© 2014 IJEDR | Volume 3, Issue 1 | ISSN: 2321-9939

Energy Efficient Distributed Data Compression in Wireless Sensor Networks

[1]Neha Verma, [2]Manju Kumari

[1]M.Tech Student,[2]Assistant Professor
[1] Department of Electronics and Communication,
[1]YMCA University of Science and Technology
Faridabad- 121006, Haryana, India

***Abstract* – For wireless sensor networks with a large number of energy constrained sensors, it is very important to find a method to organize sensors in clusters to minimize the energy used to communicate information from all nodes to the processing center. In this paper, we look at data compression method which can have significant impact on the overall energy dissipation of these networks. Based on our findings that the conventional direct transmission uses much more minimum-transmission-energy than cluster based transmission, and spatial correlation may not be efficient for large sensor networks, we propose in-network compression and optimum clustering to be implemented together for large WSNs.**

***Index Terms* – Optimum Clustering; Sensor Networks; Spatial Correlation; Slepian-Wolf Coding; Data compression.**

I. INTRODUCTION

Wireless sensor networks consist of hundreds to thousands of low-power multifunctioning sensor nodes, operating in an unattended environment, with limited computational and sensing capabilities. In order to take advantage of these of wireless sensor nodes, we need to account for certain constraints associated with them. In particular, minimizing energy consumption is a key requirement in the design of sensor network protocols and algorithms. Since the sensor nodes are equipped with small, often irreplaceable, batteries with limited power capacity, it is essential that the network be energy efficient in order to maximize the life span of the network [6, 7].

Since a large number of low-power nodes have to be networked together, conventional techniques such as direct transmissions from any specified node to a distant base station have to be avoided. Clustering and generating a sink node can be benificial but as the hops from one cluster to another is increases the energy efficiency decreases. On the other hand, utilizing a conventional multihop routing schemes will also result in an equally undesirable effect. Thus a novel approach of combining compression and routing is required at both clustering and in-cluster environment.

II. PROBLEM FORMULATION

The basic idea behind this work is study the major problems in applying energy efficient coding for data aggregation in cluster-based WSNs with an objective to optimize data compression so that the total amount of data in the whole network is minimized and to estimates the number of clusters needed to efficiently utilize data correlation of sensors for a general sensor network.

To improves the energy utilization of sensor network, we need to maximizes the network lifetime and make the wireless sensor network fault tolerant to some extent.. Once deployed it is often difficult to charge or replace the batteries for these nodes. The capacity of batteries is not expected to improve much in the future[9]. In this paper we explore energy consumption trade-offs associated with lossless data compression. The data compression techniques extend the life time of sensor network. Also by reducing data size less band width is required for sending and receiving data.

III. DISTRIBUTED SOURCE CODING

Considering a wireless sensor network in which many sensor nodes sense a common event independently, and send their sensed data (readings) to a base station to do information processing. Since their readings usually are highly correlated, that is to say much redundancy exists in their readings. It is straightforward that they can avoid send many redundant information to base station if the sensor nodes can communicate each other. However, in some cases, it is difficult for all sensors to communicate each other; in the cases that sensors can communicate each other, and then communication among sensor nodes consumes energy, which usually is the most significant issue in designing wireless sensor networks. [5]

So here comes the idea: can the redundancy be reduced without the direct communication among sensor nodes in the networks? The answer is yes after Slepian and Wolf proposed the famous Slepian-Wolf coding theorem in 1973. This paper is the theoretical base of distributed source coding (DSC) compression. **Slepian-Wolf coding** is a distributed source coding technique that can completely remove data redundancy without requiring inter-sensor communication and therefore a promising technique for data aggregation in WSNs [4].

Moreover, DSC has been used in video compression for applications which require low complexity video encoding, such as sensor networks, multiview video camcorders, and so on. With deterministic and probabilistic discussions of correlation model of two correlated information sources, DSC schemes with more general compressed rates have been developed. In these non-asymmetric schemes, both of two correlated sources are compressed[1,2].

Consider X to be a discrete random variable taking values in the set X = {1,2,...,M}.
Denote the probability distribution of X by:
PX(x) = Pr[X = x],x is element of X.
Now,let X = { X1,X 2,……X n} be a sequence of n realizations of X,so that the probability distribution for the random n - vector X is given by: PX(x) = Pr[X = x] = product from i= 1 to n, PX(xi), x =(x1,x2………xn) element of Xn.

We regard X as a block of n successive characters from the output of an information source producing characters independently with letter distribution PX(x). In a typical long block, we have letter 1 occurring nPX(1) times, letter 2 occurring nPX(2) times etc. The probability of such a long typical sequence is, therefore, $P_T = P_X(1)^{nPx(1)} \ldots . P_X(M)^{nPx(M)}$

$= \exp[\, nP_X(1)\log P_X(1)] \ldots . \exp[\, nP_X(M)\log P_X(M)] = \exp[-nH(X)]$

We define these exp[nH (X)] to be the typical sequences and these constitute the set of sequences, is most likely to occur. Each of these typical sequences is equally likely and occur with probability exp[−nH (X)].
Where, $H(X) = -\sum_1^M P_X(i) log P_X(i)$

Under a certain deterministic assumption of correlation between information sources, a DSC framework in which any number of information sources can be compressed in a distributed way has been demonstrated by X.

IV. IMPLEMENTING SLEPIAN WOLF CODING

Consider the case of two random sources X1 and X2 that are correlated as shown in fig. 1. Intuitively, each of the sources can code their data at a rate greater or equal to their respective entropies R1= H(X1), R2 = H(X2) respectively. If they are able to communicate, then they could coordinate their coding and use together a total rate R is equal to the joint entropy,
R1+ R2 =H(X1,X2).

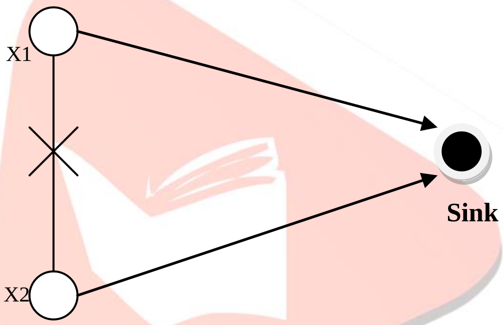

Fig. 1: Two correlated sources X1 & X2 send data to one sink.

This can be done, for instance, by using conditional entropy, that is, R1 = H(X1) and R2 = H(X2|X1), since X1 can be made available at node 2 through explicit communication. Slepian and Wolf showed that two correlated sources can be coded with a total rate H(X1,X2) even if they are not able to communicate with each other. Figure 2. shows the Slepian–Wolf rate region for the case of two sources.

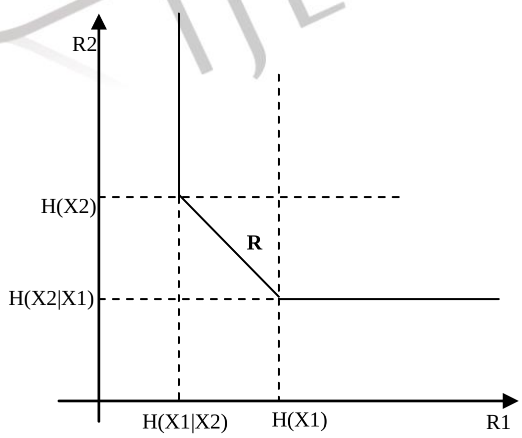

Fig. 2: Slepian-Wolf region for rate allocation of correlated sources.

This result can be generalized to the N-dimensional case. Consider a network consisting of N sensor nodes uniformly distributed in a region of interest, where each node i produces reading Xi and all the readings constitute a set of jointly ergodic sources denoted by: $X = (X1, X2, \ldots\ldots\ldots\ldots, XN)$ with distribution $p(X1, X2, \ldots\ldots\ldots\ldots\ldots.., XN)$ which corresponds to the spatial correlation structure known by each node a priori.

According to Slepian-Wolf Theorem, the nodes can jointly encode their data without inter-node communication, with a rate (in bits) lower-bounded by their joint entropy:
$H(X1, X2 \ldots\ldots\ldots\ XN)$ as long as their respective rates are under the constraints given by,
$R(U) \geq H\big(X(U)\big|X(UC)\big)$ For all $U \subseteq \{1,2 \text{ to } .N\}$ (1)
where {1,2 to N} is a set of indices of sensor nodes in the network, UC is the complementary set of U, H(X) is the entropy of X and $R(U) = \sum Ri$ for $i \in U$ (2)
$X(U) = \{Xj | j \in U\}$ (3)
For example, consider a simple case of two sensor nodes producing readings X1 and X2 . Their individual rates should be subject to:
$R1 \geq H(X1 | X2)$
$R2 \geq H(X2 | X1)$
$R1 + R2 \geq H(X1, X2)$

According to chain theory[10], under the above constraints, it is always possible to find a rate allocation for the two nodes, which makes the total rate (bits) of two nodes equal to their joint entropy, e.g.,
$R = R1 + R2 = H(X1) + H(X2 | X1)$ (4)

In general, for an arbitrary ordering of N nodes (e.g., in the ascending or descending order of nodes' ID numbers), there exits a rate allocation (vector) {Ri},
Where i = 1,2…….N such that the number of generated bits from all nodes can achieve the value of their joint entropy, e.g., $Ri = \sum_{i=1}^{N} H(X1, X2 \ldots.., XN)$
Where $R1 = H(X1)$ (5)
$\therefore\ Ri = H(Xi | X\,i-1, X\,i-2, .., X1),\ 2 \leq i \leq N$ (6)

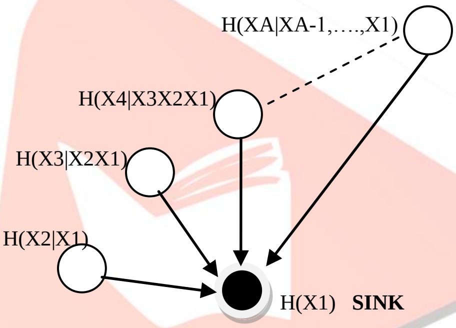

Fig. 3: Implementation of Slepian–Wolf Coding within a cluster

Therefore, a cluster of nodes A can be encoded with H (X1,X2, ….., X|A|) bits using Slepian-Wolf coding without communicating with each other, and there always exits an optimal rate allocation to achieve this local maximum compression performance as shown in figure 3. Thus we see from above discussion that Slepian and Wolf presented a surprising result. It seems strange;but it works.

V. SIMULATION RESULTS AND ANALYSIS

We consider a network with N sensor nodes uniformly deployed in a 100m×100m sensing region and the data sink located at the left-bottom edge of the region. The simulation results are based on the various experiments and each experiment uses a different randomly-generated topology.

a. The correlation structure

For the correlation structure, we assume that the observations $X_1, X_2, \ldots\ldots, X_N$ at N sensor nodes are modeled as an *N*-dimensional random vector $X = [X_1, X_2 \ldots.., X_N]^T$, Which has a multivariate normal distribution with mean zero and covariance matrix K ,i.e., the density of X is

$f(X)=\exp((-1/2).X^T.K^{-1}.X)/(\sqrt{2}\pi).\,|K|^{1/2}$ (7)

And the differential entropy of ($X_1, X_2 \ldots.., X_N$) is
$H(X_1, X_2 \ldots.., X_N) = (1/2)\log[(2\pi e)^N.\det(K)]$bits (8)

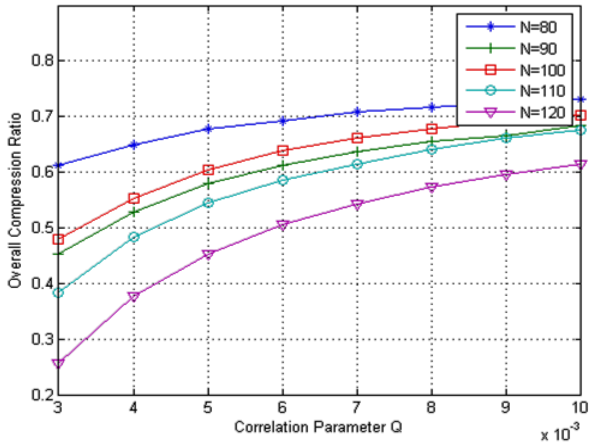

Fig. 4: Impacts of the extent of correlation & network size on overall compression ratio

Figure 4. shows the effects of the extent of correlation and the network size on the compression performance. The compression performance is measured in an overall compression ratio, which is the total amount of data produced in the network after clustered Slepian-Wolf coding is applied over the total number of bits generated by all nodes without using this distributed source coding scheme.

Let the observations $X_1, X_2,\ldots\ldots,X_N$ at N sensor nodes are has equal probability ,then each of the source code their data with $\log_2 N$ bits without using this distributed source coding scheme. The network size or the total number of sensor nodes N is set to be {80, 90, 100, 110, 120}. The parameter θ in the covariance model is set to be {0.01, 0.009,……………, 0.003 } where θ = 0.01 indicates low correlation and θ = 0.003 indicates high correlation.

In the case of high correlation, a better compression performance is achieved because the Slepian-Wolf coding can remove more redundancy caused by the high spatial correlation among the readings of different sensor nodes. In addition, the compression performance is improved as the network size or the density of sensor nodes increases. This behavior is due to the fact that the denser sensor deployment results in more sensor nodes residing within cluster while Slepian-Wolf coding can completely get rid of the highly redundant data generated by these sensor nodes in closer proximity to each other.

b. The optimal no. of sensors:

The figure 5. shows the effect cluster size on optimum number of sensors and the relattion used is:

Sopt = ((2*(N^1/2)*(1-Q))/Q)^2/3.

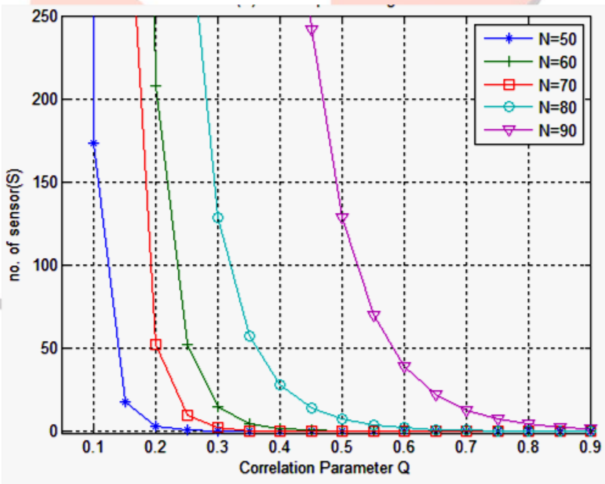

Fig. 5: Analytical curves for number of sensor S and correlation parameter Q

It shows that low correlation levels require small cluster sizes and high correlation levels require larger cluster sizes. Also the number of clusters is depending upon to sensor network size.

c. The optimal number of clusters

The optimal number of cluster depends on the number of sensors in the entire sensor network and the degree of correlation.

It can be expressed as : $\mathbf{K_{opt}} = \frac{\mathbf{N}}{\mathbf{S_{opt}}}$ (9)

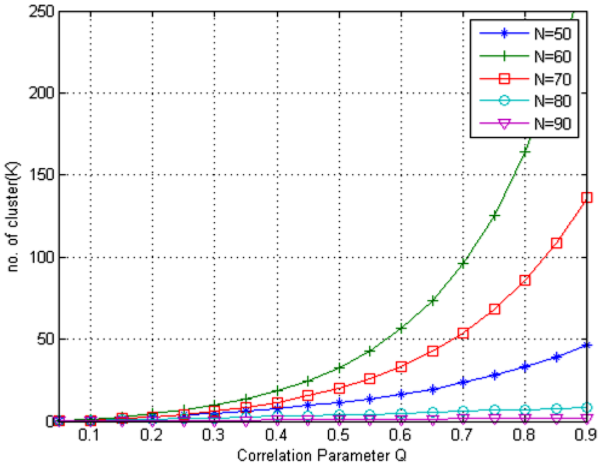

Fig. 6: Analytical curves for number of cluster k and correlation parameter Q

Figure 6. shows how different number of clusters and cluster size perform across a range of correlation levels and sensor network sizes. As expected low correlation levels require small cluster sizes while high correlation levels require larger cluster sizes. Also the number of clusters is relative to sensor network size. In other words, the larger the sensor network (N) the more clusters are required to apply optimal data aggregation.

d. Performance of optimal intra-cluster rate allocation

We now investigate the performance of optimal intra-cluster rate allocation with respect tthe intra-cluster communication cost of a cluster of different network size.

Figure 7. shows the intra-cluster communication cost with the optimal rate allocation and rate without using this distributed source coding scheme, respectively. The optimal rate allocation scheme first arranges nodes in cluster A in ascending order of their distance to the cluster head, thus the rate assigned to the node i can be expressed by

$$Ri = \sum_{i=1}^{N}\sum_{j=1}^{N} H\left(Xi \,\middle|\, \{Xj \mid d(j,1) \leq d(i,1), j \in A\}\right) \qquad (10)$$

Intra-cluster communication cost with the optimal rate allocation $=\sum_{i=1}^{|A|} Ri.\, d(i,1)$

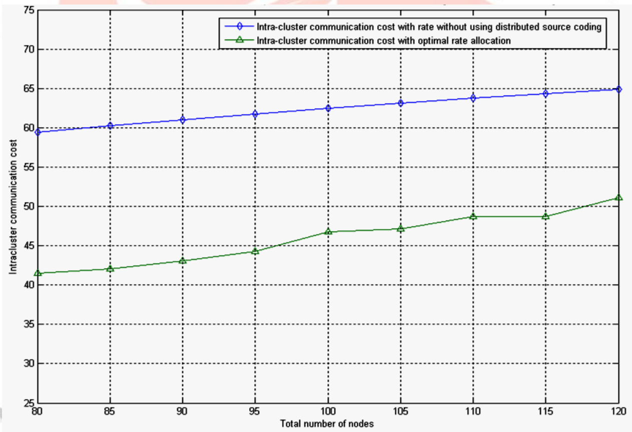

Fig. 7: Intra-cluster communication cost with optimal rate allocation and rate without using distributed source coding

We used parameter $\theta = 0.006$ to model moderate spatial correlation. The result shown is an intra-cluster communication cost of a cluster with different network sizes. As expected, the optimal intra-cluster rate allocation results in less communication cost compared with rate without using this distributed source coding scheme because former scheme jointly considers rate assignments and transmission distances between the cluster members and the cluster head.

e. Effect on overall compression ratio with normal distribution

For the correlation structure, we assume that the observations $X_1, X_2, \ldots\ldots, X_N$ at N sensor nodes are modeled as an *N*-dimensional random vector $X = [X_1, X_2 \ldots\ldots, X_N]^T$, Which has a multivariate normal distribution with mean zero and covariance matrix K ,i.e., the density of X is $f(X) = \exp((-1/2).X^T.K^{-1}.X)/(\sqrt{2\pi}).|K|^{1/2}$

The aim to find out the effect on overall compression ratio with normal distribution is to get the amount of data distributed on a network when we apply the routing and compression techniques to our data and sensor node. This is shown by the Figure 8.

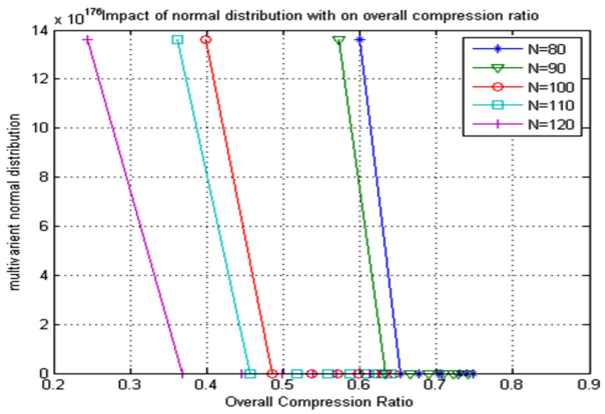

Fig. 8: Impact of normal distribution with on overall compression ratio

Overall we interpret that if we compress the data using slepian wolf coding is used the data distribution will reduced with increasing N. Also if node localization is implied, the cost and energy at each node will also be minimzed due to reduction in ideal nodes and removal of inactive nodes in WSN.

VI. CONCLUSION

The most important factor of designing WSN is how to improve the energy efficiency. The distributed source coding scheme will significantly decrease the whole network energy consumption by decrease the complexity of sensor encoder and compress the amount of transmitted data. The simulation results demonstrate that the clustered Distributed Source Coding can significantly reduce the total amount of data in the whole network while the transmission cost within cluster can be remarkably reduced by performing the optimal intra-cluster rate allocation. We also interpret that if we compress the data using slepian wolf coding the data distribution will reduced with increasing N.

We have proposed a distributed algorithm for organizing sensors into a hierarchy of clusters with an objective of minimizing the total energy spent in the system to communicate the information gathered by these sensors to the information-processing center. We have found the optimal parameter values for these algorithms that minimize the energy spent in the network. In a contention-free environment, the algorithm has a time complexity of O (k1 + k2 +… + kh) which is significant improvement over the many O (n) clustering algorithms available and thus makes this new algorithm suitable for networks of large number of nodes.

VII. ACKNOWLEDGMENT

The authors would like to thank everyone in department of electronics and Communication Engineering, YMCA for their support, creative criticism and encouragement.

REFERENCES

[1] Heshun Ouyang; Yi Yang; Yubo Deng; Wenqiang Guo, "Energy efficient data compresion algorithm and strategy for data uploading in Wireless Sensor Networks" , Automatic Control and Artificial Intelligence Journal (ACAI 2012), pp. 991 – 994, March 2012.

[2] Chih-Chung Lin, "A novel data compression method using improved JPEG-LS in wireless sensor networks", Advanced Communication Technology (ICACT)'10 Journal, pp. 346 - 351Volume:1, 2010.

[3] Tavli, B; Bagci I . E. ; Ceylan O. "Optimal data compression and forwarding in wireless sensor networks", in Proc. of Communications Letters, IEEE, pp. 408 - 410 May 2010.

[4] Meng-Han Li; Chung Lin; Chi-Cheng Chuang; Ray-I Chang "Error-Bounded Data Compression Using Data, Temporal and Spatial Correlations in Wireless Sensor Networks" Multimedia Information Networking and Security (MINES) Journal, pp. 111 - 115 , Nov 2010.

[5] Jun Zheng, Pu Wang and Cheng Li, "Distributed Data Aggregation Using Slepian–Wolf Coding in Cluster-Based Wireless Sensor Networks", in Proc. of IEEE INFOCOM'10, June 2010.

[6] Pu Wang, Rui Dai and Akyildiz, I . F. "Collaborative Data Compression Using Clustered Source Coding for Wireless Multimedia Sensor Networks", in Proc. of IEEE Communications Letters, pp. 210-219, March 2010.

[7] R. Rajagopalan and P.K. Varshney, "Data-aggregation techniques in sensor networks: a survey," IEEE Comm. Surveys & Tutorials, vol. 8, no. 4, pp. 48–63 , 2006.

[8] F. Akyildiz et al., "Wireless Sensor Networks: A Survey," Elsevier Sci. B. V. Comp. Networks, vol. 38, no. 4, pp. 393–422, Mar. 2002.

[9] Kimura, N. and Latifi S. "A survey on data compression in wireless sensor networks", in Proc. of Information Technology: Coding and Computing Journal, pp 8 – 13, Vol. 2, 2008.

[10] S. Bandyopadhyay and E. Coyle, ''An energy efficient hierarchical clustering algorithm for wireless sensor networks,'' in Proc. of IEEE INFOCOM'03, San Francisco, CA, pp. 1713-1723, vol. 3, Apr. 2003.