

A Modified Lempel Ziv Welch compressive data collection in Wireless Sensor Networks

M.Vijayalakshmi^{#1}, A.Helen Victoria^{*2}

^{*1} Assistant Professor, IT Department, SRM University, Kattankulathur.

¹ vijayalakshmi.m@ktr.srmuniv.ac.in

^{*2} Assistant Professor, IT Department, SRM University, Kattankulathur.

² ahelenvictoria.a@ktr.srmuniv.ac.in

Abstract— Nowadays wireless sensor networks have wide applications in many fields such as medical, industrial, military, etc. Sensors are having limited processing capabilities and energy is an important constraint in wireless sensor networks as it determines the lifetime of the network. In a large scale wireless sensor networks, the sensor nodes have to collect more data when they are moving towards the sink. So the sensor nodes energy, nearer to the sink may get drained off more quickly and an alternate path has to be chosen which increases the delay in the network. Compressive data gathering is one of the techniques which reduce the data size, balance the energy in large scale wireless sensor networks. The existing technique takes more computations and increase in complexity on compressing the data. So a modified compressive data gathering protocol is designed which uses a lossless compression algorithm called modified Lempel Ziv Welch compression algorithm which is a simple and fast method for compressing the data. A tree is constructed and parent-children nodes are assigned and child node carries the compressed data and intermediate parent node aggregates and compresses it until it reaches the sink. The original data is reconstructed at the sink using Modified Lempel Ziv Welch decompression algorithm

Index Terms - Wireless sensor network, data aggregation, MLZW compression and decompression.

I. INTRODUCTION

Wireless sensor networks consist of numerous autonomous sensors deployed in an environment to monitor physical or environmental conditions such as temperature, humidity etc. They transfer the data to a co-ordinating node for processing which acts as gateway to pass the data to sink. Each sensor node may consist sensing unit, processing, memory and microcontroller. The size of sensor nodes may vary from grain size to a shoe-box size depending upon the application in which it is used. Similarly, cost may vary from few to hundreds of dollars. The architecture and components of a wireless sensor node is shown in Figure: 1.1. Size and cost constraints of sensor results in resource constraints such as energy consumption, memory, bandwidth and speed.

A wireless sensor network is an emerging technology nowadays, which plays an important role in creating a smart environment. It has huge applications in many fields such as environmental monitoring, natural hazard detection, military applications, health monitoring, forest monitoring, precision agriculture etc.

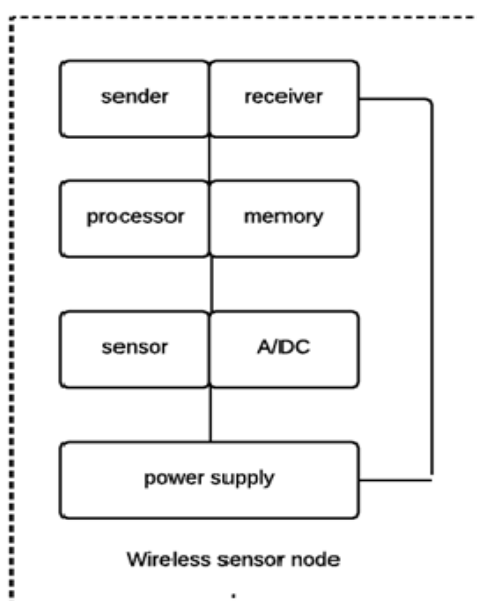


Figure 1.1 Simple Architecture of Wireless Sensor Node

Hundreds of nodes scattered throughout a field assemble together, establish a routing topology, and transmit data back to a collection point. The application demands for robust, scalable, low-cost and easy to deploy networks are perfectly met by a wireless sensor network. If one of the nodes should fail, a new topology would be selected and the overall network would continue to deliver data. If more nodes are placed in the field, they only create more potential routing opportunities

Data aggregation is a type of data gathering technique in wireless sensor network. With the help of data aggregation, we reduce the energy consumption by eliminating redundancy. There is extensive research in the development of new algorithms for data aggregation. Types of data gathering technique are tree based aggregation, cluster based aggregation, and chain based aggregation. There are many data aggregation algorithms proposed such as LEACH, TAG, and PEGASIS etc...

The data collected at children nodes are aggregated in parent nodes in the tree based approach, cluster head in clustering technique, a leader node in chain based aggregation. Data aggregation techniques are very useful for data gathering and to conserve the energy, maintaining the QOS and extend the network lifetime

Tree Topology integrates the characteristics of Star and Bus Topology. Earlier we saw how in Physical Star network Topology, computers (nodes) are connected by each other through a central hub. And we also saw in Bus Topology, workstation devices are connected by the common cable called Bus. After understanding these two network configurations, we can understand the tree topology better. In Tree Topology, the numbers of Star networks are connected using Bus. This main cable seems like a main stem of a tree, and other star networks as the branches. It is also called Expanded Star Topology. Ethernet protocol is commonly used in this type of topology.

The compressive data gathering is an emerging and useful method of collecting data from sensor nodes to sink to reduce the data size on large scale wireless sensor networks. The sensor nodes near to the sink need to collect more number of data from the nodes that are placed far away. At each level amount of data gathered becomes large and the sensor nodes may not have enough energy to transmit the data to the sink and hence the node failure occurs. As a result if must find an alternate path to reach sink takes some time, which increases the delay in the network. Thus the compressive data gathering saves the energy and reduces the delay.

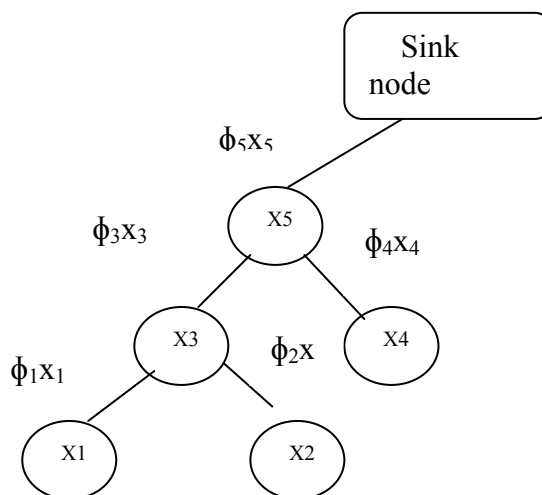


Fig. 1.2 Compressive Data Gathering

II. RELATED WORK

In [1] the author does in network aggregation with the objective of reducing the delay and save energy. Interference model is developed by considering the interference constraints on links with Scheduling policies such as myopic and non-myopic policies are proposed. This model achieves good performance when network size is small. Suitable for Bluetooth and FH-CDMA Networks. The objective in [2] is to maximise the quality of data at the sink under deadline and energy constraints. An optimisation framework is designed. That maximizes the aggregated data from the sink within a deadline. An optimal data aggregation policy and scheduling policy are provided. Polynomial time algorithm is proposed which uses local information at each hop. The sink provides a constraint on how long data can be gathered from the predecessor nodes and transmitted to the sink. Maximised the aggregated information and accuracy. [3] constructs the shortest path tree (SPT) find the shortest path to sink. Load balanced latency efficient data aggregation scheduling is proposed. Latency minimized and load balanced assignment problem (LMLBA) solves parent-children assignment problem. Compared with SPT and MSL (Minimum Sleep Latency) algorithm, LMLBA realises better trade-off between min sleep latency and balanced load and effective data aggregation scheduling for duty cycled WSNs.

A Probabilistic model is proposed in [7] which data communication over link is successful with certain probability. Cell based path scheduling (CPS) algorithm is proposed schedules multiple super nodes on multiple paths concurrently. Zone based path scheduling (ZPS) speeds up continuous data collection forming data transmission pipeline. CDG technique along with pipeline technology gives efficient network capacity. A delay efficient data aggregation scheduling with SINR constraints in [6] in which graph based interference model and reduced routing graph is proposed. Uses Breadth first search in the proposed graph. The compressive scheduling algorithm reduces the delay better when compared to previous distributed algorithm by merging the links that has been sent to the sink. Improved approximation ratio of proposed algorithms.

FAST approach in [12] deals with tree construction and scheduling under protocol interference model Connected 3-hop dominating sets (C3DS) based structure is designed. A distributed collision free TDMA schedule is used for scheduling. The algorithm outperforms other approaches, provides parallel transmissions and gave an upper bound agg latency with $12R+\Delta-2$ timeslots. S. Ji and Z. Cai (2013) proposed the paper Distributed data collection in large-scale asynchronous wireless sensor networks under the generalized physical interference model. Its objective is avoiding collision and energy efficiency. R° -PCR career sensing range is derived to avoid collision and transmission overhead. Distributed data collection (DDC) algorithm is proposed. Apply R° -PCR to DDC and Distributed data aggregation is proposed. Delay and capacity of DDC and DDA under Poisson node distribution model. Through R° -PCR fewer transmissions with more data per transmission is achieved and the delay is reduced in DDA.

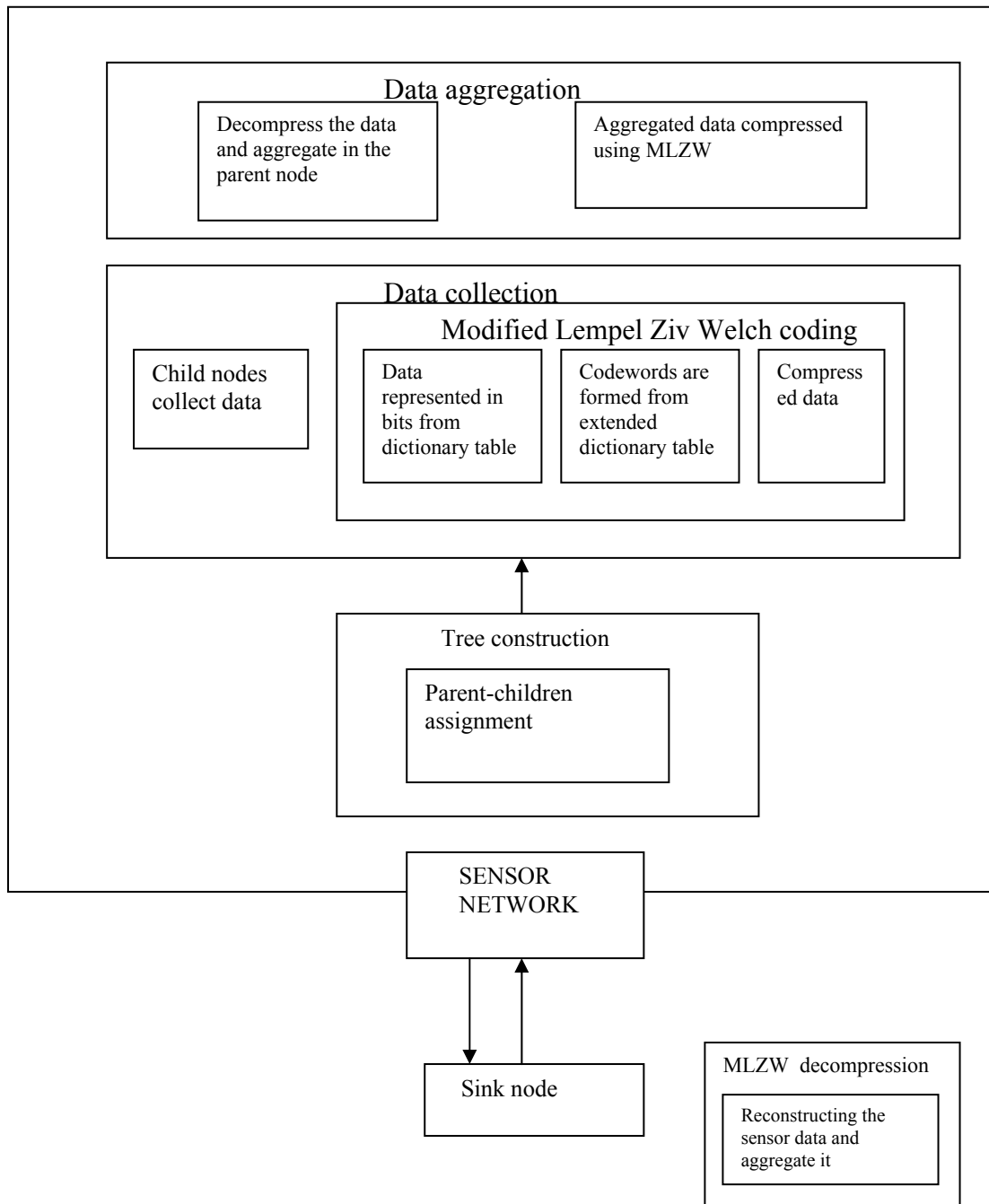
Agriculture Data Aggregation (GRIDA) scheme in [17] aggregates the data by eliminating the repeated values in the farming field. This method reduces power consumption, packet delivery rate and reduce the delay. The aggregation node may consume data as it moves up towards the sink. Cognitive Path Planning (CPP) determines path planner for path selection in [18]. Data collector (DC) moves in the network and compares the path by analysing smoothing and safeness by knowing its obstacles. Avoids node failure as it selects its right path by avoiding obstacles and if data collector fails leads to latency. The MA is considered as an ant and an ant colony optimization-based dynamic energy efficient mobile agent routing (ADEEMA) algorithm [16]. Routing Optimal Degree (ROD) algorithm is proposed to analyse the performance of chosen route. A reinitialising rule is proposed on node failures due to change in topology. If mobile agent fails entire breakdown of the network. Modified Multi Itinerary Planning (MIP) in [3] where multiple mobile agents with set of interested nodes makes the data to reach sink with good speed and minimum spanning tree used for path selection. Compressive sensing theory in [4] compresses the original data to some transformation domain. These weighed sums are received at the sink as Projection based compressive data gathering. They can be decoded by solving convex optimisation problem. It distributed the transmission load throughout the network. The future work is if used along with scheduling may reduce the latency. [13] proposes a Minimum Spanning Tree Projection (MSTP) in which MSTP and e-MSTP outperforms previous schemes such as non-CS, plain-CS, and hybrid -CS. The future work is that optimal solution can be found by finding the position of projected nodes which will minimise transmission cost and load balancing.

The objective of [5] is to reduce the computational cost, reduce the delay and energy consumption. CDG (Compressed Data Gathering). Aggregated data is compressed and weighed encoded sum is sent to sink. Forwarding Tree Construction and Scheduling (FTCS): Multiple forwarding trees are constructed and encoded sum received in node is called projection based PCDG. Link scheduling algorithm: forwarding trees links are scheduled using TDMA. Reduced when compared to the decentralised FCTS. This joint model achieves minimum latency with less transmission overhead. Energy load is balanced. A Modified Huffman coding algorithm represents the length of the code exponentially in [6]. Real data along with delay is given to ADC and converted to binary. This binary value is represented as positive and negative integers. It compresses with high compression ratio but Look up table is large

An improved LZW algorithm is proposed in [19] uses Dictionary look up based algorithm where repetited substrings have a codeword. A dictionary table is maintained in which dictionary capacity selected based on capacity of nodes. Pre-processing carried out for changing other document files to text files as traditional algorithm transmits only text files. Thereby Reduces amount of storage, improve the compression ratio and reduce the size of dictionary. It is suitable for sparse network with less no of nodes. Mobile Agent Path Design Algorithm (MAPDA) generate binary sparse random matrices from the graph. Another algorithm called Orthogonal Matching Pursuit (OMP) algorithm for signal recovery and reconstruction also proposed. Though it has less computation than Gaussian matrix, there still exists a computational complexity and better approach for achieving good throughput has to be designed. The compressive sensing theory in [2] is used in finding sparse random matrices and mobile agent collects those measurements and reach the sink. Though it is energy balanced, this type of compression involves complex computations and data packet size may get increased in collecting such measurements

III. ARCHITECTURE OF PROPOSED SYSTEM

The architecture in fig 3.1 describes that the user sends a request to the base station or sink node which disseminates the message by sending beacon signal to the wireless sensor network. The interested nodes form a tree structure with child nodes and parent nodes by considering the parameters for parent-child assignment. The sensed data is locally collected in child nodes and a lossless compression algorithm such as MLZW (Modified Lempel-Ziv-Welch) is applied to the collected data. The compressed data are then aggregated in parent node. Though we use compressed data collection, some upper aggregated nodes perform more aggregation operations which may consume energy and may lead to node failure due to lack of energy. So an energy balanced technique is proposed which perform reconstruction mechanism on energy lagging co-ordinating nodes by distributing energy in uniform from the neighbouring non transmitting nodes. The final aggregated data is sent to the sink nodes, which perform decoding of the aggregated data and send to the user through internet.



Algorithm1: Tree Construction and Parent Children Assignment Algorithm

Input: Set of sensor nodes

Output: Binary tree construction and parent-children assignment for nodes Begin

Let $V = \{V_1, V_2 \dots V_m\}$ be the set of sensor nodes and S be the sink node (root node)

Now connect the left and right node to S. E be the edge between the nodes that are connected.

Successive nodes are attached to left and right child of each node until leaf nodes are reached.

Root nodes now send HELLO messages to the sensor nodes that are immediately below S. Each node

Calculates hop count and forwards the incremented hop count to successive nodes.

Now all nodes computes QOS value says Q_t with parameters such as hop count, RSSI, transmission power and channel bandwidth.

Each node compares its Q_t value with neighbouring nodes and the node with better Q_t value will be set as parent node otherwise set as child node

End

Table :1 Initial Dictionary Table

Str/no	Bits	Code
0	000	No
1	001	n1
2	010	n2
3	011	n3
4	100	n4
6	110	n6
7	111	n7
8	1000	n8
9	1001	n9
#	00000]0
A	00001	s1
B	00010	s2
C	00011	s3
D	00100	s4
E	00101	s5
F	00110	s6
G	00111	s7
H	01000	s8
I	01001	s9
J	01010	s10
K	01011	s11
L	01100	s12
M	01101	s13
N	01110	s14
O	01111	s15
P	10000	s16
Q	10001	s17
S	10011	s19
T	10100	s20
U	10101	s21
V	10110	s22
W	10111	s23
X	11000	s24
Y	11001	s25
Z	11010	s26

The Initial dictionary table shown in Table: 1 contains code and bits for the corresponding number and the character. This table will be stored in source, intermediate nodes and sink.

Example: Let us apply MLZW algorithm for the input string word “MALAYALAM”

3.3 COMPRESSION AND DECOMPRESSION

3.3.1 MLZW COMPRESSION

It is a lossless compression algorithm and dictionary based. It is a simple process. It replaces strings of characters with single code. It starts with a dictionary of single characters and gradually extends the table with the input of characters of a string. It is good for text compression. It may use binary or ASCII values for characters. Algorithm 1 shows how the data is compressed using MLZW algorithm.

Table: 2 MLZW Compression Table

S/N	C	Output Code bit	Extended Dictionary S code
Null	M		
M	A	s13 01101	MA s27
A	L	s1 00001	AL s28
L	A	s1s2 01100	LA s29
A	Y	s1 00001	AY s30
Y	A	s25 11001	YA s31
AL	A	s28 11100	ALA s32
A	M	s1 000001	AM s33

The compression table is shown in Table: 2 first read the null character before the string and enters in the compression table and read the first character say M of the string and output the code for the S/N and then read the next character say A and enter the combination of MA and its code in the extended dictionary table. Whenever MA combination repeated the corresponding codeword will be used on further entries. A will be entered in S column and the next character from the string in C column and the process continues as such till the end of string.

Un encoded length = 9 symbols × 5 bits/symbol = 45 bits

Encoded length = (6 codes × 5 bits/code) + (1 codes × 6 bits/code) = 36 bits

$$\text{compression ratio} = \frac{\text{no of bits after compression}}{\text{no of bits before compression}} \quad (1)$$

$$= \frac{36 \text{ bits}}{45 \text{ bits}}$$

$$= 0.8$$

Algorithm 2: MLZW Compression Algorithm

Input: Sensed data is applied with modified Lempel Ziv Welch algorithm

Output: Compressed data with corrupted data elimination

Begin

Initialise the dictionary with all possible individual characters and S is empty

String S = get input symbol

C=next character in the char stream

If the string S+C present in the dictionary?

S=S+C (Extend S with C)

Else

output the codeword for S

add the string S+C to the dictionary.

Output the compressed string

End If

End

3.2. MLZW DECOMPRESSION

The decoding algorithm works by reading a value from the encoded input and outputting the corresponding string from the initialized dictionary. In order to rebuild the dictionary in the same way as it was built during encoding, it also obtains the next value from the input and adds to the dictionary the concatenation of the current string and the first character of the string obtained by decoding the next input value, or the first character of the string just output if the next value cannot be decoded (If the next value is unknown to the decoder, then it must be the value that will be added to the dictionary this iteration, and so its first character must be the same as the first character of the current string being sent to decoded output). The decoder then proceeds to the next input value (which was already read in as the "next value" in the previous pass) and repeats the process until there is no more input, at which point the final input value is decoded without any more additions to the dictionary. In this way the decoder builds up a dictionary which is identical to that used by the encoder, and uses it to decode subsequent input values. MLZW decompression table shown in Table 3.

Example: Now we can decompress the string s13 s1 s12 s1 s25 s28 s1 s13

Table: 3 MLZW Decompression Table

Input next		Output Code string		Extended Dictionary	
s13	s1	s13	M	MA	s27
s1	s12	s1	A	AL	s28
s12	s1	s12	L	LA	s29
s1	s25	s1	A	AY	s30
s25	s28	s25	Y	YA	s31
s28	s1	s28	AL	ALA	s32
s1	s13	s1	A	AM	s33
s13		s13	M	M	s34

Algorithm 3: Decompression Algorithm

Input: Compressed data is applied with decompression algorithm.

Output: Original data reconstructed at destination.

Begin

Output the string with the first codeword

While (there are more codeword)

{

If (current codeword in the dictionary)

Output the string of current codeword

Else

Output previous output +current output first character

Insert in the dictionary previous output +current output first character

End if

}

End

3.4. Modified Compressive Data Gathering

Modified Compressive Data Gathering uses MLZW algorithm on the sensor nodes and aggregated data are carried by the intermediate nodes. The total aggregated data is received by sink. Algorithm 4 shows the step by step procedure of MCDG.

Algorithm 4: Modified Compressive Data Gathering

Input: Data sensed from sensor node

Output: compressed and aggregated data at sink

Begin

Let N be the number of nodes, s_i , ϕ_i and x_i be the sensed data, compressed data and node id respectively

Let Z_C be the total aggregated and compressed data received at sink, Z_D be the decompressed data and Z be the average of all sensed data from the sensor network.

At node i , s_i is compressed to ϕ_i $\phi_i || x_i$ are compressed

Now transmit $\phi_i || x_i$ to intermediate node say j and aggregates its information along with compressed information received from node i are aggregated

Successive upper nodes are aggregated until sink is reached and the total aggregated data at the sink be

$$Z_C = \phi_i || x_i + \phi_j || x_j + \dots + \phi_n || x_n$$

$$Z_D = s_i \dots s_n$$

$$Z = \frac{\sum_{i=1}^n s_i}{n}$$

End

3.5. PERFORMANCE EVALUATION

Our proposed work is examined and compare with the existing compression technique. NS2 simulator is used to evaluate the performance. The graph is plotted for PDR, Energy consumption and delay.

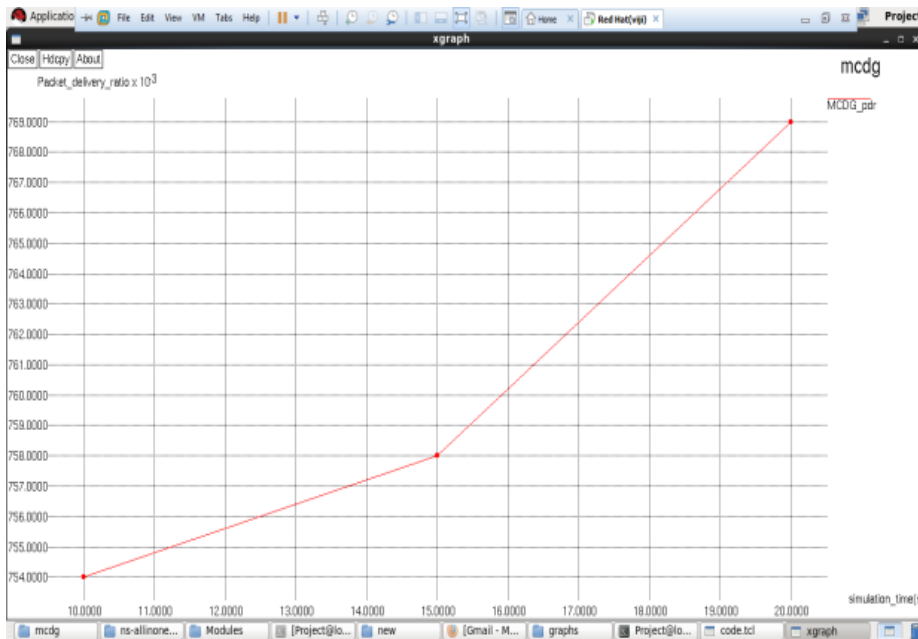


Figure: 3 MCDG_PDR Vs MCDG Delay

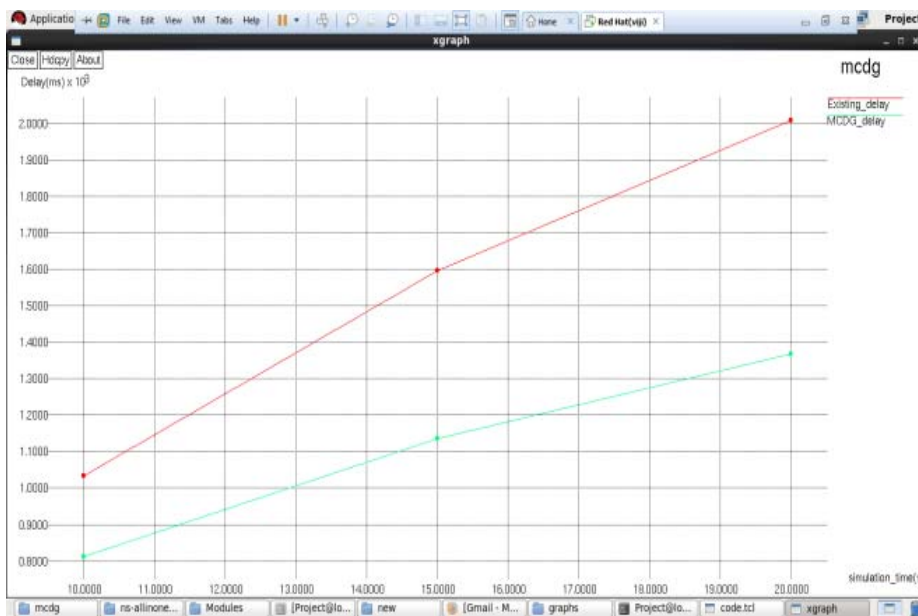


Figure: 4 Existing Delay VsVMCDG Delay

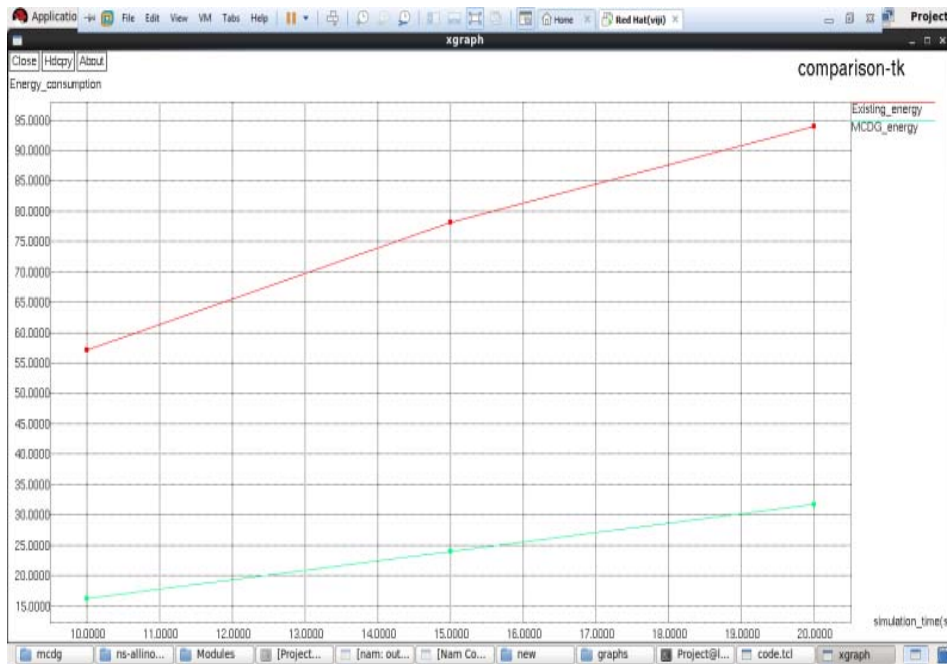


Figure: 5 Existing Energy vs MCDG_Energy

The packet delivery ratio in the proposed work is shown in Figure: 3 with simulation time in X axis and PDR on Y axis and values are plotted. The graph is drawn in Figure: 4 with simulation time in X axis and delay on Y axis. The values for the delay in the existing work and proposed work are plotted and compared in this graph. The delay has been reduced up to 28% comparing with existing CDG. The graph is drawn in Figure: 5 with simulation time in X axis and energy in Y axis. The values for the energy in the existing work and proposed work are plotted and compared in the above graph. The energy consumption is reduced by 69% compared with existing CDG.

IV. CONCLUSION

Compressive data gathering in existing system suffers from increase in complexity on compressing the data which uses the compressive sensing theory. The theory involves matrix multiplications which increase microprocessor computations inside sensors which consume additional energy and the modified compressive data gathering overcomes these drawbacks by simplifying the compression technique and maintains the quality of data without losing data on compression by lossless compression algorithm. The experimental results show that the delay and energy consumption has been reduced on comparing with the existing work. In phase II compression on heterogeneous data is to be designed. An energy reconstruction algorithm can be proposed to reconstruct energy on the nodes when their energy got drained on more transmissions.

REFERENCES

- [1] S. M. Metev and V. P. Veiko, Laser Assisted Microtechnology, 2nd ed., R. M. Osgood, Jr., Ed. Berlin, Germany: Springer-Verlag, 1998.
- [2] J. Breckling, Ed., The Analysis of Directional Time Series: Applications to Wind Speed and Direction, ser. Lecture Notes in Statistics. Berlin, Germany: Springer, 1989, vol. 61.
- [3] S. Zhang, C. Zhu, J. K. O. Sin, and P. K. T. Mok, "A novel ultrathin elevated channel low-temperature poly-Si TFT," IEEE Electron Device Lett., vol. 20, pp. 569-571, Nov. 1999.
- [4] M. Wegmuller, J. P. von der Weid, P. Oberson, and N. Gisin, "High resolution fiber distributed measurements with coherent OFDR," in Proc. ECOC'00, 2000, paper 11.3.4, p. 109.
- [5] R. E. Sorace, V. S. Reinhardt, and S. A. Vaughn, "High-speed digital-to-RF converter," U.S. Patent 5 668 842, Sep. 16, 1997.
- [6] (2007) The IEEE website. [Online]. Available: <http://www.ieee.org/>
- [7] M. Shell. (2007) IEEEtran webpage on CTAN. [Online]. Available: <http://www.ctan.org/tex-archive/macros/latex/contrib/IEEEtran/>
- [8] FLEXChip Signal Processor (MC68175/D), Motorola, 1996.
- [9] "PDCA12-70 data sheet," Opto Speed SA, Mezzovico, Switzerland.
- [10] A. Karnik, "Performance of TCP congestion control with rate feedback: TCP/ABR and rate adaptive TCP/IP," M. Eng. thesis, Indian Institute of Science, Bangalore, India, Jan. 1999.
- [11] J. Padhye, V. Firoiu, and D. Towsley, "A stochastic model of TCP Reno congestion avoidance and control," Univ. of Massachusetts, Amherst, MA, CMPSCI Tech. Rep. 99-02, 1999.
- [12] Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification, IEEE Std. 802.11, 1997.

AUTHOR PROFILE

M.Vijayalakshmi, working as Assistant Professor in IT Department in SRM University, Kattankulathur, and Kancheepuram. she has done her PG in MIT, Chennai. Her Area of interests are Wireless Sensor Networks, Big Data Analytics and Cloud Computing.

A.Helen Victoria, working as Assistant Professor in IT Department in SRM University, Kattangalathur, Kancheepuram. she has done her PG in Anna University, Chennai. Her area of Interests are Wireless sensor networks, Big Data Analytics and Cloud Computing.