BEARS | SinBerBEST

Efficient Power Disaggregation for intra-Building Smart Grid Applications

Zhaoyi Kang and Professor Costas J. Spanos
Department of Electrical Engineering and Computer Sciences
University of California, Berkeley

Supported by:

Funded by: NATIONAL RESEARCH FOUNDATION

Motivation

Smart grids suffer from performance degradation when metering becomes excessively detailed. Power disaggregation is a data mining technique that can be used to "extract" the behavior of individual appliances from a single, aggregate power signal, thus lowering sensor cost and network burden. Moreover, power disaggregation provides a more user friendly interface than high density network when privacy is a concern.

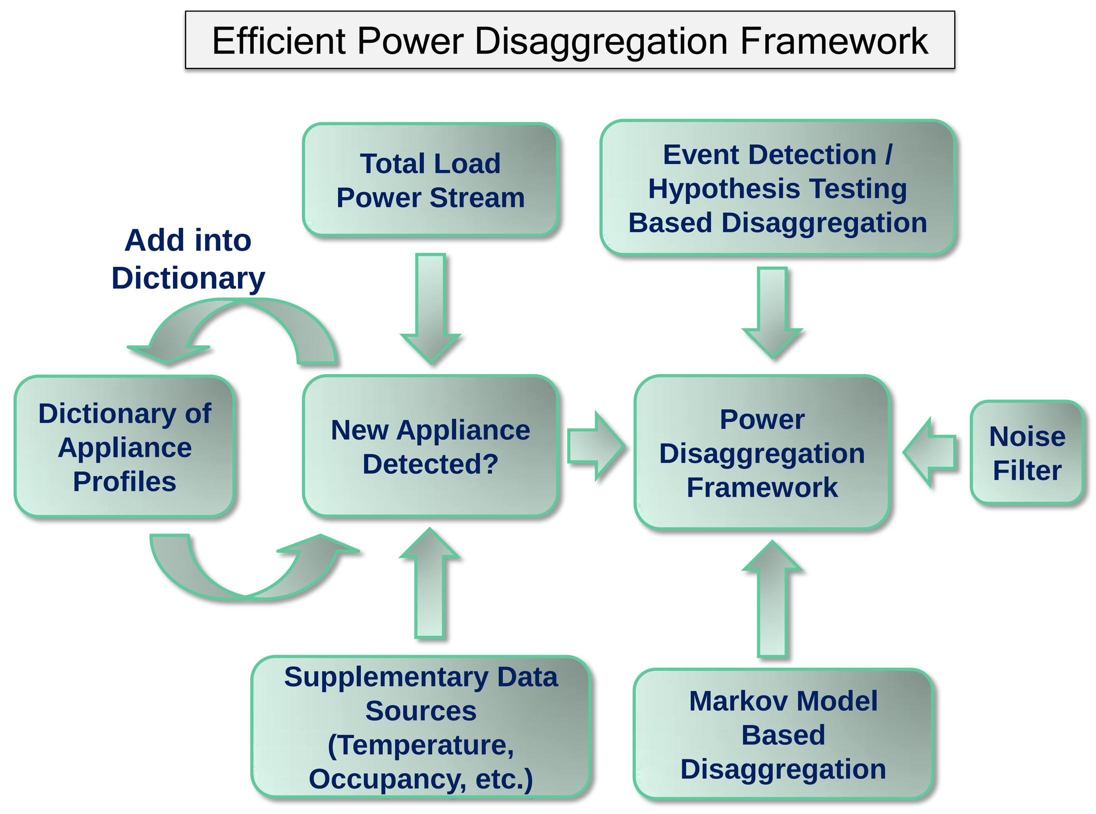
In this work, power disaggregation methods are explored in terms of accuracy, consistency and the potential for real-time implementation.

2012 Main Objectives

These are the aspects of power disaggregation that we have been working on this year:

- Design a non-intrusive power disaggregation framework;
- Improve the efficiency and accuracy of Markov Model based power disaggregation;
- Improve the efficiency and accuracy of Event Detection;
- Propose possible breakthrough based on previous methodologies;
- Examine the robustness of various methods while deploying in practical applications.

The Problem Framework

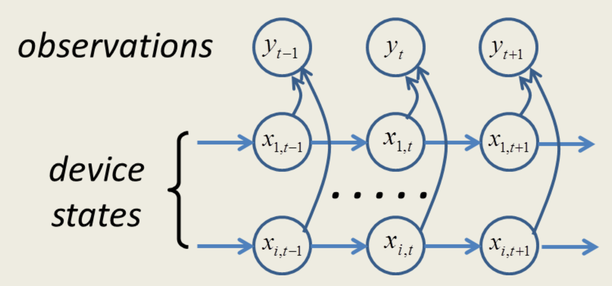


Two directions in Power Disaggregation

- **Markov Model Based Method:** view data stream as if generated from a Markov Chain
  - Benefit from early development in Hidden Markov Model (HMM);
  - Systematic theory for statistical inference and parameter learning;
- **Event Based Method:** estimate the state by detecting ON/OFF events
  - Benefit from the development in Non-intrusive Load Monitoring (NILM) model;
  - Event Detection is a core component;

Hidden Markov Model Power Disaggregation

- **Hidden Markov Model (HMM):** $\{x_{i,t}\}$ are the hidden state variables for the i[th] appliance, $\{y_t\}$ are observations:
- HMM includes the most information from the dataset



$\Pr(O_t|s_t = i) \sim N(O_t|\mu_t, \sigma^2)$ the observation distribution

$\mu_t = \mu_\infty + (\mu_0 - \mu_\infty)e^{-(t-t_0)/\tau}$ has exponential decay curve

$\Pr(x_t = i|x_{t-1} = j) = \pi_{j,i}$ the transition probability

- **Statistical Inference:** Standard Viterbi Algorithm;

Likelihood Function $\delta_t(i) = \max_j \delta_{t-1}(j)\pi_{j,i} P(O_t|s_t = i)$

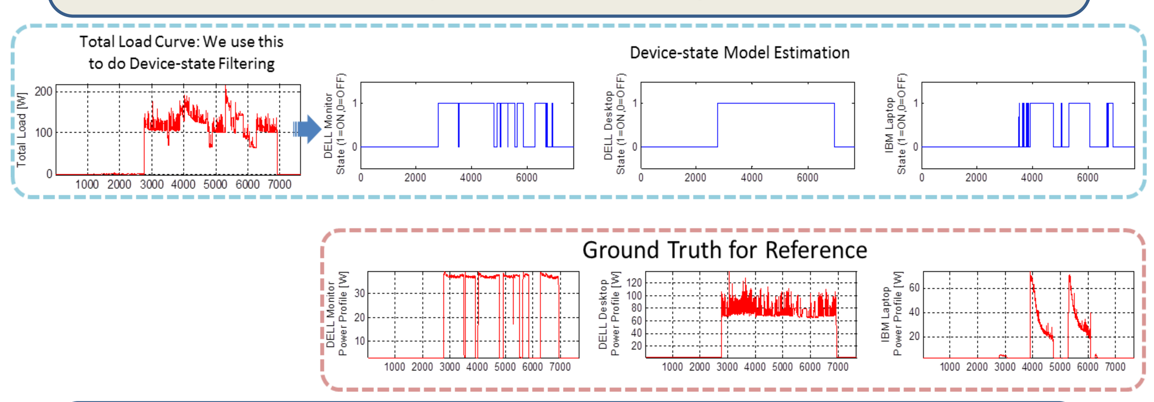
State Function $\psi_t(i) = arg\max_j \delta_{t-1}(j)\pi_{j,i}(\tau)$

Back Tracking $P_T = \max_j \delta_T(j)$ $P_{t-1} = \psi_t(P_t)$, for $t = T, \cdots, 2$

Note: $\delta_t(i)$ is the likelihood function of state i at step t

Simulation Results of HMM Power Disaggregation

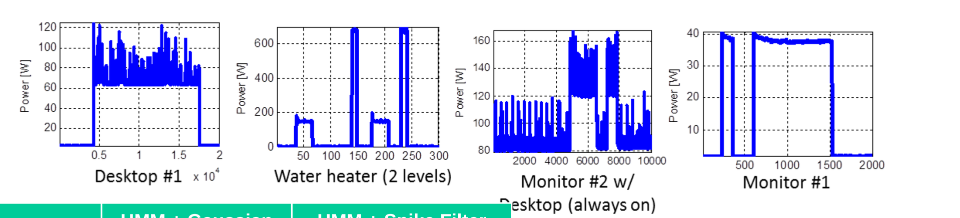
- **Example:** "appliances" are a desktop, a laptop and a monitor;
- Parameters of each appliance studied in advance
- **Results shown below:** (left is the total load)



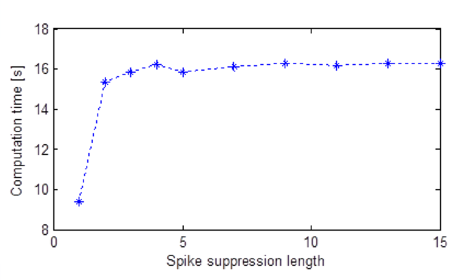
- Captures most important switching, **>85%** overall accuracy
- **Problems:** several false switches (hard to implement in practical application although have high accuracy);
- **Possible Improvement:** 1) Noise reduction; 2) Exploit Persistency;

Noise Cancellation by Spike Filtering

- **Spike Filtering (SF):**
  - Useful when the noise distribution is **highly skewed**
  - Trying SF on data stream composed by 5 devices



Appliance ID	HMM + Gaussian (Viterbi)	HMM + Spike Filter (Viterbi)
Monitor #1	25.7% / 49.1%	32.7% / 54.5%
Monitor #2	56.7% / 77.1%	59.8% / 87.8%
Desktop #1	96.0% / 94.2%	96.5% / 95.1%
Water Heater – Pump	94.3% / 94.7%	94.3% / 95.1%
Water Heater – Boiler	100% / 100%	100% / 100%
Computation	9.5 sec / 130,000	Left Figure



- Increases modestly computational burden;
- Can also modestly improve estimation accuracy;

Persistent State Duration by Hidden Semi-Markov Model

- **Markov Property:** first order dependency $P(s_t|s_{1:t-1}) = P(s_t|s_{t-1}) = p$
- Extend to semi-Markov case: higher order dependency;

$$P(s_t|s_{1:t-1}) = P(s_t|s_{t-1}, \ldots, s_{t-d}) = P(s_t|duration = d)$$

- **State Duration statistics** (Gamma distribution)
- **Benefits:** flexibility in describing non-linear or non-stationary data streams in Markov Model context.
- Modified Viterbi algorithm (higher order Markov Model)

Likelihood Function
$$\delta_t(i,d) = \begin{cases} \max_{j\neq i} \max_{2\le\tau\le D_j} \delta_{t-1}(j,\tau)\pi_{j,i}(\tau)P(O_t|s_t = i) & d = 1 \\ \delta_{t-1}(i, d-1)P(O_t|s_t = i) & d > 1 \end{cases}$$

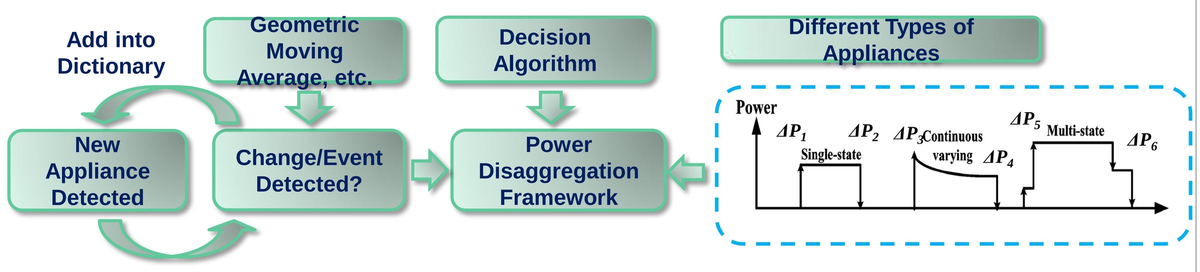
State Function
$$\psi_t(i,d) = \begin{cases} arg\max_{j\neq i} \max_{2\le\tau\le D_j} \delta_{t-1}(j,\tau)\pi_{j,i}(\tau) & d = 1 \\ (i-1, d) & d > 1 \end{cases}$$

$\delta_t(i,d)$ function is the likelihood function at the i[th] state given that it has duration time $= d$

- For higher order Markov Model, inference is very time consuming;
- Accuracy improvement is limited, if not worse;
- State Duration Model (usually Gamma distribution) is difficult to train.

Framework of Event Detection Power Disaggregation

- **Motivation:** State is constant within Segments, HMM is over-trained;



- **Two blocks:** *Change detection* & *State inference*

- **Change Detection:**
- Geometric Moving Average: adaptively capture the mean and variance with decay;
- **Threshold:** |Score|>3;

Mean: $\beta_0 = \frac{1}{d}\sum_{\tau=t-1}^{t-d} p_\tau e^{-\frac{\tau-t}{\omega}}$, **variance**

$$\sigma_0 = \left[\frac{1}{d}\sum_{\tau=t-1}^{t-d}(p_\tau - \beta_0)^2 e^{-(\tau-t)/\omega}\right]^{1/2}$$

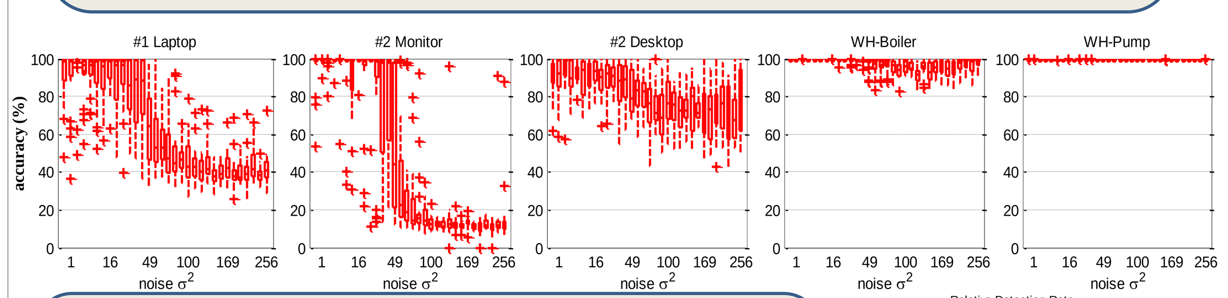
Score Function: $score = \frac{p_t - \beta_0}{\sigma_0}$

- **State Inference:**
- **Assumptions:** One appliance switch at each time;
- Calculate the most likely on/off appliance depend on the change ΔP_l and the duration statistics $f_1(\tau_l)$ (ON) or $f_0(\tau_l)$ (OFF);

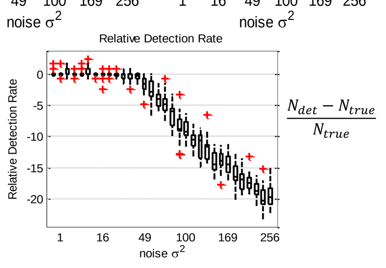
$$S_l = \begin{cases} arg\max_i P(\Delta P_l|S_l = i)f_0(\tau_l) & if\ \Delta P_l > 0 \\ arg\max_i P(\Delta P_l|S_l = i)f_1(\tau_l) & if\ \Delta P_l < 0 \end{cases}$$

Simulation of Event Detection Power Disaggregation

- **Simulated Power Stream** based on real data (Slide #7), with add-on background/sensor noise (with variance σ_{bn}^2);
- **Model of appliance:** $p(t, t_0) = \beta_s e^{-\frac{t-t_0}{\omega}} + \beta_e\left(1 - e^{-\frac{t-t_0}{\omega}}\right) + \epsilon$, $\beta_s =$ [60 38 65 150 690], $\beta_e =$ [23 35 65 145 685], $\epsilon \sim N(0, \sigma_i^2 + \sigma_{bn}^2)$;
- First Question: how does σ_{bn}^2 impact our disaggregation results



- Result gets worse as σ_{bn}^2 increase (lower than Markov Model Estimation), Reasons:
- States mixed together with larger σ_{bn}^2;
- Change detection degrades for larger σ_{bn}^2 (Left figure, change point detection rate decreases)



Other Improvement and Combination of Two Methods

- Alternative change detection algorithms

- Generalized Likelihood Ratio Test (GLRT)
- Exact event position search
- Computational intensive

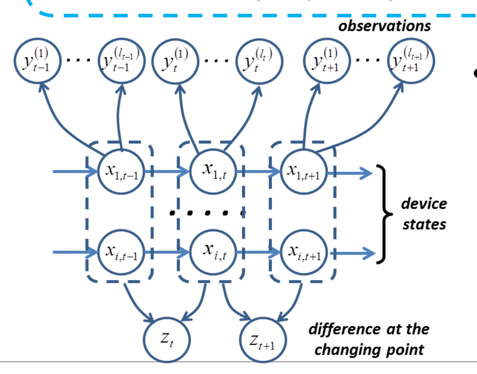
$$g_k = \max_{1\le j\le k} \sum_{i=j}^{k}\left[\frac{\hat{\nu}_j(y_i - \mu_0)}{\sigma^2} - \frac{\hat{\nu}_j^2}{2\sigma^2}\right]$$

$$|\hat{\nu}_j| = \left(\frac{1}{k-j+1}\sum_{i=j}^{k}|y_i - \mu_0| - \nu_m\right)^+ + \nu_m$$

$$\Lambda_n = \frac{f_1(Y_1, Y_2, .., Y_n)}{f_0(Y_1, Y_2, .., Y_n)} = \frac{L(Y_1, Y_2, ..., Y_n;\ H_1)}{L(Y_1, Y_2, ..., Y_n;\ H_0)}$$

$$\Lambda_n \approx \Lambda_{n-1} + \ln\frac{f_2(X_n; \mu_0 + \epsilon_2)}{f_2(X_n; \mu_0 - \epsilon_1)}$$

- Sequential Probability Ratio Test (SPRT)
- Fast
- Simple to implement



- Combination of the Two methods
  - Event detection is fast and inclusive to new items
  - Markov Model is more accurate but unnecessarily time consuming;
  - Detection for catching event and Markov model for decision

Conclusion and References

- Conclusions:
- Proposed a framework for efficient power disaggregation in commercial building;
- Analyze the performance or both Markov Model method and Event Detection and proposed ways to improve them;
- **Future Goals**: Work on improvement of the current method
  1. Improve the detection efficiency by more robust statistics, analyze the trade-off between accuracy and fast response;
  2. Data fusion, include other information for decision making;
  3. Implement the combined model and maximize the benefits

References:
1) G. W. Hart, Nonintrusive Appliance Load Monitoring, *Proceedings of the IEEE*, Vol. 80, pp. 1870-1891, 1992.
2) M. Zeifman et al., Viterbi Algorithm with Sparse Transitions (VAST), *IEEE Symposium on Computational Intelligence Applications In Smart Grid (CIASG), 2011*.
3) J.Z. Kolter and M.J. Johnson. Redd: A public data set for energy disaggregation research. In Workshop on Data Mining Applications in Sustainability (SIGKDD), San Diego, CA, 2011.
4) M. Zeifman et al. Nonintrusive Appliance Load Monitoring: Review and Outlook, IEEE Trans. Cons. Elec. Vol. 57, no. 1, 2011
5) M. Johnson and Alan Willsky, Bayesian Nonparametric Hidden Semi-Markov Models, *preprinted*