On Load Elasticity

S. Keshav and C. Rosenberg
University of Waterloo, Canada
{keshav, cath}@uwaterloo.ca

1. Introduction

Grid operators today meet the twin goals of system reliability and revenue maximization by provisioning the system conservatively at close to the peak load and charging customers for the resulting high costs of the necessary infrastructure. This maximizes revenues while maintaining reliability. Most grids are regulated monopolies or have nearly monopolistic market share in their home market, so their operators have had no incentive to reduce capital spending that can be recouped from a large and captive customer base.

Unfortunately, this business-as-usual approach does not take some externalities into account. For instance, coal, which was used to generate two-thirds of all electricity in the US in 2008, causes air pollution, radioactive emissions, and an enormous carbon footprint. With the looming danger from anthropogenic climate change, there is tremendous political pressure to decommission coal plants, reducing baseload generation capacity. Similarly, widespread fears of nuclear proliferation are causing some countries to hold back on the deployment of additional nuclear power plants (as in the United States) or even decommission existing plants (as in Germany). In view of these externalities, the inherent inefficiency of uncontrolled peak loads, and past experience in demand management, *demand response* has emerged as a critical feature of the smart grid [1].

Demand response (DR) refers to the use of pricing to cause electricity consumers to intentionally modify the time at which they consume electricity, their peak demand level, or their total electricity consumption [2]. It can be achieved either through incentive-based programs, where consumers are paid to participate in a program where utilities directly control their load, or market-based programs, where customers respond to price signals that reflect overall system load. In either case, DR reduces the peak-to-average demand ratio, increases the power factor of generators, and allows generators to defer capacity increases. Moreover, it decreases generation costs by reducing the size of spinning reserves and, in some cases, the use of expensive energy sources [2]. Finally, it can increase system reliability–averting load-shedding or blackouts–by reducing load when system stability is in jeopardy.

The underlying assumption made by any DR scheme is that load is *elastic*, that is, some loads can be time-shifted with no loss of utility. Although this is a critical assumption, we are not aware of prior work that analytically examines load elasticity; hence this communication.

2. The concept of load elasticity

It is illustrative to first study the concept of elasticity both in economics and in the context of the Internet.

Economic concept of elasticity

The economic definition of demand elasticity measures the degree to which the total revenue (the product of price and demand) is affected by a change in price [3]. Demand is said to be elastic at a particular price point when a decrease in price increases the total revenue. For example, with elastic demand, a 1% decrease in price at a particular price point would increase demand by more than 1% so that the total revenue increases. Otherwise, it is said to be inelastic (see Figure 1). A customer with elastic demand at a particular price point can be thought to be responsive to a price signal, in contrast to a customer with inelastic demand, who does not reduce demand proportional to the percentage increase in price.

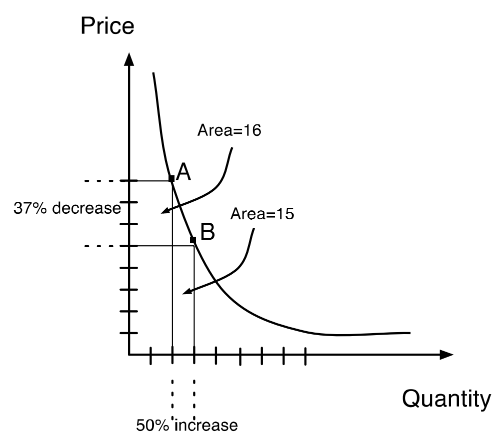

Figure 1: Demand Elasticity. A 37% decrease in price increases demand by 50%, but the total revenue at point B is lower than at point A, so demand is said to be *inelastic* at point A.

Network formulation

The widely-accepted definition of elasticity in a communication network, due to Shenker [4], is in terms of the utility to a distributed application (i.e., an application that has been implemented as two or more parts with a network connection between them) as a function of the connection's *bandwidth* (the network equivalent of power that is measured in bits per second rather than watts) available to it. If this utility exhibits a diminishing marginal increase as a function of bandwidth, then the application is said to be elastic (see Figure 2).

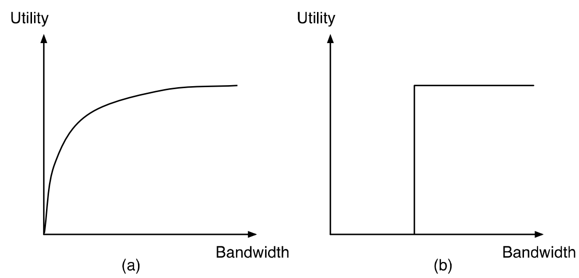

Figure 2: Elastic (a) and inelastic (b) applications

It is instructive to contrast an elastic application with an inelastic one. An inelastic application gets no utility at all when its connection's bandwidth decreases below some threshold and no additional increase in utility when this bandwidth exceeds this threshold. Thus, for such an application, the marginal increase in utility as a function of bandwidth is zero everywhere, except at the threshold value itself.

Note that the Internet definition of elasticity does not consider the duration of the connection. This obscures the fact that the connection must eventually transfer a data item from source to destination. In other words, although the bandwidth associated with a connection may vary over time, a certain number of bits must eventually be transferred over the connection. In the electrical grid, this corresponds to a load that must receive a certain amount of energy, though the power may vary over time. The Internet definition of elasticity must therefore be augmented with the constraint that for an elastic application the area under the bandwidth vs. time curve, which is the size of the data item, is conserved.

Note also that the Internet formulation of elasticity, unlike the economic formulation, does not take pricing into account.

Elastic electrical loads

We now consider how to apply these concepts of elasticity to electrical loads. Intuitively, we think of an electrical load as being elastic if it can be modified in some way, such as, for example, in response to a price or congestion signal, without overly reducing the comfort of the consumer [5]. This allows the system operator to manipulate demand to achieve some system objective such as reducing peak demand or transmission line congestion, without sacrificing customer comfort, or perhaps compensating for the reduction in comfort with a payment. Note the inherent three-way trade-off between user comfort, payments to the user, and the system objective (in some cases utilities may mandate load reduction without payments). For example, an electric vehicle (EV) owner may be insensitive to its charging rate as long as the EV is charged before some loose deadline. Therefore, reducing its charging rate can help to achieve the system objective of reducing peak load, which makes the EV's load elastic. In contrast, the power given to a refrigerator today can be neither diminished nor time shifted. We draw upon this intuition to develop a definition of elasticity next.

3. Quantifying elasticity

We define a *load profile* π to be a continuous function of time that represents an appliance's load as a function of time. We define the utility to a customer of a load profile $U_c(\pi)$ to be the benefit to the customer from a particular load profile. This generalizes the Internet concept of the utility as a function of the bandwidth to the utility as a function of the bandwidth over a time period.

A nominal load profile π^* is *inelastic* if $U_c(\pi)= 0$ for all $\pi^* \neq \pi$. In other words, the slightest change in the profile causes the customer's utility to drop to zero – for example, a load due to a television that would be damaged if there a drop in either voltage or drawn current. In practice, we expect no load to be completely inelastic, because all devices have some built-in margins to deal with short-term fluctuations.

In contrast, a nominal load profile π^* is said to be *purely elastic* when there exists a set Π of other profiles such that the customer's utility from all members of Π is equal to $U_c(\pi^*)$. That is, the customer is indifferent to all profiles in Π.

We can generalize this further. Assume that a nominal load profile π^* with utility $U_c(\pi^*) = a$ is such that there exists at least one other load profile π' such that $U_c(\pi')$ = a - ε. Then, we call the load profile ε-*elastic*.

4. Load types

We now classify load types by their degree of elasticity, drawing on a classification of connection types in data networks such as the Internet.

In the Internet, packets are sent on a connection from the source to a destination. Generally speaking, the quality of a connection diminishes and the utility of the connection to the application user is reduced as the number of bits per second that the network provisions for that connection (also called its bit rate) decreases. Three types of dependency of utility on the bitrate are well known [6]:

- Constant Bit Rate (CBR): Such applications are inelastic with respect to the bit rate. Their marginal increase in utility with respect to the bit rate is zero everywhere except at a particular threshold value. In other words, the source receives zero utility unless the network allocates enough resources to carry at least the threshold bitrate.
- Variable Bit Rate (VBR): Such applications generate traffic in bursts, rather than in a smooth stream. A VBR source is typically modeled as having an intrinsic long-term bit rate (called its average rate) with occasional bursts of limited duration at a rate as high as a specified peak rate. An application receives zero utility if the network allocates resources less than the average rate to the connection, and its utility increases and then saturates as the network increases its allocated resources to the peak rate and then beyond.
- Available Bit Rate (ABR): Such applications have a non-zero marginal gain in utility everywhere. Recall that there is an implicit assumption that the area under the bandwidth profile is conserved. A typical example is a file transfer where a source wants to send some number of bits to a recipient, the sooner the better.

This inspires us to classify loads as Fixed Power, Variable Power, and Available Power loads.

- Fixed Power (FP): Such loads are inelastic with respect to their load profile. Their utility is zero everywhere except when served using their nominal load profile.
- Variable Power (VP): Such loads are ε-*elastic*. That it, they have a preferred profile, but their utility does not change much for profiles 'close to' that profile.
- Available Power (AP): This refers to purely elastic loads whose utility does not change despite certain changes in the load profile. An example of such a change could be (a) the area under the profile curve is conserved and (b) the demand is satisfied before a given deadline (these implicitly define Π).

Traces show that a consumer's aggregate electrical load can be roughly partitioned into two portions. The base load is the load from always-on devices such as set-top boxes, safety lighting etc. This is typically fairly low and can be modeled as Fixed Power loads. Demand sharply increases when heavy-load devices such as air-conditioners, refrigerators, electric ovens, and baseboard heaters are turned on. These loads can be modeled as AP (or as VP under certain conditions). Finally, new loads, such as EV, could be modeled as VP or AP.

4. Conclusion

This paper discusses and formalizes the notion of load elasticity in the electrical grid, and compares it with the notion of elasticity used in economics and in the Internet. We use these concepts to describe electrical loads as falling into three natural categories: Fixed Power, Variable Power, and Available Power, that mimic the well-known categories of CBR, VBR, and ABR, respectively, in the field of computer networking.

References

[1] Ipakchi, Ali, and Farrokh Albuyeh. "Grid of the future." *Power and Energy Magazine, IEEE* 7, no. 2,pp. 52-62, 2009.

[2] US Department of Energy, "Benefits of demand response in electricity markets and recommendations for achieving them," Report to Congress, 2006.

[3] Samuelson, P.A., "Economics," *McGraw Hill*, 11th Ed. 1980.

[4] Shenker, S., "Fundamental design issues for the future Internet." *IEEE J. Selected Areas in Communications*, v.13, no. 7, pp. 1176-1188, 1995.

[5] Srikantha, P., C. Rosenberg, and S. Keshav, "An analysis of peak demand reductions due to elasticity of domestic appliances." In *Proc. of the 3rd International Conference on Future Energy Systems*, p. 28, 2012.

[6] Keshav, S., "An Engineering Approach to Computer Networking," *Addison-Wesley*, 1997.

Srinivasan Keshav is a Professor, Canada Research Chair, and Cisco Chair in Smart Grid in the Cheriton School of Computer Science at the University of Waterloo. He received a PhD in Computer Science from UC Berkeley in 1991.

Catherine Rosenberg is a Fellow of the IEEE, Professor, and Canada Research Chair in the Department of Electrical and Computer Engineering at the University of Waterloo. She received a Doctorat en Sciences from Université de Paris XI Orsay, France.