

# Data-driven persistent monitoring of Indoor Air Systems

**Sambuddha Ghosal**

[Grad. RA, ME, Iowa State University]

**Chao Liu, PhD**

[Post-doc Fellow, ME, Iowa State University]

**Ulrike Passe**

[Associate Professor, Architecture, Iowa State University, Associate Member, ASHRAE, AIA]

**Shan He**

[Post-grad. Associate, Architecture, Iowa State University]

**Soumik Sarkar, PhD**

[Assistant Professor, ME, Iowa State University]

## ABSTRACT

*Persistent monitoring of Indoor Air Quality (IAQ) within and around buildings and structures is critical to reduce risk of indoor health concerns. Specifically, IAQ issues in large integrated buildings may stem from inadequate ventilation and/or faults in the complex HVAC systems that together with control and communication systems can be considered as complex Cyber Physical Systems (CPSs). We propose a data-driven framework for monitoring distributed complex CPSs that reliably captures cyber and physical sub-system behaviors as well as their interaction characteristics. Using such learning methods, we aim to identify the anomalies and faults at an early stage such that necessary mitigation measures can be pursued in time. A fault in the HVAC system may be due to both physical and cyber anomalies affecting the operational goals of the building system. The proposed technique involves modeling of cyber and physical entities using probabilistic graphical models that capture individual characteristics of the sub-system and causal dependencies among different sub-systems. The proposed model can be trained using nominal historical data and then can be used to monitor the HVAC system and IAQ during regular operation. We validate our method with a case study on an integrated “zero-energy” (low energy/high performance) building, the Interlock House experimental test bed that is developed and maintained by the Center for Building Energy Research (CBER) at Iowa State.*

## I. INTRODUCTION

As humans have progressed into the 21<sup>st</sup> century, there has been a growing trend in people’s daily lives to spend more time inside their homes and office spaces rather than being out and about. According to the national human activity pattern survey, Americans spent 87% of their time indoors and 6% in vehicles on average (Klepeis et al., 2001). As such, to help improve the quality of air within closed spaces has always been a subject of extensive studies and research since the late 20<sup>th</sup> century and have continued to this day into the early 21<sup>st</sup>. Indoor air quality, as the term suggests refers to the quality of air indoors or inside closed spaces. HVAC systems are now employed almost in every building, be it residential/industrial or commercial to maintain the quality of air inside closed spaces. In highrise buildings where such quality control is of utmost importance to maintain proper and healthy environment for the large amount of occupants staying inside. Maintaining IAQ involves maintaining the indoor SPM (suspended

All the authors are affiliated with the Iowa State University, Ames, Iowa. Sambuddha Ghosal is a PhD student in the Department of Mechanical Engineering, Chao Liu is a post-doctoral research fellow in the Department of Mechanical Engineering, Ulrike Passe is an Associate Professor of Architecture, Shan He is a post-graduate associate in the Department of Architecture and Soumik Sarkar is an Assistant Professor of Mechanical Engineering.

particulate matter) levels, detecting unwanted and foreign substances present in the air that may cause health concerns serious or otherwise and also maintaining a comfort zone for the inhabitants, which includes maintaining the proper ambient temperature and humidity levels inside the occupied space. Pollutants and potentially harmful substances that may be present in air include tobacco smoke (allowed in certain few public places like casinos), CO, NO<sub>2</sub> and many volatile organic compounds (VOCs), to name a few other than SPM.

Sometimes, these HVAC systems that maintain IAQ in occupied spaces may malfunction or may deliberately be tampered with. In such cases there arises a risk of potential hazard for the inhabitants. Data driven modeling can potentially alleviate these issues (Choi et al., 2011). In order to promptly detect such cases and alleviate the risks that such events carry by taking prompt action, we propose a method of anomaly detection. In this data-driven framework for such a system-wide anomaly detection, we use symbolic dynamics (Rao et al., 2009) to develop a spatiotemporal feature extraction scheme for seeking out and representing causal interactions between the different measurements. Symbolic Dynamic Filtering (SDF) describes different types of data with a uniform symbolic representation that involves pre-processing and data space partitioning for relevant variables (which for our case study are sensor time-series data). Features captured by SDF are used in formation of spatiotemporal pattern network (STPN), a causal graphical modeling concept proposed only recently (Sarkar et al., 2014) (Jiang et al., 2015). The causal modeling is followed by unsupervised learning of various system-level nominal patterns (Fiore et al., 2013). On learning said models, inference schemes for detection of low-probability events i.e. anomalies are developed (Liu et al., 2016).

In this study, for anomaly detection purposes and considering health concerns of the inhabitants, we consider the case of excess CO<sub>2</sub> generation and build up inside the experimental test space only. Experiments are carried out where we recreate a scenario where our algorithm detects the condition inside the closed space as anomalous when CO<sub>2</sub> concentrations exceed the safe level characterized by nominal historical data. This can be seamlessly extended to any other air pollutants (viz. CO or NO<sub>2</sub> and/or contaminants) if situations demand that. Apart from CO<sub>2</sub> levels, the algorithm also takes into account humidity and temperature levels inside the test space and generates a metric that can characterize the conditions to be anomalous based on the inputs from all three sensor modalities.

## II. BACKGROUND AND PRELIMINARIES

### A. Spatiotemporal pattern network (STPN)

It has been recently shown that symbolic dynamic filtering (SDF) can be extremely effective for extracting key textures from timeseries data for anomaly detection and pattern classification (Rao et al., 2009). 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 (Sarkar et al., 2014) that captures key behavior of the underlying process.

The discretization or symbolization process is called partitioning. Let  $\mathbf{X}$  denote a set of partitioning functions,  $\mathbf{X} : X(t) \rightarrow \mathcal{S}$ , that transforms a general dynamic system (time series  $X(t)$ ) into a symbol sequence  $S$  with an alphabet set  $\Sigma$ . There are various approaches proposed in the literature, depending on different objective functions, such as uniform partitioning (UP), maximum entropy partitioning (MEP), maximally bijective discretization (MBD) (Sarkar et al., 2013) and statistically similar discretization (SSD) (Sarkar et al., 2016). 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,  $\infty D$ -Markov machine and the learning schemes can be found in (Sarkar et al., 2014).

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

**Definition.** A PFSA based STPN is a 4-tuple  $W_D \equiv (Q^a, \Sigma^b, \Pi^{ab}, \Lambda^{ab})$ : ( $a, b$  denotes 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)  $\Sigma^b = \{\sigma_0, \sigma_1, \dots, \sigma_{|\Sigma^b|-1}\}$  is the alphabet set of symbol sequence  $S^b$ ;
- (3)  $\Pi^{ab}$  is the symbol generation matrix of size  $|Q^a| \times |\Sigma^b|$ , the  $ij^{th}$  element of  $\Pi^{ab}$  denotes the probability of finding the symbol 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  $a = b$ , crosssymbol generation matrices are called relational patterns (RPs) i.e., when  $a \neq b$ .

(4)  $\Lambda^{ab}$  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  $\Pi^{ab}$ .

Further details and descriptions on STPN can be found in (Liu et al., 2016).

## B. Restricted Boltzmann Machine (RBM) – boosting and system-wide learning approach

RBM has grabbed a lot of recent attention in the Deep Learning community (Hinton et al., 2006), (Roux et al., 2008) for unsupervised feature extraction. The basic structure of RBM is shown in an unsupervised learning layer in Figure 1. As an energy based model (Hinton et al., 2006), 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 realvalued 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.

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.

Considering the weak learner with STPN in interpretation of causality in distributed complex systems, RBM is applied as a *boosting approach* to form a strong method in learning characteristics in abundant STPNs. Also, RBM is applied in this work with the purpose of capturing multiple operating modes in distributed complex systems.

## III. PROPOSED METHODOLOGY

### A. Proposed Framework

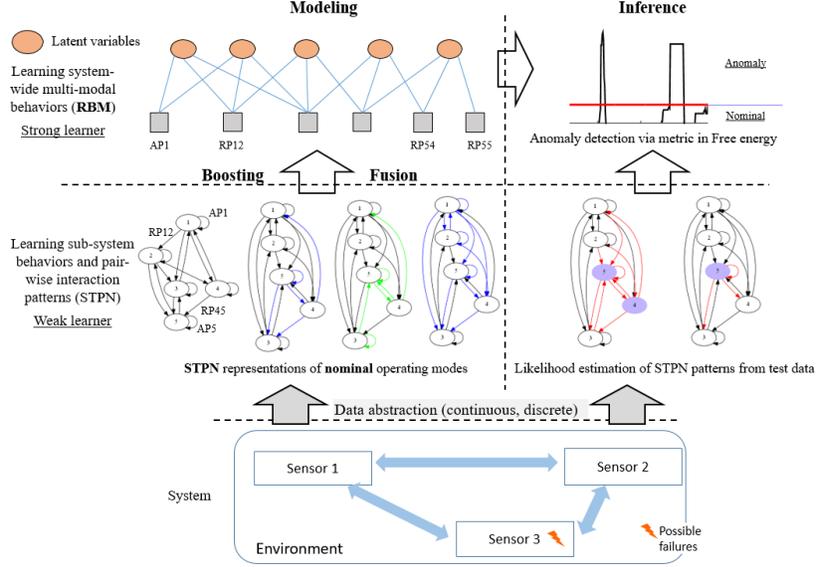
The proposed data-driven framework for system-wide anomaly detection is shown in Figure 1. The general idea is to form an unsupervised spatiotemporal graphical modelling approach that can recognize the system-wide anomalous patterns in distributed complex systems and detect anomaly only with nominal data.

As cycles are included in STPNs, causality reasoning is more difficult. Therefore, the discovered graphs in STPN, in representation of continuous time series and discrete event logs, are treated as *weak learners*. Then, RBM is applied as a boosting approach to learn multiple graphs (discovered by STPN with an online approach) for a more accurate *strong learner*. Moreover, multiple nominal operating modes are typically present in distributed complex systems, and as such, characteristics of graphical models in these modes (represented by different STPNs) can be effectively captured by RBM. With unsupervised learning, the trained RBM is used to detect anomaly via identifying a low probability event (in representation of free energy), based on the assumption that the anomaly influences the causality and induces different patterns in STPN. The process of discovering graphs with STPN, representing nominal conditions with RBM, and detecting anomaly via unsupervised learning is noted as STPN+RBM model. The steps of learning the STPN+RBM model are:

(1) Learn APs and RPs (individual node behaviors and pairwise 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  $\Lambda^{ij}$ , 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 behaviour with nodes in the visible layer corresponding to APs and RPs,
- (5) The RBM is trained using binary vectors generated from training subsequences,
- (6) Online anomaly detection is implemented via the anomaly metric obtained with the trained RBM.



**Figure 1** Methodology: A data driven framework for anomaly detection.

## B. Metric for anomaly detection

In order to train the system-wide RBM, the causal metrics ( $\Lambda^{ab}$ ) 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 (Liu et al., 2016). Although in this paper we convert the  $\Lambda^{ij}$  metrics to binary values for the ease of RBM training, it is not mandatory for this training process. 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. The energy function for an RBM is defined as:

$$E(\mathbf{v}, \mathbf{h}) = -\mathbf{h}^T \mathbf{W} \mathbf{v} - \mathbf{b}^T \mathbf{v} - \mathbf{c}^T \mathbf{h} \quad (2)$$

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  $\mathbf{v}$ :

$$e^{-F(\mathbf{v})} = \sum_{\mathbf{h}} e^{-E(\mathbf{v}, \mathbf{h})} \quad (3)$$

Another expression for free energy estimation is:

$$F(\mathbf{v}) = -\sum_i v_i a_i - \sum_j \log(1 + e^{b_j + \sum_i v_i w_{ij}}) \quad (4)$$

By applying the concept of *free energy* of RBM, the STPN+RBM model is an energy based probabilistic graphical model. This is the basic idea of applying STPN+RBM model for anomaly detection.

The anomaly metric is defined based on the free energy of the system; the higher is the free energy, higher is the probability of anomaly.

**Anomaly Detection Process:** With the defined free energy, a baseline is defined such that 95% of the nominal data have their free energy lower than the baseline.

The baseline is applied as the anomaly detection threshold in this work. In the case that free energy detected is higher than the threshold, the condition will be detected and labeled as an anomaly.

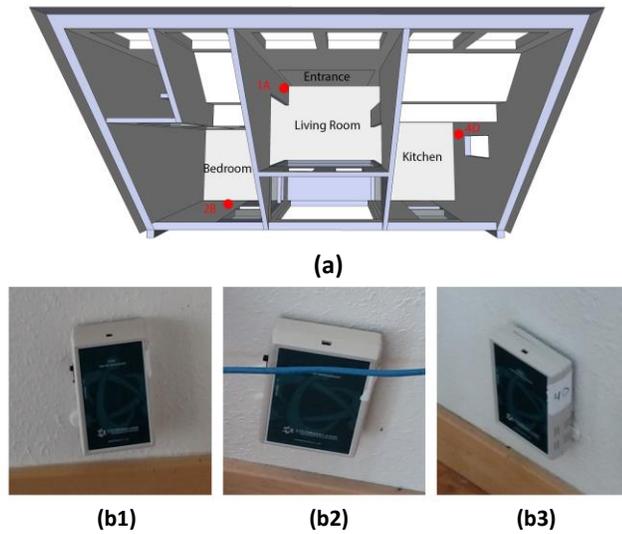
## IV. EXPERIMENTAL SETUP, RESULTS AND DISCUSSIONS

### A. EXPERIMENT DESCRIPTION

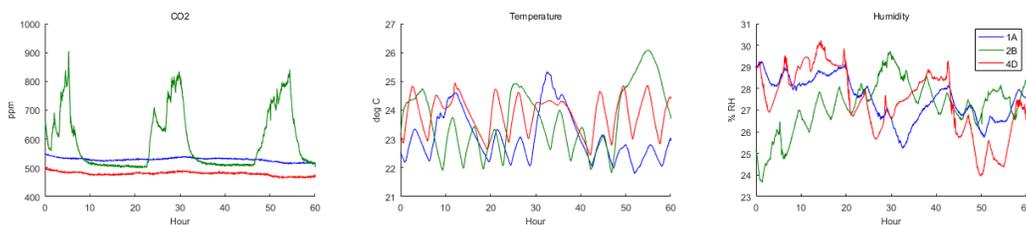
A 70m<sup>2</sup> NSF community lab called Interlock House is chosen to carry out the experiments. The Interlock House was built for the 2009 DoE Solar Decathlon and has been monitored for more than 4 years with the aid of about 120 sensors installed permanently. A data driven HVAC system control is fed by real time measurements. In the following experiments, temporarily mounted sensors are deployed to validate the algorithm which will be used to improve HVAC system control in the future. A schematic for the Interlock House (top view) with the temporary sensors in place is shown in Figure 2(a).

To obtain data for nominal and anomalous conditions, CO<sub>2</sub> sensors were temporarily placed strategically at three locations inside the living space of the Interlock house as shown in Figure 2(a) and Figure 2(b). Sensors chosen for gathering both nominal and anomaly data are CM-0210 CO<sub>2</sub>, Relative Humidity, Temperature dataloggers manufactured by CO2Meter.com. Meter specifications for the dataloggers are given as follows: It is a non-dispersive infrared (NDIR) sensor with a CO<sub>2</sub> measuring range of 0-10,000 ppm with repeatability of  $\pm 20$  ppm and an accuracy of  $\pm 30$  ppm. The inbuilt temperature sensor has a range of 0-120°C with repeatability of  $\pm 0.1$ °C and an accuracy of  $\pm 0.5$ °C. Inbuilt RH (relative humidity) sensor has a range of 0-100% with repeatability of  $\pm 0.1$ % and an accuracy of  $\pm 3$ %. Nominal and anomalous data were collected on two different occasions, described as stated below:

- a. **Nominal Data:** The sensors were configured to record data at every 1-minute interval for a period of 60 hours for recording nominal data, as shown in Figure 3. During this period, there were two people present in the house and the HVAC system was in operation, thus maintaining the proper air and ventilation requirements of the living space. Fresh air intake was maintained properly and adequately all throughout the data logging phase.
- b. **Anomalous Data:** The anomaly conditions were simulated for a shorter but sufficient period of time (10 hours) as required by our algorithm for detection. Excess CO<sub>2</sub> concentration was simulated by using a CO<sub>2</sub> fire extinguisher which was used from time to time to increase the CO<sub>2</sub> levels inside the living space. Data collected during the anomaly condition is shown in Figure 4. As seen in the figure, the faults were injected primarily at time instances of 2, 5, 7 and 8 hours. The CO<sub>2</sub> injection was carried out nearer to the sensors 2B and 4D than sensor 1A, to study how effectively CO<sub>2</sub> is able to dissipate through the entire volume of the space inside the Interlock house; resulting in sensor 1A registering very few cases as anomalous. During the data collection period, there were also two people present inside the living space and in addition, the HVAC system maintaining the ventilation and air-conditioning was turned off to simulate extreme conditions.



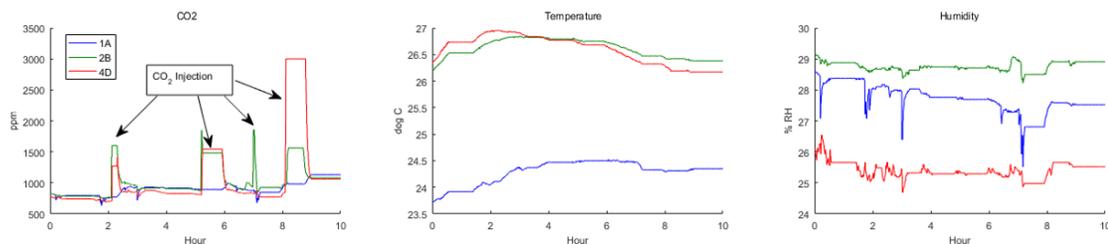
**Figure 2** Location of the sensors (CM-0210 CO<sub>2</sub>, Temperature, Humidity dataloggers) inside house: **(a)** Schematic showing sensor locations inside the Interlock House (top view); **(b1)** Sensor 1A located near the entrance; **(b2)** Sensor 2B located in the bedroom; **(b3)** Sensor 4D located in the kitchen space. The collected nominal data is shown in Figure 3.



**Figure 3** Data collected in nominal condition. Sensors 1A, 2B, and 4D correspond to the locations label in Figure 2.

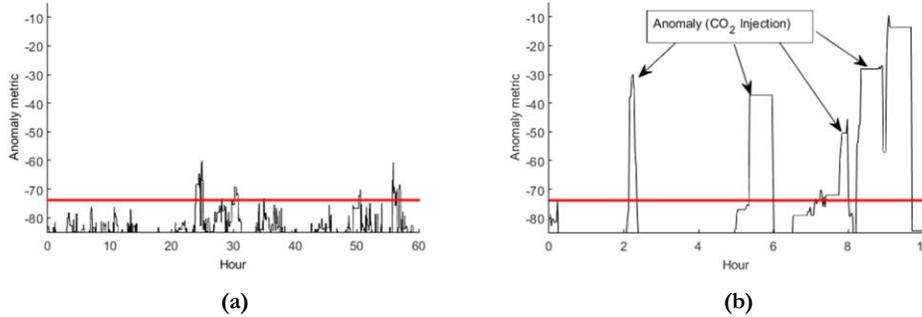
## B. RESULTS AND DISCUSSIONS

Using the proposed framework, the model is trained with nominal data, and the anomaly metric in nominal data is shown in Figure 5(a). In the nominal condition, although there are fluctuations, the model detects most of them as nominal correctly. It validates the robustness of the proposed framework. The abrupt change in humidity around 24th and 56th hours cause the increase in the anomaly metric, and the anomaly decreases in a short time.

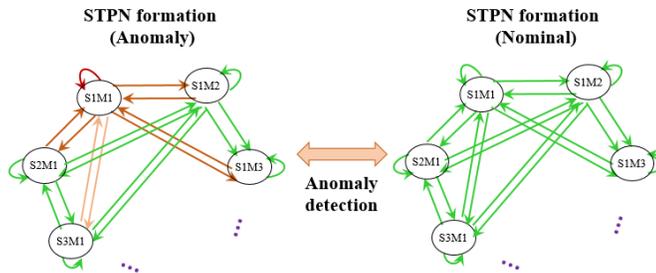


**Figure 4** Data collected in anomalous condition, where CO<sub>2</sub> is injected at specific time instances to simulate faults in the system. Sensors 1A, 2B, and 4D correspond to the locations label in Figure 2.

Based on the threshold built in the nominal data, anomaly detection is carried out and the results are shown in Figure 5(b). The anomalies detected (with free energy above the baseline) are consistent with the injected faults as shown in Figure 4. Therefore, the proposed framework can capture the anomaly correctly.



**Figure 5** (a) Free energy estimated with trained STPN+RBM model in nominal condition. The baseline in red is computed by detecting 95% cases as nominal (the remaining 5% are noted as outliers); (b) Free energy estimated with trained STPN+RBM model in anomalous condition. The baseline obtained in Figure 6(a), and can be applied as a criterion for anomaly detection.



**Figure 6** Representing fault diagnosis after anomaly detection with STPN. (S refers to a sensor and M refers to a modality, where modality refers to either of humidity, temperature or CO<sub>2</sub> concentration) i.e. there are 3 modalities within each of the 3 sensors that were temporarily installed in the Interlock House

In Figure 6, interconnectivity between the STPN nodes is shown. Each node corresponds a sensor of a certain modality (e.g., temperature, humidity and CO<sub>2</sub> concentration). Hence, there are 9 nodes in total (3 sensors each for the 3 modalities considered in this case study). When conditions are nominal the links between the nodes are shown using green arrows to signify no-fault scenarios. When a fault arises in any of the nodes, as shown in the Anomaly scenario in Figure 6, causal relationships between the faulty node and other nodes break down. In this example, S1M1 (Sensor 1 of Modality 1) is such a node that represents CO<sub>2</sub> concentration at a certain location and thus, source of a detected anomaly can be isolated using the proposed technique.

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 complex system with continuous temporal information from sensors and actuators. The proposed framework can effectively capture multiple nominal modes with a unified model, which greatly reduces the complexity of the modeling. The approach can correctly detect the anomaly and is also seen to be quite robust as it is designed to identify only persistent anomalies. After detection, it can also isolate and diagnose the root-cause as shown in Figure 6.

## V. CONCLUSION AND FUTURE WORK

As is evident from the results, the method of anomaly detection discussed in this paper can easily capture faults if and when they occur in a complex system, such as a HVAC. Anomalies in distributed complex systems vary in mechanisms, feature and duration, which makes anomaly detection difficult, especially when there are multiple faults

in the system. This method takes care of these multi-fault scenarios and comes up with a single metric to showcase such fault for the controller to take appropriate steps as a corrective measure.

In the case study presented in this paper, we have worked with CO<sub>2</sub> as the primary source of air contaminant, primarily because of safety concerns for the people living inside the Interlock House. This method can be extended to multiple conditions where there might be other contaminants or potential health hazard substances like CO, NO<sub>2</sub> and/or excess SPM present and address a wide range of nominal operating modes and unforeseen anomalous situations without comprehensive labeled training data in distributed systems.

The results show that the proposed framework can capture multiple diverse nominal modes (CO<sub>2</sub> level, temperature and humidity) within a single probabilistic graphical model, and detect anomalies by identifying low probability events. This case study validates the accuracy and robustness of the proposed method.

Future work will include capturing and detecting multiple fault scenarios, including SPM measurements along with presence of another pollutant gas, either CO or NO<sub>2</sub> with simulated and real fault data, using the graphical model for root-cause analysis for different anomalies, and detection of simultaneous multiple faults in complex HVAC systems.

## ACKNOWLEDGMENTS

The authors would like to acknowledge Evan Jeanblanc from Iowa State University Center for Building Energy Research, Hannah Wiltamuth and Jacob Ahee from Iowa Department of Natural Resource. This paper is based upon research partially supported by the National Science Foundation under Grant No. CNS-1464279. The authors would like to express their sincere gratitude to Mike Wassmer from Live to Zero LLC for helping in data acquisition and providing access to the Interlock House that is a part of the Iowa NSF EPSCOR project funded under grant number EPS-1101284

## REFERENCES

- Liu, C., Ghosal, S., Jiang, Z. and S. Sarkar. 2016. An unsupervised spatiotemporal graphical modeling approach to anomaly detection in distributed cps. *Proceedings of the International Conference of Cyber-Physical Systems, (Vienna, Austria)*.
- Klepeis, Neil E., et al. 2001. The National Human Activity Pattern Survey (NHAPS): a resource for assessing exposure to environmental pollutants. *Journal of exposure analysis and environmental epidemiology* 11.3: 231-252.
- Choi, A., Zheng, L., Darwiche, A. and O. Mengshoel. 2011. A tutorial on bayesian networks for system health Management. *Machine Learning and Knowledge Discovery for Engineering Systems Health Management*, vol. 10.1: 1–29.
- Rao, C., Ray, A., Sarkar, S. and M. Yasar. 2009. Review and comparative evaluation of symbolic dynamic filtering for detection of anomaly patterns. *Signal, Image and Video Processing*, vol. 3, no. 2: 101–114.
- Jiang, Z. and S. Sarkar. 2015. Understanding wind turbine turbine interactions using spatiotemporal pattern network. in *Proceedings of ASME Dynamics Systems and Control Conference*.
- Fiore, U., Palmieri, F., Castiglione, A. and A. De Santis. 2013. Network anomaly detection with the restricted boltzmann machine. *Neurocomputing*, vol. 122: 13–23.
- Sarkar, S., Sarkar, S., Virani, N., Ray, A. and M. Yasar. 2014. Sensor fusion for fault detection and classification in distributed physical processes. *Frontiers in Robotics and AI*, vol. 1: 16.
- Sarkar, S. August 2015. Hierarchical symbolic perception in dynamic data driven application systems. Ph.D. dissertation, The Pennsylvania State University.
- Sarkar, S., Srivastav, A. and M. Shashanka. 2013. Maximally bijective discretization for data-driven modeling of complex systems. *American Control Conference (ACC)*. IEEE: 2674–2679.
- Sarkar, S. and A. Srivastav. 2016. A composite discretization scheme for symbolic identification of complex systems. *Signal Processing*, vol. 125: 156 – 170
- Bang-Jensen, J. and G. Z. Gutin. 2008. *Digraphs: theory, algorithms and applications*. Springer Science & Business Media.
- Hinton, G. and R. Salakhutdinov. 2006. Reducing the dimensionality of data with neural networks. *Science*, vol. 313.5786: 504–507.
- Roux, N.L. and Y. Bengio. 2008. Representational power of restricted boltzmann machines and deep belief Networks. *Neural Computation*, vol. 20.6: 1631–1649.