ISSN: 2278 – 7798
International Journal of Science, Engineering and Technology Research (IJSETR)
Volume 2, Issue 4, April 2013

Hybrid Data Reduction Scheme for Energy Saving in Wireless Sensor Networks

M. Rajalakshmi M.E, R. Karthika M.E

Abstract*- In this paper efficient energy saving scheme and corresponding algorithm is designed and developed in order to provide reasonable energy consumption and to improve the network lifetime for wireless sensor network systems. A hybrid data-gathering protocol that dynamically switches between the event-driven data-reporting and data-driven data-reporting schemes is proposed. Typically, communication is the most energy-consuming operation apart from sensing and processing. In fact, it dominates the total energy consumption, so one should try to minimize the nodes' transmissions to reduce the total energy usage at the sensor nodes. Therefore, this metric can characterize the protocol in terms of energy consumption. The data-driven is used to reduce the amount of data delivered to the sink node. The novel aspect of this approach is that sensor nodes that seem to detect an event of interest in the near future, as well as those nodes detecting the event, become engaged in the adaptive transmission based data-reporting process. The purpose of deploying a WSN is to collect relevant data for processing and reporting. In particular, based on data reporting, WSNs can be classified as time-driven, event-driven and data driven. As such, the proposed protocol accurately analyzes the environment being monitored using only moderate resource consumption. The method makes a tradeoff between the energy consumption and data accuracy. The two algorithms named parameter-based Event Detection (PED) for dynamic switching and Parameter-based Area Detection (PAD) for area coverage are proposed and the effectiveness of the algorithm in a wireless scenario is analyzed.

***Index Terms*- Adaptive sampling, data driven, energy efficient, event driven, Moderate resource consumption, PED, PAD, Wireless sensor networks.**

I. Introduction

A wireless sensor network (WSN) [1] consists of spatially distributed autonomous sensors to monitor physical or environmental conditions such a temperature, sound, vibration, pressure and humidity. The WSN as shown in Fig.1 is built of nodes, each sensor network nodes has radio transceiver, a microcontroller and a battery. A sensor node might vary in size and correspondingly its cost. Size and cost constraints on sensor nodes result in corresponding constraints on resources such as energy, memory, computational speed and communications bandwidth.

The topology of the WSNs can vary from a simple star network to an advanced multi- hop wireless mesh network. The propagation technique between the hops of the network can be routing or flooding. Applications of WSN lie in the area of air pollution, greenhouse, area monitoring, machine health monitoring, water/wastewater monitoring, agriculture, and structural monitoring [8].

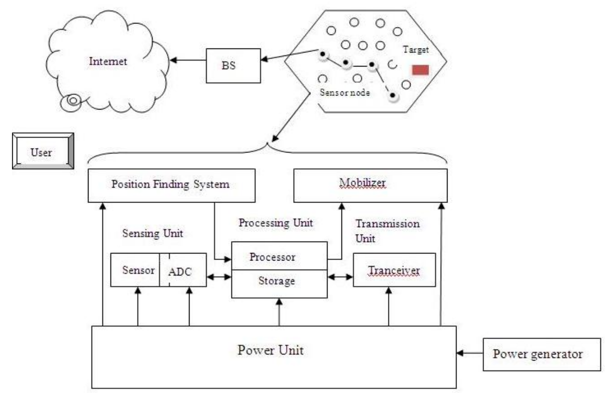

Fig. 1. Components of a Sensor Node.

A WSN consists of sensor nodes deployed over a geographical area for monitoring physical phenomena like temperature, humidity, vibrations, seismic events, and so on. 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 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.

The purpose of deploying a WSN is to collect relevant data for processing and reporting. In particular, based on data reporting, WSNs can be classified as time-driven, when sensor nodes periodically sense the environment and transmit the data of interest continuously over time, or as event-driven, when sensor nodes react immediately to sudden and drastic changes in the value of a sensed attribute due to the occurrence of a certain event [8]. As communication energy is a major contributor to total energy consumption and is determined by the total communication amount and transmission distance, the event-driven data-reporting scheme may lead to reduced energy consumption and, to increase network lifetime. However, as the amount of received data determines the accuracy level, use of the time-driven data-reporting scheme is more appropriate when higher accuracy is required. Therefore, it is essential to make a tradeoff between accuracy level and resource consumption.

Data-driven approaches can be divided according to the problem they address. Specifically, data-reduction schemes address the case of unneeded samples, while energy-efficient data acquisition schemes are mainly aimed at reducing the energy spent by the sensing subsystem. However, some of them can reduce the energy spent for communication as well. In addition, in this case, it is worth discussing here one more classification level related to data-reduction schemes. All these techniques aim at reducing the amount of data to be delivered to the sink node. However, the principles behind them are rather different. In-network processing consists in performing data aggregation (e.g., computing average of some values) at intermediate nodes between the sources and the sink. In this way, the amount of data is reduced while traversing the network towards the sink. The most appropriate in-network processing technique depends on the specific application and must be tailored to it. Data compression can be applied to reduce the amount of information sent by source nodes. This scheme involves encoding information at nodes, which generate data, and decoding it at the sink. There are different methods to compress data. As compression techniques are general (i.e. not necessarily related to WSNs), we will omit a detailed discussion of them to focus on other approaches specifically tailored to WSNs.

In this paper, we propose a hybrid data-gathering protocol that dynamically switches between the event-driven data reporting scheme and the data-driven data-reporting scheme. The proposed protocol shown in fig.2. behaves as event-driven, meaning that an event of interest triggers data dissemination by sensor nodes. However, from the point at which an event occurs to the point at which the event becomes invalid, the sensor nodes detecting the event continuously send data to an observer, thereby enabling accurate analysis of the environment. The novel aspect of our approach is that not only the sensor nodes that are detecting an event of interest but also those nodes that will potentially detect the event in the near future become engaged in the time-driven data-reporting process. This capability enables data from potentially relevant areas to be proactively gathered without requiring observer intervention. As such, the proposed protocol accurately analyzes the environment being monitored using only moderate consumption of valuable resources.

We envision that many real-life WSN applications will benefit from the proposed data-gathering protocol. For instance, consider a forest fire detection system. As soon as a fire breaks out, it is immediately reported to an observer because the sensor nodes in the burning area react to the rapid increase in temperature. By continuously receiving raw temperature data from not only the burning area but also the neighboring areas in which no fire has yet been detected but may soon travel, an observer is able to accurately identify where the fire originated, forecast where it may be heading, and determine where the fire extinguishing operation should focus. Similarly, an earthquake detection system is able to gather advance readings of seismic waves produced both at the epicenter and in surrounding areas and quickly raise an alarm for earthquake-prone buildings.

The contributions of this paper are twofold. First, we have developed an adaptive hybrid data-gathering protocol that allows a WSN to dynamically switch between the data-driven data-reporting scheme and the event-driven data reporting scheme based on node context; thus, it is able to more accurately analyze environments than the event-driven data-reporting scheme and use less energy than the time driven data-reporting scheme.

Second, we have evaluated the proposed protocol using significant simulation studies. One may argue that a combination of reactive and proactive data-reporting schemes may not be new in the literature. However, the proposed protocol differs from those of previous works in that time and space domain variables are taken into account when a WSN adapts its data-gathering process.

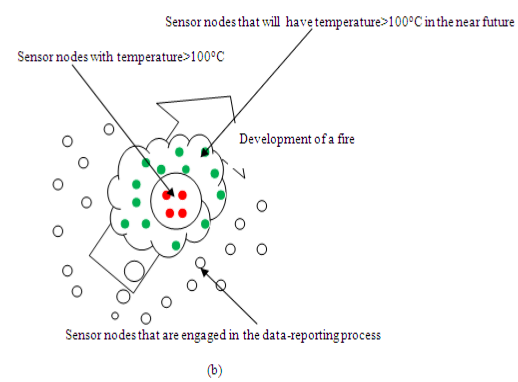

Fig. 2 proposed hybrid wsn.

II. Related Works

Most of the existing data-gathering techniques used in WSNs focus on routing messages in an energy- efficient manner. To minimize energy consumption, routing protocols proposed in the literature for WSNs employ some well-known routing tactics as well as tactics specific to WNSs,e.g., data aggregation and in-network processing, clustering, different node role assignments, and data-centric methods.

A typical WSN consists of hundreds to thousands of sensor nodes deployed over an area the dense deployementment of sensor nodes leads to high correlation of the data Sensed by the neigh- boring nodes. Hence, the main idea of data aggregation and in-network processing is to combine the data from different nodes en route by eliminating redundancy, thereby minimizing the number of transmissions. This process ultimately saves energy and prolongs the network lifetime.

Cluster-based routing [9],[10] divides the entire system into distinct clusters, compresses data arriving from nodes that belong to the respective clusters, and sends an aggregated message to the base station. Challenges faced by such a clustering-based approach include how to select the cluster heads and how to organize the clusters. In particular, as it is more likely that cluster heads will drain their batteries faster than cluster members, leading to non uniform depletion of energy in the sensor network

HEED **is** proposed to prolong the network lifetime and energy efficiency. HEED is a clustering based data collection protocol, where the residual energy of the node is used for the selection of cluster head. This method keeps the node from getting drained due to high data processing task. HEED only helps in reducing the excessive load, but the regular data transmission burden remains same. So significant energy saving is not possible. Moreover clustering burdens also may not reduced. HEED can achieve energy conservation only if the network is static. But our proposed scheme proved to be efficient in both the static and dynamic environments.

With regard to the dynamic switching of data-reporting schemes, APTEEN [5] and SINA are previous studies that are closely related to the current research. APTEEN uses a hard and a soft threshold that allow sensor nodes to transmit data only when the sensed attribute is within the range of interest. It also parameterizes the maximum time period between two successive reports sent by a sensor node (i.e., count time) so that the node is forced to transmit sensed data if it has not done so for a long period of time. As such, APTEEN combines reactive and proactive data-reporting schemes. However, in APTEEN, only a parameter in the time domain determines when a proactive data-reporting scheme is used. As a consequence, every sensor node in the network participates in a proactive data-reporting process regardless of its relevance to an event of interest. On the other hand, our approach considers variables in the space and time domains to make such decisions. In doing so, sensor nodes that detect an event of interest or those nodes that are likely to detect the event in the near future become engaged in a proactive data-reporting process. This capability can significantly eliminate unnecessary data dissemination from sensor nodes that are irrelevant to the event .

SINA [7] selects the most appropriate data distribution and collection method based on the nature of the queries and the cur- rent network status. Specifically, individual sensor nodes make autonomous decisions about whether they should participate in the information-gathering process based on the probability of a given response. Furthermore, sensor nodes are able to defer sending data for a period of time. These methods maximize the response quality in terms of number and responsiveness while minimizing network resource consumption.

A new hybrid system for data gathering is the need of the time. In the proposed method Nodes that detect an event of interest or those nodes that are likely to detect the event in the near future become engaged in a proactive data reporting through spatio-temporal correlation of data. Since no clustering is required, the over heads are completely avoided. The proposed work doesn’t rely on another node to reach the cluster. This will further improve the energy efficiency. Here also the clustering overheads are reduced to a great extent.

III. Hybrid Data-Gathering Protocol

The essential elements of the hybrid data-gathering protocol proposed in this paper are that: 1) it switches dynamically between the event-driven data-reporting scheme and the time driven data-reporting scheme, and 2) sensor nodes that will detect the events in the near future, which typically are in close proximity to those nodes detecting the events, are also engaged in the time-driven data-reporting process.

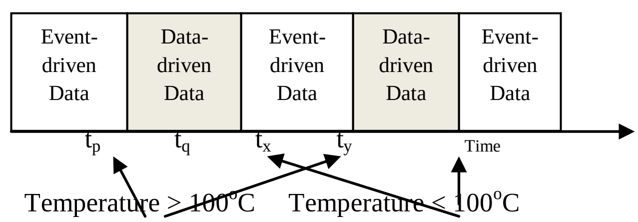

Fig . 3. D ynamic switching of data-reporting schemes.

Fig. 3 illustrates the key protocol characteristics. Under normal conditions, sensor nodes respond only when the measured temperature is above 100 °C. However, once sensor nodes realize that the abnormal phenomenon is not transient (e.g., at tp and tx), they switch to the data-driven data-reporting scheme and continuously disseminate temperature data to an observer. Furthermore, they notify other nodes of their changes so that neighboring nodes continuously disseminate data as well. Similarly, when the temperature goes below 100 °C (e.g., at tq and ty), the nodes switch back to the event-driven data-reporting scheme.

The proposed system by default is in the event driven mode. Event driven data gathering is mainly applied to detect the occurrence of an event. If an event occurs the system initiates the task of reporting. This helps in wireless sensor node to preserve the battery energy. The system is in idle listening mode until an event occurs. On the occurrence of an event, the system switches to data driven reporting scheme. To fulfill these requirements, there must be a mechanism to determine in a timely manner when to switch between the two data-reporting schemes and which sensor nodes to involve in the time-driven data-reporting process. Depending on these answers, several policies can exist. In this paper, we present two algorithms: the PED algorithm and the PAD algorithm.

A.PED Algorithm

As depicted in Fig.4, the PED algorithm is given a threshold value, a threshold variable, and two counter-variables controlling the aggressiveness level for changes in data-reporting schemes. The threshold value determines the occurrence of an event of interest. That is, when the value of the sensed attribute is beyond the threshold value, a sensor node must report it to a base station. The threshold variable is the average of the values of the sensed attribute over a time interval. The two counter-variables, p and q, are defined as follows:

1) p: the number of consecutive time intervals with increasing (for starting the time-driven data-reporting scheme) or decreasing (for stopping the time-driven data-reporting scheme) slope of the threshold variable;
2) q: the number of consecutive time intervals that the threshold variable is above (for starting the time-driven data-reporting scheme) or below (for stopping the time driven data-reporting scheme) the threshold value, regardless of slope.

Two pairs of constants, P_{start} and Q_{start}, and P_{stop} and Q_{stop}, are parameters of the algorithm such that $P_{start} \leq Q_{start}$ and $P_{stop} \leq Q_{stop}$. The PED algorithm works as follows. Suppose that the current data-reporting scheme of a sensor node is event-driven. A sensor node periodically computes the average of the sensed attribute over a recent time window and updates the two counter-variables, p and q, accordingly.

If $(p \geq P_{start})$ or $(q \geq Q_{start})$, a sensor node switches to the time-driven data reporting scheme and reports the sensed attribute continuously over time. To eliminate the risk of a transient response, the two counter-variables are reset when the average value of the sensed attribute goes below the threshold value. Similarly, when the time-driven data-reporting scheme is used, the tests on P_{stop} and Q_{stop} are executed to determine when to switch back to the event-driven data-reporting scheme. Note that the failure of tests on both P_{start} and Q_{start} does not affect the normal behavior of the sensor networks. That is, the data capturing an event are still propagated all the way to a base station.

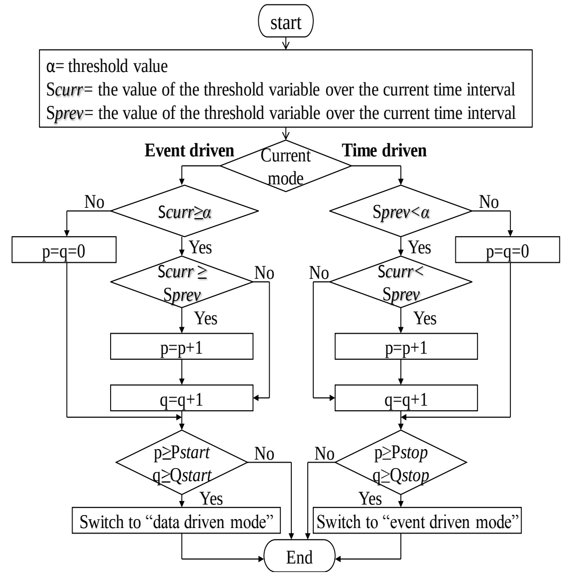

Fig. 4.Flow chart of PED Algorithm

As P (e.g., Pstart and Pstop) requires that the value of the sensed attribute be monotonically increasing (or decreasing) for P consecutive time epochs, the use of P enables a rapid response to a change in an attribute, and prevents a response to a transient change due to, e.g., sensor malfunctioning. If the test on P fails, the test on Q (e.g., Qstart and Qstop) is applied, which is more forgiving (e.g., the requirement that the slope is monotonically increasing or decreasing is relaxed). Therefore, Q permits some fluctuation of measurements and is more conservative. Overall, smaller values of P and Q lead to more aggressive action.Therefore, continuous data dissemination starts because $q \geq Q$start . Similarly, at t16, both p and q become 3. Consequently, continuous data dissemination ends because $p \geq P$stop.

B.PAD Algorithm

Once a sensor node switches to the time-driven data reporting scheme, it broadcasts its change to engage neighboring sensor nodes in the continuous data dissemination. The PAD algorithm, which is based on two configurable parameters, determines the range of the neighborhood: time-to-live (TTL) and valid time (VT).

TTL represents the number of hops within which sensor nodes must switch to the time-driven data-reporting scheme. The use of TTL in the PAD algorithm is similar to that in computer network technology, where TTL specifies the number of hops that a message can travel to before it should be discarded. When a sensor node receives a broadcast message containing a TTL, value that is greater than zero, it switches to the time-driven data-reporting scheme and rebroadcasts a TTL value decremented by one. This process continues until the TTL value becomes zero . This approach is based on the argument that the nearer a sensor node is located to the sensor nodes that detect an event, the more likely it is that the sensor node will be relevant to that event in the near future since sensor nodes are formed according to proximity. For instance, the current temperature measured by sensor nodes located close to the fire may not be high enough to trigger an event, but it would be much higher than those of nodes tens of miles away from the fire. Therefore, the possibility that the closely located sensor nodes will detect an event in the near future increases.

VT is the other important parameter of the PAD algorithm, and it specifies how long a sensor node should use the time driven data-reporting scheme regardless of the result of its PED algorithm. Note that VT is used only for those sensor nodes that are switched to the time-driven data-reporting scheme by TTL. The necessity of VT arises from the fact that sensor nodes near the area where an event occurs may not yet detect the event. Therefore, without VT, they could immediately switch back to the event-driven data-reporting scheme, losing the chance to acquire important information in advance.

We believe that our approach behaves well in multiple event-detection environments, in which sensor nodes are capable of sensing different attributes through simple modification of the PAD algorithm. For instance, a sensor node executes the PED algorithm for each sensed attribute and broadcasts its switching to the time-driven data-reporting scheme to neighboring nodes accordingly. By sending not only TTL and VT but also the type of attribute that caused the switching, the sensor node allows neighboring nodes to determine what type of data they need to collect in a data-driven manner.

IV. ADAPTIVE TRANSMISSION ALGORITHM

In a wireless sensor node transmission of data consumes highest amount of energy. By reducing the no. of transmissions with respect to the change in data, helps in reducing the energy consumption and maintaining the required data accuracy. The transmission interval between two consecutive data is determined by change in data magnitude between two consecutive samplings. The sampling frequency is maintained constant. The decision to transmit or not to transmit the data is based on the change in data. Based on the data pattern, an apt level of transmission frequency is determined on line. We propose an adaptive transmission algorithm that estimates online the optimal transmission frequencies for sensor nodes. This approach, which requires the design of adaptive measurement systems, minimizes the energy consumption of the radios and, incidentally, that of the network while maintaining a very high accuracy of collected data.

Algorithm : ATA

1. Store the *W* initial samples that come from the process in the *Dataset;*
2. Set T_f as $T_{max.}$ (Transmission Frequency)
3. Estimate Δ (change in data) from acquired samples
4. calculate T_f as $T_f = \Delta T_{med.}$
5. Check if $T_f \geq T_{max}$, if so set $T_f = T_{max.}$
6. Check if $T_f \leq T_{min,}$, if so set $T_f = T_{min.}$
7. Else $T_f = \Delta T_{med.}$
8. Calculate Δ

ATA does not assume any hypothesis with regard to the nature of the signal (e.g., stationary); moreover, its computational load is acceptable for mid complexity WSN units. ATA identifies the minimal transmission frequency online, which guarantees the reconstruction of the sampled signal; thus, it reduces the power consumption of the measurement phase by adapting the transmission frequency to the real needs of the physical phenomena under observation. By decreasing the number of transmitted samples, ATA also reduces the amount of data that will be transmitted and, as a consequence, the energy that the radio consumed. In addition, the proposed approach can be integrated with other techniques for energy conservation by acting at different abstraction levels (e.g., data aggregation and/or compression).

V. EXPERIMENTAL RESULTS

In order to evaluate the effectiveness of our approach, we implemented the hybrid data-gathering protocol on MATLAB, an event-driven simulator. The simulator can operate in either the deterministic mode to produce replicable results while testing an application or in the probabilistic mode to simulate the nondeterministic nature of the communication channel and the low-level communication protocol of the sensor nodes.

We used two metrics to analyze and compare the performance of the proposed protocol from the perspective of energy consumption and the ability to provide a dataset for accurate analysis. They are as follows.

1) The total number of data transmissions in the network (N_{comm}); among three sensor node operations (i.e., sensing, computing, and communicating), communication is the most energy-consuming operation. In fact, it dominates the total energy consumption, so one should try to minimize the nodes' transmissions to reduce the total energy usage at the sensor nodes. Therefore, this metric can characterize the protocol in terms of energy consumption.
2) The ratio of sensor nodes in Neighboring Area 1 that be- come engaged in continuous data dissemination during the current time interval (R_{nodes}). This metric represents how many sensor nodes that will potentially detect an event over the next time interval are engaged in continuous data dissemination during the current time interval. As this metric value increases, it becomes more likely that accurate analysis of the environment being monitored will be possible,

ISSN: 2278 – 7798
International Journal of Science, Engineering and Technology Research (IJSETR)
Volume 2, Issue 4, April 2013

especially for cases in which an event of interest moves.

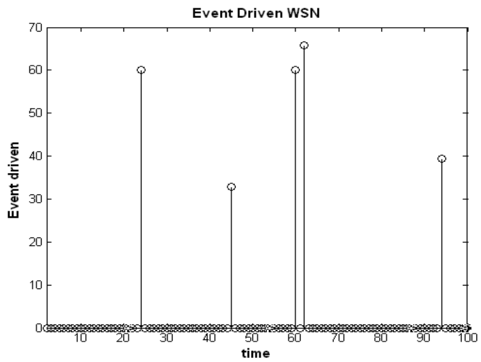

Fig. 5.Event driven Data gathering

For a rare and asynchronous event, reporting event driven data gathering complies well. This scheme reports to the sink on the occurrence of an event. This method conserves a significant amount of energy at the cost of accuracy. This scheme can lessen the data amount and requires low band width. Fig. 5. Shows event driven data gathering with low dynamic signal, the more dynamic the data, lesser the energy conservation. If the signal is dynamic lesser will be the energy conservation.

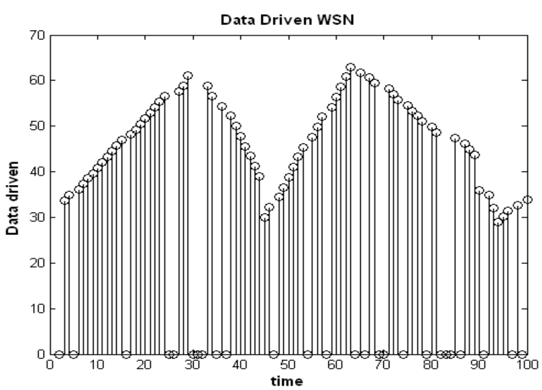

Fig. 6.Data driven Data gathering

This scheme reports to the sink on the occurrence of an change in data. This method conserves a significant amount of energy at the cost of accuracy. This scheme can lessen the data amount and requires low bandwidth. Fig. 6.shows data driven data gathering with data integrity.

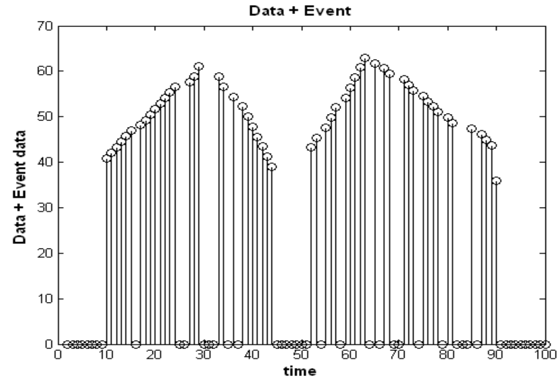

Fig. 7.Proposed Hybrid Data Gathering

When the data gathering schemes are combined, the amount of data can be reduced to a considerable level at the cost of minimal reduction in data accuracy as shown in Fig.7. This scheme initiates data driven data gathering only if the data goes beyond threshold value, otherwise it remains in event driven data gathering scheme. This method helps in analyzing the data when the data value is greater than threshold value.

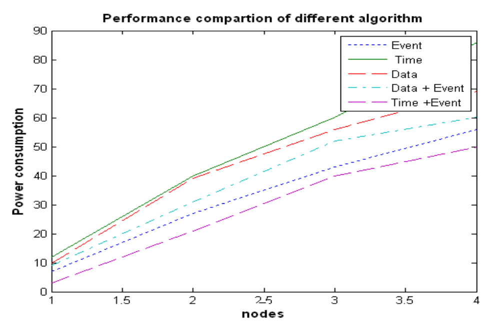

Fig. 8.Performance comparison of different Data Gathering with respect to power.

The results of hybrid data gathering outperforms data driven data gathering in energy conservation and event driven data gathering in data accuracy. The results are encouraging us towards energy efficient data collection with data integrity. The system enhances the lifespan of WSN's. This scheme can be applied to any type of WSN's topology .In this, no clustering is needed and cluster overheads are avoided and the computational load is very low which is very much apt for energy constrained WSN's.

V.CONCLUSION

The hybrid algorithm for data gathering is designed and simulated. From the above results we observe that the modified PED and PAD algorithms using only in moderate resource consumption. The hybrid algorithm achieves data accuracy and adequate data for detailed analysis at moderate resource consumption. The hybrid data gathering scheme achieves 40% energy reduction than the time driven data gathering. Hybrid algorithm outperforms data driven data gathering in energy conservation and event driven algorithm in data accuracy. The tradeoff between data accuracy and energy conservation is left to the user. Based on the application requirement, the threshold for the switching between event driven and time driven can be adjusted. The adaptive transmission scheme further reduces the data amount by adapting itself to the change in data parameter. The proactive inclusion of neighboring sensor nodes further reinforces the data accuracy and direction of parameter intensification.

REFERENCES

[1]. Akyildiz IF, Su W, Sankarasubramaniam Y and Cayirci E,"Wireless sensor networks: a survey". Springer Computer Networks. Vol 38 pp: 393-422, March 2002.

[2] J. Al-Karaki and A. Kamal, "Routing techniques in wireless sensor networks: A survey," *IEEE Wireless Commun.*, vol. 11, no. 6, pp. 6– 28, Dec. 2004.

[3] Byoung-Dai Lee, "An Energy-Efficient Hybrid Data-Gathering Protocol Based on the Dynamic Switching of Reporting Schemes in Wireless Sensor Networks," *Systems Journal,IEEE.*, vol. 6, no. 6, pp 378-387, Sep.2012

[4] M. Iiayas and I. Mahgoub, Eds., Handbook of Sensor Networks: Compact Wireless and Wired Sensing Systems. Boca Raton, FL: CRC Press,2005.

[5] A. Manjeshwar and D. P. Agarwal. 2002. APTEEN: A hybrid protocol for efficient routing and comprehensive
information retrieval in wireless sensor networks. Parallel and Distributed Processing Symposium, Proceedings International, IPDPS, pp. 195-202.

[6] S. Lindsey, C. Raghavendra. 2002. PEGASIS: PowerEfficient Gathering in Sensor Information Systems. IEEE Aerospace Conference Proceedings, Vol. 3, 9-16 pp. 1125- 1130.

[7] C. Srisathapornphat, C. Jaikaeo, and C. Shen, "Sensor information networking architecture," in *Proc. Int. Workshop Parallel Process.*, 2000, pp. 22–30.

[8] M. Iiayas and I. Mahgoub, Eds., Handbook of Sensor Networks: Compact Wireless and Wired Sensing Systems. Boca Raton, FL: CRC Press, 2005.

[9] Q. Fang, F. Zhao, and L. Guibas, "Lightweight sensing and communication protocols for target enumeration and aggregation," in *Proc. 4th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, 2003, pp. 165–176.

[10] W. Heinzelman, A. Chandrakasan, and H. Balakrishnam, "Energy-efficient communication protocol for wireless microsensor networks," in *Proc. 33rd Hawaii Int. Conf. Syst. Sci.*, 2000, pp. 10–19.