Evaluating Demand Response Programs By Means Of Key Performance Indicators

George Thanos, Marilena Minou, George D. Stamoulis
Network Economics and Services Group,
Athens University of Economics and Business,
76 Patission Str. Athens 11362, Greece
{gthanos, minou, gstamoul}@aueb.gr

Tanuja Ganu, Vijay Arya, Dipanjan Chakraborty
IBM Research−India
Bangalore
India
{tanuja.ganu, vijay.arya, cdipanjan}@in.ibm.com

Jan van Deventer
Department of Computer Science, Electrical and Space Engineering
Lulea University of Technology, Sweden
Jan.van.Deventer@ltu.se

***Abstract*—Demand response (DR) has received significant attention in recent years and several DR programs are being deployed and evaluated worldwide. DR systems provide a wide range of economic and operational benefits to different stakeholders of the electrical power system including consumers, generators and distributors. DR can be achieved through a number of different mechanisms such as direct-load-control, incentives, pricing signals, or a combination of these schemes. Due to the remarkable variation in demand response systems, it becomes a challenge to evaluate and compare the effectiveness of different DR programs holistically. In this work, we define a number of different performance metrics that could be used to evaluate DR programs based on peak reduction, demand variation and reshaping, and economic benefits.**

I. INTRODUCTION

We are part of a society that relies heavily on electrical energy to function. Our appetite for electricity continues to grow leading to different problems, most important of which is the mismatch between electricity generation, distribution and consumption. Electricity demand varies throughout the day and there are times when the demand peaks may exceed the capabilities of generation and distribution. Similarly, the deployment of variable renewable sources such as wind can lead to periods where generation exceeds the consumption and distribution capabilities of the power system. The consequences of these mismatches include inefficient generation, deterioration of grid assets, increased prices & greenhouse gases, and possibly planned and unplanned outages for millions of people lasting several days over a large area.

Traditionally, the mismatch between electricity generation and consumption has been handled by controlling generation. However, this approach, also known as known as "load/demand-following" may not always be feasible, economical, or environmental-friendly. Progress in cheap and high-speed communication between consumption, distribution and generation has enabled the complementary approach of "supply-following", wherein a large number of responsive loads are shifted or curtailed to help handle the mismatch between production and consumption. This idea crystallizes into the concept of Demand Response (DR) systems and several projects all over the world are implementing such systems in different forms and in different domains. DR systems form an important part of the future Smart Grid picture and one of the aims of deploying advanced metering infrastructure is essentially to enable DR [2][3][4][5].

Demand response can be achieved through a number of different mechanisms such as direct-load-control, incentives, pricing signals, or a combination of these schemes. The design of a DR system varies depending on a number of factors such as the type of generation, distribution, consumption, and demography. For example, certain geographical areas may have a large number of rivers or wind to power turbines while other regions may be depending on oil to produce electricity. In the former case, customers might be given incentives to temporarily store excess generation, while in the later time of use prices might be used to discourage consumption when oil prices are high. Additionally, some locations may have a single electricity distributor while others may have multiple distributors and this might determine if a DR system is implemented with the help of an energy aggregator. The electricity consumers may be industrial companies, apartment buildings, or individual houses/villas. In the latter cases, the age and lifestyle of the consumers might have influence the design of DR system [6][7][8][9].

With different types of DR systems and implementations, it is important to understand the performance of proposed and implemented DR solutions. For example, how effective are different communication mechanisms in terms of influencing the consumers to avoid peak load? How efficient are the financial incentives for different stakeholders? How effective are different incentive schemes for certain types of consumers (e.g. residential)? Are the predictions of baseline electricity consumption correct or even possible for certain consumer types? Additionally, the acquisition and communication of information, which is part of the DR solution, brings its own set of problems. The information must be secure, trustworthy, and tamper-proof. The system must not allow information to be misused (i.e., privacy of the stakeholder must be guaranteed). Lastly, the storage of all customer information must be done intelligently, both for efficiency of storage reasons as well as the processing of the data.

In this context, this paper presents different candidate metrics that may be used to evaluate and compare the effectiveness of DR programs. This work is part of the EU FP7 WATTALYST project [1], which aims to understand how consumers respond to DR signals by increasing/decreasing their demands and how their participation is influenced by external and internal factors. Another goal of the project is to understand effective methods of conveying the DR signals to the users. In particular, the project will focus on interface design; communication means (in-house displays, SMS messages), message emphasis (environmental, economical) and customized messages based on gender, age and profile.

II. Background and Motivation

There are various types of DR systems studied in literature [1]. DR systems are mainly divided in two categories – 1) Price Responsive DR, and 2) Controllable/Incentive-based DR. In Price Responsive DR, the end user is exposed to dynamic electricity price. DR programs such as Real Time Pricing (RTP), Critical Peak Pricing (CPP), and Price to Retailer (PTR) belong to this category. In Controllable/ Incentive based DR programs, end consumers alters their demand pattern against certain specified incentives or under a designated agreement with the load serving entity (LSE). Various direct and indirect load control (DLC/ILC) DR programs belong to this category. Some DR programs are also specific to certain consumer segments such as commercial or residential and certain loads types such as air conditioners or washing loads etc. Due to the remarkable variation in demand response systems, it becomes a challenge to evaluate and compare the effectiveness of different DR programs holistically.

Hence it is important to identify DR metrics that can be used to assess the efficiency and economic performance of DR systems. Some of these might be easily quantifiable such as the actual peak reduction, the demand price elasticity (further divided in self-elasticity, cross-elasticity, elasticity of substitution), whereas others may not, such as customer acceptance and their participation rate (dependent to the level of comfort). The choices are numerous and one of the challenges is to make the most appropriate selection and define the metrics and the success criteria of a DR program by taking into account all of the following:

- Environmental, demographical, physical and contextual characteristics such as types of premises and profile of users, weather conditions, national/local characteristics, idiosyncrasies and legislation etc.
- Different energy network architectures, ICT infrastructure, levels of AMI implemented and DR programs available
- The availability (or not) of historical or statistical data
- The gains of different players and the total improvement of economic efficiency (social welfare). Also, the degree of the improvement (comparison) after DR adoption and whether its magnitude can justify the capital investments both from the network and consumer side.

With various aspects to consider and many solutions available, we need to measure the performance of key issues such as

- Effectiveness of the DR programs
- Which designs are most economically effective and sustainable as well as financially rewarding
- Which information media are most effective in influencing the customer
- Models for extracting user profiles from sensor data
- Privacy and security aspects of conducting such evaluations

In this paper, we discuss these issues and outline, wherever appropriate, some of the candidate measurement indices for use in DR evaluation. We first describe some basic primitives. Thereafter, we discuss indices used to evaluate the DR programs. Lastly, we cover indices that are relevant for large-scale deployment of DR programs.

III. KPIs For Evaluating DR Programs

Although DR is one of the important approaches to match demand, production and distribution, its design and implementation remains ambiguous. With different aspects to consider and many possible solutions, we need to measure the performance of key issues in various aspects as previously discussed. In the next subsections we will focus and elaborate on a number of performance metrics, namely: basic KPIs for measuring electricity consumption and peak reduction, KPIs for measuring the demand variation and reshaping after DR is employed, economically-related KPIs and briefly outline some of the functional ones.

A. Primitive KPIs

Energy consumption is assessed by measuring power consumption over time: $e = \int_{t_i}^{t_f} p(t)\partial t$

where e is the energy consumption (Joules) and $p(t)$ is the power delivered (Watts). t_i an t_f are the initial and final times of the period of measurement in seconds. This period can be an hour, a day, a week, a month or a year depending on the interest.

Average power over the period is then: $\bar{p} = \frac{e}{t_f - t_i}$

Multiplying average power by 1 second leads to the energy consumed in Joules. Dividing the latter with 3600000 provides the average energy consumed in kWh.
The power variance over that period is:

$$\sigma_p = \frac{1}{(t_f - t_i)} \sqrt{\int (p(t) - \bar{p})^2 \partial t}$$

The variance indicates how much the power consumption has differed from the average. When the load is distributed smoothly over time, the differences between the peaks of the power consumption *p(t)* and the average are small, leading to smaller variance.

In a demand and supply setting, prediction is necessary to attempt match production and demand. A key performance index for the prediction can be the variance σ_Δ of Δ_{PA}, where Δ_{PA} is the difference between the predicted power consumption p_p and the actual or real one p_a.

$$\sigma_\Delta = \frac{1}{(t_f - t_i)} \sqrt{\int (\Delta_{PA}(t) - \bar{\Delta}_{PA})^2 \partial t}$$

The bigger this variance is, the worse is the prediction. It is interesting to monitor this index with and without DR feedback to the consumers.

B. KPIs for Peak reduction Quantification

Generally, the main goal of DR programs is to reduce the peak demand. This can be measured via the following metrics:

- Change in the total electricity consumption per day. The original consumption is measured before starting the DR program and the new consumption is measured after the program inception. This is measured as

$$\frac{orig.consumption - new.consumption}{orig.consumption}$$

- Change in total electricity consumption during the peak hours. This is measured as

$$\frac{orig.peak.consumption - new.peak.consumption}{orig.peak.consumption}$$

- Change in total electricity consumption during the off-peak hours. This is measured

$$\frac{orig.offPeak.consumption - new.offPeak.consumption}{orig.offPeak.consumption}$$

C. KPIs for Demand Variation Analysis and Demand Reshaping

The demand variation is defined as the result of the subtraction/addition of the real demand to the baseline. The real demand is measured via smart metering techniques but the baseline is always estimated since it reflects the demand that should have occurred if no DR message had been sent. A consumer may also be rewarded proportionally based on the difference between actual and baseline consumption.

A calculation method of the baseline is therefore required for every DR event within a program. But the accuracy of different baseline calculation methods obviously varies, even when the same method is applied to different groups of customers. This leads to an inaccuracy of DR event performance indicators, which can make the comparison of results between different DR programs difficult.
Generally, the main goal of DR programs is to reduce the electricity demand during the peak load hours and to increase the demand during off-peak hours so that electricity demand and supply can be matched[1]. The hours during which DR messages are sent are called event periods. Depending on the concrete type of DR programs, the event period can include the whole day (RTP-Real Time Pricing, TOU-Time Of Use tariff), high market price hours (CPP-Critical Peak Pricing, CPR-Critical Peak Rebate) or periods of grid congestion (incentive programs). Since the main objective of DR programs is to influence electricity consumption patterns, the main performance indicator that can be used is the change in energy consumption during event periods. A more powerful objective is the correlation between energy production costs and demand. Depending on the generation mix and the matching procedures in the energy market, energy costs vary in time. If DR can help to reshape the demand curve to the inverse of the energy costs curve, it can become a useful tool to decrease overall energy costs.

A typical metric to assess how a DR event's performance leads to reshaping of the demand profile is the RMS (Root Mean Square) of the difference between the real demand and the reference curve within the event control period. This value is to be compared with the RMS associated to the baseline, so as to assess the performance of the DR event. But in order to compare RMS values, both curves should imply the same amount of energy, so that their mean value is the same. This fact represents DR actuations where the overall energy demand is not decreased. However, this situation is nearly impossible, because the average demand of the baseline and the real demand do not have a direct relation. The usual effect is that the average demand is lower than the average baseline, so the reference curve of the baseline should be normalized with the ratio of average demand and average baseline.

This leads us to the introduction of another metric, the mentioned ratio of average demand and average baseline during the control event. This indicator is also a figure to monitor the performance of a DR event, but to be accurate it must also include a certain amount of time following the

[1] From grid operation point of view the ideal situation is a flat demand profile. This objective is of course utopic because of the limitations in the operation of generation and mainly those associated to demand management.

end of the control event. The reason is the possible rebound associated with operation of appliances with high thermal inertia such as HVAC systems. In the case of DR programs where the control events are consecutive, for instance in Time of use (TOU) programs with dynamic prices, the rebound effect should be included within each control period. If events are considered in a daily basis this condition is usually taken care of, because the time period susceptible for rebound effect is usually the afternoon (at least for cooling systems, in which the peak period is close to noon).

Additionally, when customers are recruited for DR programs by LSEs or third party energy aggregators, a number of KPIs related to demand dispatch such as uncertainty, variation, and delay of demand shed become relevant as these factors determine the quality of response from the customers. Such metrics are useful in DR design and evaluation to determine the type of incentives to provide and the type of customers to recruit in order to achieve a certain amount of demand reduction. These include:

- *Uncertainty of demand shed*: Let $X_1, \dots, X_n$ denote the time series that records the demand varied by a customer, i.e., the difference between the predicted demand (baseline) and the actual demand, during the past n DR events. One can measure the uncertainty of demand reduction using the following methods:

  i. *Variance*: Var(X). Higher the variance, higher the uncertainty.

  ii. *Entropy*: Let X be discretized into k bins with thresholds $b_0, b_1, \dots, b_k$ and let $p_k = \Pr(X_i \in [b_{k-1}, b_k])$. Then $H(X) = - \sum_{i=1}^{k} p_k \log (p_k)$ is an estimate of the uncertainty in the reduction. Higher entropy implies higher uncertainty.

  iii. *Risk*: Let $r_k = \Pr(X \geq b_k)$, i.e., the probability that the customer's reduction is at least b_k. r_k is an indicator of the uncertainty or risk associated with the demand reductions of a customer. Lower probability implies higher risk. Other parameters could also be taken into account to measure the risk associated with a demand reduction.
  One can also consider negative demand sheds in the time series (i.e. rebounds) if the customer increases his demand during a DR event.

- *Delay responsiveness of demand shed*: Let $\{D_1, \dots, D_n\}$ denote the time series that records the time required by a customer to shed his demand during the past n DR events. One can measure E[D], Var(D), and $\Pr(D \geq d)$ and treat them as indicators of delay responsiveness of a customer, that is how quickly the demand reduction can be dispatched. This helps classify customers/loads into different categories and helps utilities dispatch demands at optimal times during peak periods.

D. Economic-related KPIs

To quantify the DR effects on user demand and user comfort, as well as the relevant benefits in aggregate, the key indicators to be used relate to price elasticity, rate of participation, and discomfort caused to users.

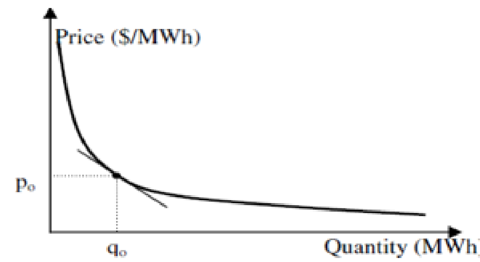

Figure 1. Price elasticity.

Price elasticity can be decomposed into three types:

1. *Self-elasticity*: measures the demand reduction in a certain time interval due to the price of that interval. It is always negative; usage goes down as price goes up. For example, if a customer's price elasticity is 0.15, then a doubling (100% change) of price results in a 15% reduction in electricity usage or other things equal. Higher elasticity values translate into increased price response by customers.
2. *Cross-elasticity*: measures the effect of time in a certain interval on the electricity consumption during another interval. Namely, it measures the consequences of reduced electricity usage on other goods. If a customer buys less electricity, then he has more money for spending them on other goods and services.
3. *The elasticity of substitution*: measures the rate at which the customer substitutes off-peak consumption for peak usage in response to a change to the ratio of peak to off-peak prices. It can have a positive value (or zero) and is commonly used in analyzing price response among large industrial and commercial customers.

Several studies conducted so far have used average price elasticity to observe the behavioral changes of residential, industrial and commercial customers. Figure 2 summarizes the results of studies that estimated the price response exhibited by customers that participated in voluntary programs that involved time-varying prices.

Various factors may influence customers' price elasticity, including the nominal level of prices. For example, some customers may be relatively irresponsible when prices are low but find it worthwhile to reduce load at very high prices. This characteristic of price elasticity has important implications for the design and evaluation of time-varying pricing and DR programs.

Studies of large customers' response to time-varying price changes find that there are large differences in price elasticity across business categories and various market segments. As for the residential customer response to time-varying prices studies often report that price elasticity is driven in part by the number of electricity devices present in the home. Climate plays also a significant role, as well as the residents' characteristics and the events that they affect, when they are at home and likely to shut off devices or reduce usage.

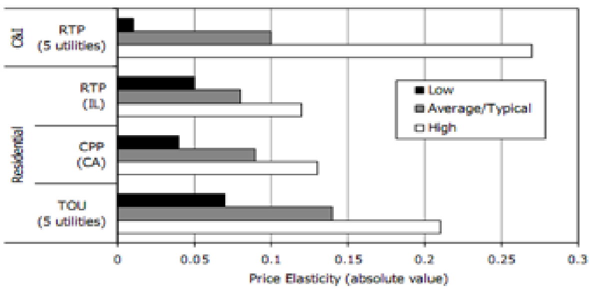

Figure 2. Customer Response to Time-Varying Prices: Price elasticity Estimates.

Finally, great role in the evaluation of DR programs have the customer acceptance and their rate of participation in dynamic pricing and DR programs. Important factors in the consumer's decision to participate and enrol include the level and type of incentives offered, program requirement and conditions (e.g. duration and frequency of curtailments), assessment of risks and value (e.g. financial consequences for failures), effectiveness of program design and implementation (e.g. marketing, technical assistant) [1]. Besides in some DR programs (e.g. where customers do not directly respond to prices) their response is typically measured by the amount of load reduced.

The following provides a synopsis of some metrics that are candidates for measuring DR economics as a whole.

- *Demand Price Elasticity (Self Elasticity)*: it measures the sensitivity of customer's demand to price changes. It can be calculated using the following formula:

$$Self\ Elasticity = \frac{\delta e}{\delta p}$$

where δe is change in energy demand and δp is change in price.

- *Customer Responsiveness*: It is an indicator that measures how many customers have responded to a DR program following a DR signal sent to them, like a change in price. It can be measured as the total number of signals sent back by the customers as an absolute number or a percentage. The term "signal" as a feedback is defined in each case based on the particular context of the DR program (for example the GUI utilised). Furthermore, at the time of the reaction of the user, the related context-specific aspects can be observed and necessary metadata stored for further use (e.g. the customer responsiveness on weekends).
- *Absolute or Relative Load Impact*: Further to the above metric this one is used in order to specify the intensity of customer's response and can be measured as the number of kW of load curtailed or the percentage (%) of customer's total load that is curtailed.
- *Absolute Discomfort Impact*: It indicates how much the customer's comfort has changed. It is a simple but important indicator and there are various options of measuring it, depending also on the definition of "customer's comfort" in the specific case. An example of discomfort is the temporary change of temperature to save energy (momentarily stopping air conditioning in summer or heating in winter) as perceived by the consumer [3] .
- *Discomfort level against total energy reduction constraint*: considering the case that there is an objective set by the environmental/energy manager or administrator of achieving a specified % reduction of total consumption (expressed as a "hard" constraint) in a specific household or office premises, etc., this metric measures the level of discomfort caused by the specified reduction. This may vary based on the different reduction strategies utilised in order to achieve the target reduction. i.e., Discomfort level *X* achieved for reduction

(*NewTotalConsumption* - *Orig.TotalConsumption*) [%]

- *Total energy reduction against discomfort level constraint*: Compared to the aforementioned metric this one investigates the inverse case i.e., the challenge here for the energy administrator of the building is to achieve the maximum reduction of energy consumption without exceeding a specific discomfort level/threshold. i.e. Given a Discomfort level *X*,

(*NewTotalConsumption* - *Orig.TotalConsumption*) [%] is achieved.

E. KPIs for quantifying the net economic benefits

Furthermore, in order to quantify the potential net economic benefits of DR programs the following KPI can be utilised:

- Net Economic benefit for a player; that is, the difference of profits after and before the DR program.

Thus, special methods should be developed to measure the economic and financial benefits of the various energy value chain players arising from the load/demand reduction due to the adoption and application of different DR programs under different types of markets. There are many factors and externalities among players to consider when determining the economic benefits, which can be short-term (e.g. peak reduction leading to less frequent usage of costly backup generators) and long-term (e.g. stability of the distribution and transmission networks resulting in lower maintenance costs and better network planning). However, at the end of the day, it all comes down to estimating and comparing the total associated costs for each value chain player and the resulting difference in revenues before and after the deployment of the DR systems and programs in order to obtain the net Economic benefit for this player.

Finally, an overall measure of the economic impact of a DR program on the entire society (including both players and users), can be given by:

- Difference in Social Welfare. This is the sum of the net economic benefits of all players in the value chain, plus the sum of net benefits of all users. It can be easily seen, however, that this amounts to the different of the sum of user utilities after and before DR plus the difference of the total cost for energy production and distributions.

IV. Conclusions and Further Work

DR programs are an important part of future smart grids and given their vast potential in helping reduce peak load and matching demand with supply, it is important to understand the effectiveness of different DR programs and define metrics which could be used to evaluate the performance of proposed programs.

Key performance indices are part of a set of tools that can reveal the efficiency of a DR system: efficiency of communication between the stakeholders, efficiency of flattening the peak load, efficiency towards sustainability of the DR program, efficiency of influencing the customer, and so on. In this work, we presented different types of KPIs including KPIs for Demand Variation Analysis and Demand Reshaping, Economic KPIs, and some of the KPIs for quantifying the net economic benefits.

Further to the KPIs discussed, in the course of the WATTALYST project we are planning to investigate a number of related KPIs related to different aspects of performance of DR systems such as those related to storage and retrieval of consumption data. These include (i) KPIs related to data management infrastructure: Examples include the Average Response Time [s], Average Data throughput [MB/s]. Both metrics must be evaluated in two directions with respect to the data flow: i.e. Data Upload and Data Retrieval.(ii) Mining and Knowledge Extraction. The Benchmarks will include traditional relational database systems (RDBS) and systems tailored for storing mass data like the Hierarchical Database Format (HDF5).

In order to evaluate particular DR programs and assess the appropriateness of the KPIs presented (and refine them as necessary), four trial sites have being established as part of the WATTALYST project. The data collected by sensors and smart meters along with feedback messages to and from the end consumers from these sites will be used to evaluate various parameters that influence customer participation as well as the impact of the key performance indices in order to establish their support in the design and maintenance of DR systems.

Acknowledgment

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 288322.27 (WATTALYST).

Finally, we would like to thank all the partners of WATTALYST (Modelling and Analysing Demand Response Systems) project [1] who have contributed with their analysis and discussions in the production of these results.

References

[1] WATTALYST project http://www.wattalyst.org/

[2] Benefits of demand response in electricity markets and recommendations for achieving them. Technical report, U.S. Dept. of Energy, 2006

[3] Van Deventer, J., Gustafsson, J., and Delsing, J., Controlling district heating load through prices, Systems Conference (SysCon), 2011 IEEE International, 461-465.

[4] NETL 2011. Demand Dispatch—Intelligent Demand for a More Efficient Grid. Prepared by National Energy Technology Laboratory for US Dept of Energy.

[5] Brooks, A.; Lu, E.; Reicher, D.; Spirakis, C.; Weihl, B., Demand Dispatch, IEEE Power and Energy Magazine, vol.8, no.3, pp.20-29, May-June 2010.

[6] Y. V. Makarov, C. Loutan, J. Ma, and P. de Mello. Operational impacts of wind generation on California power systems. IEEE Transactions on Power Systems, 24(2):1039–1050, May 2009.

[7] D. Kirschen. Demand-side view of electricity market. IEEE Transactions on Power Systems, 18(2):520–527, May 2003.

[8] J. C. Smith, M. R. Milligan, E. A. DeMeo, and B. Parsons. Utility wind integration and operating impact: State of the art. IEEE Transactions on Power Systems, 22(3):900–908, August 2007.

[9] N. Ruiz, I. Cobelo, and J. Oyarzabal. A direct load control model for virtual power plant management. IEEE Transactions on Power Systems, 24(2):959–966, May 2009.