Distributed Event Identification for WSNs in Non-Stationary Environments

K. Ali[2], S.B. Ali[4], I.H. Naqvi[1], M.A. Lodhi[3]

[1]Department of Electrical Engineering, LUMS Syed Babar Ali School of Science and Engineering, Pakistan
[2]Department of Computer Science and Engineering, Michigan State University, USA
[3]Department of Electrical and Computer Engineering, Rutgers, USA
[4]TU München, Germany
Email: alikamr3@msu.edu, syedbilal.ali@tum.de, ijaznaqvi@lums.edu.pk, masad.lodhi@rutgers.edu

***Abstract*—This paper proposes a novel scheme to estimate the percentage contribution of different attributes in a detected event (a process termed as event identification) for streaming multi-attribute data in WSNs. The proposed event detection and identification algorithm takes into account correlation among sensed attributes as well as the spatio-temporal correlations with similar attributes measured by neighboring nodes. Moreover we update our statistical parameters in an iterative manner such that the dynamics of non-stationary environments are taken into account. We test our leave one out (LOO) event identification approach with simulations on both synthetic and real data sets and an implementation on off-the-shelf WizziMotes. The experimental results show that our detection scheme outperforms state of the art schemes by showing detection rates (DRs) of more than 98% and false positive rates (FPRs) of less that 2%. Moreover, our event identification approach effectively determines the contribution of both correlated and uncorrelated attributes in an event of interest. The identification has also been shown to be in strong agreement with previous computationally complex benchmark PCA based event identification approaches.**

***Keywords*: Outlier detection, event detection, event identification, non-stationary environment, clustering, wireless sensor networks.**

I. INTRODUCTION

A Wireless Sensor Network (WSN) is comprised of low power and memory constrained sensor nodes. Each node measures different environmental attributes (like acceleration, temperature, humidity, etc.), processes that information and transmits important inferences to the base station. WSNs are being used in a variety of real world applications related to industrial, business, and military domains, such as environmental and habitat monitoring, object and inventory tracking, elderly patients' monitoring, battlefield observations, industrial safety and control etc. [1][2]. Since these measurements form the basis of event detection (or an alarm), it is important to correctly detect readings that deviate from the normal data.

In order to ensure ambiguity free decisions, outlier detection techniques are deployed. Outliers are sensed data measurements that significantly deviate from the normal pattern [1]. The application of outlier detection in filtering false data, finding faulty sensor nodes [3], malicious attacks and identification of events of interest has attracted significant attention of the research community in recent years. A robust outlier detection scheme takes into account the temporal correlation between different attributes at a single node and the spatial correlation between similar attributes among neighbouring nodes. An appropriate outlier detection technique is unsupervised, on-line (avoids storage and processing of batches of sensor readings) and distributed (minimizes communication overhead). Classification based algorithms like Support Vector Machines (SVMs) have been extensively used for this purpose due to their ability to incorporate spatio-temporal and attribute correlations of data [4]. However high computational cost limits their applicability to environments with strict latency requirements. Extensive surveys on characterization and classification of outlier detection techniques [1][5][6][7] suggest that clustering based algorithms do not require prior knowledge of data distribution, can be used in an incremental model, are computationally inexpensive and can achieve high detection and low false positive rates.

Hyperellipsoidal clustering has been regarded as one of the popular clustering based methods for outlier detection in (WSNs). The algorithm presented in [8] lays a mathematical foundation for estimating a hyperellipsoidal boundary between normal data and outliers. A modified version of [8] has been proposed in [9], namely Iterative Data Capture Anomaly Detection (IDCAD), which incrementally updates the elliptical boundary and is capable of supporting outlier detection on streaming data in WSNs. Although these techniques are very effective for outlier detection, little work has been done on making these clustering techniques suitable for *dynamic* or *non-stationary* environments. The data distribution in non-stationary environments can contain changing trends or periodicity. The application of outlier detection techniques on this raw data leads to an increase in False Positive Rates (FPRs) and degradation of Detection Rates (DRs).

Moreover, hyper-ellipsoid clustering has not yet been used for *Event Detection* and *Identification* despite all its advantages. An *event* can be characterized as an unexpected change in environmental conditions or a hazardous condition for example a fire or gas leakage [5]. Formally, a sequence of outliers that show correlation in time and space correspond to an event. An algorithm for event detection and identification by projecting the hyperellipsoids on to single attribute subspaces has been proposed in [4], but their identification algorithm does not accommodate correlation between multiple attributes and is susceptible to single attribute faults.

This paper introduces an online clustering based joint event detection and identification scheme for streaming multi-attribute data in WSNs. The proposed event detection and identification algorithm takes into account correlation among sensed attributes as well as the spatio-temporal correlations with similar attributes measured by neighboring nodes. To update our statistical parameters, we employ an iterative approach which is capable of efficiently tracking the dynamics of non-stationary environments. Our iterative outlier detection technique gives very low False Positive Rates (FPRs) and high Detection Rates (DRs) when compared to previously proposed on-line FFIDCAD and Effective-N techniques [9]. All three methods have been applied on the data sets used in [9] for a fair comparison. Our leave one out (LOO) event identification approach quantitatively determines the contribution of individual attributes in an event of interest. The algorithms are simulated on both synthetic and real data sets and are implemented on CC430 based WizziMotes [10] running DASH 7 protocol [11].

The rest of the paper is organized as follows. Section II discusses the our outlier detection scheme and its extension to non stationary environments. Section III explains the event detection procedure. The proposed leave one out event identification algorithm has been presented in Section III-A. The results on synthetic and real data sets along with the hardware implementation of the proposed method have been presented and discussed in Section IV. Finally, Section V concludes this paper.

II. On-line outlier detection: Prior Art and the Proposed Scheme

Consider a WSN collecting streaming multi-attribute data and let X=$\{\mathbf{x}_1, \mathbf{x}_2, \cdots \mathbf{x}_k\}$ be the first k samples of data collected at a node; where each sample $\mathbf{x}_i$ is a $d \times 1$ vector in $\Re^d$. Each element in the vector represents an attribute a_j for example temperature, pressure etc. Let the arithmetic mean, the second moment, and the covariance of the incoming on-line data be m_X, m_{X^2} and S_X respectively. During the outlier detection phase, the hyper-ellipsoid clustering algorithm encapsulates the data samples by using the following equations:

$$e_X(m_X, S_X^{-1}, t) = \{x_i \epsilon \Re^d | \underbrace{\sqrt{(x_i - m_X)^T {S_X}^{-1} (x_i - m_X)}}_{D_i = Mahalanobis\ \ distance\ \ of\ \ x_i} \leq t\} \tag{1}$$

$$e_X(m_X, m_{X^2}, t) = \{x_i \epsilon \Re^d | \underbrace{\sqrt{(x_i - m_X)^T (m_{X^2} - (m_X\, m_X^T))^{-1} (x_i - m_X)}}_{D_i = Mahalanobis\ \ distance\ \ of\ \ x_i} \leq t\} \tag{2}$$

where e_X is the set of data samples whose Mahalanobis distance from the mean of X, $D_i \leq t$ and t is the effective radius of the hyper-ellipsoid [12]. It is known that if t is chosen such that $t^2 = ({\chi_d}^2)^{-1}_{0.98}$ [9], it encapsulates up to 98% of the data points. The data samples $\mathbf{x}_i$ for whom $D_i > t$ are not enclosed in the hyper-ellipsoidal boundary are identified as outliers. Eq. (2) represents e_X entirely in terms of means thus making it possible to write Eq. 1 in an *incremental* or *on-line* form. For a newly arrived data sample $\mathbf{x}_{k+1}$, we propose that $m_{X,k+1}$ and $m_{X^2,k+1}$ can be updated in the following manner:

$$m_{X,k+1} = \frac{k m_X + \mathbf{x}_{k+1}}{k+1} \tag{3}$$

$$m_{X^2,k+1} = \frac{k m_{X^2} + \mathbf{x}_{k+1}\mathbf{x}_{k+1}^T}{k+1} \tag{4}$$

Eqs. (3) and (4) can be used to update the hyperellipsoidal boundary and the set of normal data points e_X after the newly arrived measurement has been received. If $\mathbf{x}_{k+1}$ does not belong to the updated e_X then it is declared an outlier. It is beneficial to update means instead of covariance S_X because incrementing means requires lower complexity and achieves higher accuracy as compared to the iterative update of covariance matrix.

A. Clustering in Non Stationary Environments

Non stationarities like environmental trends and random noise cause variations in the sensor measurement readings and these variations are responsible for poor detection and higher false positive rates of outlier detection algorithms. We present an effective on-line technique for updating the mean of incoming data which makes use of *independent forgetting factors* (IFF), leading to dramatic improvement in both the FPR and DR. The proposed approach uses different *forgetting factors* for both the previous mean $M_{X,k}$ and the newly readings X_{k+1} and takes into account the deviation of latest readings with respect to previous mean:

$$X_{k+1}^{'} = X_{k+1} - M_{X,k} \tag{5}$$

$$M_{X,k+1} = \lambda_m M_{X,k} + (1 - \lambda_n) X_{k+1}^{'} \tag{6}$$

where $X_{k+1}^{'}$ is a normalized version of the newly arrived data sample $X_{k+1}, M_{X,k}$ is the mean updated on the previous iteration and $M_{X,k+1}$ is the mean updated using independent forgetting factors λ_m and λ_n. Although the proposed IFF based technique has been designed for non-stationary environments, it works well even for stationary environments. In other words, $M_{X,k+1}$ and $X_{k+1}^{'}$ remain bounded in case of a stationary environment ($X_k = X_0$, $\forall k$). After some mathematical analysis, the updated mean at the k^{th} iteration can be written as:

$$M_{X,k+1} = X_0 \left[1 + \epsilon \sum_{i=0}^{k} \lambda_m^i \right] \tag{7}$$

$$X_{k+1}^{'} = -\epsilon X_0 \left[\sum_{i=0}^{k} \lambda_m^i \right] \tag{8}$$

where $\epsilon = \lambda_m - \lambda_n$. For large values of k, $X^{'}_{k+1}$ reduces to:

$$\lim_{k \longrightarrow \infty} X^{'}_{k+1} = \left[\frac{-\epsilon X_0}{1 - \lambda_m} \right] \tag{9}$$

which is a constant and thus proves boundedness under stationary environments. In order to give more weight to the newly arrived data sample, λ_n should be less than λ_m. It would help to effectively track the changes in the environment. We have chosen $\lambda_m = 0.9$ and $\lambda_n = 0.84$ in our simulations and experiments.

We observe that $X^{'}_{k+1}$ is a measure of how the data is varying with respect to its mean. This technique combined with our outlier detection technique tracks the distribution of sensed data such that the variations (e.g. *noise, trends, periodicity*) are bounded tightly by the incremental hyper-ellipsoids without affecting weights of the outliers which leads to a remarkable minimization of FPRs and increase in DRs. The technique has been applied on *three* different real data sets and the results are compared with the results of IDCAD, FFIDCAD and Effective-N approaches presented in [9].

III. Event Detection and Identification

An event is defined as a sequence of outliers correlated in both time and space. To detect an event, we store the outliers that have been detected via the iterative outlier detection algorithm. Let k be the most recent iteration at node Q, and let o_{k+1} be the first detected outlier. We define an event array E_Q to be an array of outliers that fulfills the following conditions:

1) The outlier o_{k+1} detected at $k{+}1^{th}$ iteration is pushed into E_Q if and only if E_Q is empty or there is a difference of less than T iterations between o_{k+1} and the last entry of E_k.
2) If no outlier is detected for T iterations, E_Q is emptied.
3) If the number of outliers in E_Q equals n_0, a local event is declared at node Q.

These conditions ensure that the outliers based on which an event is declared are not too apart in time i.e. they are temporally correlated, and that they are significant in number. Once a local event is declared, we check if similar events occurred at the surrounding nodes. The similarity is found by clustering E_Q at each node and calculating the Bhattacharya similarity coefficient S [13][14] between clusters:

$$S = e^{-\frac{1}{8}(m_1 - m_2)^T (\frac{V_1 + V_2}{2})^{-1}(m_1 - m_2)} \times \sqrt{\frac{\sqrt{det(V_1)det(V_2)}}{det(\frac{V_1+V_2}{2})}} \tag{10}$$

where m_1, m_2 are means and V_1, V_2 are the covariance matrices of hyper ellipsoid cluster at individual nodes, and S is their similarity on a scale of 0 to 1. An event is declared if S between node Q and its surrounding nodes is greater than some threshold (e.g. 0.5).

A. *Leave One Out (LOO) Event Identification algorithm*

The proposed *leave one out* (LOO) event identification algorithm projects the hyper-ellipsoid (containing the data set e_X) on to a subspace X^q which contains all but one (q^{th}) attributes. Hence if X is a d dimensional subspace, X^q is a $(d{-}1)$ dimensional subspace. As attributes are often correlated, a subspace of multiple attributes ensures that this correlation is incorporated towards the decision. It is also more robust to errors that might occur in the measurement of individual attributes. LOO algorithm works in three steps:

STEP 1: Defining the Projection: We begin with projecting E_Q, its covariance S_E and mean m_E onto the subspace X^q. Let $E_Q = \{o_1, o_2, \cdots o_{n_0}\}$ where o_i is a $d{\times}1$ data sample identified as an outlier. The projection of E_Q onto q^{th} $d-1$ dimensional subspace is given as:

$$E^q_{loo} = P^q_{loo} E_Q, \quad \forall q = 1 \cdots d$$

where $P^q_{loo} \in \Re^{d \times d}$ is a diagonal matrix which has the q^{th} diagonal element equal to zero and all other diagonal elements equal to one. Similarly S_E (a $d \times d$ covariance matrix) and m_E (a $d{\times}1$ column vector representing the mean of E_Q), can be projected on to X^q to get S^E_{loo} and m^E_{loo} respectively.

STEP 2: Mahalanobis Distance of the Projections: After finding out the above projections, the Mahalanobis distances from the projetected means, D^q, can be found through Eq. (11):

$$D^q_i \left(m^E_{loo}, S^E_{loo}, t\right) = \sqrt{(o^q_i - m^E_{loo})^T \; S^E_{loo\,-1} \, (o^q_i - m^E_{loo})} \tag{11}$$

where $o^q_i, m^E_{loo}, S^E_{loo}$ are the projections of outlier o_i, m_E and S_E respectively, on the subspace X^q. For a particular value of q, D^q is an array of the Mahalanobis distances of each column in E^q_{loo} for example $D^1 = \{D^1_1, D^1_2, \cdots D^1_{n_0}\}$.

STEP 3: Decision Formulation: The event identification decisions are carried out through a decision arrays. For q^{th} attribute, the decision array is defined as:

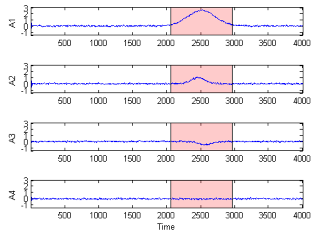
$$R^q = \sqrt{(D_i)^2 - (D^q_i)^2} \quad \forall i = 1, 2, \cdots, n_0 \tag{12}$$

The elements of R^q signify the contribution of attribute q towards each of the n_0 elements in E_Q. Mean value of this array, $\overline{R^q}$, gives a measure of the average contribution of q^{th} attribute towards the event. The attribute that contributes most towards the event is given by:

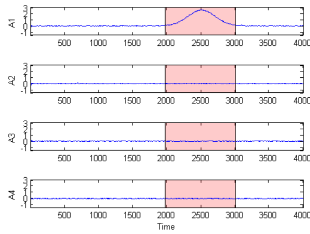
$$max(\overline{R^q}) \;\text{ for }\; q = 1, 2, \cdots d \tag{13}$$

IV. Simulation Results and Analysis

The performance evaluation of outlier detection, event detection and event identification has been performed on both synthetic and real data sets.



(a) Synthetic correlated data-set



(b) Synthetic uncorrelated data-set

Fig. 1. All 4 dimensions of synthetic data set plotted separately.

TABLE I. COMPARISON OF CLUSTERING TECHNIQUES (SYNTHETIC DATA SET)

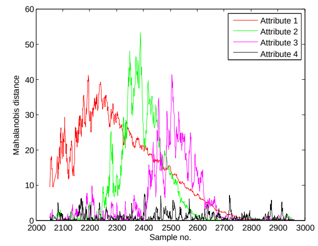
Clustring Technique	DR	FPR
DCAD (batch clustering)	67.4%	0.18%
IDCAD (online) [9]	95.89%	6.4%
FFIDCAD (online) [9]	96.77%	7.05%
Effective-N (online) [9]	97.8%	7.05%
Purposed online approach	98.97%	6.3%

A. Synthetic Data Set

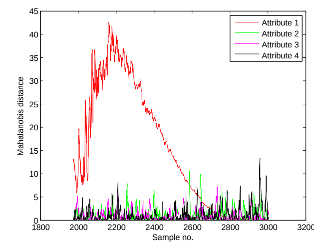
The synthetic data consisted of two normally distributed 4-dimensional data sets consisting of 4000 samples each. Each dimension, corresponding to a single attribute, is plotted with respect to time in Figure 1. In the first data set, an event is introduced in the 1^{st} attribute centered at the 2500^{th} sample. The second and third attribute are correlated to the first attribute and hence show some variations in the same interval when the event occurs. The 4^{th} attribute is uncorrelated to the other three. In the second data-set (see Figure 1b), all attributes are independent and an event occurs only in the 1^{st} attribute.

1) Outlier Detection: Three outlier detection techniques have been tested on the Dataset:

- *Cumulative clustering or DCAD [8]:* In DCAD outlier detection, all samples until the current (say n^{th}) sample are clustered and the Mahalanobis distance for the n^{th} sample is calculated using Eq. (1). The drawback of this technique is that it requires storage of all n samples which may not be suitable for memory constrained nodes.
- *FFIDCAD [9]:* In this scheme. the covariance matrix is estimated from the covariance matrix of the previous iteration and substituted into Eq. (1) to make the decision. A loss of precision occurs during the estimation of the covariance matrix and results in a relatively degraded DR and FPR performance in comparison with the proposed technique.
- *Proposed IFF based Clustering:* The proposed outlier detection scheme makes use of independent forgetting factors which makes it robust to non stationary environments. The results for FPR and DR for the three mentioned techniques are summarized in Table I. The results show that the proposed IFF based outlier detection outperforms other on-line schemes.

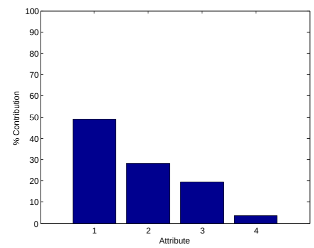


(a) Correlated Dataset

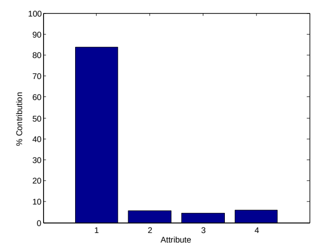


(b) Independent Dataset

Fig. 2. Contribution of each attribute over the duration of the event



(a) Correlated Dataset



(b) Independent Dataset

Fig. 3. The bar graphs represent the overall percentage contribution of each attribute towards the event

2) Event Detection and Identification: The Event Detection algorithm mentioned in section III has been applied to this synthetic data set. Keeping $T = 5$ and $n_o = 20$, an event gets recorded from the 2010^{th} sample to the 2700^{th} sample as indicated by the shaded region in Figure 1. Figure 2 shows the decision array R^q for $q = 1, 2, 3, 4$ for both correlated and independent data sets respectively. The contributions of each attribute over the duration of the event is shown. For correlated data set, the event is initiated in attribute 1 which solely contributes towards the event till the 2250^{th} sample after which the effect of attribute 1 slowly dies down and attribute 2 and 3 contribute significantly towards the event. Attribute 4 does not make a significant contribution towards the event throughout the duration of the event. $\overline{R^q}$ which is the normalized mean contribution towards the event by each attribute is represented by Figure 3. For the correlated data set, attribute 4 contributes the least whereas attribute 1 contributes the most towards the event followed by the contributions of attributes 2 and 3. For the independent data set, only attribute 1 shows a significant contribution towards the event.

B. Real Data Sets

We tested our algorithms on $three$ real data sets. First data set was taken from a multi-hop WSN deployment using TelosB motes. It consists of humidity and temperature measurements taken over a 6 hour period at 5 second intervals [15]. The other two data sets, as used by the authors of [9], have been chosen for comparison of our IFF based technique with FFIDCAD and Effective-N. First among them contains the data from epochs 25000 to 30000 of node 18 of the IBRL (Intel Berkeley Research Lab) dataset [16]. The second data set contains the

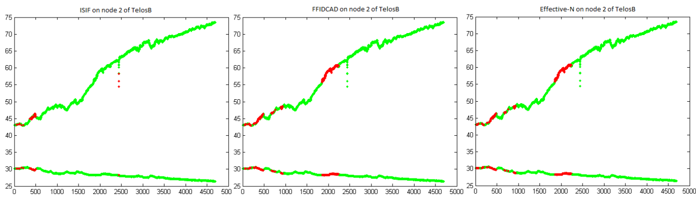


Fig. 4. Outlier Detection algorithms (IFF based clustering, FFIDCAD and Effective-N from Left to Right) applied on TelosB Node 2 (Time Series)

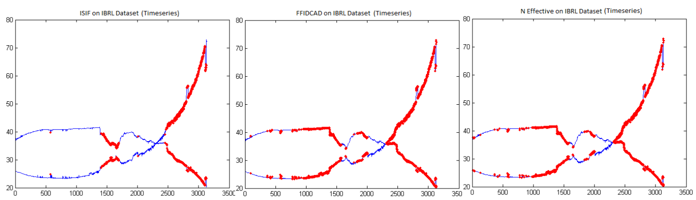


Fig. 5. Outlier Detection algorithms (IFF based clustering, FFIDCAD and Effective-N from Left to Right) applied on IBRL Data Set (Time Series)

TABLE II. COMPARISON OF CLUSTERING TECHNIQUES (REAL DATA SET (TELOSB NODE 1))

Clustring Technique	DR	FPR
DCAD (batch clustering)	81.82%	0%
IDCAD (online) [9]	94.12%	1.69%
FFIDCAD (online) [9]	88.64%	1.625%
Effective-N (online) [9]	88.64%	1.538%
Purposed online approach	95.45%	2.24%

data collected at station 10 of the network deployed on the rock glacier at Le Genepi in Switzerland in a 12 days collection period starting from October 10^{th} 2007 [17].

IFF based clustering, FFIDCAD and Effective-N have been simulated on all of the above data sets. For IFF based technique, $\lambda_m = 0.9$ and $\lambda_n = 0.84$ have been used in Eq.(6) throughout all simulations. Figures 4 and 5 show the outliers (shown in red) detected in TelosB (Node 2) and IBRL respectively. From left to right, the outlier detection schemes are IFF, FFIDCAD and Effective N. According to [15] node 2 did not contain any anomaly. It is evident that the proposed IFF based technique (left most) gives significantly smaller false positives as compared to FFIDCAD and Effective-N especially in the IBRL and TelosB results in Figures 4 and 5. Tables II and III provide a comparison of DRs and FPRs for all of the schemes (FFIDCAD, Effective N and IFF based clustering) on real data sets. The tables provide great insight on the performance of these schemes and can be used for quantitative comparison. The results show that the proposed scheme can achieve 7.68 % better DR than both FFIDCAD and Effective N schemes. The DR of the proposed scheme is significantly better than the batch clustering methods. The results given in Table III confirms that IFF based outlier detection scheme is better than the rest of the schemes as it gives the highest DR (100 %) and the lowest FPR (1.17 %) amongst online schemes.

The proposed event detection and identification algorithms have also been applied on Nodes 1 and 3 of TelosB data set. The simulations have been performed using $T = 5$ and $n_o = 20$. Figs. 6 & 7 show the three steps involved in event identification i.e. outlier detection, event detection and event identification. The results of our identification approach

TABLE III. COMPARISON OF CLUSTERING TECHNIQUES (REAL DATA SET (TELOSB NODE 3))

Clustring Technique	DR	FPR
DCAD (batch clustering)	97.65%	0%
IDCAD (online) [9]	94.12%	1.96%
FFIDCAD (online) [9]	95.29%	2.07%
Effective-N (online) [9]	95.29%	1.86%
Purposed online approach	100%	1.17%

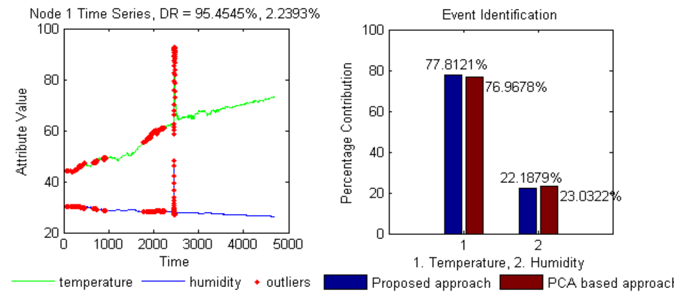


Fig. 6. Outlier Detection, Event Detection and Identification Results for Node 1 of TelosB Dataset

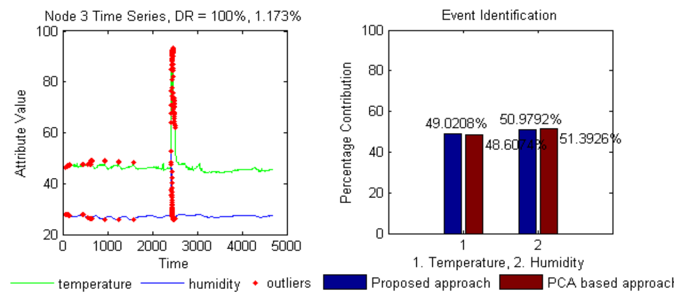


Fig. 7. Outlier Detection, Event Detection and Identification Results for Nodes 3 of TelosB Dataset

are in strong agreement with the computationally complex benchmark principal component analysis (PCA) based event identification method used in [18] as shown in Figs. 6 & 7.

C. Hardware Implementation

For testing our proposed event identification technique we use CC430 based WizziMotes [10] running DASH7 protocol [11]. In order to test event identification on our motes, we require measurements of multiple attributes. The tested attributes are illumination and temperature for both correlated and uncorrelated events. Two sensor nodes have been used for testing purposes. One of them is used as sink (or the cluster head) and other was used as source node. Each node continuously checks the readings for anomalous behavior on the basis of outlier detection algorithm. Once a local event has been declared upon reception of a series of outliers, the child (or source) node sends this information to the sink which checks for an event. If the measurements at the sink nodes also point to a local event, global event is declared. The node then checks for the the individual contributions of the attributes in the detected event using algorithms described in previous sections. The identification ratios are displayed on the PC through a UART interface on the *sink* node.

We have implemented two different scenarios. In the first scenario we test the performance of our algorithms when anomalous attributes are uncorrelated. In order to test that we turn off the lights of the room in which the motes have been placed. The other attribute (temperature) is not affected by the suddenly turning off the lights. The results showed approximately 0.5% and 99.5% contribution for temperature and illumination respectively confirming the effectiveness of the proposed scheme. Thereafter, we test the case where multiple correlated attributes contribute toward an event. Two incandescent bulbs in a close vicinity of both the motes are illuminated such that the bulb was touching the LM35 temperature sensor. It served as a perfect experiment for that purpose as both the light intensity and temperature readings increased in a correlated manner. The results showed an average contribution of 76.34% and 23.66% for illumination and temperature attributes respectively showing that the light intensity and temperature readings increased in a correlated manner. These results show that the proposed scheme gives accurate identification ratios in cases where both independent as well as correlated attributes contribute towards an event. For future work we are considering the implementation of above mentioned algorithms on a WSN with higher node density to analyze the impact of these techniques on the overall lifetime of the network.

ACKNOWLEDGEMENT

This research work has been supported by LUMS Startup/FIF grant. The authors thank Nauman Shahid for discussions on the topic and guidance.

V. CONCLUSION

This paper shows the effectiveness of the leave one out (LOO) event identification scheme to autonomously compute the percentage contributions of different attributes in correlated as well as uncorrelated events. The scheme has been tested on both synthetic and real data sets and implemented on CC430 based WizziMotes running DASH 7 protocol. The results are in strong agreement with the computationally extensive benchmark PCA based methods. The underlying outlier detection scheme makes use of independent forgetting factors and is robust to dynamics of non-stationary environments. We show that our proposed technique achieves detection rates (DRs) close to 100% and false positive rates (FPRs) as low as 1.17%. It outperforms previously proposed methods for trendy environments with more than 8% improvement in detection rate and and 90% improvement in FPRs.

REFERENCES

[1] Y. Zhang, N. Meratnia, and P. Havinga, "Outlier detection techniques for wireless sensor networks: A survey," *Communications Surveys & Tutorials, IEEE*, vol. 12, no. 2, pp. 159–170, 2010.

[2] S. Rajasegarar, C. Leckie, and M. Palaniswami, "Hyperspherical cluster based distributed anomaly detection in wireless sensor networks," *Journal of Parallel and Distributed Computing*, vol. 74, no. 1, pp. 1833–1847, 2014.

[3] A. B. Sharma, L. Golubchik, and R. Govindan, "Sensor faults: Detection methods and prevalence in real-world datasets," *ACM Transactions on Sensor Networks (TOSN)*, vol. 6, no. 3, p. 23, 2010.

[4] N. Shahid, S. Ali, K. Ali, M. Lodhi, O. Usman, and I. Naqvi, "Joint event detection & identification: A clustering based approach for wireless sensor networks," in *IEEE Wireless Communications and Networking Conference (WCNC)*, (Shanghai, P.R. China), pp. 1703–1708, apr 2013.

[5] N. Shahid, I. Naqvi, and S. Qaisar, "Characteristics and classification of outlier detection techniques for wireless sensor networks in harsh environments: a survey," *Artificial Intelligence Review*, pp. 1–36, 2012.

[6] V. J. Hodge and J. Austin, "A survey of outlier detection methodologies," *Artificial Intelligence Review*, vol. 22, no. 2, pp. 85–126, 2004.

[7] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Computing Surveys (CSUR)*, vol. 41, no. 3, p. 15, 2009.

[8] S. Rajasegarar, J. C. Bezdek, C. Leckie, and M. Palaniswami, "Elliptical anomalies in wireless sensor networks," *ACM Transactions on Sensor Networks (TOSN)*, vol. 6, no. 1, p. 7, 2009.

[9] M. Moshtaghi, C. Leckie, S. Karunasekera, J. Bezdek, S. Rajasegarar, and M. Palaniswami, "Incremental elliptical boundary estimation for anomaly detection in wireless sensor networks," in *IEEE International Conference on Data Mining (ICDM)*, pp. 467–476, 2011.

[10] "Wizzi mote." Available: http://www.wizzilab.com/solutions/wizzikit/, 2014.

[11] "Dash 7." Available: http://www.indigresso.com/wiki/, 2014.

[12] M. Moshtaghi, S. Rajasegarar, C. Leckie, and S. Karunasekera, "Anomaly detection by clustering ellipsoids in wireless sensor networks," in *IEEE International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*, pp. 331–336, 2009.

[13] J. Bezdek, T. Havens, J. Keller, C. Leckie, L. Park, M. Palaniswami, and S. Rajasegarar, "Clustering elliptical anomalies in sensor networks," in *IEEE International Conference on Fuzzy Systems (FUZZ)*, pp. 1–8, 2010.

[14] T. Kailath, "The divergence and bhattacharyya distance measures in signal selection," *IEEE Transactions on Communication Technology*, vol. 15, no. 1, pp. 52–60, 1967.

[15] S. Suthaharan, M. Alzahrani, S. Rajasegarar, C. Leckie, and M. Palaniswami, "Labelled data collection for anomaly detection in wireless sensor networks," in *IEEE Intelligent sensors, sensor networks and information processing (ISSNIP)*, pp. 269–274, 2010.

[16] "Ibrl-web." Available: http://db.lcs.mit.edu/labdata/labdata.html, 2006.

[17] "Sensorscope web." Available: http://lcav.epfl.ch/op/edit/sensorscope-en, 2007.

[18] J. Gupchup, A. Terzis, R. Burns, and A. Szalay, "Model-based event detection in wireless sensor networks," *arXiv preprint arXiv:0901.3923*, 2009.