INTERNATIONAL JOURNAL FOR ADVANCE RESEARCH IN ENGINEERING AND TECHNOLOGY
WINGS TO YOUR THOUGHTS.....

Improving/Enhancing Aggregation of Data Using Data Mining In WSN

Prerna Kharbanda[1], Er. Neeraj Madaan[2]

[1] Student
Haryana Engineering College, Jagadhri
pkharbanda35@gmail.com
[2] Sr. Lecturer
Haryana Engineering College, Jagadhri
neeraj.k.madaan@gmail.com

***Abstract**: WSN consists of spatially distributed autonomous sensors used to monitor physical or environmental conditions, such as temperature, sound, pressure, etc. and to cooperatively pass their data through the network to a main location. The development of wireless sensor networks was motivated by military applications such as battlefield surveillance; today such networks are used in many industrial and consumer applications, such as industrial process monitoring and control, machine health monitoring, and so on. Sensor nodes are capable of sensing and transmitting. They collect huge amount of data in a highly decentralized manner. The data collected contain all the information about the region. But sometimes users need only the specific information and for them rest of the information is treated as irrelevant. So, here we filter out that irrelevant data for the benefit of the users. In future, same can be used to extract the desired information from the set of large information.*

***Keywords:** Wireless Sensor Networks, Data Mining, Anomalies, Network Simulator, Network Animator and Xgraph.*

1. INTRODUCTION

We have witnessed the emergence of wireless sensor networks (WSNs) as a new information-gathering paradigm, in which a large number of sensors spread over a field and extract data of interests by reading real-world phenomena from the physical environment.[1] Nowadays sensors are very essential for today life to monitor environment where human cannot get involved very often.[2] A sensor network basically consist of a large number of sensor nodes. These sensor nodes are deployed either inside the phenomenon or very close to it. The sensor nodes position need not be pre-determined. This allows randomly deployment in inaccessible terrains or disaster relief operations. This also means that sensor network protocols and algorithms must possess self organizing capabilities. Another unique feature is the cooperative effort of sensor nodes. These sensor nodes use their processing abilities to locally carry out simple computations and transmit only the required and partially processed data. The sensor network collects the massive amount of data. To manage these data the appropriate data analysis is required. Therefore the two disciple sensor network and data mining can be combined. Knowledge from sensor data (Sensor-KDD) is important due to many application of crucial important to our society and large scale sensor system need to process heterogeneous and multisource of information from diverse type of instruments. The raw data of sensor need to be efficiently manage and transform to usable information through data fusion, which in turn must be induced tactical decision or strategic policy.[4]

It is more challenging for sensor network to sense and collect a large amount of data which are continuous over time, which in turn need to be forwarded to destination for further decision making process. Sensor data in the form of cluster act as a nucleus job of data mining. A clustering in wireless sensor network involves selecting cluster heads and assigning cluster members (sensors) to it for efficient data relay.[2] In data mining grouping a similar data is known as clustering which is a preparatory step for future data analysis.

IN a wireless sensor network, where sensors are geographically distant from each other, it may not be practical to require sensors to directly coordinate with each other to form a communication network due to the energy restriction. One possible solution is to employ a mobile robot, which can travel to all sensors, to download the data and finally return to its base station (starting position). In order to communicate with each sensor, the robot must present physically within its effective range, which is specified by a disk.[7] Sensor nodes can be classified into static sensor nodes and mobile sensor nodes. Currently, a wireless sensor network has focused on fixed sensor networks, in which the nodes are static. These Static sensor nodes cannot change position by themselves, after they have been placed in the sensing area. On the

other hand, mobile sensor nodes can change position autonomously that depending on their mission requirements. They are able to dynamically adjust network topology and promote the performance of sensor networks. Wireless Sensor Networks (WSN) is used in many real world applications like environmental and trajectory monitoring, traffic control. It is also used in several real life applications, especially in physical phenomena such as climate, Building structure and response to earthquakes.[2]

1.1 Challenges of Data Mining

Challenges in science and engineering, from the data mining perspective will focus on the following issues: (1) information network analysis, (2) discovery and usage of patterns and knowledge, (3) mining of stream, (4) Data mining of moving object , RFID , and data from sensor networks, (5) spatial, temporal and multimedia data mining, (6) text, Web, and other unstructured data mining, (7) cube-oriented multi-dimensional online analytical mining, (8) visual data mining, and (9) data mining by integration of sophisticated scientific and engineering domain knowledge.[3]

1.2 Anomalies of Wireless Sensor Network (WSN)

- WSN face rigorous resource constraints in communication bandwidth, power supply, and storage and processor capacity.
- Wireless sensors networks are typically highly limited in terms of sensing, computation, communication, battery life, and the actions they can perform.
- Higher-capability mobile robots may be dispatched to gather more accurate temperature or humidity readings.
- Mostly wireless sensor networks consist of a large number of static, low-power, short-lived and unreliable sensors.

1.3 Ad hoc on demand distance vector routing protocol

The Ad hoc On-Demand Distance Vector (AODV) protocol is one of the most popular reactive routing protocols. It is pure on demand routing protocol. This protocol enables dynamic, self-starting, multi hop routing among the mobile nodes in the mobile ad hoc networks. AODV allows mobile nodes to respond to link breakages and changes in network topology in a timely manner. The best thing about the AODV is that AODV provides the loop-free route and also by using the link state routing technique it removes the "counting to infinity" problem and provides quick convergence when the ad hoc network topology changes. AODV uses destination sequence number for maintaining each route entry. This destination sequence number is created by the destination. A requesting node always selects the route which has the greatest sequence number. AODV has three message types. Which are: Route Requests (RREQs), Route Replies (RREPs), and Route Errors (RERRs). When a sender wants to communicate with a node, it creates a RREQ packet and broadcast it to find a route to the destination.[9]

(a) Advantages of AODV

1. AODV protocol is a flat routing protocol it does not need any central administrative system to handle the routing process.
2. AODV tries to keep the overhead of the messages small. If host has the route information in the Routing Table about active routes in the network, then the overhead of the routing process will be minimal. The AODV has great advantage in overhead over simple protocols which need to keep the entire route from the source host to the destination host in their messages. The RREQ and RREP messages, which are responsible for the route discovery, do not increase significantly the overhead from these control messages. AODV reacts relatively quickly to the topological changes in the network and updating only the hosts that may be affected by the change, using the RRER message. The Hello messages, which are responsible for the route maintenance, are also limited so that they do not create unnecessary overhead in the network.
3. The AODV protocol is a loop free and avoids the counting to infinity problem, which were typical to the classical distance vector routing protocols, by the usage of the sequence numbers.
4. The AODV protocol will perform better in the networks with static traffic with the number of source and destination pairs is relatively small for each host.

(b) Disadvantages of AODV

1. Intermediate nodes can lead to inconsistent routes if the source sequence number is very old and the intermediate nodes have a higher but not the latest destination sequence number, thereby having stale entries.
2. Multiple Route Reply packets in response to a single Route Request packet can lead to heavy control overhead.
3. The periodic beaconing leads to unnecessary bandwidth consumption.[10]

2. RELATED WORK

In a wireless sensor network, where sensors are geographically distant from each other, it may not be practical to require sensors to directly coordinate with

each other to form a communication network due to the energy restriction. When data sensed by the nodes have to be transferred to the other nodes or sink then data being send also contain some irrelevant data that is also send along with the data desired. This results in the problem of irregularities in data. So, we have to detect those irregularities in order to improve the efficiency of the network.

The problem of irregularities detection is to find those sensory values that deviate significantly from the norm. This problem is especially important in the sensor network setting because it can be used to identify abnormal or interesting events or faulty sensors. We break this problem into two smaller problems.

One is to detect irregular patterns of multiple sensory attributes and the other to detect irregular sensory data of a single attribute with respect to time or space. The irregular multi-attribute pattern detection problem has the assumption that there are some normal patterns among multiple sensory attributes, which is true in some natural phenomena. Once these normal patterns are broken somewhere, the irregularity is detected and reported. In contrast, the irregular single-attribute sensor data detection problem examines the temporal and spatial characteristics of a sensor node and detects any irregularity in comparison with the node's previous data or the data of the neighbour nodes.

2.1 RESEARCH GOALS

1. Detection of sensor data irregularities is useful for practical applications as well as for network management, because the patterns found can be used for both decision making in applications and system performance tuning & it also avoid bulky data.
2. To identify abnormal or interesting events or faulty sensors.

2.2 METHODOLOGY

A new approach named pattern variation discovery is used to solve this problem. Our approach works in the following three steps:

1. Selection of a reference frame. This frame consists of the directions along which we want to look for irregularities among multiple sensory attributes. An analyst can explicitly specify the reference frame. It is also possible to discover the reference frame that results in a lot of irregularities.
2. Definition of normal patterns. This definition can be models of multiple sensory attributes or constraints among multiple attributes.
3. Discovery of irregularity. Whenever a normal pattern is broken at some point along the reference frame, irregularity appears. That is, the pattern variation happens.

2.2.1 Detection of sensor data irregularities

For example, we want to discover the irregular distribution pattern among multiple sensory attributes along time. Then, for each time point, we can put the values of a group of sensory attributes at a series of sensor nodes into a matrix, which represents a distribution status. The problem then becomes to discover the irregular matrix among a set of matrices. An irregular matrix represents that, at the corresponding time point, the distribution pattern of all the sensory attributes on all the nodes are irregular. Detection of irregularities is tightly interrelated to modeling of sensor data. Therefore, we propose to detect irregular single-attribute sensor data with respect to time or space by building models.

3. SIMULATION TOOL

The simulation tool used here is ns2. Ns2 stands for network simulator. **Ns** (from **network simulator**) is a name for series of discrete event network simulators, specifically **ns-1**, **ns-2** and **ns-3**. All of them are discrete-event network simulator, primarily used in research and teaching. ns-3 is free software, publicly available under the GNU GPLv2 license for research, development, and use.

NS-2:

NS-2 stands for Network Simulator version 2. NS-2 is a discrete event simulator for networking research. This simulator works at packet level. NS2 is simply an event driven simulation tool that has proved useful in studying the dynamic nature of communication networks. NS-2 uses a TCL as scripting language. Simulation of wired as well as wireless network functions and protocols (e.g., routing algorithms, TCP, UDP) can be done using NS2. In general, NS2 provides users with a way of specifying such network protocols and simulating their corresponding behaviors. Due to its flexibility and modular nature, NS2 has gained constant popularity in the networking research community since its birth in 1989. Ever since, several revolutions and revisions have marked the growing maturity of the tool, thanks to substantial contributions from the players in the field. Among these are the University of California and Cornell University who developed the REAL network simulator,1 the foundation which NS is based on. Since 1995 the Defense Advanced Research Projects Agency (DARPA) supported development of NS through the Virtual InterNetwork Testbed (VINT) project [9]. Currently the National Science Foundation (NSF) has joined the ride in development. Last but not the least, the group of researchers and developers in the community are constantly working to keep NS2 strong and versatile.

Volume 2, Issue VII July 2014
ISSN 2320-6802

INTERNATIONAL JOURNAL FOR ADVANCE RESEARCH IN ENGINEERING AND TECHNOLOGY

WINGS TO YOUR THOUGHTS.....

3.1 Basic Architecture

Figure 1 shows the basic architecture of NS2. NS2 provides users with executable command ns which take on input argument, the name of a Tcl simulation scripting file. Users are feeding the name of a Tcl simulation script (which sets up a simulation) as an input argument of an NS2 executable command ns. In most cases, a simulation trace file is created, and is used to plot graph and/or to create animation.

NS2 consists of two key languages: C++ and Object-oriented Tool Command Language (OTcl). While the C++ defines the internal mechanism (i.e., a backend) of the simulation objects, the OTcl sets up simulation by assembling and configuring the objects as well as scheduling discrete events (i.e., a frontend). The C++ and the OTcl are linked together using TclCL. Mapped to a C++ object, variables in the OTcl domains are sometimes referred to as handles. Conceptually, a handle (e.g., n as a Node handle) is just a string in the OTcl domain, and does not contain any functionality. Instead, the functionality (e.g., receiving a packet) is defined in the mapped C++ object (e.g., of class Connector). In the OTcl domain, a handle acts as a frontend which interacts with users and other OTcl objects. It may define its own procedures and variables to facilitate the interaction. Note that the member procedures and variables in the OTcl domain are called instance procedures (instprocs) and instance variables (instvars), respectively. Before proceeding further, the readers are encouraged to learn C++ and OTcl languages. NS2 provides a large number of built-in C++ objects. It is advisable to use these C++ objects to set up a simulation using a Tcl simulation script. However, advance users may find these objects insufficient. They need to develop their own C++ objects, and use a OTcl configuration interface to put together these objects. After simulation, NS2 outputs either text-based or animation-based simulation results. To interpret these results graphically and interactively, tools such as NAM (Network AniMator) and XGraph are used.

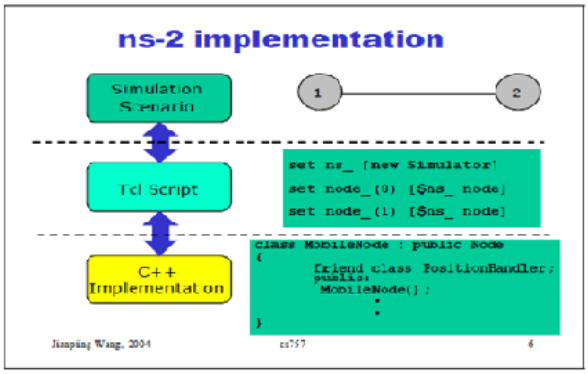

Figure 1: Ns2 implementation

4. RESULTS

We work in the field of data mining to support the better performance of the system by providing user the required information and rejecting/filtering the irrelevant data.

4.1 Parameters analyses in thirty nodes network with two different scenarios with respect to time:

1. Packets with Anomalies
2. Packets without Anomalies
3. Energy Consumption with Anomalies
4. Energy Consumption without Anomalies

4.2 Graphical Results:

1. Scenario One (For simulation time=25 sec):

1: Packets with & without Anomalies

Table 1: Table showing values of packets with and without anomalies for simulation time=25 sec

Simulation time	Packets with anomalies	Packets without anomalies
5	600	10
10	1500	200
15	1800	400
20	2050	600
25	2300	900

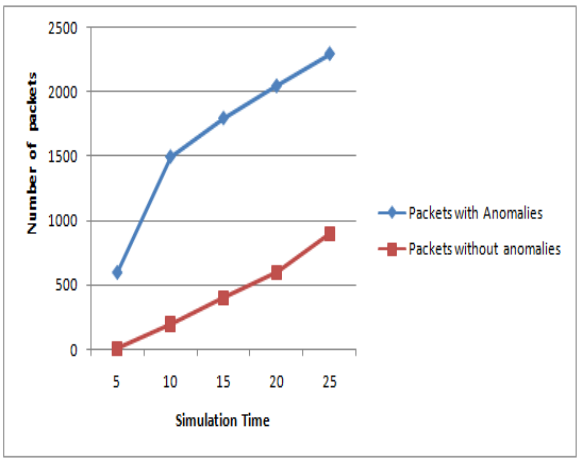

Figure 2: Graph showing packets with anomalies and without anomalies for simulation time=25 sec

From above we can say that Table 1 shows packets with and without anomalies with respect to the simulation time=25 sec. Figure 2 shows the graph that represents the values shown in Table 1.

2: Energy Consumption with & without Anomalies

INTERNATIONAL JOURNAL FOR ADVANCE RESEARCH IN ENGINEERING AND TECHNOLOGY
WINGS TO YOUR THOUGHTS.....

Table 2: Table showing values of energy with and without anomalies for simulation time=25 sec

Simulation Time	Energy with anomalies	Energy without anomalies
5	125	25
10	260	90
15	340	160
20	440	240
25	530	310

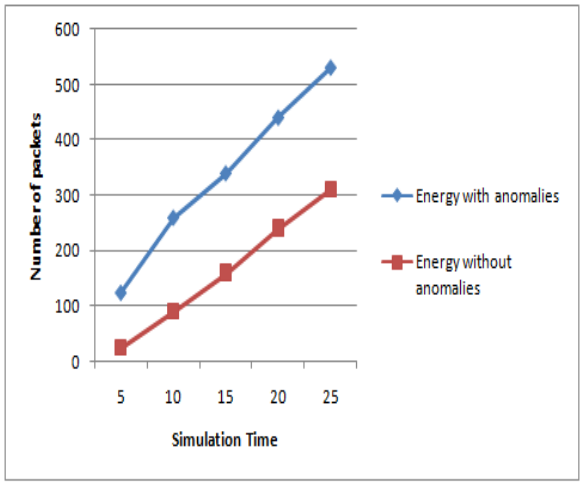

Figure 3: Graph showing energy with and without anomalies for simulation time=25 sec

From above we can say that Table 2 shows energy with and without anomalies with respect to the simulation time=25 sec. Figure 3 shows the graph that represents the values shown in Table 2.

2. Scenario Two (For simulation time =40):

1: Packets with & without Anomalies

Table 3: Table showing values of packets with and without anomalies for simulation time=40 sec

Simulation Time	Packets with anomalies	Packets without anomalies
10	1500	200
20	2000	600
30	2600	1150
40	3100	1600

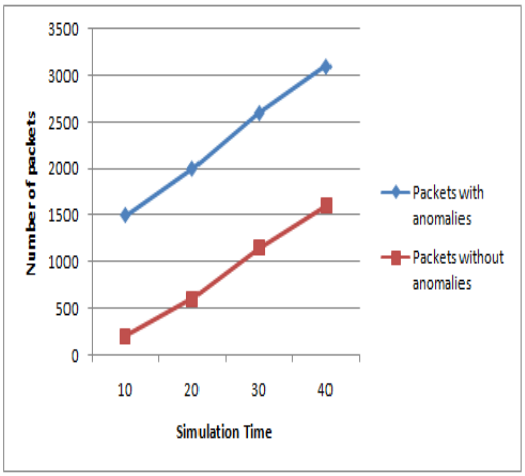

Figure 4: Graph showing packets with anomalies and without anomalies for simulation time=40 sec

From above we can say that Table 3 shows packets with and without anomalies with respect to the simulation time=40 sec. Figure 4 shows the graph that represents the values shown in Table 3.

2: Energy Consumption with & without Anomalies

Table 4: Table showing values of packets with and without anomalies for simulation time=40 sec

Simulation Time	Packets with anomalies	Packets Without anomalies
10	240	90
20	440	245
30	650	360
40	850	510

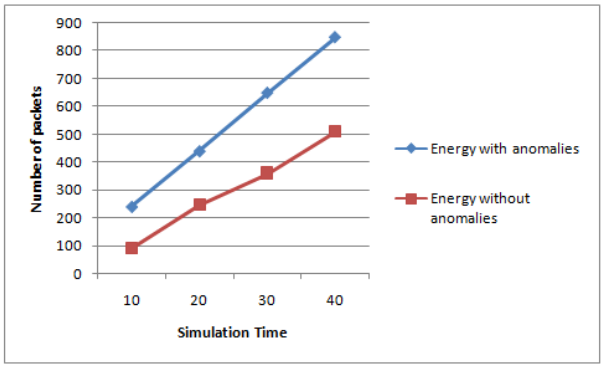

Figure 5: Graph showing energy with anomalies and without anomalies for simulation time=40 sec

From above we can say that Table 4 shows packets with and without anomalies with respect to the

simulation time=40 sec. Figure 5 shows the graph that represents the values shown in Table 4.

5. CONCLUSIONS AND FUTURE WORK

Sensor nodes are capable of sensing and transmitting. They collect huge amount of data in a highly decentralized manner. The data collected contain all the information about the region. But sometimes users need only the specific information and for them rest of the information is treated as irrelevant. So, here we filter out that irrelevant data for the benefit of the users. In future, same can be used to extract the desired information from the set of large information. AODV is being compared the proposed AODV. Proposed AODV has better performance as efficiency of the system is being increased by filtering out the unnecessary data. I compared proposed aodv with existing aodv protocol at different simulation time after that the result shows that the packets with anomalies are being filtered out using the data mining concept. And the proposed aodv is more energy efficient than the existing aodv. For getting the best results I have added the filtering concept.

Future aspects for the proposed system are bright. We can use the concept of clustering for filtering out the data with anomalies. Other mining concepts are also available.

REFERENCES

[1] Miao Zhao, Member, IEEE, and Yuanyuan Yang, Fellow, IEEE "Bounded Relay Hop Mobile Data Gathering in Wireless Sensor Networks" IEEE TRANSACTIONS ON COMPUTERS, VOL. 61, NO. 2, FEBRUARY 2012.

[2] S. Nithyakalyani (Department of Computer Science and Engineering, K.S.R College of Engineering, TamilNadu, India) and S. Suresh Kumar (Vivekananda College of Technology for Women, Tamilnadu, India) "Data Relay Clustering Algorithm for Wireless Sensor Networks: A Data Mining Approach" Journal of Computer Science 8 (8): 1281-1284, 2012 ISSN 1549-3636 © 2012 Science Publications.

[3] Laxmi Choudhary, Banasthali University, Jaipur "CHALLENGES FOR DATA MINING" IJREAS Volume 2, Issue 2 (February 2012) ISSN: 2249-3905.

[4] Khushboo Sharma, Manisha Rajpoot, Lokesh Kumar Sharma Department of Computer Science and Engineering, Rungta College of Engineering and Technology Kohka Road, Kurud, Bhilai, India "Nearest Neighbour Classification for Wireless Sensor Network Data" International Journal of Computer Trends and Technology- volume2Issue2- 2011

[5] Tzung-Cheng Chen, Tzung-Shi Chen, *Member, IEEE*, and Ping-Wen Wu " On Data Collection Using Mobile Robot in Wireless Sensor Networks" IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART A: SYSTEMS AND HUMANS, VOL. 41, NO. 6, NOVEMBER 2011.

[6] JANG-PING SHEU, KUN-YING HSIEH+ AND PO-WEN CHENG+ Department of Computer Science National Tsing Hua University Hsinchu, 300 Taiwan ,+Department of Computer Science and Information Engineering National Central University Chungli , 320 Taiwan "Design and Implementation of Mobile Robot for Nodes Replacement in Wireless Sensor Networks" JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 24, 393-410 (2008).

[7] Bo Yuan, Member, IEEE, Maria Orlowska, and Shazia Sadiq "On the Optimal Robot Routing Problem in Wireless Sensor Networks" IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 19, NO. 9, SEPTEMBER 2007.

[8] Rouhollah Maghsoudi, Somayye Hoseini, Yaghub Heidari, Department of Computer, Nour Branch, Islamic Azad University, Nour, Iran Payame Noor University of Shahrerey ,Tehran, Iran, sce.hoseyny@gmail.com Department of Electrical, Nour Branch, Islamic Azad University, Nour, Iran, " Surveying Robot Routing Algorithms with Data Mining Approach" The Journal of Mathematics and Computer Science Vol .2 No.2 (2011) 284-294.

[9] Neelam Khemariya and Ajay Khuntetha (2013), "An Efficient Algorithm for Detection of Blackhole Attack in AODV based MANETs", International Journal of Computer Applications (0975 – 8887) Volume 66– No.18, March 2013

[10] Rahul Sharm1, Naveen Dahiya and Divya Upadhyay (2013), "An Analysis for Black Hole Attack in AODV Protocol and Its Solution", IJCSMC, Vol. 2, Issue. 4, April 2013, pg.391 – 395.