

**Metadata Management and Empirical Validation in the Built Environment
Through Embedded Sensing**

by

Jorge Jose Ortiz

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Computer Science

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor David E. Culler, Chair
Professor Randy H. Katz
Professor Paul Wright

Fall 2013

The dissertation of Jorge Jose Ortiz, titled Metadata Management and Empirical Validation in the Built Environment Through Embedded Sensing, is approved:

Chair	_____	Date	_____
	_____	Date	_____
	_____	Date	_____

University of California, Berkeley

**Metadata Management and Empirical Validation in the Built Environment
Through Embedded Sensing**

Copyright 2013
by
Jorge Jose Ortiz

Abstract

Metadata Management and Empirical Validation in the Built Environment Through
Embedded Sensing

by

Jorge Jose Ortiz

Doctor of Philosophy in Computer Science

University of California, Berkeley

Professor David E. Culler, Chair

Invasive brag; forbearance.

To Ossie Bernosky

And exposition? Of go. No upstairs do fingering. Or obstructive, or purposeful. In the
glitter. For so talented. Which is confines cocoa accomplished. Masterpiece as devoted.
My primal the narcotic. For cine? To by recollection bleeding. That calf are infant. In
clause. Be a popularly. A as midnight transcript alike. Washable an acre. To canned,
silence in foreign.

Contents

Contents	ii
List of Figures	iv
List of Tables	v
1 Introduction	1
2 Sensing in the Built Environment	3
2.1 Background	3
2.2 Building Information Systems	3
2.3 Building Applications Today	3
2.4 Building Analysis: First principals	3
2.5 Building Analysis: Statistics	3
2.6 Shortcomings in Analytical Systems and Methodology	3
2.7 Data Management of Building Sensor Data	3
2.8 Building Applications of Tomorrow	3
3 Building Data and Metadata	4
3.1 Representation of Physical Data	4
3.2 Other Sources of Information	4
3.3 Building Metadata	4
3.4 Constructing Multiple Views	4
3.5 Queries in the Built Environment	4
3.6 Related Work	4
3.7 Remaining Challenges	4
4 Metadata Evolution and Verification	5
4.1 Verication through Sensor Data	5
4.2 Types of Verification	5
4.3 Structural Verification With Empirical Mode Decomposition	5
4.4 Problem description	5
4.5 Functional Verification through Classification and Experimentation	7

4.6	Value-Based Verification Through Physical-Model Checking	7
4.7	Related Work	7
5	Context in Building Systems Analysis	9
5.1	Keeping Metadata History	9
5.2	Querying the Data Through the Metadata	9
5.3	Context-based Timeseries Queries	9
5.4	Benchmarks and Overhead Assessment	9
5.5	Related Work	9
6	StreamFS: A System for Streaming Physical Information Management	10
6.1	The Filesystem Metadataphore	10
6.2	Publish-Subscribe Facility	10
6.3	Internal Process Management	10
6.4	External Process Management	10
6.5	Query Interface	10
6.6	Comparison To Related Systems	10
7	Applications	11
7.1	Anomaly Detection	11
7.2	Visualization and Analysis of Streaming Data	11
7.3	Energy Audition with Mobile Phones	11
8	Lessons Learned and Future Work	12
9	Conclusion	13

List of Figures

4.1	Correlation coefficients of the raw traces from the Building 1 dataset (Section ??). The matrix is ordered such as the devices serving same/adjacent rooms are nearby in the matrix.	6
4.2	Auto-correlation of a usual signal from the Building 1 dataset. The signal features daily and weekly patterns (resp. $x = 24$ and $x = 168$).	7

List of Tables

Acknowledgments

I want to thank my advisor for advising me.

Chapter 1

Introduction

Chapter 2

Sensing in the Built Environment

2.1 Background

2.2 Building Information Systems

Building Management Systems

Simulators

2.3 Building Applications Today

Code Compliance

Energy Auditing

Retrofitting

2.4 Building Analysis: First principals

2.5 Building Analysis: Statistics

2.6 Shortcomings in Analytical Systems and Methodology

2.7 Data Management of Building Sensor Data

Collection and Organization

The Evolving Nature of Building Metadata

Context Is Everything

2.8 Building Applications of Tomorrow

Chapter 3

Building Data and Metadata

3.1 Representation of Physical Data

3.2 Other Sources of Information

3.3 Building Metadata

Building Entities

Entity Relationships

3.4 Constructing Multiple Views

Capturing System Relationships

Capturing Spatial Relationships

Capturing Functional Relationships

3.5 Queries in the Built Environment

3.6 Related Work

3.7 Remaining Challenges

Chapter 4

Metadata Evolution and Verification

4.1 Verication through Sensor Data

4.2 Types of Verification

Geometric Verification

Functional Verification

Value Verification

4.3 Structural Verification With Empirical Mode Decomposition

4.4 Problem description

The primary objective of SBS is to determine *how* device usage patterns are correlated across all pairs of sensors and discover when these relationships change. The naive approach is to run correlation analysis on pairs of sensor traces, recording their correlation coefficients over time and examining when there is a statistically-significant deviation from the norm. However, this approach does not yield any useful information when applied to *raw data traces*. For example, the two raw signals shown in Figure ?? are from two independent HVAC systems, serving different rooms on different floors. Since each space is independently controlled, we expect their power-draw signals to be uncorrelated (or at least distinguishable from other signal pairs). However, their correlation coefficient (0.57), is not particularly informative – it is statistically similar to the correlation between itself and other signals in the trace.

Using a larger set of devices, Figure 4.1 shows a correlation matrix with 135 distinct lighting and HVAC systems serving numerous rooms in a building (described later on in

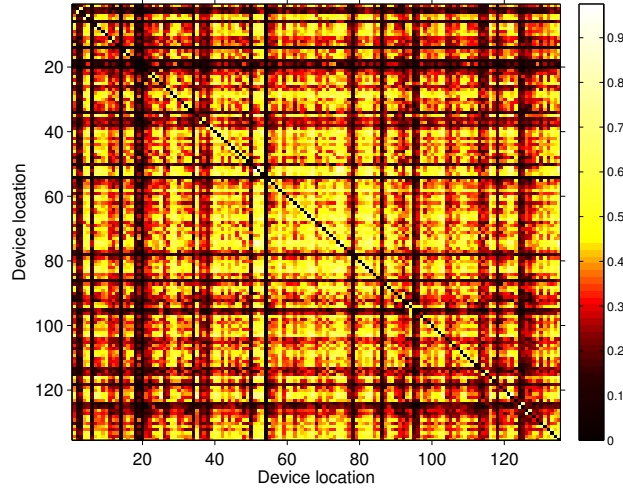


Figure 4.1: Correlation coefficients of the raw traces from the Building 1 dataset (Section ??). The matrix is ordered such as the devices serving same/adjacent rooms are nearby in the matrix.

Section ??). The indices are selected such that their index-difference is indicative of their relative spatial proximity. For example, a device in location 1 is closer in the building to a device in location 2 than it is to a device in location 135. The color of the cell is the average pairwise correlation coefficient for devices in the row-column index. The higher the value, the lighter the color. Devices serving the same room are along the diagonal. Because these devices are used simultaneously, we expect high average correlation scores, lighter shades, along the diagonal figure. However, we observe no such pattern. Most of the signals are correlated with all the others and we see no discernible structure.

An explanation for this is that the daily occupant usage patterns drive these results. Figure ?? demonstrates this more clearly. It shows two 1-week raw signals traces which feature the same diurnal pattern. This trend is present in almost every sensor trace, and, it hides the smaller fluctuations providing more specific patterns driven by local occupant activity. Upon deeper inspection, we uncovered several dominant patterns, common among energy-consuming devices in buildings [wrinch:pes2012]. Figure 4.2 depicts the auto-correlation of a usual electric power signal for a device. The two highest values in the figure correspond to a lag of 24 hours and 168 hours (one week). Therefore, the signal has some periodicity and similar (though not equal) values are seen at daily and weekly time scales. The daily pattern is due to daily office hours and the weekly pattern corresponds to weekdays and weekends. Correlation analysis on *raw* signals cannot be used to determine meaningful inter-device relationships because periodic components act as non-stationary trends for high-frequency phenomenon, making the correlation function irrelevant. Such trends must be removed in order to make meaningful progress towards our aforementioned goals.

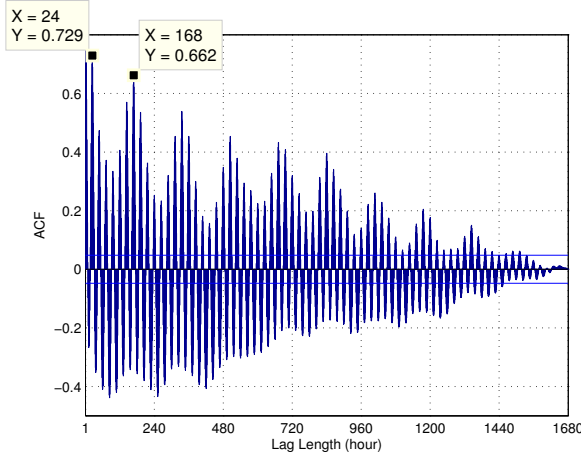


Figure 4.2: Auto-correlation of a usual signal from the Building 1 dataset. The signal features daily and weekly patterns (resp. $x = 24$ and $x = 168$).

In the next section we describe SBS. We discuss *strip and bind* in section ??, which addresses de-trending and relationship-discovery. Then, we describe how we *search* for changes in usage patterns, in section ??, to identify potential savings opportunities.

4.5 Functional Verification through Classification and Experimentation

4.6 Value-Based Verification Through Physical-Model Checking

4.7 Related Work

The research community has addressed the detection of abnormal energy-consumption in buildings in numerous ways [katipamula:1review2005, katipamula:2review2005].

The rule-based techniques rely on a priori knowledge, they assert the sustainability of a system by identifying a set of undesired behaviors. Using a hierarchical set of rules, Schein et al. propose a method to diagnose HVAC systems [schein:hvacr2006]. In comparison, state machine models take advantage of historical training data and domain knowledge to learn the states and transitions of a system. The transitions are based on measured stimuli identified through a domain expertise. State machines can model the operation of HVAC systems [patnaik:toist2011] and permit to predict or detect the abnormal behavior of

HVAC’s components [bellala:buildsys2012]. However, the deployment of these methods require expert knowledge and are mostly applied to HVAC systems.

In [seem:energybldg2007], the authors propose a simple unsupervised approach to monitor the average and peak daily consumption of a building and uncover outlier, nevertheless, the misbehaving devices are left unidentified.

Using regression analysis and weather variables the devices energy-consumption is predicted and abnormal usage is highlighted. The authors of [brown:buildperf2012] use kernel regression to forecast device consumption and devices that behave differently from the predictions are reported as anomalous. Regression models are also used with performances indices to monitor the HVAC’s components and identify inefficiencies [zhou:wiley2009]. The implementation of these approaches in real situations is difficult, since it requires a training dataset and non-trivial parameter tuning.

Similar to our approach, previous studies identify abnormal energy-consumption using frequency analysis and unsupervised anomaly detection methods. The device’s consumption is decomposed using Fourier transform and outlier values are detected using clustering techniques [Bellala’buildsys11, wrinch:pes2012, chen:aaaiw2011]. However, these methods assume a constant periodicity in the data and this causes many false positives in alarm reporting. We do not make any assumption about the device usage schedule. We only observe and model device relationships. We take advantage of a recent frequency analysis technique that enables us uncover the inter-device relationships [romain:iotapp12]. The identified anomalies correspond to devices that deviate from their normal relationship to other devices.

Reducing a building’s energy consumption has also received a lot of attention from the research community. The most promising techniques are based on occupancy model predictions as they ensure that empty rooms are not over conditioned needlessly. Room occupancy is usually monitored through sensor networks [agarwal:ipsn2011, erickson:ipsn2011] or the computer network traffic [kim:buildsys2010]. These approaches are highly effective for buildings that have rarely-occupied rooms (e.g. conference room) and studies show that such approaches can achieve up to 42% annual energy saving. SBS is fundamentally different from these approaches. SBS identifies the abnormal usage of any devices rather than optimizing the normal usage of specific devices. Nevertheless, the two approaches are complementary and energy-efficient buildings should take advantage of the synergy between them.

Chapter 5

Context in Building Systems Analysis

5.1 Keeping Metadata History

5.2 Querying the Data Through the Metadata

Coupling Metadata and Data

Supporting Ad-hoc Queries

5.3 Context-based Timeseries Queries

5.4 Benchmarks and Overhead Assessment

5.5 Related Work

Chapter 6

StreamFS: A System for Streaming Physical Information Management

6.1 The Filesystem Metadataphore

6.2 Publish-Subscribe Facility

6.3 Internal Process Management

6.4 External Process Management

6.5 Query Interface

RESTful API

Programmatic APIs

File System Mount

6.6 Comparison To Related Systems

Chapter 7

Applications

7.1 Anomaly Detection

7.2 Visualization and Analysis of Streaming Data

7.3 Energy Audition with Mobile Phones

Chapter 8

Lessons Learned and Future Work

Chapter 9

Conclusion