peak to admit for ***this billing cycle***?

- An online control problem
  - Control windows may be in the minutes (or even seconds)
- Complementary problem: how to operate cost-effectively within a specified power capacity (as determined by cap-ex optimization)

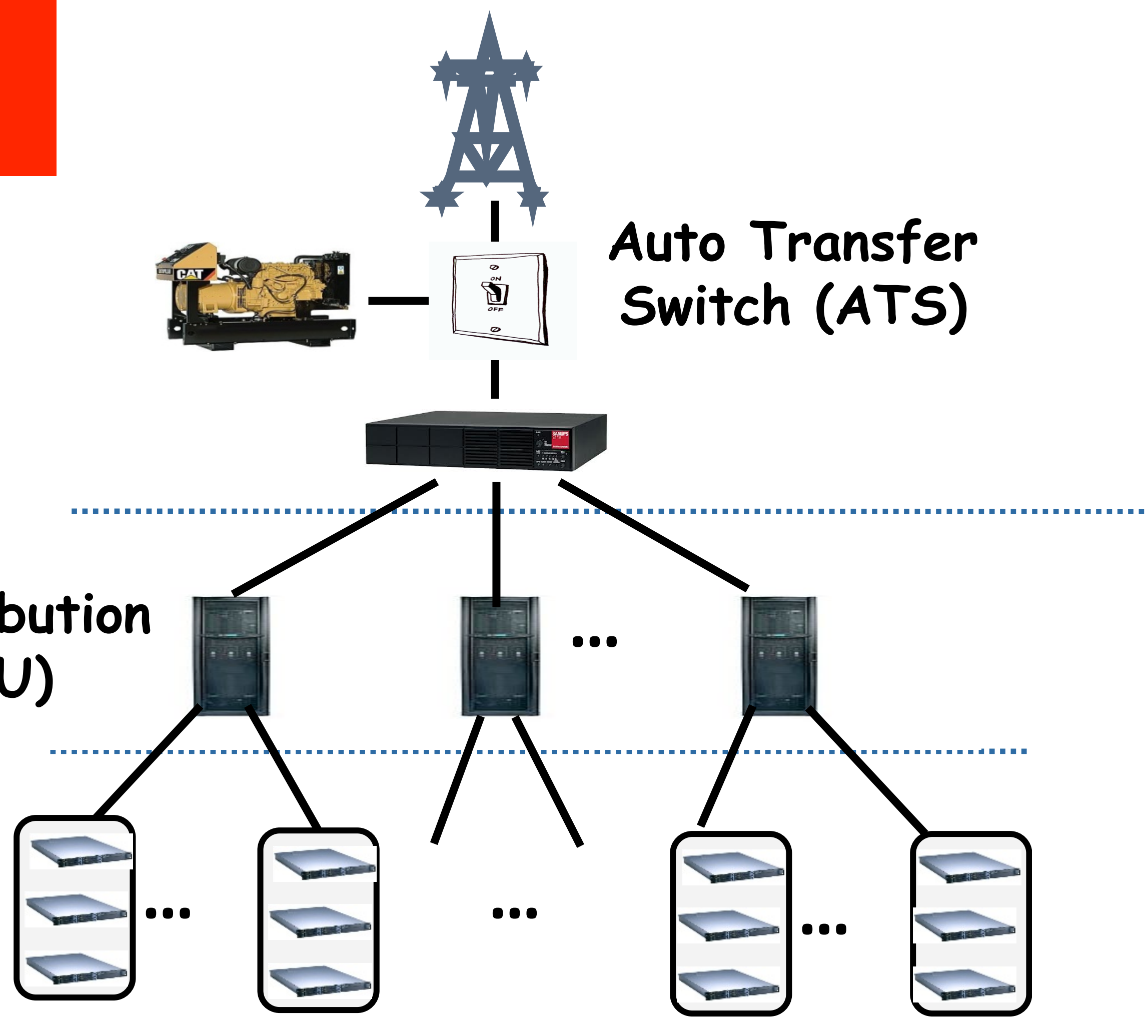
# Demand Response: An Important Set of Techniques for Optimizing Power Costs

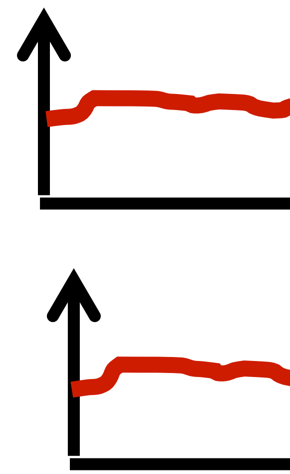


# Demand Response Knobs in a Data Center

emporal Knobs
FS, Load scheduling
or deferral)

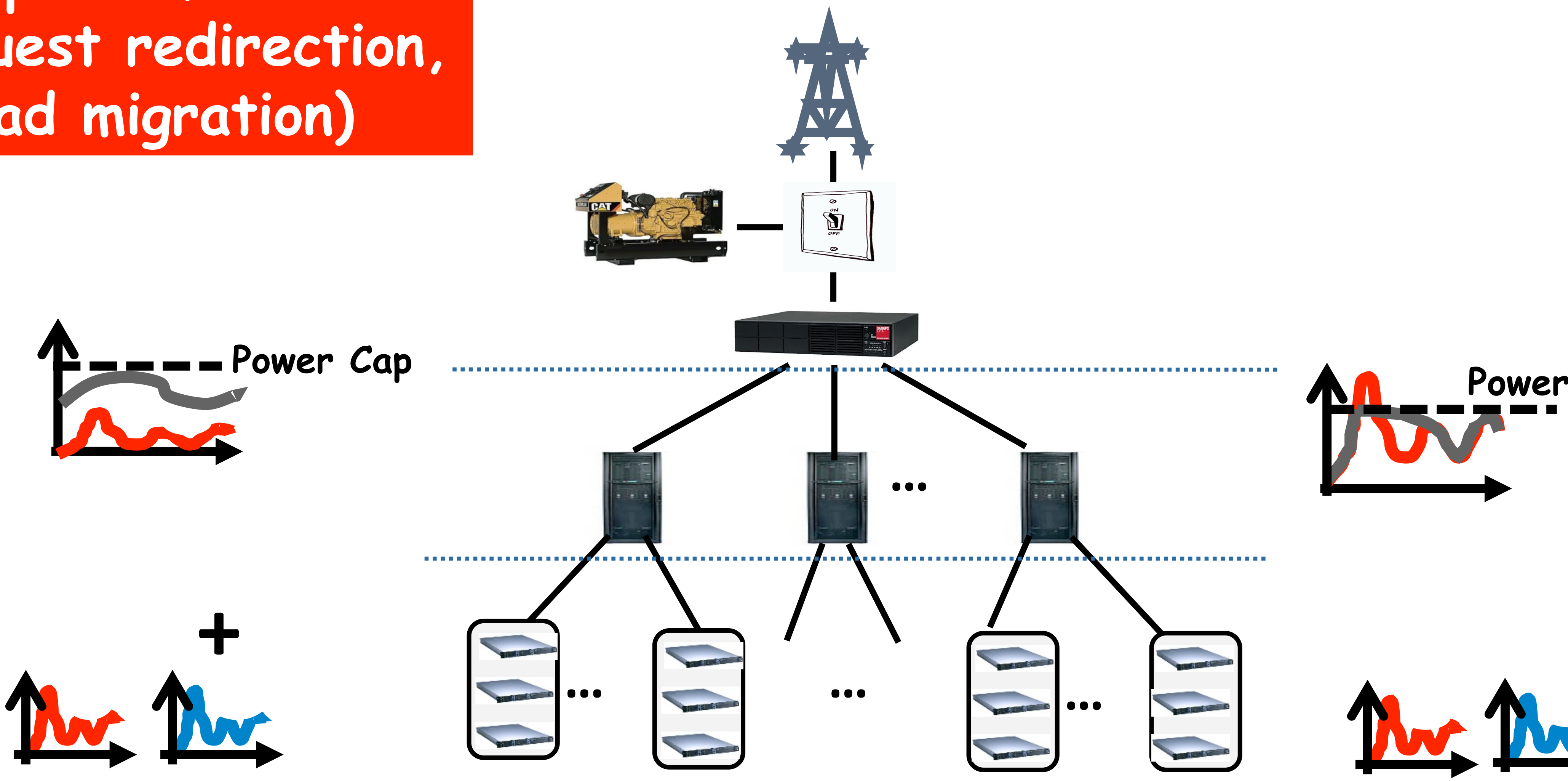
| | |
|---|---|
| 1$/W | Diesel Generator |
| 0.6$/W | UPS |
| 0.3$/W | Power Distribution Unit (PDU) |
| 0.2$/W | Server Racks |



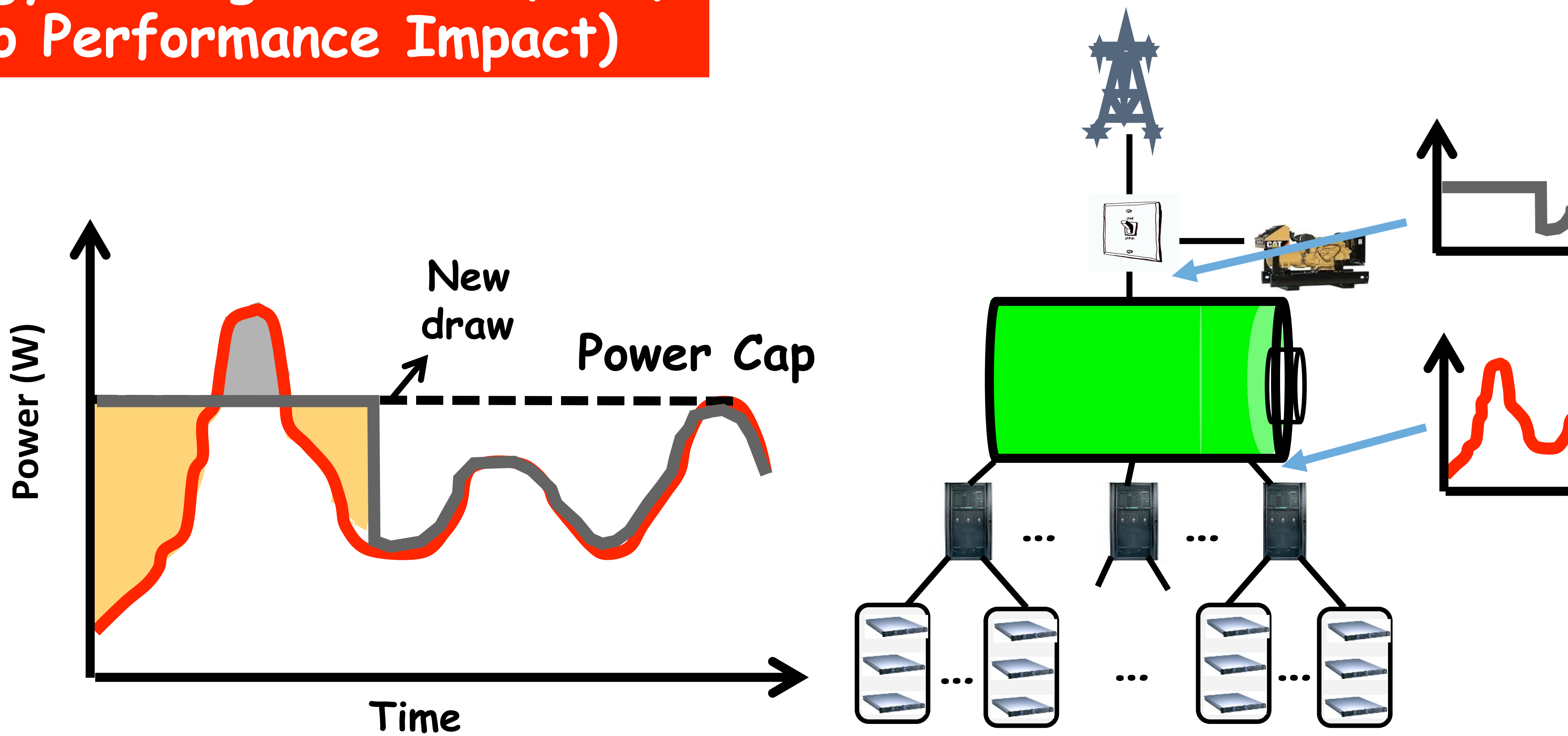
# Demand Response Knobs in a Data Center

Spatial Knobs
quest redirection,
oad migration)



# Demand Response Knobs in a Data Center

rgy Storage Device (ESD)
lo Performance Impact)



# Overview of our Work

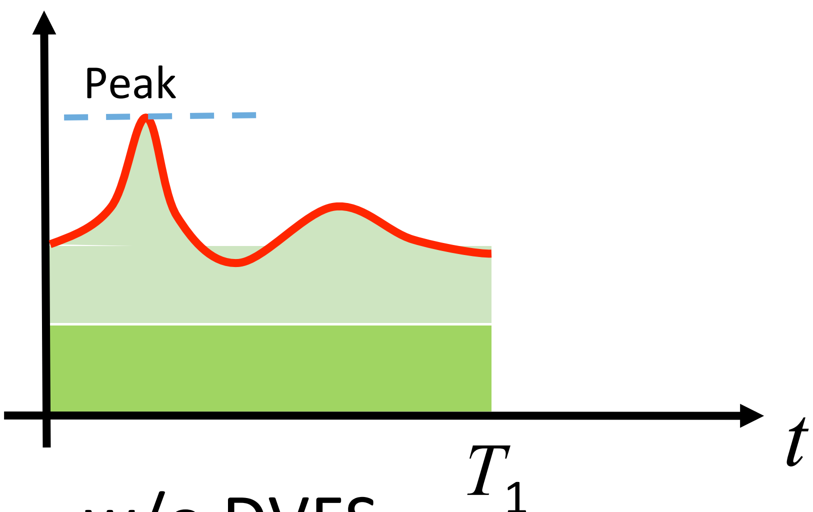
- This talk
  - Op-ex optimization using IT control knobs for a peak-based utility pricing scheme
- Other work (happy to discuss offline)
  - Cap-ex improvements via provisioning of batteries and local generation sources
  - Op-ex optimization:
    - Real-time utility pricing schemes
    - Control of batteries and local generation sources

# A Simple Model for IT-based DR

espite their diversity, IT knobs can be viewed
*ropping* and/or *delaying* some power demand at the co
performance degradation / revenue loss.

ample: DVFS/Scheduling

ver
ver

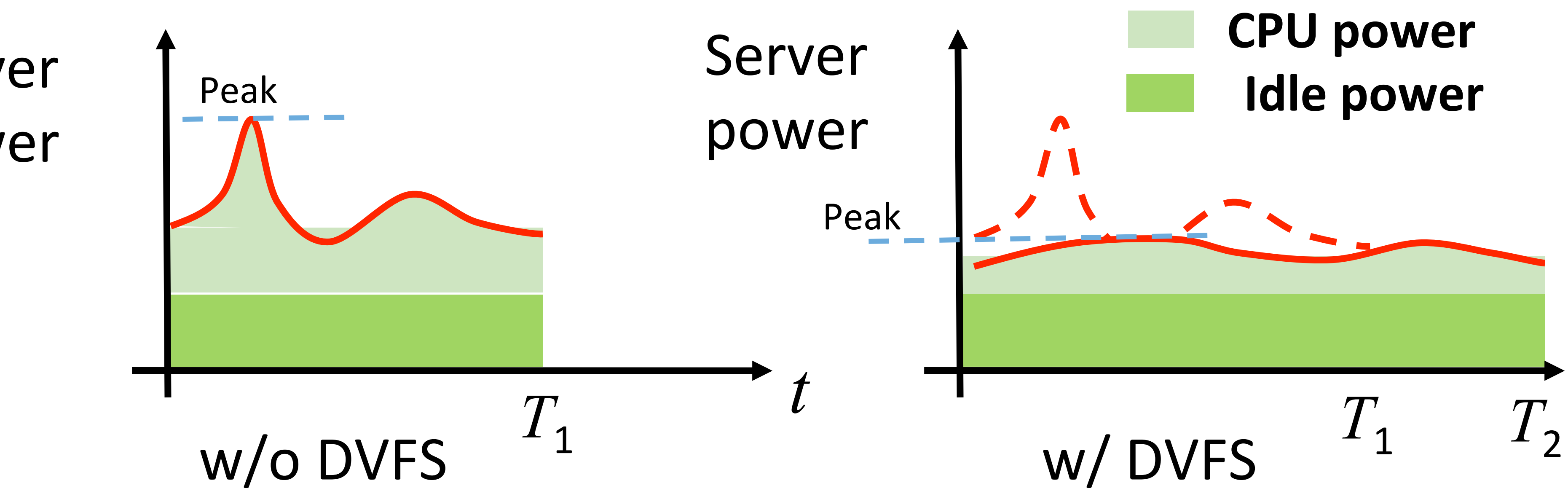



w/o DVFS

# A Simple Model for IT-based DR

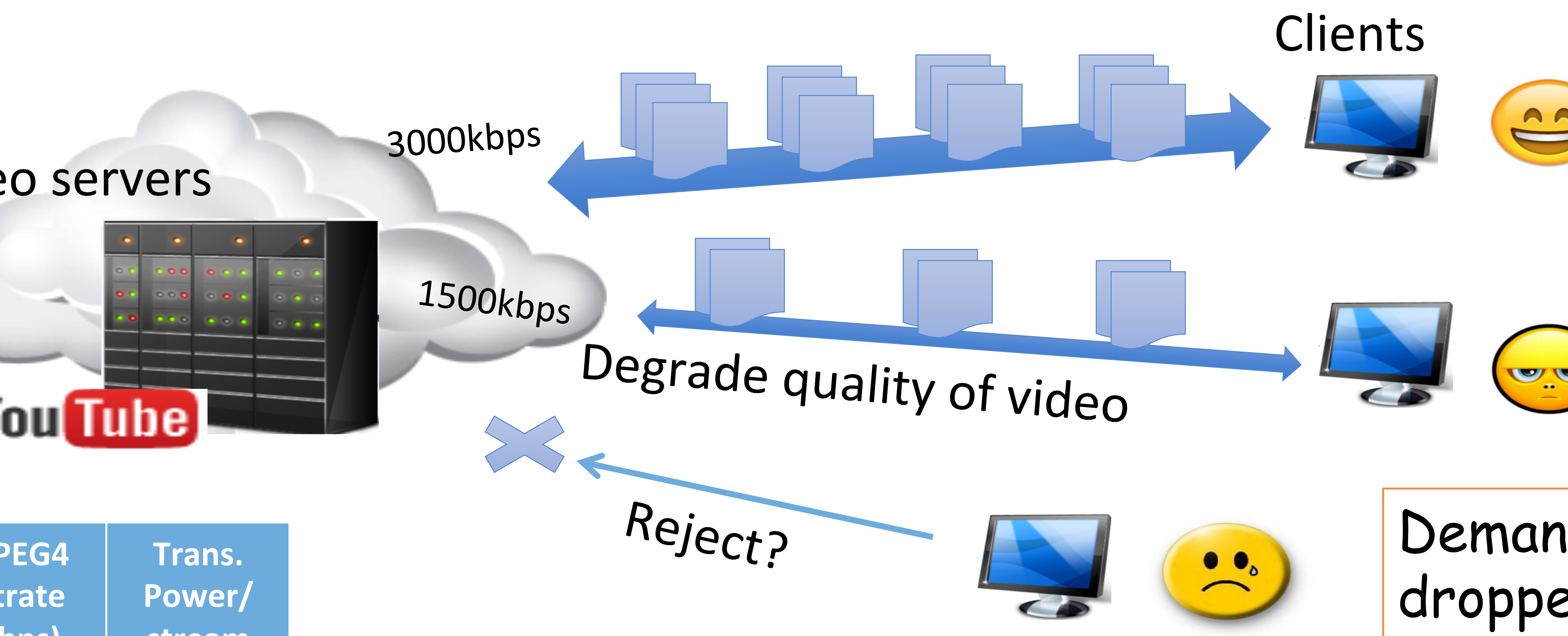
espite their diversity, IT knobs can be viewed
*ropping* and/or *delaying* some power demand at the co
performance degradation / revenue loss.

ample: DVFS/Scheduling




$T_1 < T_2$
Delay

# Example 1: MPEG Video Server



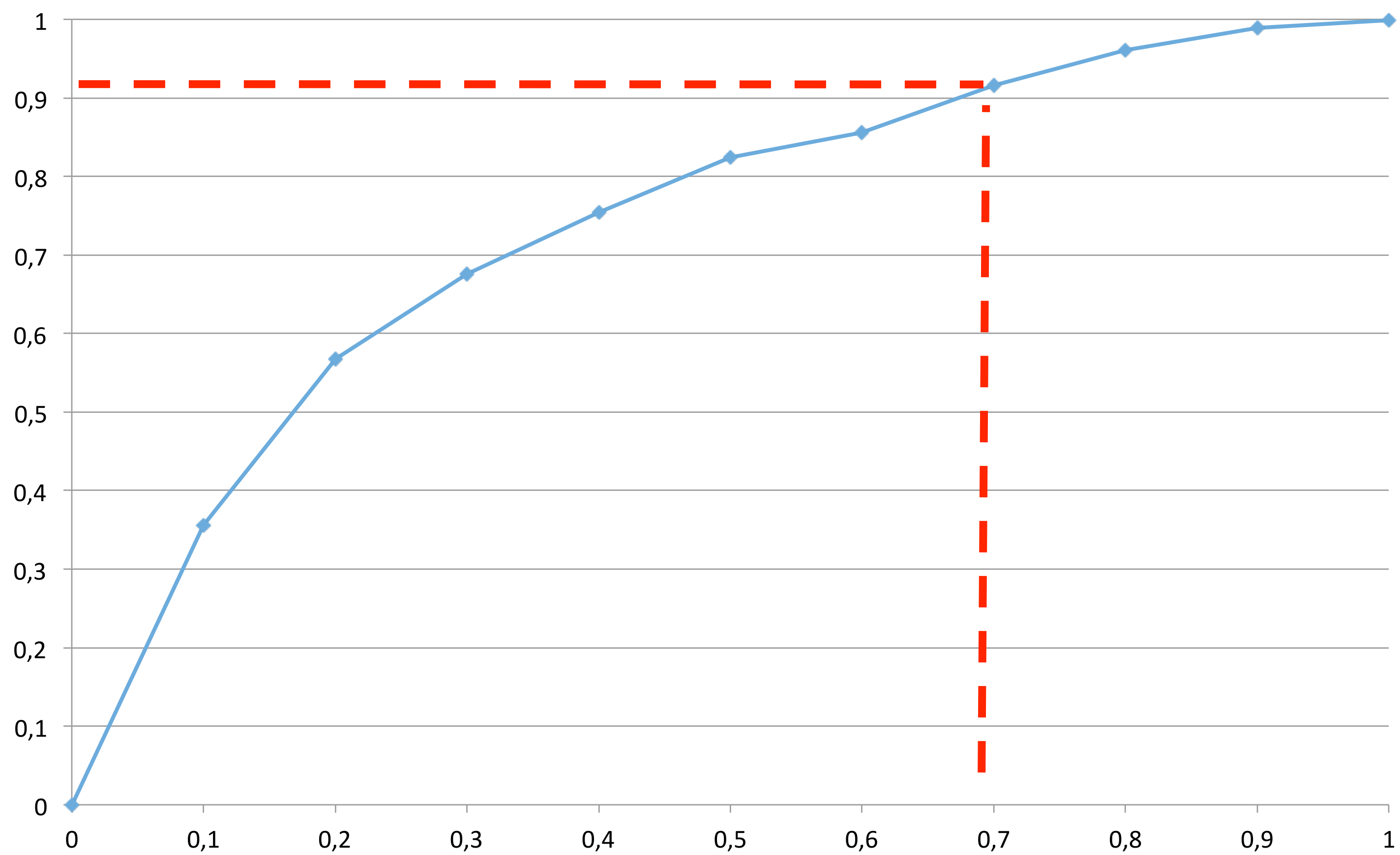

| PEG4 trate bps) | Trans. Power/ stream (watt) |
|---|---|
| 000 | 3.0 |
| 500 | 1.5 |

*Source:* Y. Sharrab et al., MMSys’13 and Wikipedia

*E.g., Source on revenue impact:* “Video stre quality impacts viewer behavior,” Krishna a Sitaraman, IEEE/ACM TON, 2013

# Example 2: Search Engine

## Concave Quality Profile of Bing Search



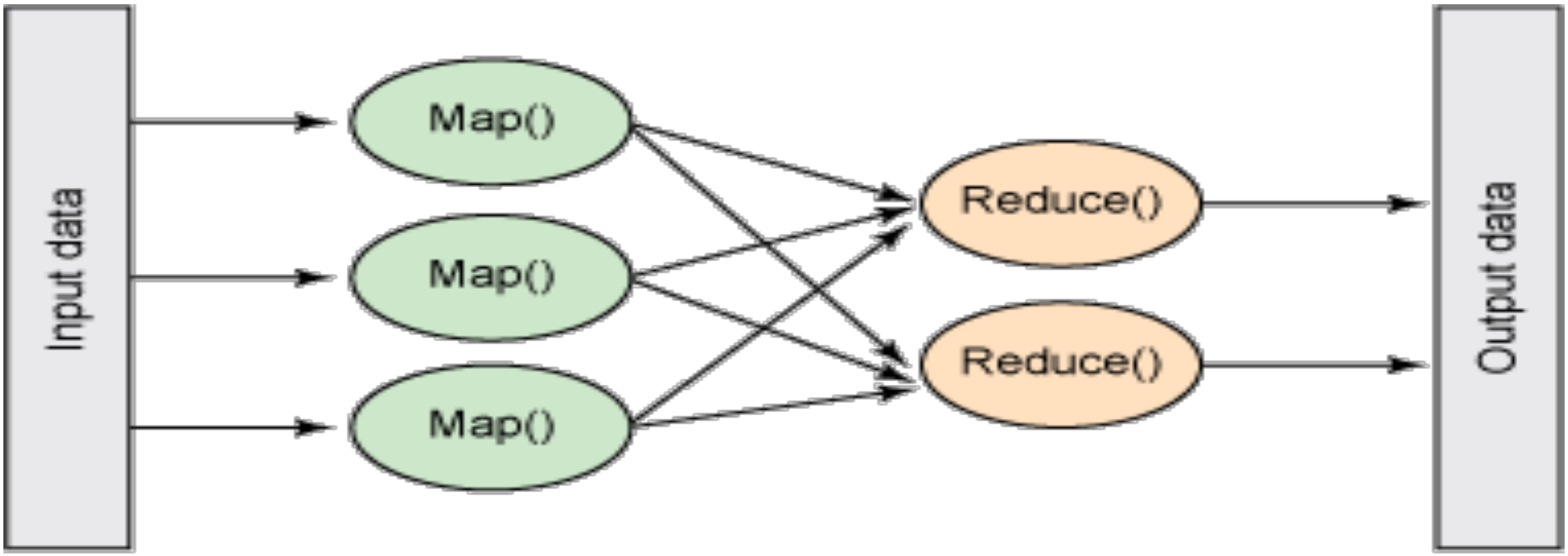

bing

Degrade quality of query

Demand dropped

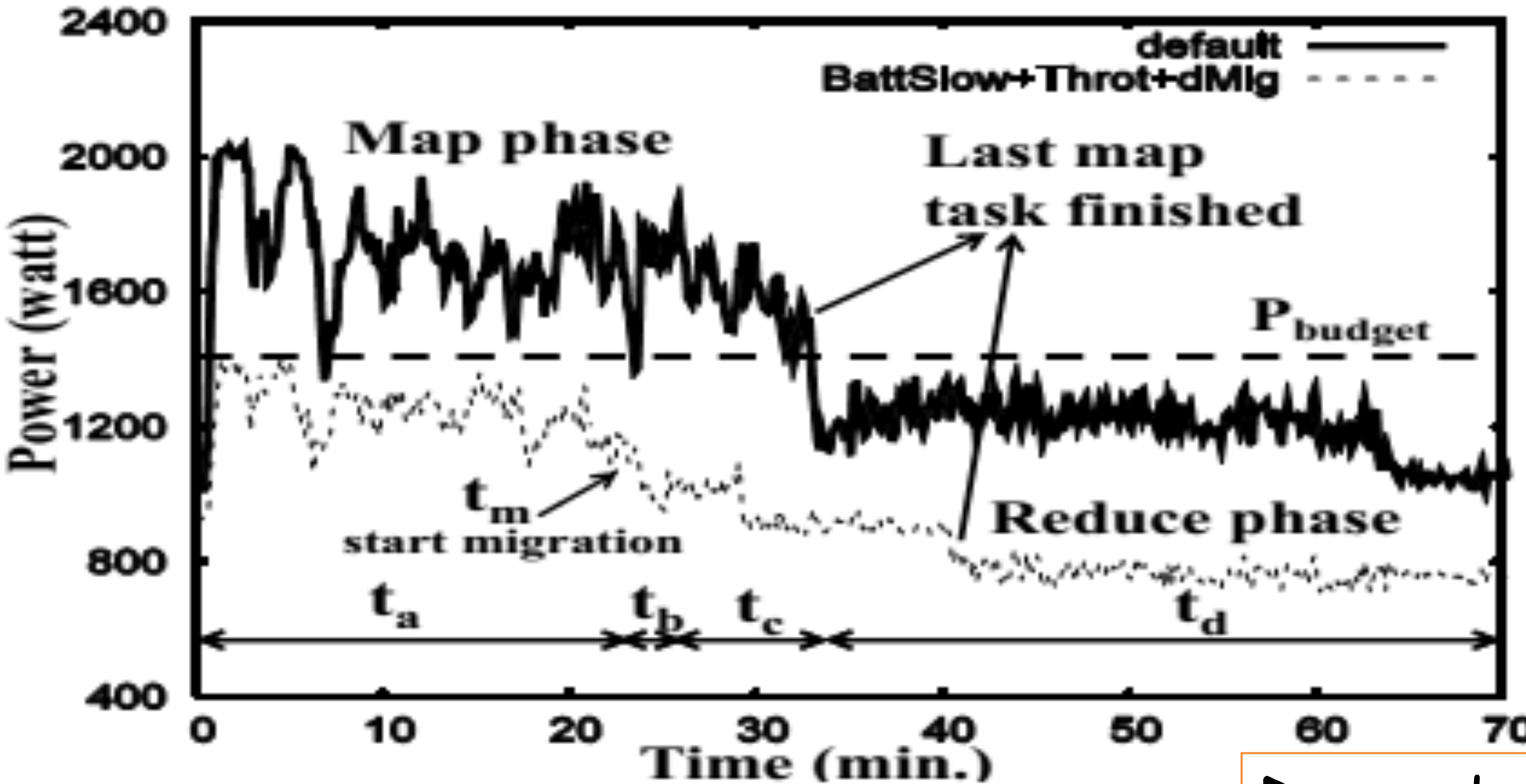
*ource:* Y. He et al., *SOCC'12*

# Example 3: Delay-tolerant, Batch




Split Shuffle Merge

# Example 3: Delay-tolerant, Batch




**MapReduce power profile**

*Source:* S. Govindan et al., *ASPLOS'12*

Demand delay

# Op-ex Optimization Problem

How to use IT-based dropping or delaying of power demand to optimize op-ex vs. performance/revenue loss trade-off?

# Much Related Work for Real-time Pricing

**al-time pricing**

| | Adversarial power demands | Stochastically known power demands |
|---|---|---|
| sing IT-based DR | Z. Liu et al., Sigmetrics'13, robust optimization, avoid coincident peak | |
| sing batteries | | R. Urgaonkar et al., Sigm'11, Lyapunov optimization, distance from optimal inversely prop. to battery size<br>P. Van de Ven et al., Energy'11, residential energy storage, MDP |

# Much Related Work for Real-time Pricing

**al-time pricing**

| | Adversarial power demands | Stochastically known power demands |
|---|---|---|
| sing IT-based DR | Z. Liu et al., Sigmetrics'13, robust optimization, avoid coincident peak | **Current work (SDP formulation)** |
| sing batteries | | R. Urgaonkar et al., Sigm'11, Lyapunov optimization, distance from optimal inversely prop. to battery size<br>P. Van de Ven et al., Energy'11, residential energy storage, MDP |

# But Less for Peak-based Pricing

k-based pricing

| | Adversarial power demands | Stochastically known power demands |
|---|---|---|
| sing IT-based DR | Current work: CR=2 for time-varying energy prices; CR=2-1/T for fixed energy prices | Current work (SDP formulation, gSBB heuristic) |
| sing batteries for DR | A. Bar-Noy et al., WEA'08, threshold-based, CR of $H_n$ (=7.84 if 30-min time-slot) | Current work (SDP formulation, gSBB heuristic) |

# But Less for Peak-based Pricing

**k-based pricing**

| | Adversarial power demands | Stochastically known power demands |
|---|---|---|
| sing IT-based DR | **Current work: CR=2 for time-varying energy prices; CR=2-1/T for fixed energy prices** | **Current work (SDP formulation, gSBB heuristic)** |
| sing batteries for DR | A. Bar-Noy et al., WEA'08, threshold-based, CR of $H_n$ (=7.84 if 30-min time-slot) | **Current work (SDP formulation, gSBB heuristic)** |

# Online "Dropping" of Power Demand

Lets begin by assuming that the "knob" available to the data center is that of dropping part of the power demand

- Dropped demand never returns

Recall examples of a video streaming server and a search engine

# Demand Response to Optimize Peak-based Utility Bill

**How to determine the peak demand to admit in an *online* fashion?**

# Offline Formulation for Dropping Demand

## Demand dropping

- $l_{drop}(x)$ : Dropping demand v.s. Performance/Revenue loss
- Discretized optimization horizon T: A billing cycle (typically a mont
- Known demand time series $\{p_t\}_{t=1}^{T}$

Peak cost

Energy cost

Dropping cost

$$\min_{\{A_t\},\{D_t\}} \beta y_{max} + \sum_{t=1}^{T} \{\alpha_t a_t + l_{drop}(d_t)\}$$

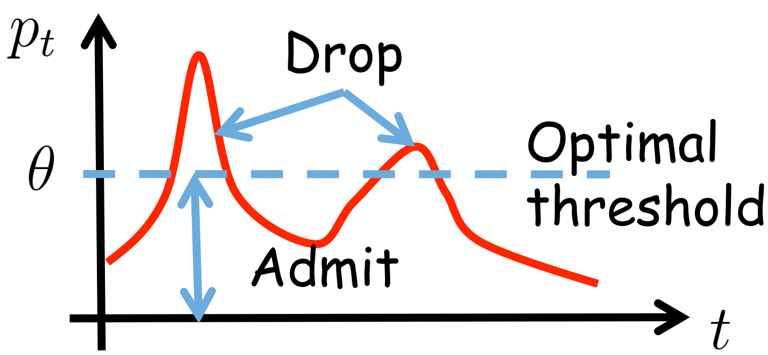
$s.t.$ $\quad p_t - a_t - d_t = 0, \forall t$ — New demand either admitted or dropp

$\quad y_{max} \geq a_t, \forall t$ — Peak of admitted demand

# Online Control: $ON_{\text{Drop}}$

No information about future demand

Peak charge + time-varying energy price

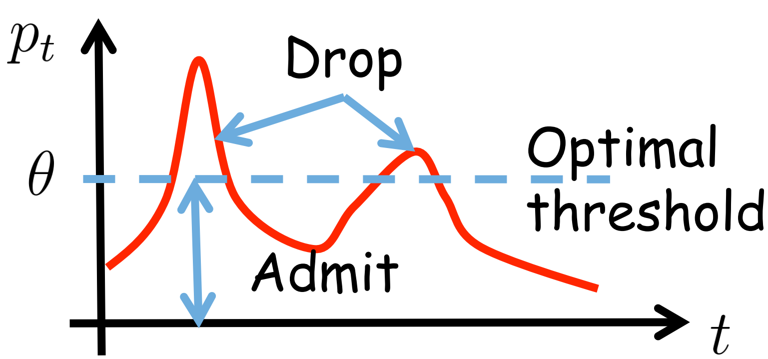


# Online Control: $ON_{Drop}$

No information about future demand

Peak charge + time-varying energy price $l_{drop}(x) = k_{drop}x$

**emma**. The optimal solution has a demand dropping threshold $\theta$ of the follo
rm: If we denote as $\hat{p}_t$ the $t$-th largest demand value in $\{p_t\}_{t=1}^T$ and as $\hat{\alpha}_t$
orresponding energy price, then $\theta = \hat{p}_n$
here $\beta - \sum_{t=1}^{n-1}(k_{drop} - \hat{\alpha}_t) \geq 0$ and $\beta - \sum_{t=1}^{n}(k_{drop} - \hat{\alpha}_t) \leq 0$ .

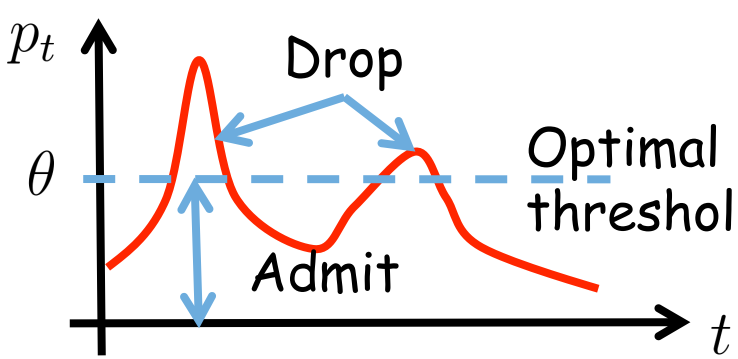


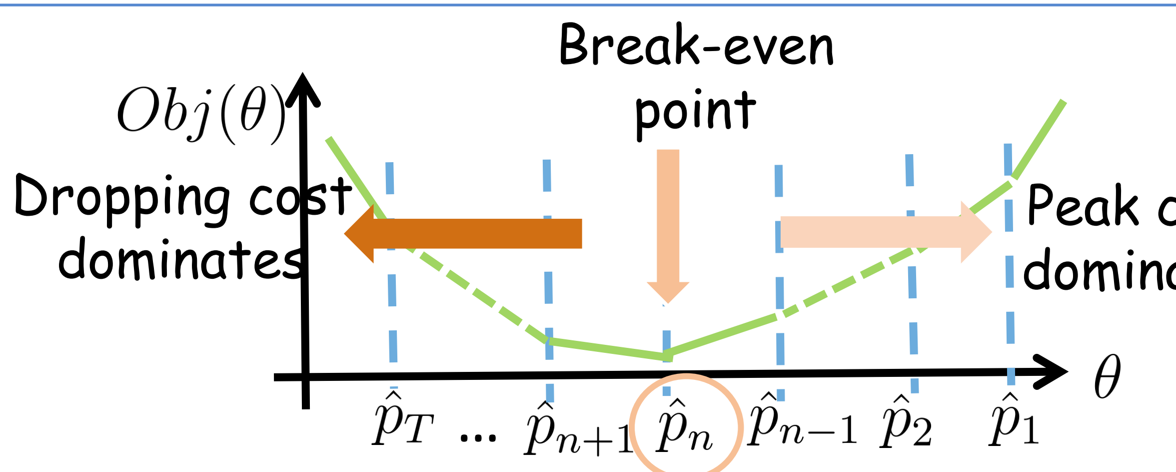
# Online Control: $ON_{\text{Drop}}$

No information about future demand

Peak charge + time-varying energy price $l_{drop}(x) = k_{drop} x$

**emma**. The optimal solution has a demand dropping threshold $\theta$ of the follo
rm: If we denote as $\hat{p}_t$ the $t$-th largest demand value in $\{p_t\}_{t=1}^T$ and as $\hat{\alpha}_t$
orresponding energy price, then $\theta = \hat{p}_n$
here $\beta - \sum_{t=1}^{n-1}(k_{drop} - \hat{\alpha}_t) \geq 0$ and $\beta - \sum_{t=1}^{n}(k_{drop} - \hat{\alpha}_t) \leq 0$ .



# Online Control: *ON*$_{\text{Drop}}$

Decision-making. Admit , drop .

$\theta_0 = 0$

At time *t*, sort $p_1, p_2, ..., p_t$ into $\hat{p}_1, \hat{p}_2, ..., \hat{p}_t$ such that $\hat{p}_1 \geq \hat{p}_2 \geq ... \geq \hat{p}_t$ .

Update $\theta_t$ as follows: Find index $n$ such that $\beta - \sum_{t=1}^{n-1}(k_{drop} - \hat{\alpha}_t) \geq 0$ and $\beta - \sum_{t=1}^{n}(k_{drop} - \hat{\alpha}_t) \leq 0$ ; set $\theta_t = \hat{p}_n$ .

Decision-making. Admit $\min\{p_t, \theta_t\}$ , drop $[p_t - \theta_t]^+$.

*): The CR of *ON*$_{\text{Drop}}$ can be improved if $\theta_0$ can be trained using historical data.

**Theorem**. *ON*$_{\text{Drop}}$ offers a competitive ratio of 2 under peak-based pricing.

# Stochastic Control for Dropping Demand

In many cases, workloads can be predicted

- Often via Markovian models

Can develop a SDP that leverages such predictive models

Offline formulation:

$$\min_{\{a_t\},\{D_t\}} E\left\{\beta y_{max} + \sum_{t=1}^{T}(\alpha_t a_t + l_{drop}(d_t))\right\}$$

Unconventional state space due to *sum + max*

Stochastic dynamic programming?

↓

Sum + Max

↓

Sol: Track peak-so-far by state $y_t$

$$y_{t+1} = \max\{y_t, a_t\}$$

# Stochastic Control for Dropping Demand

$SDP_{\text{Drop}}$ optimality rules:

$$V_t(y_t, p_{[t-1]}, \alpha_{[t-1]}) = \min_{\{A_t\},\{D_t\}} E\left\{\alpha_t a_t + l_{drop}(d_t) + V_{t+1}(y_{t+1}, p_{[t]}, \alpha_{[t]}) \right.$$
$$\left. \mid P_{[t-1]} = p_{[t-1]}, \Lambda_{[t-1]} = \alpha_{[t-1]}\right\}$$

$$s.t. \quad y_{t+1} = \max\{y_t, a_t\}$$
$$p_t - a_t - d_t = 0$$

**Lemma**. Under stage-independent demand $SDP_{\text{Drop}}$ has the following threshold-based optimal control policy :

$$(a_t^*, d_t^*) = \begin{cases} (\phi_t p_t, p_t - \phi_t p_t), & \text{if} \phi_t \leq 1 \\ (p_t, 0), & \text{if} \phi_t > 1 \end{cases}$$

# Making the model a bit more complex

**What if dropping alone does not capture DR behavior?**

**Recall example of MapReduce …**

# Offline Problem Formulation

Demand delaying $l_{delay}(x,t)$: Delay up to $\tau$ time slots

Demand dropping $l_{drop}(x)$

Peak cost

Energy cost

Dropping cost

Dela

$$\min_{\ldots\},\{D_t\}} \beta y_{max} + \sum_t \{\alpha \sum_{i\in h^+(t)} a_{i,t} + \sum_{i\in h^+(t)} l_{drop}(d_{i,t}) + \sum_{i\in h(t)} l_{delay}(a_{i,t}, t - \ldots$$

| | |
|---|---|
| $s.t.\ p_t - a_{t,t} - d_{t,t} = r_{t,t+1}, \forall t$ | New demand either admitted or dropp |
| $r_{i,t} - a_{i,t} - d_{i,t} = r_{i,t+1}, i \in h(t), \forall t$ | Pending demand either admitted or dr |
| $r_{t-\tau,t} - a_{t-\tau,t} - d_{t-\tau,t} = 0, \forall t$ | Delayed for $\tau$ time slots: Admit immed |
| $r_{i,T+1} = 0, i \in h(t)$ | No more delay at the end of billing cyc |
| $y_{max} \geq \sum_{i\in h^+(t)} a_{i,t}, \forall t$ | Peak of admitted demand |

# Stochastic Control

## SDP formulation

- Track all pending demand if $l_{delay}(x,t)$ is non-linear w.r.t. $t$
- Curse of dimensionality: $O(RL_p^{2(\tau+2)}L_\alpha T)$

$${}_t = (r_{t-\tau,t}, r_{t-\tau+1,t}, \ldots, r_{t-1,t}, y_t)$$

$${}_t(s_t, p_{[t-1]}, \alpha_{[t-1]}) = \min_{\{A_t\},\{D_t\}} E\{\alpha_t a_t^+ + l_{drop}(d_{t,t}) + \sum_{i \in h(t)} l_{delay}(a_{i,t}) +$$

$$V_{t+1}(s_{t+1}, p_{[t]}, \alpha_{[t]}) \mid P_{[t-1]} = p_{[t-1]}, \Lambda_{[t-1]} = \alpha_{[t-1]}$$

# Stochastic Control: Curse of Dimensionalit

te vector

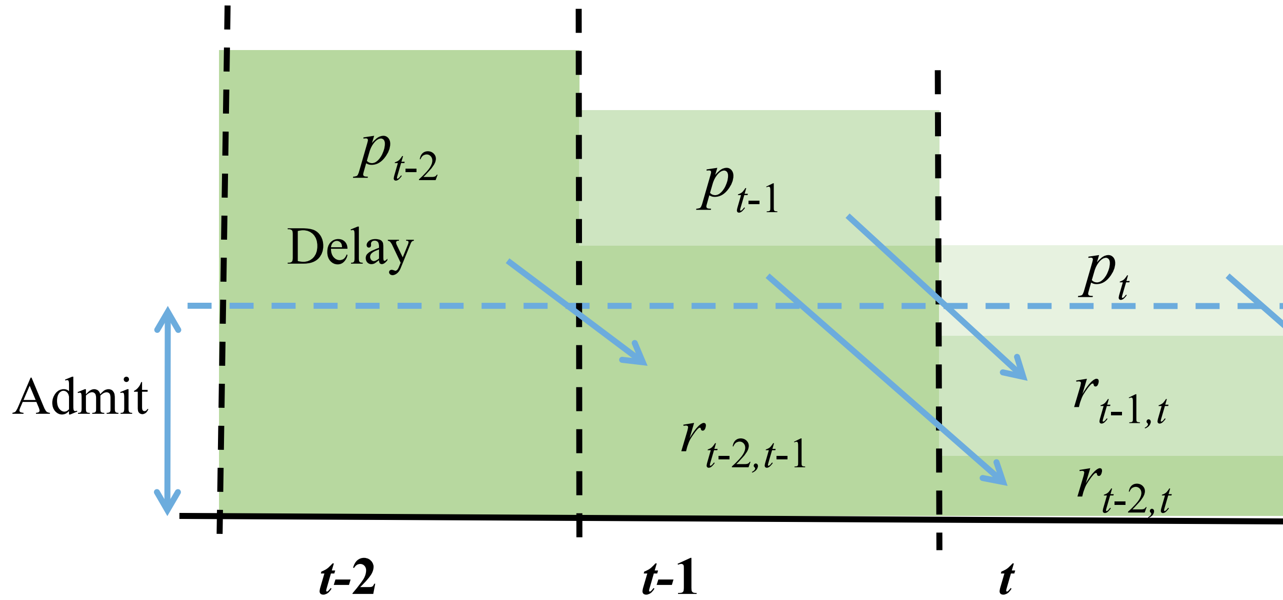
$t-\tau,t, r_{t-\tau+1,t}, \ldots, r_{t-1,t}, y_t, p_t)$

| Delay / time slots | Num. of states |
| --- | --- |
| 0 | $L_p^2$ |
| 1 | $L_p^3$ |
| 2 | $L_p^4$ |
| $\tau$ | $L_p^{2+\tau}$ |

**Curse of dimensionality**

$L_p$: Discretization level

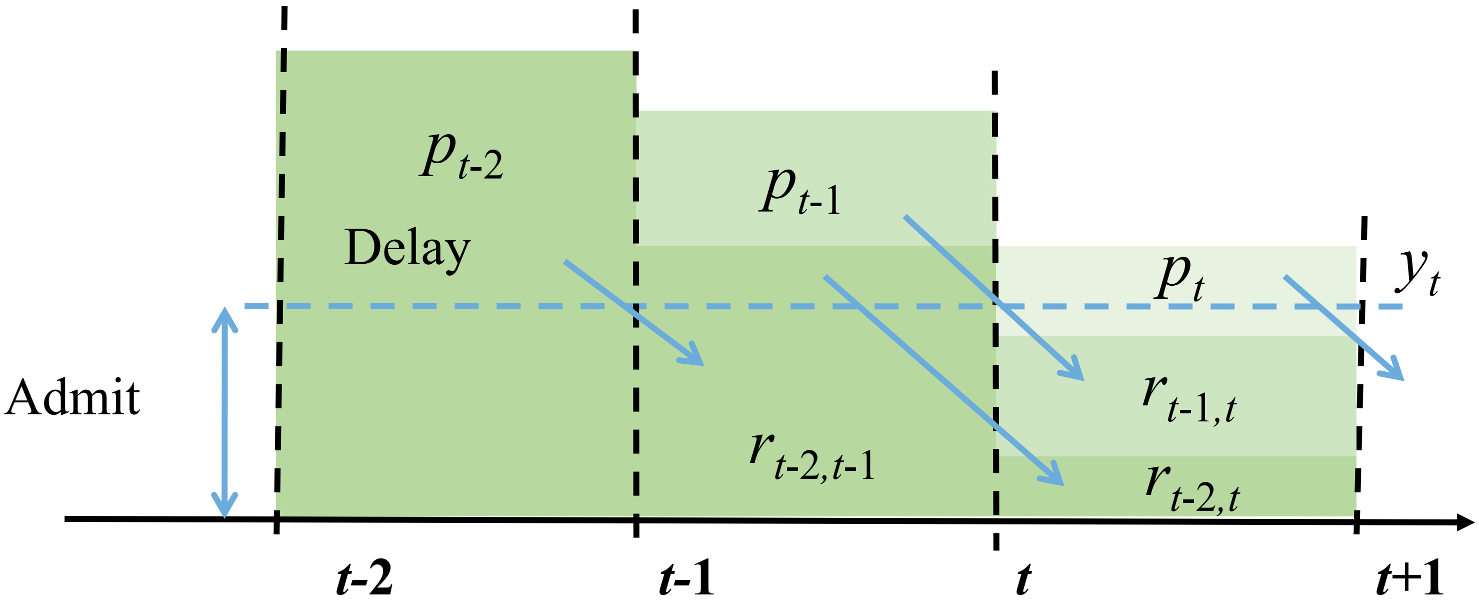
Quadratic delay c
$l_{delay}(x,t) = kxt$




Delay cost 0 $k(r_{t-2,t-1} - r_{t-2,t})1^2$ $kr_{t-2,t}2^2 + kr_t$

**Need to track all pending demands!**

# Stochastic Control: Linear delay cost




Linear delay cost
$l_{delay}(x,t) = kxt$

$\tau = 2$

| Delay cost | | | |
|---|---|---|---|
| | 0 | $k(r_{t-2,t-1} - r_{t-2,t})1$ | $kr_{t-2,t}2 + kr_{t-1,t}1$ |
| | 0 | $k(r_{t-2,t-1} - r_{t-2,t} + r_{t-2,t})1$ | $kr_{t-2,t}1 + kr_{t-1,t}1$ |
| | 0 | $kr_{t-2,t-1}1$ | $k(r_{t-2,t} + r_{t-1,t})1$ |
| | 0 | $kr_{t-1}$ | $kr_t$ |

Equivalent transformation

Only one state va
needed!

# Scalable Approx. for SDP

SDP$_{\text{Lin}}$

- Linear approximation for $l_{\text{delay}}()$
- $O(RL_p{}^5 L_\alpha T)$

$$s_t = (r_t, y_t)$$

$$V_t(s_t, p_{[t-1]}, \alpha_{[t-1]}) = \min_{\{A_t\},\{D_t\}} E\{\alpha_t a_t + l_{drop}(d_t) + l_{delay}(r_t) + V_{t+1}(s_{t+1}, p_{[t]},$$

$$| \; P_{[t-1]} = p_{[t-1]}, \Lambda_{[t-1]} = \alpha_{[t-1]}\}$$

$$s.t. \quad y_{t+1} = \max\{y_t, a_t\}$$

$$r_{t+1} = (p_t - d_t) - a_t - r_t$$

What if SDP does not scale?

A scalable technique based on
a “gSBB” model for power demand

# gSBB-based Control

- Raw demand is modeled as "generalized stochastically bounded burstiness" curve

$$\{(\gamma, \phi(\gamma\tau^*)) \mid \gamma > \mu\}$$

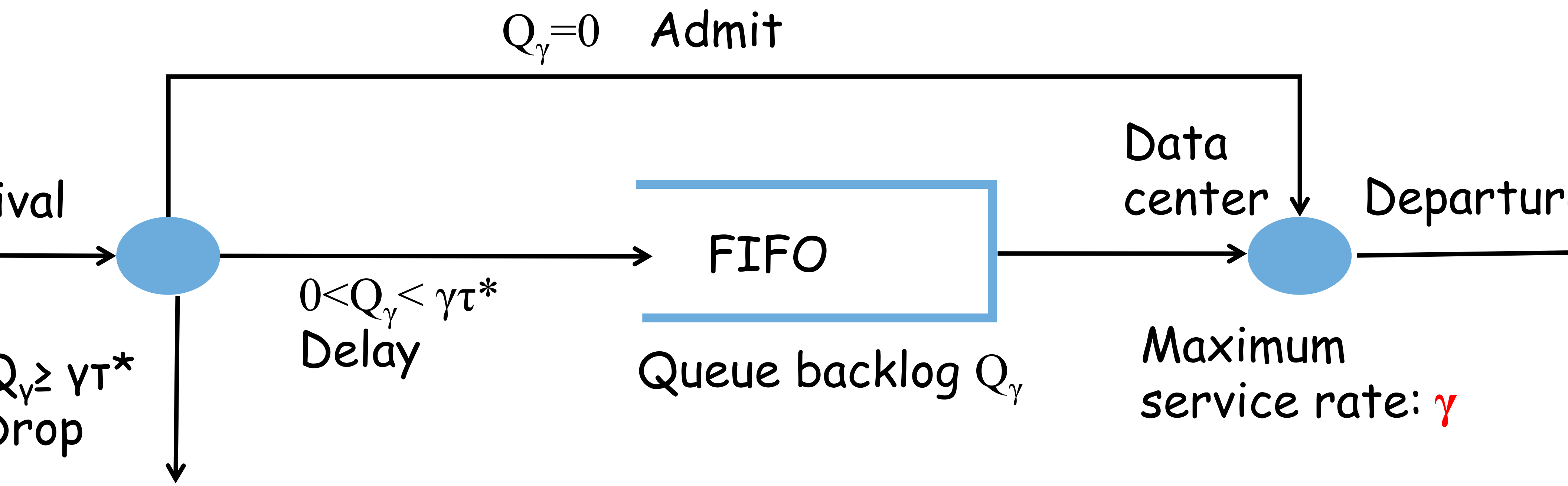
- A queue whose arrivals are the "raw" demands and is served at rate $\gamma$ will have backlog $Q_\gamma$ such that

$$Pr(Q_\gamma \geq \gamma\tau^*) \leq \phi(\gamma\tau^*)$$

# gSBB-based Control

Control loop

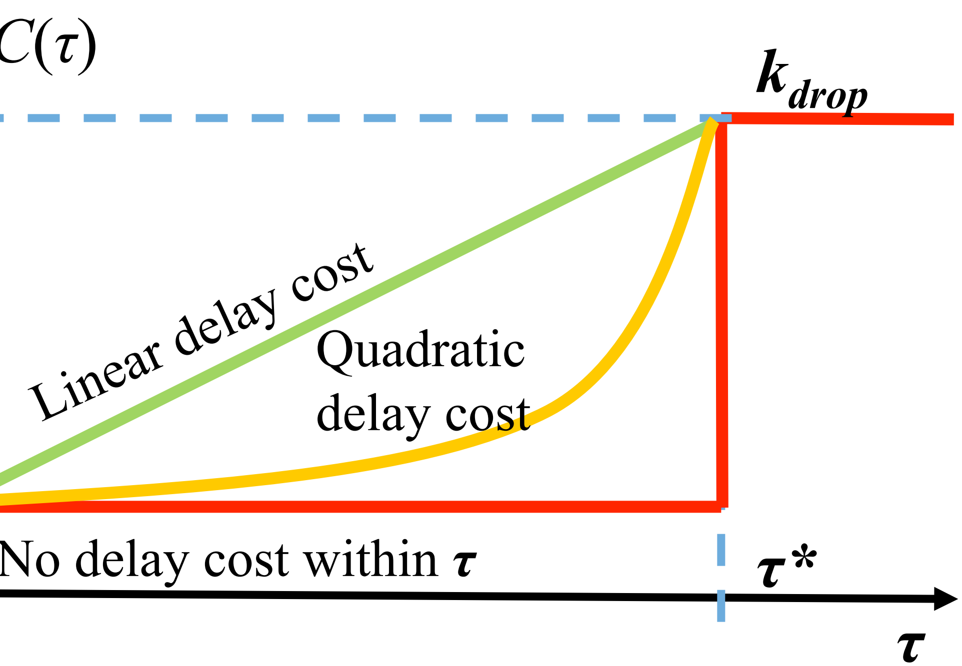
**How to obtain γ ?**



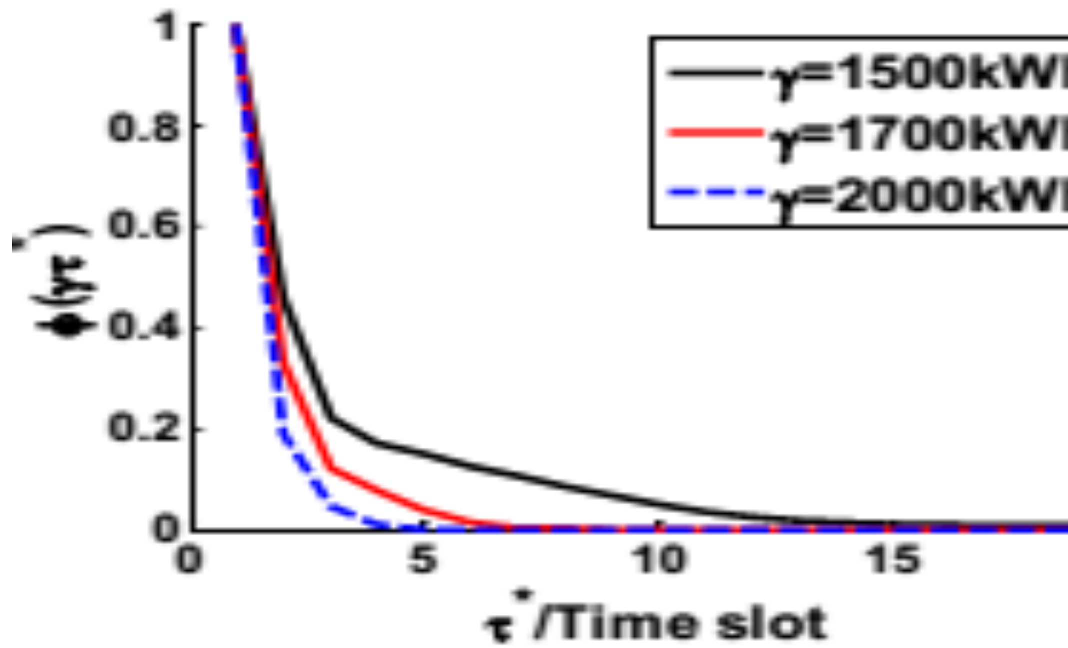
# gSBB-based Control

Objective $\min_{\gamma>\mu} T\mu \int_0^\infty C(\tau)dF_\gamma(\tau) + \beta\gamma$

$-\phi(\gamma\tau)$




Examples of C($\tau$)

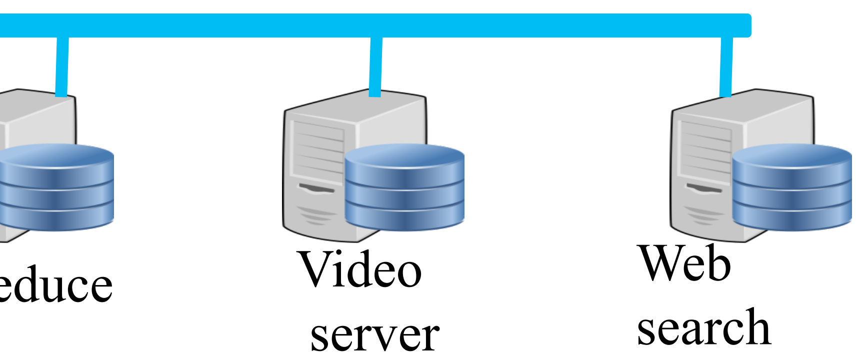


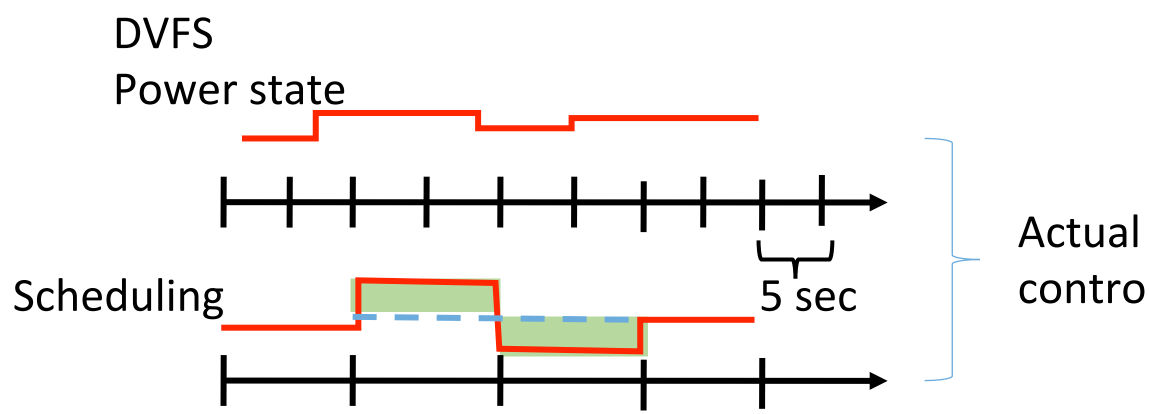

Example of $\phi(\gamma\tau^*)$

# Selected Simulation Results

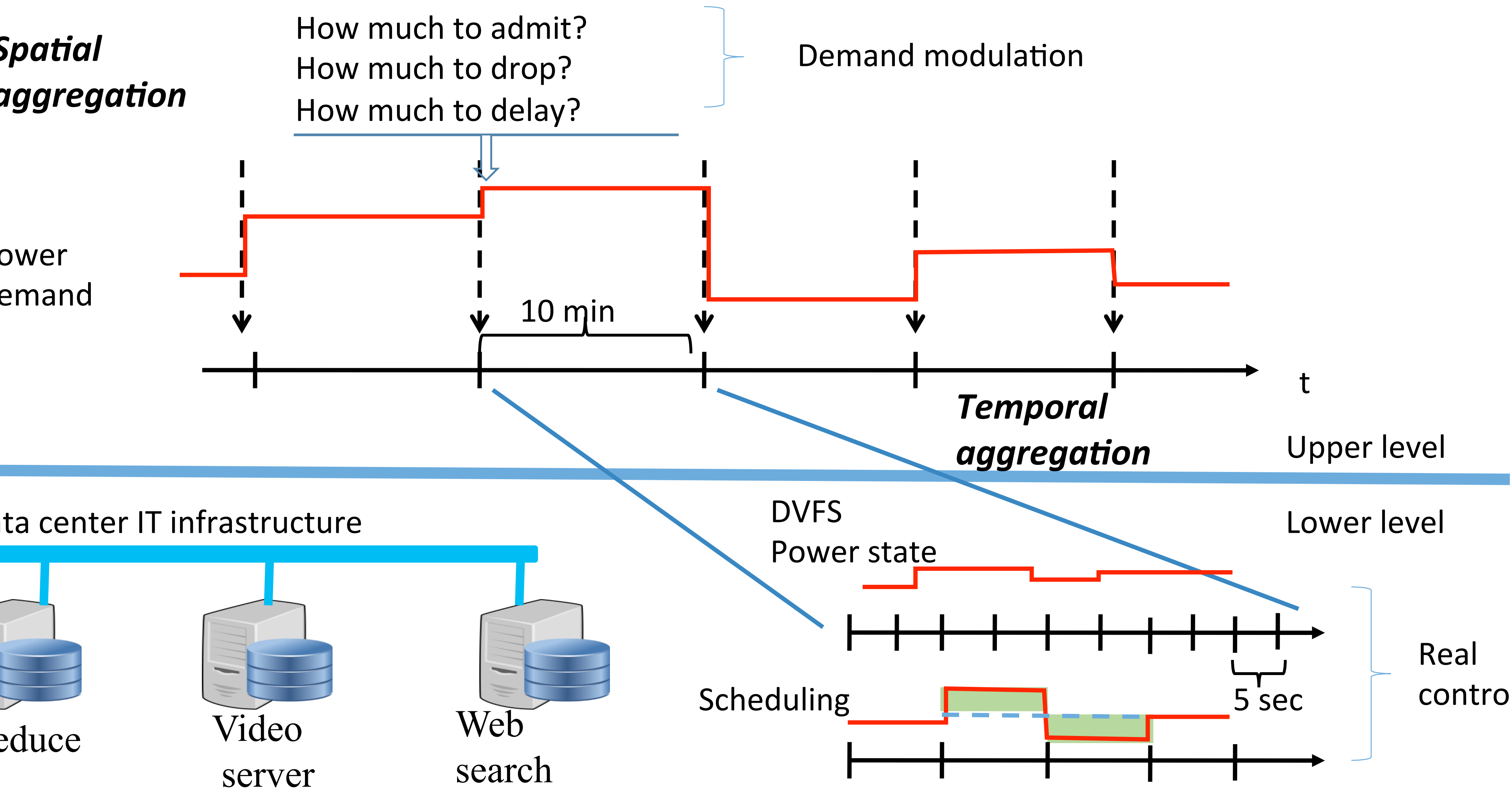
- Cost benefits of demand response via abstract demand dropping and delaying
- From abstract control to real control: A case study

# A Hierarchical Demand Response Framework



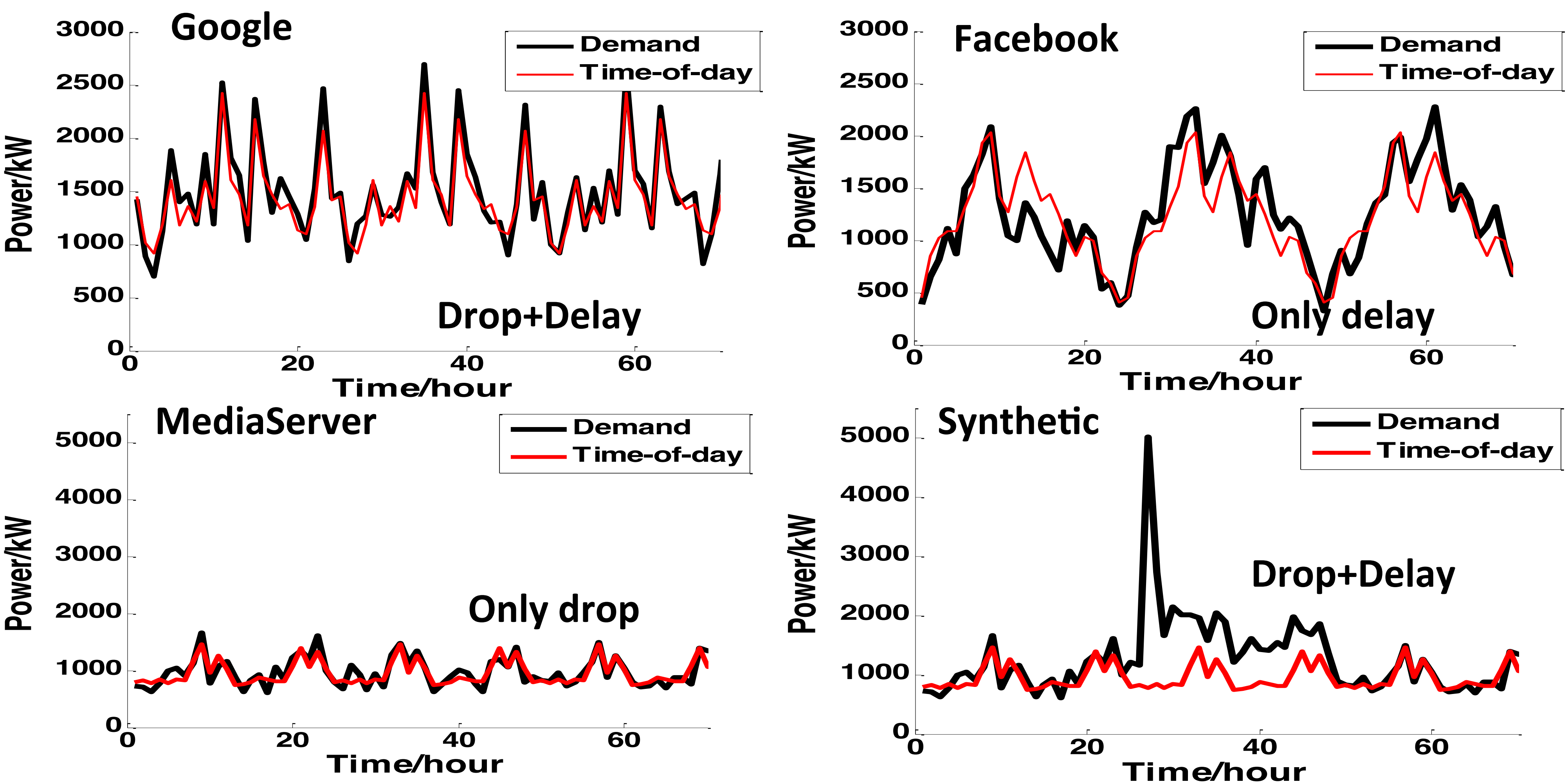
# A Hierarchical Demand Response Framework



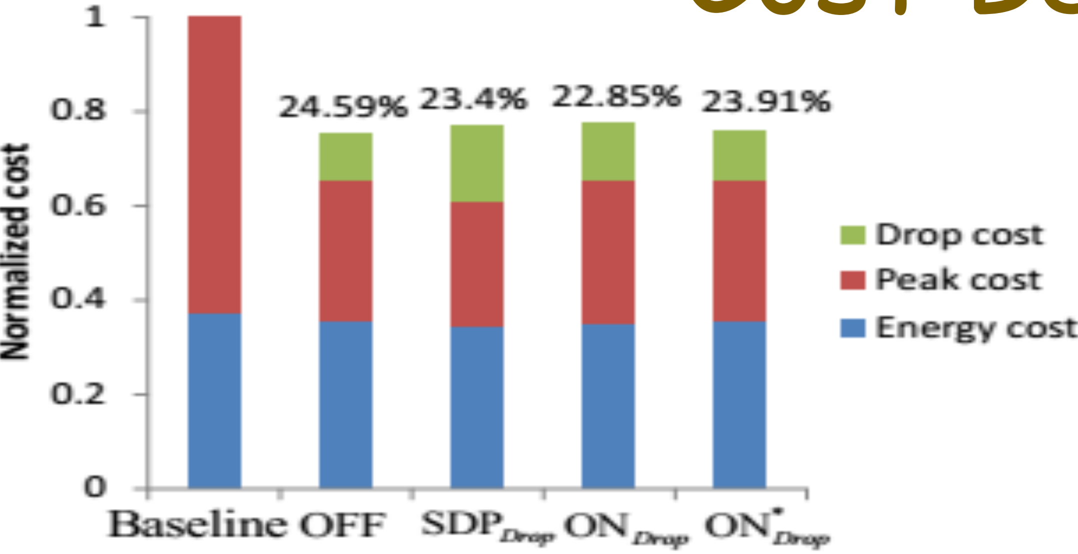
Traces

- Google, Facebook, MediaServer, Synthetic

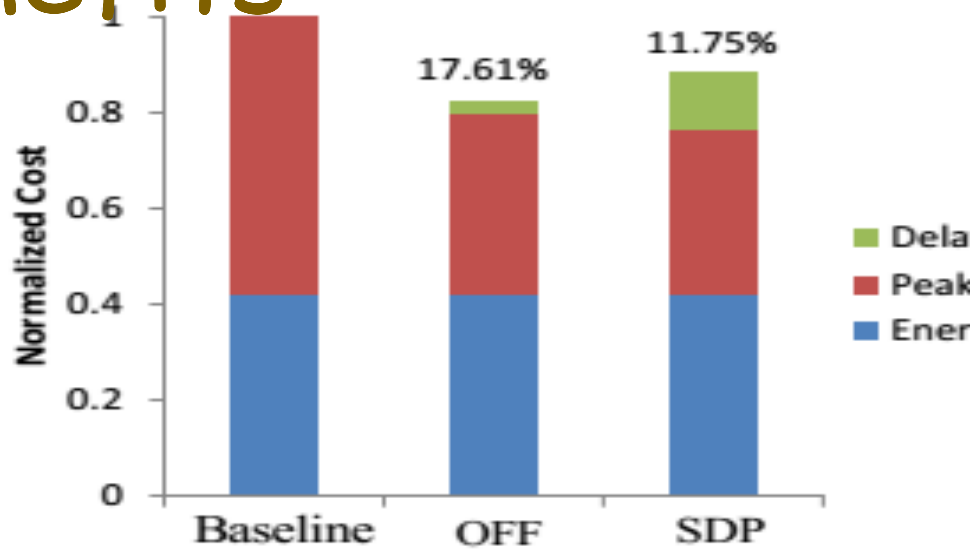
eak-based pricing



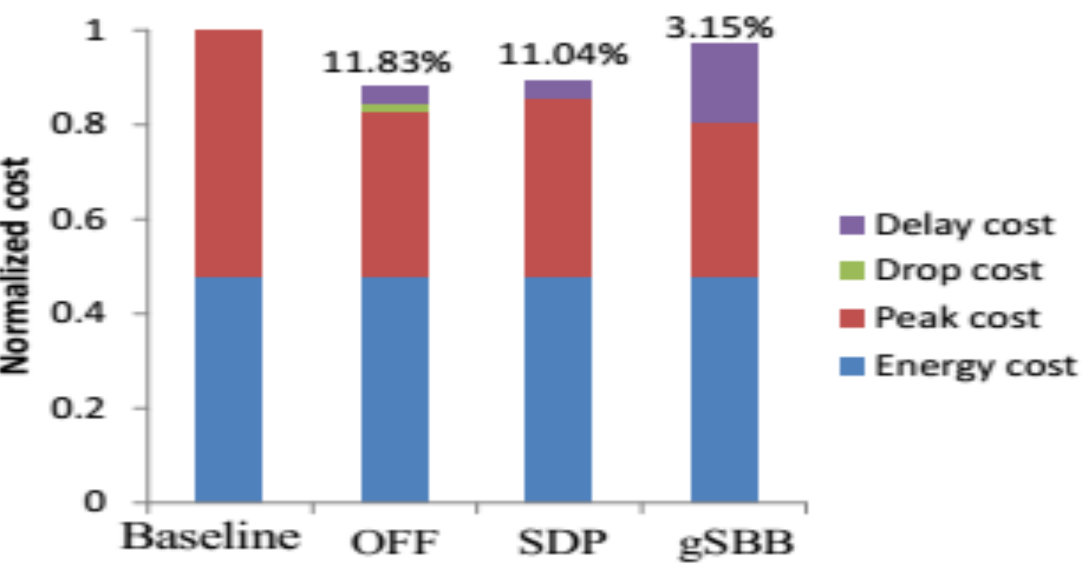
# Cost Benefits



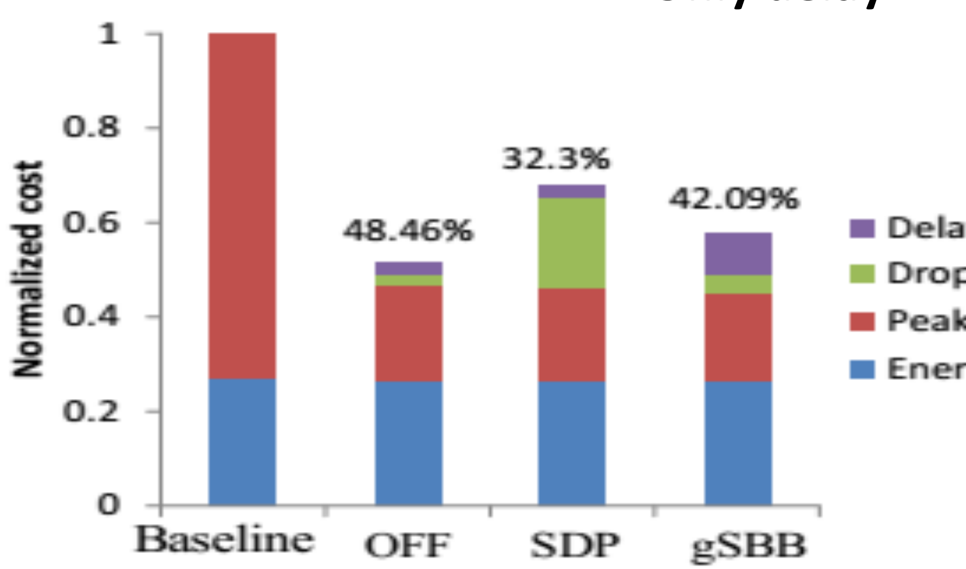

**(a) MediaServer** **Only drop**




**(b) Facebook** **Only delay**

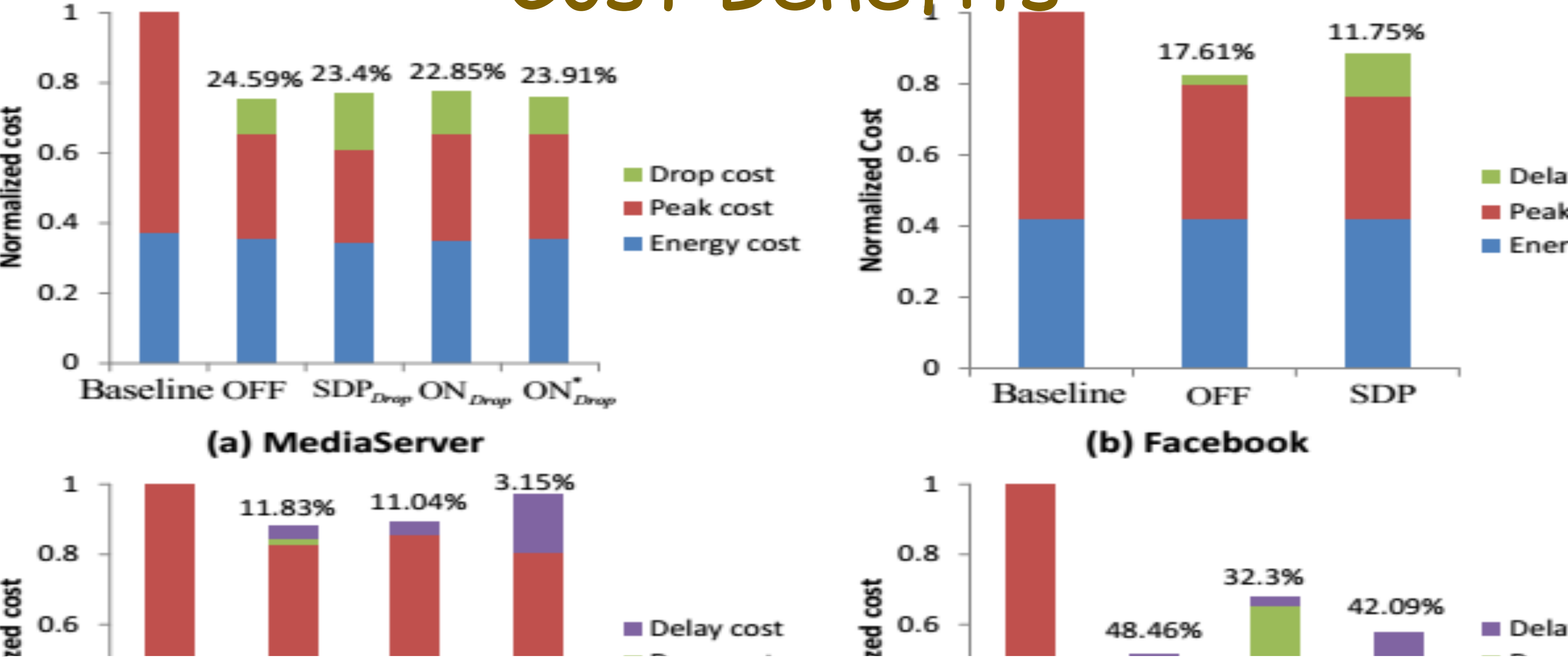



**(c) Google** **Drop+Delay**



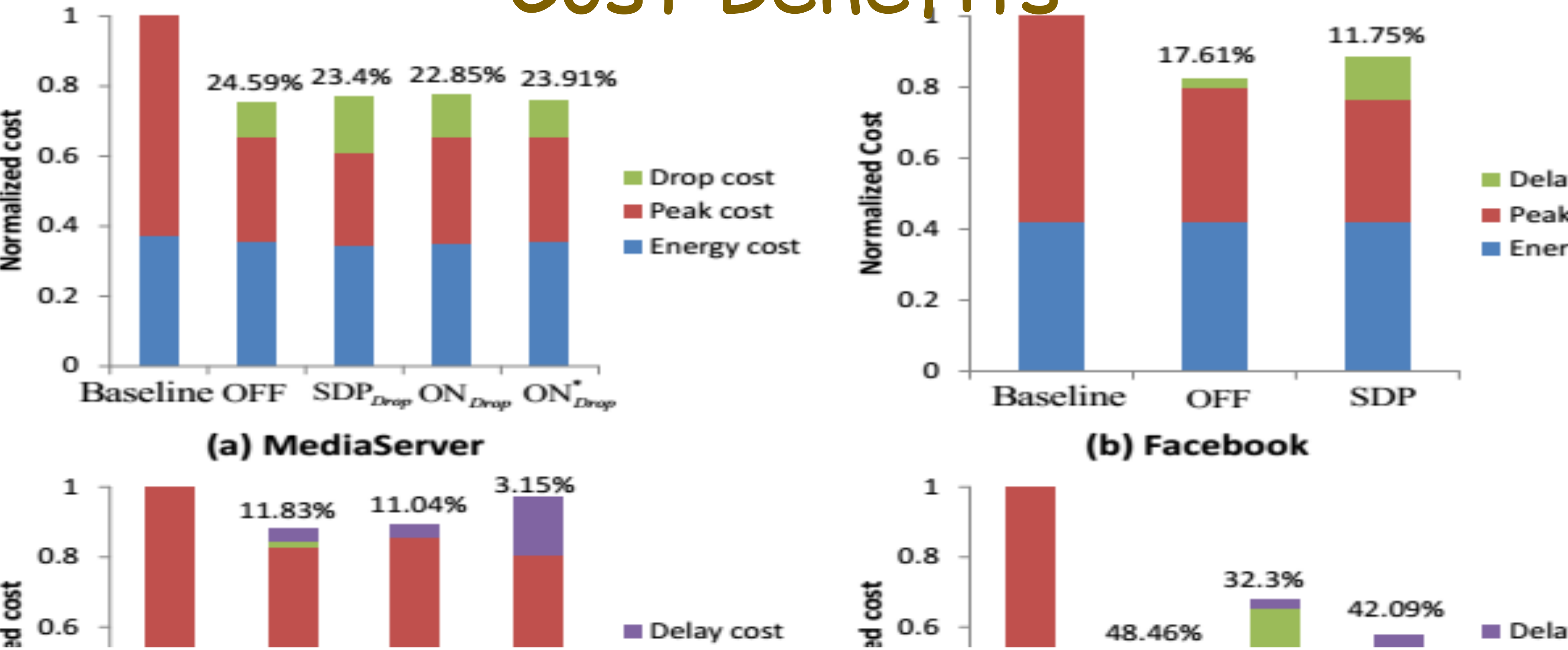

**(d) Synthetic** **Drop+Delay**

# Cost Benefits




**Under peak-based pricing, our approaches provide significa cost savings for real-world workloads w/o losing much "raw demand.**

# Cost Benefits



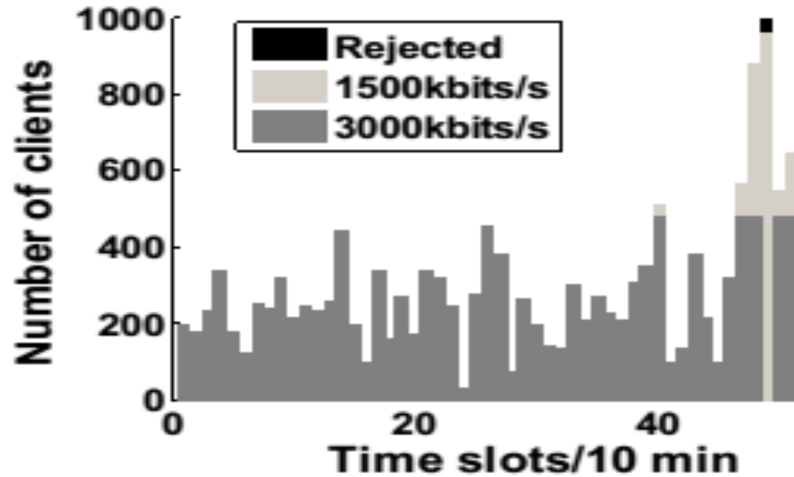

***SDP*** **and** $ON_{Drop}$ **achieve near-optimal cost-saving except for *Synthetic* (flash crowd); gSBB is able to handle workload unpredictability.**

# A Case Study: Media Server

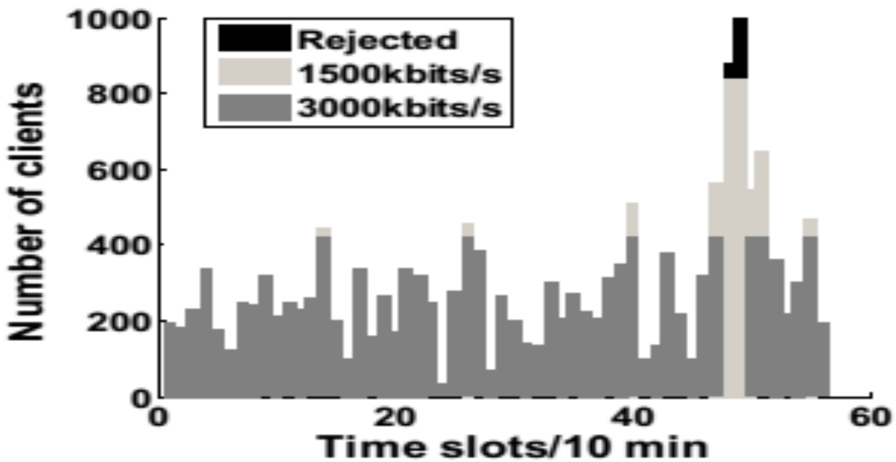
**anslation**.

Degrade QoS (trans. bitrate) to meet dropping cision as much as possible;

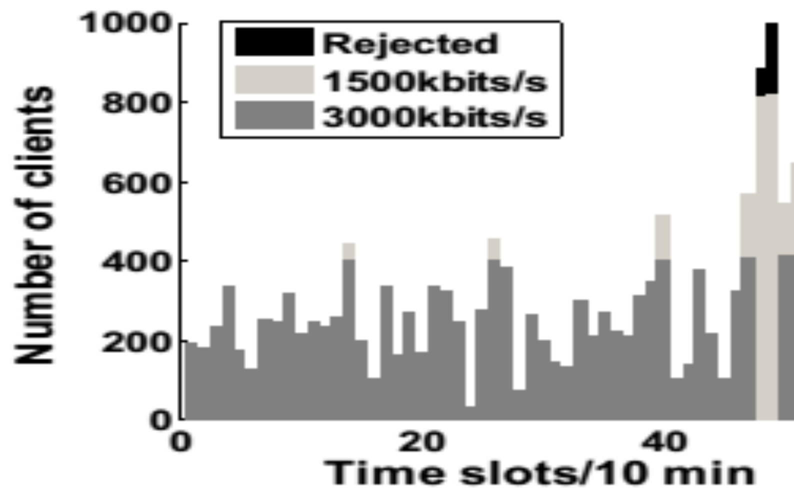
Otherwise, reject some requests to compensate the target dropping power that is unmet.




(a) OFF




(b) $\mathrm{SDP}_{Drop}$




(c) $\mathrm{ON}_{Drop}$

# Conclusions and Open Problems

- Peak power draw significantly impacts both cap-ex and op-ex
- Algorithms and empirical case studies from our work on such optimization using IT knobs
  - Key idea: Abstract myriad IT knobs as dropping or delaying power demand at the cost of performance/revenue loss
  - Results for both adversarial inputs and stochastically known inputs
- Plenty of scope for more work (both theoretical and empirical) on op-ex optimization for peak-based pricing schemes, e.g.,:
  - Competitive analysis for real-time pricing using batteries for DR
  - Competitive analysis for peak-based pricing using IT knobs and/or batteries when using both "dropping" and "delaying" of power demand
- More details at: http://www.cse.psu.edu/~bhuvan

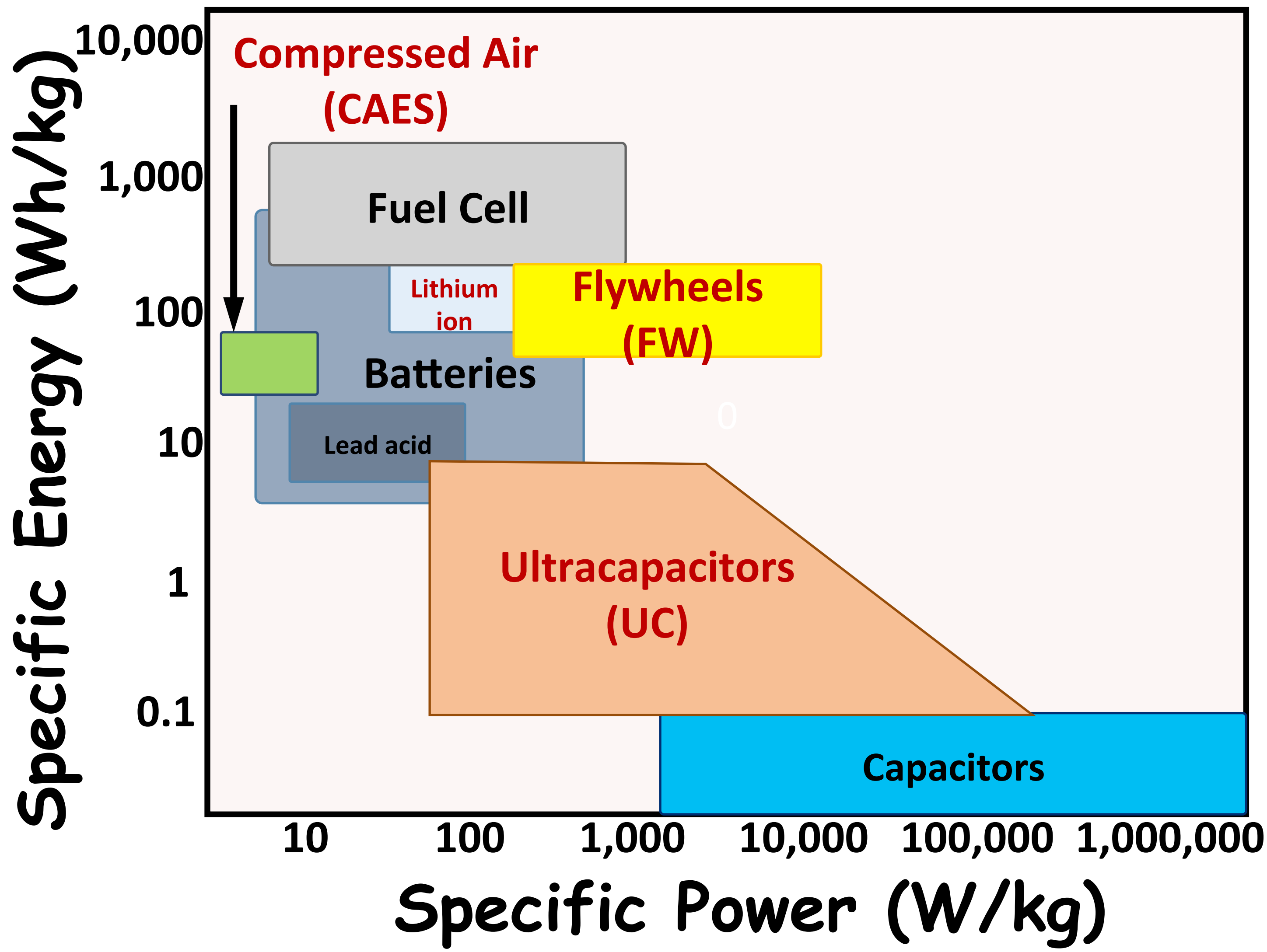
# ESDs in Current Datacenters

Cost

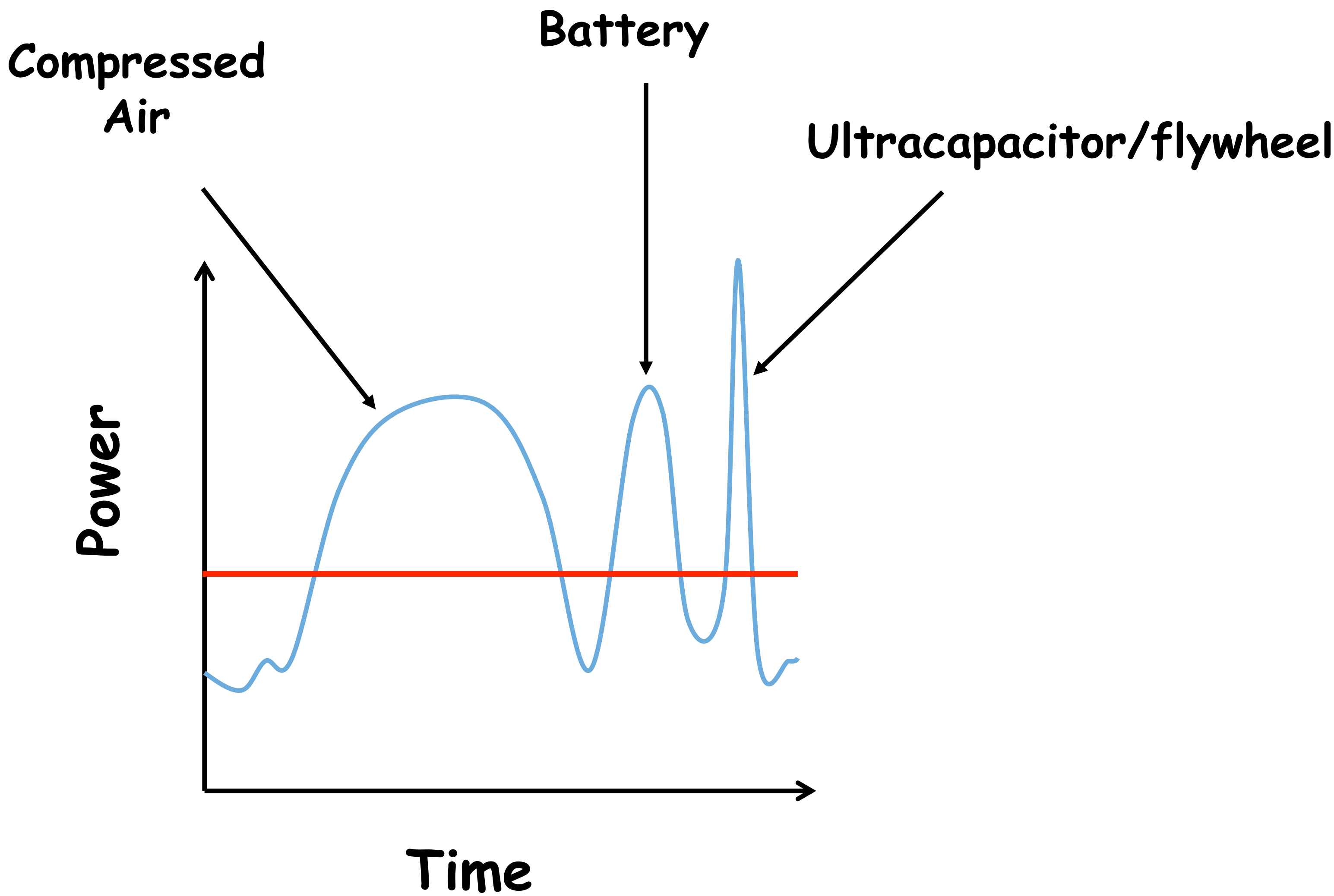
Why restrict ESDs to any one level of the datacenter power hierarchy (e.g., central or server)?

Why restrict to single ESD technology (e.g., Lead acid battery)?

# Ragone Plot


Compressed Air
(CAES)
Fuel Cell
Lithium ion
Flywheels
(FW)
Batteries
Lead acid
Ultracapacitors
(UC)
Capacitors
10,000
1,000
100
10
1
0.1
10
100
1,000
10,000
100,000
1,000,000
Specific Energy (Wh/kg)
Specific Power (W/kg)

# Hybrid ESD solution may be desirable



# Multi-level Multi-technology ESDs