International Journal of Advanced Computer Research (ISSN (print): 2249-7277 ISSN (online): 2277-7970)
Volume-3 Number-3 Issue-11 September-2013

A Variable Length Distributed Source Coding Algorithm for WSNs

K S Shivaprakasha[1], Muralidhar Kulkarni[2]

Abstract

Ubiquitous computing has become an integral part of the modern communication system. Wireless Sensor Networks (WSNs) are networks of small devices called sensors deployed in a geographical area to sense some physical entity. The sensed information will be then conveyed to the central entity called the Base Station (BS). WSN nodes being generally deployed in harsh environments, recharging or replacing of batteries become an infeasible task. Thus it is desirable to represent the information using as minimum number of bits as possible. One such technique to perform this is source coding. In this paper an attempt is made in proposing a novel algorithm called Variable LengthDistributed Source Coding (VLDSC) which is an improvement over the existing Huffman coding technique. The results validate the betterment in the performance of the proposed algorithm.

Keywords

Distributed Source Coding, Huffman Coding, DSC using Syndromes, Variable Length Distributed Source Coding (VLDSC).

1. Introduction

Wireless Sensor Networks (WSNs) have become one of the most important research areas in the modern communication system. WSNs find applications in vast areas like military, disaster management, habitat monitoring, environmental monitoring etc [1, 2]. WSN consists of hundreds of small devices called motes. A mote consists of a sensor to sense the information, a processor to process the same, transmitting and receiving antennas, memory unit and power units. The sensed information will be communicated to the BS which in turn is connected to the physical world. A typical WSN is as shown in figure 1.BS being facilitated with the fixed infrastructure, is not energy constrained.

K S Shivaprakasha, Department of Electronics and Communication Engineering, National Institute of Technology Karnataka (NITK), Surathkal, INDIA.
Muralidhar Kulkarni, Department of Electronics and Communication Engineering, National Institute of Technology Karnataka (NITK), Surathkal, INDIA.

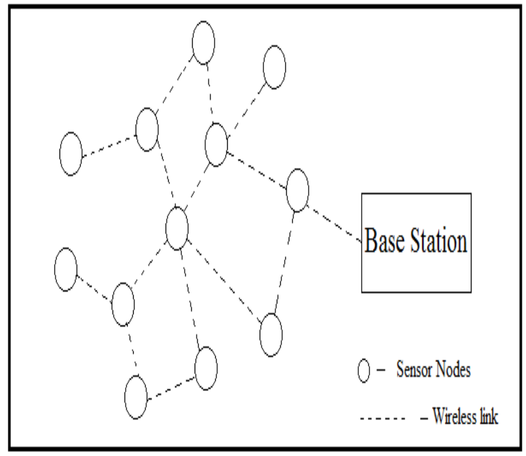

Figure 1: Basic Wireless Network

However motes suffer severe energy restrictions as recharging the batteries is difficult in WSNs. Thus the representation of the data should be done with minimum number of bits. Source coding is one of the promising techniques to achieve data compression before transmission [3].

Source Coding (SC) can be either centralized or distributed. In centralized approach there will be an encoder to compress the information and decoder at the receiver to retrieve the information. Whereas in distributed algorithms the compression is done in the distributed manner amongst motes with or without mutual exchange of information. Distributed approach suites WSN better as it also imparts load distribution amongst all nodes.

An ample number of Distributed Source Coding (DSC) algorithms have been proposed in the literature. The performance of the DSC algorithm is also dependent on the underlying routing algorithms. Most of the studies have concluded that the cluster based algorithms perform better [4]. In cluster based algorithms, the entire geographical area is divided into clusters consisting of a set of nodes with a respective Cluster Head (CH). CH monitors the data transmissions in the corresponding cluster.

Figure 2 presents a WSN incorporating a hierarchical architecture. Different clusters and the corresponding CHs and cluster members have been highlighted.

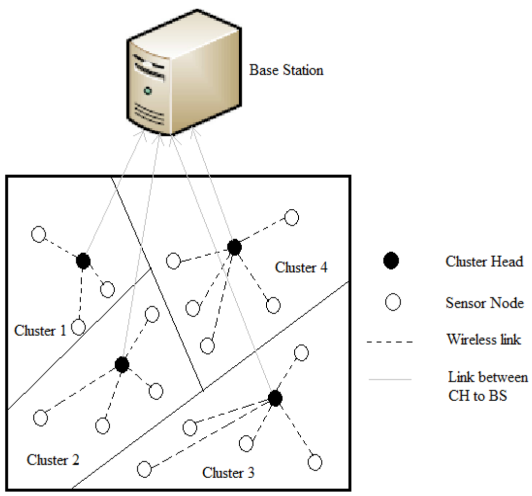

Figure 2: Basic CH Communication in DSC

Nodes with higher energy become the CHs. CHs collect the data from all its cluster members, combine the same and then transmit the fused data to the BS. SC can further improve the performance. Huffman variable length distributed source coding and Distributed Source Coding using Syndromes (DISCUS) algorithms were proved to offer a good compression ratio. However the algorithms can be improved for better efficiency. In this paper an attempt has been made in improvising Huffman variable length distributed source coding so as to assure a better compression ratio and thus enhancing the network lifetime. The performance is validated through simulations.

The rest of the paper is organized as follows: section 2 gives an insight on the previous works on DSC proposed in the literature. Section 3 presents a detailed description on the Huffman coding for WSNs. Section 4 details the proposed VLDSC algorithm. Simulation results and analysis have been carried out in section 5. Finally Section 6 gives the concluding remarks.

2. Previous Works

In WSNs, nodes are densely deployed; hence correlation exists amongst the data sensed by nodes. SC is a very important problem in information and communication systems. WSN devices are powered by batteries. Most of the WSNs are placed in hostile area, where battery recharging and replacing is very difficult. Hence better energy efficient data communication technique is required. DSC is one such technique [3]. DSC generally shifts the complexity to the decoder side.In the cluster based communication, one of the nodes in the cluster is considered as the CH. The CHs gather the information from all nodes in the cluster and send it to the BS [5, 6]. Slepian and Wolf have given a limit on the rate of transmission and have theoretically showed that separate encoding (with increased complexity at the joint decoder) is as efficient as joint encoding for lossless compression [7, 8].

In [9], authors have proposed the design for two estimators based on the correlation information so as to achieve a practical DSC. DISCUS algorithm achieves DSC by modeling the source coding problem into a channel coding one and thus exploring the usage of syndrome bits to represent the side information [10, 11]. An improvement over the syndrome based DSC has been proposed in [12]. It not only offers a higher compression ratio but also is rate adaptive. It considers integer valued data for compression.

3. Huffman Algorithm for DSC

In [13] authors have proposed a novel algorithm using Huffman coding to encode the data to be sent by the sensor in a WSN. Let m_i be the data to be sent in the i^{th} time instant. Let r_i be the binary equivalent of the data consisting of R bits where R indicates the resolution of the ADC. Then the corresponding codeword d_i can be represented as a combination of two bit sequences $<s_i a_i>$ where s_i indicates the number of bits n_i used to represent the data and a_i gives the corresponding representation.

The term d_i can be calculated as $d_i = r_i - r_{i-1}$ i.e the difference in the measured entity in a sensor node. The representation of s_i for different values of d_i is as shown in the table 1. Table 1 is retained from the baseline JPEG algorithm.

Table 1: Huffman Variable Length Codes used

n_i	s_i	d_i
0	00	0
1	010	-1, +1
2	011	-3, -2 ,+2, +3
3	100	-7, . . . -4, +4, . . . +7
4	101	-15,. . . ,-8,+8,. . . ,+15
5	110	-31,. . . ,-16,+16,. . . ,+31
6	1110	-63,. . . ,-32,+32,. . . ,+63
7	11110	-127,. . . ,-64,+64,. . . ,+127

8	111110	-255,. . . ,-128,+128,. . . ,+255
9	1111110	-511,. . . ,-256,+256,. . . ,+511
10	11111110	-1023,. . ,-512,+512,. . ,+1023
11	111111110	-2047,..,-1024,+1024,. ,+2047

a_i can be calculated as follows:

- If $d_i > 0$, a_i is the 2's complement representation of the sequence d_i
- If $d_i < 0$, a_i is the 2's complement representation of the sequence d_i-1
- If d_i=0, a_i is not transmitted

The transmitter transmits the tuples <s_i, a_i> for every d_i. Simulation results show a good compression achieveddespite its simplicity.

4. Variable Length Distributed Source Coding (VLDSC) Algorithm

The algorithm discussed in section 3 offers a better performance and simplicity in design. However some modifications can be incorporated so as improve its performance and thus enhancing the network lifetime. This section provides an overview of the proposed VLDSC algorithm for WSNs.

The following assumptions were made in the proposed VLDSC:

- Nodes are assumed to be stationary.
- BS has the complete network information.
- BS does cluster formation and CH selection based on M-TRAC algorithm [14].
- CH allocates schedule for their cluster members based on TDMA.
    - Scheduling is done in such a way that node in the common transmission range of all other nodes in the cluster will be given the chance to transmit first. However it is always true as the transmission within the cluster is direct.
    - Other nodes listening to the transmission of the first sender, transmits the CH, the difference between its data and the data of the first node
    - The CH collects data from all members, encodes using the proposed algorithm and then sends the encoded sequence to the BS.
- Transmission between the nodes and CH is direct and that of CH to the BS is multihop.
- Nodes will have the continuous monitoring of data and the data is conveyed to the BS either
    - When there is a significant change in the sensing entity (event driven). In this case only the node sensing the significant change initiates a request for transmitting and the normal M-TRAC algorithm is used for communication. As there is only one sender, no DSC is required.
    - Periodically, in which case nodes transmit the data as per the TDMA policy. First node sends the n-bit data uncoded. Listening to the data transmitted by the first node, others transmit the difference to the CH. CH performs encoding as per the proposed algorithm.

As per our algorithm we consider to transmit the difference in the measured entity w.r.t the data transmitted by the first node. As the nodes are likely to be densely deployed, the measured entity will be correlated and thus the difference can be 0, ±1, ±2, ±3 and so on. It is clear from our assumption that the difference being 0 or ±1 would be more likely to occur than the differences being ±2, ±3 and so on. Thus it is advantageous to encode the smaller difference with less number of bits and the length is an increasing function w.r.t the difference.

The following improvements over the algorithm proposed in [13] have been done. The algorithm considers the compression of the data from a single sensor node. The auto- correlation of the data sensed by a sensor is explored for the data compression. However it can be modified for a cluster based query based/time driven WSN where the data sensed by the nodes within a cluster are more likely to be correlated. Thus an aggregation has to be performed at the CH. CHs collect all data from its cluster members, say M_1, M_2, M_3, M_4, M_5 if there are 5 nodes in the cluster. All these 5 data are highly spatially correlated. Thus the first data can be sent as it is whereas for the other messages the difference sequences d_2, d_3, d_4 and d_5 can be transmitted to the CH. CH then performs the aggregation using the procedure as explained in the previous algorithm.

The proposed compression algorithm also has to assure a simple computational requirement as the computational energy requirement may override the amount of energy saved in the reduction of the size of the data representation. In the algorithm proposed in [13], each of the nodes has to compute the 2's complement of either d_i or d_i-1 before transmission. However this computation can be avoided if there is a priori tabulated code vectors, a_i for each of the difference vector d_i. It is as shown in table 2.

Thus if both the sender and receiver agrees upon this table during network initialization, the CH, need not has to compute anything. Instead it has to just transmit an entry corresponding to the difference from the table.

Table 2: Modified Huffman Variable Length Codes used

n_i	s_i	d_i	a_i
0	00	0	---
1	010	-1	0
		+1	1
2	011	-3	00
		-2	01
		+2	10
		+3	11
3	100	-7	000
		-6	001
		-5	010
		-4	011
		+4	100
		+5	101
		+6	110
		+7	111
…	…..	….	….
…	…..	….	….

Also the proposed algorithm is an improvement compared to DISCUS too on the following grounds. In DISCUS, it is assumed that the maximum Hamming distance amongst the data sensed by nodes as 1. As in general the data being sensed in a WSN is analog, in our proposed algorithm, we consider the analog representation of the data. Thus the difference in the sensed entity of the nodes in a cluster can be either 0 (no difference), ±1 (difference of 1 either in positive or negative direction), ±2 and so on.

In DISCUS, nodes transmit 'n' bit data to a central entity and the central entity does the DSC, by transmitting 'n' bits for the first data and just '(n-k)' bits of syndrome to represent the remaining data. The receiver then decodes the data from the syndrome as the maximum allowed Hamming distance is 1. However the transmission in the last mile (i.e nodes to a central entity say CH) is uncompressed.

In the proposed algorithm we have explored the advantage of using wireless medium for transmission in WSN. All nodes in the range receive the data transmitted by the first node as the first node to transmit data is assumed to be in the common range of all other nodes in the cluster. Thus instead of CH receiving complete data from all nodes in the cluster and then compressing, we propose each of the nodes itself can directly transmit the compressed data by sending just the difference in the sensed entity. This achieves better energy efficiency by not only reducing the overall network traffic but also assuring a better load balancing as CH is relieved from getting overburdened.

In DISCUS, a fixed length (n-k) bits are used for sending the information and thus saving 'k' bits per data. In the proposed algorithm we propose a variable length coding where for every non negative value for 'l' there exists 2^l codes of length 'l'.Thus differences which are more likely to occur will be coded with lesser bits which will further add to the energy conservation in the network.

The encoding procedure is as follows

- The first node in the time schedule transmits its data to the CH.
- Hearing to the data transmitted by the first node, further nodes transmit the difference signal to the CH
- CH computes the pair <$s_i a_i$> for all nodes in the cluster. Where s_i and a_i can be obtained from table 2.
- The fused information is communicated to the BS from the CH.
- Each node has to be active only for two time slots. One is for the first time slot to get the information about the first data sent by a node and the other is at its slot to transmit the difference signal.

5. Simulation Results and Analysis

To validate our algorithm, we have considered the sensirion [15] sensor nodes. Sensirion SHT11 module is used for the analysis. It measures the temperature and relative humidity. The module equipped with a 14-bit ADC. From the datasheetsof the SHT11, 12 bit resolution is used for humidity and 14 bit for the temperature.

We have considered 500 samples of temperature and humidity for the analysis. We considered a cluster of

10 nodes for the validation. The results are detailed in table 3.

Table 2: Compression Ratio in VLDSC

	Temperature	Humidity
Original Size	7000 bits	6000 bits
Compressed Size	2388 bits	1994 bits
Compression Ratio	65.88%	66.76%

It is seen from the table that the proposed VLDSC offers a very high compression ratio and thus assures better energy efficiency.

6. Conclusion and Future Scope

Energy aware data communication is the most vital criterion in the design of a WSN. DSC along with routing assures a better data representation in a sensor network. In this paper an improvement over the Huffman Coding for WSNs has been proposed resulting in VLDSC which offers a better compression. The proposed algorithm has been validated for its performance through simulations.Although the work highlights the advantage of using a DSC for WSNs, selection of appropriate scheduling policy also has an impact on its performance. As a part of future work, an appropriate scheduling algorithm has to be devised for the proposed algorithm. Also the performance of the proposed algorithm can be compared with any of the existing ones to further validate the results.

References

[1] WaltenegusDargie, Christian Poellabauer, "Fundamentals of Wireless Sensor Networks Theory and Practice", John Wiley and Sons, 2010.
[2] Krishnamachari, L., Estrin, D., & Wicker, S. "The impact of data aggregation in wireless sensor networks" In Distributed Computing Systems Workshops. Proceedings. 22nd International Conference on (pp. 575-578). IEEE, 2002.
[3] Pier Luigi and MichaeGastpar, "Distributed Source Coding", Academic Press, Elsevier, 2009.
[4] K.S. Shivaprakasha, MuralidharKulkarni, "Energy Efficient Routing Protocols for Wireless Sensor Networks: A Survey", International Review on Computers and Software (IRECOS), Vol 6, No 6, pp 929-943, 2009.
[5] Xiong, Z., Liveris, A. D. & Cheng, S. "Distributed Source Coding for Sensor Networks", Signal Processing Magazine, IEEE, 21(5), pp 80-94, 2004.
[6] Arjmandi, H., &Lahouti, F. "Resource Optimized Distributed Source Coding for Complexity Constrained Data Gathering Wireless Sensor Networks" Sensors Journal, IEEE, 11(9), pp 2094-2101, 2011.
[7] A.D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," IEEE Trans. Inform. Theory, Vol. IT-22, pp. 1–10, Jan. 1976.
[8] Stankovic, V., Stankovic, L., & Cheng, S. "Distributed Source Coding: Theory and Application", Proceedings of European Signal Processing Conference, 2010.
[9] Barceló-Lladó, J. E., Pérez, A. M., &Seco-Granados, G. "Enhanced Correlation Estimators for Distributed Source Coding in Large Wireless Sensor Networks" Sensors Journal, IEEE, 12(9), pp 2799-2806, 2012.
[10] S. S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): Design and construction," Proc. IEEE Data Compression Conference, Snowbird, UT, March 1999.
[11] Pradhan, S. S., &Ramchandran, K. "Distributed source coding: Symmetric rates and applications to sensor networks", In Data Compression Conference. Proceedings. DCC 2000, pp. 363-372. IEEE, 2000.
[12] Chen, J., Khisti, A., Malioutov, D. M., &Yedidia, J. S. "Distributed source coding using serially-concatenated-accumulate codes", InInformation Theory Workshop, 2004. IEEE, pp. 209-214. IEEE, 2004.
[13] Francesco Marcelloni, Massimo Vecchio, "A Simple Algorithm For Data Compression In Wireless Sensor Networks", IEEE Communications Letters, Vol. 12, No. 6, June 2008.
[14] K S Shivaprakasha, M Kulkarni, Nishant Joshi D, "Improved Network Survivability using Multi-threshold Adaptive Range Clustering (M-TRAC) Algorithm for Energy Balancing in Wireless Sensor Networks", Journal of High Speed Networks, Vol 19, pp 99-113, 2013.
[15] SHT15 - Digital Humidity Sensor (RH&T). Available: http://www.sensirion.com/en/products/humidity-temperature/humidity-sensor-sht15/.

K S Shivaprakasha received his BE (Electronics & Communication) from Bahubali College of Engineering, Visvesvaraya Technological University with IX rank to the University and MTech (Digital Electronics and Communication Systems) from MalnadCollege of Engineering, Visvesvaraya Technological University with I rank to the University in 2004 and 2007 respectively. Presently he is pursuing his PhD at National Institute of Technology Karnataka, Surathkal in the field of Wireless Sensor Networks. He is currently working as Associate Professor in the Department of Electronics and Communication Engineering, Bahubali College of Engineering, Shravanabelagola, Karnataka, India. His areas of interest include Wireless Sensor Networks, Mobile Adhoc Networks, Information Coding Theory and Cryptography. He has more than 15 publications to his credit.

MuralidharKulkarni received his B.E. (Electronics Engineering) degree from University Visvesvaraya College of Engineering, Bangalore University, Bangalore, M. Tech (Satellite Communication and Remote Sensing) from Indian Institute of Technology, Kharagpur (IIT KGP) and PhD from JMI Central University, New Delhi in the area of Optical Communication networks.

He has 31 years of experience which includes 5 years in industry and 26 years of teaching experience. He has held the positions of Scientist in Instrumentation Division at the Central Power Research Institute, Bangalore (1981-1982), Aeronautical Engineer in Avionics group of Design and Development team of Advanced Light Helicopter(ALH) project at Helicopter Design Bureau at Hindustan Aeronautics Limited(HAL), Bangalore (1984-1988), Lecturer (Electronics Engineering) at the Electrical Engineering Department of University Visvesvaraya College of He has 31 years of experience which includes 5 years in industry and 26 years of teaching experience.

He has held the positions of Scientist in Instrumentation Division at the Central Power Research Institute, Bangalore (1981-1982), Aeronautical Engineer in Avionics group of Design and Development team of Advanced Light Helicopter(ALH) project at Helicopter Design Bureau at Hindustan Aeronautics Limited (HAL), Bangalore (1984-1988), Lecturer (Electronics Engineering) at the Electrical Engineering Department of University Visvesvaraya College of Engineering, Bangalore (1988-1994) and Assistant Professor in Electronics and Communication Engineering (ECE) Department at the Delhi College of Engineering (DCE), Govt. of National Capital territory of Delhi, Delhi (1994-2008). He has served as Head, Department of Information Technology and Head, Computer Center at the Delhi College of Engineering (University of Delhi), Delhi. Currently, he is a Professor and Head in the Department of Electronics and Communication Engineering (ECE) Department, National Institute of Technology Karnataka (NITK), Surathkal, Karnataka, India.

He is currently the Coordinator of the Centre of Excellence for Wireless Sensor Networks, Dept. of Electronics and Communication Engineeing, National Institute of Technology Karnataka. Dr. .Kulkarni's teaching and research interests are in the areas of Digital Communications, Fuzzy Digital Image Processing, Optical Communication and Networks, and Wireless Sensor Networks. He has published several research papers in the above areas, in national and International journals of repute. For various contributions his Biography has been listed in the Marquis, Who's Who in Science & Engineering (2008). He has also authored/coauthored four very popular books in Microwave & Radar Engineering, Communication Systems, Digital Communications and Digital Signal Processing.