www.ioirp.com

International Journal of Innovative Research in Computer Science and Engineering (IJIRCSE)
ISSN: 2394-6364, Volume – 1, Issue – 11. October 2015

SCALABLE LARGE SCALE WIRELESS SENSOR NETWORK USING DATA COUPLED CLUSTERING

R.JAYANTHI
PG Student, Department of Computer Science and Engineering,
Vivekanandha Institute of Engineering and Technology for Women,
Namakkal, TamilNadu.
jayanthicse37@gmail.com

R.RAJAGOPAL
Asst.Professor, Department of Computer Science and Engineering,
Vivekanandha Institute of Engineering and Technology for Women,
Namakkal, TamilNadu.
rajagopalrmail@gmail.com

***Abstract*— Self-organizing algorithms (SOAs) and Second-Order Data-Coupled Clustering (SODCC) algorithm for wireless sensor networks (WSNs) developed to increase the lifetime, to minimize unnecessary transmissions or to maximize the transport capaci ty. The general framework is formulated where the data from the resulting clusters ensures the well-posedness of the signal processing problem in the cluster. The second-order data-coupled clustering (SODCC) algorithm used second-order statistics. SODCC is scalable and has comparable or enhanced message complexity than other SOAs and likewise the performance of SODCC is improved at any compression rate and desires no prior adjustment of any parameter. SODCC associates well to former energy efficient clustering algorithms in terms of energy consumption.**

Keywords*— WSN, *SOA, SODCC, cluster algorithm, energy efficient, scalable, message complexity, performance, energy consumption.

I. INTRODUCTION

Wireless sensor networks (WSNs) aim at the correct recovery of the data measured by the sensors and gathered by a Data Fusion Center (DFC). A modest size and transmit policy is not suitable for a Large Scale Wireless Sensor Network (LS-WSN defined as having high amount of resource limited nodes and large datasets); it causes the network collapse due to interferen ce or blocking [13]. Therefore, to maintain the network operation, complex or even ingenious techniques are needed. By clustering a WSN different goals are sought:

- load balance,
- fault tolerance or
- Network connectivity.

The measured data gave the reason to exist of LS-WSNs and their recovery is of little importance to popular clustering algorithms. One of the alternatives is the data-driven clustering algorithms [3], where the decision criteria are based on characteristics of the measured data.

The self-organization of the network is a desirable feature for the WSN partitioning algorithm as no consolidated or peripheral entity is required. The early propositions of Self-Organized Algorithms (SOA) concentrated on maximizing the network lifetime since a limited energy budget, minimizing the amount of control message interchange and adapting to entrance or exit of sensor nodes into the WSN [6]. However, to the best of knowledge, no challenge to use the data correlations to initiative a self-organizing partitioning algorithm has been made. Concerning in-network processing, the disseminated source coding with rate-distortion control in the data compression is one of the most researched scattered processing techniques for WSNs. Even though it has been proposed for some years ago, its practical applications have been scarce.

For example, it has been applied to distribute video coding, but its main hindrance is the huge amount of inter-node mechanism communication generated. A dissimilar in-network strategy is used in where a distributed Principal Component Analysis (PCA) is realistic to data after clusters partitioned according to the *k*-means principle, wherever no information characteristics are attended [2]. In the present work the aim is to propose an algorithm that upturns the process efficiency (correctness of reassembled measurements and lifetime probability) of WSNs by means of both data-coupled clustering then in-network processing techniques. The core difference among the data-coupled and data-driven clustering algorithms is the usage of the restrained data. Though data-coupled algorithms partition the network by obeying with the necessities of the in-network processing algorithm to recover the measured field with the maximum possible reliability, data-driven algorithms do not abide to this constraint [1].

The additional condition has the outcome of redefining the ability limits by decreasing the quantity of transmitted packets at huge distances. Using the present proposal the network is able to operate more efficiently and the network lifetime is increased. The main contributions of this work are:

- The proposal of a general outline for the self organized data-coupled algorithms.
- The Second-Order Data-Coupled Clustering (SODCC) algorithm.

The practice of both proposed algorithms in LS-WSN authorities for the best network partitioning in terms of renovation quality of the sensed field and scalability [10].

A. Self Organized Data Coupled Algorithm

First, a general framework is proposed to combine both clustering also in- network processing algorithms and pursues a clustering configuration that helps the further processing. The Fast Subspace Decomposition (FSD) algorithm that estimates the dimension of the signal subspace of a dataset by using second-order statistics [5]. The FSD is used by the SODCC algorithm. The data coupled clustering algorithms use some characteristic of the sensed data as part of the decision criterion and guide for the network partitioning [12]. The clustering algorithm combines the network configuration to the identified data by forming clusters in which the in-network processing algorithm has a well-posed problem to solve.

B. Second Order Data Coupled Clustering

Self-organized data-coupled clustering algorithm presented that the algorithm and the sensed fields are coupled by means of a second order statistic. The goal of the clustering algorithm is that nodes through advanced spatio-temporal correlation between them belong to the same cluster [7]. The algorithm's result principle is constructed on the dimension of the signal subspace of the dataset gathered by collections of nodes. To confirm a well-posed correlation conditions in each cluster, the necessary and sufficient condition is that the indication subspace dimension wants to be subordinate that the number of nodes [9].

II. PROPOSED SYSTEM

Self-organizing algorithms (SOAs) for wireless sensor networks (WSNs) generally pursue to upturn the lifetime, to minimize unnecessary transmissions or to maximize the transport capacity. SOA defines that it is adjusted according to the shape and size of a network and it is the automation technology for planning, configuration, management and optimization. In this work, we formulate a general outline where the statistics from the resulting clusters ensures the well-posedness of the signal processing problem in the cluster.

Well-posedness defines that it has a solution and the solution is unique. Large Scale Wireless Sensor Network (LS-WSN) has high amount of resource limited nodes and large datasets. WSN aim at the correct recovery of data sensed by sensors and gathered by a Data Fusion Center (DFC) [4]. A simple algorithms and policy is not suitable for LS-WSN and there is network collapse due to interference or blocking. Therefore to maintain network operation, SOA and SODCC algorithms are used. WSN recover the data and SOA algorithm is used to self organize the cluster network. Self organization adjusted according to the size and shape of a sensor to form a network. Cluster head is created based on the characteristics of data and the data are coupled to its specified cluster head based on data coupled clustering.

Load balance defines that the distribution of workloads across multiple resources in network [8]. Fault tolerance allows a system to remain operating correctly in the event of failure of some of its components. Network connectivity is the process of connecting various parts of network to one another. The project is to develop the SODCC algorithm that uses second-order statistics. This shows that SODCC is accessible and has comparable or improved message complexity than other well-known SOAs. SODCC is scalability because of network is enlarged to accommodate the growth of data in each cluster head. If the cluster head was not able to accommodate the data, it is enlarged by using SODCC. And then, the data are gathered in DFC and distributed to particular sensor node.

The scalability is validated by connecting and transferring data between sensor nodes and also shows that the performance of SODCC is comparative to additional SOAs, improved at any compression rate and needs no prior adjustment of any parameter [11]. Finally, SODCC compares well to other energy efficient clustering algorithms in terms of energy consumption though surpassing in statistics reconstruction.

A. System Architecture

In system architecture, the sensor nodes collect the data and based on data coupled clustering, the data are coupled to its particular head node. And then, data are stored in data fusion center. Based on request of receiver, the data are distributed.

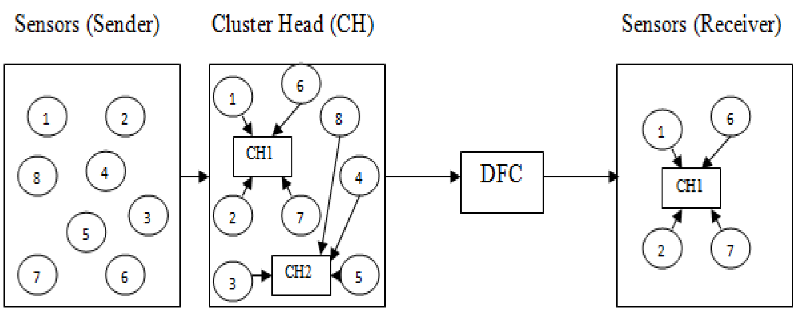

Fig.1. System Architecture

B. Data Flow Diagram

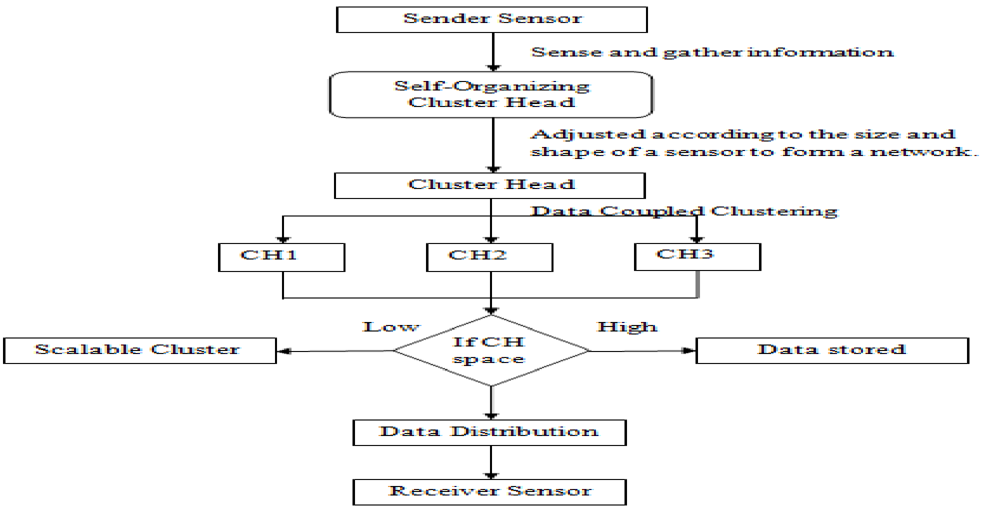

Fig.2. Data Flow Diagram

III. MODULE DESCRIPTION

- Cluster Self-organizing
- Data-coupled clustering
- Scalable cluster
- Data Distribution

A. Cluster self organizing

Ordered clustering is one technique for finding community structures in a network. The technique arranges the network into a order of groups according to a specified weight function. The data can then be represented in a tree structure recognized as a dendrogram. Ordered clustering can either be agglomerative or divisive depending on whether one proceeds over the algorithm by accumulation links to or eliminating links from the network, respectively.

Deliberate a LS-WSN with *N* role-free nodes (nodes that are neither CH nor sensor nodes). In this, role-free nodes decide randomly to turn into CH, with an independently and in a fully distributed fashion. The new CHs request their first neighbors (nodes that interact with a unique hop communication) to form share of their clusters; only role-free nodes answer.

ALGORITHM 1: CLUSTER INITIALIZATION

```
Random decision to turn into CH
If CH then
REQUEST to first neighbors (answer from role free nodes)
UNION message to selected neighbors
End if
```

Each CH assigns the role of sensor nodes to its selected neighbors and finishes the cluster initialization. Role-free nodes repeat Algorithm1 once a timeout $T1$ is reached. Once a larger timeout $T2 > T1$ is reached, all remaining role-free nodes turn into CH and cluster initialization is finished. Two key design decisions must be taken as

- Criteria to select neighbors and
- Maximum cluster size for the cluster initialization.

The extreme cluster size is required as no data structure is considered in this stage and the only aim is the initial cluster seeding. However, if any CH receives fewer answers from neighbors than the maximum allowed, that cluster is established with fewer nodes.

B. Data coupled clustering

The present work aims to propose an algorithm that increases the process effectiveness (correctness of reconstructed measurements and life expectancy) of WSNs by using both data-coupled grouping and in-network handling techniques. The key difference between the data-coupled and data-driven clustering algorithms is the usage of the restrained data. Though data-coupled algorithms partition the network by complying with the requirements of the in-network processing algorithm to recover the measured field with the highest possible fidelity, data-driven algorithms do not tolerate to this constraint. The supplementary condition has the effect of redefining the capacity limits by decreasing the quantity of transmitted packets at enormous detachments. Through the present proposal the network is able to operate more efficiently and the network lifetime is increased. The main contributions of this work are: the proposal of a general outline for the self organized data-coupled algorithms and the Second-Order Data-Coupled Clustering (SODCC) algorithm. The key for the growing phase of SODCC in Algorithm2 is as the cluster gathers data and it computes the dimension d. If FSD is unable to determine a dimension smaller than cluster size, an aggregation of neighboring clusters is directed by the CH; otherwise, the cluster breaks rising. Clusters with high spatio-temporal correlation estimate a dimension lower than the cluster size (d < N), i.e., the dimension of the blare subspace is at least 1 and signal and noise subspaces are distinguishable. In this instance, the cluster size fulfills the clustering algorithm convergence criterion. On the additional side, clusters with short spatio-temporal correlation do not meet the convergence condition i.e., signal and the noise subspaces are non separable. The cluster must grow in size (fusion with another cluster) to achieve the convergence principle; fusion is compulsory for the designated cluster, even if it has already met the convergence condition.

ALGORITHM 2: SECOND ORDER DATA COUPLED CLUSTERING

```
If CH then
N ← Cluster size
D ←dimension of cluster data
If d >= N then
    FUSION with selected CH; decision to choose neighbor CH
If neighbor CH then
    Gather all data from the CH; update N
End if
Else
Send data to CH
End if
```

The aggregation criterion impacts on the final cluster configuration. SODCC aims to increase the spatio-temporal correlation of the collected data to each cluster. This impartial is attained when the aggregation criterion seeks to, e.g., minimize the remoteness between CHs, decrease the cluster area; exploit the quality of the wireless channel between the CHs. SODCC is stretchy sufficient such that it permits additional restrictions to be included into the aggregation criterion (e.g., energy effectiveness). Node organization is not a condition for SODCC. Every node starts the algorithm just after it is switched on. If original nodes are combined to the LS-WSN, they start the procedure as new.

C. Scalable cluster

In this module extern cluster to make use nearby more similar cluster node. In this section, a self-organized data-coupled clustering algorithm is presented as the algorithm and the sensed fields are coupled by means of a second order statistic. The goal of the clustering algorithm is that nodes with higher spatio-temporal correlation between them belong to the similar cluster.

ALGORITHM 3: SCALABLE CLUSTER

```
If Network then
N ← Network size
D ←dimension of cluster head
If d >= N then
    Decision to enlarge network
    Gather sensed data to form CH; update N
End if
```

The algorithm’s choice principle is based on the dimension of the signal subspace of the dataset assembled by collections of nodes. To confirm a well-posed association matrix in each cluster, the necessary and sufficient condition is that the indication subspace dimension desires to be subordinate that the number of nodes. To assess the scalability of SODCC in a LS-WSN scenario, the average value of the cluster size (N) is computed to check whether if SODCC is scalable.

D. Data Distribution

The Data Distribution Service for Real-Time Systems (DDS) is a leading data-centric publish and subscribe grace of infrastructures standard from the Object Managing Group (OMG) for the Internet of Things (IoT) and Industrial Internet of Things. The DDS specification describes two levels of interfaces:

- A lower DCPS (Data-centric publish-subscribe) smooth that is besieged towards the efficient distribution of the proper information to the proper recipients.
- An optional higher DLRL (data local restoration layer) flat, which permits for a simple integration of DDS into the application layer

DDS handles transfer chores: message addressing, data marshalling and demarshalling (so subscribers can be on dissimilar platforms from the publisher), distribution, stream control, retries, etc. Any node can be a publisher, subscriber, or both simultaneously. DDN is that the live conversational data is cached in real time rather than content cached at ordered periods so it's classically much slighter and much more up to date. The data is cached in a hierarchy of topics to allow for easy subscription to subsets of data. The data comes from an originating server, typically called data foundations, such as a catalogue or an initiative service bus. Instead of requesting the data, applications (used by customers, workers, machineries) contribute to the data. If data is previously cached, the end user or machine will get the current version (or formal) of the information and then any consequent informs are pushed as the data changes. An example of where a DDN is more suitable is for firm altering data where Internet or movable connectivity is unreliable or bandwidth is unpredictable. If data is altering rapidly, say for instance financial data or sporting odds, and end user loses connectivity, upon reconnection, DDN will have cached the data worker are subscribed to and it will be sent only the data that are missed at the similar time as lost the connection, rather than needing to load the whole page or application over.

IV. CONCLUSION AND FUTURE WORK

In this work, the general framework is proposed for LS-WSN that uses a combination of network partitioning and clustering algorithms to achieve a self-organized as well as data-coupled network configuration. The criterion to form the clusters confirms the well-posedness of the in-cluster signal processing. To demonstrate the general framework the clustering algorithm which is based in a second order statistics of SODCC and SOA. The performance of SODCC message complexity is compared and also showed that SODCC is scalable and has similar or better message complexity than other well-known self-organizing algorithms.

SODCC has the limitation of lack of adaptability and lack of support for multiple sensors. If the cluster head will unable to enlarge its adaptation, the compression technique is used to adapt. The performance and other parameters also improved as much as compared to SODCC. Hence, the future work is to overcome these drawbacks with the help of Distributed Compressive Projections Principal Components (DCPCCA).

References

[1] Caione.C, Brunelli.D and Benini.L (2012), 'Distributed Compressive Sampling for Lifetime Optimization in Dense Wireless Sensor Networks', IEEE Trans. on Industrial Informatics, vol. 8, no.1, pp. 30–40.

[2] Cidean M.I, Morgado.E, Ramiro-Bargueño.J, and Caamaño.A.J (2013), 'Self-organized distributed compressive projection in large scale wireless sensor networks', in Proc. IEEE 24th PIMRC, pp. 2000–2004.

[3] El Gamal.H (2005), 'On the scaling laws of dense wireless sensor networks: the data gathering channel', IEEE Trans. In. Theory, vol. 51, no. 3, pp. 1229–1234.

[4] Fowler J.E (2008), 'Compressive-projection principal component analysis for the compression of hyperspectral signatures', in Proc. IEEE Data Compression Conf., Snowbird, UT, pp.83–92.

[5] Fowler J.E (2009), 'Compressive-projection principal component analysis', IEEE Trans. Image Process., vol. 18, no. 10, pp. 2230–2242.

[6] Fowler J.E (2009), 'Compressive-projection principal component analysis and the first eigenvector', in Proc. IEEE Data Compression Conf., J. A. Storer and M. W. Marcellin, Eds., Snowbird, UT, pp. 223–232.

[7] Franceschetti.M, Dousse.O, Tse.D.N.C and Thiran.P (2007), 'Closing the gap in the capacity of wireless networks via percolation theory', IEEE Trans.Inf. Theory, vol. 53, no. 3, pp. 1009–1018.

[8] Gedik.B, Liu.L and Yu.P.S (2007), 'ASAP: An Adaptive Sampling Approach to Data Collection in Sensor Networks', IEEE Transactions on Parallel and Distributed Systems, vol. 18, no. 12, pp. 1766–1783.

[9] Gupta.P and Kumar P.R (2000), 'The capacity of wireless networks', IEEE Trans. Inf. Theory, vol. 46, no. 2, pp. 388–404.

[10] Heinzelman.W, Chandrakasan,A and Balakrishnan. H (2002), 'An Application Specific Protocol Architecture for Wireless Micro sensor Networks', IEEE Transactions on Wireless.Comm, vo1 (4), pp.660-670.

[11] Hung C.C, Peng W.C and Lee W.C (2012), 'Energy-aware set-covering approaches for approximate data collection in wireless sensor networks', IEEE Trans. Knowl. Data Eng., vol. 24, no. 11, pp. 1993–2007.

[12] Jia.X and Richards J.A (1999), 'Segmented principal components transformation for efficient hyperspectral remote-sensing image display and classification', IEEE T.G.Remote Sens., vol. 37, no. 1, pp. 538–542.

[13] Lathauwer.L L.D, Moor B.D and Vandewalle.J (2001), 'Independent component analysis and (simultaneous) third-order tensor diagonalization. Signal Processing, IEEE Transactions on, 49(10):2262– 2271.

[14] Lee J.B, Woodyatt.S, and Berman.M (1990), 'Enhancement of high spectral resolution remote-sensing data by a noise-adjusted PCT', IEEE T.G.Remote Sens., vol. 28, no. 3, pp. 295–304.

[15] Xu.G and Kailath.T (1994), 'Fast subspace decomposition', IEEE Trans. Signal Process, vol. 42, no. 3, pp. 539–551.