Restructuring the retail market to include load flexibility

Anna Scaglione

Arizona State University
anna.scaglione@asu.edu

Student: Kari Heirsonn
PostDoctoral students: Dr. Mahnoosh Alizadeh, Dr. Masood Parvania

June 8, 2016

Project objective

Research Theme

Remove barriers for the large scale integration of Demand Response (DR) in Power Systems Operations

Main results

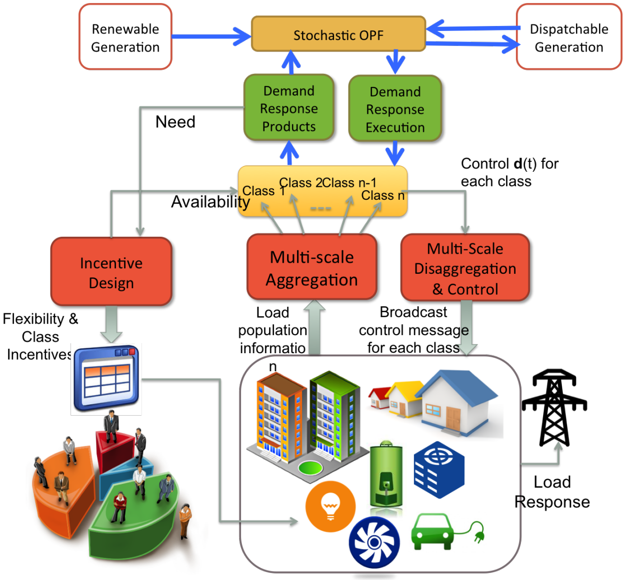
- Large-scale population model-predictive scheduling
- An economic retail mechanism to incentivize customers to *sell flexibility*
- Price responsive Electric Vehicle charging $\rightarrow$ coupling of grid and transportation system
- Valuation of ramping capability at the Energy Market

Recent results

Questions we approached the last quarter

- What kind of Reserve Capacity is DR?
  Comparison with storage
- How would a for profit Aggregator behave in the market?
  The need for retail competition, benefits of differentiated pricing
- Pricing for ramp?
  Continuous time price for electricity (...will skip for lack of time)

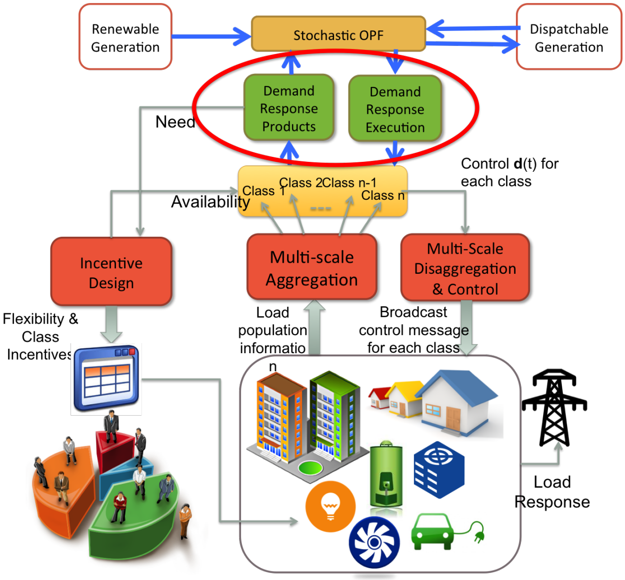
Vision



Renewable Generation
Stochastic OPF
Dispatchable Generation
Demand Response Products
Demand Response Execution
Need
Class 1
Class 2
Class n-1
Class n

Availability
Control **d**(t) for each class
Incentive Design
Multi-scale Aggregation
Multi-Scale Disaggregation & Control
Flexibility & Class Incentives
Load population informatio n
Broadcast control message for each class
Load Response

Understanding DR products



Aggregate DR Abstractions

- Battery Model
  - e.g. [Lambert,'06], [Lamadrid, '11],[Papavasiliou, '10],[Nayar'12]...
  - Good for everything, easy to characterize and understand, but coarse
- Dynamical System Model
  - Mostly HVACs, e.g. [Chong, 85],[Callaway, '09],[Koch, '11], [Mathieu '12],[Chassin, '12],[Mein,'14]..
  - somewhat restrictive
- Dynamical System Model with stochastic input + constraints:
  - Deferrable Loads, EVs and HVACs [Alizadeh,'12 - '14]
  - Similar representations but different classes of DR resources remain hard to compare

DR Abstractions

Let $\boldsymbol{L}$ be the power load per unit of time h over a certain horizon:

$$\boldsymbol{L} = [L(t^0), L(t_0 + h), \ldots, L(t_0 + (T-1)h)]$$

New inspiration [Taylor,'14]:

- Flexible trajectories $\boldsymbol{L}$ are confined in convex polytopes $\boldsymbol{\mathcal{L}}$:

$$\boldsymbol{\mathcal{L}} = \{\boldsymbol{L} | \boldsymbol{A}\boldsymbol{L} \leq \boldsymbol{b}\}$$

- Minkowski sums of the individual polytopes:

$$\boldsymbol{\mathcal{L}} \triangleq \boldsymbol{\mathcal{L}}^2 \oplus \boldsymbol{\mathcal{L}}^2 = \{\boldsymbol{L} | \boldsymbol{L} = \boldsymbol{L}^1 + \boldsymbol{L}^2, \boldsymbol{L}^1 \in \boldsymbol{\mathcal{L}}^1, \boldsymbol{L}^2 \in \boldsymbol{\mathcal{L}}^2\}$$

  are the feasible region for the population

Minkowski sums

- In general complex, but tractable (approximately) in the case of general polytopes [Taylor,'14]
- Easy in case of equal parameters $\rightarrow$ our previous work directly computed Minkowski sums, clustering loads based on their features
- Idea: First divide and then conquer
  - Direct aggregation via clustering for single population
  - Minkowski sum of heterogeneous populations

Common Model

- For Electric Vehicles (EVs) and Deferrable Appliances (DAs)

Definition

The **service time** is the minimum time the load needs to receive the required energy. The **slack time** is the time difference between the time appliance can wait and the service time.

Unified model

At each time $t \in \mathcal{T}$ the load state is its current pair (u_r, u_s) where $u_r =$ remaining service time and $u_s =$ remaining slack time

Note: The time resolution for state/actions space is h/m, $m \geq 1$ where h is the time resolution h for load profile

Common state space model

- The state $(u_r, u_s) \in \boldsymbol{\mathcal{U}} = \{0, \ldots, N_r - 1\} \times \{0, \ldots, N_s - 1\}$
- Load i is plugged in (arrives) at a random discrete time t_i^a and is unplugged (departs) at discrete time $t_i^d > t_i^a$ which is its deadline in $\mathcal{T}$
- State evolution for load i The matrix $\boldsymbol{u}^i \rightarrow$ row $\boldsymbol{u}_t^i = (u_r^i(t), u_s^i(t))$
- If $\mathrm{L^i}$ is the service time (in h/m units):

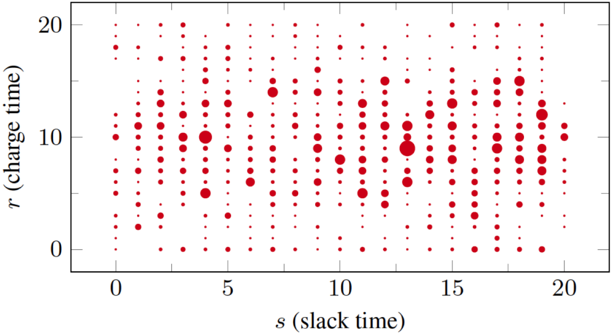
$$\boldsymbol{u}_{t<t^a}^i = 0, \quad \boldsymbol{u}_{t^a}^i = (\mathrm{L^i}\,,\; t_i^d - t_i^a - \mathrm{L^i})$$

L^i for EV is the duration to charge the battery, for DA is simply the length of the load profile vector $\boldsymbol{\ell} = (\ell_1, \ldots, \ell_{\mathrm{L^i}})$

Dimensionless State Space

- With $\boldsymbol{a_u}$ denoting the number of arrivals in state u the state population tensor $\boldsymbol{n} \in \mathbb{N}^{N_r \times N_s \times T}$ dynamics are:

$$\boldsymbol{n_u} = \boldsymbol{a_u} + \sum_{\boldsymbol{u'} \in \mathcal{U}} \boldsymbol{D_{u',u}} - \sum_{\boldsymbol{u'} \in \mathcal{U}} \boldsymbol{D_{u,u'}}$$



Minkowski sum

- The aggregate feasible set

$$\boldsymbol{\mathcal{L}} = \{\boldsymbol{L} \mid \boldsymbol{L} = \sum_{\boldsymbol{u} \in \boldsymbol{\mathcal{U}}} \sum_{\boldsymbol{u}' \in \boldsymbol{\mathcal{V}}(\boldsymbol{u})} R(\boldsymbol{u}, \boldsymbol{u}') \dot{\boldsymbol{D}}_{\boldsymbol{u},\boldsymbol{u}'},$$

$$\boldsymbol{D}_{\boldsymbol{u},\boldsymbol{u}} = \boldsymbol{0} \ \forall \boldsymbol{u} \in \boldsymbol{\mathcal{U}}, \ \boldsymbol{D} \in \mathbb{N}^{(N_r \times N_s)^2 \times T},$$

$$\sum_{\boldsymbol{u}' \in \boldsymbol{\mathcal{U}}} \dot{\boldsymbol{D}}_{\boldsymbol{u},\boldsymbol{u}'} = \boldsymbol{n}_{\boldsymbol{u}} \ \forall \boldsymbol{u} \in \boldsymbol{\mathcal{U}}\}$$

- The only difference between EVs and DAs:
  $\rightarrow$ $\boldsymbol{\mathcal{V}}(\boldsymbol{u})$ = the set of possible moves
  $\rightarrow$ $R(\boldsymbol{u}, \boldsymbol{u}')$ = power per unit time required to change state

A notion of DR Reserve Capacity

Metrics to evaluate our state-space

- The energy requirement that is scheduled to be consumed:

$$\boldsymbol{E} = -\sum_{\boldsymbol{u}\in\mathcal{U}} R(\boldsymbol{u},\boldsymbol{0})\boldsymbol{n_u}.$$

- The aggregated slack that is stored in the system:

$$\boldsymbol{S} = \sum_{\boldsymbol{u}\in\mathcal{U}} u_s \boldsymbol{n_u}.$$

- Maximum ramp down or up for a single period given our schedule $\boldsymbol{D}$:

$$\boldsymbol{L}^+ = \sum_{(\boldsymbol{u},\boldsymbol{u}')\in\mathcal{W}^+} R(\boldsymbol{u},\boldsymbol{u}')\boldsymbol{n_u}, \quad \boldsymbol{L}^- = \sum_{(\boldsymbol{u},\boldsymbol{u}')\in\mathcal{W}^-} R(\boldsymbol{u},\boldsymbol{u}')\boldsymbol{n_u}$$

$$\mathcal{W}^+ = \{(\boldsymbol{u},\boldsymbol{u}')\,|\,u'_r = \max(u_r-1,0),\, u'_s = \max(u_s+u_r-u'_r-1,0)\}$$

$$\mathcal{W}^- = \{(\boldsymbol{u},\boldsymbol{u}')\,|\,u'_s = \max(u_s-1,0),\, u'_r = \max(u_r+u_s-u'_s-1,0)\}$$

The set $\mathcal{W}$ denotes the most favorable moves from any point $\boldsymbol{u}\in\mathcal{U}$ with respect to either of the objectives.

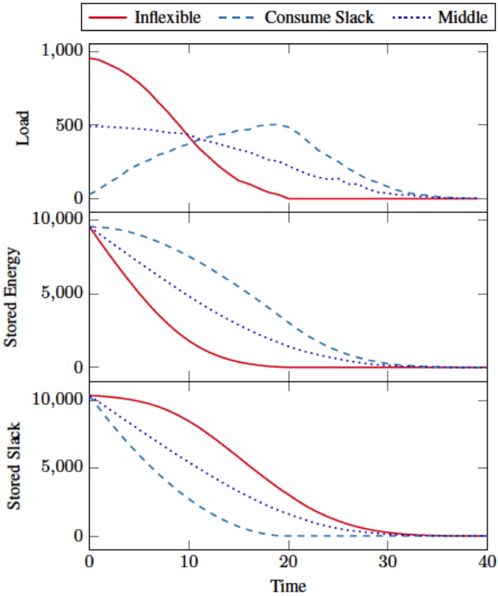
Strategies

Boundaries of the DR reserve (no new arrivals $\boldsymbol{a} = \mathbf{0}$):

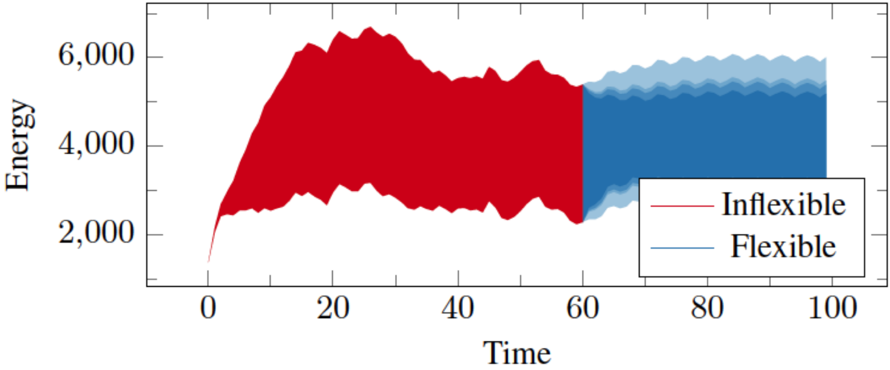
Policy 1: Move along the r axis till zero, then along the s dimension like inflexible loads.

Policy 2: Move everyone along the s axis and delay service as long as possible.

Middle strategy: maximizes the minimum of $\boldsymbol{L}^{+} - \boldsymbol{L}$ and $\boldsymbol{L} - \boldsymbol{L}^{-}$, leaving the most potential for short-term deviations up or down

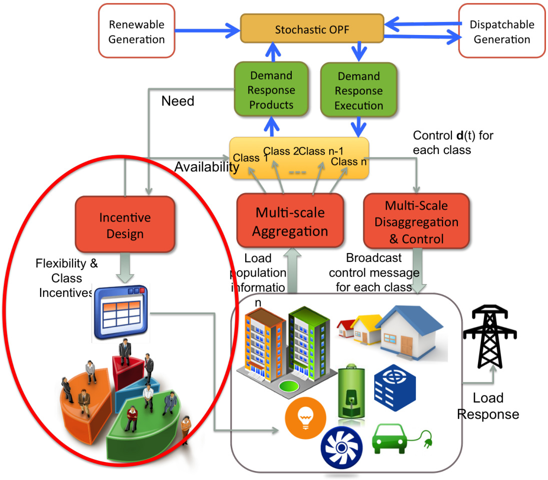


Stochastic Infeed



The (negative) energy stored in the system bounded by Policy 1 and 2. Red region ($0 \leq t \leq 60$) past and present, blue regions ($t > 60$, 95%, 90% and the 50%percentiles) estimated future behavior of the bounds given non stationary Poisson arrivals $\boldsymbol{a}$.

Retail market deregulation: a menace?



Retailers can design their own dynamic retail pricing tariffs to shift consumer load to times with cheaper wholesale prices

Retail markets are imperfect!

- Empirically, we see retail prices do not converge to wholesale prices and are much higher than regulated rates [Puller, 2013] [Puller, 2015]
- Retailer's goal = maximize profit + retain customers in long run

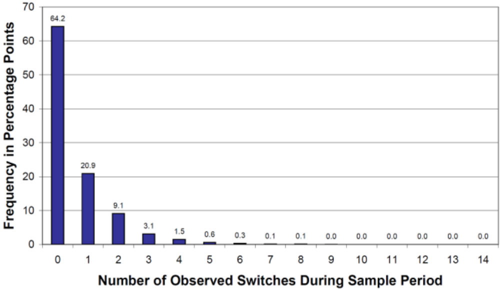


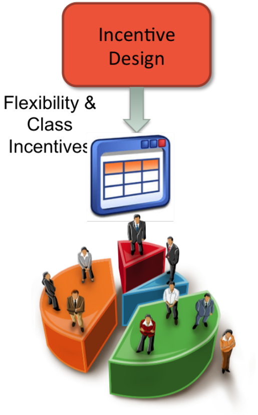
Figure 6: Frequency of Switches Per Household

Our claim

Aside from market inefficiency, this can result in operational reliability concerns for the grid since load can be **shifted away from supply**

Analysis elements

- Model end-users response to prices
- Model retailers' price design problem under 2 scenarios
- Benefits of differentiated pricing over dynamic retail pricing



Customer price response model

- Rational price-takers for daily operations + each day is 2 epochs
- Retail price = (p_1, p_2)
- Each customer i is characterized by a triplet $(d_{i,1}, d_{i,2}, w_i) \in \mathbb{R}^3_+$
- Customer load at $i = 1, 2$

$$L_{i,t} = d_{i,t} + \delta_{i,t}, \quad t = 1, 2$$

where

$$(\delta_{i,1}, \delta_{i,2}) \succeq 0, \quad \delta_{i,1} + \delta_{i,2} = w_i \quad (*)$$

- Customers minimize their cost over each two epoch run:

$$\begin{aligned} \min_{\delta_{i,1}, \delta_{i,2}} \quad & p_1 L_{i,1} + p_2 L_{i,2} = p_1(d_{i,1} + \delta_{i,1}) + p_2(d_{i,2} + \delta_{i,2}) \\ \text{s.t.} \quad & (*) \end{aligned}$$

which leads to a discrete price-response model:

$$(\delta_{i,1}, \delta_{i,2}) = \begin{cases} (w_i, 0), & \text{if } p_1 \leq p_2, \\ (0, w_i), & \text{if } p_1 > p_2, \end{cases}$$

with preference given to epoch 1 in case of a price tie.

Aggregate customer load

- Customer population $i \in \mathcal{N}$
- Given a certain ask price (p_1, p_2), the electricity demand of the entire customer population is:

$$(L_1, L_2) = \begin{cases} \left(\sum_{i \in \mathcal{N}} d_{i,1} + w_i, \sum_{i \in \mathcal{N}} d_{i,2}\right), & \text{if } p_1 \leq p_2, \\ \left(\sum_{i \in \mathcal{N}} d_{i,1}, \sum_{i \in \mathcal{N}} d_{i,2} + w_i\right), & \text{if } p_1 > p_2, \end{cases}$$

- Analogous to that of one large customer with load parameters

$$D_1 = \sum_{i \in \mathcal{N}} d_{i,1}$$
$$D_2 = \sum_{i \in \mathcal{N}} d_{i,2}$$
$$W = \sum_{i \in \mathcal{N}} w_i$$

Retailer profit model

- Wholesale market prices for the two epochs $\rightarrow (s_1, s_2)$
- Assumptions: access to perfect forecasts of (s_1, s_2) + price-taker + all fixed costs sunk and labor costs constant
- Retailer profit is:

$$\pi(p_1, p_2) = (p_1 L_1 + p_2 L_2) - (s_1 L_1 + s_2 L_2),$$

- Inserting the price response of customers:

$$\pi(p_1, p_2) = \begin{cases} (p_1 - s_1)(D_1 + W) + (p_2 - s_2) D_2, \\ \qquad\qquad\qquad\qquad \text{if } p_1 \le p_2, \\ (p_1 - s_1) D_1 + (p_2 - s_2)(D_2 + W), \\ \qquad\qquad\qquad\qquad \text{if } p_1 > p_2, \end{cases}$$

- The retailer cannot just simply solve $\max \pi(p_1, p_2)$

Scenario #1

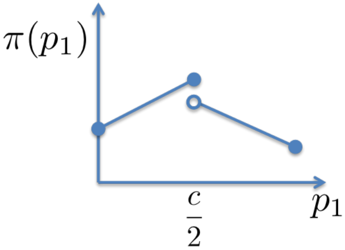
- Customers switching decisions a function of average daily prices rather than instantaneous prices or the bill
- Representative of customers with variable daily load [Puller,2015]
- Retailer designs retail prices such that their average is the same as that of the competitors:

$$p_1 + p_2 = c$$

Scenario #1

- Without loss of generality, assume $D_2 > D_1$.
- Retailer profit is a piece-wise linear function with discontinuity at $\frac{c}{2}$:

$$\pi(p_1) = \begin{cases} \underbrace{p_1(D_1 - D_2 + W) - s_1(D_1 + W) + (c - s_2)D_2}_{E(p_1)}, & \text{if } p_1 \le \frac{c}{2} \\ \underbrace{p_1(D_1 - D_2 - W) - s_1 D_1 + (c - s_2)(D_2 + W)}_{T(p_1)}, & \text{if } p_1 > \frac{c}{2} \end{cases}$$



Scenario #1

- If $W < D_2 - D_1$ (recall $D_2 > D_1$):

$$\max_{p_1} \pi(p_1) = \max\left\{ \underbrace{E(0),}_{\text{load shifted to epoch 1}} \quad \overbrace{T\left(\frac{c}{2}+\varepsilon\right)}^{\text{load shifted to epoch 2}} \right\},$$

$$p_1^{\text{opt}} \in \left\{0, \frac{c}{2}+\varepsilon\right\}, p_2^{\text{opt}} = c - p_1^{\text{opt}}$$

Danger!

Suppose $s_2 < s_1$. Demand is shifted away from the cheap whole-sale market supply iff

$$E(0) > T\left(\frac{c}{2}+\varepsilon\right) \Rightarrow 0 < W < \frac{(3c/2 - 2s_2 - \varepsilon)D_2 - (2s_1 - c/2 - \varepsilon)D_1}{c/2 - s_2 + s_1 - \varepsilon}$$

Intuition $\rightarrow$ can make more money by overcharging for inelastic load

Scenario #2 - Aggregate revenue constraint

- Customers switch to minimize average bill
- If keeping all customers has the same value to the retailer, he will keep average customer bill at competitive industry standard b

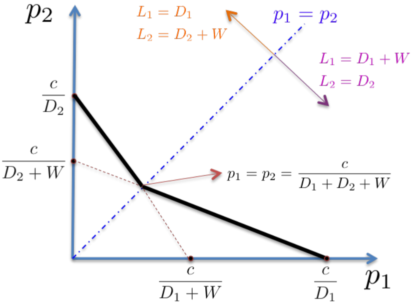


Figure: The set of feasible retail price pairs (p_1, p_2) shown by a black line.

$$p_1 L_1 + p_2 L_2 = |\mathcal{N}| b, \qquad \text{profit} = |\mathcal{N}| b \text{ - wholesale market costs.}$$

Scenario #2 - Discussion

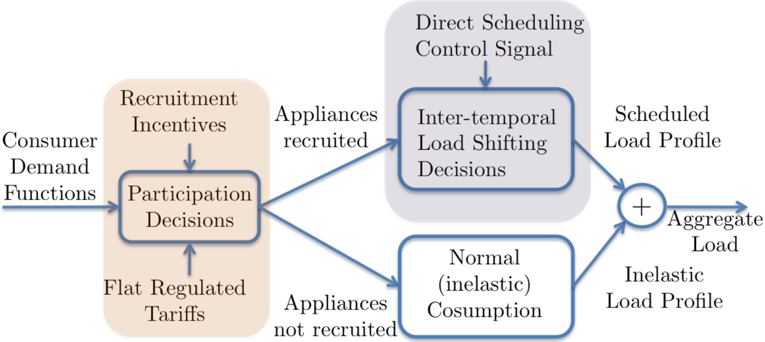
- If the revenue cannot change the retailer has no specific incentive to do anything sensible!
- Approximation of one large customer here is too extreme
- $(d_{i,1}, d_{i,2}, w_i)$ vary across individual customers
- A shared price (p_1, p_2) can lead to different bills for individuals
- Retailers may decide to operate in the high wholesale cost regime and accept the (comparatively small) short-run loss in profit in return for making its biggest revenue generating customers happy

Should we keep retail markets regulated to ensure reliability then?

The need for differentiated pricing

- Differentiate loads based on flexibility → discount for more flexibility + direct control
  see, e.g., [Kefayati an Baldick, 2011], [Bitar and Low, 2012], [Alizadeh et al, 2013]

Challenges = more monitoring and control
Benefits = ↑ reliability + ↓ incentive to overcharge inelastic load



Monopolistic setting - flexibility discount

- Rate = bundle (f, p)
  = (regulated flat tariff, discount for flexibility)
- How much flexible load to offer to the retailer for control?

$$\omega^*(p) = \max_{0 \leq \omega \leq W} U(\omega) = p\omega - I(\omega).$$

- The retailer's profit just from serving flexible loads is:

$$\pi(p) = (s_2 - s_1)^+ \omega^*(p) - p\omega^*(p),$$

  Retailer solves for $\max_p \pi(p)$
  - No short-run incentives to shift load in the opposite direction of supply availability,
  - No long run incentives to favor a subset of customers over others
  - Challenge: in the long-run, measuring the average daily wholesale costs without load shifting to fix flat tariff

Differentiated pricing with competitive bundles

- Each retailer posts bundle (f_i, p_i)
- The profit of each retailer:

$$\pi_i(f_i, p_i | f_{-i}, p_{-i}) =$$

$$\text{choice indicator} \times \text{(flat billing revenue - wholesale costs - discounts)}$$

$$= b_i(f_i, p_i | f_{-i}, p_{-i}) \big[(D_1 + D_2 + W) f_i - s_1 (D_1 + W) - s_2 D_2$$

$$+ \left((s_2 - s_1)^+ - p_i \right) \omega^*(p_i | b_i^* = 1) \big]$$

$b_i(f_i, p_i | f_{-i}, p_{-i})$ is a binary indicator function of service selection

Differentiated pricing with competitive bundles

- To decide the value of b_i, the customer solves:

$$i^* = \text{argmin}_i \ (D_1 + D_2 + W)f_i - U_i(\omega^*(p_i | b_i^* = 1)),$$

b_{i*} is set to 1, while for all $i \neq i^*$, $b_i = 0$.

The equilibrium in an ideal setting:

$$f^* = \frac{s_1(D_1 + W) + s_2 D_2}{D_1 + D_2 + W}, \ p* = (s_2 - s_1)^+$$

$\rightarrow$wholesale prices passed on to customers like ideal dynamic pricing

Important property

Market non-idealities don't affect reliability because the billing and control modules are separate

Conclusions

- New mechanisms to add more degrees of freedom in our infrastructure
- Such mechanisms will be need to:

1. make financial sense in terms of initial investment and operational costs
2. should be backward compatible with most existing physical assets and practices
3. not threaten system reliability
4. their reliable performance should not depend on unrealistic assumptions such as perfect knowledge of customer behavior, social welfare maximizing private firms, or perfect competition.

We argued that under imperfect retail competition, differentiated pricing is preferable over regular dynamic pricing tariffs for tapping into load flexibility while protecting reliability