Egyptian Informatics Journal 23 (2022) 137–147

Contents lists available at ScienceDirect

Egyptian Informatics Journal

journal homepage: www.sciencedirect.com

A Hierarchical Energy Conservation Framework (HECF) of Wireless Sensor Networks by Temporal Association Rule Mining for Smart Buildings

Farhan Sabir Ujager [a,*], Azhar Mahmood [b], Muhammad Usman [c], Muhammad Siraj Rathore [b]

[a] Faculty of Engineering & IT, Northern University, Nowshera, Pakistan
[b] Capital University of Science and Technology (CUST), Islamabad, Pakistan
[c] Shaheed Zulfiqar Ali Bhutto, Institute of Science and Technology, Islamabad, Pakistan

ARTICLE INFO

Article history:
Received 17 June 2021
Revised 25 August 2021
Accepted 20 September 2021
Available online 2 October 2021

Keywords:
Energy conservation
Wireless Sensor Network (WSN)
Hierarchal temporal association rule mining and Frequent patterns

ABSTRACT

The challenge of extending the sensor's energy consumption is a key research issue in Wireless Sensor Networks (WSNs). Recently, association rule mining has proven to be a potential candidate to prolong the lifetime of sensor nodes in WSN. However, temporal correlations of the contextual values are not taken into account which is useful for the sensors to conserve their energy. Similarly, association rules mining at different tiers of the network has not been considered to reduce the number of transmission messages, by avoiding redundant data which is the major cause of the energy drain of sensors. In this paper, a novel Hierarchical Energy Conservation Framework (HECF) is proposed which aims to conserver's energy at each layer of a network by using the Hierarchical Temporal Association Rule Mining in multistory buildings. In hierarchal setup, each floor of the building can conserve energy locally at the local sink and conserve entire network energy at the global sink by using temporal association rule mining at different tiers of the network. The HECF is ideal for large multistory buildings where energy conservation is a major issue along with effective monitoring and system performance. The result shows that HECF outperformed other classical energy conversation methods such as LEACH-C and RR-Schedule-Buffer in terms of energy consumption. It extends 16% network lifetime, also 20% less number of messages during data transmission, which is a remarkable improvement for sensors energy conservation.

1. Introduction

Recent advances in the field of microelectronics and wireless communications have led to the development of large scale WSNs. These sensor networks have the capability to sense the environment, compute the data locally, and relay them to the control location for further processing. The major cause of the energy drain is the transmission of sensed data by the sensor nodes. By reducing redundant data transmission, and minimizing the transmission messages, the lifetime of sensor nodes can be extended, and hence WSN [1].

WSNs have gained importance in many application domains such as habitant monitoring, environment monitoring, military and disaster management, and object tracking. Among all application areas, one of the existing emerging application areas is smart multistory building [2–5]. Sensors based smart building yields many benefits such as

- The improved residential and commercial environment
- The improved lifespan of electric appliances and devices
- Efficient energy conservation for the usage of associated facilities of the building
- Improved assets reliability
- Most importantly observing, scheduling and reacting to the status of smart buildings from centralized and decentralized locations

* Corresponding author.
E-mail address: farhan@northern.edu.pk (F.S. Ujager).

Peer review under responsibility of Faculty of Computers and Information, Cairo University.

Production and hosting by Elsevier

https://doi.org/10.1016/j.eij.2021.09.001
1110-8665/© 2022 Published by Elsevier B.V. on behalf of Faculty of Computers and Information, Cairo University.

All these benefits are dependent on efficient energy management and an improved lifetime of the WSN network. Recently, knowledge extraction from WSN has received great attention in the data mining community. Different approaches [6–8] based on classification; clustering and association rules mining have been successfully used on WSN. However, WSN creates new research challenges due to continuous, heterogeneous, and huge amounts of data generation where conventional data mining techniques are not applicable. Association rules mining [9] was introduced to discover correlation among frequent itemsets that appeared in the transactional database. Association rules mining discovers associations from the raw data generated from the sensor nodes of WSN. These associations are temporal, spatial, or among different attributes [7,10–12] of sensed data of sensor nodes.

The association rules based mining approaches have either employed centralized or distributed rule generation architectures [7,11] for energy conservation of WSN. However, it has been observed that distributed association rule mining architecture yields better results in terms of response from the sink to a node and overall energy efficiency of WSN. Similarly, several association rules mining mechanisms have been proposed which are based on the spatial correlation [10,13,14] of sensor nodes, temporal and data attribute correlation of sensed data are considered for the generation of association rule for energy conservation. To best of our knowledge, none of the previously proposed techniques have employed tier level based associations together to mine useful patterns of the sensed data for the energy conservation of WSN. Similarly, the usefulness of the generated association rule has been analyzed with respect to support and confidence parameters , however, usefulness can be further analyzed with other advanced features of association rule mining such as Lift. Lift yields better rules as it considers confidence measure along with support value. Furthermore, association mining 'Lift' parameter has been proposed for several applications such as in building energy conservation [15,16]. Such analysis can be a new dimension for better energy efficiency to the WSN as well.

Modern buildings are complex in design and to provide a variety of services (such as management, security, and comfort) there is a need of more systematic service level designing such as hierarchical management. The majority of modern buildings are multistory and WSN is an integral part to provide a variety of services. Modern buildings have no energy provision limitations to the normal electric appliances, however, there are often some hard to reach locations in a building where it becomes difficult to provide power to the sensors, for example, sensors that are used for detecting sewerage pipe fault, water flow and supply fault detections. Direct power provision at such locations; replacing sensors batteries is difficult [17]. Similarly, Pipe and duct sensors are deployed in thin pipes and ducts for the monitoring of temperature, pressure and humidity. Another deployment location is open-air deployment which also needs remote battery backups instead of direct power supply. Thus in smart buildings WSN life prolongation is equally important as compared to other WSN deployments.

There is a need for hierarchical WSN deployment as it provides multiple benefits [18] such as hierarchical WSN is highly scalable in terms of designing and implementation, especially in multistory complex buildings. Similarly, it provides efficient manageability, as each tier performs its predefined functionalities which are consistent throughout each tier as shown in Fig. 1.

In the proposed HECF, we have considered scenarios of modern multistory buildings, as shown in the Fig. 2. In order to achieve hierarchal energy conservation, three algorithms (based on temporal associations) at different tiers of sensor network (sink, cluster head and node respectively) are proposed to mine useful patterns for conserving energy. The proposed algorithms at the sensor node and cluster head (CH) significantly reduce the redundant data transmission, where cluster is a group of sensors. CH is a sensor node in the cluster of sensor nodes and it perform additional tasks of transmitting the data of the cluster to the gateway along with the sensing. They have limited amount of battery backup for the said tasks. CH selection is based on an election which depends upon energy level and distance of the sensor from the sink [19]. Different sensor nodes take the responsibility of being CH and as discussed earlier sensor nodes are deployed in certain locations where direct power provision is not possible. The energy conservation is equally important at the CH level. A scheduling algorithm (temporal association based) is proposed for local and global sink of WSN, to achieve the optimum monitoring scheduling of sensor nodes to conserve WSN energy. Moreover, in this proposed framework, we have utilized advance rule validation measure (Lift) [20], for the generation of stronger sensor scheduling rules for conserving energy to prolong the lifetime of sensors in smart multi-story buildings.

Rest of the paper is organized as follow; Section II presents the related work and Section III describes the formal definitions of association rule mining and detailed discussion on proposed HECF framework. Section IV describes the implementation and simulation setup of HECF framework along with dataset details to validate the HECF framework. Section V provides simulation results and a comparative analysis of HECF with other approaches. In the end Section VI concludes the research study and its findings.

2. Related work

The related work section consists of two subsections (i) Literature review, and (ii) Comparative analysis of the related work.

2.1. Literature review

Data mining is a well-known computing domain, employed to extract interesting patterns and useful information from the given data. This extracted information is useful in taking critical decisions for the overall system improvement. Several studies have been conducted for the energy conservation of sensors by employing association rule mining from the data mining domain.

The reviewed literature in this study either follows the reduction in the transmission or the reduction in sensing by the scheduling of the sensor nodes. This study focuses to employ association rule mining based approaches to find an association among different aspects of sensors for energy conservation.

A novel mechanism [7] to learn behavioral pattern for Point-of-Coverage (POCWs) is proposed to reduce the frequency of event reporting at the sink. Three energy aware data preparation mechanisms proposed based on temporal event association of targets at the node level. However, spatial and data attribute correlations have not been considered in this approach for the generation of rules. The behavioral activities of the targets are monitored and stored in the buffers at each node [8]. The temporal association mining is performed in distributed manner to find out frequent nodes (nodes having support and confidence equal or above-given threshold values).

A mechanism [10] is proposed to compress the data gathered from the sensor nodes for fast and efficient mining so to study the behavioral patterns of the targets. A tree-based data compression mechanism is used for data mining which is based on spatial correlations. This proposed mechanism also employed flash memories at the node to study targets behavioral patterns. A novel algorithm, namely Intelligent Collaborative Event Query (ICEQ) is proposed [21] which appropriately selects sensor nodes at the boundary of an event for efficient monitoring instead of allowing all sensor nodes to monitor an event, based on association rule

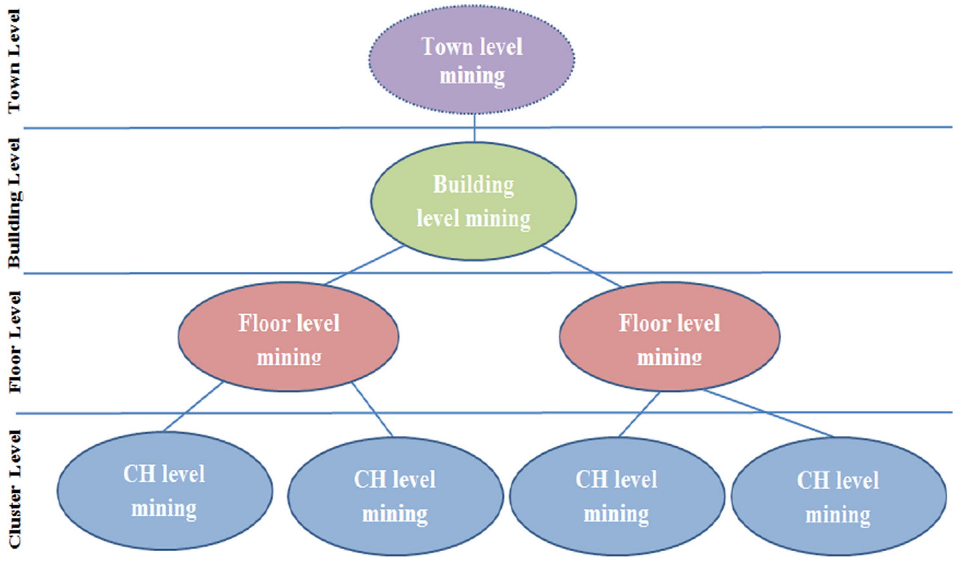

Fig. 1. Hierarchical sensor network architecture for the proposed framework.

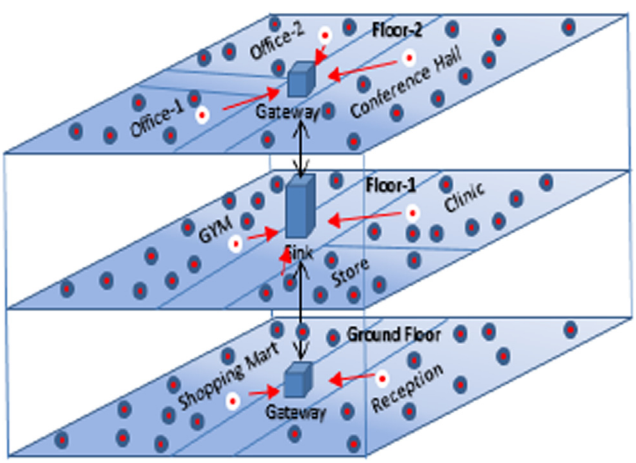

Fig. 2. Sensor network based multistory building.

mining. For efficient rules generation, tree-based mining is performed at the sink (centralized mining). The novelty of this proposed algorithm is that correlations are found based on queries generated and frequent sensor nodes for the selection. An adaptive data reduction approach has been proposed in [22] to increase the lifetime of sensor network. The proposed technique identify the redundant data by utilizing similarity functions and based on the score of similarity index sampling rate of sensors are adjusted to ensure the sampling rate adaptive.

A context aware framework [11], based on sensor node attributes (contextual) association rule mining, which uses distributed architecture for extracting rules at node and network level to improve the energy consumption of WSN. A Distributed Data Extraction (DDE) mechanism is proposed [19] to predict the missing and corrupted data from sensor by the application of association rule mining (temporal and data association). Additionally, the proposed mechanism reduces data size, which has significance in overall efficient network performance. A genetic algorithm based solution to reduce the amount data sent to a sink efficiently by using association rule mining in a distributed mining architecture [13]. Association rule mining at the node level extracts correlations among data attributes. The sink genetic algorithm uses extracted patterns, energy consumption and network energy equilibrium to compute an energy-efficient spanning tree for routing the sensed data at the sink. In [23], a prediction based data reduction technique has been proposed to reduce temporal redundant data. This proposed technique uses Principal Component Analysis (PCA) to reduce transmitted data, however, PCA may result in loss of sensing data thus results in low accuracy.

WSN energy efficiency technique is proposed in [24] for the occupancy detection system in a building. In this proposed approach different components of the sensor's duty cycling have been evaluated for sensors lifetime maximization. However, variation in duty cycling of sensors may results in the inaccuracy of sensing data. An energy-efficient sleep and wake (duty cycle) scheduling technique has been proposed in [25] for sensing nodes to reduce the power consumption of WSN. This proposed technique determines a optimal radius of sensors (which have balance energy levels) and fuzzy matrix similarity correlation makes sensing nodes to sleep which transmit redundant data to conserve the energy.

A centralized (data mining at sink) prediction based tracking technique [26], called prediction based tracking by sequential association rules patterns (PTSPs). It is a framework designed to achieve significant reductions in the energy dissipated by the object tracking sensor network while maintaining acceptable missing rate levels. A novel multilevel object traction (MLOT) mechanism [27] is proposed for energy efficient and real time tracking of objects in wireless sensor network by mining the movement logs at the sink.

2.2. Comparative analysis of related work

Several techniques have been proposed for extracting useful patterns for the energy conservation of sensor network. A comparative analysis of the proposed techniques which are discussed in this article is presented in this subsection. The comparative analysis is performed on certain defined parameters, these parameters are listed as:

- Association rule mining architecture is an important aspect, as centralized mining architecture doesn't yield better results because of delayed response from the sink to sensing nodes [28].
- Mining dimensions define the levels at which association mining is being performed for efficient response
- Algorithm type of association rules mining
- Correlation (Temporal, Spatial, Data) basis to extract the patterns for the generation of rules
- Rules validation standards (Support, Confidence, Lift, Levenger)
- Number of nodes in a network, that defines the size of network
- Data Routing mechanism as it is the most energy hungry process

  Close study of the comparative analysis in Table 1 reveals that the proposed solution should enjoy the following features:
- An ability to perform association mining in distributed manner (i.e. at different hierarchal tiers) for efficient response
- An ability to employ efficient association rule mining algorithm
- An ability to employ advance rule measuring standards for better rules generation
- An ability to use all correlation types for better extraction of patterns

3. Proposed energy conservation framework

This section of the article will talk about the proposed "A Hierarchical Energy Conservation Framework" (HECF) of WSN and the definitions associated with HECF.

3.1. Formal definition of association rules

Temporal association rule mining is being employed at different tiers of sensor network (i.e. Sensor node, CH and Sink). At sensor node and CH, temporal association mining will help in the reduction of redundant data transmission to CH and from CH to sink. Similarly, temporal associations of the sensing nodes will be used for the scheduling of the sensor nodes in the clusters.

In this section, we present the definitions of temporal association mining and association rules that will help to reduce the overall energy consumption of sensor network in multistory building.

The proposed HECF uses the parameters which are defined in for generating association rules, these parameters are Duty_Cycle (DCs) for sensing node, Time Window (λ) and minimum support, S. Another parameter is introduced as Duty_Cycle Cluster Head (DCCH) for cluster head. Time Window (λ) is the allowed time for a sensor to sense the environment and store in the buffer, there is no delay between two adjacent windows as the sensor is sensing the data continuously. λ is the time unit of DCs, it consists of number of time windows (λs) as shown in Fig. 3. Sensor node does not continuously send redundant data messages to the CH. After performing temporal association mining (over DCs time) it sends data that fulfills minimum defined support (S) to CH. Similarly, DCCH is the period of time within which CH receives data from all sensors in the cluster. CH performs temporal association mining over DCCH and sends temporal frequent values along with floor and cluster ID to the sink.

The definitions of the concepts and correlation rules are given as follow:

Definition 1. Let DC_s is a set of time windows of a sensor to sense the environment, $DC_s = \{\lambda_{1,} \lambda_2 \ldots, \lambda_n\}$. The size of DCs is defined, $Size\ (DC_s) = |\lambda_n - \lambda_1|$, where λ_1 is the starting time window and λ_n is the ending time window.

Definition 2. Let V_i be a set of observed values by a particular sensor in the network, $V_i = \{v_1, v_2, \ldots v_n\}$. These observed values are stored in the buffer with the corresponding time windows λ_i in $DC_{s.}$

Definition 3. Let C_i *(Cluster)* be a set of sensors deployed in a room, office or hall, $C_i = \{S_1, S_2 \ldots S_n\}$. Where CH_i is a sensor (cluster head) belongs to C_i. The selection of CH_i among all the sensors in a cluster is defined later.

Definition 4. The support of rule is defined as the support of the observed value $(X \cup Y)$ from V_i respect to time windows λ_i in $DC_{s.}$, while the confidence of the rule is defined as

$$C(X \rightarrow Y) = Support\ (X \cup Y)\ /\ Support\ X$$

Lift of a rule is defined as

Table 1
Critical analysis table of proposed existing techniques.

Ref	Association Mining	Mining levels	Data mining Algo;	Data Association	Prediction	Threshold values		Data Routing	Contribution	Limitation
						Support	Confidence			
[7]	Distributed	Sink & node	Apriori	Temporal	No	0–90%	–	Single hop	Data Reduction	Additional memory
[8]	Distributed	Sink, CH & Node	Apriori	Temporal	No	10–90%	–	Multi hop	Data Reduction	Additional memory
[10]	Centralized	Sink	Tree based	Temporal and Spatial	No	10–90%	–	Multi hop	Data compression	Additional memory
[21]	Centralized	Sink	Tree based	Data attributes	No	–	60%	Multi hop	Min. node selection to monitor	Additional query messages
[11]	Distributed	Sink, CH & Node	Apriori	Data attributes	No	–	50–100%	Multi hop	Min. node selection to monitor	Spatio-temporal association missing
[19]	Distributed	Sink, CH & Node	Apriori	Temporal & data attributes	Yes	–	66–80%	Multi hop	Predicting missing value	Spatial association missing
[13]	Centralized	Node	Apriori	Data Attributes	No	–	–	Multi hop	Energy aware routing	Large query messages
[26]	Centralized	Sink	Apriori	Temporal	Yes	–	50%	Multi hop	Future movement of objects	Centralized mining is expensive
[27]	Centralized	Sink	Apriori	Temporal	Yes	–	60%	Multi hop	Future movement of objects	Centralized mining is expensive

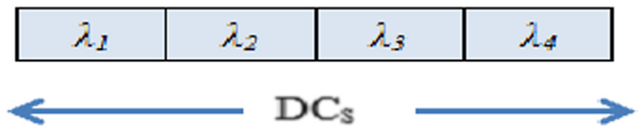

Fig. 3. Duty Cycle of sensing node (DCs) of over four window slots.

Lift ($X \rightarrow Y$) = *Support* ($X \cup Y$) / *Support* (X). *Support* (Y)

Definition 5. Let minimum support *(min_s)* and minimum confidence *(min_c)* are the threshold respectively at the sensor node, if and only if during same DCs, support >= min_s and confidence >= min_c than a rule $\lambda_i(X) \rightarrow \lambda_j(Y)$ is called as temporal association rule. Where X, Y ∊ V_i are the observed values by the sensor. X and Y will be called as temporal frequent sensed value at sensor.

Definition 6. Let SV_i = {*Temporal frequent sensed value*}, $SV_i \subseteq V_i$, be defined as a set of temporal frequent sensed value on specific DC_s to be sent to the CH by a single sensor.

Definition 7. Let SV_{all} be a set of observed values of all sensors in the cluster *(Ci)*, $SV_{all} = \{SV_1, SV_2, \ldots SV_n\}$, where n is the number of sensor nodes in the cluster. SV_{all} is defined as the set of frequent temporal values in a cluster.

Definition 8. Let minimum cluster head support (min_{CH_s}) and minimum cluster head confidence (min_{CH_c}) are the threshold respectively at the CH, if and only if during same DC_{CH}, support >= min_{CH-s} and confidence >= min_{CH-c} than a rule $S_i\ (SV_P) \rightarrow S_j\ (SV_Q)$is called as sensor temporal association rule at CH. Where SV_P, SV_Q ∊ EV_{set} *(Extracted Value set)*, $EV_{set} \subseteq SV_{all}$ (*of all sensor nodes in a cluster*), it is defined as a set of extracted values, sent by a particular CH to the sink.

3.2. Proposed energy conservation framework

In this section, the proposed HECF is presented as a framework design for the energy conservation of sensors applicable for multistory building in hierarchical manner. The objectives of the proposed framework as shown in Fig. 4 are:

- To reduce redundant data transmission of the sensed values from the sensor node to the CH by finding temporal correlations among the sensor node values. So it transmits only the temporal frequent values.
- To reduce the data transmission from the CH to the local sink, by finding the temporal correlations among the sensor nodes of the cluster in a given time window.
- To schedule sensor nodes (among which only a few would remain active) in specific clusters on the basis of temporal correlation rules mined at the sink and learns these rules for effective and efficient energy conservation of sensors at each floor or rooms of the building.
- To achieve global energy efficiency of the whole multistory building by setting smart energy efficient hierarchical sensor network.

As discussed earlier the Hierarchal Energy Conservation Framework is being proposed for multistory buildings. Where sensing nodes from each cluster transmits only temporal frequent sensed data to the cluster head, cluster head then transmit the sensor nodes only temporal frequent data to the local sink node. The local sink node is responsible for the generation and learning of temporal association rules to schedule the sensor nodes for the environmental monitoring. The benefit of local scheduling is that: efficient scheduling can be performed in time for energy efficiency of sensors locally, as less number of sensing nodes to be scheduled and prompt response to the sensing nodes will be available. Similarly, local sinks have direct energy supply and energy consumption of multi-hop data transmission from CHs to global sink is greatly reduced as the local sink nodes stand-in as a gateway node to the global sink with predefined routes. The temporal association rules are synced to the global sink for global view and scheduling of the whole building is performed floor wise. The sensor network is divided into network tiers, which are, sensor node level, cluster level and network level as shown in Fig. 4. The sensor nodes are geographically distributed and are divided into clusters. The formation of clusters is conducted by the sink node. Local sink performs cluster formation, by using the technique [29]. Cluster formation is done according to energy efficient manner by load balancing among the sensing nodes and it considers gateway routing. Energy efficient cluster formation and gateway routing are desirable for the proposed framework for overall energy efficiency.

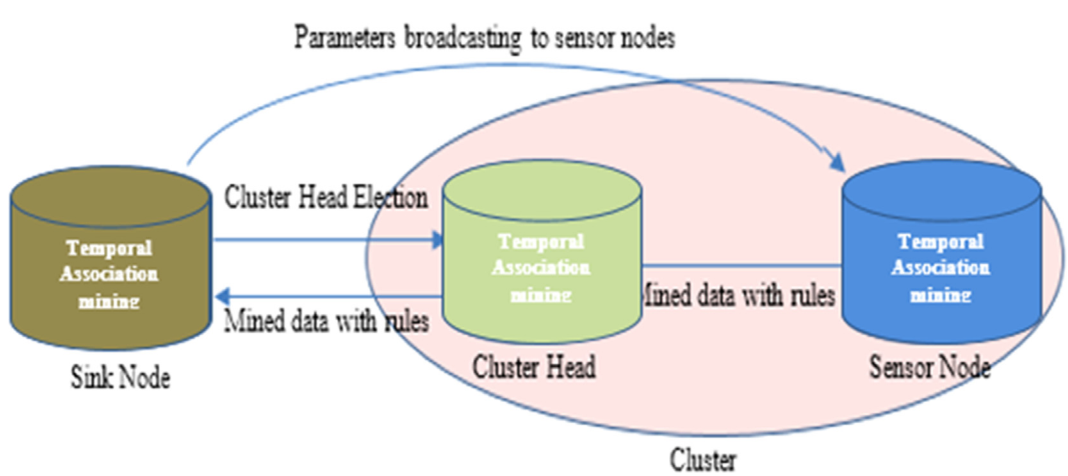

Fig. 4. Hierarchical Energy conservation framework (HECF).

After the formation of cluster, a sensor node is elected as a cluster head (CH) which will be responsible for receiving data. The process of election of CH in each cluster will be conducted by the sink node by using technique [19]. All sensor nodes will participate in the election process. However only that sensor node will be elected as a CH, which has minimum hop count to the local sink and maximum amount of energy in the cluster. Maximum amount of energy is required as all the sensor nodes will transmit data to the CH, and CH will transmit the mined data (where the redundant data were removed) to the sink. Additionally, each cluster will have secondary CH, in case the primary CH has reached the minimum energy level, secondary cluster head will take over the responsibility of cluster head as recommended. Similarly, minimum hop counts are desirable for reducing energy consumption, as this is the major cause of energy consumption of sensors.

Local sink node broadcasts the CH identity to all the sensor nodes of the corresponding cluster and along Duty_Cycle of sensing node (DCs). DCs is the duration of time, after the expiry of this duration; all sensor nodes will transmit temporal frequent sensed values according to the Definition.5 to the corresponding CHs as shown in Fig. 4. Likewise, local sink will also transmit Duty_Cycle cluster head (DCCH) to all the cluster heads on the respective building floors.

The purpose of defining DCCH for clusters is to mine frequent temporal sensed value of sensor nodes in a specific cluster according to the Definition.8. The hierarchal architecture helps in reducing the overall broadcasting of control and scheduling messages, as these messages will be broadcasted only to the clusters of particular floors instead of all the clusters in minimum hop transmission fashion.

The sensor nodes will sense the data in a defined window (λ) of DCs. Each sensor node will have a memory buffer to store the sensed data with respective assigned window. The size of the buffer is equal to the DCs as shown in Fig. 3. It stores all the sensed readings for a given DCs according to Definition 2.

Sensor nodes are capable of performing temporal association rule mining on the sensed data as mentioned in Definition.5. For example, a sensor node S_1 is sensing the sequence of readings; Medium, Medium, High, High, High in a given DCs as shown in Fig. 5. Sensor node S_1 will send data 'Medium' against the generated rule (i.e. $\lambda1 \rightarrow \lambda2$), similarly 'High' against the generated rule ($\lambda3 \wedge \lambda4 \rightarrow \lambda5$). As the sensed value ($SV$) has to meet the minimum support criteria to be sent to the CH. DCs of an arbitrary sensor (S_1) has been represented in the Fig. 5 as an example. S_1 is transmitting temporal frequent sensed value instead of transmitting redundant messages to the CH (according to the definition.5 and definition.6).

Algorithm.1 is proposed for mining of temporal frequent sensed values (SV) based on temporal association (among the time windows λ_i) at Sensor node. Algorithm.1 will mine the temporal frequent sensed value over a DCs based on the minimum defined support and confidence. The outcome of algorithm.1 is set of temporal frequent sensed values (SV_i) of a sensing node. After the expiry of each DCs, data from every sensing node of the cluster will be accumulated at the CH of the respective cluster with cluster_ID and Floor_ID.

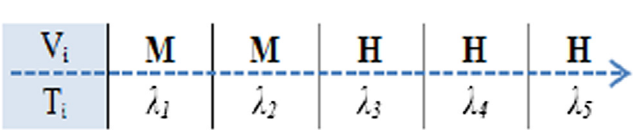

Fig. 5. DCs observed values of an arbitrary sensor node.

Algorithm 1: Mining frequent value (SV_i) based temporal association at Sensor node (S_i)

Input: Set of Observed values ($\boldsymbol{V_i}$) and set of corresponding total time Windows ($\boldsymbol{\lambda_i}$) in a **DCs**

Output: Set of temporal frequent sensed values ($\boldsymbol{S_V}$) along with time association to be sent to the cluster head

```
SVi={temporal frequent sensed values}
FOR λi = 1 to n
    IF [Vi+1] ==[ Vi]
        THEN temp = Vi+1
    IF support (temp) ≥ min_s & confidence(temp) ≥ min_c
        THEN [temp]ϵ SVSet
        ELSE Discard the value
END
RETURN SVi(λ)(along cluster ID and Floor ID)
```

All sensors of the cluster, have sent temporal frequent sensed values (SV_i) over their respective Duty cycles (DCS_i) as shown in Fig. 6. DCCH contains multiple DCs. After the expiry of DCCH, cluster head will mine the temporal association among the sensing nodes along with the temporally extracted frequent values (EV_{set}) according to Definition.8.

A set of extracted temporarily associated frequent values will be formed at the CH (EV_{set}). Algorithm 2 is proposed to mine the set of extracted frequent values (EV_{set}) from the sensing nodes of a specific cluster. The proposed algorithm.2 will take temporal frequent value (SV_i) of each sensor (S_i) to mine the extracted temporal frequent value (EV_{set}) of the CH.

The local sink node consists of four components data storage component, temporal association generation component, rules learning component and rule triggering component.

Algorithm 2: Creating a set of temporal frequent values (EV_{set}) along the correlation of Sensor nodes in the cluster

Input: Set of Observed values ($\boldsymbol{SV_{all}}$) of all sensors (**m**) in a cluster and corresponding DC_{si} in a DC_{CH}.

Output: Set of frequent values ($\boldsymbol{EV_{set}}$) along the correlation of Sensor nodes in the cluster

```
EVset1 ={frequent values along the association of Sensor
  nodes in the cluster}
FOR Si = 1 to m
FOR SVi = 0 to n
IF support ([Si][SVi]) ≥ minCH_s & confidence([Si] [SVi
  ] ≥ minCH_c
THEN [SVi][λ] ϵ EVset
ELSE Discard
END
END
RETURN EVset (along cluster ID and Floor ID)
```

F_ID	C_ID	Sensor	DC_{S1}	DC_{S2}	DC_{S3}	DC_{S4}…
1	2	S_1	SV_1=M	SV_2=L	SV_3=L	SV_4=L
1	2	S_2	SV_1=L	SV_2=L	SV_3=L	SV_4=M
1	2	S_3	SV_1=M	SV_2=M	SV_3=M	SV_4=H
1	2	S_4	SV_1=M	SV_2=M	SV_3=M	SV_4=H
1	2	S_5	SV_1=M	SV_2=L	SV_3=L	SV_4=L

Fig. 6. DC_{CH} of Temporal frequent sensed values (SV_i) of sensors at the CH over each DC_s.

Temporal association rule generation component generates the rules by employing proposed Algorithm 3.a. by using lift measure. The proposed algorithm 3.a will help to schedule the sensing nodes based on the generated rules in a specific cluster.

Rules learning component is designed to incorporate miss executed rules for energy conservation. The introduction of this component will help to understand the usage patterns of sensors. The prominent usage patterns will be highlighted with high priority value.

Algorithm 3.a: Temporal rules generations

1: *Rule_Set*(*C_ID*)={Temporal association Rules generation with global time (γ)}
2: FOR*Global_* γ = 0 **to** *n* (where *n* is total time Windows in a day)
3: *Compute* Lift (EV_{set})
4: IF ($Rule_i \geq Min$ Lift)
5: THEN
Rule$_i$ is valid
6: ELSE
Rule$_i$ discard
7: *Rule_Set*(*C_ID*) = $Rule_i$
8: END
9: RETURN *Rule_Set*(*C_ID*) (along cluster ID and Floor ID)

Rules are arranged in the learning component in ascending order with respect to priority (priority is based is based on Lift measure of the rule). This task is performed by the proposed Algorithm 3.b.

As shown in Table 2, the learning component has rules along with time duration and corresponding priority of each rule. The triggering component will trigger the rule when a rule meets the criteria. For example, if a rule is generated as (10:00 to 10:15 PM: $S1$ ^ $S3 \rightarrow S5$), when the global clock (γ) (at local sink) will have this time (10:00 to 10:15 PM), this rule will be triggered. If the sensing values of the sensing nodes and of the rule are same, then the rule is executed. As mentioned in the example; sensing nodes $S5$ and $S3$ will be turned to sleep mode by the local sink for this duration of time (15 min) to conserve energy and the priority of the rule will be incremented.

In case that the values of the sensing nodes are different on a specific global time (γ), the rule will not be triggered and the priority of the rule will be decremented. It is important to note that the validity of the rule is based on the priority (Lift measure for the proposed framework). If the priority of a rule is equal to or greater than a given threshold value, otherwise will happen.

Algorithm 3.b: Rules learning

1: Let *Rule$_i$_Priority* = Lift of rule
2: *Global_Time* (*G_T*) = Current Time assigning current time of day
3: *G_γi* = Temporal association rule (time value) of a specific cluster
4: IF ((*G_T* == *G_* γ*i*) && check the sensor values are valid w.r.t rule)
5: THEN *Rule$_i$_priority* ++
6: ELSE *Rule$_i$_priority* --
7: END

Similarly, local sink is responsible for the energy conservation of the whole floor, dealing with multiple clusters on a single floor.

Table 2
Learning component with rules priority.

S.No	Rules	Time	Priority
1	$S1 \rightarrow S2$	9:00 AM	94
2	$S2$^$S3 \rightarrow S1$	9:15 AM	92
3	$S2 \rightarrow S1$	9:00AM	87

At the global level, local sinks of each floor transmit temporal association rules along with the sensed data to the global sink via gateways to monitor and schedule all the floors from a single view point, i.e. the global sink. The global view is helpful in achieving a better global analysis, better understanding of the problem areas, and better scheduling of the sensor operations at the floors of the building and thus better energy efficiency.

4. Implementation & Simulation setup

In this section, implementation of the proposed model and simulation setup has been discussed. NS2 and MATLAB v.7 is used for the implementation, data extraction, network clustering and for result comparison in terms of network life time, message reduction and average energy consumption.

In the proposed HECF, we have considered multistory building of floor area 250 m × 250 m for each floor as shown in Fig. 2. Sensor nodes are grouped into clusters. The communication of the sensor nodes to cluster head is a single hop, as clusters are divided rooms wise.

There are eight living areas (Ground floor: Shopping Mart and Reception) (Floor #1: Store, Clinic and Gym) (Floor #2: Office-1, Office-2 and Conference room). Numbers of sensor nodes are deployed in the three stories building as described in the Table 4. Total of 54 sensor nodes have been deployed on each floor that monitor the atmospheric information (humidity, temperature, light and voltage) of each living area. The communication network architecture allows the sensor nodes sending data to the cluster heads of their respective clusters, cluster heads then send data to the gateway nodes, which in turn will send data to the sink.

For the proposed model, we have considered Mica 2 sensors for their better efficiency. Battery power model a sensor node (Mica 2) is defined in Eq. (1). Sensor nodes are powered by 2 AA batteries of 1.5 V. We have considered the realistic model of power supply of sensor nodes in terms of voltage by considering 2.7 V instead of 3 V. Eq. (1) gives the total amount of power that a sensor node has for performing different tasks in joule (J) energy unit.

$$Power = (CPU\ Current\_Active + Current\_Operation) \times Time \times Voltage \quad (1)$$

Current consumption model of Mica 2 Model is taken from the data sheet [30]. Time and current values of Mica 2 are given in Table 3. By using this data sheet as mentioned in Table 3, we have computed the results of the proposed framework. By using Eq. (1) the following power values for transmitting and receiving 1 bit of information are computed, as mentioned.

$$\boldsymbol{Power_{TX} = (8 + 20) \times 10^{-3}\ A \times 2.7\ Volts \times 416 \times 10^{-6} / 8\ bits = 4\ 06\ \mu J/bit}$$
$$\boldsymbol{Power_{RX} = (8 + 15) \times 10^{-3}\ A \times 2.7\ Volts \times 416 \times 10^{-6} / 8\ bits = 2.\ 34\ \mu J/bit}$$

While considering the communication distance from sensor nodes to cluster head and from cluster head to gateway node, the power consumption model for sensor nodes modifies to accommodate the communication distance.

Table 3
Mica.2 Current consumption model [30].

Mode	Current	Mode	Current
CPU		**Radio**	
Active	8.0 mA	Rx	7.0 mA
Idle	3.2 mA	Tx (-20dBm)	3.7 mA
ADC Noise Reduce	1.0 mA	Tx (-19dBm)	5.2 mA
Power-down	103 µA	Tx (-15 dBm)	5.4 mA
Power-save	110 µA	Tx (-dBm)	6.5 mA
Standby	216 µA	Tx (-dBm)	7.1 mA
Extended-Standby	223 µA	Tx (dBm)	8.5 mA
Internal-Oscillator	0.93 mA	Tx (+dBm)	11.6 mA
LEDs	2.2 mA	Tx (+dBm)	13.8 mA
Sensor board	0.7 mA	Tx (+dBm)	17.4 mA
EEPROM		Tx (+10 dBm)	21.5 mA
Read	6.2 mA	Communication range (868/916 MHz)	150 m
Read Time	565 µs		
Write	18.4 mA	Communication range (315/433 MHz)	300 m
Write Time	12.9 ms		

Table 4
sensor nodes distribution cluster wise.

S.No	Clusters	No. of Clusters
1	Shopping Mart	24
2	Reception	08
3	Store	12
4	Clinic	12
5	Gym	24
6	Conference Hall	12
7	Office-1	24
8	Office-2	24

$$ActualPower_{TX} = (CPUActivePower \times data) + (PTX \times Datasent \times Distance^2) \quad (2)$$

$$ActualPower_{RX} = (CPUActivePower \times datareceived) \quad (3)$$

By using Eqs. (2) and (3), transmission power (**Actual_Power $_{\mathbf{TX}}$**) and receiving power (**Actual_Power $_{\mathbf{RX}}$**) of sensor nodes are computed. Distance of a sensor node from other sensor node for transmitting of data is taken in to account in Eq. (2) and similarly amount of data to be sent is also considered. In Eqs. (2) and (3), CPU active power into data is the power consumed by a sensor node while processing unit of sensor node is active for any task related to data, regardless of whether transmission or receiving of data, reading or writing data in memory of data. PTX is the transmission power required for transmitting data messages. While distance is the communication distance between any two sensor nodes for data routing purpose. For our simulations, we have used fixed length data packet i.e. 1 Mbits information. The distance values varies according to the deployment their (x,y) coordinates that help to find the communication distance in a 250 × 250 grid area of each floor. The transmission and receiving powers are computed at (+10 dBm) noise level.

For the evaluation of the HECF, we have used sensor nodes data of wireless sensor network of Intel Berkeley Lab [31]. This dataset consists of 2.3 million sensed data readings. This dataset is produced from 54 sensing nodes. In the simulation experimentation setup 54 sensing nodes have been placed on each floor of the building. Sensed data has time stamped topological information along with the humidity, temperature, light and voltage readings. Reading duration for every new transaction is 31 s, i.e. Window size (λ). The DC_s for the proposed solution is four. All the environment monitoring readings are in numeric real format. Light sensed data reading is considered as homogenous data for HECF. As association rule mining cannot be performed on real values for this reason the dataset is preprocessed, discretization is applied by binning method. The real values are discretized into four bins with Very_-Low, Low, Medium and High values.

The simulation parameters are summarized in Table 5.

5. Results

The aim of this section is to analyze the simulation outcomes of the proposed HECF and compared with the exiting standard energy efficient RR-Schedule-Buffer [32] and LEACH-C [33]. The objectives of the results analysis are to understand the behavior of HECF over following energy consumption vital parameters of sensor network.

- Impact of frequency of dead sensor nodes
- Impact of number of cluster formation
- Impact of number of messages transmission
- Impact on sensor network life with and with HECF scheduling
- Impact on sensor network life for different values of Duty_Cycle of sensor
- Impact on sensor network life for different values of sensing window size of sensor

Since HECF is designed for hierarchical multi-hop transmission where it considers energy consumption of associated nodes within the cluster; residual energy balance and generated advance rules (using Lift) are employed for energy distribution balance, as in every round it improves the balance dynamically and CH and multi-hop routing can be calculated intelligently. It can be observed from Fig. 7 by applying the proposed HECF method (hierarchical transmission mode). HECF has employed data correlation mining approach at the sensor level to reduce the overall continu-

Table 5
Simulation parameters.

Simulation Parameters	Values
Simulation application	NS2 and MATLAB v.7
Network size (sensor nodes)	54 × 3 (each floor) = 162
Type of sensor nodes	Mica.2
Energy level of sensor nodes	PowerTX = 4.06 µJ/bit PowerRX = 2.34 µJ/bit
Building	Size of each floor = 250 m × 250 m Number of floors = 3
Window size (λ) sensor node	21 sec, 31 sec, 41 sec
Duty_Cycle (DCs)	2λs, 3λs, 4λs

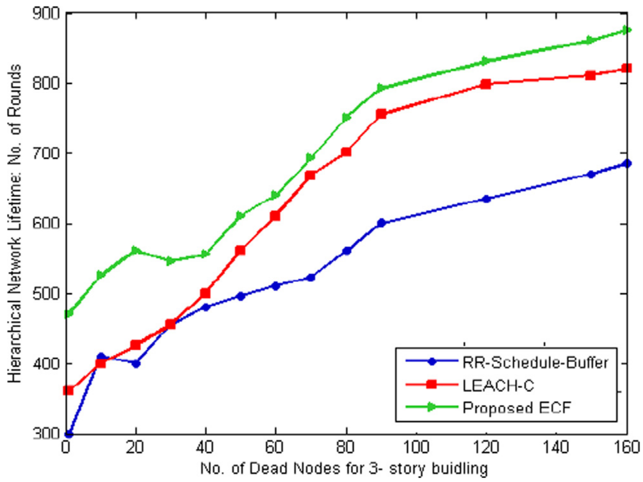

Fig. 7. Hierarchical network lifetime Vs. Number of dead nodes.

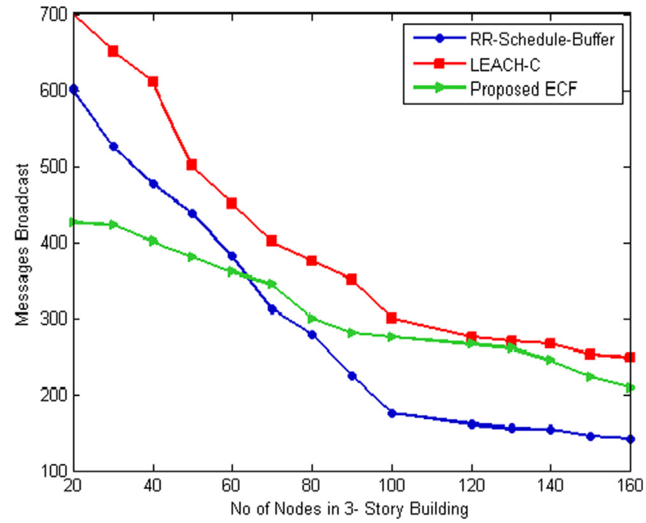

Fig. 9. Number of messages broadcast of multi-hop hierarchical transmission.

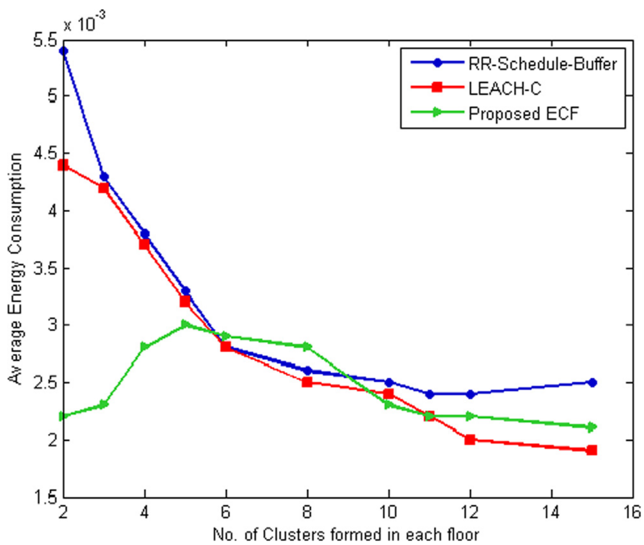

Fig. 8. Energy consumptions Vs Number of clusters.

ous message transmission to the CH which results in less energy consumption at the sensor level, thus increased overall network life. Similarly, CHs in HECF is not transmitting the redundant data to the sink (data correlation), thus saving CH's energy with less message transmission to increase the overall network life. Sensor data association mining at different tires in HECF not only helped in reducing the overall message transmission but also utilized in the scheduling (that is, sleep and wakeup) of sensors in network, thus further reduction in network message transmission. Thus for the same number of data transmission rounds, HECF has more live sensor nodes and hence longest network lifetime among the two methods.

Fig. 8 shows efficient energy consumption within each cluster during data extraction and transmission. HECF considered advance association rules for each cluster. Also multi hop routing is achieved in hierarchical fashion for all the three floors, keeping in mind that the usage, active modes, range of data values show less energy consumption and less number of messages that need to be broadcasted. All these contribute ultimately to extending the overall network lifetime as compared to other two methods.

In Fig. 9, it is observed that as the number of sensing nodes is increasing the number of broadcasting messages are decreasing. Relatively stable trend has been observed in the number of messages and number of sensing nodes increase for the proposed HECF. The stable trend is the behavior projection of HECF, as the message broadcasting has been controlled by the DC_S. Variations from linear trend is the effect of HECF scheduling, where different numbers of sensing nodes are active, as number of message broadcast depends on active sensors. HECF is monitoring the environment by scheduling and making less number of sensor nodes active for the said purpose thus conserving energy of sensor network.

In Fig. 10, energy consumption (in days) of each cluster is presented of the proposed HECF against when scheduling is performed on sensor network of a multistory building. The proposed solution considerably reduces energy requirement of sensing nodes in each cluster, thus increases the life of each cluster.

Selection of Duty_cycle of sensor (DC_S) has effects on the energy consumption of a cluster. The results in Fig. 11 represent that by the increase in the DC_S value more energy is conserved for each cluster but increase in the DC_S size will have negative impact on the accuracy of the monitored sensed values.

In this study, it is observed that when DCS = 3, it yields better energy conservation with acceptable level of accuracy of the sensed data than others. As the larger duty cycles will have more sensing windows, only the most frequent value will be sent over

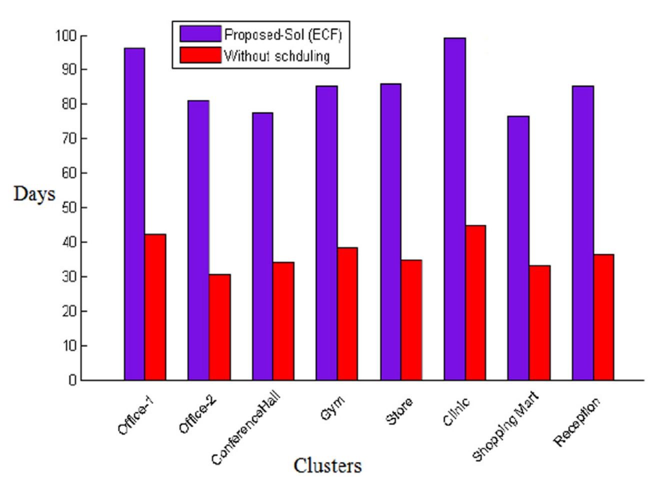

Fig. 10. Energy consumptions (in days) of each cluster.

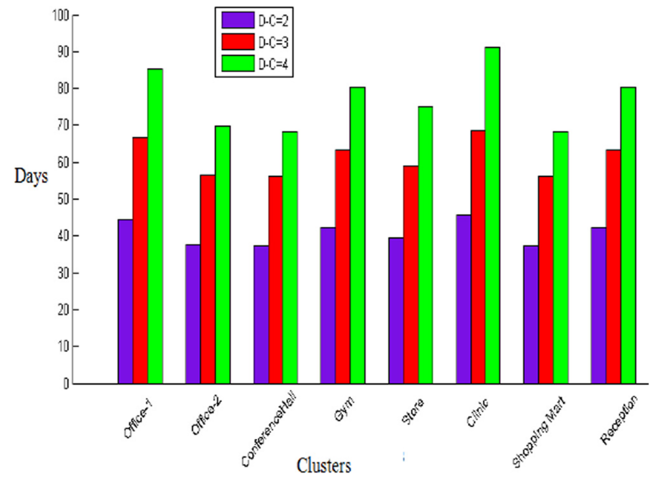

Fig. 11. Energy consumptions (days) by each cluster Vs Effect of different DC_S.

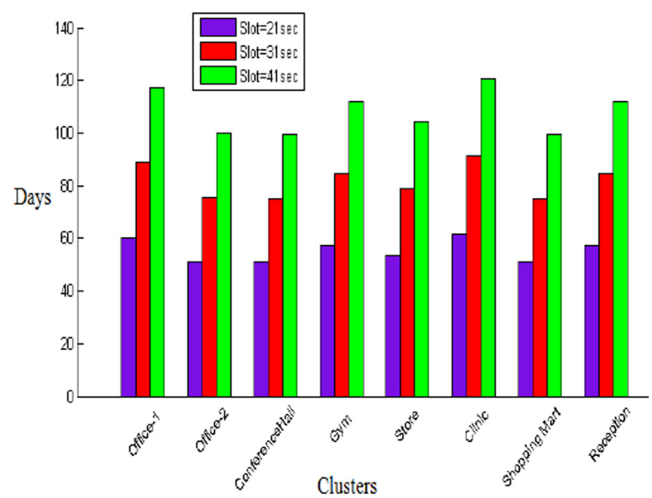

Fig. 12. Energy consumptions (days) by each cluster Vs Effect of different window sizes in DC_S.

that DCs and other sensed values will be ignored leading to less message transmission. While smaller DCs value is more accurate as in smaller number of sensing windows the sent frequent value will be most likely to be the same with other sensed values.

Fig. 12 illustrates the impact of the window size, the granular level of sensing time, on the proposed solution. It is observed that by increasing the window size more energy of the clusters can be conserved, as less number of message will be transmitted. However, having a larger window size will have negative impact on the accuracy of sensing data. As for the larger sensing window value the environment condition might vary. The window size should have an appropriate size with accuracy and less number of message transmissions. It is found that at window size = 31 sec, the proposed solution finds optimum tradeoff between window size and accuracy.

6. Conclusion

In this article, a framework for the energy conservation (HECF) for WSN is proposed. By using the concept of hierarchical data transmission and association rule mining at each level of sensor network, HECF largely reduces the energy consumption of multi-story building's sensor network and consequently prolonged the network lifetime. In HECF, sink node is capable of dividing sensor nodes into clusters on the basis of temporal and spatial associations among the sensor nodes. Additionally, it equips the sink node with storage component for contextual data and rules storage, learning component for optimum rules validity over the time and triggering component for the execution of rules when criteria is met. Association rules are generated by using Apriori algorithm along with advance rules validation measure such as Lift and network parameters for effective transmission. Simulation result shows that the proposed HECF achieves 16% better energy conservation as compared to traditional LEACH-C and RR-Schedule-Buffer methods. Additionally, HECF achieves 20% less number of messages during data transmission, which is a remarkable improvement for sensors energy conservation. However, HECF needs to be further evlauted over the tradeoff between the energy conservation of WSN and the accuracy of the sensed data.

It would be interesting to evaluate HECF in context-aware and complex enviorments such as smart health vicinities and smart cities. Another future direction is the mathematical modeling of the upper and lower bounds of the hierarchical energy conservation of WSN with respect to smart buildings, this would further standardize the proposed framework.

References

[1] El-Sayed WM, El-Bakry HM, El-Sayed SM. Integrated data reduction model in wireless sensor networks. Appl Comput Inf 2020. doi: https://doi.org/10.1016/j.aci.2019.03.003.
[2] Grindvoll H, Vermesan O, Crosbie T, Bahr R, Dawood N, Revel GM. A wireless sensor network for intelligent building energy management based on multi communication standards-a case study. J Inf Technol Constr 2012.
[3] Vo M-T, Tran V-S, Nguyen T-D, Huynh H-T. Wireless sensor network for multi-storey building: Design and implementation. In: Computing, Management and Telecommunications (ComManTel), 2013 International Conference on. p. 175–80.
[4] Chen Y-C, Hung H-C, Chiang B-Y, Peng S-Y, Chen P-J. Incrementally mining usage correlations among appliances in smart homes. In: *Network-Based Information Systems (NBiS), 2015 18th International Conference on*. p. 273–9.
[5] Sembroiz D, Careglio D, Ricciardi S, Fiore U. Planning and operational energy optimization solutions for smart buildings. Inf Sci 2019;476:439–52.
[6] Cuzzocrea A, Gaber MM, Shiddiqi AM. Distributed Classification of Data Streams: An Adaptive Technique. In: in *International Conference on Big Data Analytics and Knowledge Discovery*. p. 296–309.
[7] Samarah S, Boukerche A, Habyalimana AS. Target association rules: A new behavioral patterns for point of coverage wireless sensor networks. IEEE Trans Comput 2011;60(6):879–89.
[8] Das A, Das G. Distributed extraction and a novel association rule mining mechanism for WSN: An empirical analysis. In: in *Emerging Trends and Applications in Computer Science (ICETACS), 2013 1st International Conference on*. p. 252–5.
[9] Agrawal R, Imieliński T, Swami A. Mining association rules between sets of items in large databases. SIGMOD Rec. 1993;22(2):207–16.
[10] Boukerche A, Samarah S. In-network data reduction and coverage-based mechanisms for generating association rules in wireless sensor networks. IEEE Trans Veh Technol 2009;58:4426–38.
[11] Chong SK, Gaber MM, Krishnaswamy S, Loke SW. Energy conservation in wireless sensor networks: a rule-based approach. Knowl Inf Syst 2011;28(3):579–614.
[12] K DD, K Sonwalkar P. Analysis of sequential rule mining for energy conservation. Energy 2018;1(8). doi: https://doi.org/10.26808/rs.ca.i8v1.04.
[13] Abirami T, Priakanth P. Energy efficient wireless sensor network using genetic algorithm based association rules. Int J Comput Appl 2014;91.
[14] Muangprathub J, Boonnam N, Kajornkasirat S, Lekbangpong N, Wanichsombat A, Nillaor P. IoT and agriculture data analysis for smart farm. Comput Electron Agric 2019;156:467–74.
[15] Abediniangerabi B, Shahandashti S, Makhmalbaf A. A data-driven framework for energy-conscious design of building facade systems. J Build Eng 2020:101172.
[16] Ashouri M, Haghighat F, Fung BCM, Lazrak A, Yoshino H. Development of building energy saving advisory: A data mining approach. Energy Build 2018;172:139–51.
[17] Fujitsu. (2015). Fujitsu Develops Technology for Low-Cost Detection of Potential Sewer System Overflows. Available: http://www.fujitsu.com/global/about/resources/news/press-releases/2015/0210-03.html.
[18] Samijayani ON, Fauzi I, Zasyi M. Multi-Sensing Wireless Sensor Network for Smart Building System. Adv Sci Lett 2017;23:3660–4.

[19] Mahmood A, Ke S, Khatoon S. An Efficient Distributed Data Extraction Method for Mining Sensor Networks Data. arXiv preprint arXiv:1306.4069 2013.
[20] Tew C, Giraud-Carrier C, Tanner K, Burton S. Behavior-based clustering and analysis of interestingness measures for association rule mining. Data Min Knowl Disc 2014;28:1004–45.
[21] Zhu R. Intelligent collaborative event query algorithm in wireless sensor networks. Int J Distrib Sens Netw 2011;8:728521.
[22] Harb H, Makhoul A, Jaber A, Tawil R, Bazzi O. Adaptive data collection approach based on sets similarity function for saving energy in periodic sensor networks. Int J Inf Technol Manage 2016;15(4):346. doi: https://doi.org/10.1504/IJITM.2016.079603.
[23] Diwakaran S, Perumal B, Devi KV. A cluster prediction model-based data collection for energy efficient wireless sensor network. J Supercomput 2019;75:3302–16.
[24] Lasla N, Doudou M, Djenouri D, Ouadjaout A, Zizoua C. Wireless energy efficient occupancy-monitoring system for smart buildings. Pervasive Mob Comput 2019:101037.
[25] Wan R, Xiong N. An energy-efficient sleep scheduling mechanism with similarity measure for wireless sensor networks. Human-centric Computing and Information Sciences 2018;8(1). doi: https://doi.org/10.1186/s13673-018-0141-x.
[26] Samarah S, Al-Hajri M, Boukerche A. A predictive energy-efficient technique to support object-tracking sensor networks. IEEE Trans Veh Technol 2011;60(2):656–63.
[27] Tseng VS, Lu EH-C. Energy-efficient real-time object tracking in multi-level sensor networks by mining and predicting movement patterns. J Syst Softw 2009;82(4):697–706.
[28] Mahmood A, Shi K, Khatoon S, Xiao M. Data mining techniques for wireless sensor networks: A survey. Int J Distrib Sens Netw 2013;9:406316.
[29] Kuila P, Jana PK. Energy efficient clustering and routing algorithms for wireless sensor networks: Particle swarm optimization approach. Eng Appl Artif Intell 2014;33:127–40.
[30] Polastre J, Hill J, Culler D. Versatile low power media access for wireless sensor networks. In: *Proceedings of the 2nd international conference on Embedded networked sensor systems*. p. 95–107.
[31] I. B. R. lab. (2004, 28 April). Intel Lab Data. Available: http://db.csail.mit.edu/labdata/labdata.html.
[32] Singh P, Gupta RK. A novel data preparation approach for target based association rule mining. Wireless Sensor Network 2014;06(11):249–55.
[33] Shi S, Liu X, Gu X. An energy-efficiency Optimized LEACH-C for wireless sensor networks. In: *Communications and Networking in China (CHINACOM), 2012 7th International ICST Conference on*. p. 487–92.