International Journal of Engineering Research and Management (IJERM) – Volume 3 Issue 6- June 2017

Efficient Data Aggregation using Buffered Strategy to Prevent Congestion

Puja Mandlapure, Varsha Jadhav

Kajal Oswal , Aishwarya Nalawade

pujamandlapure1@gail.com,111varshajadhav93@gmail.com

kajaloswal10@gmail.com, aishnalawade2103@gmail.com

Abstract*— Network including sensors have peer to peer connection and are deployed to be used in the practical applications. Sensors sense the information and simply forward data packet but do not modify the payload. Due to which huge number of data is gathered which needs analysis to be done. Many of the practical applications require specific information in order to have accurate results. For this users should have access to information of their interest from huge gathered data i.e. they require only certain aggregate functions. For example, getting information of average temperature of some particular place .In this case, we need to carry out distributed computation where we have to deal with aggregate functions. This overall process is carried out in In-Network computation. In-Network computation helps in improving communication efficiency as it allows the participation of intermediate node in computation of global functions. But in-network computation suffering the packet loss goes very disadvantageous. So hop by hop packet recovery for good reliability is essential. For achieving the same this paper has introduced a node caching procedure called as Active caching. In in-network aggregation each node collects the data of node under it and aggregate its own data and creates a new packet. But packet drops, unreliability, additional delay due to re-transmission are some of the challenges here. These challenges can be work out of using the new strategy.

Keywords*— data aggregation, in-network aggregation, congestion, Active cache.

I. INTRODUCTION

Wireless Sensor networks are collection of sensor nodes which sense the data and forward sensed data to base station[1]. Some of the properties of sensor nodes are that they are battery driven, an efficient utilization of power is essential in order to use networks for long duration hence it is needed to reduce data traffic in the sensor networks by reducing the amount of data that is to be send to base station. Data transmission and its aggregation are two simultaneous processes that should be carried out efficiently so that good throughput and energy efficiency is achieved. Power needed for processing is more as compared to transmitting data. It is preferable to do in-network processing inside network and reduce packet size and hence increase in battery life. This paper has detailed information about what challenges are faced during data transmission and aggregation. Many of the problems like energy constraints, redundancy, low data rate etc. are some of the challenges. We have tried to fix the challenge of transmission delay of data also the packet loss. Further part of this paper contain the active cache concept which is the proposed strategy we have used with distributed data aggregation which add efficiency in the existing results of throughput and packet drop ratio. Aim of efficient data aggregation is to collect the data in such a way that many of the challenges are reduced . To this end, we develop a new network architecture for in-network computation for a class of generalized maximum functions. We show that aggregation with wireless broadcast can substantially reduce the delay while satisfying the reliability constraint and the additional buffer capacity will also give efficient results and outcomes .These outcomes are compared to the existing system and simulation results are shown . Since WSN has limited energy resources, limited memory storage and battery power therefore data collection should be in energy efficient manner so that the network energy lifetime is increased.

1.1 Data Aggregation and Data transmission:

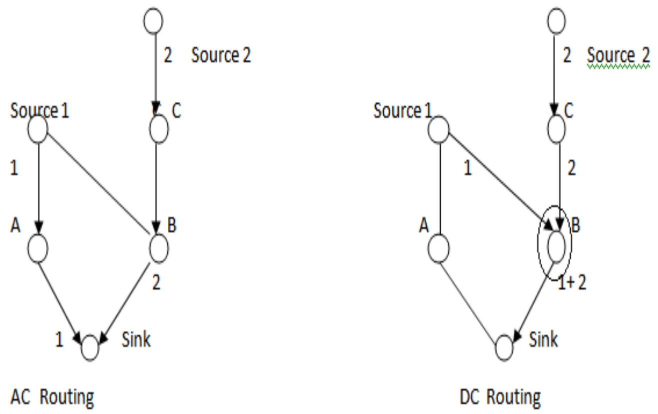
Data aggregation is the fundamental process which includes collecting and aggregating the useful data. Data aggregation is considered as one of the fundamental processing procedures for saving the energy. Data is to be gathered from the sensors at the source place to the sink. Data aggregation is basically categorized as *data centric* and *address centric*.

1.1.1 Address centric routing protocol

In this each source independently sends data to the shortest link possible to the sink.(end to end routing).The source with the address specified in the query, sends its data directly to the BS. It only contains receiving and sending.

1.1.2 Data centric routing protocol

The query is broadcast to all the nodes within its range in the network. Data packets are combined from the neighboring nodes are combined at a cylindrical node (Intermediate node) before sending it to Base station. Data centric approach is used in In-network computation where aggregation functions are calculated. Aggregation function calculation is actually when the intermediate node transmit single aggregate packet even if it addresses multiple input packet.



II. RELATED WORK

We have done the analysis of the data aggregation and congestion problem during that we came across many protocols and methods which are used to improve the data transmission. Redundancy Elimination for Accurate Data Aggregation in Wireless Sensor Networks (READA).

During the transfer of data and its aggregation there should be accuracy in data collection. This paper introduces A new algorithm for minimizing the redundancy in the data called Redundancy Elimination for Accurate Data Aggregation (READA).One more paper related to our research was the Active Caching: A Transmission Method to Guarantee Desired Communication Reliability In Wireless Sensor Networks In this paper[2],Active Networking (AN) technology is used to make congestion control more responsive to detect or recover congestion in WSN. They have designed a simple Active Backpressure (BP) mechanism to allocate bandwidth Proportional to the Size of tree (ABPS). In ABPS each data packet contains a program that tell nodes how to react to congestion, and quickly converges to a fair and efficient rate. They have implemented the ABPS in tiny OS itself .This paper discusses why we choose hop-by- hop control and active network. In previously implemented protocols they mainly focus on the end to end delay recovery and source limiting which requires more use of resources to deal with the congestion .So the Hop by hop strategy have an advantage over this. Hop-by-hop control is worthy in a network which may suffer a large delay and a high error rate. Each sensor detects congestion by monitoring the channel utilization and the buffer-occupancy level[3].

Aggregation [4] is the first step is comprising of First monitoring system followed by event detection. Here basically

1) We focused on the delay performance of in-network computation; 2) we consider reliability constraints in lossy wireless networks; 3) we investigate the effect of wireless broadcast on the delay performance.

Congestion control: Congestion control involves *Congestion detection*, congestion *notification* and *congestion avoidance [5]*. Different protocols are involved in various tasks to control congestion of data during its transfer. Congestion is the major problem which is focused here.

In congestion many protocols work at node level congestion Recovery of lost packets due to congestion can be in two way:

1 *End to End loss recovery*

2 *Hop by hop loss recovery*

In end to end in case of lost packets the retransmission request is send to the source . This causes end to end delay as here channel error accumulates exponentially over multi-hops FLUSH and RMST are some examples .In hop by hop (HBH) congestion recovery mechanism the lost packet retransmission request is send to intermediate node .In (HBH) the retransmission is possible from intermediate nodes because they cache the data packets into storage this reduce end to end delays.The well known HBH mechanisms are PSFQ and RMST .

III. PROBLEM STATEMENT

The problem of congestion can cause packet loss , queuing delay and blocking of new connections. When the source sends the huge amount of data ,it leads to link-level as well as node-level congestion causing packet loss. So transfer of data by avoiding congestion is the major challenge. After in-network computation the packet has more amount of information and if the packet get lost then there is the wastage of network bandwidth. The purpose is to reduce the node level congestion by adding cashes at intermediate-node. The experimental results show that the proposed framework can achieve good throughput and improved packet drop ratio. Hence this document holds detailed description of the system and mathematically describes the working of all components.

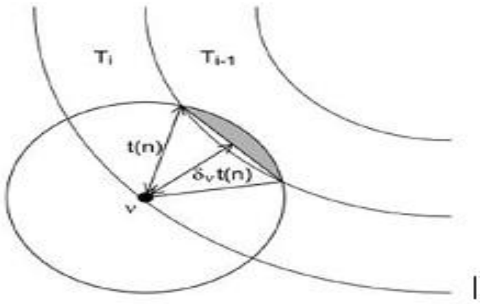
IV. EXISTING SYSTEM

In order to make Data aggregation efficient the existing system aimed at reducing the unnecessary redundant packets in the collected data. The distributed in network aggregation is the method behind this.

In-Network Aggregation: This is the technique which reduced the disadvantages in the direct delivery[6]. In direct delivery One problem is that a large number of packets must be sent to the host node. Since each node sends its own data to the host, there must be at least one packet of data sent per node. Additionally, since some nodes may not be able to communicate directly with the host, their data packets must be forwarded by other nodes until they reach the host. A related problem is that the size of each packet is relatively small since it only contains readings from one sensor. The increased number of small packet transmissions necessary to propagate needed data to the host severely limits the life of the network. In order to conserve both energy and bandwidth, it is useful to move the integration and filtering of sensor data into the network itself. In-network aggregation is a mechanism for reducing the overall amount of power and bandwidth required to process the users query by allowing sensor readings to be aggregated by intermediate nodes. As packets propagate from source nodes they may be combined together into fewer packets containing aggregate values. Because the user is not interested in individual values there is no loss in the quality of the result returned. A reduction in the likelihood of packet collisions and A reduction in the amount of redundancy received at the host node. In the existing system aggregation is simplified using the concept of tiered structure[7]. This algorithm describes distributed data aggregation. It is explained in detail.

Algorithm With Tired Structure

We assume that n wireless sensor nodes are uniformly deployed over a disk of radius 1. The network is a time-slotted TDMA system. At the beginning of each time-slot, each sensor node generates a packet with the sensed information. A time-slot is further divided into mini slots, and in each mini-slot, a single packet can be transmitted .Routing is simplified using the tiered structure Every node u in is a parent of node v in Ti+1 if its distance is no greater than t(n). Transmissions are scheduled from the outermost tier to the sink tier by tier one at a time, so that nodes in Ti can transmit only after all nodes in Ti+1 finish their transmissions. We group nodes in each tier into mutually exclusive subsets such that all nodes in a subset can transmit simultaneously. Let H(i,j) denote the jth subset in Ti, and let hi denote the total number of subsets in each tier such that all nodes in Ti can finish a single transmission in hi minis lots.



for $i = 1/\delta t(n)$ to 1 **do**
  for $j = 1$ to h_i **do**
    Each node ν in $H(i, j)$ broadcasts its (aggregated) information $(1 + r_b(n))$ times.
    if node $\mu \in T_{i-1}$ receives the packet **then**
      Node μ does aggregation and updates its information
    end if
  end for
end for

This Algorithm, each node has at least nt(n)2(4.0)parents and the maximum hop distance from a node to the sink is 1/t(n). Suppose that node v is located in Ti as shown in Fig. The number of parents of node v in Ti-1 is no smaller than the number of nodes in the shaded area. Since nodes are uniformly distributed, it can be easily shown that the number of nodes in the shaded area is bounded by nt(n)2.

***Motivation to have proposed system*:** In distributed In-network computation in case of Packet loss this strategy goes very disadvantages, as after several In-network computation of single packet contains much more information and loss of this affect the computation result. Therefore packet transmission needs a special care. Though re-transmitting is one of the solution but that results in additional delay. For limiting such additional delay, we have carried out a survey where we analyzed and worked on the delay performance of function computation also ensure reliability of packet in noisy channel .So that information is timely aggregated and transmitted with reliable transmission and less delay to improve sensor network reliability.

V. PROPOSED SYSTEM

In this paper we have used the concept of distributed in network aggregation for achieving some of the good parameters like reliability, network efficiency etc. In the course of collecting information, in network computations at intermediate forwarding nodes can substantially increase network efficiency by reducing the number of transmissions. On the other hand, it also increases the amount of the information contained in a single packet and makes the system vulnerable to packet loss. Instead of retransmitting the lost packets, which incurs additional delay, we have developed a wireless system architecture that exploits the diversity of the wireless medium for reliable operations i.e. in addition to this distributed in network data aggregation Buffered strategy is introduced and used as an additional thing to improve the delay performance. Whenever the data transmission is done each packet has the aggregated information. That packet get processed at each node. In the network nodes have limited storage so if one node is transmitting the data forward at that time it stores the incoming packet .As the new packet will received the storage buffer will get full and hence the storage overflow .This situation is called as node level congestion. Loosing that packet near the sink node will be the wastage of network resources ,The condition will be more adverse if tit happens in multi-hop network and again urge for the retransmission of that data packet will consume the additional energy consumption. And hence more energy wastage so to avoid all this the additional cache will prevent the storage buffer overflow by storing the extra unprocessed packet in the cache called active cache. Wireless broadcasting with the buffer cache [8] mechanism is an effective strategy to improve delay performance while satisfying reliability constraint and the additional buffering using the cashes will avoid the retransmission too.

System Design:

Architecture of System design is more clear in the diagram shown below. The following diagram describes the workflow of our system. Here nodes are aggregating data using distributed broadcast algorithm which we have discussed in the previous section(Tiered Algorithm)[9] and have additional cache. As we can see here each and every node caches the data to it's active cache. For aggregation purpose TDMA strategy is used.

These both strategies working together will reduce the congestion and also ensures data reliability.

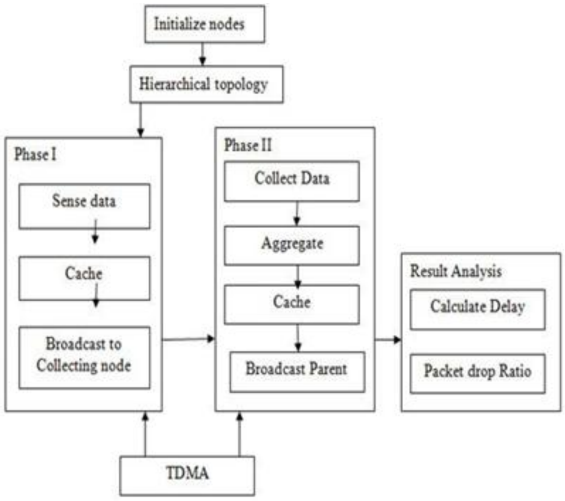


Figure: System Architecture.

Taking in consideration the second strategy. These diagrams will help in describing it in detail.

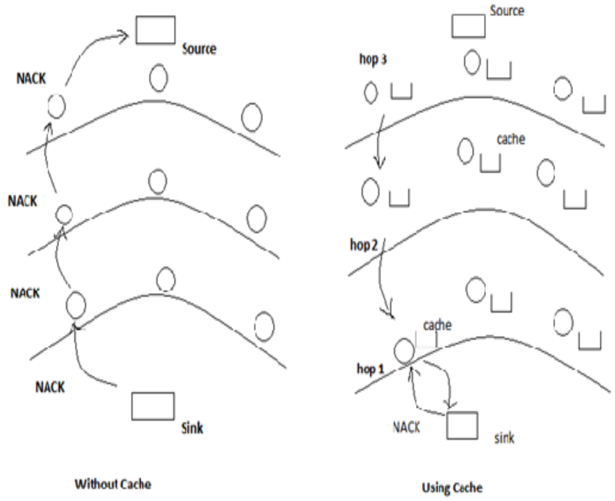


Figure2 : Diagrammatic view of proposed system

As shown in the diagram the firstly the network nodes don't have cache and on the other side there are the nodes caching the data. As the proposed schemes caches data packets at intermediate nodes over routing paths computed by CR to retransmit lost packets at intermediate nodes during multi-hop transmissions. Then in the first system of network the receiver will ask for the retransmission to previous node if the data packet is not available the previous node will ask the node previous to it. This will increase the overhead at system level. But the buffer strategy minimized the disadvantage .And hence worked out to the problem of reliability.

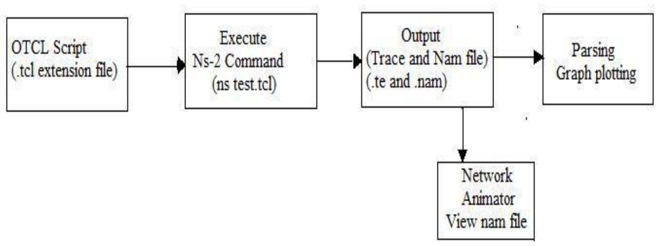
In the first and second diagram the scenarios with buffer and without buffer is shown. In the later case the retransmissions are reduced at very much extent. Suppose the situation when the packet is lost in between in multi hope transmission. When packet loss happens between a source node *n*1 and the caching node *n*5, the caching node requests retransmission to the source node[10]. When packet loss happens between the caching node and a destination node *n*7, the destination node requests retransmission to the caching node.

Implementation Details

Given a network of n sensor nodes uniformly and independently distributed on a disk of radius 1, each node has an identical transmission range t(n) and has the same order of parents x(n) . The sink is located at the center with antennas. Straight-line routing has been employed, thus achieving, and all the paths from a node to the sink have asymptotically the same length. Each and every node is collecting data and caching at individual level. They are using the distributed broadcasting algorithm. This distributed broadcast TDMA scheduling algorithm that allows each node to assign itself a time slot based on the two hop neighborhood information the main objective of this algorithm is to reduce the number of message transaction taking place between the nodes without compromising the convergence time of algorithm. Here each node is aggregating data in cache to prevent the packet drop out of buffer.

Evaluation: Simulation Result analysis

The Introduced strategy is checked for its performance on the simulator. For this we used the Network simulator 2 tool on red hat operating system. In this section, we validate the analysis through simulations and compare the performance of active caching (AC) with that of the in-network aggregation. Whole simulation study is divided into two part one is create the node i.e. NS-2 output it called NAM (Network Animator) file, which shows the nodes movement and communication occurs between various nodes in various condition or to allow the users to visually appreciate the movement as well as the interactions of the mobile node and another one is graphical analysis of trace file (.tr). Trace file contain the traces of event that can be further processed to understand the performance of the network.



Performance Metrics :

End-To-End Delay: End-To-End delay is the average time that takes by a data packet to reach its destination. This metric is calculated by subtracting time that first data take to traverse the network from time at which first data packet arrived to destination. This is a time the generate data packet by sender and it received by receiver at destination in application layer and it is measured in seconds. All delays in network is cause by node mobility, packet, retransmission and due to weak signal strength between nodes connection tearing and its making is also be included.

Packet Delivery Fraction: Packet delivery Fraction is defined as the ratio of data packets received by the destinations to those generated by the sources. Mathematically, it can be defined as: PDR= S1÷ S2 Where, S1 is the sum of data packets received by the each destination and S2 is the sum of data packets generated by the each source. Graphs show the fraction of data packets that are successfully delivered during simulations time International Journal of Advanced Science and Technology versus the number of nodes. Performance of the DSDV is reducing regularly while the PDF is increasing in the case of OLSR and ZRP. OLSR is better among the three protocols.

Packet Loss:

Packet loss in a communication is the difference between the generated and received packets. Packet Loss is calculated using awk script which processes the trace file and produces the result.

Residual Energy:

The energy model represents the energy level of nodes in the network. The energy model defined in a node has an initial value that is the level of energy the node has at the beginning of the simulation. This energy is termed as initial Energy_. In simulation, the variable "energy" represents the energy level in a node at any specified time. The value of initial Energy is passed as an input argument. A node looses a particular amount of energy for every packet transmitted and every packet received. As a result, the value of initial Energy in a node gets decreased. The current value of energy in a node after receiving or transmitting routing packets is the residual energy. Data Transmission is established between nodes using UDP agent and CBR traffic. Residual energy of the node is evaluated by accessing inbuilt variable "energy" in find Energy procedure at different times.

Jitter:

Jitter in IP networks is the variation in the latency on a packet flow between two systems, when some packets take longer to travel from one system to the other. Jitter results from network congestion, timing drift and route changes. *Jitter* is delay that varies over time. To measure Jitter, we take the difference between samples, then divide by the number of samples (minus 1).

For example, let us take collection of 5 samples with the following latencies: 136, 184, 115, 148, 125 (in that order).

The average latency is 142 - (add them, divide by 5). The 'Jitter' is calculated by taking the difference between samples.

136	to	184,	diff	=	48
184	to	115,	diff	=	69
115	to	148,	diff	=	33
148	to	125,	diff	=	23

(Notice how we have only 4 differences for 5 samples). The total difference is 173 - so the jitter is 173 / 4, or 43.25.

Energy consumption:

In simulation, the variable "energy" represents the energy level in a node at any specified time. The value of initial Energy is passed as an input argument. A node looses a particular amount of energy for every packet transmitted and every packet received. As a result, the value of initial Energy_ in a node gets decreased. The energy consumption level of a node at any time of the simulation can be determined by finding the difference between the current energy value and initial Energy_ value. If an energy level of a node reaches zero, it cannot receive or transmit anymore packets.

Simulations and Performance Analysis

Here we have some simulation results based on Graphs. All parameters are compared over the varying number of total nodes. The parameters like :

1)Packet delivery
2)Delay
3)Dropping Ratio
4)Average Energy Consumption
5)Residual energy
6)Jitter
7)Throughput

Following graphs show the improved behavior using the active cache strategy. As shown below in the graphs(fig. 6) by using indications viz. the black line ------ and the existing reading using the red line -----.

1)Packet Delivery has been increased as it can be clearly observed shown by using the black line ------ .Packet Delivery is more as compared to previous results

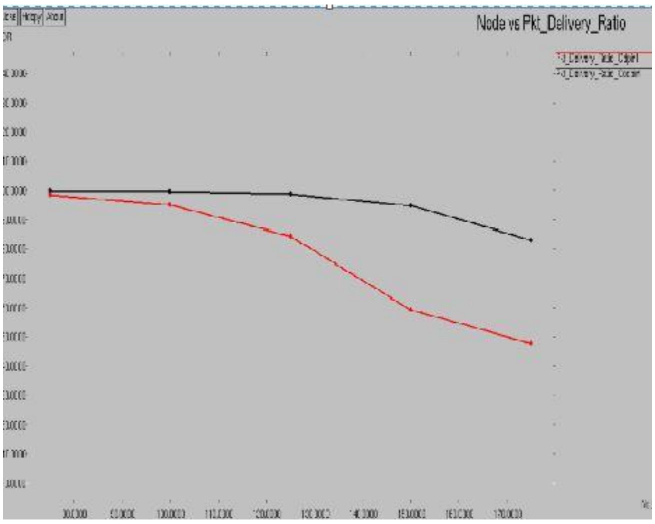


Fig. 6.1

2)Delay is low ,so is the improvement.

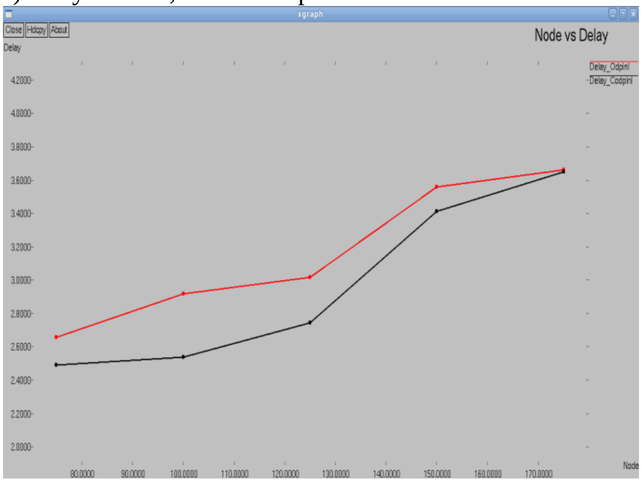


Fig. 6.2

3) Dropping Ratio this low as can be viewed below in graph. Using black line.

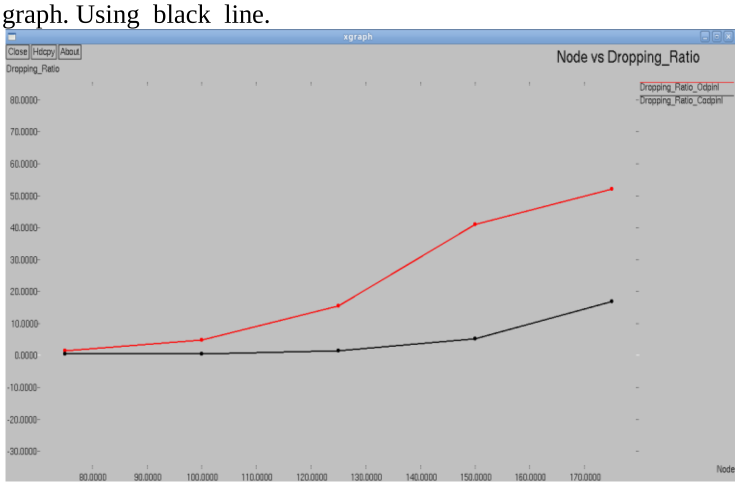


Fig. 6.3

4) Average Energy Consumption: Reduced energy consumption is shown using the back line .

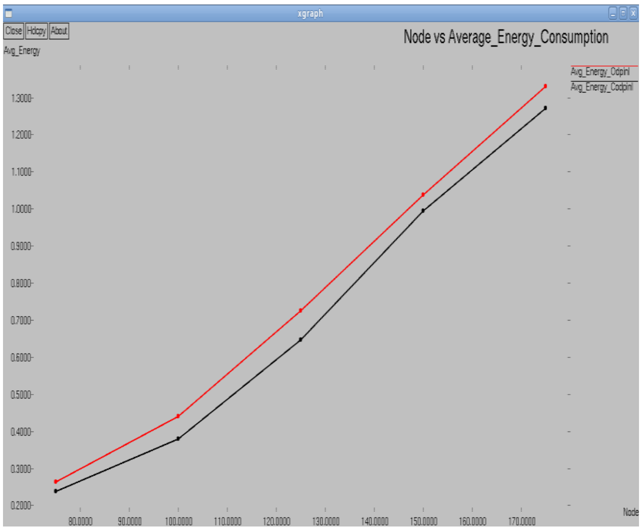


Fig. 6.4

5)Average Residual Energy: Residual energy is more as shown at the beginning and this is going decreasing down.

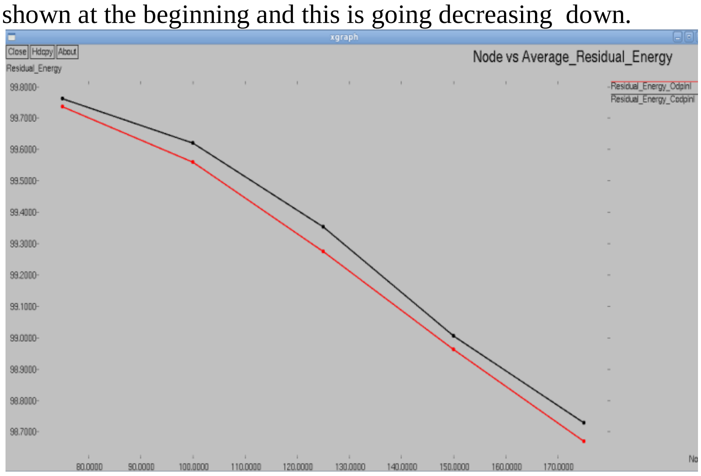


Fig. 6.5

6) Jitter: Latency Variation:

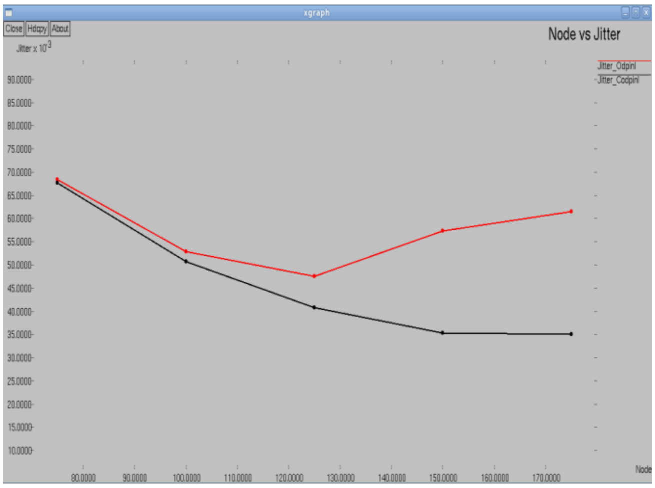


Fig. 6.6

7)Throughput **:** It is increased more than previous.

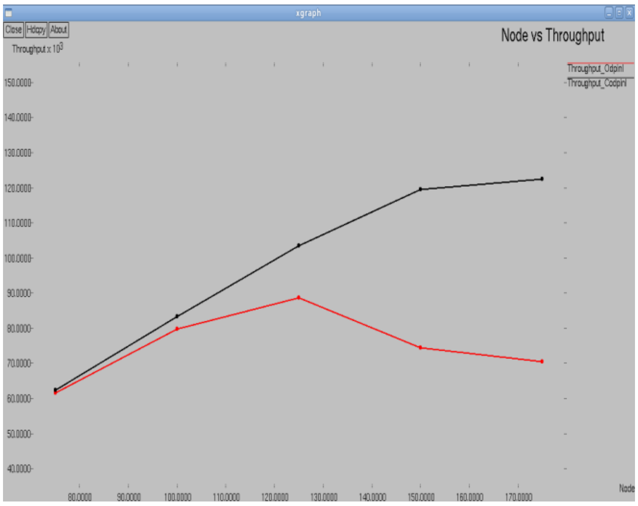


Fig. 6.7

This simulation is conducted in three different scenarios. In the first scenario the comparison of the existing and proposed is done for the varying no of packets, different node intervals and then for variable packet size And all the graphs show the efficient results. As above graphs show the reading in varying no of packets .

We have got following results as output: Table of simulation :

VI. CONCLUSION

Number of packets send	:3100
Number of packets received	3089
Packet delivery ratio:	99.6452
Delay	2.57985
Throughput	84350.3
Dropping Ratio	0.3548
Total Energy consumption	40.6641
Average Energy Consumption	0.48661
Total residual energy	9959.34
Average residual energy	99.5934

Common sensor nodes within the clusters, in cluster based WSN, ensure periodically or continuously, the event and the data packets are transmitted to the respective cluster head. The packet loss is recovered by retransmitting the lost packet from source. But this retransmission increases delay in network. Again network suffers from congestion due to duplicate data packets from different sensor nodes. So a proposed model of data broadcasting will over-come the requirement of retransmitting the data while data aggregation at collecting node will remove the duplicated at a from network. In a wireless sensor network in network aggregation can significantly improve efficiency when the goal of the network is to compute a global function. Since the loss of an aggregated packet is far more harmful than an un aggregated packet a higher level of protection is required or reliable operations in lossy wireless environments ,so wireless broadcasting issue fulfill for protecting aggregated information. Hence AC ensures the recovery reliable data transmissions. And also provides more flexibility between the End-to-End and memory requirements for a given reliability than the existing recovery mechanisms.

REFERENCES

[1] W.R. Heinzelman, J. Kulik, and H. Balakrishnan "Adaptive Protocols for Information Dissemination in Wireless Sensor Networks," Proceedings of the Fifth Annual ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom '99), Seattle, Washington, August 15-20, 1999, pp. 174-185.

[2] "READA: Redundancy Elimination for Accurate Data Aggregation in Wireless Sensor Networks" *Wireless Sensor Network*, 2010, 2, 300-308 doi:10.4236/wsn.2010.24041 Published Online April 2010 (http://www.SciRP.org/journal/wsn)

[3] Ying Ouyang, Fengyuan Ren, Chuang Lin, Tao He, Chao Li, Yada Hu, Hao Wen "*A Simple Active Congestion Control in Wireless Sensor Network*"

[4] R. Rajagopalan and P. K. Varshney, "Data-Aggregation Techniques in Sensor Networks: A Survey," IEEE Communication Surveys and Tutorials, Vol. 8, No. 4, December 2006, pp. 48-63.

[5] Xiaoyan Yin, Xingshe Zhou, Rongsheng Huang, Yuguang Fang, and Shining Li, A Fairness-Aware Congestion Control Scheme in Wireless Sensor Networks, *IEEE Transactions on Vehicular Technology, Vol. 58, No.9,* November 2009.

[6] I. Solis and K. Obraczka, .In-network aggregation trade-offs for data collection in wireless sensor networks,. *INRG Technical Report 102*, 2003, http://inrg.cse.ucsc.edu/techreports/tr102.pdf.

[7] On the Delay Performance of In-Network Aggregation in Lossy Wireless Sensor Networks Changhee Joo, Senior Member, IEEE, and Ness B. Shroff,Fellow, IEEE, IEEE/ACM TRANSACTIONS ON NETWORKING, VOL. 22, NO. 2, APRIL 2014.

[8] IEEE COMMUNICATIONS LETTERS, VOL. 13, NO. 6, JUNE 2009 Active Caching: A Transmission Method to Guarantee Desired Communication Reliability in Wireless Sensor Networks Dae-Young Kim and Jinsung Cho, *Member, IEEE*

[9] On the Delay Performance of In-Network Aggregation in Lossy Wireless Sensor Networks Changhee Joo, Senior Member, IEEE, and Ness B. Shroff, Fellow, IEEE, IEEE/ACM TRANSACTIONS ON NETWORKING, VOL.22, NO. 2, APRIL 2014.

[10] IJCSIS November/December 2016 issue (Vol. 14 No. 11) " Efficient data aggregation using buffered strategy to prevent congestion ".