Suma, R, et. al. International Journal of Engineering Research and Applications
www.ijera.com
ISSN: 2248-9622, Vol. 11, Issue 1, (Series-IV) January 2021, pp. 33-42

RESEARCH ARTICLE OPEN ACCESS

Incremental Entropy Component Analysis based Energy Minimization Approach in MANET

Suma, R* and C. Siddaraju**
**(Department of Computer Science, Dr. Ambedkar Institute of Technology, Bangalore, India.*
*** (Professor, Department of Computer Science, Dr. Ambedkar Institute of Technology, Bangalore, India.*

ABSTRACT
Due to the explosive growth of applications in the field of mobile ad hoc networks, and also the amount of data available in various forms, it is the need to develop reliable communication algorithms among the various sensor nodes. It is highly challenging to develop algorithms that are robust against the dynamic movements of the network nodes which results in a varied number of nodes in a clustered network of nodes. Besides, the design of an algorithm is further complicated as the nodes are suffering from retaining their energy due to communication among the nodes to transfer data. In this context, we have attempted to provide a solution based on a multivariate data analysis technique called principal component analysis to eliminate the redundant information that exists within the data itself and hence reduces the data transfer overhead, thereby minimizing the energy. Unlike other methods that use the standard procedure to select the eigenbasis based on the strength of eigenvalues, we explored Renyi entropy component analysis to select the best eigenbasis which helps in reducing the reconstruction error. The proposed method is validated on the dataset which is introduced in other similar recent works and proves the worthiness of the proposed approach. We have integrated the conventional incremental principal component analysis model with the Renyi entropy component analysis to select the best eigenbasis. The comparative analysis is also provided to exhibit the performance of the proposed method.

Keywords **-** Energy Minimization, Incremental Principal Component Analysis, Mobile Adhoc Networks, Renyi Entropy Component Analysis.

Date of Submission: 12-01-2021 Date of Acceptance: 27-01-2021

I. INTRODUCTION

The data transfer among the network of nodes is one of the critical factors in mobile ad-hoc networks (MANET). Several attempts are made by the research community to optimize data transfer among the nodes. The data transfer is a critical factor since the nodes are randomly moving in different directions and the network infrastructure is highly dynamic and random. Achieving accurate and efficient data handling is one of the desired factors in mobile ad-hoc networks and as well to retain the energy among the nodes in the infrastructure-less network. It is a known fact that if the nodes are kept with more idle time, the lifetime of the node will be increased and hence attempts are made by the researchers to develop energy-efficient data transfer techniques suitable to the MANET environment.

An ample amount of research has been developed in recent days to the development of data processing techniques appropriate for sensor networks [9, 31]. The wireless sensor networks (WSN) are known to be constrained by limited resources, in terms of energy, network data throughput, and computational power, and is true in the case of MANET also. The network capacity is inherently limited and hence the communication module is one of the constrained resource [27]. Also, wireless communication is an energy-consuming task and it is identified in many situations as the primary factor of lifetime reduction [1]. The data gathering schemes design which limit the transmitted data is recognized as a dominant issue for wireless sensor networks [9, 24, 31]. Similarly, data aggregation is a precarious factor in MANETs too. MANET although similar to wireless sensor networks on some of the aspects, the way it operates in the layered network architecture makes it different from wireless sensor networks.

In static sensor networks and networks with stable links, data aggregation can be performed by routing along with fixed structures such as trees or network backbones [7, 10, 17, 21]. However, in MANETs, nodes mobility causes design issue as much complex problem and also incur communication overhead complexity due to random movement of nodes in the network topology [14]. Although flooding is one kind of solution which involves broadcasting of data from each node to every other node, the cost of data transfer is of $\boldsymbol{O}(N^2)$,

where N is the number of nodes in the network. Therefore, in this paper, we explore a well-known multivariate data analysis technique called principal component analysis and the selection of principal component vectors is made based on entropy retained in the basis vectors unlike conventional approaches explored for data aggregation/ feature selection/ dimensionality reduction, etc.

The emphasis of this paper is on computing order and duplicate unresponsive data aggregates and sending them to every node in a mobile ad-hoc network (MANET). We are specifically motivated by data aggregation requirements in extremely large scale mobile sensor networks such as networks of UAVs, military networks, the network of mobile robots, and dense vehicular networks, where the number of nodes is often several thousand. In order and duplicate insensitive synopsis, the same data can be aggregated multiple times but the result is unaffected. The MAX, MIN, and BOOLEAN OR are the natural examples of such duplicate insensitive data aggregation.

In the wireless sensor network domain, it is found that data aggregation is one of the best solutions for optimal resource management including energy management in wireless sensor networks. The data aggregation is simply the combination of data that is sensed by various sensor nodes. Data aggregation technique used to decrease the amount of data transmitted to the base station tremendously. Generally, in MANETs, many sensor nodes communicate using wireless links and collaborate with each other. The data composed of each of the nodes interface with the gateway node after carrying out the aggregation of the data by diverse nodes. It is necessary to secure the data collected by such nodes. The security problems such as integrity of data, privacy, and newness in data aggregation become crucial when the MANET is implemented in a militant environment because of sensor node failures. The secure data aggregation schemes are categorized into hop by hop aggregation and end to end aggregation [25].

The problem of data aggregation has been well considered by the research community in the context of static sensor networks. It has been shown in the network aggregation techniques with spanning trees, network backbones are efficient and reliable solutions for the problem. However, in the context of a mobile ad-hoc network, such static routing structures possess instability and could potentially experience a high communication burden for maintenance. Flooding, neighborhood gossip, and spatial gossip are structure-free techniques for data aggregation. The Random walks and their cover times are also addressed to data aggregation problem with different types of static graphs [6, 16]. Different kinds of protocols have been devised to routing packets for facilitating data aggregation. Generally, users require competent aggregate functions. Generally, a sensor network contains a few thousands of low-cost sensors where each node is responsible to act an information-source, sensing, and gathering data from the environment for a given task [22].

Some of the basic issues to be considered for data aggregation are as follows:

- Nodes sense features over the complete network and route to nearby nodes.
- A node can receive different kinds of the same message from several neighboring nodes.
- Communication is usually performed in the aggregate.
- Neighboring nodes report similar data.
- Conglomerate data received from different sources and routes to remove redundancy.

Data aggregation play very foremost role wireless sensor network. In recent times and also in current time researchers designed diverse data aggregation techniques through which the lifetime of the network increased significantly. Some of the techniques are depicted below:

- Cluster-based aggregation
- Tree-based aggregation
- Network aggregation
- Centralized aggregation

Cluster-Based approach: Here, the entire network is separated into different clusters. A Cluster Head is designated in each cluster among different sensor nodes or cluster members. The nodes selected as a Cluster Heads are accountable for the data aggregation process received from group members and then communicate the outcome to the Base Station.

Tree-Based Approach: This approach is apprehended by building an aggregation tree where Base Station is designated as roots and sensor nodes are considered as leaves. Every node has a parent node whose data are forwarded. The flow of data starts from sensor nodes (leaves) up to the Base Station (roots) and the aggregation is done by parent nodes [19].

In-**Network Aggregation,** the following are the approaches:

- With size reduction: each node combines and compresses the data packets obtained from its neighbors to reduce the packet length which will be transmitted towards Base Station.
- Without size reduction: is defined as the process of assimilating data packets received from diverse neighbors into a single data packet. The process of aggregating data packets received from diverse neighbors into a single data packet but unlike with the size reduction process, it is without processing the value of data.

Centralized Approach: In this strategy, the node sends data to a central node via the shortest possible route. These data are aggregated by the central node (header node) to reduce the redundancy.

The collection of data from a MANET is a challenging problem. Due to limited resources such as energy, computational power, data storage, and bandwidth, it is necessary to develop a suitable technical decision while forwarding all the sensed information directly to a sink that does the corresponding processing. However, one need to consider the statistical characteristics.

We consider a agreed group of sensor nodes, processing nodes, and sink nodes with paths leading from the sources to sinks through relay nodes. Generally, the node is battery-powered. Sources gather environmental parameters and form observations, which are forwarded to consumers for auxiliary processing. Two nodes exchange observations if they are within the communication range of each other. Because the communication-range of a node is limited, nodes that are not in the communication range of some sink have their measurements delivered through relays. A relay can also perform sensing or receive data directly from a source, thus injecting such sensed/received data into the incoming observation from a preceding relay. A relay keeps memory and communication resources for each stream of observations. A source reserve resources only for sensing and forwarding.

We study two distributed multivariate contextual information compression mechanisms, namely incremental PCA (i-PCA) and kernel entropy component analysis (k-ECA), which enhance the functionality of the Principal Component-based Context Compression model. Given the resource constraints in a MANET, i-PCA exploits the vector nature of the exchanged observations. The i-PCA makes use of the statistical interdependencies between the contextual parameters (components) for achieving energy efficiency. It aims to reduce the number of the components earlier to transmission once it discovers dependencies among them (learning phase). The compression in i-PCA is accomplished by the projection (prior to transmission) of the observations to a space of reduced dimensions. The aim of i-PCA and k-ECA is to prolong the lifetime of a MANET by improving energy efficiency through optimized compression. The eigenbasis of this space consists of the determined Principal Components (PC). PCs are obtained by the Principal Component Analysis (PCA) [11] method.

Several models exploit preexisting knowledge on MANET topology to operate in an energy-efficient way. In the case of WSN, the model in [15] is based on de Bruijn and Voronoi diagrams for routing and data aggregation. This model requires training to construct the routing tables and, therefore, computational effort. Besides, the model in [26] uses data aggregation for removing data redundancy based on the sensed information attributes. The idea behind the model in Ren et al. [26] is that pieces of data produced by the same type of sensors demonstrate significant redundancy; thus, only data samples with the same attribute can be aggregated. This leads to the dynamic routing of pieces of data corresponding to the same attribute toward the sink. However, the entire knowledge of WSN topology is required to achieve attribute-aware data aggregation. In i-PCA and k-PCA, the entire topology and number of nodes of the MANET are completely unknown. In Arroyo-Valles et al. [4], the nodes transmit only the statistically important messages and discard the rest. This imposes extra computation cost to infer which message is statistically important. Moreover, the discussed model will not take into consideration the energy cost per CPU instruction for evaluating the degree of significance for each received message. Furthermore, the model in [3] uses linear extrapolation approaches for forwarding univariate contextual values. In [12], each sensor node performs PCA by projecting its local data along the principal components and applies a clustering algorithm on this projection. Then, the node communicates compressed data to a Data Collection Point (DCP), which, by performing PCA on such data, produces global principal components. The DCP projects its data along the global principal components which are sent back to the sensor nodes to assist in subsequent compression. However, such a method considers a central process to define the groups, which may be overloaded if nodes were required to react to the definition of clusters, thus requiring DCP to communicate with all nodes [12].

The proposed model defines an adaptive scheme for estimating the span of the forwarding period for observations to minimize energy consumption due to data transmission and reception. The i-PCA and kernel-PCA deal with the statistical relations among contextual parameters and estimate the current observation vector in the upstream nodes every time a forwarding decision takes place. The considered models control not only the compression period but also the online learning phase.

II. BACKGROUND ON PRINCIPAL COMPONENT ANALYSIS AND KERNEL ENTROPY COMPONENT ANALYSIS

2.1 Principal Component Analysis

The PCA is the best compression scheme that minimizes the mean squared error between the original images and their reconstructions for any given level of compression [13, 28]. The PCA is a

technique in statistical data analysis for data approximation and lossy compression.

The eigenspace model is constructed as follows:

Let T = 1, 2,... be a discretized time domain accounting for the sampling period at which the sensor measurements are collected. Each sensor (source) generates a stream of observations.

Let $\mathbf{x}[t] = [(x_1(t), x_2(t), \ldots, x_N(t)] \in R^N$ be the *N*-dimensional (column) vector of measurements (observation) sensed by a source at time $t \in T$. Let $X_N \times T$ be a matrix with elements $x_{kt} = x_k[t]$ containing column-wise *T* observations of $\mathbf{x}[t]$, $1 \le t \le T$.

Let $\bar{x}$ be the average of all N training samples and represented as a column vector. Let **U** be a training matrix of size NxT defined as:

$$\mathbf{X} = [x_1[t], -\bar{x}, x_2[t], -\bar{x}, \ldots, x_N[t], -\bar{x}] \quad ..(1)$$

The training matrix **X** given in (1) contains one-dimensional column vectors of the reshaped mean-centered data. Each column vector is then represented as a point in a high dimensional vector space. However, a more number of training examples are required to get reliable and robust estimation about the nature of data distribution, which is often called as *a curse of dimensionality*. Besides, as the dimension of the covariance matrix **Q** defined by $\mathbf{XX}^T$ is very large (normally M>>N), computation of the eigenvectors of such a huge matrix is prohibitive and hence we have to initially compute the eigenvectors of $\bar{Q} = \mathbf{X}^T\mathbf{X}$ which is of size NxN and subsequently the eigenvectors of **Q** are computed.

The eigenvectors, $\bar{e_i}$ and $\bar{\lambda_i}$ be the corresponding eigenvalues $\bar{Q}$ are determined by solving the well-known eigen-structure decomposition problem:

$$\bar{\lambda_i}\,\bar{e_i} = \bar{Q}\,\bar{e_i} \quad \cdots (2)$$

The eigenvectors, e_i and the corresponding eigenvalues λ_i of Q are now computed as follows [20].

$$\lambda_i = \bar{\lambda_i} \quad \cdots (3)$$

$$e_i = \frac{1}{\sqrt{\lambda_i}} U\,\bar{e_i} \quad \cdots (4)$$

However, we can have at most N number of eigenvectors using Eq. (4). Though all the T eigenvectors are required for exact data re-construction, only a small number, $k << N$, is generally good enough to obtain the primary features of the objects. The *k* eigenvectors, corresponding to the *k* largest eigenvalues, constitute the *eigenspace*. It should also be noticed that the first **N** eigenvectors $\bar{Q}$ are the first N eigenvectors vectors of **X** [20]. Thus eigenspace analysis can radically lessen the dimension (T) of the data to the eigenspace dimension (*k*) while keeping several of the most effective features that summarize the original data. We have shown in Fig. 1, a 2-D example of the PCA to illustrate dimensionality reduction. The eigenvector e_1 is the principal component associated with the largest eigenvalue λ_1 for the dataset **X**, and the projection of X_i's on e_1 results in Y_i's which minimizes the error between the data points X_i's and their projections Y_i's.

For instance, we can keep (1) the *Q* largest eigenvalues, (2) all PCs whose eigenvalues exceed an absolute threshold, and (3) the largest PCs such that a specified fraction of "energy" in the eigenvalue spectrum is retained. The portion α of retained variance with the first *Q* PCs, which characterizes the accuracy of the approximation, is

$$\alpha = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{N} \lambda_i} \quad \ldots\ldots (5)$$

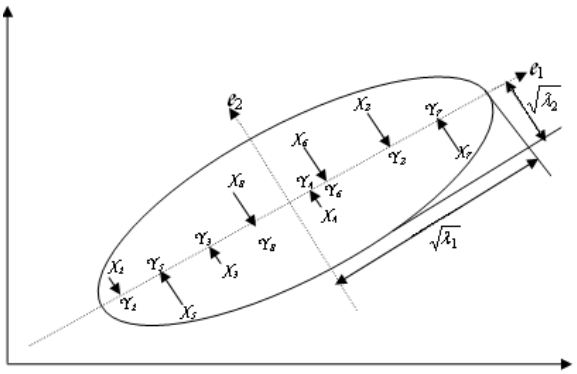

Fig. 1. A 2D example of PCA.

It is common practice in data analysis to retain the first *k* PCs such that α = 0.9; that is, to conserve 90% of the variance of the original signal. Having ordered column-wise matrix, approximations the set of vectors $\{w_i\}_{1 \le i \le k}$ by decreasing order of their eigenvalues λ_i Eigen matrix are obtained by:

$$\widetilde{x[t]} = ee^T x[t] \quad \ldots\ldots (6)$$

The *compressed* column vector **z** $[t] = W^\top \mathbf{x}[t]$ refers to the vector of coordinates of $\hat{x}[t]$ in $\{\mathbf{w}_i\}_{1 \le i \le k}$. This dimensionality reduction is lossy; thus, to measure the accuracy of $\hat{x}[t]$, the reproduction error $e[t]$ and the relative reproduction error are:

$$e[t] = \|\hat{x}[t] - x[t]\|$$
$$\eta[t] = \frac{e[t]}{\|x[t]\|}, x[t] \neq 0 \quad \ldots(7)$$

2.2 Incremental Principal Component Analysis

PCA is a batch algorithm; that is, $X_N \times T$ must be provided to the algorithm to produce the principal

components. The incremental PCA (i-PCA) technique incrementally updates the *current* eigenbasis (if necessary) on the reception of a new observation. Many algorithms have been proposed for IPCA [5, 8, 23, 30].

We review here the method described in [8]. Consider a eigen-basis of k principal components in an $N \times T$ matrix e[t] at time $t = T$ derived from the first T observations **x**(1), **x**(2), .., **x**(T) with $\tilde{x}(t) = \frac{1}{T}\sum_{k=1}^{T} x(k)$. Let y[t+1] ϵ R^N be the new observation. Let Z[t+1] = W^T(y[t+1] - $\tilde{x}(t)$) be its projection.

The reproduction error is **e**[t + 1] = (**y**[t + 1] $-\tilde{x}(t) - W$ **z**[t + 1]. The vector W.**z**[t + 1] lies entirely within the subspace defined by the eigenvectors. The i-PCA estimates the new eigenvectors given the new observation **y**[t + 1]. Specifically, the i-PCA method in [8] includes an additional eigenvector if necessary. That is because, based on [8], the new eigen-basis might consist of either $r = Q$ or $r = Q$+ 1 eigenvectors. The (updated) new mean value and (updated) sample covariance matrix are then:

$$\bar{x}[t+1] = \frac{1}{T+1}(T\bar{x}[t] + y[t+1])$$

$$C[t+1] = \frac{T}{T+1}c(T) + \frac{T}{(T+1)^2}(y[t+ \quad \text{... (8)}$$

$1-xt\ yt+1-xtT$

Consider that $r = Q + 1$. If it turns out that the additional eigenvalue is relatively small, it will be discarded (and the corresponding eigenvector at a later stage); thus, $r = Q$. The new eigenvectors must be a (Q+ 1) × (Q+ 1) rotation matrix R[t + 1] of the current eigenvectors plus some new orthogonal unit vector. Based on [8], the unit error vector of **e**[t+1] is an obvious choice for the additional unit vector, such that,

$$e[t+1] = \begin{cases} \dfrac{e[t+1]}{\|e[t+1]\|} & if \|e[t+1]\| \geq 0 \\ 0 & Otherwise \end{cases} \quad ..(9)$$

Hence, the new $N \times r$ matrix W[t + 1] of eigenvectors is:

W[t+1] = [W[t]$\hat{e}[t+1]$]R[t+1]. .... (10)

with $r = Q$+ 1. The rotation matrix R[t+ 1] is the solution of the following "eigenproblem" of size (Q+1) × (Q+ 1)

$$\left(\frac{T}{T+1}\begin{bmatrix}\Lambda[t] & \mathbf{0}\\ \mathbf{0}^\top & 0\end{bmatrix} + \frac{T}{(T+1)^2}\begin{bmatrix}\mathbf{z}[t+1]\mathbf{z}^\top[t+1] & \gamma\mathbf{z}[t+1]\\ \gamma\mathbf{z}^\top[t+1] & \gamma^2\end{bmatrix}\right)R[t+1] = R[t$$

, ... [11]

w $\gamma = (\hat{\mathbf{e}}[t+1])^\top(\mathbf{y}[t+1] - \bar{\mathbf{x}}[t]), \mathbf{0}$

here

is a Q-dimensional zero column vector, and [t] is the diagonal $Q\times$ Q matrix of the Q first eigenvalues λ_k associated with each PC $\mathbf{w}_k$ from W[t]. The solution of this problem yields the new eigenvalues directly in the (Q+ 1) × (Q+ 1) diagonal matrix $\Delta'[t+1]$, and the corresponding new eigenvectors are computed from W[t + 1] in Equation (10).

2.3 The Incremental PCA Model

In this section, we describe the adoption of the IPCA method [8] for enhancing the basic PCA model. The proposed model hereafter referred to as i-PCA, relaxes the assumptions of the basic PCA model, reduces the induced computational load and storage of the node, and updates the current eigenbasis according to new observations. i-PCA adopts online learning of the current eigenbasis on a new observation at node i. The i-PCA achieves simultaneously possible online learning and compression. At the beginning (i.e., t = 1, up to t = T), node i undergoes a learning phase as in basic model; that is, it receives, stores, and forwards the T first observations x[1],x[2],...,x[T] to node j. At t = T, both nodes produce locally the Q first principal components using PCA (i.e., the (initial) N × Q matrix W[t] and mean(x)).

Consider a new observation x[t] received by node i at t > T . Node i decides whether x[t] results in a change of the current eigen-basis or not. It projects x[t] onto the subspace a tolerance threshold defined by {wk}1≤k≤Q thus producing the error θ ∈ R+, then node e[t]= (xi[transmits the (compressed)] − mean(x)) − Wz[t]. If η[t] is less than z[t] to node j. Node j can then reconstruct the observation from its locally stored eigen-basis; that is, xˆ[t] = Wz[t]+ mean(x). If η[t] > θ (i.e., x[t] is not accurately represented by the current eigen-basis), then: node i forwards x[t] and updated x[t] to node j and both nodes update the eigen-basis through Equation (10) and maintain the new r first principal components with r = Q or r = Q+ 1 with respect to α.

In i-PCA, there is no explicit learning and compression phase other than the initial learning phase. Node i incrementally learn the principal components from incoming observations if necessary and forwards compressed vectors once there is no change in the eigenbasis to θ. Node j reconstructs the observation with error e[t] if η[t] ≤ θ; otherwise, e[t] is +∞ (i.e., the error e[t] refers to the reconstruction of the observation from its locally stored eigenbasis). However, the selection of eigenbasis is based on the Renyi entropy component analysis which is described in the following section.

2.4 Kernel Entropy Component Analysis

The kernel PCA is one of the best data transformation approach known in the machine learning domain. Let **X** = [x_1, x_2,…,x_N], where x_i Є

R^d, i=1,…,N. The basic idea of kernel PCA is to map the input data **X** onto a feature space F via a nonlinear mapping ϕ and then perform a linear PCA in F.

The non-linear map from input space to feature space is given by ϕ: $R^d \rightarrow F$ such that $x_t \rightarrow \phi(x_t)$, t=1,…,N. Let $\boldsymbol{\phi} = [\phi(x_1), \phi(x_2), \dots, \phi(x_N)]$. To perform PCA in F, we need to find an expression for the projection $P_{\alpha_i}\boldsymbol{\phi}$ of $\boldsymbol{\phi}$ onto a feature space principal axes α_i, or onto a subspace, E_l spanned by the top *l* eigenvectors, which is achieved implicitly via the kernel function.

The estimated covariance matrix of the mapped data $\phi(x_i)$ in kernel PCA is defined as:

$$C = \frac{1}{N}\sum_{i=1}^{N} \Phi(x_i).\Phi(x_i)' \quad \dots(12)$$

And the corresponding eigenvalue problem is:

$\lambda w = Cw$.

As $Cw = \frac{1}{N}\sum_{i=1}^{N}(\Phi(x_i).w).\Phi(x_i)'$, all solutions w with $\lambda \neq 0$ lie in the span *of* $\phi(x_1), \phi(x_2), \dots, \phi(x_N)$. i.e., the coefficients α_i (i=1,…,N) exist such that

$$w = \sum_{i=1}^{N} \alpha_i \phi(x_i) \quad \dots(13)$$

Then the following set of equations can be considered:

$\lambda(\phi(x_i).w) = (\phi(x_i).Cw)$ for all i=1,…,N …(14)

Upon substituting (12) and (13) into (14) and by defining the NxN kernel matrix K by $K_{ij} = k(x_i, x_j) = \phi(x_i).\phi(x_j)$ produces an eigenvalue problem that can be expressed in terms of the dot products of two mappings.

Solve $N\lambda\alpha = K\alpha$

For nonzero eigenvalues λ_i and eigenvectors $\alpha_i = (\alpha_1, \dots, \alpha_N)'$ subject to normalization condition $\lambda_i\alpha^i.\alpha^i = 1$.

In kernel PCA, the top-most eigenvectors corresponding to the first few largest eigenvalues are used to transform the high dimensional feature vector to a lower dimensional feature vector. In the case of kernel ECA, it is the energy content that is being used in order to estimate the eigenvalues and Eigen vectors which are used for data transformation. The following section present the details in detail.

2.4.1 Kernel Entropy Component Analysis for face feature extraction

The Renyi quadratic entropy is given by

$$H(p) = -\log \int p^2(x)dx, \quad \dots (15)$$

where p(x) is the probability density function generating the data set, or sample, $S = x_1, x_2, \dots, x_N$. Because of the monotonic characteristic of the logarithmic function, we consider the quantity:

$$V(p) = \int p^2(x)dx \quad \dots (16)$$

The estimation of V(p) will be done using the Parzen window density estimator as given below.

$$\hat{p}(x) = \frac{1}{N}\sum_{x_t \in S} k_\sigma(x, x_t) \quad \dots (17)$$

Here, $k_\sigma(x, x_t)$ is the kernel centered at x_t and width governed by the parameter σ.

Hence,

$$\hat{V}(p) = \frac{1}{N}\sum_{x_t \in S}\hat{p}(x_t)$$
$$= \frac{1}{N}\sum_{x_t \in S} \frac{1}{N}\sum_{x_{t'} \in S} k_\sigma(x_t, x_{t'}) \quad \dots(18)$$
$$= \frac{1}{N^2}\mathbf{1}^T\mathbf{K}\mathbf{1}$$

Here, the element (t, t') of the N x N kernel matrix K is $k_\sigma(x_t, x_{t'})$ and **1** is an (N x 1) vector containing all ones.

Hence, the Renyi entropy is represented in terms of the eigenvalues and eigenvectors of the kernel matrix, which is decomposed as $\mathbf{K} = \mathbf{EDE}^T$. Here, **D** is the diagonal matrix having the eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_N$ and **E** is a matrix with the respective eigenvectors $\alpha_1, \alpha_2, \dots, \alpha_N$ as columns.

Rewriting the above, we have:

$$\hat{V}(p) = \frac{1}{N^2}\sum_{i=1}^{N}(\sqrt{\lambda_i}\alpha_i^T\mathbf{1})^2 \quad \dots (19)$$

We can notice here that the eigenvalues and the associated eigenvectors contribute more to the estimation of entropy, which is the fundamental difference between Kernel ECA and kernel PCA. That is Eq. (19) discloses that the Renyi entropy estimator is composed of projections onto all the kernel PCA axes, however, only a principal axis α_i for which $\lambda_i \neq 0$ and $\alpha_i^T\mathbf{1} \neq 0$ subsidises to the entropy estimate. Hence, kernel ECA is defined as an *l*-dimensional data transformation method got by projecting $\boldsymbol{\phi}$ onto a subspace E_l spanned by *l* kernel PCA axes contributing maximum to the Renyi entropy estimate of the information. Therefore E_l is composed of a subclass of kernel PCA axes but not essentially those conforming to the top *l* eigenvalues. Hence in kernel ECA, for principal components, the projections of x are computed onto the eigenvectors in F which contributes more to Eq. (19).

2.5 Time and Storage Complexity

We first report on the time and storage complexity of PCA and i-PCA. The optimal linear scheme of PCA for reducing a set of T observations of dimension N to dimension Q involves computing the sample covariance matrix with $\boldsymbol{O}(TN^2)$ time

complexity. The i-PCA is executed online using only one N-dimensional observation at a time with storage requirements $\boldsymbol{O}(Q^2)$. Also, i-PCA is quite attractive in our case because its complexity is $\boldsymbol{O}(Q^3)$; this corresponds to the solution of the eigen problem for a Q×Q matrix.

The basic model needs **O**(TN) storage and $\mathbf{O}(TN^2)$ time for the learning phase. It also requires **O**(Q) time for a compression period of length. The i-PCA model requires ***O***(TN) storage and ***O***(TN2) time for the initial learning period of length T. After this period, i-PCA needs only $\boldsymbol{O}(Q^2)$ storage, and its time complexity depends on the probability $p = P(\eta[t] > \theta)$. Hence, it needs $\boldsymbol{O}(pQ^3)$ time for a horizon of length after the initial period.

Finally, it should be stressed that the operation of the proposed schemes typically relies on a static topology with established routing paths. If a certain node fails, then a learning phase has to be initiated on the alternative path (determined through the underlying routing protocol) to render upstream nodes capable of handling the compressed stream. A similar remedy (i.e., rerouting) is assumed for the scenario of energy-failing nodes.

III. PERFORMANCE AND COMPARATIVE ANALYSIS

In this section, we study the performance of i-PCA integrated with k-ECA and also compared with the basic PCA model in terms of load gain and energy efficiency achieved due to the Renyi entropy component analysis. Moreover, we compare our models with the non-PCA-based model called as Threshold-sensitive Energy Efficient sensor Network (TEEN) model presented in [18] for completeness reasons, although the TEEN is proposed for wireless sensor networks. The TEEN is intended to be responsive to sudden changes of a measurable quantity. TEEN is based on a hierarchical structure of nodes: cluster nodes and simple sensor nodes. The nodes communicate the contextual data only to their immediate cluster-head, thus saving energy. A cluster-head node i agrees on forwarding a contextual value xk[t] to the upstream cluster-head node j based on two predefined tolerance thresholds: (1) the hard threshold k, which is an absolute value of the k-th contextual parameter, and (2) the soft threshold θk, which is a small relative change in the sensed xk[t] value. Node i forwards xk[t] to node j at t when xk[t] > k and the change in xk[t] is equal or greater than θ_k, where k is used for reducing the number of transmissions of contextual components by allowing node i to transmit only when the sensed contextual value xk[t] is in the range of interest. Besides, θ_k reduces the number of transmissions by further eliminating all transmissions that might have occurred when there is a minor change in xk[t], with xk[t] > k. TEEN, however, handles univariate contextual information. To objectively compare the performance of our models with TEEN, we apply TEEN for N-dimensional observations x. For comparison reasons, the soft threshold equals the tolerance threshold $\theta=\theta_k$, k=1,...,N in the experiments. Besides, k is defined as the average value of the lowest and the highest possible values of the k-th parameter, similarly to the performance evaluation of TEEN presented in [18].

As a reference model in terms of energy consumption, we refer to a model in which node i unconditionally forwards all received observations to node j without any additional processing. The reference model incurs no data inaccuracies (zero error).

3.1 Performance Metrics and Parameters

The total cost c_t (in Joules) at time t is the cost incurred by node i transmitting x[t] or z[t] and by node j receiving x[t] or decompressing z[t]. Cost c_t is recursively calculated as follows:

$$c_t = c_{t-1} + c_t^R + c_t^T + c_t^I + c_t^0 \quad \text{... (20)}$$

where c_t^R and c_t^T are receive and transmit costs for x[t] or z[t], respectively. The c_t^I is the energy cost for the CPU instructions for the adopted algorithms. The c_t^0 is the cost for node i transiting from idle to standby operational modes. We denote as the total cost for the REF model in which $c_t = 0$, since no processing is applied to observations. We define as load gain ratio $a \in [0,\infty)$ up to time T the communication load gain obtained from a model to reference model; that is,

$$\alpha = \frac{c_T}{c'_T} \quad \text{.... (21)}$$

A value of a < 1 indicates a reduction of communication costs concerning REF. The lower the value is, the lower energy consumption is attained by a certain model. Moreover, we assess the benefit of a model by taking into account the mean relative reproduction error $\bar{\eta}$ up to time T; that is,

$$\bar{\eta} = \frac{1}{T}\sum_{t=1}^{T}\eta(t) \quad \text{.... (22)}$$

Also, let m_t refer to the number of transmitted values at t by node i (i.e., m_t = Q) if compression is in effect at instance t; otherwise m_t = N. For a basic reference model, we obtain m_t = N, ∀t. We define as mean data compression ratio δ up to T the quantity:

$$\delta = \frac{1}{T}\sum_{t=1}^{T}\frac{N}{m_t} \quad \text{... (23)}$$

A high δ value indicates the efficiency of a model in transmitting low-dimensional information.

3.2 Datasets

We have considered the data proposed as detailed below. The Trace T_0 contains readings from real sensors distributed in a certain region for monitoring in-field temperature, relative humidity,

and wind direction/speed. Trace T_0 has been collected from an experimental sensor deployment of the Sensor and Computing Infrastructure for Environmental Risks (SCIER) project. The observations that are processed in the nodes are of N = 7 dimensions. The Trace T_1 refers to real sensor readings taken from a commercial vessel to monitor the ship's health condition. Trace T_1 was collected from production installations of the Mari-Brain system used for online remote monitoring of ships based on sensor network installed onboard. The observations that are processed in the nodes are of N = 29 dimensions.

Table-1. Characteristics of Datasets.

Data set T_0	**Value**
Dimension of observation	N = 7
Number of observations	4800
Number of hours of sensing	400
Sampling period	5 minutes
Predictability	high (avg. Hurst exponent 0.87)
Stationarity	medium
Sensor types	humidity, wind speed, temperature
Data set T_1	**Value**
Dimension of observation	N = 29
Number of observations	25920
Number of hours of sensing	2160
Sampling period	5 minutes
Predictability	low (avg. Hurst exponent 0.52)
Stationarity	low
Sensor types	propeller shaft torque, tank leak/fuel gauges, fuel flow speed, inclinometers, ground speed, depth, water speed, vessel head, wind angle, wind speed

3.3Energy Model

We adopted the energy consumption model from the Mica2 sensor board [He et al. 2004]. This energy model assumes energy of two AA batteries that approximately supply 2,200mAh with an effective average voltage of 3V. It has 20mA if running a sensing application continuously, which leads to a lifetime of 100 hours. The energy costs for lone CPU instructions (energy per instruction) and transmitting/receiving contextual values (energy per bit) are given in Table-2.

Table-2. Energy Costs

Node operation mode	Energy cost
Instruction execution	4nJ/instruction
Idle – Stand by	9.6–0.33mJ/s
Transmitting – Receiving	720–110nJ/bit

Characteristics of i-PCA: We study certain characteristics of the proposed model, giving some insight into its behavior. Figure 1 shows the Probability Density Function (PDF) of the compression duration $\Delta t*$ for different θ values and T = 50 for trace T0. The i-PCA shortens the compression period once tolerance is relatively low and vice versa.

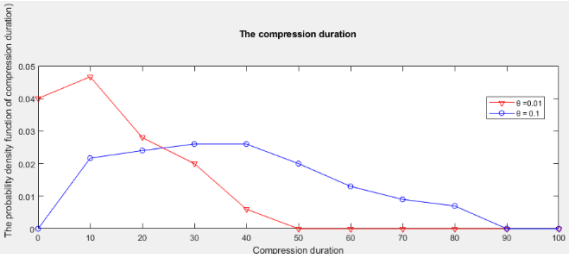

Fig. 2. The probability density function of duration t∗ of i-PCA for different θ values and T = 50 for Trace T0.

Load Gain: Figures 3 and 4 present the a value against T for i-PCA (coupled with Renyi entropy component analysis) and the basic PCA model, taking the average for all θ values for traces T_0 and T_1, respectively. The value for i-PCA and the basic model is obtained for all and θ values, respectively. The i-PCA is very energy efficient for all T for both traces because it exploits the initial horizon only for constructing the (initial) eigenbasis and then updates the eigenbasis to IPCA. The *a* value for basic PCA increases with T as expected because, during the learning phase, basic PCA transmits/receives uncompressed observations. The proposed model is very energy efficient because, after the initial learning period, it transmits only compressed observations and even with high θ, it possesses energy efficiency, which indicates its applicability for MANET applications with low and high data accuracy requirements.

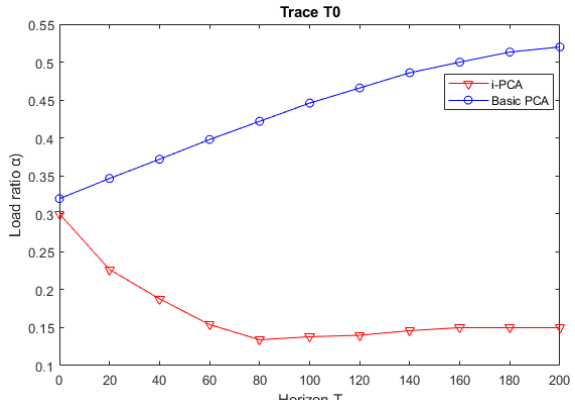

Fig. 3. Load gain α against horizon T for all models for Trace T_0.

We have made a comparative analysis with TEEN which assumes the highest load gain values for both traces compared to the proposed model since it does not exploit statistical dependencies among contextual values as the proposed models do and are agnostic of the error at the receiver end. The TEEN has the load gain 7.60, 16.6, 25.00 38.03 43..34 48.98, and 6.00 7.95 8.34 9.46 11.23 13.32 respectively for T0 and T1 traces.

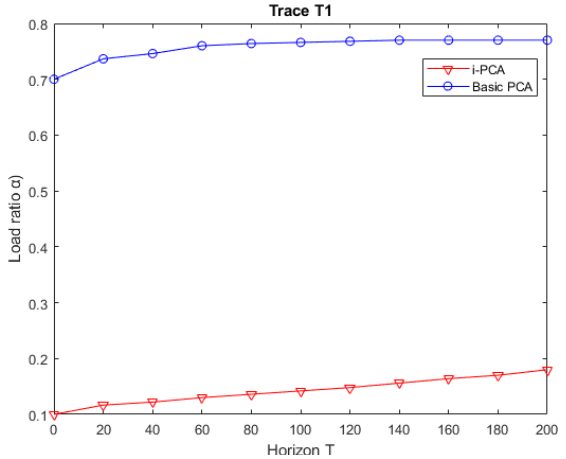

Fig. 4. Load gain α against horizon T for all models for Trace T_1.

IV. CONCLUSION

In this work, we have proposed an incremental principal component analysis based model coupled with kernel entropy component analysis for energy minimization suitable to mobile ad-hoc networks. The data aggregation issue has been addressed here and successfully comes out with a new model that consumes less energy due to the dimensionality reduction of the data i.e., in terms of the number of bits being transferred through sensor nodes where they are constrained by battery life. The redundant information in a large amount of data is identified by exploring a well-known multivariate data analysis technique called principal component analysis and a variant of principal component analysis called incremental principal component analysis is explored suitable to the mobile ad-hoc network environment. Besides, we have also integrated the incremental principal component analysis with Renyi entropy component analysis which is useful in estimating the best eigenbasis, thereby accurate reconstruction of the data is possible at the receiving end (destination node). The proposed approach results in a low energy consumption model with controlled data accuracy (based on tolerance θ). Experimental results indicating the load gain which in-turn helps in saving the energy consumption are presented by considering the dataset used by many researchers. The comparative analysis with the recently proposed approaches reveals the performance of the proposed method.

REFERENCES

[1] Akyildiz, I., W. Su, Y. Sankarasubramaniam, and E. Cayirci. Wireless sensor networks: a survey. Computer Networks, 38(4):393–422, 2002.

[2] Anagnostopoulos, C and S. Hadjiefthymiades, 2014: Advanced Principal Component-Based Compression Schemes for Wireless Sensor Networks, ACM Transactions on Sensor Networks, Vol. 11(7), pp. 7.1-7.34.

[3] Anagnostopoulos, C., T. Anagnostopoulos, and S. Hadjiefthymiades. 2010. An adaptive data forwarding scheme for energy efficiency in wireless sensor networks. In Proceedings of IEEE Conference on intelligent Systems. 396–401.

[4] Arroyo-Valles, R., A. Marques, and J. Cid-Sueiro. 2009. Optimal selective transmission under energy constraints in sensor networks. IEEE Transactions on Mobile Computing 8, 11, 1524–1538.

[5] Artac, M., M. Jogan, and A. Leonardis. 2002. Incremental PCA for on-line visual learning and recognition. In Proceedings of the 16th International Conference on Pattern Recognition. 781–784.

[6] Ercal G, Avin C. On the cover time of random geometric graphs. Autom Lang Program. 2005;3580(1):677–89.

[7] Gnawali O, Fonseca R, Jamieson K, Moss D, Levis P. Collection tree protocol. In: Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems, SenSys '09. New York: ACM; 2009. p. 1–14.

[8] Hall, P., D. Marshall, and R.Martin. 1998. Incremental eigenalysis for classification. In Proceedings of British Machine Vision Conference. 286–295.

[9] Ilyas, M., I. Mahgoub, and L. Kelly. Handbook of Sensor Networks: Compact wireless and Wired Sensing Systems. CRC Press, Inc. BocaRaton, FL, USA, 2004.

[10] Intanogonwiwat C, Govindan R, Estrin D, Heidemann J, Silva F. Directed diffusion for wireless sensor networking. IEEE Trans Netw. 2003;11(1):2–16.

[11] Jolliffe. I. 2002. Principal Component Analysis (2nd ed.). Springer Series in Statistics. Springer.

[12] Kargupta, H., W. Huang, K. Sivakumar, and E. L. Johnson. 2001. Distributed clustering using collective principal component analysis. Knowledge and Information Systems 3, 4, 422–448.

[13] Kirby, M and L. Sirovich, "Application of the Karhunen- Loève procedure for the characterisation of human faces" IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 12, pp. 831-835, Dec. 1990.

[14] Kulathumani V, Arora A, Sridharan M, Parker K, Lemon B. On the repair time scaling wall for manets. IEEE Commun Lett. 2016;PP(99):1–4.

[15] Lee, H., A. Klappenecker, K. Lee, and L. Lin. 2005. Energy efficient data management for wireless sensor networks with data sink failure. In Proceedings of IEEE International Conference on Mobile Ad-hoc and Sensor Systems 7, 210.

[16] Lovascz L. Random walks on graphs: A survey. Combinatorics, Paul Erdos 80. 1993.

[17] Madden S, Franklin JM, Hellerstein J, Hong W. TAG: A Tiny Aggregation Service for Ad-hoc Sensor Networks. SIGOPS Oper Syst Rev. 2002;36(SI):131–46.

[18] Manjeshwar A and D. Agrawal. 2001. TEEN: A protocol for enhanced efficiency in wireless sensor networks. In Proceedings of 1st International Workshop on Parallel and Distributed Computing Issues in Wireless Networks and Mobile Computing. 23–27.

[19] Mousam Dagar and Shilpa Mahajan, "Data Aggregation in Wireless Sensor Network: A Survey", International Journal of Information and Computation Technology, Volume 3, Number 3, 2013. ISSN 0974-2239

[20] Murakami, H and B. V. K. V. Kumar, "Efficient Calculation of Primary Images from a Set of Images," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-4, no. 5, pp. 511-515, Sept. 1982, doi: 10.1109/TPAMI.1982.4767295.

[21] Naik V, Arora A, Sinha P, Zhang H. Sprinkler: A Reliable and Energy Efficient Data Dissemination Service for Extreme Scale Wireless Networks of Embedded Devices. IEEE Trans Mob Comput. 2007;6(7):777–89.

[22] Nanthini.D and R.A.Roseline, "Aggregation Protocols in Wireless Sensor Network- A Survey", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 3 Issue 7, July 2014.

[23] Oja, E and J. Karhunen. 1985. On stochastic approximation of the eigenvectors and eigenvalues of the expectation of a random matrix. Journal of Mathematical Analysis and Application 106, 69–84.

[24] Pattem, S., B. Krishnamachari, and R. Govindan. The impact of spatial correlation on routing with compression in wireless sensor networks. Proceedings of the third international symposium on Information processing in sensor networks, pages 28–35, 2004.

[25] Rakesh, K R.Ranjan, S.P.Karmore, "Survey on Secured Data Aggregation in Wireless Sensor Network" IEEE Sponsored 2nd International Conference on Innovations in Information Embedded and Communication Systems 2015.

[26] Ren, F., J. Zhang, Y. Wu, T. He, C. Chen, and C. Lin. 2013. Attribute-aware data aggregation using potential based dynamic routing in wireless sensor networks. IEEE Transactions on Parallel Distributed Systems 24, 5, 881–892.

[27] Scaglione, A and S. Servetto. On the Interdependence of Routing and Data Compression in Multi-Hop Sensor Networks. Wireless Networks, 11(1):149–160, 2005.

[28] Sirovich, L. and Kirby, M. (1987) Low-Dimensional Procedure for the Characterization of Human Faces. Journal of the Optical Society of America A, 4, 519-524.

[29] Shekar, B.H., M. Sharmila Kumari, Leonid M. Mestetskiy, Natalia F. Dyshkant, Face recognition using kernel entropy component analysis, Neurocomputing, vol. 74(6), 2011, Pages 1053-1057.

[30] Weng, J., Y. Zhang, and W. Hwang. 2003. Candid covariance free incremental principal component analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 25, 8, 1034–1040.

[31] Zhao, F and L.J. Guibas. Wireless Sensor Networks: An Information Processing Approach. Morgan Kaufmann, 2004.