Outlier tolerant adaptive sampling rate approach for wireless sensor node

Sunil Kumar Selvaraj1, Venkatramana Bhat Pundikai2
1Department of Computer Science and Engineering, Mangalore Institute of Technology and Engineering, Affiliated to Visvesvaraya Technological University, Belagavi, India
2Department of Computer Science and Engineering, Nitte Mahalinga Adyanthaya Memorial (NMAM) Institute of Technology, Nitte, India

Article Info
Article history:
Received Jan 25, 2021
Revised Apr 19, 2022
Accepted May 20, 2022

Keywords:
Adaptive sampling approach
Clustering
National instruments wireless sensor node 3202
Outlier
Wireless sensor network

ABSTRACT
Nowadays the wireless sensor network (WSN) has been used for variety of applications and still lot of research in progress around the corner for the betterment of the wireless sensor network technology. In this paper, one such issues related to energy consumption in sensor node due to fixed sampling interval of sensing unit and its impact on redundant data is discussed with a possible solution. The association of sampling interval and its impact on energy dissipation in sensor node enforces the need for study on energy efficient adaptive sampling interval approach. The lack of serious consideration of outlier in sensor data degrades the performance of the existing adaptive sampling interval approach. The result of the proposed approach of in-network clustering algorithm shows the better efficiency towards detecting the outlier in real time. The results also showcase the better efficiency of proposed approach in terms of rapid optimization of sampling interval compared to simple variance based approach.

This is an open access article under the CC BY-SA license.

Corresponding Author:
Sunil Kumar Selvaraj
Department of Computer Science and Engineering, Mangalore Institute of Technology and Engineering, Affiliated to Visvesvaraya Technological University
Belagavi, India
Email: sunil@mite.ac.in

1. INTRODUCTION
The popularity and increased usage of wireless sensor network (WSN) for various remote applications makes the wireless sensor network as one of the emerging technologies of recent days [1]. The wireless sensor network comprised of sensor nodes connected via wireless technologies, and each sensor node is comprised of microcontroller, sensing unit, radio transceiver and limited power supply unit. The limited energy source for remotely deployed sensor node poses the various challenges for the smooth functioning of deployed wireless sensor network [2]. Most of the wireless sensor network challenges are related to energy efficient data collection and transmission to base/sink station.

The sampling rate at which samples are collected by sensing unit of the sensor node plays a reasonable impact on the energy dissipation of sensor node. The higher sampling interval rate leads to large measurement data from sensing unit and correspondingly higher the energy dissipation. The collection of larger amount of data due to high sampling rate always ends up with a better analysis, but which in turn increases the chances of redundancy among data. Determining the optimal value for sampling rate plays a vital role in sensor node towards minimizing the energy dissipation without compromising the quality of data. Instead of a fixed sampling interval rate, the value for sampling interval rate needs to be adaptive with respect to the situation under monitor and must be sensitive to variance on real time measured data.

Journal homepage: http://ijeecs.iaescore.com
In adaptive sampling interval approach, the variable sampling interval value will be used and the value for sampling interval keeps changing according to the variance among measured data. The limited energy at the sensor node poses the need of adaptive sampling interval approach for dynamically adjust the sampling interval rate without compromising the quality of data and minimizes energy consumption. Most of the existing adaptive sampling approach relies on the detection of variance or change in successive data to adjust the sampling interval rate for the next round of measurement without removing the outlier data.

The outlier means, wrong or improper measurement data due to sensor node implicit problems or external effect on object under observation. Some of the implicit reasons for outlier in sensor node readings are power failure, connection issue, malfunction of equipment. Similarly some of the explicit reason for occurrence of outlier on sensor node measured data are raining, fire or due to any natural calamities and man mishandling of equipment. The unavoidable common occurrence of outlier in the sensor node measurement data leads to the need of outlier tolerant sensor network mechanism or algorithms. The non consideration of a possible outlier on adaptive sampling interval approach may ends up in improper values for sampling interval, which in turn highly impacts the quality of data and very importantly high energy consumption. This paper discusses one such work towards, the outlier tolerant adaptive sampling interval approach and showcases the impact of outlier on adaptive sampling interval approach.

Limited power supply on the sensor node has been a major issue and has pushed a lot of research towards building an energy efficient wireless sensor network [3]. Even various studies have been carried to achieve energy efficient sensor network in terms of energy efficient routing, data aggregation and sampling frequency based on sleep or wake method [4]. The applications of structural monitoring using energy hungry sensor changed the orientation of energy consumption in wireless sensor network. Some of the energy hungry sensor poses challenges of higher energy consumption than the radio module [5].

Lots of research work was being carried out on energy efficient adaptive sampling interval approaches without compromising the quality of data. Most of the research work on the adaptive sampling interval approach concentrated by incorporating various method such as kalman filter [6], temporal variation based correlation analysis [7], reverse sigmoid function [8], association rule between sensing parameter [9], Kruskal-Wallis test [10], cumulative sum (CUSUM) test [11] for detecting the change in data to adjust the sampling interval. Andre et al. [12] used simple standard deviation to detect the change and dynamically adjusted sampling interval for the specific region of interest participating a sensing device instead of the entire sensing unit. Harb and Makhoul [13] compare and analyse the Bartley test, jaccard similarity and Euclidean distance based approaches for detecting the change and dynamically adjusting the sampling interval. In continuation of advancement in CUSUM test for change detection, upper and lower limit threshold concepts are used to minimize the impact of small variation captured by CUSUM test [14].

Some of the work used the concept of the prediction model to reduce the number of samples and inversely reducing sampling interval rate [15]. The lack of consideration of available energy into consideration may put the above approaches at risk for real field wireless sensor network applications. In continuation of research, various approaches have been proposed on considering the available energy while adjusting the sampling interval in accordance to the sampling interval approach [16]-[19]. The rash environmental factors and limited energy constraint of wireless sensor network poses the risk for high probability of outlier data [20]. The impact of outlier data on the end application pushes the outlier detection and removal into important challenging task in Wireless sensor network [21]. Many researchers have been carried out for detecting and removing outlier in wireless sensor network using various approaches based on classification, clustering and temporal, special correlation analysis. In classification based approaches sensor data are predefined into various types of data and upcoming data are classified accordingly [22], [23]. In clustering based approach, the exploratory analysis is carried out for grouping the data based on their similarity and formed cluster with less density is considered as the presence of outlier [24], [25].

In temporal and special co-relation based outlier detection, sensor data are compared with the previous data and neighbour sensor data to conclude the change as normal data or outlier [26]. Even though many researches were carried out on outlier detection in wireless sensor network, the emergence of IoT and streaming wireless sensor data poses the another challenge of outlier detection in real time streaming data. Yu et al. [27] used the dynamic time wrapping concept to detect the anomalous behaviour in real time streaming data. Kontaki et al. [28] used the distance as a measure for forming the clustering of real time streaming data and density function to evaluate the formed cluster outlier as non outcluster head. Yadav and Ahamad [29] applied the machine learning algorithm for predicting the outlier in wireless sensor network. Aalaoui et al. [30] applied the efficient cluttering approach by electing the cluster head in optimized way by choosing the cluster head with minimum distance to the base station.
2. SYSTEM DESIGN AND PROPOSED APPROACH

In the proposed approach, the real time in-network clustering algorithm is used for detecting the outlier and change in successive sensor readings. In the proposed system, three sensor nodes are used, formally represented as Sensor Node={S_1, S_2, S_3}, where as S_1=programmed with a clustering based approach, S_2=programmed with a variance based approach, S_3=Programmed with a fixed sampling interval. All the three sensor nodes are interfaces with vernier temperature probe and send temperature data to national instruments wireless gateway node 9791.

The architectural block diagram of the proposed setup as shown in Figure 1, the gateway node upon collecting the Sensor reading from sensor node S_1 and S_2 computes the new sampling interval based on the change detection using clustering and variance approach respectively. The proposed clustering based adaptive sampling interval approach consists of the following stages of operation

- Distance based network cluster formation on real time sensor readings.
- Density based approach for evaluating the formed cluster.
- Nullifying the impact of outlier as change in sampling interval adjustment.
- Adjusting the sampling interval based on the detection of second dense cluster.

![Figure 1. Architectural block diagram of the proposed system](image-url)

In the distance based network clustering, sensor readings are collected in a period of 100 readings represented as S_1={X_i,X_{i+1},...,X_{100}}. The gateway node forms the primary cluster C_1 by considering the first sensor reading X_1 as cluster head CH_1 for primary cluster, represented as CH_1←X_1. The gateway node collects the successive sensor reading (X_{i>1}) and compares with the predetermined cluster head CH_1 of primary cluster C_1. If the successive readings (X_{i>1}) distance measure Dist(X_i, X_{i+1})=|X_i - X_{i+1}| is lesser than the predefined application based custom threshold of ε then the successive reading under consideration would be put into the first cluster as represented in (1).

\[ C_1 \leftarrow \forall X_i \text{ if } Dist(X_i, X_{i+1}) < \varepsilon \]  \hspace{1cm} (1)

The first non near sensor reading in the series denoted as X_j with respect to the first cluster head CH_1 is considered as the second cluster head CH_2, henceCH_2←X_j | Dist(X_j, CH_1)>ε, such that CH_2==Null. Once the two cluster heads are determined, the following clustering function is applied to the successive reading to push the upcoming reading to the respective cluster:

\[ Cluster(X) = C_1 \leftarrow X_i \text{ if } Dist(X_i, CH_1) < \varepsilon \]  \hspace{1cm} (2)

Outlier tolerant adaptive sampling rate approach for wireless sensor node (Sunil Kumar Selvaraj)
The formation of a second cluster head and cluster is the clear indication of change in successive sensor readings. Nevertheless, the formation of a second cluster is not quite enough to conclude the detected change as a pattern of event data or outlier.

In the density based approach for evaluating the formed cluster, the density metric of the cluster is used for determining the change as event data or an outlier. Outliers generally does not sustain for longer duration, which keeps it different from the event data. The above concept is considered for setting the predefined custom threshold of Minpts (30) values in successive readings to make the reading as event data and lesser to that value as outlier reading. The equation: \( \forall X \in C_2 \rightarrow \text{Outlier}; \) if \(|C_2|<\text{Minpts}\), which represents cluster as outlier in accordance to the above condition and similarly the equation: \( \forall X \in C_2 \rightarrow \text{Pattern of Change}; \) if \(|C_2|>\text{Minpts}\) represents the formed cluster as event data.

Nullifying the impact of outlier as change in sampling interval adjustment means the non detection and removal of outlier in period of sensor readings which may lead to unnecessary decrease of sampling interval adjustment, as illustrated by considering the example of Outlier_Period\(_i\)\(_1\);\(_2\)\(_i\)\(_1\)\(_i\)\(_i\)=\(\{X_1,X_2,X_3,X_4,\ldots,X_{100}\}\), in simple Statistical based adaptive sampling approach, the adjustment of sampling interval based on the higher variation among the period of data is shown in (4),

\[
\text{Variance(Period}_i\text{)} > \text{Threshold} \downarrow \Delta\text{Sampling}\_\text{Interval} (4)
\]

Conversely in the proposed approach, once the second cluster is representing the outlier data, and then the second cluster is not considered as representing the actual change in successive sensor readings and correspondingly does not reduce the Sampling interval as show (5).

\[
|C_2| < \text{MinPts} \uparrow \Delta\text{Sampling}\_\text{Interval} (5)
\]

In adjusting the sampling interval based on the detection of cluster representing the variations in sensor readings, the second cluster is formed representing the change as event data, then the sampling interval would be reduced. Considering the example Period\(_i\);\(_1\)\(_2\)=\(\{X_i,X_i+1,\ldots,Y_i,Y_{i+1},\ldots,Y_{50},Y_{50},Y_{100}\}\), whereas \(X=\) continuous sensor readings in a period., \(Y_i=\) represent varied sensor readings in a period. If the change in sensor readings \(Y_i\) occurs more than a MinPts, then dense second cluster is formed. The formation of a second cluster indicates the variation and poses the need of reducing sampling interval at point in time of variation to capture the variation under progress. Consequently, formation of dense second cluster and reduction of sampling interval is shown (6).

\[
|C_2| \geq \text{MinPts} \rightarrow \downarrow \Delta\text{Sampling}\_\text{Interval} (6)
\]

### 3. EXPERIMENTAL SETUP AND STUDY

The proposed approach is tested on real wireless sensor network test bed comprised of 3 National Instruments WSN 3202 Programmable node, interfaced with vernier temperature sensor and one NI WSN 9791 gateway node. In this experimental study, one NI WSN 3202 node is programmed with proposed clustering based adaptive sampling approach, another NI WSN 3202 node is programmed with simple variance based adaptive sampling approach and remaining NI WSN 3202 node is programmed with fixed sampling interval rate to measure the impact of proposed approach. All the NI WSN 3202 nodes are interfaced with separate vernier temperature probe and each probe is inserted into pot soil for capturing the real time variation of soil temperature. All the sensor nodes are initialized to send the temperature data to NI WSN 9791 gateway node at fixed transmission interval rate of 5 sec per sample despite of varying sampling interval. The gateway node upon collecting data from first sensor node, which is programmed with clustering approach, applies the proposed in-network real time clustering approach to detect the change in real time series data. In this experiment, MinPts of 30 values are considered as the predefined threshold for considering the cluster as dense and detection of change. The value of 0.5 for distance is considered for computing near and non near points with respect to the cluster head. The gateway node determines sampling interval accordingly and send the new sampling interval to first sensor node. With respect to second node, gateway node collects the period of 100 sensor readings from second node and computes the variance to detect the change. Based on the change, the new sampling interval is determined and sends to the second node. In the variance based sampling interval approach of second node; gateway node tries to detect the change only after collecting of period of 100 sensor readings. Whereas with respect to the first sensor node; the gateway node would detect the change in real time by employing the proposed in-network clustering approach. The third
sensor node is programmed with fixed sampling interval and sends data to gateway at preset transmission rate. The collected temperature time series data from all the sensor node and varying sampling interval rate of each node are stored for further analysis of proposed approach. The outlier values are induced manually by turning off the power source to sensing device and the operation of the proposed approach of detecting and removal of generated outlier is observed. The proposed approach is shown in algorithm 1.

Algorithm 1: Outlier Tolerant in network clustering
Input: Period of Sensor Readings $X_i$, Period $j = X_i, X_{i+1}, X_{i+2}, \ldots, X_{100}$
Result: Sampling Interval (SI)

1. Initialize:
   - $j = 1$
   - $i = 1$
   - $\text{Max}_{\text{interval}} = 30$
   - $\text{Min}_{\text{interval}} = 5$

2. While ($j \geq 1$) do:
   - $\text{CH}_1 \leftarrow X_i \in \text{Period}_j$
   - $C_1 \leftarrow \text{CH}_1$
   - While ($i < 100$) do:
     - $\text{Distance}(D) = |\text{Period}_j(X_{i+1}) - \text{CH}_1|$
     - if ($\text{Distance}(D) < \varepsilon$) then
       - $C_1 \leftarrow C_1 \cup \text{Period}_j(X_{i+1})$
     - if ($\text{Distance}(D) > \varepsilon$) && if ($\text{CH}_2 == \text{NULL}$) then
       - $\text{CH}_2 \leftarrow \text{Period}_j(X_{i+1})$
       - $C_2 \leftarrow \text{CH}_2$
     - if ($\text{Distance}(D) > \varepsilon$) && if ($\text{CH}_2 != \text{NULL}$) then
       - $C_2 \leftarrow C_2 \cup \text{Period}_j(X_{i+1})$
     - if ($|C_2| > \text{MinPts}$) then
       - $\text{Change}_{\text{detected}} \leftarrow 1$
     - if ($\text{Change}_{\text{detected}} == 1$) then
       - $\text{SI} = \text{SI} / 2$
     - if ($\text{SI} < \text{Min}_{\text{interval}}$) then
       - $\text{SI} = \text{Min}_{\text{interval}}$
     - if ($i == 99$) && if ($\text{Change}_{\text{detected}} == 0$) then
       - $\text{SI} = \text{SI} + 2$
   - if ($\text{SI} > \text{Max}_{\text{interval}}$) then
     - $\text{SI} = \text{Max}_{\text{interval}}$
   - $i \leftarrow i + 1$
   - $j \leftarrow j + 1$

4. RESULTS AND DISCUSSION

In this section, we showcase the impact of proposed approach of outlier tolerant adaptive sampling interval. Experiments were conducted for collecting temperature data comprised of change in temperature and induced outlier for the study of proposed approach as shown in below Figure 2. The time series temperature data collected by sensor node comprised of outlier data denoted as value 300 and change of event as a temperature value of 55 are captured at various points as shown in Figure 2. Even though actual changed temperature value was around 30 degree to 32 degree with respect to normal value 30 degree Temperature, changed temperature in the range of 31 to 32 degree is represented as a value of 55 degree for ease interpretation.

![Figure 2. Real time temperature measurement along with outlier measurements](image-url)
In the proposed approach sensor readings are collected in period of 100 values and clusters are formed on period of sensor readings for detecting the outlier and change of event. From experimental study, the formation of second non dense cluster indicating the outlier as part of period of sensor readings is shown in below Figure 3. Similarly, Figure 4 shows the period of sensor readings with the formation of dense second cluster indicating the presence of change in event as variation in temperature data.

The formation of non dense second cluster representing the outlier, the proposed approach would not decrease the sampling interval by considering change as outlier. Whereas the formation of dense second cluster representing the change of event is detected and sampling interval is adjusted accordingly as shown in Figures 2 and 5. The study on impact of outlier on sampling interval adjustment is showcased in Figure 6 by considering the simple statistical variance based sampling interval approach. With respect the non detection of outlier in data in simple statistical variance based approach leads to unnecessary reduction in sampling interval as shown in Figures 2 and 6.

The sampling interval value during the run time of experimental study of both clustering approach versus simple statistical variance based approach is shown in Figure 7 respectively. As shown in Figure 7, the variance based approach adjust the sampling interval by reduction considering the outlier data as changed data, where as in clustering approach, outlier data are detected and sampling interval would not be adjusted by reduction on outlier data. The Figure 7 clearly shows the impact of outlier on sampling interval adjustment and need of outlier detection in real time. With respect to the problem formulated of higher energy dissipation due to lower value of sampling rate, the results obtained in Figure 7 shows the reduction in energy consumption in proposed clustering approach compared to the simple statistical based approach without outlier detection module.

![Figure 3](image3.png)

Figure 3. Period of sensor measurements containing outlier measurement represented as non-dense cluster

![Figure 4](image4.png)

Figure 4. Period of sensor measurements containing change in temperature measurements and formation of the dense second cluster to detect the measurable change in measurements
Outlier tolerant adaptive sampling rate approach for wireless sensor node (Sunil Kumar Selvaraj)

Figure 5. Sampling interval adjustment in clustering approach on sensor readings comprised of outlier and non outlier measurements

Figure 6. Sampling interval adjustment in variance based approach on sensor readings comprised of outlier and non outlier measurements

Figure 7. Rate of change in sampling interval values with respect to clustering approach and variance approach
5. CONCLUSION

The proposed work results clearly showcase the impact of outlier on adaptive sampling interval approach and need of real time in-network outlier detection for optimization of sampling interval. The results were satisfactory on the experimental test-bed towards real time detection of outlier and rapid optimization of sampling interval. We would like to carry out the above proposed approach on more real field deployed devices for further study on energy dissipation and data quality issues.

ACKNOWLEDGEMENTS

Authors like to sincerely acknowledge Vision Group on Science and Technology (VGST), Department of IT, BT, and S&T, Government of Karnataka, Bangalore, for sanctioning research grant for the Department of Computer Science and Engineering, Mangalore Institute of Technology and Engineering, Moodabidri, under KFIST- L2 grant schema vide GRD No. 339 dated.

REFERENCES

Outlier tolerant adaptive sampling rate approach for wireless sensor node (Sunil Kumar Selvaraj)

BIOGRAPHIES OF AUTHORS

Mr. Sunil Kumar Selvaraj currently working as a Senior Assistant Professor in the Department of Computer Science and Engineering, Mangalore Institute of Technology and Engineering. He received his B.E degree from KVGCE, Sullia and M.Tech Degree form MIT, Manipal, currently pursing his Ph.D. study on Wireless Sensor Network Domain at VTU, Belgaum. He can be contacted at email: sunil@mite.ac.in.

Dr. Venkatramna Bhat P. B.E, M.Tech, Ph.D is currently working as a Professor in the Department of Computer Science and Engineering, NMAM Institute of Technology, Nitte, India. He has received his Ph.D. on Wireless Sensor Network Domain from Manipal University, Manipal. He can be contacted at email: pv.bhat@nitte.edu.in.