HYBRID DATA AGGREGATION TECHNIQUES IN WIRELESS SENSOR NETWORK

*1Ms. Dharani R, *2 Ms. Lavanya S.

*1 M.Phil Research Scholar, Department of Computer Science Auxilium College (Autonomous),
Vellore, TamilNadu, India

*2Assisant Professor, Department of Computer Science Auxilium College (Autonomous),
Vellore, TamilNadu, India

---***---

Abstract - ***Wireless Sensor Network is most important building block for communications. Wireless sensor network faces some critical challenges like security, fault tolerance, scalability, heterogeneity, energy efficiency. Energy efficiency is one of the most studied challenges in Wireless Sensor Network by researchers. Different protocols or algorithms are designed to improve the energy of Wireless Sensor Network. Data aggregation is very crucial techniques in Wireless Sensor Network. With the help of data aggregation is energy consumption can be eliminated to a certain extended when Wireless Sensor Network is deployed in remote areas or hostile environment. Wireless Sensor Network have the most challenging task of life increases life time. So with help of data aggregation techniques its life time can be increased.***

Key Words: *Wireless sensor network, Data Aggregation, Data Communication, fault Tolerance, Security. Energy Efficient.*

I. INTRODUCTION

Wireless Networking which is comprised on number of numerous Sensors and they are interlinked or connected with each other for performing the same function collectively or cooperatively for the sake of checking and balancing the environmental factors. This type of Networking is called as Wireless Sensor Networking.

A Wireless Sensor Network consists of Sensor nodes deployed over a geographical area for monitoring physical phenomena like temperature, humidity, vibrations, and seismic events. Typically, a Sensor node is a tiny device that includes three basic components: a sensing subsystem for data acquisition from the physical surrounding environment, a processing subsystem for local data processing and storage, and a Wireless communication subsystem for data transmission. In addition, a power source supplies the energy needed by the device to perform the programmed task.individual values from each sensor. [6]For ex., In a Traffic Monitoring application where Noise levels are monitoring, the observer may be interested in the average value.

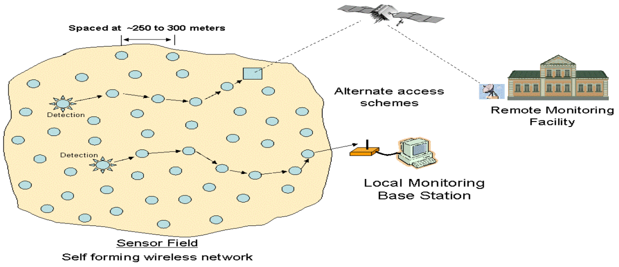

Spaced at ~250 to 300 meters
Detection
Detection
Alternate access schemes
Remote Monitoring Facility
Local Monitoring Base Station
Sensor Field
Self forming wireless network

Fig 1: wireless Sensor Network Architecture

Total working of Wireless Sensor Networking is based on its construction. Sensor Network initially consists of small or large nodes called as Sensor nodes. These nodes are vary in size and totally depend on the size because different sizes of Sensor nodes work efficiently in different fields

1.1Applications of WSN

- ***Area Monitoring:-***The WSN is deployed over a region where some phenomenon is to be monitored. A military example is the use of Sensors detects enemy intrusion; a civilian example is the geo-fencing of gas or oil pipelines.

- ***Environmental/Earth Monitoring:-***It has evolved to cover many applications of WSNs to earth science research. This includes sensing volcanoes, oceans, glaciers, forests etc. Some of the major areas are listed below

- ***Air Quality Monitoring:-*** The degree of pollution in the air has to be measured frequently in order to safeguard people and the environment from any kind of damages due to air pollution. In dangerous surroundings, real time monitoring of harmful gases is an important process because the weather can change rapidly changing key quality parameters.

- ***Interior Monitoring:-*** Observing the gas levels at vulnerable areas needs the usage of high-end, sophisticated equipment, capable to satisfy industrial regulations. Wireless internal monitoring solutions facilitate keep tabs on large areas as well as ensure the precise gas concentration degree.
- ***Exterior Monitoring:-*** External air quality monitoring needs the use of precise Wireless Sensors, rain and wind resistant solutions as well as energy reaping methods to assure extensive liberty to machine that likely to have tough access.
- ***Air Pollution Monitoring:-*** Wireless Sensor Networks have been deployed in several cities (Stockholm, London and Brisbane) to monitor the concentration of dangerous gases for citizens. These can take advantage of the ad hoc Wireless links rather than wired installations, which also make them more mobile for testing readings in different areas.
- ***Landslide Detection:-*** A landslide detection system makes use of a Wireless Sensor Network to detect the slight movements of soil and changes in various parameters that may occur before or during a landslide. Through the data gathered it may be possible to know the occurrence of landslides long before it actually happens.

II. RELATED WORK

The issues in Wireless Sensor Network have to be resolved are high energy consumption and low packet delivery rate. Energy conservation is a key issue in the design of systems based on Wireless Sensor Networks. It overcomes the problem in high energy consumption by introducing effective sleep awake mechanism to conserve energy. Energy efficiency is a critical feature of Wireless Sensor Networks (WSNs), because Sensor nodes run on batteries that are generally difficult to recharge once deployed. For target tracking one of the most important WSN application type's energy efficiency needs to be considered in various forms and shapes, such as idle listening, trajectory estimation, and data propagation.

A power source supplies the energy needed by the device to perform the programmed task. This power source often consists of a battery with a limited energy budget. In addition, it could be impossible or inconvenient to recharge the battery, because nodes may be deployed in a hostile or unpractical environment. On the other hand, the Sensor Network should have a lifetime long enough to fulfill the application requirements. In many cases a lifetime in the order of several months, or even years, may be required. External power supply sources often exhibit a non-continuous behavior so that an energy buffer (a battery) is needed as well. In any case, energy is a very critical resource and must be used very sparingly. Therefore, energy conservation is a key issue in the design of systems based on Wireless Sensor Networks.

2.1 Data Aggregation in Wireless Sensor Network

Data aggregation is the process of collecting and aggregating the useful data. Data aggregation is considered as one of the fundamental processing procedures for saving the energy. In WSN, data aggregation is an effective way to save the limited resources. A Sensor nodes that generates data, based on its sensing mechanisms observation and transmit sensed data packet to the base station (sink). This process basically direct transmission since the base station may locate very far away from Sensor nodes needs. More energy to transmit data over long distances so that a better technique is to have fewer nodes sends data to the base station. These nodes called aggregator nodes.

2.2 Process of Data Aggregation

Data aggregation is defined as the process of aggregating the data from multiple sensors to eliminate redundant transmission and provide fused information to the base station. All the aggregation nodes collect data from their children nodes and calculate the aggregation value. Then only the aggregated value is forwarded towards the data sink. The aggregate value may be average, maximum, minimum, summation, etc, which is calculated according to the application requirements. Data generated from neighboring Sensors is often redundant and highly correlated. Data aggregation usually involves the fusion of data from multiple Sensors at intermediate nodes and transmission of the aggregated data to the base station. The main objective of data aggregation is that remove redundant data broadcasting and improves the lifetime of energy in WSN.

2.3 Architecture of Data Aggregation

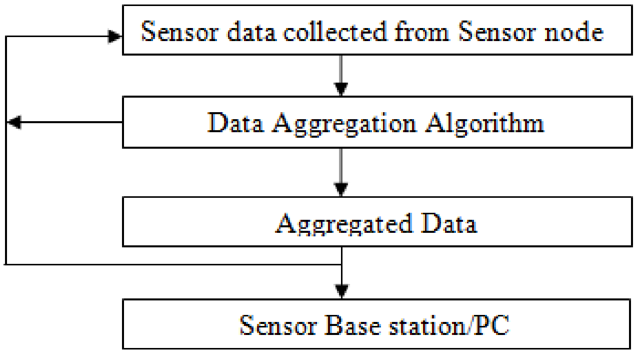

Sensor data collected from Sensor node
Data Aggregation Algorithm
Aggregated Data
Sensor Base station/PC

Fig 2: Architecture of Data Aggregation

2.3 ISSUES IN WIRELESS SENSOR NETWORKS

Wireless communication technologies the application interconnected by tiny sensor nodes become more and more attractive .However ,Killer application that take advantage of these technologies is yet to appear due to unsolved issues.[2]

Scalability: Potential application system might contain thousands, if not millions of sensor nodes .the cost such system is unaffordable. The affordability requirement imposes budget constraints on hardware components at each sensor node that directly affect the communication bandwidth.

Network Lifetime: The long life time requirement of application and the limited capacity of batteries create a wide gap between the node power consumption and the node power supply. Current solution include: Power Aware Protocols, Low Hardware Design, Power Saving Sleep, Mode and Transmission Range Optimization.

Functionality: In WSN some nodes are excepted to take multiple responsibilities of collecting various types of data processing and fusing data to improve communication efficiency, and relaying data via Multi Hop Transmission. This is the challenging requirement to the resource constrained of scalability of network life time for available technologies to make tradeoffs and optimization of several pairs of parameters such as the Network Scale and System Throughput.

III. DESIGN OF A DATA AGGREGATION

- ***No Data Aggregation:-*** In No Data Aggregation scheme, sensor devices are unaware of other neighboring nodes. Each sensor upon detecting an event attempts to send the amount of information collected, however small it may be, to the end nodes (sink).Sensor devices do not apply any data aggregation technique and simply forward the data packets toward the sink node.

- ***Network Data Aggregation:-***This scheme is highly suitable for environments where events have localized phenomenon, occurring in a fixed region of space. Such environments will hold for a wide variety of sensor network applications, since may external events are localized themselves. In this scheme, the sensor network environment is divided into pre-defined set of grids or regions. Each region or grid is responsible for observing and reporting events that occurs inside the region to the sink nodes. Also each sensor devices only inside the region. Only one sensor, the data aggregator, sends the critical information received either from other sensor devices or by itself to the sink nodes.

- ***Grid –Based Data Aggregation:-*** Grid-based Data Aggregation is highly suitable for mobile environments where the time duration of an event at a particular place is very small. Such scenarios will hold for a variety of sensor network applications like military surveillance, weather forecasting etc., in network Data Aggregation, the sensor network environment is divided into pre-defined set of grids or region.Each region or grid is responsible for observing and reporting events that occur insides the region to the sink nodes. One sensor device based on geographical positions with respect to either the sink or the center of the grid is chosen as data aggregator. All other sensors inside the grid are aware of this information. During event detection, all other sensors are supposed to send the event information to this data aggregator. The data aggregator after collecting data from other sensors sends only the critical information to the sink node.

3.1 Flow Chart for Data Aggregation Process

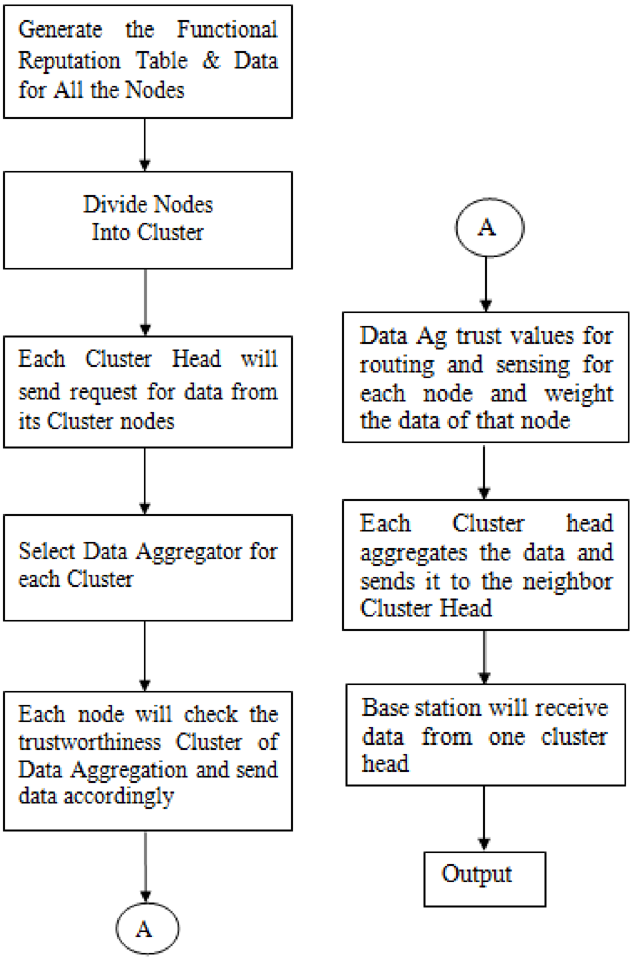

Generate the Functional Reputation Table & Data for All the Nodes
Divide Nodes Into Cluster
Each Cluster Head will send request for data from its Cluster nodes
Select Data Aggregator for each Cluster
Each node will check the trustworthiness Cluster of Data Aggregation and send data accordingly
A
A
Data Ag trust values for routing and sensing for each node and weight the data of that node
Each Cluster head aggregates the data and sends it to the neighbor Cluster Head
Base station will receive data from one cluster head
Output

Fig 3: Data Aggregation Process

IV. PROPOSED ANT COLONY OPTIMIZATION AND SVM

The Ant Colony Optimization (ACO) algorithm is a metaheursitic initially proposed by Marco Dorigo in his PhD dissertation in 1992. "The original idea comes from observing the exploitation of food resources among ants, in which ants' individually limited cognitive abilities have collectively been able to find the shortest path between a food source and the nest".It is firstly used to solve traveling salesman problem (TSP). Because of the characteristics of distributing computing, self-organization and positive feedback, ACO has been used in prior works for routing in Sensor Networks "Node Potential" is the heuristic used to evaluate the potential of next hop selection based on three factors: the candidate's distance to the sink node, its distance to the nearest aggregation node and its data correlation with the current node.

In this algorithm, random searching for the destination (sink node) is needed in early iterations. Use a simpler heuristic by only considering the distance to the sink node. An algorithm composed of path construction, path maintenance, and aggregation schemes including synchronization scheme, loop-free scheme, and avoiding collision scheme.

4.1 Algorithm Overview

In ACO algorithms, a colony of artificial ants is used to construct solutions guided by the pheromone trails and heuristic information. The original idea of ACO comes from observing the exploitation of food resources among ants. Ants explore the area surrounding their nest initially in a random manner. As soon as an ant finds a source of food (source node), it evaluates the quantity and quality of the food and carries some of it to the nest (sink node). During the back tracking, the ant deposits a pheromone trail on the ground. The quality of deposited pheromone, which may depend on the quantity and quality of the food, will guide other ants to the food source. The pheromone trails are simulated via a parameterized probabilistic model. The pheromone model consists of a set of parameters. In general, the ACO approach attempts to find the optimal routing by iterating the following two steps:

- Solutions are constructed using a node selection model based on a predetermined heuristic and the pheromone model, a parameterized probability distribution over the solution space.
- The solutions that were constructed in earlier iterations are used to modify the pheromone values in a way that is deemed to bias the search toward high quality solutions.

The algorithm runs in two passes: forward and backward. In the forward pass, the route is constructed by a group of ants, each of which starts from a unique source node. In the first iteration, an ant searches a route to the destination randomly. Later, an ant searches the nearest point of the previously discovered route. This could take much iteration before the ant can find a correct path with a reasonable length. A solution is flooding the sink node ID from the sink to all the Sensor nodes in the Network before any ant starts. The points where multiple ants join are aggregation nodes. In the backward pass every ant starts from sink node and travels back to the corresponding source node by following the path discovered in the forward pass. Pheromone is deposited hop by hop during the traversal.

4.2 FORWARD PASS

Each ant is assigned a source node. After that, an ant starts from the source node and moves towards the sink node using ad-hoc routing. The forward pass ends only if all the ants have arrived at the sink node. Single ant-based solution construction uses following steps:

- If the node has been visited in the same iteration, follow a previous ant's path
- Use a node selection rule
- If all the neighbors have been visited, use the shortest path
- If no neighbor nodes, backtrack to the previous node
- If no neighbor nodes and the previous node is dead, record the Network
- Lifetime and exit the program.

The current node sends the packet. The selected node receives the packet. Both Nodes update the residual energy after transmission. If the current node does not have enough energy to send, this transmission fails. The Network is maintained afterwards. Transmission failure is mostly prevented by doing a receiving and sending energy check in the node selection step.

4.3 BACKWARD PASS

Ants start from the sink node and move towards their source nodes. The ants follow the paths discovered in the forward pass. Before an ant arrives at its source node, the algorithm repeats:

- Retrieve the previous node in the path solution.
- Transmit the packet.

- If transmission fails, maintain the Network and terminate this ant.

- Encourage or discourage the node selection in the forward pass by depositing.

4.4 DATA AGGREGATION IN FORWARD PASS

Each Sensor node maintains two queues to store packets: a receiving queue and a sending queue. The packet sending process includes:

- Remove all the packets from the receiving queue.

- For"SinkDistNoAggre", push all the packets into the sending queue.

- For other aggregation algorithms, use the predefined function to aggregate all the received packets into one packet and push it into the sending queue.

- Among all the ants arrived at this node, select the earliest ant as the aggregating ant.

- The aggregating ant will finish the rest of the routing construction in this iteration.

- All the later arrived ants become aggregated ants. They remember the aggregating ant.

- Each aggregated ant shares its path with the aggregating ant. The aggregating ant updates its subsequent hops with all the aggregated ants.

Probability Calculation

When a node is ready to send a packet, it calculates the probability of all the neighbors using the equation below.

$$p_k(i,j)=\frac{\tau(i,j)\times\eta(i,j)^{\beta}}{\sum_{u\in N_i}\tau(i,u)\times\eta(i,u)^{\beta}}$$

In equation (1), an ant *k* having data packet in node *i* chooses to move to node *j* until the sink node, where *τ* is the pheromone, *η* is the heuristic, *Ni* is the set of neighbors of node *i*, and *β* is a parameter which determines the relative importance of pheromone versus distance (*β*>0). Value *η* is calculated using equation (2). Multiple factors can be used and each one is weighted.
After all the ants finish the forward passing and before they are going backward, the pheromone values on all the edges in the Network evaporate at rate ρ. The value is consistently reduced. Equation (3) shows how the evaporated pheromone value is calculated.

$$\tau_{ij} = (1 - \rho) \times \tau_{ij} \quad (2)$$

In the backward pass, each ant deposits or reduces the pheromone value on its own solution path. This step is different from the conventional ACO algorithm, in which pheromone is always deposited using the same rate. Encouraging or discouraging a node choice in the forward pass depends on the comparison of performance in the forward pass with the one of the best iteration found so far. The new pheromone is calculated using equation (4). Equations (5) and (6) are used to support equation (4).

$$\tau_{ij} = (\tau_{ij} + \rho\Delta\, \tau_{ij}) \times e0 \qquad (3)$$

$$\tau_{ij} = [\zeta + (h_i - h_j)] \times \Delta\omega_j \, (4)$$

$$\Delta\omega_j = \Sigma \qquad (5)$$

In equation (4), ρ is the pheromone decay parameter, τij is the pheromone value on the edge between nodes i and j, and e0 is the encouraging or discouraging rate derived from the forward pass. A path resulting in less energy consumption and smaller total hop-count is preferred. The best iteration is one with the least energy consumption and hop-count among all previous iterations. It is used as a control to calculate the e0 in the current iteration. If the forward pass is a failed path exploration or used more hop-count and energy consumption than the best iteration, the path is discouraged. Very small amount of pheromone is deposited on the edge to differentiate from those links not been visited, and e0 is set to a predetermined "Punish Rate," which is a relatively low rate between 0 and 1. If the forward pass found a path using the same hop-count and energy consumption as the best iteration, e0 is set to a relatively higher rate between 0 and 1-- the "encourage Rate." If the forward pass found a path with the same hop-count but less energy consumption than the best iteration, e0 = 1.5 × encourage Rate. If the forward pass found a path using less hop-count and energy consumption than the best iteration, e0 = hop-count difference × encourage Rate.

(1)

In equation (5), *ζ* is a positive number, h_i is the hop-count between node *i* and the sink, and h_j is the hop-count between node *j* and the sink. If the value of (*hi* - h_j) is greater than zero, it can be concluded that node *j* is closer to the sink node than node *i*. Therefore, the algorithm rewards the path from node *i* to node *j* by depositing more pheromone. If the value equals to zero, it means that both nodes *i* and *j* have the same hop-count to the sink, then the algorithm lays little pheromone on the path. If the value is less than zero, the algorithm does not lay pheromone on this path. In equation (6), R_j is the total hop-counts of these sources before vising node *j*. Therefore, $\Delta\omega_j$ is the total hop-counts of some sources to the sink through node *j*. The less the total hop accounts, the larger amount of pheromone is added on the path from node *i* to node *j*, as shown in equation (5).

This means that more ants are encouraged to follow this path. For an aggregation node, it updates the pheromone levels of its all neighbors by equation (4) when an ant moves to it. If a node does not have ants visit it within a limited time, its pheromone is evaporated according to equation (3).

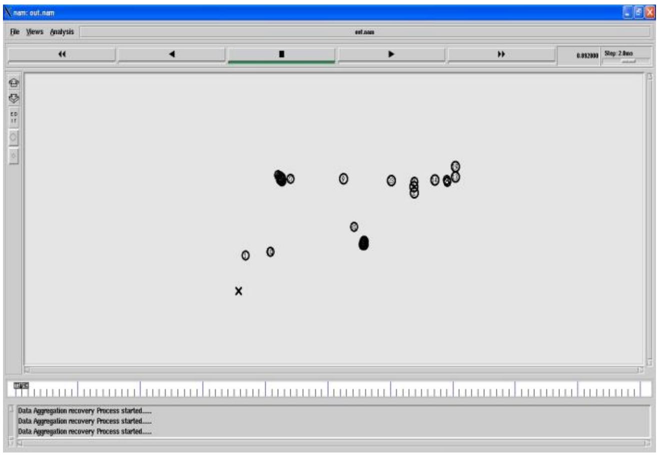
EVALUATION RESULT:

Nowadays, the WSN is a hot research topic. Many network details in WSNs are not finalized and standardized. Building a WSNs test bed is very costly. Running real experiments on a test bed is costly and difficulty. Besides, repeatability is largely compromised since many factors affect the experimental results at the same time. It is hard to isolate a single aspect. Moreover, running real experiments are always time consuming. Therefore, WSNs simulation is important for WSNs development. Protocols, schemes, even new ideas can be evaluated in a very large scale. WSNs simulators allow users to isolate different factors by tuning configurable parameters. [2] Consequently, simulation is essential to study WSNs, being the common way to test new applications and protocols in the field. This leads to the recent boom of simulator development. However, obtaining solid conclusions from a simulation study is not a trivial task. There are two key aspects in WSNs simulators:

EXPERIMENTS AND RESULTS

NS-2[NS-2_wiki,NS is the abbreviation of Network simulator version two, which first been developed by 1989 using as the REAL network simulator. Now, NS-2 is supported by Defense Advanced Research Projects Agency and National Science Foundation. NS-2 is a discrete event network simulator built in Object-Oriented extension of Tool Command Language and C++. People can run NS-2 simulator on Linux Operating Systems or on Cygwin, which is a Unix-like environment and command-line interface running on Windows. NS-2 is a popular non-specific network simulator can used in both wire and wireless area. This simulator is open source and provides online document.

However, this simulator has some limitations. Firstly, people who want to use this simulator need to familiar with writing scripting language and modeling technique; the Tool Command Language is somewhat difficult to understand and write. Secondly, sometimes using NS-2 is more complex and time-consuming than other simulators to model a desired job. Thirdly, NS-2 provides a poor graphical support, no Graphical User Interface (GUI) the users have to directly face to text commands of the electronic devices. Fourthly, due to the continuing changing the code base, the result may not be consistent, or contains bugs. In addition, since NS-2 is originally targeted to IP networks but WSNs, there are some limitations when apply it to simulate WSNs. Firstly, NS-2 can simulate the layered protocols but application behaviors. However, the layered protocols and applications interact and cannot be strictly separated in WSNs.

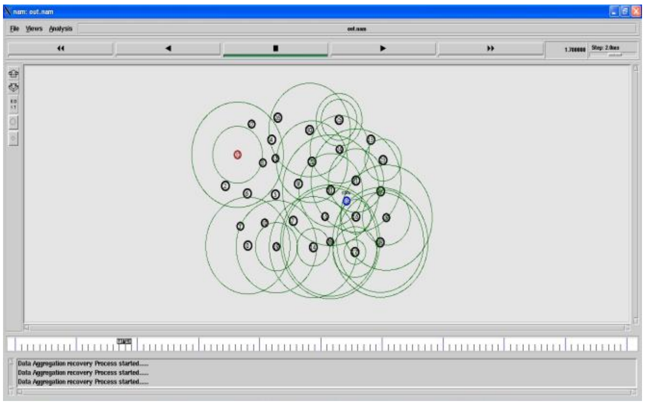


```
Data Aggregation recovery Process started.....
Data Aggregation recovery Process started.....
Data Aggregation recovery Process started.....
```

Red- Clustering; Blue & black- Sink; black – No of Node

Fig 4: Node Creation

For the experimental results presented in this section, a network of sensors randomly distributed in a 1000m * 1000m field is considered. The number of sensors in the network, i.e. the network size, is kept at 100, 200, 300 and so on. Each sensor has an initial energy of 1.Joule and the base station is located at (250, 330). Each sensor generates packets of size 1000 bits. The energy model for the sensors is based on the first order radio model.



```
Data Aggregation recovery Process started.....
Data Aggregation recovery Process started.....
Data Aggregation recovery Process started.....
```

Red- Clustering; Red & black- Sink; black – No of Node

Fig 5: Clustering Node

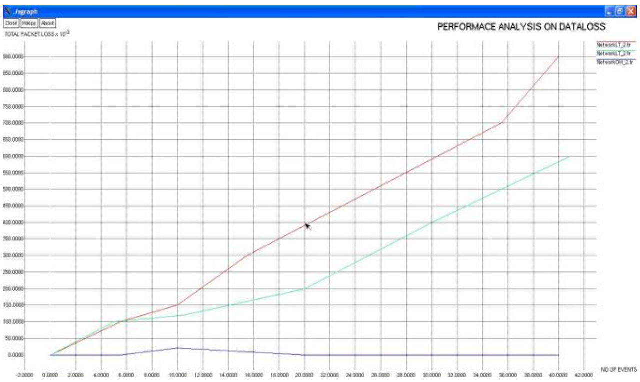


Fig 6: Energy Consumption

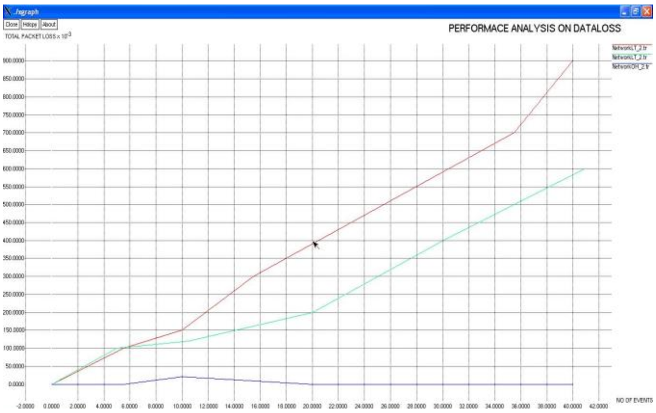


Fig 7: Data Loss

Conclusion

The data aggregation process is an important issue in WSN which indirectly affect network performance hence good optimization techniques is required. An efficient data aggregation not only enough along with the hybrid data aggregation and cluster approach used .Hence can save resource and increases the network life time., by using the optimization technique. An Ant Colony Optimization technique used to reach the destination through shortest path and to get fruitful message SVM classification is proposed.

Future Work:

In this method it is possible to prolong the Network life time and also get fruitful message. Since clustering and SVM based classification increases the overhead of cluster head, the use of distributed algorithm for electing the cluster head will further reduce the processing overhead, which also improves the network life time.

REFERENCES:

1. E.Fasolo, M.Rossi, J.Widmer, and M.Zorzi, *“In-Network Aggregation Technique for Wireless Sensor Netwroks:A Survey”*, IEEE Wireless Communications 2007.
2. Jun Yuea, Weiming Zhang, Weidong Xiao, Daquan Tang, Jiuyang Tang, *“Energy Efficient and Balanced Cluster-Based Data Aggregation Algorithm for Wireless Sensor Networks”* International Workshop on Information and Electronics Engineering (IWIEE),2012.
3. Rajeev Gupta, and Krithi Ramamirtham, *“Query Planning for Continuous Aggregation Queries over a Netwok of Data Aggregations”* IEEE Transcations on Knowledge and Data Engineering, June 2012.
4. Jang-Ping Sheu, Prasan Kumar Sahoo, Chang-Hsin Su and Wei-kai Huc, *“Efficient Path Planning and Data Gathering Protocols for Wireless Sensor Networks”*, Computer Communications, Elsevier Journal,2010.
5. Ozlem Durmaz Incel, Amitabha Ghosh, Bhaskar Krishnamachari, and Krishna ant Chintalapudi, “*Fast Data Collection in Tree-Based Wireless Sensor Networks*”, IEEE Transaction on Mobile Computing,2012.