Multi-armed Bandit Based Tariff Generation Strategy for Multi-Agent Smart Grid Systems

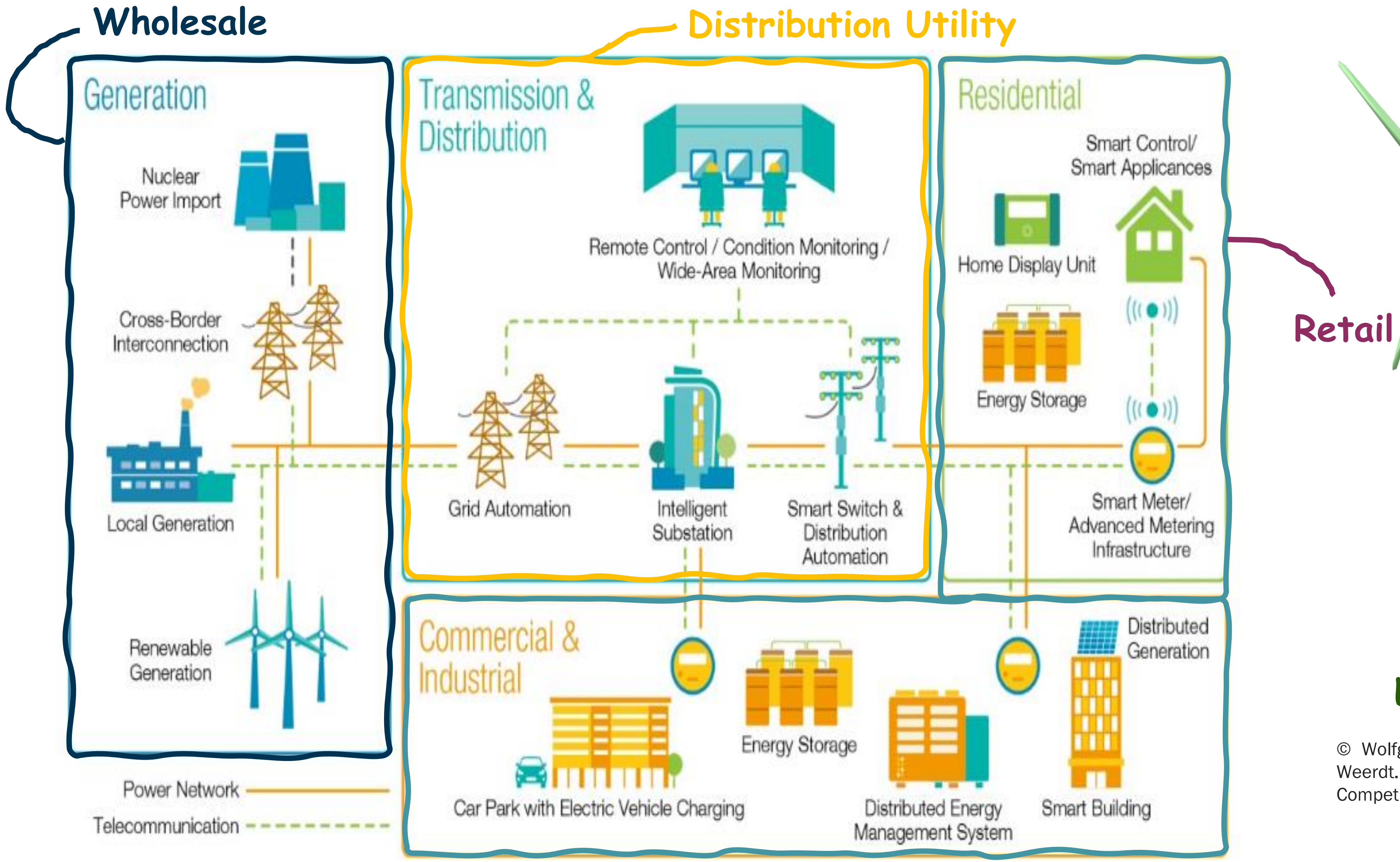
Sanjay Chandlekar[1], Easwar Subramanian[2] and Sujit Gujar[1]

1 INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY HYDERABAD

2 tcs TATA CONSULTANCY SERVICES

EMAS 2023

Introduction



Energy Brokers

© Wolfgang Ketter, John Collins, Mathijs de Weerdt. The 2020 Power Trading Agent Competition.

© "Overview of Smart Grid Technology And Its Operation and Application". https://www.elprocus.com/overview-smart-grid-technology-operation-application-existing-power-system. Accessed 7 June 2022.

Note that,

- Having all the customers under broker's portfolio doesn't lead to the highest profit
  - Having more and more customers under the portfolio increases broker's revenue; however …
  - The higher market share attracts higher peak-demand penalties and grid-imbalance penalties
  - Such penalties are proportional to broker's market-share and huge monetary loss for broker
  - Thus, broker should not aim for full market-share

Research Questions

1. What is the optimal customer market-share?
2. How to offer tariffs to achieve and maintain such optimal market-share?

Previous Work and Our Contribution

Previous Work

- In our previous work in IJCAI'22, we designed a tariff strategy inspired by the game theory literature that decides the optimal market share for various player configurations
- Using the knowledge of the optimal market-share, we constructed a heuristic-based techniques to achieve and maintain the optimal market share during the simulation
- This strategy was deployed in our broker VidyutVanika, which helped us to win the annual PowerTAC tournaments in 2021 and 2022 by maximizing our revenues in the tariff market
- In this work, we aim to follow a more scientific approach to offer tariffs to achieve and maintain the optimal market share, essentially replacing our heuristics-based strategy with a learning-based strategy to achieve similar performance

Game Theoretical Analysis (5 Player Game)

VV / Opp	(TT, VV18, VV20,C)	(TT, VV18, VV20,A)	(TT, VV18, A,C)	(TT, A, VV20,C)	(A, VV18, VV20,C)
0%	-0.893	-0.298	-0.169	-0.156	1.737
15%	-0.199	-0.017	-0.205	-0.146	1.581
30%	0.112	-0.049	0.106	0.044	1.898
45%	-0.083	0.041	0.159	0.143	1.808
60%	-0.312	0.027	-0.288	-0.102	1.741
75%	-0.493	-0.228	-0.373	-0.409	1.025
100%	-0.498	-0.561	-0.188	-0.188	0.996

The above utility matrix can be easily solved, and equilibrium strategy can be found.

For 2-Player, 3-Player, and 5-Player configurations, target optimal market shares for GenerateTariffs-EXP3 are 51%, 40.8%, and 32.3%, respectively.

Formula to calculate utility values for $\forall s_i \in S_1$ *and* $\forall s_{-i} \in S_2$

$$u_i(s_i, s_{-i}) = \frac{1}{T}\sum_{i=1}^{T} x_i - \frac{1}{n}\sum_{k=1}^{n}\left(\frac{1}{T}\sum_{i=1}^{T} y_{ik}\right)$$

- x_i denotes the final cash of VV21 in game i
- y_{ik} denotes the final cash of opponent broker k in game i
- n denotes the number of opponent brokers
- T denotes the number of games in the set, T = 5 for the current experiment

Tariff Module GenerateTariffs-EXP3

GenerateTariffs-EXP3

- The proposed strategy is modeled as a Markov Decision Process (MDP) consisting of a tuple <S, A, P, R>
    - State space S
    - Action space A
    - Transition Probabilities P
    - Reward R
- However, the model does not know the transition probabilities. To learn the optimal action in each state (called a policy) in the absence of transition probabilities, we use Contextual MAB techniques along with the EXP-3 algorithm.

State Space [7 States] (Market-share buckets)

OMS: Optimal Market Share

CMS : Current Market Share

$OMS = 0.85 * market\ share\ suggested\ by\ GT\ module$

0 - $|OMS - CMS| <= OMS*0.1$

1 - $(OMS - CMS) > OMS*0.1$ & $(OMS - CMS) <= OMS*0.4$

2 - $(OMS - CMS) > OMS*0.4$ & $(OMS - CMS) <= OMS*0.7$

3 - $(OMS - CMS) > OMS*0.7$

4 - $(-OMS + CMS) > OMS*0.1$ & $(-OMS + CMS) <= OMS*0.4$

5 - $(-OMS + CMS) > OMS*0.4$ & $(-OMS + CMS) <= OMS*0.7$

6 - $(-OMS + CMS) > OMS*0.7$

EMAS 2023

Action Space [5 Actions]

$$New\ Tariff = Old\ Tariff + \texttt{step}$$

Action 0 - $\texttt{step} = 0.0$ [Maintain]

Action 1 - $\texttt{step} = -0.02$ [Lower 1]

Action 2 - $\texttt{step} = -0.04$ [Lower 2]

Action 3 - $\texttt{step} = 0.02$ [Higher 1]

Action 4 - $\texttt{step} = 0.04$ [Higher 2]

EMAS 2023

Reward

```
If        |OMS - CMS| <=  0.05    -->   reward = 1.00
else If   |OMS - CMS| <=  0.20    -->   reward = 0.50
else If   |OMS - CMS| <=  0.35    -->   reward = 0.25
else                              -->   reward = 0
```

EXP-3 Algorithm

EXP3: Exponential-weight algorithm for Exploration and Exploitation

1: Initialize/Load table$[|S|][|A|]$

2: $prob(s,i,t) = (1-\gamma)\frac{table(s,i,t)}{\sum_{a=1}^{|A|} table(s,a,t)} + \frac{\gamma}{|A|}, \forall i \in \{1,2,...,|A|\}$

3: Sample next action act stochastically from $[prob(s,1,t), prob(s,2,t), ..., prob(s,|A|,t)]$

4: Observe reward $r(s,act,t)$ for taking action act in state s at t

5: Update the reward:
$\hat{r}(s,a,t) = r(s,a,t)/prob(s,a,t)$, if $a = act_t$
$\hat{r}(s,a,t) = 0$, otherwise

6: $table(s,i,t+1) = table(s,i,t) * e^{\gamma * \hat{r}(s,i,t)/|A|}, \forall i \in \{1,2,...,|A|\}$

EMAS 2023

PowerTAC: Experiments and Results

Q-Tables

Action / State	Maintain	Lower1	Lower2	Higher1	Higher2
0	33.64	16.99	10.35	28.90	14.85
1	361.41	30.35	11.11	18.30	167.86
2	18.07	4.04	3.59	22.95	32.24
3	2.02	1.74	1.32	3.66	7.94
4	40.02	27.96	19.82	20.95	10.69
5	4.33	7.35	15.81	4.31	2.36
6	1.17	2.46	4.47	1.13	1.32

2-Player Configuration

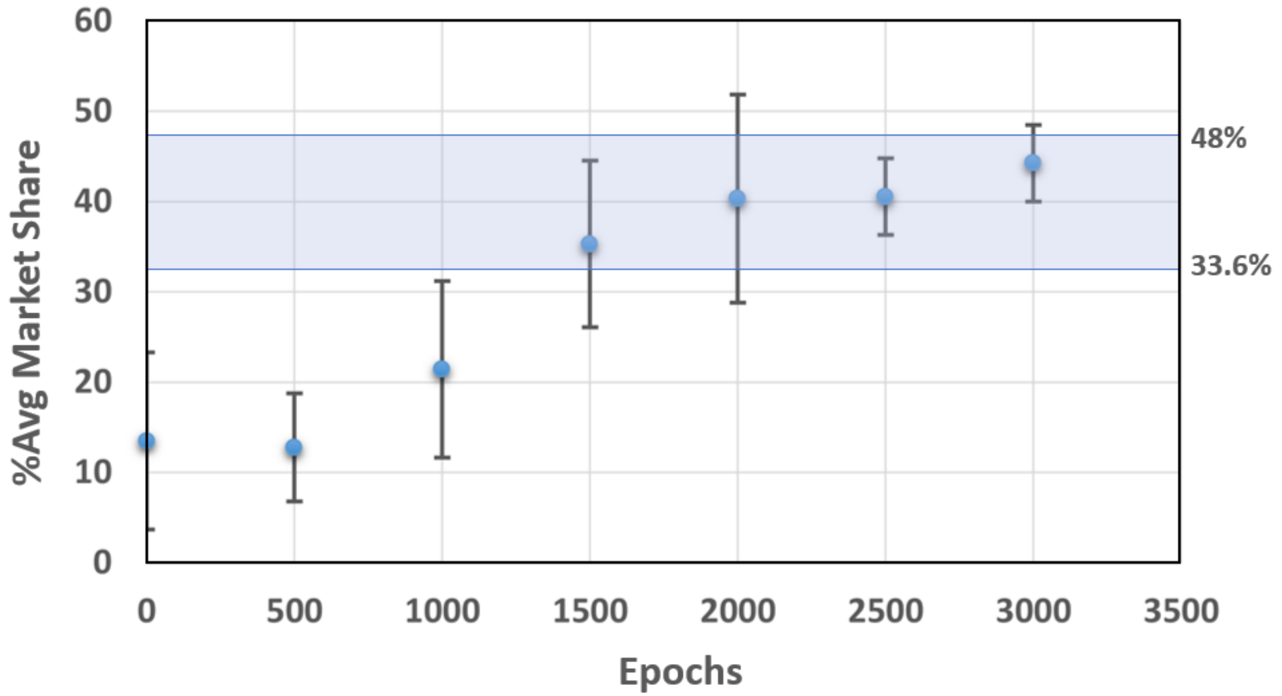
Action / State	Maintain	Lower1	Lower2	Higher1	Higher2
0	31.88	31.01	12.77	31.08	16.02
1	56.31	31.82	9.31	46.91	60.29
2	11.92	5.25	3.81	10.75	35.35
3	4.11	1.20	1.30	4.82	24.37
4	26.51	46.61	49.11	23.80	11.94
5	2.91	6.58	4.56	2.74	1.43
6	1.45	2.68	8.40	1.96	1.32

3-Player Configuration

Action / State	Maintain	Lower1	Lower2	Higher1	Higher2
0	2.30	1.54	1.41	2.00	1.36
1	4.83	2.00	1.47	4.23	2.19
2	2.90	1.45	1.07	6.24	6.40
3	1.72	1.00	1.04	5.85	33.89
4	2.14	5.61	1.78	1.62	1.51
5	1.66	3.60	1.81	1.22	1.10
6	1.36	12.27	10.62	1.44	1.69

5-Player Configuration

Results



Market Share Maintained by GenerateTariffs-EXP3 w.r.t Number of Epochs of Training for 3-Player Configuration

EMAS 2023

Thank you

Contact: sanjay.chandlekar@research.iiit.ac.in