Anomaly Detection Using Projective Markov Models in a Distributed Sensor Network

Sean Meyn
Electrical and Comp. Engg. and the Coordinated Sciences Laboratory
University of Illinois at Urbana-Champaign

Amit Surana, Yiqing Lin, and Satish Narayanan
United Technologies Research Center
411 Silver Lane, East Hartford, CT.

***Abstract*— The paper develops application of techniques from robust and universal hypothesis testing for anomaly detection and change-point detection in dynamic, interconnected systems. This theory is extended using the concept of *projected Markov models* originally proposed by Claude Shannon. Also presented is a detailed application to anomaly detection from people movements patterns in buildings.**

I. Introduction

This paper deals with detection of anomalies in complex systems. An interest of the authors is applications to problems concerning building security monitoring in the presence of people traffic.

A. Distributed, Universal, and Robust Detection

What do the words in this section title mean? The first adjective is most familiar to a controls audience: In building applications there are sensors distributed to monitor heat, light, air quality, and video that can track the movement of occupants. A distributed detection algorithm will construct local as well as global alarms to improve robustness, and these solutions have low complexity. Robustness, of course, means that the algorithms will be effective even in the face of large modeling error. *Universality* is less well-known in the controls community. This means that the anomaly detection algorithm does not require a model of the anomalous behavior.

The statistical theory of robust detection and universal detection contains beautiful mathematics and practical algorithms. One goal of the paper is to reveal some of this beauty, and also explain some of the pitfalls in direct application of universal detection methods.

B. Modeling People Movement

Modeling and predicting occupancy traffic can be useful in a variety of building applications, including: building security monitoring of unusual traffic patterns, information awareness to first responders in case of fire [22], and control of ventilation for energy efficient building operation [13]. Buildings today have unprecedented access to data [3]. The resulting large and heterogeneous data streams contain valuable information that is largely untapped.

Behavioral modeling is valuable in the applications to security to help identify unexpected patterns of behavior. In this paper we are not considering the detailed movements of an individual, as in the work of [1], [29], [9], although we do use concepts from this work in the construction of a simulator for use in the experiments reported here.

In recent experimental studies we have considered the aggregate behavior of people in the building: The rate of flow, and distribution over the building. On-line information combined with historical data is the basis of the detection algorithms applied at United Technologies Research Center (UTRC) [15], [14] and described in this paper.

The techniques introduced in this paper are most closely related to those of [28]. A “tracklet model” is fit to data from movement and email exchanges collected over approximately one year in a building. The authors propose the KL metric for an empirical Markov model, and show that this can provide an indicator of disruptions, such as holidays or a change in staff.

The models used in this work are similar to those found in the operations research literature to model queues. Smith and Towsley in [24] introduced a queueing model for estimating evacuation times in buildings. These techniques have been developed recently in, for example, [26], [27], [7], [18].

In this paper we extend the model of Smith and Towsley for applications to anomaly detection during normal building operation. Experiments conducted at UTRC show that occupancy flow and external arrivals to the building can sometimes be captured accurately using simple probability distributions. The resulting semi-empirical Markov model is ideally suited for the proposed detection methods.

C. Projective Markov Models

In fact, a Markov model for the entire building is too complex for direct application of these methods. We propose instead the construction of local models for (possibly overlapping) regions of the building. Moreover, various models are constructed in each region. During a training stage a model of normal operation in the building is created based on historical sensor data. Similar models are built in real-time, based on real-time data from the sensor network, for use in the detector.

Based on this we obtain a family of detectors that can be fused at a central location for centralized decision making, or used for local detection. This approach admits an appealing information-theoretic, geometric interpretation.

The semi-empirical model is compared to the use of empirical models based on binning. As predicted by theory [5], [10], we find that the detector based on a semi-empirical model exhibits lower variance, and thus exhibits significantly lower false alarm rates.

D. Organization

This paper is organized in five sections. In Sec. II we describe robust and universal hypothesis testing frameworks for anomaly detection and change-point detection. Sec. III deals with the empirical and semi-empirical Markov models of occupant movement in a building environment based on a projective approach due originally to Shannon, Mori and Zwanzig. Contained in Sec. IV-A is a description of the simulation framework in which occupant activity in an office building is used to generate test cases for the anomaly detection problem. The performance of detectors based on empirical and semi-empirical models is compared on these test cases which include anomalous occupancy patterns such as crowd formation, excess of flow into a region, or periods of no activity. Finally, we conclude in Sec. V with recommendations for future work.

II. Anomaly Detection

In this section we describe generally the approaches to anomaly detection applied in this project. The foundation of hypothesis testing, robust, universal, or otherwise, is entropy. However, the meaning of entropy depends upon modeling assumptions and goals.

Regardless of the model, the notation is fixed as follows. A stochastic process denoted $\boldsymbol{Z}$ is observed, evolving on a state space Z. In this section it is assumed that Z is a finite set. Based on a finite set of observations $Z_1^T = (Z(1), \dots, Z(T))$, our goal is to determine if these observations indicate anomalous behavior. A test can be described as a region $\mathcal{R} \subset \mathsf{Z}^T$ such that an alarm is sounded if $Z_1^T \notin \mathcal{R}$. We denote the test by $\phi = \mathbb{I}_{\mathcal{R}}$ so that the alarm is sounded if and only if $\phi(Z_1^T) = 1$.

The performance of a detector is framed in the Neyman-Pearson setting. For a given test ϕ we let $p_{\text{MD}}(\phi)$ denote the probability of a missed detection (the probability that $\phi = 0$, conditional on the event that an anomaly has occurred), and we let $p_{\text{FA}}(\phi)$ denote the probability of a false alarm (the probability that $\phi = 1$, conditional on the event that an anomaly *did not* occur). When $\boldsymbol{Z}$ is modeled as a stationary process then it is convenient to consider the error exponents, defined by

$$
\begin{aligned}
\eta_{\text{MD}}(\phi) &= -\lim_{T\to\infty} \frac{1}{T} \log(p_{\text{MD}}(\phi)) \\
\eta_{\text{FA}}(\phi) &= -\lim_{T\to\infty} \frac{1}{T} \log(p_{\text{FA}}(\phi))
\end{aligned}
\tag{1}
$$

The construction of an effective region is considered first in the i.i.d. setting.

A. Approaches for I.I.D. Model

Suppose that $\boldsymbol{Z}$ is an i.i.d. sequence whose marginal distribution π^0 is known. The most famous test is the log-likelihood ratio (LLR) test, which requires that there be a model for anomalous behavior. In this case we assume that $\boldsymbol{Z}$ has marginal π^1 in the event of anomalous behavior. Letting $L = \log(d\pi^1/d\pi^0)$ denote the LLR, the decision region is given by $\mathcal{R} = \{Z_1^T \in \mathsf{Z}^T : T^{-1}\sum_{t=1}^T L(Z(t)) \geq c\}$, for some threshold c. That is, the test becomes,

$$
\phi(Z_1^T) = \mathbb{I}\Big\{\frac{1}{T}\sum_{t=1}^{T} L(Z(t)) \geq c\Big\} \tag{2}
$$

In the absence of a model for anomalous behavior there remains a useful detector. In the i.i.d. setting we restrict to tests defined by a region in the space of probability distributions $\mathcal{P}(\mathsf{Z})$, on Z, based on the empirical distributions. Assume that Z is a finite set, and denote the empirical distribution by

$$
\Gamma_T(z) =: \frac{1}{T}\sum_{t=1}^{T} \mathbb{I}\{Z(t) = z\}, \quad z \in \mathsf{Z}.
$$

For any set $\mathcal{K} \subset \mathcal{P}(\mathsf{Z})$, with $\pi^0 \in \mathcal{K}$, we obtain a test via

$$
\phi(Z_1^T) = \mathbb{I}\{\Gamma_T \notin \mathcal{K}\} \tag{3}
$$

This can also be expressed $\phi = \mathbb{I}_{\mathcal{R}}$ for some set $\mathcal{R}$, but the lifting to the space of probability distributions is extremely valuable. One value is a reinterpretation of the LLR test. On letting $\mathcal{K}$ denote the half-space $\mathcal{K} = \{\mu \in \mathcal{P}(\mathsf{Z}) : \mu(L) < c\}$ we see that the LLR test is precisely (3).

A test that is asymptotically optimal (for large T) with respect to Bayesian or Neyman-Pearson error criteria is obtained when $\mathcal{K}$ is defined as a relative entropy neighborhood [6], [20]. Define the Kullback-Leiber (KL) distance (or relative entropy) by,

$$
D(\mu\|\nu) = \sum_{z\in\mathsf{Z}} \mu(z) \log\frac{\mu(z)}{\nu(z)}, \quad \mu, \nu \in \mathcal{P}(\mathsf{Z}) \tag{4}
$$

and define for any $\eta > 0$ the “divergence neighborhood”,

$$
\mathcal{Q}_\eta(\nu) = \{\mu : D(\mu\|\nu) < \eta\}. \tag{5}
$$

The universal test is obtained using $\mathcal{K} = \mathcal{Q}_\eta(\nu)$ for some value of η. This test is optimal in an asymptotic sense: As T tends to infinity, it achieves the maximal error exponent $\eta_{\text{MD}}(\phi)$ over all tests that satisfy the false-alarm constraint $\eta_{\text{FA}}(\phi) \geq \eta$ (see Hoeffding [8] for initial results, and extensions in [31], [30], [11], [19], [10]).

B. Markov Models

The emergence of the empirical distributions on Z can be explained as follows. Suppose that $\gamma^{(T)}$ and $\pi^{(T)}$ are two distributions on Z^T that are i.i.d., with marginals γ, π, respectively. Then, the KL distance only depends on the marginals, $T^{-1}D(\gamma^{(T)}\|\ \pi^{(T)}) = D(\gamma\|\ \pi)$. If $\gamma^{(T)}$ and $\pi^{(T)}$ describe the distribution of a stationary Markov processes, then this KL distance no longer depends only on the marginals alone. Assuming that the divergence is finite, it is not difficult to show that the divergence is a function of the *bivariate distributions* $\pi^{(2)}$, $\gamma^{(2)}$ (the distribution of $(Z(t), Z(t+1))$ under $\pi^{(T)}$ or $\gamma^{(T)}$, respectively.) Moreover, as $T \to \infty$, the normalized divergence converges to the *rate function* of Donsker and Varadhan, $T^{-1}D(\gamma^{(T)}\|\ \pi^{(T)}) \to J(\gamma^{(2)}\|\pi^{(2)})$, where

$$
J(\gamma^{(2)}\|\pi^{(2)}) := D(\gamma^{(2)}\|\pi^{(2)}) - D(\gamma\|\pi) \tag{6}
$$

We can then define a universal test based on the bivariate empirical distributions:

$$\phi(Z_1^T) = \mathbb{I}\{\Gamma_T^{(2)} \notin \mathcal{K}\} \tag{7}$$

where $\mathcal{K} = \{\gamma^{(2)} : J(\gamma^{(2)} \| \pi^{(2)}) < \eta\}$. For an irreducible finite state space Markov model the arguments in [31], [30] can be extended to show that this test is again optimal for the asymptotic Neyman-Pearson hypothesis criterion: It achieves the maximal error exponent $\eta_{\text{MD}}(\phi)$ over all tests that satisfy $\eta_{\text{FA}}(\phi) \geq \eta$.

C. Free Lunch?

The universal test is optimal in the asymptotic Neyman-Pearson hypothesis testing problem. In particular, in the i.i.d. model the universal test is as effective as the LLR test in which π^1 is known. These tests are also highly robust — this is again most easily seen in an asymptotic setting. Suppose that $\boldsymbol{Z}$ is stationary, but not i.i.d., yet the test is based on an i.i.d. model using the marginal of $\boldsymbol{Z}$. Provided π lies in the interior of $\mathcal{K}$, and assuming Sanov's Theorem holds for $\boldsymbol{Z}$ with a positive exponent, then the false alarm probability will vanish exponentially fast (i.e., $\eta_{\text{FA}}(\phi) > 0$).

The distribution π or $\pi^{(2)}$ can be obtained from historical data, and this training phase does not require that the true observations satisfy a Markov or i.i.d. assumption.

However, when π^1 is not known, then there is a cost for this missing prior information.

The following limit is established in [10], by way of results in [5, Section III.C]. Suppose that $\boldsymbol{Z}$ is indeed i.i.d. with marginal π^0. We then have

$$\lim_{T\to\infty} T\mathsf{E}[D(\Gamma_T\|\pi^0)] = \tfrac{1}{2}(|\mathsf{Z}| - 1).$$

This is good news, because the error measured by $D(\Gamma_T\|\pi^0)$ vanishes at rate $1/T$, rather than $1/\sqrt{T}$ as seen in the Central Limit Theorem. The bad news is that the rate of decay is very slow if the cardinality $|\mathsf{Z}|$ is very large. A formula for the variance is also obtained in [10]:

$$\lim_{n\to\infty} T^2\mathsf{E}[\big(D(\Gamma_T\|\pi^0) - \mathsf{E}[D(\Gamma_T\|\pi^0)]\big)^2] = \tfrac{1}{2}(|\mathsf{Z}| - 1).$$

The universal test based on ordinary divergence will be unreliable when the time horizon T is less than the square root of the alphabet size $|\mathsf{Z}|$. In [10] this is addressed by introducing a relaxation of the divergence (interpreted as a *Statistical Support Vector Machine*.) Here we apply multiple models and assume some prior information regarding anomalous behavior.

D. Multiple Models and Partial Information

1) Partial and Distributed Information: Suppose that we observe only a few function of the process $\boldsymbol{Z}$. For the purposes of discussion we restrict to the special case in which $\boldsymbol{Z}$ is a two-dimensional process, $Z(t) = (Z_1(t), Z_2(t))$, taking values in $\mathsf{Z} \times \mathsf{Z}$. We might have access to only one component of $\boldsymbol{Z}$, which is the case of partial information. Or, we might form two tests based on the individual observations $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$. In this case we are building a test from multiple models. The motivation for this is in applications to distributed detection, to reduce computational complexity, or to reduce variance.

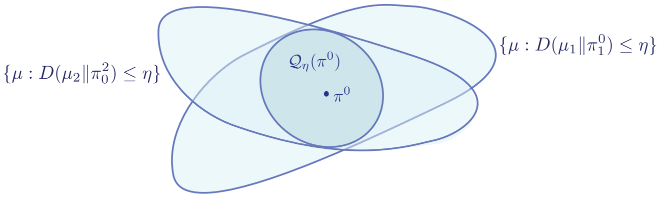
$\{\mu : D(\mu_2\|\pi_0^2) \leq \eta\}$
$\mathcal{Q}_\eta(\pi^0)$
π^0
$\{\mu : D(\mu_1\|\pi_1^0) \leq \eta\}$

Fig. 1. Given several local models, the intersection of the associated rejection regions provides a better approximation to the universal detection region $\mathcal{Q}_\eta(\pi^0)$.

In i.i.d. or Markov models we can investigate the performance tradeoffs faced using multiple models through simple information-geometric arguments.

For simplicity we restrict to the i.i.d. model in this subsection. Let π_i^0 denote the distribution of $Z_i(t)$, $i = 0, 1$. For a given value of $\eta > 0$ we have the inclusion

$$\mathcal{Q}_\eta(\pi^0) \subset \{\mu_i : D(\mu_i\|\pi_i^0) \leq \eta\}$$

This observation is illustrated in Fig. 1. Suppose that the results from two individual universal tests are given, with potentially conflicting information. Consider the test that sounds an alarm if either of the two tests is positive. This is equivalent to taking the intersection of the two decision regions. Fig. 1 shows that this intersection provides a better approximation to the universal detection region $\mathcal{Q}_\eta(\pi^0)$ than either individual test.

The performance of the resulting test depends in part on the nature of the anomaly.

2) Multiple Models of Anomalous Behavior: Suppose now that we have not one but several models of potential anomalous behavior. For purposes of discussion we again restrict to the i.i.d. model, and we assume that just two models of anomalous behavior are given, denoted $\pi^{1\,1}, \pi^{1\,2}$. Denote the corresponding LLRs by $L_i = \log(d\pi^{1\,i}/d\pi^0)$, $i = 1, 2$.

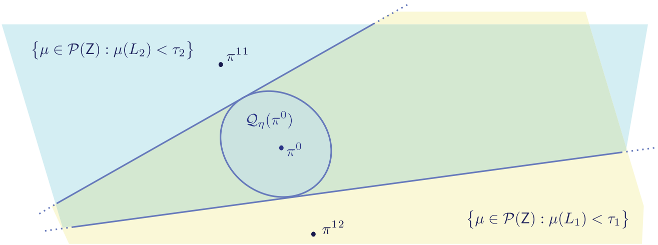
$\{\mu \in \mathcal{P}(\mathsf{Z}) : \mu(L_2) < \tau_2\}$
π^{11}
$\mathcal{Q}_\eta(\pi^0)$
π^0
π^{12}
$\{\mu \in \mathcal{P}(\mathsf{Z}) : \mu(L_1) < \tau_1\}$

Fig. 2. Given two models of anomalous behavior, an alarm is triggered if either of the corresponding LLR tests are positive. The intersection of the two corresponding half spaces in $\mathcal{P}(\mathsf{Z})$ contains the universal detection region $\mathcal{Q}_\eta(\pi^0)$.

Each of the LLR test (2) can be computed, and the resulting outcomes can be fused by considering again the information-geometric view. For each i, the corresponding test (2) can be expressed as (3) with $\mathcal{K}$ equal to a half-space $\mathcal{K}_i = \{\mu \in \mathcal{P}(\mathsf{Z}) : \mu(L_i) < c_i\}$. Let ϕ_u denote the universal test using $\mathcal{K} = \mathcal{Q}_\eta(\pi^0)$, and ϕ_ℓ denote the test using $\mathcal{K} = \cap_{i=1,2}\mathcal{K}_i$. The geometry illustrated in Fig. 2 combined with the results surveyed in Sec. II-C can be summarized in the following:

Proposition 1: Consider the i.i.d. model in which anomalous behavior may be i.i.d. with one of two priors. Then,

(i) $\phi_\ell(Z_1^T) \le \phi_u(Z_1^T)$ for any observation sequence, so that the probability of false alarm is greater using the universal test, and the probability of missed detection is reduced.

(ii) With $\{c_i\}$ chosen consistently with Fig. 2, so that $c_i = \eta$ for each i, the error exponents are identical using ϕ_ℓ or ϕ_u. In particular, $\eta_{\text{FA}} = \eta$ in either case.

(iii) For finite samples, the variance and bias using ϕ_ℓ or ϕ_u may be very different. □

E. Change Detection

Suppose that $\boldsymbol{Z}$ is a stochastic process, along with an unknown time τ, such that the process is generated by one law for $t < \tau$, and another for $t \ge \tau$. Again for simplicity we restrict to the (conditionally) i.i.d. setting in which $\boldsymbol{Z}$ is generated by two i.i.d. processes $\boldsymbol{Z}^1$ and $\boldsymbol{Z}^0$ with

$$Z(t) = \begin{cases} Z^0(t)\,, & t < \tau \\ Z^1(t)\,, & t \ge \tau \end{cases} \tag{8}$$

Let $\mathcal{T}$ denote a test: $\mathcal{T} = T$ means that the test declares $\tau \le T$. One formulation of optimality is to minimize the mean of $\mathcal{T}$, subject to a constraint on the probability of false alarm p_{FA} at time $\mathcal{T}$.

A useful test can again be constructed based on a likelihood ratio test, or a more complex decision region. The test most commonly considered is called CUSUM. In the *window-limited* form of the algorithm, a threshold $c \in \mathbb{R}$ and maximal time-horizon $\delta_0 \ge 1$ are given, and a test is defined by

$$\begin{aligned} \mathcal{T} &:= \min\{T \ge 1 : R(T) \ge c\} \\ R(T) &:= \max_{T-\delta_0 < T_0 \le T} \sum_{t=T_0}^{T} L(Z(t)) \end{aligned} \tag{9}$$

Lai has shown that this test is approximately optimal for appropriate choice of c and δ_0. The setting is asymptotic, as the change-point time tends to infinity, and the values of c and δ_0 can be predicted from the solution to the binary hypothesis testing problem described in Sec. II-A. The surveys [12], [25] give references, and an account of the field up to 2004.

We used (9) in experiments conducted at UTRC, but found that a fixed horizon was equally effective in this application. We define the moving-horizon test statistic,

$$R(T) := \sum_{t=T-\delta_0+1}^{T} L(Z(t)) \tag{10}$$

For fixed δ_0, performance is assessed using the standard binary hypothesis testing setting: An ROC curve of p_{FA} vs. p_{MD} is estimated through simulation or bounds.

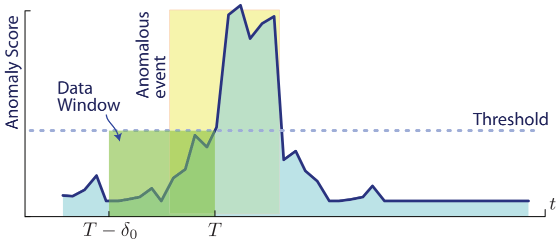

Fig. 3. Time T at which the test statistic $R(T)$ reaches a threshold.

Fig. 3 shows typical behavior of anomaly score $R(t)$ as a function of time. Also shown is a period of time during which anomalous behavior was occurring, and the time window used in the test statistic (10).

III. Projective Modeling of People in Buildings

A Markov model for the entire building is too complex for direct application of the anomaly detection methods described in the previous section. We propose instead the construction of local models for zones of the building, with possible spatial dependencies among neighboring zones. The choice of zones depends on the number, location, type and accuracy of sensors in the building, and the spatial scale (coarse or fine) at which we want to capture the occupancy evolution.

This approach has a long history. Shannon proposed the use of low dimensional Markov models to describe complex systems in his seminal 1948 paper [23]. It is argued in [16] that this technique can be regarded as a component of the approach to reduced-order modeling pioneered by Mori [17] and Zwanzig [32]. See [16] for history and extensions of this approach.

We adopt the following notation throughout the remainder of the paper. The building is divided into N zones that are organized as a graph, and we let W denote the $N \times N$ incidence matrix, with $W_{ii} = 1$. The set of neighbors of zone i is denoted by $N_i = \{j : W_{ij} = 1\}$. The flow from zone j to i at time t is denoted $S_{ji}(t)$, $j \in N_i$, and the external arrivals to zone i in time-slot t is denoted by $A_i(t)$ (provided there is an external entrance at this zone). We also reserve the index $i = N+1$ for the 'outside world', so that $S_{i\,N+1}(t)$ is the number of occupants in zone i who at time t transition from this zone out of the building. For each $i = 1, \cdots N$, and each $t \ge 0$, the occupancy in zone i is denoted by $Z_i(t)$, and we denote by $\bar{Z}_i$ an upper bound on the occupancy level.

These variables are related through the flow balance equations,

$$\begin{aligned} Z_i(T+1) = Z_i(T) &+ \sum_{j \in N_i} (S_{ji}(T+1) - S_{ij}(T+1)) \\ &+ A_i(T+1) - S_{i\,N+1}(T+1) \end{aligned} \tag{11}$$

Two approaches for modeling have been used in this research. The first is based on binning, leading to an *empirical Markov model*. The second semi-empirical approach is motivated by similar models used in queueing theory.

A. Empirical Markov Model

This approach begins with a finite partition of the observation space $\mathsf{Z} = \cup_{i=1}^{m_z} \mathsf{Z}_i$. The empirical distribution of the bivariate process is obtained using the sample path averages,

$$\hat{\pi}^{(2)}(k,\ell) = \frac{1}{T} \sum_{t=1}^{T-1} \mathbb{I}(Z(t) \in \mathsf{Z}_k) \mathbb{I}(Z(t+1) \in \mathsf{Z}_\ell) \qquad (12)$$

where $k, \ell \in \{1, \dots, m_z\}$ are the indices corresponding to the m_z bins. The marginal is given by $\hat{\pi}(k) = \sum_\ell \hat{\pi}^{(2)}(k,\ell)$. A Markov transition probability matrix is obtained using Bayes' rule, $\widehat{P}(k,\ell) := \hat{\pi}^{(2)}(k,\ell)/\hat{\pi}(k)$.

These models are easily constructed based on historical data. The empirical distribution is also the basis of a universal detector. Suppose that the model (π, P) has been learned, and new data is observed. We then evaluate $J(\hat{\pi}^{(2)} \| \pi^{(2)})$ using the definition (6), and compare this to a threshold to obtain a universal test, exactly as in the i.i.d. setting. However, as discussed in Sec. II-C, this approach is expected to be plagued by high false alarms unless T is much larger than m_z. This can be achieved by constructing several models at a range of spatial scales by dividing building into zones at different scales.

We can in addition incorporate additional structure on the model to reduce variance. An approach found to be highly successful in this application is described in the next section.

B. Semi-Empirical Markov Model

The semi-empirical modeling is an extension of the evacuation model of Smith and Towsley [24]. In this approach the random variables appearing in (11) are modeled by predefined distributions. It is assumed that the distribution of the net occupancy flow $Y_{ij}(t) = S_{ji}(t) - S_{ij}(t)$, conditioned on the current occupancy values, is a two-sided (asymmetric) geometric distribution. The distribution of each $A_i(t)$ is assumed to be exponential. All of these random variables are assumed to be independent, conditioned on the prior occupancy $Z(t-1)$. Under these conditions the occupancy process $\boldsymbol{Z}$ is a Markov chain, whose dynamics are computed from the flow balance equations (11). The semi-empirical model is highly flexible, and naturally captures spatial correlation between neighboring zones. Moreover, experiments conducted at UTRC show that these approximations are reasonable for an office building during normal operations.

Moreover, techniques of [5], [10] can be applied to show that variance in the semi-empirical approach will grow with the dimension of the parameterized family of models, and not with the size of the state space on which the model is built.

IV. Experimental Results and Analysis

In this section we illustrate the application of the anomaly detection framework of Sec. II in the building context of Sec. III. The anomalies considered are related to the spatiotemporal distribution of people, including phenomena such as crowd formation in regions where gathering is not typical, excess of flow into a region, or periods of no activity in areas where high traffic is typical. Activity of occupants on a single floor of an office building is used to generate test cases for the anomaly detection problems. See Fig. 4 for a map of the floor, along with the location of digital video cameras. People count data from these cameras is used to calibrate agent based models for simulating normal and anomalous occupancy patterns.

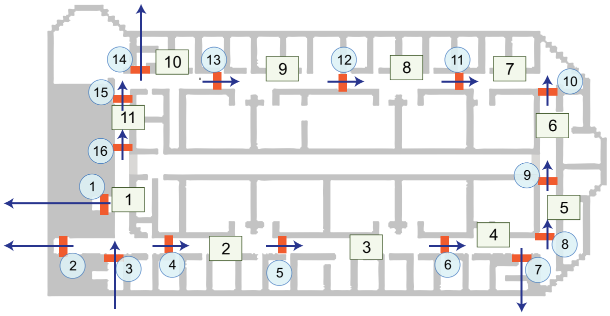

Fig. 4. Sensor layout for people traffic problem. The vertical and horizontal dark bars denote the location of the 14 video sensors, with the arrows indicating the direction of positive occupancy flow. Each lightly colored numbered block represents a zone whose boundary is formed by the surrounding sensors.

A. Simulation Environment

The agent-based simulation model we have employed is analogous to the cellular automata models [21], [2], [4]. The agent based model takes into account the building geometry and simulates each individual's movement. In order to simulate each individual's movement, agent decisions must be modeled on several levels, such as itinerary, path choice, and walking behavior [1], [29], [9]. Itinerary decision determines the time of arrival and entrance, the number of stops, and the location and dwell time of each stop, for each individual. Path choice is the decision on which path to follow, given the individual's destination and set of alternatives. Walking behavior is determined by factors such as average speed, average space taken by each individual, and conflict resolution rules at times when conflicts between agents arise.

Along with the behavior of occupants is a model of building topology: In our implementation, the building floor was spatially discretized into cells of size 2×2 ft. Walking speed was taken to be uniformly 4ft/sec, and each agent was assumed to follow a shortest path to his/her destination. Based on these assumptions, a model was fit to data obtained from a video sensor network in the physical building.

The construction of a model was based on a two step procedure. In the first step, an "optimal" flow of occupancy was computed in which all constraints imposed by the agent based model are satisfied; the objective was the L_1-norm of the difference between the given sensor data and the model output. This calculation can be performed using a linear program since the constraints in this model are linear. These constraints include flow conservation, prescribed initial and final occupancy, and non-negativity of occupancy and occupancy flow, for each zone. In the second step, a heuristic procedure was followed to generate agent itineraries consistent with the occupancy flow obtained in Step 1.

Applying this calibration procedure to the people count data from video cameras collected for 25 workdays, we generated 25 sets of one-hour simulated normal occupancy patterns, corresponding to the same one-hour time period. In the experiments described in next section, anomalous patterns of behavior were generated by modifying these normal patterns.

B. Experiment and Evaluation Architecture

The following test cases were considered, where the zones are identified in Fig. 4.

Scenario 1: ***Convergence*** People converge in one zone.
Case 1: People converge to zone 6 from zones 5 and 7. Normal occupancy level in zone 6: 0-9; Anomaly: 13
Case 2: People converge to zone 5 from zones 4 and 6 in 5 minutes. Anomaly: Converge at a higher than normal rate.

Scenario 2: ***Divergence*** People diverge from one zone.
Case 3: People disperse from zone 9, moving to zones 8 and 10. Anomaly: Reduction of 20 people at zone 9.

Scenario 3: ***Idleness*** No Activity for a long period of time.
Case 4: No activity in zone 7. Normal: maximum period of no activity 24 minutes; Anomaly: 38 minutes.

Scenario 4: ***Loitering*** Oscillating sensor readings.
Case 5: Unusual movement between zones 5 and 6. Normal: 4 consecutive oscillating readings; Anomaly: 9 minutes.

Scenario 5: ***Congestion*** Higher occupancy level in combined zones.
Case 6: Sum of occupancy level in zones 5 and 6. Normal: 19; Anomaly: 21.

These patterns were chosen for two reasons: (i) they represent traffic patterns that are very difficult to label a priori as anomalies; and (ii) they are typically associated with unlawful activities such as gang gathering, hold-up, and theft or arson.

In each zone, the occupancy Z_i was chosen as a coarse variable in the definition of the empirical and semi-empirical test statistics. The corresponding test statistics are expressed by, respectively,

$$R_i(t) = J(\Gamma^{(2)}_{i,\delta_0,t} \| \pi^{(2)}_i) \tag{13}$$

$$R_i(t) = \frac{1}{\delta_0} \sum_{k=t-\delta_0+1}^{t} \log\big(\ell_{i,t,\delta_0}\big) \tag{14}$$

In (13) the distribution $\pi^{(2)}_i$ is a local model for zone i based on historical data, and $\Gamma^{(2)}_{i,\delta_0,t}$ denotes the empirical distribution for zone i at time t, based on a widow of length δ_0. In (14) we have used ℓ to denote the likelihood ratio,

$$\ell_{i,t,\delta_0} := \frac{P_{i,t,\delta_0}(Z_i(k+1) \| Z_{N(i)}(k))}{P_i(Z_i(k+1) \| Z_{N(i)}(k)])}$$

where $N(i)$ denotes the neighbors of zone i, the transition law P_i is obtained based on training data, and P_{i,t,δ_0} is the semi-empirical model obtained at time t based on the most recent δ_0 observations.

The definitions of false alarm and missed detection require some care since the detection problems considered here do not fall squarely within the binary hypothesis testing, or change detection frameworks. An inspection of Fig. 3 shows that a test statistic admits many peaks and valleys, meaning that many alarms may be raised for a single anomaly. Similarly, one false alarm will likely be followed by several others.

In order to unambiguously determine false alarm and missed detection we adopt the following conventions. We fix a delay parameter $\delta > 0$ and at any time t_0 consider the test statistic on the interval $[t_0 - \delta, t_0]$. For each experiment, and a given detection algorithm with threshold $c > 0$, the following conventions were used to define false alarm and missed detection events for a given time t_0:

1 *Missed detection*: If there is an anomaly at time t_0, and $\max_{t_0 \le t \le t_0+\delta} R(t) < c$, then this is considered a missed detection.
2 *False Alarm*: If $R(t_0) > c$ and there is no anomaly in the time period $[t_0 - \delta, t_0]$ then this is an occurrence of false alarm.

In light of the conclusion in Sec. II-D.1, the maximum of the test statistics (over a collection of local models) was used in above definition, i.e. $R(t) = \max_i R_i(t)$. Based on these definitions, the probability of false alarm p_{FA} and missed detection p_{MD} were estimated using standard Monte-Carlo.

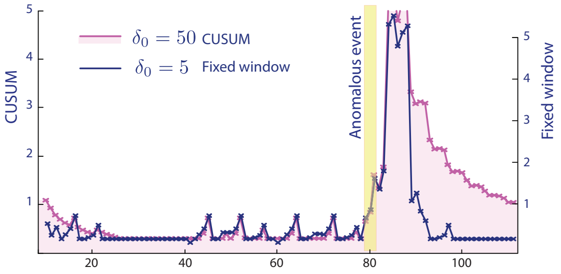

Fig. 5. The magenta curve (with associated shaded region) shows the test score using CUSUM (Eq. (9)) for $\delta_0 = 50$. Overlaid in dark blue is the score computed using (10) with $\delta_0 = 5$.

Fig. 5 illustrates why we found the test statistic (10) as valuable as the window-limited CUSUM (9). Similar experiments yielded similar results, and we regard this as an indicator of the robustness of CUSUM with respect to the time horizon δ_0.

C. Findings

In each scenario considered, the anomaly score $R_i(t)$ was computed for each zone $i = 1, 2 \cdots, 11$ using each of the test statistics: (13) for the empirical model, and (14) for the semi-empirical model.

We found that, in comparison to empirical models, semi empirical models

1 Show larger delay in detection, and
2 Exhibit reduced false alarm rate.

This conclusion is illustrated in Fig. 6, where ROC curves for the two approaches are illustrated for several values of the allowable delay δ. These curves were obtained using 10 different experiments with anomalies belonging to the five scenarios described above. It is clear from these plots that despite larger detection delay, semi-empirical models have a overall much better performance in detecting anomalies in people traffic than empirical models.

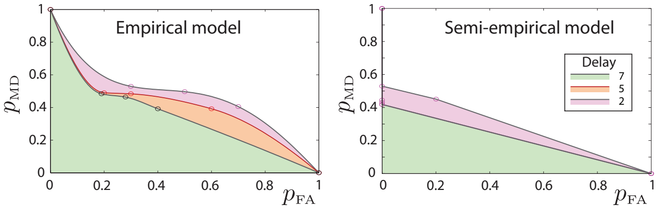

Fig. 6. ROC curve for people traffic detector based on empirical (left) and semi-empirical Markov model (right)

We found that the empirical and semi-empirical approaches are successful for detection in cases where the anomalies are related to convergence (see Fig. 7 and Fig. 9) or divergence (see Fig. 10) of occupants. Note that in these figures the test statistic is only shown for the relevant zones. As discussed in Sec. IV-B, the maximum of the test statistic ($R(t)$) over 11 zones is used for computation of the ROC curve: Fig. 8 shows one such plot of $R(t)$ for the first test case.

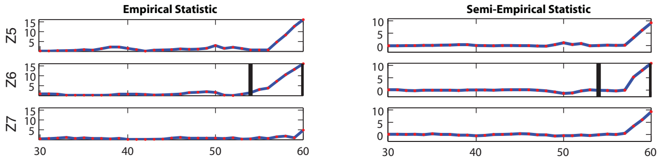

Fig. 7. Case 1: Anomaly scores using empirical and semi-empirical Markov model

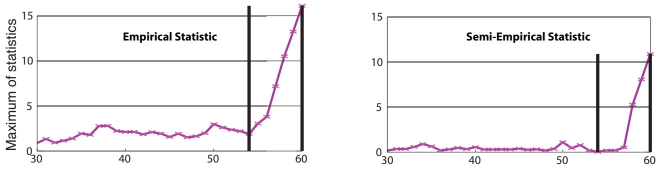

Fig. 8. The maximum of all eleven anomaly scores using empirical and semi-empirical Markov models in Case 1

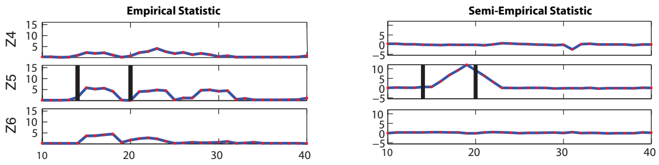

Fig. 9. Case 2: Anomaly scores using empirical and semi-empirical Markov model

These methods were not very effective in scenario 3 and 4, for which anomalies are associated with significant memory.

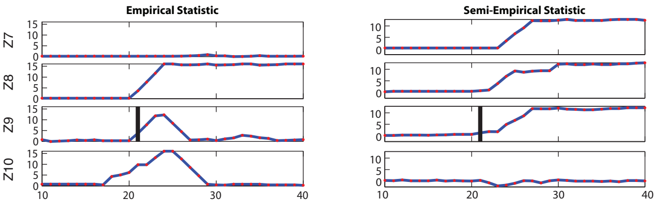

Fig. 10. Case 3: Anomaly scores using empirical and semi-empirical Markov model

To improve these algorithms the state process should be extended with an additional coarse variable such as the interval count.

The empirical approach failed to capture anomalies corresponding to scenario 5, which involves higher occupancy level in zones 5 and 6. Such anomalies are however detected using empirical models at a coarser scale, in which zones 5 and 6 are combined into a single zone (and also aggregation of the other nine zones into three: $\{1,2,3\}$, $\{4,7,8\}$, $\{9,10,11\}$). The results are summarized in Fig. 11.

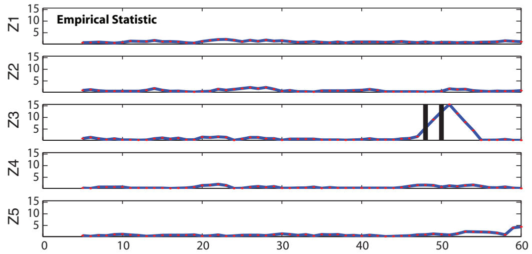

Fig. 11. Case 6: Anomaly score using empirical Markov model based on 5 zones.

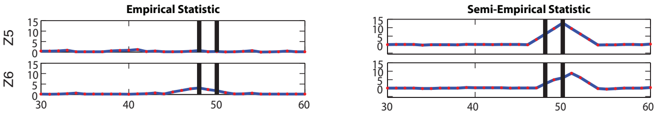

Fig. 12. Case 6: Anomaly scores using empirical and semi-empirical Markov model

The semi-empirical approach was far more effective for this scenario. As seen on the right in Fig. 12, this anomaly is flagged using the semi-empirical statistic even at fine spatial scales, since these models naturally incorporate spatial dependencies.

V. Conclusions and Future Research

In this paper we have demonstrated feasibility of an anomaly detection framework using projected Markov models and universal hypothesis testing. Semi-empirical Markov models were introduced in order to capture spatial correlations in traffic behavior in a computationally tractable manner. It was found that detection algorithms based on these models exhibit significantly lower false alarm rates and perform better than empirical models in detecting anomalies in all experiments.

In addition to validation of the anomaly detection framework in field data sets, several challenges remain to be addressed. In order to obtain more specific actionable information the detector should be able to distinguish between different types of anomalies. The knowledge of models of anomalous behavior can be used to refine anomaly detection, as described in Sec. II-D.2. However, given that anomalous behaviors do not occur frequently, learning models for such behaviors is a significant challenge. Another important extension is to automatically localize the perceived anomaly and probe for more information (e.g. query for additional data from the building control system, or deploy additional resources, such as guards). Isolating anomalous individual behavior (e.g. loitering), and coordinated group behavior among few individuals in a large population is another important challenge. These extensions are under progress and would be reported elsewhere.

ACKNOWLEDGMENTS

The funding provided by United Technologies Research Center for this work is greatly appreciated. Authors would like to thank George Mathew (UCSB) for his initial contributions to this work. The first author would also like to acknowledge a partial support by NSF under grant CCF 07-29031. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

REFERENCES

[1] M. Bierlaire, G. Antonini, and M. Webers. Behavioral dynamics for pedestrians. In *10th International Conference on Travel Behavior Research*, August 2003.

[2] V. J. Blue and J. L. Adler. Using cellular automata microsimulation to model pedestrian movements. In A. Ceder, editor, *Proceedings of the 14th International Symposium on Transportation and Traffic Theory*, pages 235–254, 1999.

[3] M. R. Brambley, D. Hansen, P. Haves, D. R. Holmberg, S. C. McDonald, K. W. Roth, and P. Torcellini. Advanced sensors and controls for building applications: Market assessment and potential R&D pathways. *Prepared for the DOE under Contract DE-AC05-76RL01830*, April 2005.

[4] C. Burstedde, A. Kirchner, K. Klauck, A. Schadschneider, and J. Zittartz. Cellular automaton approach to pedestrian dynamics — Application. In M. Schreckenberg and S. Sharma, editors, *Pedestrian and Evacuation Dynamics*, pages 87–98, 2002.

[5] Bertrand S. Clarke and Andrew R. Barron. Information-theoretic asymptotics of Bayes methods. *IEEE Trans. Inform. Theory*, 36(3):453–471, 1990.

[6] A. Dembo and O. Zeitouni. *Large Deviations Techniques And Applications*. Springer-Verlag, New York, second edition, 1998.

[7] Kun Deng, Wei Chen, Prashant G. Mehta, and Sean P. Meyn. Resource pooling for optimal evacuation of a large building. *Decision and Control, 2008. CDC 2008. 47th IEEE Conference on*, pages 5565–5570, Dec. 2008.

[8] W. Hoeffding. Asymptotically optimal tests for multinomial distributions. *Ann. Math. Statist.*, 36:369–408, 1965.

[9] S. P. Hoogendoorn, P. H. L. Bovy, and W. Daamen. Microscopic pedestrian wayfinding and dynamics modeling. In M. Schreckenberg and S. Sharma, editors, *Pedestrian and Evacuation Dynamics*, pages 123–155, 2002.

[10] D. Huang, J. Unnikrishnan, S. Meyn, V. Veeravalli, and A. Surana. Statistical SVMs for robust detection, supervised learning, and universal classification. In *Proceedings of the Information Theory Workshop on Networking and Information Theory, Volos, Greece.*, 2009.

[11] J. Huang, C. Pandit, S. P. Meyn, M. Médard, and V. Veeravalli. Entropy, inference, and channel coding. In Prathima Agrawal, Matthew Andrews, Philip J. Fleming, George Yin, and Lisa Zhang, editors, *Proceedings of the Summer Workshop on Wireless Networks*, volume 143 of *IMA Vol. Math. Appl.*, pages 99–124. Springer, New York, 2007.

[12] T.L. Lai. Sequential analysis: some classical problems and new challenges. *Stat. Sin.*, 11:303–408, 2001.

[13] K. Law, K. Dauber, and X. Pan. Energy impact of commercial building controls and performance diagnostics: Market characterization, energy impact of building faults and energy savings potential. Technical Report TIAX Reference D0180, TIAX LCC for DOE, November 2005.

[14] S. Meyn, A. Surana, Y. Lin, S. M. Oggianu, S. Narayanan, and T. A. Frewen. A Sensor-Utility-Network method for estimation of occupancy distribution in buildings. Accepted for inclusion in the 48th IEEE Conference on Decision and Control, December 16-18 2009.

[15] S. P. Meyn, Y. Lin, S. M. Oggianu, A. Surana, and I. Fedchenia. System and method for occupancy estimation and monitoring. *USPTO patent application under review*, 2008.

[16] S. P. Meyn and G. Mathew. Shannon meets Bellman: Feature based Markovian models for detection and optimization. In *Proc. 47th IEEE CDC*, pages 5558–5564, 2008.

[17] H. Mori. Transport, collective motion, and brownian motion. *Progress of Theoretical Physics*, 33:423–455, 1965.

[18] J.S. Niedbalski, Kun Deng, P.G. Mehta, and S. Meyn. Model reduction for reduced order estimation in traffic models. *American Control Conference, 2008*, pages 914–919, June 2008.

[19] C. Pandit and S. P. Meyn. Worst-case large-deviations with application to queueing and information theory. *Stoch. Proc. Applns.*, 116(5):724–756, May 2006.

[20] H. V. Poor. *An introduction to signal detection and estimation*. Springer Texts in Electrical Engineering. Springer-Verlag, New York, second edition, 1994. A Dowden & Culver Book.

[21] A. Schadschneider. Cellular automaton approach to pedestrian dynamics — theory. In M. Schreckenberg and S. Sharma, editors, *Pedestrian and Evacuation Dynamics*, pages 75–86, 2002.

[22] M. Schreckenberg and S.D. Sharma, editors. *Pedestrian and Evacuation Dynamics*. Springer, Berlin, 2002.

[23] C.E. Shannon. A mathematical theory of communication. *Bell System Tech. J.*, 27:379–423, 623–656, 1948.

[24] J. M. Smith and D. Towsley. The use of queuing networks in the evaluation of egress from buildings. *Environment and Planning B: Planning and Design*, 8(2):125–139, 1981.

[25] A. G. Tartakovsky and V. V. Veeravalli. General asymptotic Bayesian theory of quickest change detection. *SIAM Theory of Probability and its Applications*, 49(3):538–582, 2004.

[26] Robert Tomastik, Satish Narayanan, Andrzej Banaszuk, and Sean Meyn. Model-based Real-Time Estimation of Building Occupancy During Emergency Egress. 4th International Conference on Pedestrian and Evacuation Dynamics. `http://www.ped2008.com/`, February 27–29 2008.

[27] P. Wang, P. B. Luh, S. Chang, and J. Sun. Modeling and optimization of crowd guidance in building emergency evacuation. In *Proceedings of the 2008 IEEE Conference on Automation Science and Engineering*, 2008.

[28] Christopher R. Wren, Yuri A. Ivanov, Ishwinder Kaur, Darren Leigh, and Jonathan Westhues. Socialmotion: Measuring the hidden social life of a building. In *Third International Symposium on Location-and Context-Awareness*, Oberpfaffenhofen, Germany, September 2007. Springer.

[29] J. Zacharias. Choosing a path in the underground: visual information and preference. In *ACUUS 2002 International Conference Urban Underground Space: A Resource for Cities*, November 2002.

[30] O. Zeitouni and M. Gutman. Correction to: "On universal hypotheses testing via large deviations". *IEEE Trans. Inform. Theory*, 37(3, part 1):698, 1991.

[31] O. Zeitouni and M. Gutman. On universal hypotheses testing via large deviations. *IEEE Trans. Inform. Theory*, 37(2):285–290, 1991.

[32] R. Zwanzig. *Nonequilibrium Statistical Mechanics*. Oxford University Press, Oxford, England, 2001.