Fundamental Limits of Nonintrusive Load Monitoring*

Roy Dong, Lillian Ratliff, Henrik Ohlsson, and S. Shankar Sastry
Dept. of Electrical Engineering and Computer Sciences
UC Berkeley
Berkeley, CA, USA
roydong@eecs.berkeley.edu, ratliffl@eecs.berkeley.edu, ohlsson@eecs.berkeley.edu,
sastry@eecs.berkeley.edu

ABSTRACT

Provided an arbitrary nonintrusive load monitoring (NILM) algorithm, we seek bounds on the probability of distinguishing between scenarios, given an aggregate power consumption signal. We introduce a framework for studying a general NILM algorithm, and analyze the theory in the general case. Then, we specialize to the case where the error is Gaussian. In both cases, we are able to derive upper bounds on the probability of distinguishing scenarios. Finally, we apply the results to real data to derive bounds on the probability of distinguishing between scenarios as a function of the measurement noise, the sampling rate, and the device usage.

Categories and Subject Descriptors

H.1.1 [**Information Systems**]: Models and Principles—*energy disaggregation; nonintrusive load monitoring (NILM); performance bounds*

1. INTRODUCTION

Nonintrusive load monitoring (NILM) is a general term which refers to determining the energy consumption of individual devices, or statistics of the energy consumption signal, without installing individual sensors at the plug level. The goals of different NILM algorithms include event detection, i.e. determine when certain devices switch states, and energy disaggregation, i.e. recovering the power consumption signals of each device in its entirety from the aggregate signal. In many cases, we would like to have the latter for many households, but installing sensors on every plug in each house is prohibitively expensive and intrusive. For example, studies have shown that merely providing users feedback on their energy consumption patterns is sufficient to improve their consumption behaviors [12, 17, 1]. Forecasts predict that 20% savings in residential buildings are attainable with the use of personalized recommendations based on disaggregated data. Additionally, these savings are sustainable over long time periods, and are not transient effects of introducing new interfaces to users. These device-level measurements can further be used for strategic marketing of energy-saving programs and rebates, both improving efficacy of the programs and reducing costs.

NILM algorithms can help guide regulation for privacy policies in advanced metering infrastructures (AMIs) [4]. Analyzing NILM algorithms is a way to determine how much device-level information is contained in an aggregate signal. This information is critical to understanding the privacy concerns in AMIs and which parties should have access to aggregate power consumption data. Further, NILM algorithms can provide a good benchmark for defining privacy risk; the state–of–the–art NILM algorithm may be a reasonably conservative model for an adversary. For example, if we use the framework defined in [8, 9], we can analyze how much energy disaggregation an adversary can achieve with a prior on the device usage patterns and models for individual devices.

Technologies and algorithms are constantly evolving, and to the best of our knowledge, there has not yet been an attempt to analyze the fundamental limits of NILM algorithms. An understanding of the fundamental limits can provide a theoretical guarantee of privacy, if we conclude that disaggregation is impossible in a certain scenario. It can be used in the design of AMIs, by determining a minimum sampling rate, sensor accuracy, and network capacity to achieve a desired goal. Further, it may allow us to determine how many measurements actually need to be stored and transmitted.

*The work presented is supported by the NSF Graduate Research Fellowship under grant DGE 1106400, NSF CPS:Large:ActionWebs award number 0931843, TRUST (Team for Research in Ubiquitous Secure Technology) which receives support from NSF (award number CCF-0424422), and FORCES (Foundations Of Resilient CybEr-physical Systems), the European Research Council under the advanced grant LEARN, contract 267381, a postdoctoral grant from the Sweden-America Foundation, donated by ASEA's Fellowship Fund, and by a postdoctoral grant from the Swedish Research Council.

HiCoNS April 15-17, 2014, Berlin, Germany
ACM 978-1-4503-2652-0/14/04.http://dx.doi.org/10.1145/2566468.2576849.

In this paper, we study the fundamental limits of NILM algorithms. We consider a building containing a number of devices. Given the aggregate power consumption of these devices, we would like to distinguish between two scenarios, e.g. whether or not a light turns on, or whether it was a toaster or kettle that turned on. In particular, provided an arbitrary NILM algorithm, we seek bounds on the probability of distinguishing two scenarios given an aggregate power consumption signal. Additionally, once we have this theory developed for two scenarios, we generalize to find an upper bound on the probability of distinguishing between a finite number of scenarios. With this theory of the fundamental limits of NILM in hand, we address questions about the possibility of NILM in the context of AMIs. Further, using high-frequency, high-resolution measurements of power consumption signals of common household devices as the ground truth, we analyze the probability of successfully identifying common scenarios in a household. We also analyze the tradeoff between successful NILM and sensor/model accuracy, as well as sampling rate.

This paper is organized as follows. In Section 2, we review the relevant literature. We formulate the NILM problem and the model of NILM algorithms in Sections 3 and 4 respectively. In Section 5, we discuss the fundamental limits of NILM algorithms. We derive bounds on distinguishing a finite number of scenarios using a classical hypothesis testing framework. In Section 6, we focus on the case where the NILM model is deterministic with additive Gaussian noise. We derive analytical expressions for bounds on the probability of distinguishing scenarios. We apply the theory to real–data gathered on a number of household appliances in Section 7. Finally, in Section 8, we make concluding remarks.

2. BACKGROUND

The problem of NILM is essentially a single-channel source separation problem: determine the power consumption of individual devices given their aggregated power consumption. The source separation problem has a long history in information theory and signal processing and well known methods include the infomax principle [2], which tries to maximize some output entropy, the maximum likelihood principle [5], which uses a contrast function on some distribution on the source signals, and a time-coherence principle [3], which assumes time-coherence of the underlying source signals. These often lead to formulations which use some variation of a principle components analysis (PCA) or independent component analysis (ICA).

The most common applications of the source separation theory is to audio signals and biomedical signals. For these applications, it is often assumed that source signals are i.i.d. stationary processes. We note that power consumption signals are very different from these types of signals. The power consumption of a device has strong temporal correlations and are not stationary, e.g. whether or not a device is on at a given time is correlated with whether or not it was on an instant ago, and the mean power consumption signal changes with the state of the device. The algorithmic and theoretical development in source separation have therefore not been successfully applied to NILM and most methods for NILM are rather different to those developed for classical source separation.

The field of NILM is much younger than source separation and most development has focused on algorithms. We briefly outline a few approaches here. One approach has focused on the design of hardware to best detect the signatures of distinct devices [18, 13, 11], but algorithms to handle the hardware's measurements are still an open problem. Another approach which has been taken by much of the machine learning community is to use hidden Markov models (HMMs), or some variation, to model individual devices [16, 15, 20]; energy disaggregation can be done with an expectation maximization (EM) algorithm. In recent publications [8, 9], we model individual devices as dynamical systems and use adaptive filtering. These are a few examples of concrete algorithms for NILM. For a more comprehensive review, we refer the reader to [1].

The discussion presented in this paper focus on the theoretical limitations of an arbitrary NILM algorithm. To the best of our knowledge, there has not been any previous work attempting to model the NILM problem in its full generality and derive theoretical bounds. The work is inspired by recent work in differential privacy [10, 6, 19, 14]. The underlying goal of differential privacy is to model privacy in a fashion that encapsulates arbitrary prior information on the part of the adversary and an arbitrary definition of what constitutes a privacy breach. The theory of differential privacy can be extended to give similar, but weaker, bounds to those derived in this paper.

3. THE PROBLEM OF NILM

As mentioned in Section 1, NILM has a variety of end uses. For each of these potential applications, the statistics of interest may be different. Thus, when we state the problem of NILM, we remain as general as possible to accommodate all these applications.

We are given an aggregate power consumption signal for a building. Let $y[t] \in \mathbb{R}$ denote the value of the aggregate power consumption signal at time t for $t = 1, \ldots, T$, and let $y \in \mathbb{R}^T$ refer to the entire signal. This signal is the aggregate of the power consumption signal of several individual devices:

$$y[t] = \sum_{i=1}^{D} y_i[t] \text{ for } t = 1, \ldots, T \tag{1}$$

where D is the number of devices in the building and $y_i[t]$ is the power consumption of device i at time t.

There are many possible goals of NILM. For example, the energy disaggregation problem is to recover y_i for $i = 1, 2, \ldots, D$ from y. Another goal commonly studied is to recover information about the y_i from y, such as when lights turn on or the power consumption of the fridge over a week.

Generally, we will refer to the phenomena we wish to distinguish as scenarios throughout this paper.

4. MODEL OF NILM ALGORITHMS

In this section, we outline a general framework for analyzing the problem outlined in Section 3. At a high level, the framework can be summarized as follows. First, any NILM method must choose some representation for individual devices; these can be seen as functions from some input space to $\mathbb{R}^T$. Depending on the purpose of the NILM algorithm, the input space will vary; essentially, scenarios we wish to distinguish should correspond to different inputs in the input space. Then, we describe NILM algorithms as functions on the observed aggregate signal. The definition is meant to be general and hold across both generative and discriminative techniques.

4.1 Aggregate device model

Formally, let $(\Omega, \mathcal{F}, P)$ denote our probability space. As in Section 3, D denotes the number of devices and T denotes the length of our observed power signal.

Let $\mathcal{U}_i$ denote the input space for for the ith device. Inputs represent scenarios we wish to distinguish. The output space, representing the power consumption signal of an individual device, is $\mathbb{R}^T$ for every device. Then, the model associated with the ith device can be denoted as $G_i : \mathcal{U}_i \times \Omega \to \mathbb{R}^T$. Here, we have the condition that, for any $u_i \in \mathcal{U}_i$, $G_i(u_i, \cdot)$ is a random variable. Finally, let $\mathcal{U} = \mathcal{U}_1 \times \mathcal{U}_2 \times \ldots \times \mathcal{U}_D$, and let $G : \mathcal{U} \times \Omega \to \mathbb{R}^T$ be defined as $G((u_1, u_2, \ldots, u_D), \omega) = \sum_{i=1}^{D} G_i(u_i, \omega)$. Here, G denotes our aggregated system, i.e. the model of our building.

Assumption 1. Given that the input is $u \in \mathcal{U}$, the distribution of the power consumption is $G(u, \cdot)$.

We emphasize the generality of this framework. Many state-of-the-art methods can be formulated in this framework. For example, factorial hidden Markov model methods [16, 15, 20] can be thought of as single-input, single-output systems where the input is the state of the underlying Markov chains. The Markov transition probabilities become a prior on the input signal. In previous work [8, 9], we formulated the models as dynamical systems whose inputs are real-valued and correspond to the device usage. Thus, we now have a general way of expressing different models of devices in a NILM problem.

4.2 NILM algorithms

An algorithm for NILM will be a function of our observed aggregate power consumption signal. Its result will depend on the goal of the algorithm, and the end use of the algorithm output. For example, it could be the set of possible estimated disaggregated energy signals, $\{\widehat{y}_i\}_{i=1}^{D}$, or the set of possible discrete event-labels on our time-series data, or a set of statistics on the disaggregated data.

More formally, let S represent some NILM algorithm and $\mathcal{Z}$ represent its output space, discussed above. Then, the algorithm could be thought of as a function $S : \mathbb{R}^T \to \mathcal{Z}$. We will analyze a general S in the following section.

5. FUNDAMENTAL LIMITS OF NILM

In this section, we derive an upper bound on the probability of successfully distinguishing two scenarios with any NILM algorithm. Then, we extend these results to handle the case where we wish to upper bound the probabilities of distinguishing a finite set of scenarios, as well as two collections of scenarios. Note that in our framework, scenarios correspond to inputs to our device models, and we will use the two terms interchangeably.

5.1 Distinguishing two scenarios

First, fix any two inputs $v_0, v_1 \in \mathcal{U}$ which we wish to distinguish. For example, we may pick v_0 and v_1 so that they differ only in the usage of one device. In that case, we are analyzing the difference in observed output caused by whether or not, say, a microwave turns on in the morning. Alternatively, we may choose inputs that correspond to more dissimilar scenarios, such as whether or not a household uses an air conditioner at all. The choice of v_0, v_1 depends on which scenarios we wish to distinguish in our NILM algorithm.

As mentioned previously, let $S : \mathbb{R}^T \to \mathcal{Z}$ denote any NILM algorithm. Then, let $I : \mathcal{Z} \to \{0, 1\}$ be an indicator for whether or not an algorithm output satisfies some condition. For example, I could output 1 if a particular discrete phenomena, e.g. a light turning on, is detected in the algorithm output, and 0 otherwise. Or, I could output 1 if the estimated power consumption signals of individual devices lies in a certain set.

Suppose that this indicator is supposed to capture whether our algorithm believes the input is v_0 or v_1. That is, $(I \circ S)$ should output 1 if the NILM algorithm believes the input is v_1 and 0 if it believes the input is v_0. For this reason, from this point forward we will refer to I as our discriminator.

Assumption 2. $(I \circ S)$ is measurable, i.e. $(I \circ S)^{-1}(\{1\})$ is a measurable set in $\mathbb{R}^T$, with respect to the Borel field on $\mathbb{R}^T$.

We note that this is a reasonable assumption, as most, if not all, NILM algorithms in practice will be a finite composition of measurable functions.

Additionally, we note that this is a very conservative understanding of an NILM algorithm. In general, these algorithms are not be designed simply to distinguish between v_0 and v_1, and are likely not to be optimal in this regard. Thus, by analyzing an optimal $(I \circ S)$, we have a conservative upper bound on the probability of distinguishing v_0 and v_1. In particular, the scenarios v_0 and v_1 may contain additional information, so our optimal separator is allowed to use side information, such as the switching times of devices, when doing inference, making our bound more conservative.

Furthermore, we can contrast our contribution with existing work in differential privacy. Whereas differential privacy would consider any v_0 and v_1 that are adjacent, and bound the change in distributions for a fixed mechanism, here we fix a particular v_0 and v_1 and consider a bound on the performance of any mechanism.

Thus, we can formulate this in classical hypothesis testing frameworks seen in the statistics literature [7]. Our main contribution is the abstraction of the task of NILM that allows us to use well-known results in detection theory.

Let y denote our observed signal. Suppose that $G(v_0, \cdot)$ has a probability density function (pdf) f_0 and similarly $G(v_1, \cdot)$ with f_1. Let our likelihood ratio be defined as:

$$L(y) = \frac{f_1(y)}{f_0(y)} \tag{2}$$

The maximum likelihood estimator (MLE) finds the input that maximizes the likelihood of our observations. The MLE is given by:

$$\widehat{u}_{\text{MLE}}(y) = \begin{cases} v_1 \text{ if } L(y) \geq 1 \\ v_0 \text{ otherwise} \end{cases} \tag{3}$$

If we have a prior p on the probability of v_0 or v_1 as inputs, we can find the maximum a posteriori (MAP) estimate. This finds the input that is most likely given our observations and prior. The MAP is:

$$\widehat{u}_{\mathrm{MAP}}(y) = \begin{cases} v_1 \text{ if } L(y) \geq \frac{p(v_0)}{p(v_1)} \\ v_0 \text{ otherwise} \end{cases} \tag{4}$$

Note that this prior can be a discrete distribution or a density. However, for simplicity, we'll treat the prior as a discrete distribution throughout this paper; small notational changes are required for the prior to be a density.

Now, suppose we have a maximum acceptable probability of mislabeling the input v_1; let this parameter be denoted $\beta > 0$. Also, let u denote the true input. The optimal estimator with this constraint is:

$$\begin{array}{rl} \min_{\widehat{u}} & P(\widehat{u} = v_1 | u = v_0) \\ \text{subject to} & P(\widehat{u} = v_0 | u = v_1) \leq \beta \end{array} \tag{5}$$

By the Neyman-Pearson lemma, the non-Bayesian detection problem in Equation 5 has the following solution:

$$\widehat{u}_{\mathrm{NB}}(y) = \begin{cases} v_1 \text{ if } L(y) \geq \lambda \\ v_0 \text{ otherwise} \end{cases} \tag{6}$$

where λ is chosen such that $P(\widehat{u}_{\mathrm{NB}} = v_0 | u = v_1) = \beta$.

Throughout the rest of this paper, we will consider the MAP, but these can be extended to the other two cases. The probability of interest is the probability of successful NILM:

Definition 1. For the two-input case, the *probability of successful NILM* for an estimator $\widehat{u}$ is:

$$\sum_{i=0}^{1} P(\widehat{u}(y) = v_i | u = v_i) p(u = v_i) \tag{7}$$

This can be explicitly calculated given the densities and the prior. Additionally, any algorithm and discriminator ($I \circ S$) will perform worse than $\widehat{u}_{\mathrm{MAP}}$, so the MAP estimate provides an upper bound on any algorithm's probability of successful NILM.

Proposition 1. Any estimator $\widehat{u}$ will have a probability of successful NILM bounded by:

$$\sum_{i=0}^{1} P(\widehat{u}_{\mathrm{MAP}}(y) = v_i | u = v_i) p(u = v_i) \tag{8}$$

5.2 Distinguishing a finite number of scenarios

This easily extends to distinguishing between a finite number of scenarios. Let V denote a finite set of inputs. Then:

Definition 2. For the N-input case, the *probability of successful NILM* for an estimator $\widehat{u}$ is:

$$\sum_{i=1}^{N} P(\widehat{u}(y) = v_i | u = v_i) p(u = v_i) \tag{9}$$

The MAP is given by:

$$\widehat{u}_{\mathrm{MAP}}(y) = \arg\max_{v \in V} P(G(u, \cdot) = y | u = v) p(u = v) \tag{10}$$

Proposition 2. There is an upper bound to the probability of successful NILM provided by the MAP:

$$\sum_{i=1}^{N} P(\widehat{u}_{\mathrm{MAP}}(y) = v_i | u = v_i) p(u = v_i) \tag{11}$$

5.3 Distinguishing two collections of scenarios

This philosophy of deriving an upper bound extends nicely to whenever we wish to distinguish two collections of scenarios. This corresponds to distinguishing two sets of inputs.

Now, suppose we have two sets of inputs: V_0 and V_1. We can still define the probability of successful NILM in this context:

Definition 3. For the case where we wish to distinguish two sets of inputs, the *probability of successful NILM* for an estimator $\widehat{u}$ is:

$$\sum_{i=0}^{1} P(\widehat{u}(y) \in V_i | u \in V_i) p(u \in V_i) \tag{12}$$

Depending on the context, this quantity may be calculable. In other cases, it may be possible to find good approximations or upper bounds. We will see this arise in Section 6.

6. GAUSSIAN CASE

In this section, we instantiate our theory on the special case where our model is a deterministic function with additive Gaussian noise.

6.1 Two scenarios

Suppose our system takes the following form:

$$G(u, \omega) = h(u) + w(\omega) \tag{13}$$

where $h : \mathcal{U} \to \mathbb{R}^T$ is a deterministic function and w is a random variable. Furthermore, fix any two inputs v_0, v_1 which we wish to distinguish, and suppose that w is a zero-mean Gaussian random variable with covariance Σ. Furthermore, suppose our prior is $p(u = v_0) = p(u = v_1) = 0.5$.

This can encapsulate the case where the uncertainty arises from measurement noise and model error. Referring to our motivating example, suppose that the only difference between v_0 and v_1 is the presence of a toaster turning on once in v_1. The question we are asking is: can we detect the toaster turning on?

Then, let f_0 denote the Normal pdf with mean $h(v_0)$ and covariance Σ, and similarly let f_1 be the Normal pdf with mean $h(v_1)$ and the same covariance Σ. For shorthand, let $\mu_0 = h(v_0)$ and $\mu_1 = h(v_1)$.

Since the covariance matrix Σ is the same for both random variables, $\widehat{u}_{\mathrm{MAP}}$ is determined by a hyperplane. Let $a^\top = (\mu_0 - \mu_1)^\top \Sigma^{-1}$ and $b = \frac{1}{2}\left(\mu_1^\top \Sigma^{-1} \mu_1 - \mu_0^\top \Sigma^{-1} \mu_0\right)$. Then:

$$\widehat{u}_{\mathrm{MAP}}(y) = \begin{cases} v_1 & \text{if } a^\top y + b \leq 0 \\ v_0 & \text{otherwise} \end{cases} \tag{14}$$

Now, suppose the input is actually v_0. That is, y is distributed according to f_0. Then, the signed distance from y to the boundary of the hyperplane is given by $\frac{1}{\|a\|_2}(a^\top y + b)$. This is a linear function of Gaussian random variable, and is thus also a Gaussian random variable. Furthermore, the mean of this random variable will be $\frac{1}{\|a\|_2}(a^\top \mu_0 + b)$, and the variance will be:

$$\sigma^2 = \frac{1}{\|a\|_2^2} a^\top \Sigma a = \frac{(\mu_0 - \mu_1)^\top \Sigma^{-1} (\mu_0 - \mu_1)}{(\mu_0 - \mu_1)^\top \Sigma^{-2} (\mu_0 - \mu_1)} \tag{15}$$

Thus, given that the input is actually v_0, the probability that $\widehat{u}_{\mathrm{MAP}}(y) = v_0$ is:

$$\begin{aligned} &P(\widehat{u}_{\mathrm{MAP}}(y) = v_0 | u = v_0) \\ &\quad = \frac{1}{2}\left(1 - \operatorname{erf}\left(\frac{-\frac{1}{\|a\|_2}(a^\top \mu_0 + b)}{\sqrt{2\sigma^2}}\right)\right) \end{aligned} \tag{16}$$

where erf is the Gauss error function and Equation 16 is simply the 1 minus the cumulative distribution function (cdf) of the distance to the hyperplane evaluated at 0, i.e. the probability that the signed distance is positive.

The computations are exactly the same for the case where the input is v_1. Thus:

Proposition 3. By Equation 8, the probability of successfully distinguishing v_0 and v_1 with the MAP is given by:

$$\frac{1}{2}\left(1 - \operatorname{erf}\left(\frac{-\frac{1}{\|a\|_2}(a^\top \mu_0 + b)}{\sqrt{2\sigma^2}}\right)\right) \tag{17}$$

Note that, in general, disaggregation algorithms would not be designed simply to distinguish between v_0 and v_1, and are likely not to be optimal in this regard. That is, Equation 17 provides a theoretical upper bound on how good any possible disaggregation algorithm could perform in distinguishing v_0 and v_1. Also, note that $\frac{1}{\|a\|_2}(a^\top \mu_0 + b)$ will be positive if $\mu_0 \neq \mu_1$. It follows that the upper bound is always greater than 0.5 if $\mu_0 \neq \mu_1$, and the MAP achieves this upper bound. Thus, if the inputs cause different outputs from the system, there will always exist an algorithm that improves the discrimination between v_0 and v_1 over blind guessing.

6.2 N scenarios

In this subsection, we build on the development in Section 6.1 to handle the case where we wish to distinguish several inputs.

Suppose now that we have a finite set of inputs that we wish to distinguish. Consider the set $\{u_i\}_{i=1}^N$, where $u_i \in \mathcal{U}$ for each i. Again, suppose all these inputs are equally likely, i.e. $p(u = v_i) = \frac{1}{N}$ for all i. We wish to find the MAP. We carry over the assumption of Gaussian noise with variance Σ. The MAP will partition $\mathbb{R}^T$ with hyperplanes of the form given in Section 6.1.

So, suppose the actual input is u_1. We wish to ask: what is the probability the MAP will accurately identify u_1 from the other $N-1$ inputs? Let $\mu_i = h(u_i)$ for $i = 1, \ldots, N$. Then, let $a_i^T = (\mu_1 - \mu_i)^T \Sigma^{-1}$ and $b_i = \frac{1}{2}(\mu_i^T \Sigma^{-1} \mu_i - \mu_1 \Sigma^{-1} \mu_1)$. Given our observation $y \in \mathbb{R}^T$, we wish to ask the probability that $\frac{1}{\|a_i\|_2}(a_i^T y + b_i) > 0$ for $i = 2, \ldots, N$, i.e. that the input u_1 is more likely than any of the other inputs. More succinctly, define:

$$A = \begin{bmatrix} a_2^T / \|a_2\|_2 \\ a_3^T / \|a_3\|_2 \\ \vdots \\ a_N^T / \|a_N\|_2 \end{bmatrix} \qquad b = \begin{bmatrix} b_2 / \|a_2\|_2 \\ b_3 / \|a_3\|_2 \\ \vdots \\ b_N / \|a_N\|_2 \end{bmatrix} \tag{18}$$

We wish to ask the probability that $Ay + b$ is in the positive orthant of $\mathbb{R}^N$. Recall that y is distributed according to mean μ_1 and covariance Σ. Thus, the random variable $Ay+b$ has mean $A\mu_1 + b$ with covariance $A\Sigma A^T$. The probability that this random variable is in the positive orthant cannot be analytically calculated, but can be approximated with high accuracy.

This can be done for $i = 2, \ldots, N$ as well, and provide an upper bound on the probability of successful NILM. An example based on real data will be explicated in Section 7.

6.3 Linear systems

In this subsection, we specialize the previous theory to the case where all our devices are linear systems. Suppose that the dynamics of our household are of the form $y = Au + e$, and our noise e has covariance $\widehat{\sigma}^2 I$. Note that σ^2 as defined in Equation 15 is equal to $\widehat{\sigma}^2$.

Now, suppose the sets that we wish to distinguish are $V_0 = \{0\}$ and $V_1 = \{v : L \leq \|v\|_2 \leq U\}$, for some constants $0 < L \leq U$. That is, can we detect an input with magnitude in the range $[L, U]$? By Equation 12, we have the probability of successful NILM for an estimator $\widehat{u}$ is:

$$P(\widehat{u}(y) = 0 | u = 0) p(u = 0) + P(\widehat{u}(y) \in V_1 | u \in V_1) p(u \in V_1) \tag{19}$$

First, consider a fixed input $v \in V_1$. If we suppose that $u = v$, then the probability of an estimator $\widehat{u}$ distinguishing v from 0 is bounded by:

$$P(\widehat{u}(y) \neq 0 | u = v) \leq \frac{1}{2}\left(1 + \operatorname{erf}\left(\frac{\|Av\|_2}{2\sqrt{2\sigma^2}}\right)\right) \tag{20}$$

This can be seen by noting that, after a projection into one-dimension, the separating hyperplane is the point $\pm\|Av\|_2/2$. Without loss of generality, let us suppose the separating point is $\|Av\|_2/2$.

Note that this equation is an increasing function of $\|Av\|_2$. This gives us:

$$\frac{1}{2}\left(1 + \operatorname{erf}\left(\frac{\|Av\|_2}{2\sqrt{2\sigma^2}}\right)\right) \leq \frac{1}{2}\left(1 + \operatorname{erf}\left(\frac{\sigma_{\max}(A) U}{2\sqrt{2\sigma^2}}\right)\right) \tag{21}$$

where $\sigma_{\max}(A)$ is the largest singular value of A. This held for any $v \in V_1$, so measure-theoretic properties give us:

$$P(\widehat{u}(y) \in V_1 | u \in V_1) \leq \frac{1}{2}\left(1 + \operatorname{erf}\left(\frac{\sigma_{\max}(A) U}{2\sqrt{2\sigma^2}}\right)\right) \tag{22}$$

Proposition 4. In the linear system case, the probability of successful NILM is bounded above by:

$$p(u = 0) + \frac{1}{2}\left(1 + \operatorname{erf}\left(\frac{\sigma_{\max}(A) U}{2\sqrt{2\sigma^2}}\right)\right) p(u \in V_1) \tag{23}$$

These are bounds which do not depend explicitly on a model, but rather only on the sensitivity of the model. Thus, even with just knowledge of the variance of the noise and the sensitivity of our linear systems, we can still find an upper bound on the probability of successful NILM.

7. REAL DATA ANALYSIS

In this section, we take the theory from Sections 5 and 6 and use them on real data to address several different problems. We used the emonTx wireless open-source energy monitoring node from OpenEnergyMonitor[1] from several devices at 12Hz. We used current transformer sensors

[1] `http://openenergymonitor.org/emon/emontx`

and an alternating current (AC) to AC power adapter to measure the current and voltage respectively of the devices that we monitored. For each device we measured the root-mean-square (RMS) current, RMS voltage, apparent power, real power, power factor, and a UTC time stamp.

Data was recorded in a laboratory setting for a microwave, a toaster, a kettle, an LCD computer monitor, a projector, and an oscilloscope. As our sensors are highly accurate, we treat the measurements as noise free.

Problem 1. What is an upper bound for the probability of successfully detecting a toaster turning on, as a function of the modeling and measurement error?

We took measurements from a toaster. The basic signal is shown in Figure 1. We use the assumptions outlined in Section 6.1. Additionally, we assumed that the covariance of the Gaussian noise was $\sigma^2 I$, i.e. the noise was uncorrelated at each time step. Following the analysis in Section 6.1, the probability of distinguishing the toaster turning on is shown in Figure 2.

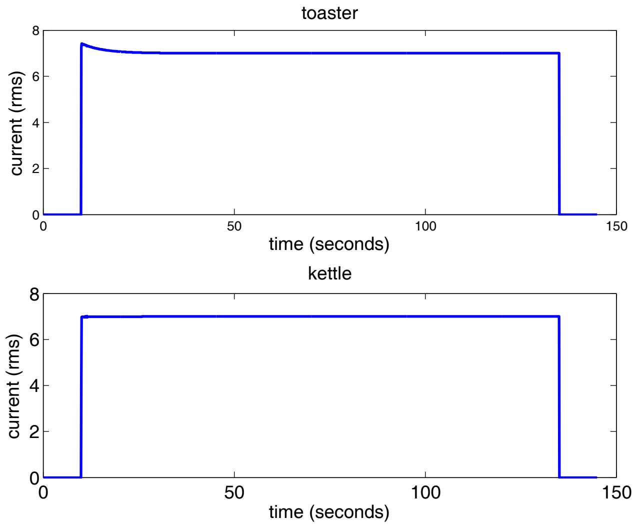

Figure 1: *Top:* The measured RMS current signal for a toaster. *Bottom:* The measured RMS current signal for a kettle.

Note that σ^2 has to grow considerably large before the optimal algorithm starts to fail to distinguish the toaster from nothingness. This is unsurprising, as the optimal algorithm would have several samples to distinguish quite separate means.

Problem 2. What is an upper bound for the probability of successfully distinguishing a toaster turning on and a kettle turning on, as a function of the modeling and measurement error?

We repeated this analysis with both a toaster and a kettle signal, depicted in Figure 1. The devices are on for exactly the same time window. The results are shown in Figure 3. As we can see, the variance on the error is orders of magnitude smaller when the probability drops to near 0.5. However, the σ^2 value is still quite large, and we likely can distinguish the two devices at 12Hz.

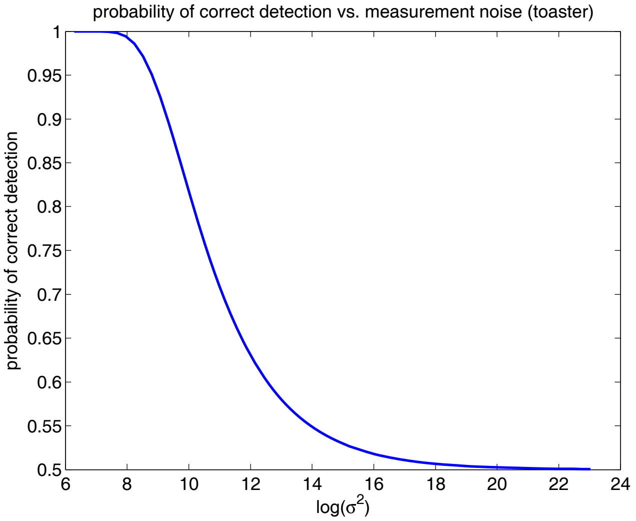

Figure 2: The probability of successful identification of a toaster as a function of modeling and measurement error.

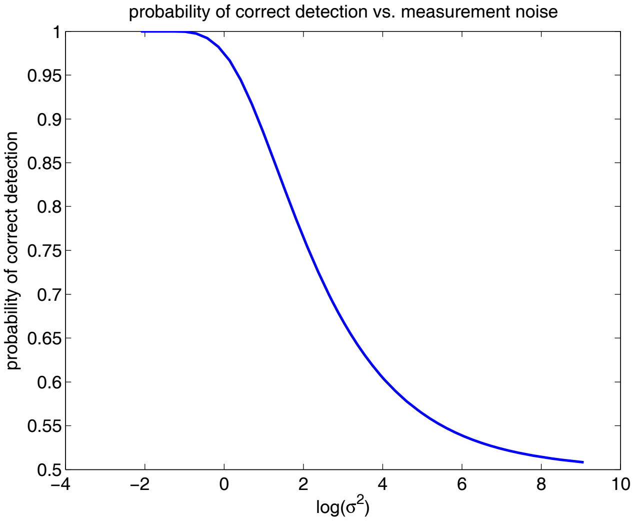

Figure 3: The probability of successfully discriminating of a toaster and a kettle as a function of modeling and measurement error.

Problem 3. What is an upper bound for the probability of successfully distinguishing a toaster turning on and a kettle turning on, as a function of the sampling rate?

The results to Problem 2 are promising, as they tell us it is very possible to distinguish two rather similar devices. However, the sampling rate of 12Hz is very high. Now, we analyze how likely we are to distinguish the two devices as the sampling rate changes. This is shown in Figure 4. We down-sampled the 12Hz signal. Additionally, if we down-sampled with rate K, we assumed it was equally likely that the signal would begin on any of the first K time-steps.

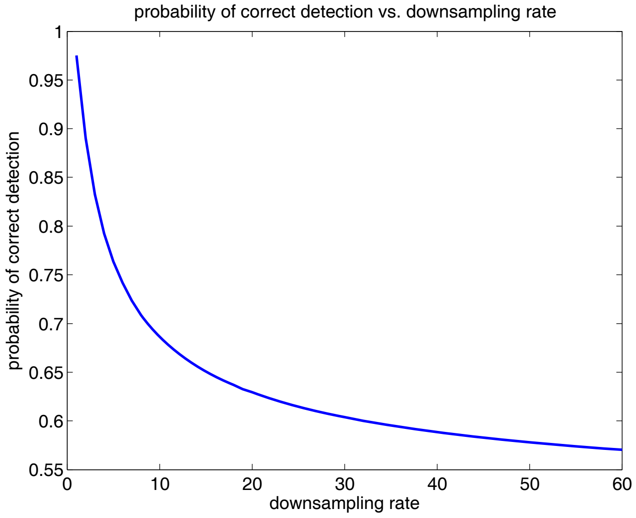

Figure 4: The probability of successfully discriminating a toaster and a kettle as a function of the sampling rate. We fixed $\sigma^2 = 1$.

It should be noted that a downsampling rate of K implies that we only receive $1/K$ as many measurements. Thus, if we sample for 1 second, our original problem would be a separation problem in $\mathbb{R}^{12000}$, whereas the downsampled problem is a separation problem in $\mathbb{R}^{\lfloor 12000/K \rfloor}$.

As expected, the probability of successful NILM decreases with the sampling rate. Additionally, the performance degrades quite quickly, and we barely perform better than guessing when the downsampling rate is 60, i.e. we sample every 5 seconds. This result allows us to determine a lower bound on the sampling rate necessary to achieve a certain effectiveness of NILM. It gives prescriptions on what hardware specifications and network capacity is needed in AMIs to achieve a certain goal.

Problem 4. What is an upper bound for the probability of successfully distinguishing several devices, as a function of measurement and modeling error?

Here, we use the results in Section 6.2. The devices in question are a microwave, a toaster, a kettle, an LCD computer monitor, a projector, and a digital oscilloscope. As before, we have one signal for each device, which is activated for the same time window for each device. That is, for each device, we have a *fixed* input. If each device is equally likely to turn on, we have the results shown in Figure 5.

Problem 5. What is an upper bound for the probability of successfully distinguishing several linear systems, as a function of the input magnitude?

Suppose we have the same 6 devices as in the previous problem. Furthermore, suppose they are linear systems, and the observed signals were a result of an input which was a pulse of magnitude 1. Then, we can use results from Section 6.3.

Suppose we wish to determine whether or not a device turned on. The input to our device is nonzero and bounded by U. We plot the upper bound from Section 6.3 on the probability of successful NILM as a function of U. The results are in Figure 6.

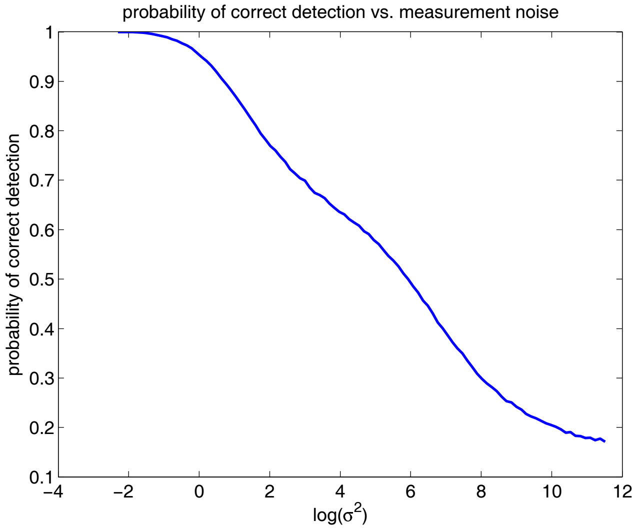

Figure 5: The probability of successfully discriminating one device from the $N-1$ other devices as a function of the measurement and modeling error.

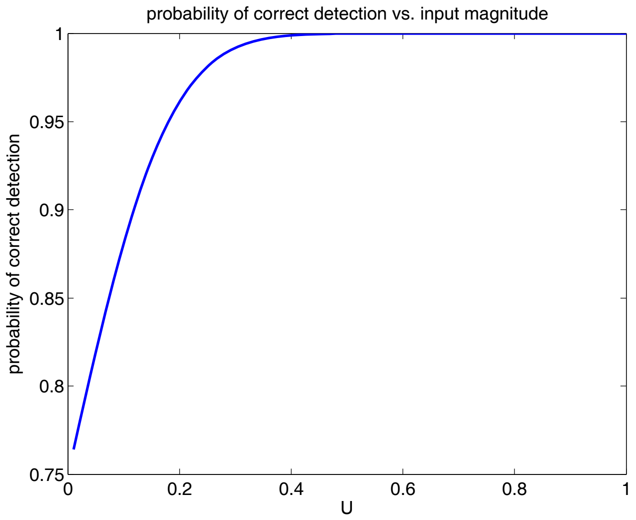

Figure 6: The probability of successfully discriminating a device turning on from the null hypothesis as a function of the input magnitude U. We fix $\sigma^2 = 1$.

8. CONCLUSION

In this paper, we explore the fundamental limits of NILM algorithms. More specifically, we derive an upper bound on the probability of distinguishing scenarios for an arbitrary NILM algorithm. First, we present the theory in its general case, and then we instantiate the theory on the case where the error is additive Gaussian noise independent of the underlying scenario. With this upper bound in hand, and our Gaussian assumption, we interpret real data we collected and discuss how the probability of successful NILM depends on the modeling and measurement error, the sampling rate, and the magnitude of the device usage.

To the best of our knowledge, this is the first paper investigating the fundamental limits of NILM. These fundamental limits are useful for several reasons. They can provide a guarantee on when NILM is impossible, which has implications for the design of privacy-aware AMIs, as well as privacy policies in AMIs. These limits are algorithm-independent, so they will hold regardless of changing technologies. These limits also can provide prescriptions for the design of AMIs, if NILM of a certain sort is desired, in terms of network capacity and sensor accuracy. Finally, it also provides a unified framework for understanding the problem of NILM.

9. ACKNOWLEDGMENTS

The authors would like to thank Alvaro Cárdenas for his helpful comments on an early draft of this document, as well as Aaron Bestick for stimulating conversations and assistance with our experimental setup.

10. REFERENCES

[1] K. C. Armel, A. Gupta, G. Shrimali, and A. Albert. Is disaggregation the holy grail of energy efficiency? The case of electricity. *Energy Policy*, 52:213–234, 2013.

[2] A. J. Bell and T. J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 1995.

[3] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines. A blind source separation technique using second-order statistics. *IEEE Transactions on Signal Processing*, 45(2):434–444, 1997.

[4] A. A. Cárdenas, S. Amin, G. Schwartz, R. Dong, and S. S. Sastry. A game theory model for electricity theft detection and privacy-aware control in AMI systems. In *Proceedings of the 50th Allerton Conference on Communication, Control, and Computing*, pages 1830–1837, 2012.

[5] J. Cardoso. Infomax and maximum likelihood for blind source separation. *IEEE Signal Processing Letters*, 4(4):112–114, 1997.

[6] K. Chaudhuri and D. Hsu. Sample complexity bounds for differentially private learning. In *COLT*, pages 155–186, 2011.

[7] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley-Interscience, 1991.

[8] R. Dong, L. Ratliff, H. Ohlsson, and S. S. Sastry. A dynamical systems approach to energy disaggregation. In *Proceedings of the 52nd IEEE Conference on Decision and Control (CDC)*, 2013.

[9] R. Dong, L. Ratliff, H. Ohlsson, and S. S. Sastry. Energy disaggregation via adaptive filtering. In *Proceedings of the 51th Allerton Conference on Communication, Control, and Computing*, 2013.

[10] C. Dwork. Differential privacy. In *Proceedings of the International Colloquium on Automata, Languages and Programming*, pages 1–12. Springer, 2006.

[11] J. Froehlich, E. Larson, S. Gupta, G. Cohn, M. Reynolds, and S. Patel. Disaggregated end-use energy sensing for the smart grid. *IEEE Pervasive Computing*, 10(1):28–39, 2011.

[12] G. T. Gardner and P. C. Stern. The short list: The most effective actions U.S. households can take to curb climate change. In *Environment: Science and Policy for Sustainable Development*, 2008.

[13] S. Gupta, M. S. Reynolds, and S. N. Patel. Electrisense: single-point sensing using EMI for electrical event detection and classification in the home. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*, Ubicomp '10, pages 139–148, New York, NY, USA, 2010. ACM.

[14] Z. Huang, S. Mitra, and G. Dullerud. Differentially private iterative synchronous consensus. In *Proceedings of the 2012 ACM Workshop on Privacy in the Electronic Society*, WPES '12, pages 81–90, New York, NY, USA, 2012. ACM.

[15] J. Z. Kolter and T. Jaakkola. Approximate inference in additive factorial HMMs with application to energy disaggregation. In *Proceedings of the International Conference on Artificial Intelligence and Statistics*, 2012.

[16] J. Z. Kolter and M. J. Johnson. REDD: A public data set for energy disaggregation research. In *Proceedings of the SustKDD Workshop on Data Mining Appliations in Sustainbility*, 2011.

[17] J. A. Laitner, K. Ehrhardt-Martinez, and V. McKinney. Examining the scale of the behaviour energy efficiency continuum. In *European Council for an Energy Efficient Economy*, 2009.

[18] S. Leeb, S. Shaw, and J. Kirtley, J.L. Transient event detection in spectral envelope estimates for nonintrusive load monitoring. *IEEE Transactions on Power Delivery*, 10(3):1200–1210, 1995.

[19] J. L. Ny and G. J. Pappas. Differentially private filtering. *arXiv:1207.4305*, July 2012.

[20] O. Parson, S. Ghosh, M. Weal, and A. Rogers. Nonintrusive load monitoring using prior models of general appliance types. In *Proceedings of the 26th AAAI Conference on Artificial Intelligence*, 2012.