Markovian Models for Home Electricity Consumption

Omid Ardakanian, S. Keshav
David R. Cheriton School of Computer Science
University of Waterloo
Waterloo, ON, Canada
{oardakan,keshav}@uwaterloo.ca

Catherine Rosenberg
Dept. of Electrical and Computer Engineering
University of Waterloo
Waterloo, ON, Canada
cath@uwaterloo.ca

ABSTRACT

Modelling home energy consumption is necessary for studying demand-response, transformer sizing, and distribution network simulation. Using an existing classification, we propose parsimonious Markovian reference models of home load for each class. We derive models for on-peak periods, off-peak periods, and mid-peak periods. These models are derived using traces based on fine-grained measurements of electricity consumption in 20 homes over four months. We validate the representativeness of our models in a specific application.

Categories and Subject Descriptors

I.6 [**Simulation and Modelling**]: Model Validation and Analysis; G.3 [**Probability and Statistics**]: Markov Processes

General Terms

Design, Measurement, Verification

Keywords

Electrical grid, Markov chain, load modelling

1. INTRODUCTION

With the rapid ongoing deployment of the smart grid, there is an increasing need for accurate and parsimonious models for home electrical loads. Accurate home load models are the basis for research in smart appliance design, distribution network sizing, modelling of demand response, smart charging of electric vehicles, and home energy management. Most attempts at home load models in the past have been constrained by the limited measurement infrastructure. Smart meters measure home load every five minutes, at best, which is too slow to capture the on-off behaviour of appliances. In path-breaking work, Richardson et al at the University of Loughborough measured 22 homes

for a year at a one-minute granularity to construct a home-load generator [11]. However, their generator is not parsimonious, in that its input is the set of appliances in each home, the appliance load model, the number of occupants, and the occupancy behaviour. It is difficult to obtain these inputs in any realistic situation. In this paper, we are motivated by the need to develop models for the sizing of transformers and storage in distribution networks using queueing theory [2], and hence we focus on Markovian models.

Over the years, electrical utilities have built classifications of home loads into a small number of representative classes. We use one of these classifications to try to derive a reference load model per class for different periods of the day. We propose an analytical model based on a k-state Markovian model. This kind of model will be very suitable for use in mathematical analysis. We use even finer-grained measurements than Richardson *et al* (every six seconds) of 20 homes over four months to develop these reference models. We show that each class in each period can be modelled accurately by a Markovian model with no more than 6 states. We provide the transition rate matrix, Q, and the power consumption matrix, R, for each of these reference models. We believe that these models could be very useful to the research community interested in studying demand-response, transformer and storage sizing, and distribution network simulation. We validate our modelling approach for the specific application of transformer sizing by comparing the loss period (i.e., the period that the aggregate load exceeds the transformer size computed by the teletraffic theory) obtained from the numerical simulation with the industry standard loss duration.

We make three contributions:

- We collect fine-grain load measurements from real usage using inexpensive off-the-shelf components in 20 homes over four months, mid-January to mid-May. These 20 homes can be partitioned into 4 classes using an existing classification.

- We create parsimonious Markovian reference models for these 4 classes for different periods of the day.

- We validate that the accuracy of our models is sufficient for our specific application.

The rest of the paper is organized as follows. We describe the need for parsimonious models and for Markovian models, as well as related work in Section 2. We present our measurement testbed in Section 3. We present an overview of the data in Section 4. Our model is presented in Section 5 and is validated in Section 6. We conclude in Section 7.

2. BACKGROUND AND RELATED WORK

It is important for a home electricity consumption model to be parsimonious because a plethora of parameters makes it difficult to cover the model space when doing analysis or simulations. However, parsimony comes at a cost. Many factors determine home load such as the time of day, environmental and geographical factors, the type of appliances in the home, the usage pattern of these appliances, the number of occupants of the house, the occupancy pattern, and the size and perhaps even the floor plan of the house. A parsimonious model must necessarily ignore one or more of these factors. To make the modelling problem tractable, we focus on parsimonious models that are suitable for a specific purpose, that is, the sizing of transformers in the distribution system [2]. This reduces the problem to constructing a representative set of continuous-time Markov models, as discussed in Section 5.

A critical modelling decision is to use continuous-time Markov processes. We are motivated by three reasons. First, Hidden Markov Models and Markov chains have been extensively used in the past to model sequential events. For example, they have been used to model natural languages, human speech, molecular evolution, and animal behaviour [10, 8, 6] In all these cases, it has been shown that Markov models combine both parsimony with descriptive power. Second, Markov models are the foundation for many types of mathematical analysis, especially queueing theory and stochastic optimization. This makes them particularly attractive for our application. Finally, home electricity usage arises from the superposition of a finite set of on-off loads from individual appliances. Such superpositions have been shown in the past to be well-modelled using Markovian models [12].

In the past, detailed models for residential loads have been presented in the power engineering, environmental studies, and civil engineering literature [11, 3, 4, 9]. However, these models suffer from three problems. First, these models tend to be highly parametrized, rather than parsimonious. Second, the data sets on which these models are based are not publicly available. Therefore, we cannot use them to create Markovian models. Finally, to the best of our knowledge, existing models group all homes into a single class. Our measurements show significant differences in demand behaviour at different homes.

Over the years, electrical utilities have built classifications of home loads into a small number of representative classes. In particular, they use this classification to size their transformers. We use one of these classifications to partition our measured home loads into 4 classes and derive per class Markovian reference models for different periods of the day.

3. TESTBED

Our first step is to obtain real measurements of the electrical load. To obtain our own load data set, we built a testbed to measure aggregate loads at 20 homes. We deployed measurement nodes (Figure 1) at 19 houses and one home-based small business covering a range of living area sizes, occupants, appliances, and energy consumption patterns. For the purpose of our small pilot study, we used a convenience sample rather than a stratified random sample. Our methodology generalizes to samples chosen using standard population sampling techniques [5].

Each measurement node consists of a Current Cost Envi

Figure 1: A measurement node

Type of Heating	House Size			
	$100m^2$	$200m^2$	$300m^2$	$400m^2$
Baseboard electric heat	3.0	4.0	5.0	6.0
Central electric heat	4.0	5.0	6.0	7.0
Gas/oil heat, no central A/C	1.0	1.5	2.0	2.5
Gas/oil heat, central A/C	1.5	2.5	3.5	4.5
For town or row houses, multiply the unit value by 0.8.				

Table 1: 'Unit values' assigned to customer homes by a major utility.

device [1] and a netbook. The Envi device measures the active power consumption (in Watts) of a house every six seconds and stores it locally in flash memory[1]. A script on the netbook queries the device every six seconds to obtain an XML file that it stores on disk. This is uploaded using a secure SSL connection to a server in our laboratory once a day. To preserve privacy of the participants in our study, logs files are anonymized before being stored in a secure directory on the file server.

4. DATASET

Given the high variability of home loads, choosing a classification for them is a challenging task. Fortunately, standard rules based on decades of field experience allow an electric utility to both predict and classify a home load based on a few simple parameters. We obtained such a parametrization, specifically used for transformer sizing, from a major utility in our area (Table 1). The key sizing parameters are the house size and the nature of the heating and cooling systems, which constitute the major loads in our geographical area. These are used to compute a 'unit value' that represents the load expected from that home.

To minimally impact participant privacy, we asked each participant to tell us their home's unit value computed using this table. We then placed homes with the same unit value in the same class. Table 2 shows the four classes so obtained

---

[1]Consequently, the device does not capture load transients that last shorter than this time.
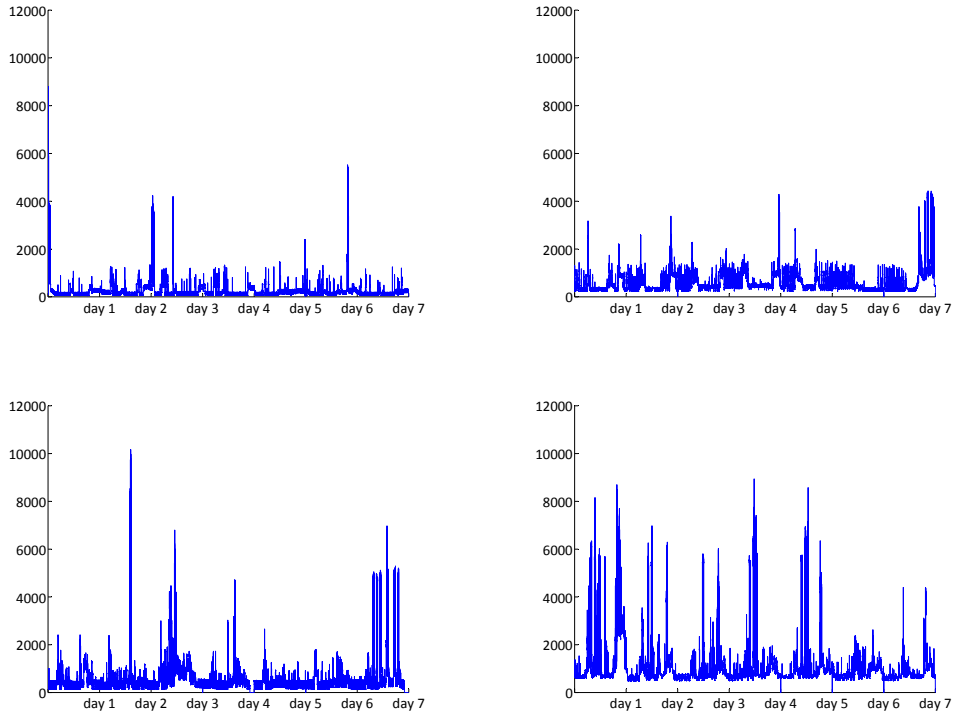
Figure 2: Load measurements from typical homes in 4 classes for one week.

Class	Unit value	Number of houses
1	1.2	8
2	2.5	7
3	3.5	3
4	4.5	2

Table 2: Number of homes in our experiment within each class.

(note that in our area homes are in general equipped with gas/oil heater).

Typical loads from one home in each class for one week (in winter) are shown in Figure 2. The main differences between classes are the amount of peak and base loads, the width of the peak period, and variability of the load (e.g., the average number and the average height of spikes in a given period). Table 3 represents the peak and base loads of each class. The base load of each class is the smallest element of the R matrix, described in Section 5. Similarly, the peak load is the largest element of the R matrix.

	Class 1	Class 2	Class 3	Class 4
Peak load (W)	3902	4580	6890	7132
Base load (W)	142	234	410	769

Table 3: Peak and base loads of each class.

5. MARKOV MODEL

This section describes our approach to building a reference model for each class of homes introduced in Section 4. We answer the following questions:

- How many reference models do we need?

- How can we choose the Markov states?

- What metric determines the goodness of a model?

- How many Markov states are needed in each model?

In this work, we assume that the homes selected for measurement in our study are a representative random sample of their assigned class. This assumption is admittedly strong, but can be removed if homes chosen for measurement were chosen from a stratified random sample, which we defer to future work.

5.1 Definition

A continuous-time Markov process is the continuous version of a Markov chain. In a Markov chain transitions happen at specific time steps; however, in a continuous-time Markov process the system remains in the previous state for a period of time before it transitions to a new state; these time periods are exponentially distributed. Dynamics of a k-states Markov process is represented by a $k \times k$ transition rate matrix (also known as the intensity matrix), Q, where q_{ij} is the rate of departing from state i to arrive at state j. Since the transition rates of each state should sum to zero,

q_{ii} is defined as[2]:

$$q_{ii} = -\sum_{j \neq i} q_{ij}$$

Similar to the approach used in [13] for modelling VBR video sources, we assign a value, R_i, to state i of a Markov process; this value represents the amount of power consumed in this state. We model the electric load of a home using a k-state continuous-time Markov process defined by the $<Q, R>$ tuple.

5.2 How many models do we need?

In Section 4, we classify the homes being measured into four classes based on the 'unit size' assigned to them. This motivates the need for at least four models. Furthermore, home load is highly sensitive to the time of day. More specifically, it is not stationary over the period of a day, and the probability that the peak power will be consumed at 6am is far lower than it is at 6pm. To deal with this issue, taking a cue from the time periods specified by the electric utility in our region, we divide a day into three periods, namely *on-peak* (7am-11am and 5pm-9pm), *mid-peak* (11am-5pm), and *off-peak* periods (12am-7am and 9pm-12am). Visual observation verifies that the home load is almost stationary in each period. We caution that the definition of these periods might be different in other geographic regions or seasons. Nevertheless, we believe that similar three periods can be identified in every region and season. Therefore, we construct three reference models (one for each period) for each class, for a total of 12 reference models.

5.3 Choosing Markov states

To find a k-state Markov model for each class in each period, we first need to choose a representative load from our measurements. This representative load must include electric loads of all homes within the corresponding class in that period; therefore, we construct it by concatenating these electric loads. For example, the representative on-peak load of a class is the concatenation of on-peak periods of all homes selected for measurements within that class for the entire measurement period. We then use the k-means clustering algorithm to find k centroids of the representative load. Values of these centroids constitute the R matrix. To increase the chance of finding the global optimum in the k-means clustering, we run 500 replicates of the k-means algorithm with random start points. Substituting values of the points in each cluster with the value of its centroid, we obtain a clustered home load. Then we use the clustered home load to compute the Q matrix from the following expressions

$$q_{ij} = \frac{\text{no. of transitions from } R(i) \text{ to } R(j) \text{ in clustered load}}{\text{total time spent in state } i \text{ before a transition to state } j}$$

5.4 How many Markov states should a model have?

An obvious question is that how many states are needed in each model. To answer this question we need to define a goodness-of-fit metric. We use the area between the cumulative distribution function (CDF) of the measured home load

	On-peak	Mid-peak	Off-peak
Class 1	6	6	5
Class 2	5	3	5
Class 3	4	5	2
Class 4	3	3	3

Table 4: Minimum number of states necessary for representing the home load of a class in a period.

and the CDF obtained from modelling the load of the class it belongs to as this metric. This represents how far two probability distributions are from each other: the greater the metric the higher the modelling error. Therefore, to find a sufficient number of states to model a class, we sum the areas for all homes within that class and study this as a function of the number of states in the model (Figures 3, 4, and 5). We define the knee or corner point of this curve as the minimal number of states of the corresponding Markov model. Table 4 summarizes how many states should be used for each class and for each period.

Thus, for example, a 5-state Markov model is a good representative of the on-peak load of the second class. It is defined by the following Q and R matrices[3]:

$$Q = \begin{bmatrix} -0.2704 & 0.0141 & 0.0684 & 0.0763 & 0.1116 \\ 0.0114 & -0.1336 & 0.0436 & 0.0513 & 0.0273 \\ 0.0146 & 0.0103 & -0.0626 & 0.0097 & 0.0279 \\ 0.1186 & 0.0711 & 0.0833 & -0.3029 & 0.0298 \\ 0.0239 & 0.0493 & 0.0650 & 0.0183 & -0.1565 \end{bmatrix}$$

$$R = \begin{bmatrix} 2720 & 878 & 325 & 4649 & 1503 \end{bmatrix}$$

To sum up, we have shown how to derive a Markovian reference model per class and per time period. We have also provided the Q and R matrices for each reference model. We now validate the representativeness of our reference models.

6. VALIDATION

Validating the representativeness of our reference models raises some subtle issues. Every model intentionally ignores certain aspects of reality. In this sense, no model is truly representative of reality. To avoid this inherent limitation of modelling, we remind the reader of the purpose of our models, which is primarily for sizing transformers in the distribution grid. We therefore claim that a model is representative if the transformer size computed using the model matches that computed using the raw measurements of home loads.

To this end, we start with home loads measured from 20 houses over a week. We divide the measured home loads into two disjoint data sets, training and test. The training data set is used to construct the reference models, whereas the test data set contains the raw home load measurements that are used in the validation process. We use our reference models to compute the transformer size required to achieve a loss of load probability of 2.74×10^{-4}, which is the standard used in industry, using teletraffic theory [7], as described in [2]. Then, using this same transformer size, we run a numerical simulation using the raw measurements as the input and compute the total duration of overload, which is a numerical estimate of the loss of load probability.

---

[2]An ergodic continuous-time Markov process has a stationary probability distribution, π, that can be easily computed from its transition rate matrix.

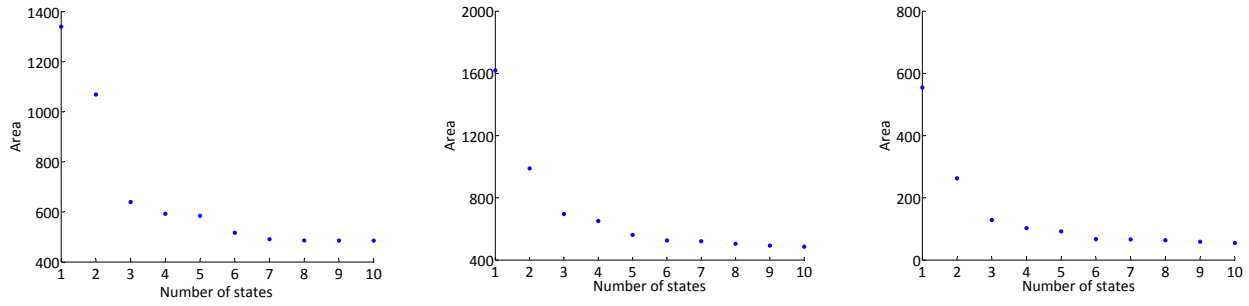[3]The matrices for the other 11 models are available on `http://blizzard.cs.uwaterloo.ca/~oardakan/models`.

Figure 3: The goodness of fit metric versus the number of states for the on-peak period of classes 1, 2, and 4. Due to lack of space, the curve for class 3 is omitted. Note that the Y axis is exaggerated for emphasis.
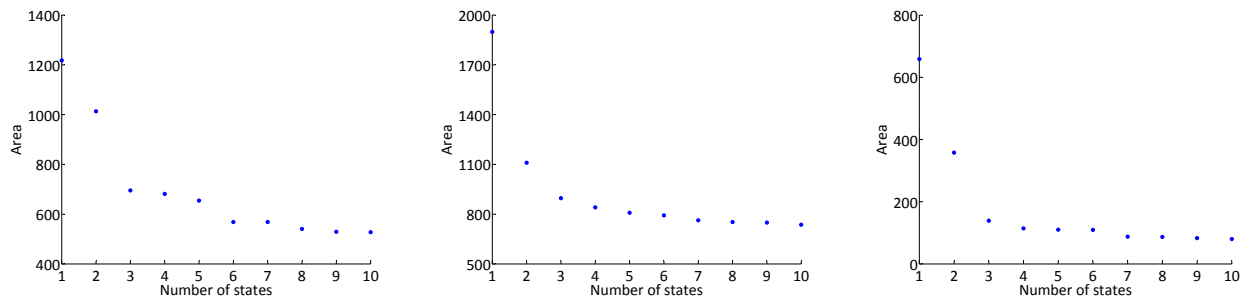

Figure 4: The goodness of fit metric versus the number of states for the mid-peak period of classes 1, 2, and 4. Due to lack of space, the curve for class 3 is omitted. Note that the Y axis is exaggerated for emphasis.
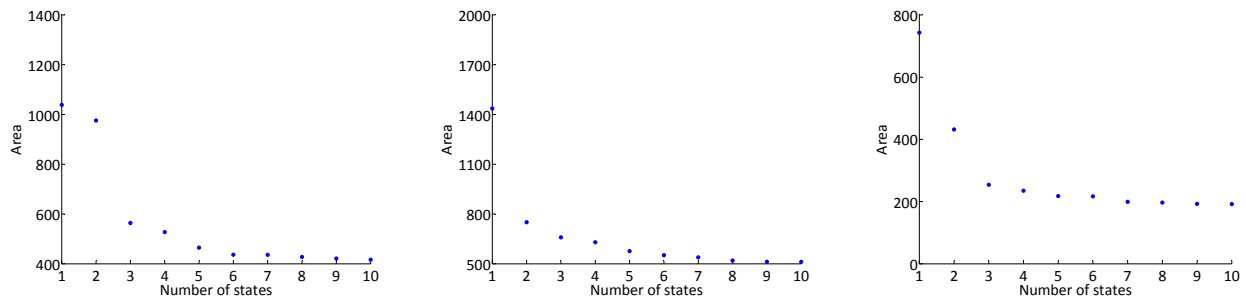

Figure 5: The goodness of fit metric versus the number of states for the off-peak period of classes 1, 2, and 4. Due to lack of space, the curve for class 3 is omitted. Note that the Y axis is exaggerated for emphasis.

	Observed (seconds)	Expected (seconds)
On-peak period	0	55.2
Mid-peak period	0	41.4
Off-peak period	0	69.1

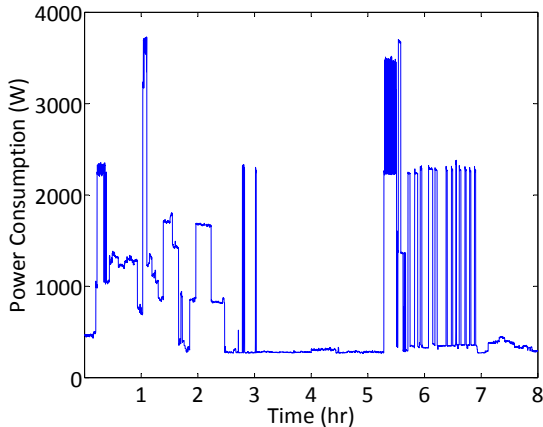**Table 5: Comparison of loss durations observed in numerical simulation and expected from the teletraffic theory.**

Figure 6: On-peak load measured from a house in class 1

Figure 7: On-peak load trace generated from the on-peak model of class 1

This is done simply by summing the actual measured load from each home at each time instant and checking if this sum exceeds the transformer sizing. If our models are representative, then the probability of overflow computed using teletraffic theory should be an upper bound on the numerical estimate. Table 5 shows the results, indicating that our reference models are representative, at least for the purpose of transformer sizing in the electrical grid. This result, though positive, does not give a good 'feel' for our models. A less analytical, but still useful, validation is by visual inspection. Figure 6 shows actual measurements of on-peak loads of a home in class 1 obtained from the test data set. We see that it is similar to the on-peak home load generated from the on-peak model of class 1 (Figure 7). This was true for all of our other models as well.

7. CONCLUSION

This paper reports very encouraging preliminary results on load modelling. Using an existing classification, we derive per class k-state Markovian reference models for different periods of the day. We show that none of these models need more than 6 states. We also show that these models are accurate enough for transformer sizing in the distribution network.

8. REFERENCES

[1] Current cost. http://www.currentcost.com/, January 2011.
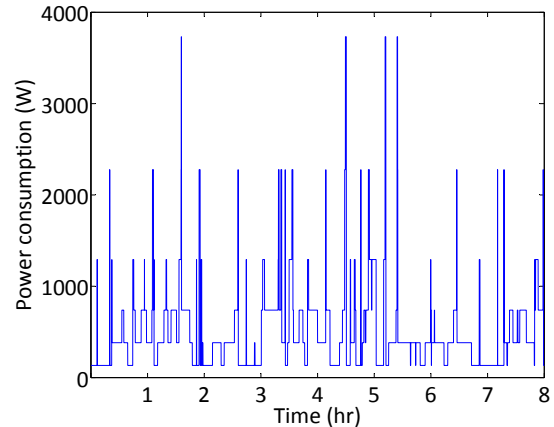[2] O. Ardakanian. Resource provisioning in the electrical grid. Thesis, University of Waterloo, 2011.
[3] M. Armstrong, M. Swinton, H. Ribberink, I. Beausoleil-Morrison, and J. Millette. Synthetically derived profiles for representing occupant-driven electric loads in Canadian Housing. *J. of Building Performance Simulation*, 2(1):15–30, 2010.
[4] A. Capasso, W. Grattieri, R. Lamedica, and A. Prudenzi. A bottom-up approach to residential load modeling. *IEEE Trans. Power Systems*, 9(2):957–964, 2002.
[5] W. Cochran. *Sampling techniques*. John Wiley & Sons Inc, 1977.
[6] R. Durbin, S. Eddy, A. Krogh, and G. Mitchison. *Biological Sequence Analysis*. Press, Cambridge U., 11 edition, 2006.
[7] A. Elwalid, D. Heyman, T. V. Lakshman, D. Mitra, and A. Weiss. Fundamental bounds and approximations for ATM multiplexers with applications to video teleconferencing. *IEEE JSAC*, 13(6):1004–1016, 1995.
[8] C. D. Manning and H. Schuetze. *Foundations of Statistical Natural Language Processing*. The MIT Press, 1 edition, 1999.
[9] J. Paatero and P. Lund. A model for generating household electricity load profiles. *International J. of Energy Research*, 30(5):273–290, 2006.
[10] L. Rabiner and B. Juang. An introduction to hidden markov models. *ASSP Magazine, IEEE*, 3(1):4 – 16, jan 1986.
[11] I. Richardson, M. Thomson, D. Infield, and C. Clifford. Domestic electricity use: A high-resolution energy demand model. *Energy and Buildings*, 2010.
[12] J. W. Roberts. Performance evaluation and design of multiservice networks. COST 224 Final Report, Commission of the European Communities, October 1992.
[13] P. Sen, B. Maglaris, N.-E. Rikli, and D. Anastassiou. Models for packet switching of variable-bit-rate video sources. *Selected Areas in Communications, IEEE Journal on*, 7(5):865 –869, June 1989.