

An unsupervised spatiotemporal graphical modeling approach to anomaly detection in distributed CPS*

Chao Liu, Sambuddha Ghosal, Zhanhong Jiang, and Soumik Sarkar[†]
Department of Mechanical Engineering, Iowa State University, Ames, IA 50011, USA
{cliu5, sghosal, zhjiang, soumiks}@iastate.edu

ABSTRACT

Modern distributed cyber-physical systems (CPSs) encounter a large variety of physical faults and cyber anomalies and in many cases, they are vulnerable to catastrophic fault propagation scenarios due to strong connectivity among the sub-systems. This paper presents a new data-driven framework for system-wide anomaly detection for addressing such issues. The framework is based on a spatiotemporal feature extraction scheme built on the concept of symbolic dynamics for discovering and representing causal interactions among the sub-systems of a CPS. The extracted spatiotemporal features are then used to learn system-wide patterns via a Restricted Boltzmann Machine (RBM). The results show that: (1) the RBM free energy in the off-nominal conditions is different from that in the nominal conditions and can be used for anomaly detection; (2) the framework can capture multiple nominal modes with one graphical model; (3) the case studies with simulated data and an integrated building system validate the proposed approach.

1. INTRODUCTION

With the advent ubiquitous sensing, advanced computation and strong connectivity, modern distributed cyber-physical systems (CPSs) such as power plants [10], integrated buildings [11], transportation networks [28] and power-grids [25] have shown tremendous potential of increased efficiency, robustness and resilience - a fact

*This material is based upon work supported by the National Science Foundation under Grant No. CNS-1464279.

[†]Corresponding author.

that is reflected in great interest in CPS research shown by the U.S. government [12], the EU [5], other countries around the world [26] and industry in general [2]. However, to realize such potential, effective modeling and analysis tools have to be developed for CPSs that are scalable, robust, flexible and adaptive. Most of the current approaches lack in these qualities as they heavily depend on domain knowledge based rules and first-principle based models that need meticulous calibration and validation.

From the perspective of performance monitoring and diagnostics of distributed CPSs, the technical challenges arise from a large number of subsystems that are highly interactive [18] and operate in diverse modes. Although physics-based modeling of individual small sub-systems, developing models that capture all possible complex interactions often become intractable. Data-driven modeling can potentially alleviate such issues [3]. However, most of such methods need examples of nominal and all possible anomalous conditions (e.g., cyber attacks and physical faults) which is intractable for real life systems. Therefore, anomaly detection approaches should have (i) the potential to recognize most of the operating modes without anomaly as nominal, and (ii) an unsupervised learning ability to distinguish the (possibly unforeseen) anomalies from the nominal modes. Moreover, physical space generates mostly continuous temporal information from sensors and actuators and the cyber space generates mostly discrete event-driven data while processing physical information. Such disparity in fundamental properties and nature of information space drives most of the current approaches to treat cyber and physical spaces separately for modeling and analysis (a detail survey can be found in [13]).

In this context, this work presents a data driven framework for system-wide anomaly detection in distributed CPSs, where a spatiotemporal feature extraction scheme is built on the concept of symbolic dynamics [15] for discovering and representing causal interactions between the subsystems. Symbolic Dynamic Filtering (SDF), as a data driven modeling of complex systems, has advanced

tages in describing different types of data with a uniform representation, named as data abstraction. Abstraction involves pre-processing and data space partitioning for relevant variables (e.g., sensor time-series, event logs) of the system [20], where such a representation helps modeling cyber and physical sub-systems together. Features captured by SDF are used in formation of spatiotemporal pattern network (STPN) [9], a recently proposed causal graphical modeling concept. The causal modeling is followed by unsupervised learning of various system-level nominal patterns [4]. Upon learning such models, the paper develops inference schemes for detection of low probability events or anomalies.

2. BACKGROUND AND PRELIMINARIES

2.1 Spatiotemporal pattern network (STPN)

Symbolic dynamic filtering (SDF) has been recently shown to be extremely effective for extracting key textures from time-series data for anomaly detection and pattern classification [15]. The core idea is that a symbol sequence (i.e., discretized time-series) emanated from a process can be approximated as a Markov chain of order D (also called depth), named as D -Markov machine [22] that captures key behavior of the underlying process. The discretization or symbolization process is called partitioning [15]. There are various approaches proposed in the literature, depending on different objective functions [19], such as uniform partitioning (UP), maximum entropy partitioning (MEP), maximally bijective discretization (MBD) [23]. This study uses simple uniform partitioning. The D -Markov machine is essentially a probabilistic finite state automaton (PFSA) that can be described by states (representing various parts of the data space) and probabilistic transitions among them that can be learnt from data. Related definitions of deterministic finite state automaton (DFSA), PFSA, D -Markov machine, xD -Markov machine and the learning schemes can be found in [22].

With this setup, a spatiotemporal pattern network (STPN) is defined below.

Definition. A PFSA based STPN is a 4-tuple $W_D \equiv (Q^a, \Sigma^b, \Pi^{ab}, \Lambda^{ab})$: (a, b denote nodes of the STPN)

1. $Q^a = \{q_1, q_2, \dots, q_{|Q^a|}\}$ is the state set corresponding to symbol sequences S^a ;
2. $\Sigma^b = \{\sigma_0, \dots, \sigma_{|\Sigma^b|-1}\}$ is the alphabet set of symbol sequence S^b ;
3. Π^{ab} is the symbol generation matrix of size $|Q^a| \times |\Sigma^b|$, the ij^{th} element of Π^{ab} denotes the probability of finding the symbol σ_j in the symbol string s^b while making a transition from the state q_i in

the symbol sequence S^a ; while self-symbol generation matrices are called atomic patterns (APs) i.e., when $a = b$, cross-symbol generation matrices are called relational patterns (RPs) i.e., when $a \neq b$.

4. Λ^{ab} denotes a metric that can represent the importance of the learnt pattern (or degree of causality) for $a \rightarrow b$ which is a function of Π^{ab} .

An illustration of STPN is shown in Fig. 1.

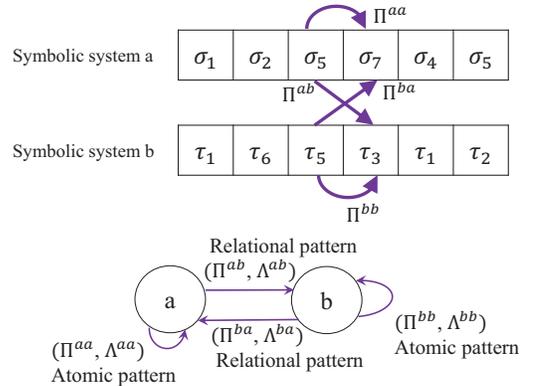


Figure 1: *Extraction of atomic and relational patterns (using D -Markov and xD -Markov machines respectively and $D = 1$, i.e., states and symbols are equivalent) to characterize individual sub-system behavior and interaction behavior among different sub-systems.*

Information based criteria are often used to compute a metric such as Λ^{ij} , e.g., transfer entropy [27] and mutual information [24]. Detailed description of mutual information based causality metric in the context of APs and RPs can be found in [9, 22].

2.2 Restricted Boltzmann Machine (RBM)

RBM has grabbed a lot of recent attention in the Deep Learning community [7, 16] for unsupervised feature extraction. The basic structure of RBM is shown in unsupervised learning layer in Fig. 2 (top-left corner). As an energy based model [7], weights and biases are learnt so that the feature configurations observed during nominal operation of the system gets low energy (or high probability). Consider a system state that is described by a set of visible variables $\mathbf{v} = (v_1, v_2, \dots, v_D)$ and a set of hidden (latent) variables $\mathbf{h} = (h_1, h_2, \dots, h_F)$. The variables can be binary or real-valued depending on the need. Now, each joint configuration of these variables determines a particular state of the system and an energy value $E(\mathbf{v}, \mathbf{h})$ is associated with it. The energy values are functions of the weights of the links between the variables (for RBM, internal links within the visible variables and the hidden variables are not considered) and bias terms related to the variables.

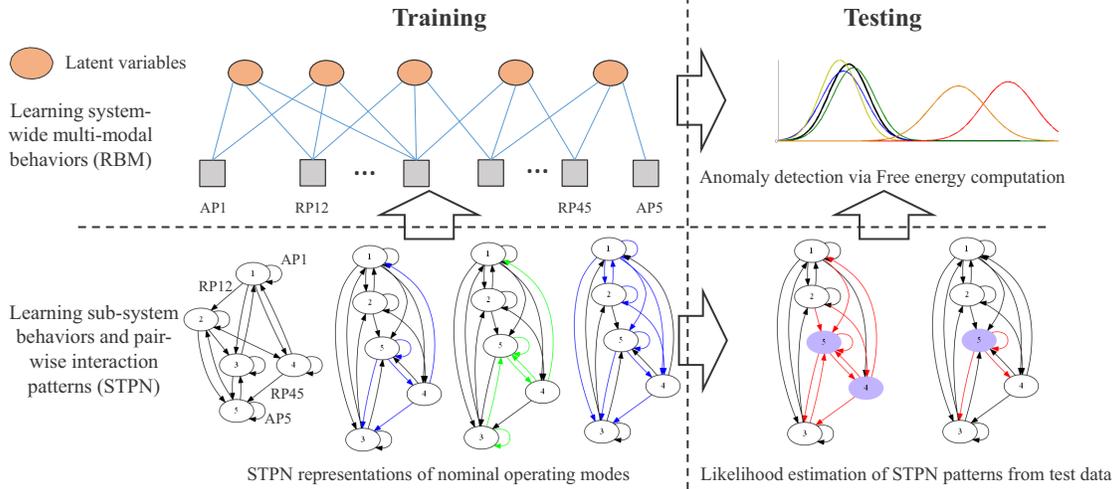


Figure 2: A data driven framework for anomaly detection in distributed CPS (i) Spatiotemporal graphical modeling learn nominal behavior of subsystems and their interactions in various operation modes, and system-wide patterns are learnt by an RBM, (ii) distribution of free energy is used to detect low probability events or anomalies

With this setup, the probability of a state $P(\mathbf{v}, \mathbf{h})$ depends only on the energy of the configuration (\mathbf{v}, \mathbf{h}) and follows the Boltzmann distribution

$$P(\mathbf{v}, \mathbf{h}) = \frac{e^{-E(\mathbf{v}, \mathbf{h})}}{\sum_{\mathbf{v}, \mathbf{h}} e^{-E(\mathbf{v}, \mathbf{h})}} \quad (1)$$

Typically, during training, weights and biases are obtained via maximizing likelihood of the training data.

3. PROPOSED METHODOLOGY

The proposed data-driven framework for system-wide anomaly detection is shown in Fig. 2. During training, the steps of learning the STPN+RBM model are:

1. Learn APs and RPs (individual node behaviors and pair-wise interaction behaviors) from the multivariate training symbol sequences
2. Consider short symbol subsequences from the training sequences and evaluate $\Lambda^{ij} \forall i, j$ for each short subsequence
3. For one subsequence, based on a user-defined threshold on Λ^{ij} , assign state 0 or 1 for each AP and RP; thus every subsequence leads to a binary vector of length L , where $L = \#AP + \#RP$
4. An RBM is used for modeling system-wide behavior with nodes in the visible layer corresponding to APs and RPs
5. The RBM is trained using binary vectors generated from training subsequences

3.1 Training STPN+RBM model of a CPS

Consider a multivariate training time series (nominal operation data in the present context), $X = \{X^{\mathbb{A}}(t), t \in$

$\mathbb{N}, \mathbb{A} = 1, 2, \dots, f\}$, where f is the number of variables or dimension of the time series. At first, symbolization and PFSA learning are performed to extract atomic and relational patterns for the corresponding STPN with f vertices and f^2 edges. In this case, let the set of symbol sequences be $S = \{S^{\mathbb{A}}\}$. Then, we define a short subsequence, $\tilde{X} = \{\tilde{X}^{\mathbb{A}}(t), t \in \mathbb{N}^*, \mathbb{A} = 1, 2, \dots, f\}$, where \mathbb{N}^* is a subset of \mathbb{N} . Essentially, different (possibly overlapping) time windows (represented by $t \in \mathbb{N}^*$) extracted from the overall training data set can be considered as the short subsequences. Similar to the previous notation, a set of symbolic subsequences for one time window is denoted as $\tilde{S} = \{\tilde{S}^{\mathbb{A}}\}$.

The next step is to compute $\Lambda^{ij} \forall i, j$ for every short subsequence extracted from the entire time series. Although information theoretic measures (as mentioned above) can be good candidates, estimation of such metrics need a large number of data points and hence, can be prohibitive at the anomaly detection phase. In this paper, we propose a new measure of $\Lambda^{ij} \forall i, j$ based on the statistical inference strategy developed in [21, 1].

Computation of this metric involves two phases, namely (i) modeling phase and (ii) inference phase.

Modeling Phase: The entire training data is considered in the context of the STPN, $W_D \equiv (Q^a, \Sigma^b, \Pi^{ab}, \Lambda^{ab})$. As the entire set of symbol sequences is denoted by S , let the set of state sequences (obtained by applying the depth D on the symbol sequence) be denoted as $\mathcal{Q} = \{Q^a, a = 1, 2, \dots, f\}$. Recall, a pattern Π^{ab} depends on the state sequence Q^a and the symbol sequence S^b . With this setup, each row of Π^{ab} is treated as a random vector. For the m^{th} row, a prior probability density function $f_{\Pi_m^{ab} | \{Q^a, S^b\}}$ for the random vector Π_m^{ab} , conditioned on the joint state-symbol sequence $\{Q^a, S^b\}$

following the Dirchlet distribution is described below

$$f_{\Pi_m^{ab}|\{\mathcal{Q}^a, S^b\}} = \frac{1}{B(\alpha_m^{ab})} \prod_{n=1}^{|\Sigma^b|} (\theta_{mn}^{ab})^{\alpha_{mn}^{ab}-1} \quad (2)$$

where θ_m^{ab} is a realization of the random vector Π_m^{ab} , namely

$$\theta_{mn}^{ab} = [\theta_{m1}^{ab} \quad \theta_{m2}^{ab} \quad \cdots \quad \theta_{m|\Sigma^b|}^{ab}]$$

and the normalizing constant is

$$B(\alpha_m^{ab}) \triangleq \frac{\prod_{n=1}^{|\Sigma^b|} \Gamma(\alpha_{mn}^{ab})}{\Gamma(\sum_{n=1}^{|\Sigma^b|} \alpha_{mn}^{ab})} \quad (3)$$

where $\alpha_{mn}^{ab} \triangleq [\alpha_{m1}^{ab} \quad \alpha_{m2}^{ab} \quad \cdots \quad \alpha_{m|\Sigma^b|}^{ab}]$ with $\alpha_{mn}^{ab} = N_{mn}^{ab} + 1$ and N_{mn}^{ab} is the number of times the symbol $\sigma_n^b \in \Sigma^b$ is emanated after the state $q_m^a \in \mathcal{Q}^a$, i.e.,

$$N_{mn}^{ab} \triangleq |\{(Q^a(k), S^b(k+1)) : S^b(k+1) = \sigma_n^b \mid Q^a(k) = q_m^a\}| \quad (4)$$

where $Q^a(k)$ is the k^{th} state in the state sequence \mathcal{Q}^a and $S^b(k+1)$ is the $(k+1)^{th}$ symbol in the symbol sequence S^b .

It follows from Eq. 3 that

$$B(\alpha_m^{ab}) = \frac{\prod_{n=1}^{|\Sigma^b|} \Gamma(N_{mn}^{ab} + 1)}{\Gamma(\sum_{n=1}^{|\Sigma^b|} N_{mn}^{ab} + |\Sigma^b|)} = \frac{\prod_{n=1}^{|\Sigma^b|} (N_{mn}^{ab})!}{(N_m^{ab} + |\Sigma^b| - 1)!} \quad (5)$$

by using the relation $\Gamma(n) = (n-1)!$.

With the Markov property of the learned PFSA, the row-vectors, $\{\Pi_m^{ab}, m = 1, 2, \dots, |Q^a|\}$, are statistically independent of each other. Therefore, it follows Eqs. 2 and 5 that the prior joint density $f_{\Pi_m^{ab}|\{\mathcal{Q}^a, S^b\}}$ of the probability morph matrix Π^{ab} , conditioned on the joint state-symbol sequences $\{\mathcal{Q}^a, S^b\}$, is given by

$$\begin{aligned} & f_{\Pi^{ab}|\{\mathcal{Q}^a, S^b\}}(\theta^{ab}|\{\mathcal{Q}^a, S^b\}) \\ &= \prod_{m=1}^{|\mathcal{Q}^a|} f_{\Pi_m^{ab}|\{\mathcal{Q}^a, S^b\}}(\theta_m^{ab}|\{\mathcal{Q}^a, S^b\}) \\ &= \prod_{m=1}^{|\mathcal{Q}^a|} (N_m^{ab} + |\Sigma^b| - 1)! \prod_{n=1}^{|\Sigma^b|} \frac{(\theta_m^{ab})^{N_{mn}^{ab}}}{(N_{mn}^{ab})!} \end{aligned} \quad (6)$$

where $\theta^{ab} = [(\theta_1^{ab})^T (\theta_2^{ab})^T \cdots (\theta_{|Q^a|}^{ab})^T] \in [0, 1]^{|Q^a| \times |\Sigma^b|}$.

Inference Phase: After modeling with the entire set of training sequences, the goal of the inference phase is to compute the metric $\Lambda^{ab}(\tilde{\mathcal{Q}})$ for a given short subsequence described by $\tilde{\mathcal{Q}}$ and \tilde{S} . The value of this metric suggests the importance of the pattern Π^{ab} or the degree of causality in $a \rightarrow b$ as evidenced by the short

subsequence. In this context, we consider

$$\Lambda^{ab}(\tilde{\mathcal{Q}}, \tilde{S}) \propto Pr(\{\tilde{\mathcal{Q}}^a, \tilde{S}^b\}|\Pi^{ab}) \quad (7)$$

where the probability of the joint state-symbol subsequence is a product of independent multinomial distributions given that the exact morph matrix is known.

$$Pr(\{\tilde{\mathcal{Q}}^a, \tilde{S}^b\}|\Pi^{ab}) = \prod_{m=1}^{|\mathcal{Q}^a|} (\tilde{N}_m)! \prod_{n=1}^{|\Sigma^b|} \frac{(\Pi_{mn}^{ab})^{\tilde{N}_{mn}}}{(N_{mn}^{ab})!} \quad (8)$$

where, the definition of \tilde{N}_{mn}^{ab} is similar to N_{mn}^{ab} in the context of the short subsequence.

With similar derivation in [21], the metric $\Lambda^{ab}(\tilde{\mathcal{Q}}, \tilde{S})$ can be obtained as follows

$$\begin{aligned} \Lambda^{ab}(\tilde{\mathcal{Q}}, \tilde{S}) &= K \prod_{m=1}^{|\mathcal{Q}^a|} \frac{(\tilde{N}_m^{ab})!(N_m^{ab} + |\Sigma^b| - 1)!}{(\tilde{N}_m^{ab} + N_m^{ab} + |\Sigma^b| - 1)!} \\ &\quad \prod_{n=1}^{|\Sigma^b|} \frac{(\tilde{N}_{mn}^{ab} + N_{mn}^{ab})!}{(\tilde{N}_{mn}^{ab})!(N_{mn}^{ab})!} \end{aligned} \quad (9)$$

where, K is a proportional constant.

Thus, with Eq. 9, importance metrics of APs (i.e., when $a = b$) and RPs (i.e., when $a \neq b$) are obtained with respect to the *short subsequences*. In order to train the system-wide RBM, the metrics can be further normalized and converted to binary states (0 for low values and 1 for high values) for APs and RPs. Note, from each subsequence, all the APs and RPs together form a binary vector of length $L = f^2$ ($L = \#AP + \#RP$, where $\#AP = f$, $\#RP = f \times (f-1)$). One such binary vector is treated as one training example for the system-wide RBM (with f^2 number of visible units) and many such examples are generated from different short subsequences extracted from the overall training sequence. Then a maximum likelihood method is used to train the RBM as mentioned in 2.2. Note, although in this paper we convert the $\Lambda^{i,j}$ metrics to binary values for the ease of RBM training, it is not mandatory for this training process.

Remark. With an optimized time lag in STPN learning, the STPN+RBM model can handle variable time lag situations and significantly reduces the complexity of the overall learning process, in comparison with other potentially spatiotemporal approaches, such as recurrent neural networks (RNNs).

3.2 Anomaly detection process

The anomaly detection process developed here uses the concept of *free energy* of RBM which is an energy based probabilistic graphical model. The energy function for an RBM is defined as

$$E(\mathbf{v}, \mathbf{h}) = -\mathbf{h}^T \mathbf{W} \mathbf{v} - \mathbf{b}^T \mathbf{v} - \mathbf{c}^T \mathbf{h} \quad (10)$$

where \mathbf{W} are the weights of the hidden units, \mathbf{b} and \mathbf{c} are the biases of the visible units and hidden units, respectively.

With the weights and biases of RBM, free energy can be obtained which is the energy that a single visible layer pattern would need to have in order to have the same probability as all of the configurations that contain \mathbf{v} [8]

$$e^{-F(\mathbf{v})} = \sum_{\mathbf{h}} e^{-E(\mathbf{v}, \mathbf{h})} \quad (11)$$

Another expression of free energy estimation is [8]

$$F(v) = - \sum_i v_i a_i - \sum_j \log(1 + e^{b_j + \sum_i v_i w_{ij}}) \quad (12)$$

During training, weights and biases are obtained such that the training data has low energy. Therefore, during testing, an anomalous pattern should manifest itself as a low probability (high energy) configuration which can be used for anomaly detection. To detect an occurrence of anomaly, short testing subsequences are converted into an f^2 -dimensional binary vector with the same *inference phase* of the training process. Multiple testing (possibly overlapping) subsequences can be used to obtain a distribution of free energy. For the nominal condition, the distribution of free energy should be close to that of the training data, while the anomalous data should result in a different distribution of free energy (with increase in expected free energy).

To compare the free energy distributions of testing data and training data, Kullback-Leibler divergence is considered [14].

Since the Kullback-Leibler (KL) divergence is a non-symmetric information measure of distance between distribution P and distribution Q , a symmetric Kullback-Leibler Distance (KLD) [14] is used which is defined as,

$$KLD(P||Q) = KL(P||Q) + KL(Q||P) \quad (13)$$

4. RESULTS AND DISCUSSIONS

4.1 Anomaly detection in simulation data

4.1.1 Synthetic data with Vector Autoregressive (VAR) Process

Vector autoregressive (VAR) process is widely applied in economics and other sciences as it is flexible and simple for multivariate time series data [6]. The basic model of VAR process $Y(t) = y_{i,t}$, $i = (1, 2, \dots, f)$, $t \in \mathbb{N}$ is defined as

$$y_{i,t} = \sum_{k=1}^p \sum_{j=1}^n A_{i,j}^k y_{j,t-k} + \mu_t, \quad j = (1, 2, \dots, f) \quad (14)$$

where p is time lag order, $A_{i,j}$ is the coefficient of the influence on i th time series caused by j th time series,

and μ_t is an error process with the assumption of white noise with zero mean, that is $E(\mu_t) = 0$, the covariance matrix, $E(\mu_t \mu_t') = \Sigma_\mu$ is time invariant.

Synthetic multivariate data is generated using the VAR process in this study. A hierarchical structure of five-vertex network is considered with different interactions among vertices. Various interaction patterns represent nominal and off-nominal conditions in a distributed CPS. Data for two case studies are generated: *Case I*: six patterns are defined, where the first one is considered as nominal condition and the others as anomalous; *Case II*: eight patterns are defined, where the first three are considered as nominal conditions, and remaining five as anomalous. The second case is to simulate the CPS scenario with multiple nominal operating modes.

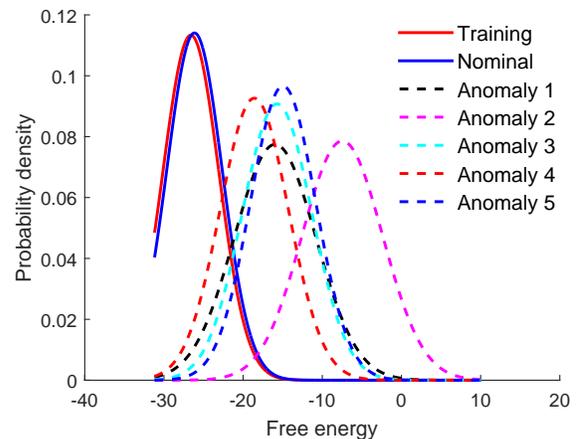


Figure 4: *Distribution of free energy of RBM with multiple testing samples from each nominal/anomalous condition*

STPNs are learnt from raw data generated by the VAR models representing different interaction patterns. For unsupervised RBM learning, only nominal patterns are used to obtain the weights and biases. Data from all patterns are used with the trained RBM to compute the free energy with respect to the different pattern inputs. Evaluation of free energy is performed with Gaussian assumption of multiple testing inputs.

4.1.2 Case Study I: Single nominal mode

While the details of STPN learning is skipped for brevity, the predefined graphical models are shown in Fig. 3. Atomic and relational patterns of the nominal condition are used for training the RBM. Testing data corresponding to all 6 patterns are used to generate the probability distribution of free energy with respect to the trained RBM and the results are shown in Fig. 4. It is evident, while the free energy distribution of the nominal pattern matches quite closely with that of the training patterns, free energy distributions of other patterns are quite different and moves to the

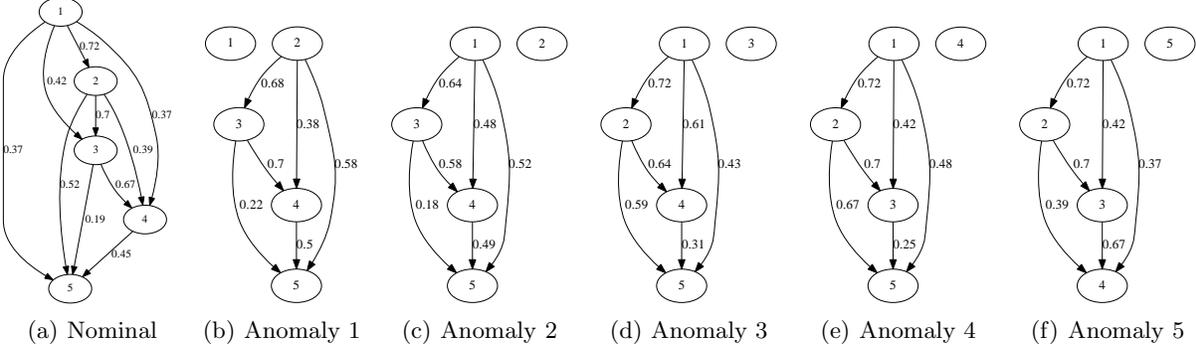


Figure 3: Graphical models defined for synthetic data generation and anomaly detection, where anomalous conditions occur due to failure of one vertex and the causal links to the failed vertex are lost

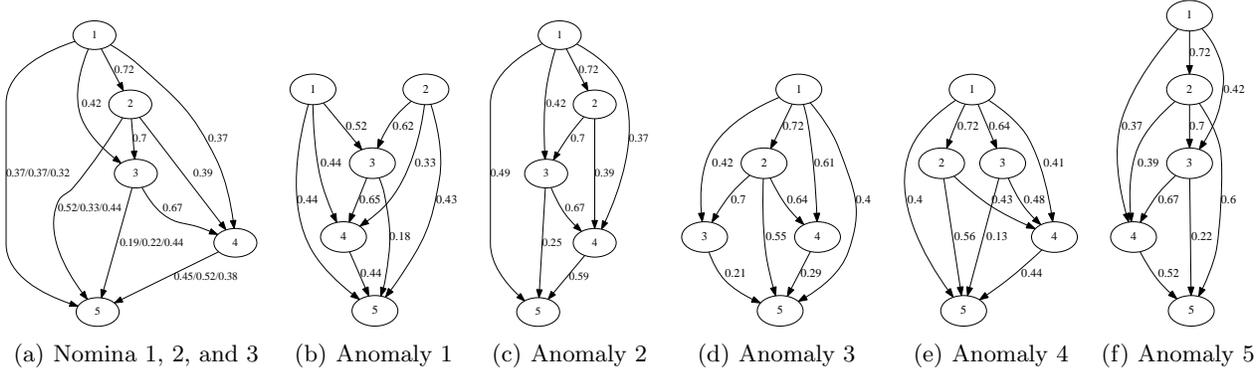


Figure 5: Graphical models defined for synthetic data generation and anomaly detection, where the nominal condition consists of (first) three graphs representing three different operation modes, and the others are considered as anomalous

right. This means such patterns have high free energy and hence low probability of occurrence. Quantitatively, the KLD metrics between the free energy distribution of the training/nominal pattern and those of all the other test patterns are 0.01, 6.82, 21.70, 7.63, 4.17, and 9.08 respectively.

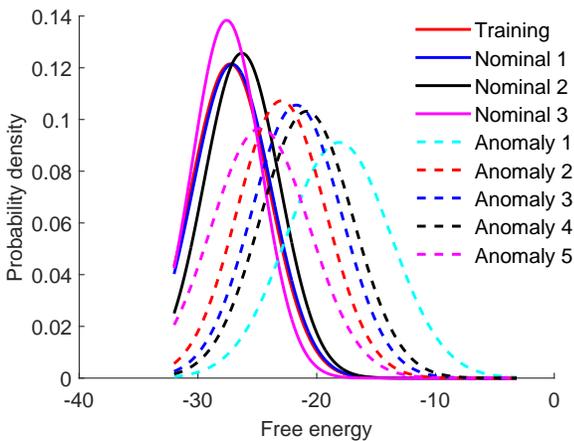


Figure 6: Distribution of free energy of RBM with multiple testing samples for each nominal/anomalous condition

4.1.3 Case Study II: Multiple nominal modes

Similar to the previous case, the predefined graphical models are shown in Fig. 5 and the free energy distributions are shown in Fig. 6. As expected, free energy distributions for first three patterns (Nominal 1, 2 and 3) are similar to that of the training data with KLD values 0.002, 0.068 and 0.040 respectively. The figure also shows that anomalous patterns can be clearly identified for conditions 4 through 8 (with KLD values 6.067, 1.461, 2.420, 3.199, and 0.906 respectively). Overall, the result clearly shows that the proposed framework can capture multiple nominal behaviors within one model while slight change in causality patterns can be detected efficiently.

4.2 Validation on a Real System - Integrated Building

The Interlock House, an Iowa NSF EPSCOR (experimental program to stimulate competitive research) (# EPS-1101284.) community lab, located at the Honey Creek Resort State Park, Iowa is a smart home designed to “interlock” with its environment, its occupants, and its active and passive energy systems. The goal of this test bed is to perform as a net-zero energy house that (ideally) produces as much energy as it consumes. With all its intelligent sensing and control systems, it can be

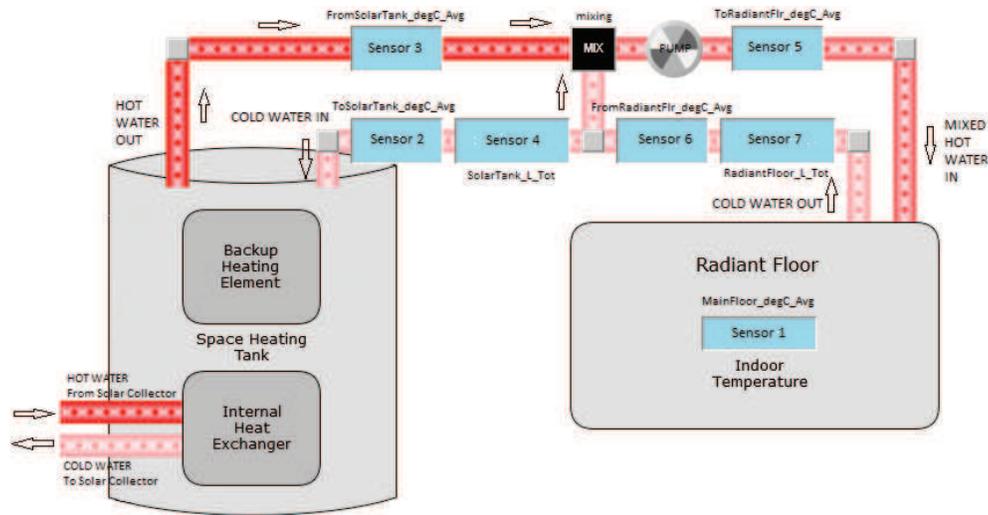


Figure 7: Radiant floor heating subsystem within the integrated building system, including 7 sensors measuring temperature, flow rate, and power consumption. Sensors 1, 2, 3, 5, and 6 are temperature sensors, sensors 4 and 7 are flow rate sensors, sensors 8 and 9 measure power consumption of the heating tank (not marked in the figure).

classified as a typical cyber-physical system, with collaborating computational elements controlling the interconnected physical sub-systems. One of the key systems for building operation is the HVAC (Heating, Ventilation & Air-Conditioning) system that is comprised of the following three independent sub-systems: (a) Radiant Floor Heating (Space Heating), (b) Domestic Water Heating, and (c) Space Cooling and Ventilation. For this paper, the Radiant Floor Heating subsystem is chosen and studied. A schematic of the Radiant Floor Heating system is shown in Fig. 7.

Sensor data related to its operation is collected for a period of 1 year from Jul., 2014 to Jun., 2015. Within this time-frame, data collected for two different seasonal operational modes (viz. winter and summer modes) is studied. For winter mode of operation, the month of Nov., 2014 is chosen and Jul., 2015 is chosen for summer mode. Typical data for two representative sensors are shown in Fig. 8, where features vary in seasons and are used for unsupervised learning. Another timeframe from Feb., 2015 to Jun., 2015 is considered for validation of the proposed approach that covers a wide range of behavior over both winter and summer, as shown in Fig. 9.

Winter Mode study (Nov., 2014): In Winter, periodic heating of the radiant floor space is required. The heating is set to start every morning at 5:00 am if the indoor temperature at that point is below 19.5 deg C. The heating turns off when the heating setpoint (plus the dead band) is reached, which was around 20 deg C, with the dead band being 1 deg C (+0.5C/-0.5C).

- This is evident from Fig.8 (a) where, we see a rise in the temperature of the water supplied from

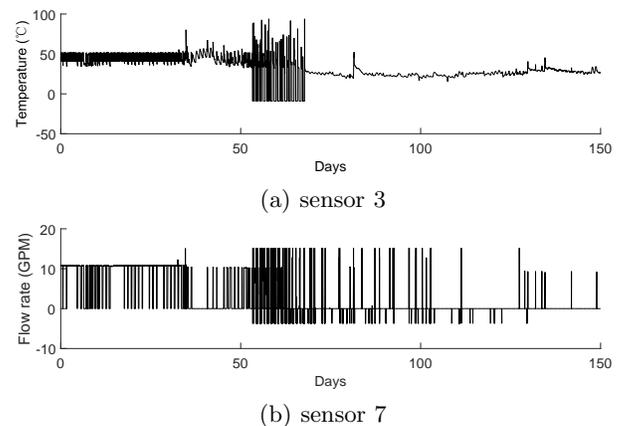


Figure 9: Data collected at the Interlock House during Feb., 2015 – Jun., 2015.

the space heating tank to radiant floor at approximately every 24 hour time interval, when demand for heating is high. The temperature of the supplied water gradually decreases as demand for heating decreases and it rises again at 5:00 am next day, if the set conditions are met.

- Consequently, it is seen in Fig.8 (b) that the flowmeter measuring flow of water returning from the radiant floor records flow of water in a similar pattern. Flow measurements are recorded at every 24 hour interval when heating is required, depending on the set conditions.
- It is observed that on the fourth day, flow measurements remained zero and floor heating was not carried out. This is because, the indoor temperature that day remained above 19.5 deg C at 5 am.

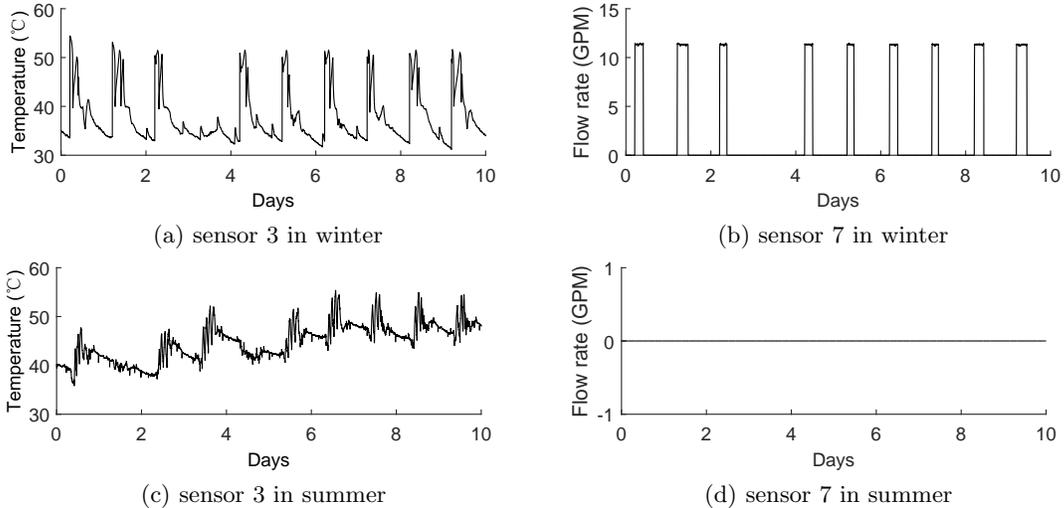


Figure 8: Data collected at the Interlock House in winter and summer, the subsystem includes 9 sensors measuring various temperatures, flow rates, and power consumptions.

Summer Mode study (July, 2014): In Summer, water is heated in the Space Heating Tank periodically, as is evident from Fig. 8 (c), but as floor heating in summer is not required, supply to the radiant floor remains 0 all throughout, as is evident from Fig. 8 (d).

Anomaly Detection: On the study conducted on the data from Feb., 2015 – Jun., 2015, we notice that anomalies creep in certain sensor observations (see Fig. 9). While the first anomaly starts around March 26, 2015 (day 55 in the plot) and continues till June with erroneous readings in several sensors, a second anomaly appears during June, 2015. However, careful observation shows that the second anomaly is extremely intermittent and only manifests itself in a few number of (three out of nine) sensor measurements.

Online anomaly detection is carried out using the proposed technique and the results are shown in Fig. 10. The algorithm detects the first anomaly quite easily that appears as a large peak in the plot. The second anomaly which is local, intermittent and not much pronounced, is also detected and appears as the second sharp peak in the plot. Another crucial observation is that we use different sets of test data for two different modes of operation (summer and winter). The proposed technique can effectively handle nominal data for all such modes without false detection.

4.3 Discussions

Anomalies in distributed CPSs vary in mechanisms, features, and duration, and this makes anomaly detection difficult, especially the collection of labeled data for all possible anomalies. The proposed framework in this work only needs nominal data, and anomaly detection is considered as a low probability event conditioned on

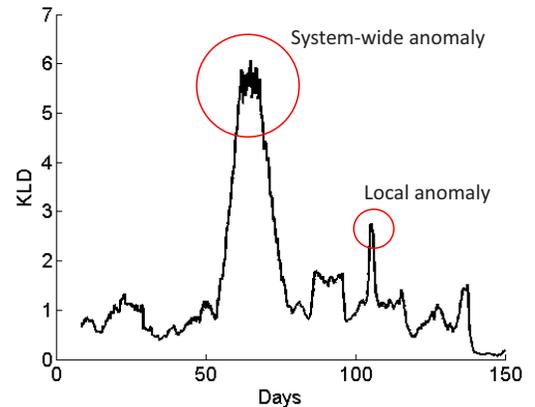


Figure 10: Online anomaly detection during Feb. 2015 – Jun. 2015.

the nominal data. The results show that the free energy distribution during an anomalous condition is different from that of the nominal condition, and a metric such as KLD can be used to quantify the change. Furthermore, it has the potential for monitoring all the way from small physical degradation to severe faults or cyber attacks, where a small KLD signifies slight change in the causal pattern (STPN) that possibly indicates early-stage degradation, localized cyber anomaly or precursor of a fault.

Multiple nominal modes: The proposed framework can capture multiple modes as nominal condition (e.g., Fig. 5 and Fig. 8), within a single model and hence reduces the modeling and reasoning complexity. Traditionally, a graphical model such as STPN can represent one operating mode at a time, and multiple STPNs are required to represent various operating modes. Moreover, it is difficult to determine the number of STPNs needed for representing all nominal modes, as a small

number of changes in the patterns can result in a larger number of combinations, which may cause dimension explosion and extreme difficulty in real life applications.

Local and Global anomalies: In the simulated cases shown in Fig. 5, only one causal connection is removed (e.g., arrow $1 \rightarrow 2$ is disconnected in STPN 4, comparing with STPNs 1, 2, and 3), and it represents a local anomaly. In the real case study, most of the measurements are abnormal during days 55-70 in Fig. 9, and it is a type of global anomaly. For both of types of anomalies, the proposed approach can distinguish them from the nominal modes.

Mixed data types: In the real case study (Section 4.2), mixed data types are used. For example, temperature measurement in Fig. 8 (a) is a representative continuous sensor observation. On the other hand, the flow rate measurement in Fig. 8 (b) appears as fairly discrete due to the on-off control schedule. The proposed approach is able to extract features from both types of data and fuse them into one model. This is a critical need for a distributed CPS with continuous temporal information from sensors and actuators of the physical space and discrete event-driven data from the cyber space.

Robustness: Although detailed false alarm and missed detection studies will be done in near future, from a qualitative standpoint, the approach is seen to be quite robust as it is designed to identify only persistent anomalies. This property is inherent as the anomaly detection is performed with short subsequences and not based on individual data points. As a consequence, no false alarms are observed when only a few sensors show bad readings for a very short time window. For example, error in readings occurs occasionally during the first 50 days (in Fig. 9), where no anomaly is detected during that period (in Fig. 10).

Remark: The proposed framework is categorized as unsupervised because unlike typical diagnostic frameworks, no labeled training data for various fault types and fault locations are required; only assumption made for the training data is that it represents nominal condition(s) with a high probability. Even with such limited training data, the proposed tool has the potential of identifying root-cause of a detected anomaly. Demonstration of such a capability remains a critical future work. Furthermore, for learning multiple STPNs with one-layer RBM, the results are acceptable in the above cases. However, differences among nominal modes may be quite large in some cases and one-layer RBM may not capture all the features in such cases. To address this issue, stacked RBM (i.e., a deep belief network (DBN)) can be used [16, 17] to learn more complex features.

5. CONCLUSIONS

This work presents a new data-driven framework for system-wide anomaly detection to address a wide-range of nominal operating modes and unforeseen anomalous situations without comprehensive labeled training data in distributed CPSs. The framework involves a spatiotemporal feature extraction scheme for discovering and representing causal interactions among the subsystems of a CPS, and a free energy estimation of system-wide patterns using an RBM. The results show that the proposed framework can capture multiple diverse nominal modes within a single probabilistic graphical model, and detect anomalies via identifying a low probability event. Multiple case studies with simulation and real data validate accuracy, robustness, ability to handle mixed data type and adaptiveness (i.e., local vs. global) of the proposed method.

While the current work is focusing on validating the method for a large variety of scenarios, quantifying false alarm and missed detection rates, further works will pursue the following: (i) using the graphical model for root-cause analysis for various anomalies, (ii) stacked RBM approach to capture more complex nominal patterns and (iii) detection of simultaneous multiple faults in distributed CPS.

Acknowledgments

The authors would like to express their sincere gratitude to Dr. Soumalya Sarkar and Mr. Adedotun Akintayo for their thoughtful comments on the framework, Prof. Ulrike Passe, Ms. Shan He (Center for Building Energy Research at Iowa State University), and Mr. Mike Wassmer (Live to Zero Inc.) for their support in the validation process using real data, and for providing access to their historic data collected as part of the building science plank research of the Iowa NSF EPSCOR project funded under grant number EPS-1101284.

6. REFERENCES

- [1] A. Akintayo and S. Sarkar. A symbolic dynamic filtering approach to unsupervised hierarchical feature extraction from time-series data. In *Proceedings of American Control Conference*, 2015.
- [2] J. Bradley, J. Barbier, and D. Handler. Embracing the internet of everything to capture your share of \$14.4 trillion. *CISCO White Paper*, 2013.
- [3] A. Choi, L. Zheng, A. Darwiche, and O. Mengshoel. A tutorial on bayesian networks for system health management. *Machine Learning and Knowledge Discovery for Engineering Systems Health Management*, 10.1:1–29, 2011.
- [4] U. Fiore, F. Palmieri, A. Castiglione, and A. De Santis. Network anomaly detection with the restricted boltzmann machine. *Neurocomputing*, 122:13–23, 2013.

- [5] J. Fitzgerald, S. Foster, C. Ingram, P. G. Larsen, and J. Woodcock. Model-based engineering for systems of systems: the compass manifesto. <http://www.compass-research.eu/Project/Publications/MBESoS.pdf>, 2013.
- [6] R. Goebel, A. Roebroek, D.-S. Kim, and E. Formisano. Investigating directed cortical interactions in time-resolved fmri data using vector autoregressive modeling and granger causality mapping. *Magnetic resonance imaging*, 21(10):1251–1261, 2003.
- [7] G. Hinton and R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313.5786:504–507, 2006.
- [8] G. E. Hinton. A practical guide to training restricted boltzmann machines. In *Neural Networks: Tricks of the Trade*, pages 599–619. Springer, 2012.
- [9] Z. Jiang and S. Sarkar. Understanding wind turbine turbine interactions using spatiotemporal pattern network. In *Proceedings of ASME Dynamics Systems and Control Conference*, 2015.
- [10] B. Kesler. The vulnerability of nuclear facilities to cyber attack. *Strategic Insights*, 10(1):15–25, 2011.
- [11] J. Kleissl and Y. Agarwal. Cyber-physical energy systems: focus on smart buildings. In *Proceedings of the 47th Design Automation Conference*, pages 749–754. ACM, 2010.
- [12] D. Kramer. White house offers encouragement for cyberphysical systems. *Physics Today*, 67(9):20–22, 2014.
- [13] S. Krishnamurthy, S. Sarkar, and A. Tewari. Scalable anomaly detection and isolation in cyber-physical systems using bayesian networks. In *Proceedings of ASME Dynamical Systems and Control Conference, San Antonio, TX, USA*, October 2014.
- [14] S. Kullback and R. A. Leibler. On information and sufficiency. *The annals of mathematical statistics*, pages 79–86, 1951.
- [15] C. Rao, A. Ray, S. Sarkar, and M. Yasar. Review and comparative evaluation of symbolic dynamic filtering for detection of anomaly patterns. *Signal, Image and Video Processing*, 3(2):101–114, 2009.
- [16] N. L. Roux and Y. Bengio. Representational power of restricted boltzmann machines and deep belief networks. *Neural Computation*, 20.6:1631–1649, 2008.
- [17] R. Salakhutdinov and G. Hinton. An efficient learning procedure for deep boltzmann machines. *Neural Computation*, 24(8):1967–2006, 2012.
- [18] T. Sanislav and L. Miclea. Cyber-physical systems-concept, challenges and research areas. *Journal of Control Engineering and Applied Informatics*, 14(2):28–33, 2012.
- [19] S. Sarkar. *Hierarchical symbolic perception in dynamic data driven application systems*. PhD thesis, The Pennsylvania State University, August 2015.
- [20] S. Sarkar, K. Mukherjee, and A. Ray. Generalization of hilbert transform for symbolic analysis of noisy signals. *Signal Processing*, 89(6):1245–1251, 2009.
- [21] S. Sarkar, K. Mukherjee, S. Sarkar, and A. Ray. Symbolic dynamic analysis of transient time series for fault detection in gas turbine engines. *Journal of Dynamic Systems, Measurement, and Control*, 135(1):014506, 2013.
- [22] S. Sarkar, S. Sarkar, N. Virani, A. Ray, and M. Yasar. Sensor fusion for fault detection and classification in distributed physical processes. *Frontiers in Robotics and AI*, 1:16, 2014.
- [23] S. Sarkar, A. Srivastav, and M. Shashanka. Maximally bijective discretization for data-driven modeling of complex systems. In *American Control Conference (ACC), 2013*, pages 2674–2679. IEEE, 2013.
- [24] V. Solo. On causality and mutual information. In *Decision and Control, 2008. CDC 2008. 47th IEEE Conference on*, pages 4939–4944. IEEE, 2008.
- [25] S. Sridhar, A. Hahn, and M. Govindarasu. Cyber-physical system security for the electric power grid. *Proceedings of the IEEE*, 100(1):210–224, 2012.
- [26] J. Sztipanovits, S. Ying, I. Cohen, D. Corman, J. Davis, H. Khurana, P. Mosterman, V. Prasad, and L. Stormo. Strategic R&D opportunities for 21st century cyber-physical systems. Technical report, Technical Report for Steering Committee for Foundation in Innovation for Cyber-Physical Systems: Chicago, IL, USA, 13 March, 2012.
- [27] M. Wibrals, B. Rahm, M. Rieder, M. Lindner, R. Vicente, and J. Kaiser. Transfer entropy in magnetoencephalographic data: Quantifying information flow in cortical and cerebellar networks. *Progress in biophysics and molecular biology*, 105(1):80–97, 2011.
- [28] D. Work and A. Bayen. Impacts of the mobile internet on transportation cyberphysical systems: traffic monitoring using smartphones. In *National Workshop for Research on High-Confidence Transportation Cyber-Physical Systems: Automotive, Aviation, and Rail*, pages 18–20, 2008.