PalArch's Journal of Archaeology of Egypt / Egyptology

# WIENER PREDICTION FOR ENVIRONMENTAL MONITORING IN WIRELESS SENSOR NETWORK OF CLUSTER STRUCTURE

*Arpana Mishra[1],Sandhya Bhardwaj[2],Rashmi Priyadarshini[3]*

[1, 2] Department of Electronics and Communication Engineering, IIMT, Greater Noida, UP

[3] Department of Electronics and Communication Engineering, Sharda University, Greater Noida, UP

E-mail: [1]shubhparna20@gmail.com

**Arpana Mishra,Sandhya Bhardwaj,Rashmi Priyadarshini: Wiener Prediction for Environmental Monitoring in Wireless Sensor Network of Cluster Structure -- Palarch's Journal Of Archaeology Of Egypt/Egyptology 17(9). ISSN 1567-214x**

**Keywords: mean-square derivation (MSD, Principal Component Analysis (PCA), Least Mean Square(LMS), Normalized Mean Square (NLMS),Recursive Least Square (RLS), Auto Regressive (AR).**

## ABSTRACT

Environmental monitoring is very important applications of wireless sensor net-works (WSNs).The lives of WSNs are of several months, or even years. However, the inherent restriction of energy carried within the battery of sensor nodes brings an extreme difficulty to obtain a satisfactory network lifetime, which becomes a bottleneck in scale of such applications in WSNs. Proposed novel frame work of data prediction, can apply in WSN to simultaneously achieve accuracy and efficiency of the data processing in clustered architectures. The main aim of the framework is to reduce the communication cost while guaranteeing the data processing and data prediction accuracy. In this framework, data prediction is achieved by implementing the dual wiener prediction algorithm with optimal step size by minimizing the mean-square derivation (MSD), in a way that the cluster heads(CHs) can obtain a good approximation of the real data from the sensor nodes. On this basis, a centralized Principal Component Analysis (PCA) technique is utilized to perform the compression and recovery for the predicted data on the CHs and the sink, separately in order to save the communication cost and to eliminate the spatial redundancy of these used data about environment [1]. All errors generated in these processes are finally evaluated theoretically, which come out to be controllable. Based on the theoretical analysis, designing is possible a number of algorithms for implementation.

Simulation results by using the real world data demonstrate that our frame work provides a cost-effective solution to such as environmental monitoring applications in cluster based WSNs..

## 1. Introduction

Internet of things (IOT) is an emerging heterogeneous networking concept aimed towards a significant impact in the today's digital world. The key vision of IoT is to bring together a massive number of smart objects towards integrated and interconnected heterogeneous networks, making the internet even more ubiquitous.

IoT framework is based on several enabling technologies including wireless sensor net-works (WSNs), cloud computing, machine learning, and peer to peer systems.

By using wireless communication and sensor technology, WSNs have advantages in applications over other casual networks, on the aspects of such as with standing ability, clustering for scalability, and self-organization properties [5–7].

Moreover, in the context of continuous monitoring, the most of data changes at a slow speed, which results in a large amount of data redundancy in space or time ,subsequently frequent communications between sensor nodes will be a waste of limited energy .Basically ,the increase of network lifetime will be proportional to the reduction in the number of transmitted data packets. Following this principle, data reduction has become one of the most enhanced solutions that is aimed to reduce the amount of data transmissions [15–18].

The most efficient way to obtain data reduction in WSN is data prediction that uses the prediction values instead of the real ones,there by avoiding the data transmission. In areal-world scenario, it is often unnecessary and yet costly to obtain the precise measurements for each sample period. Data prediction techniques focus on minimizing the number of transmitted measurements from the sensor nodes during continuous monitoring process. However, one key concern is to ensure the accuracy of the prediction with in a user-given error bound. For the periodical sensing applications especially environmental monitoring, each consecutive observation of a sensor node is temporally correlated to a certain degree. In our prediction model, the temporal co-relation is exploited to perform the prediction of data for the monitoring application based on the user-defined error tolerance. The result of using this correlation-based approach is a dual prediction protocol(Wiener filter protocol ) that has are mark able effect on reducing the frequency of data transmissions in a way that guarantees the prediction accuracy.

One alternative approach to realize at a reduction is using compressing techniques [19,20] that lead a reduction in the amount of transmitted data because the size of data is reduced. In general, we can classify the data compression schemes into two categories: lossless and loss compression. Lossless data compression demands the original data to be perfectly reconstructed from the compressed data. By contrast, lossy data compression allows some features of the original data that may be lost after the decompression operation. For highly resource constrained WSN ,lossless algorithms are usually not necessary despite the fact that they have better performance on data recover ability. To put it the other way, lossy compression is better able to reduce the amount of data to be sent over the WSN. In the case of lossy compression, the amount of compression and there construction error are the important criterions to judge the quality of compression algorithms. Our work using the Principal Component Analysis (PCA) method to compress the original data is proved to be able to obtain satisfactory results in two ways. More importantly, the error generated by the PCA compression is negligible compared to the prediction error, which ensures the user's acceptable total error bound.

In order to obtain the energy- efficient scheme for continuous environmental monitoring, we develop in the present paper a novel frame work with delicate combination of data prediction, compression, and recovery in cluster based WSNs. The main idea of the frame work is to reduce the communication cost through data prediction and compression techniques whilst the accuracy is guaranteed .First, sensor nodes collecting environmental parameters are grouped in to multiple clusters based on their physical locations. At the same time, a dual prediction mechanism using Wieiner prediction algorithm with optimal step size is implemented at sensor nodes and their respective CHs, which not only improves the prediction accuracy, but also achieves faster convergence speed during the initial stage of algorithm. Then the CHs extract the principal component of collected data by the PCA techniques after a sampling period, so redundant data can be prevented. Finally, data is successfully recovered at the Base Station (BS). Throughout the entire process, all errors are controllable and kept within the tolerable bound. After achieving data reduction, the size of recovered data at the BS is equal to that of raw sensory data collected by all nodes. It is advantageous for the BS to gain a more in-depth understanding of environment parameters .The simulation results also demonstrate that the combination of Wiener prediction algorithm and PCA technique is energy-efficient for environmental monitoring applications in cluster based WSNs.

## 2. Related work

Many models have been proposed to perform data prediction in WSN. The Auto Regressive (AR) model uses the linear regression function embedded in the sink to calculate the estimation of future sensor readings. By regularly collecting local measurements, the sensor node can compute the coefficients of the linear regression based on past real values. These coefficients are then delivered to the sink to perform time series forecasting. Within the context of AR model, the paper [21] proposed a general framework called Probabilistic Adaptable Query (PAQ) to efficiently answer queries at the sink based on a simple AR model. An adaptive model selection algorithm used in [22] allows sensor nodes to independently choose the one from a set of candidate models, which has the best performance in terms of the statistics property. The

Similarity-based Adaptive Frame work (SAF) [23] uses a simple linear time series model that consists of a time-varying function, called trend component, and a stationary auto regressive component representing the divergence of the phenomena on from the time-varying function over time. In this frame work, nodes learn these models locally (requiring no communication). When the local model is no longer a good fit for the data, the node relearns the model and transmits its coefficients to the sink. Although the proposed A R model based method shows that all errors are below the user specified threshold, the short coming of this approach is however that the communication cost is high when the error threshold is set at a small value.
Data prediction in WSN using adaptive filters has the merits of quick convergence and high prediction accuracy over the AR model. Among various filter algorithms discussed in[27] such as the Least Mean Square(LMS), the Normalized Mean Square (NLMS), and the Recursive Least Square (RLS) algorithm, the main difference among them is the weight updating method at each iteration. Santini and Romerin [28] presented an adaptive strategy that uses the standard LMS algorithm for quality-based data reduction in WSN. However, our studies presented in the current paper will show that the LMS prediction algorithm works well only for the star network, the performance degenerates for a hierarchical network unfortunately. In particular, we find that, for the fixed step size parameter the wiener method does not necessarily provide the optimal prediction performance in every iteration. In this paper, this problem is solved successfully by introducing the variable step size into the LMS algorithm.

## 3. Description of prediction algorithms

### 3.1. Model of network

Due to sensor node's nature of limited battery capacity, how to design energy-aware network architecture has been the important research issue in WSNs. The method by grouping sensor nodes in to clusters has been widely applied in WSN in order to achieve energy-efficient and long-lived objective. Cluster based structure is seen to have more advantages than other network model, including scalability, ease of data fusion and robustness. Fig.1 depicts a typical two-tier hierarchy in a WSN, in which Cluster Heads (CHs) form the higher tier while member nodes form the lower tier. The cluster composed of sensor nodes and CH is the core unit of WSN. CHs play multiple roles in the network model. CHs communicate directly with sensor nodes and forward sensor data to the base station, which means they bear the responsibility of data sensing as the common node, aggregation and processing as the intermediate node.

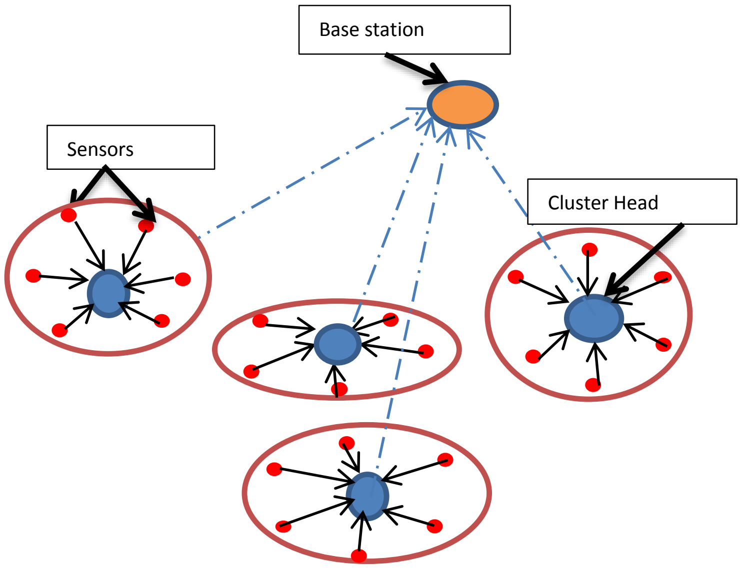


***Figure 1:*** *Network topology of Clustered WSN*

### 3.2. Dual prediction model in cluster-based WSN

There exists a basic assumption in the dual prediction (DP) model for cluster-based WSNs. In environmental monitoring applications of WSNs, failures of nodes and links can be caused by a variety of reasons such as battery drain, congestion, channel fading, and high bit error rate (BER). All those can lead to packet loss during data transmissions. Only with reliable data trans-mission, the sending node and receiving node can out put the same predictions. Some methods to solve such failure issues are presented in[43–45]. In the present paper in order to ensure data synchronization among nodes, we assume that the communications between sensor nodes is reliable and failure free. The prediction methods and models in a lossy environment in a WSN are left for future study. The main aim of the DP model is to reduce the energy consumption caused by radio transmission through minimizing the number of transmissions between the sender and the receiver. Let us consider a cluster in a WSN, which is formed by several member nodes and a cluster head. A sensor node continuously collects the reading x(n) from the surrounding environment at every sampling time n. In the case without any prediction, the node sends all readings to the CH, whereas, the node can selectively send some readings to the CH if an identical predictor is deployed both at the node and the CH. The prediction method works in the following way: First, the sensor node predicts the actual sensed reading x(n) based on the historical data recording and compares it with x(n). If the error e(n) between the predicted and actual data is over the pre-specified tolerance or threshold e_max, the actual data x(n) will be sent to the CH. It is only after receiving x(n) that the CH uses the received data as the sensed data. On the other hand, if the calculated result shows that error e(n) is below the pre-specified threshold e_max, the sensor node dose not send x(n) to the CH, but replace it with the predicted value as the latest historical data in the next prediction round. The CH will not receive any data from the node within a limited time and deduces that the error generated in the prediction does not exceed the threshold value. During this period, the CH

treats the estimated value as the sensed data as well as the sensor node, data communication does not occur between the CH and node. In this case, the prediction algorithm plays a vital role in the performance aspect of DP model.
In the performance aspect of DP model .Due to the robustness and the low computational over head, least mean square (LMS) algorithm is widely recognized in signal processing and wireless communication research field .And the standard LMS with some slight modifications can be used for prediction. Unlike other prediction algorithms, LMS does not require a perfect knowledge of the statistical properties of the previous samples, which subsequently reduces the additional computation to a minimum. This feature of the algorithm manifest itself an attraction as it is very important to reduce the additional computational over head in applications for WSN.

### 3.3. The LMS prediction algorithm

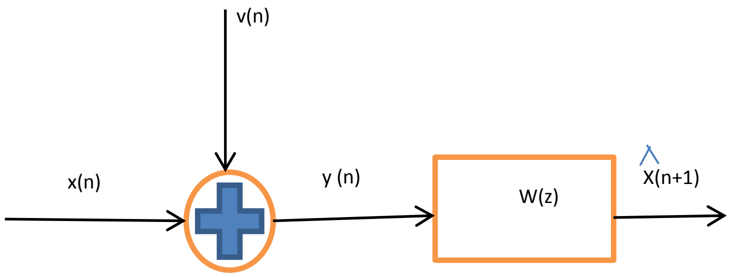


***Figure 2:*** *Prediction algorithm in Clustered WSN*

As shown in Figure 2, a sensor node generates a data stream

$X(n)=[x(n-1),x(n-2),\ldots,x(n-N)]\ T,$ (1)

Which is composed of the previous N readings from the instant n−1.
This forms the input signal to the filter.
The output y(n), which is the prediction value of the actual reading x(n) at the time n, can be obtained by

$y(n) = WT(n)X(n),$ (2)

where $W(n)=[w1(n),w2(n),\ldots,wN(n)]T$ is the weight coefficient of the filter.
The error e(n) between the output and the desired signal d(n) that the filter tries to adapt to is then computed by

$e(n)=d(n)-y(n).$ (3)

The weight coefficient at the next instant n+1 can be updated by

$W(n+1)=W(n)+\mu e(n)X(n),$ (4)

where μ is the step size parameter. The standard LMS prediction algorithm is described by the above four Equations from (1) to (4).

## 4. Conclusions and future work

In this paper, we propose a new frame work for processing environmental data in a clustered WSN, which utilizes data analysis technique, prediction. In our frame work, an optimal step size in the LMS algorithm is obtained to perform data prediction both at the node and at the cluster head, and then the cluster head applies the centralized PCA for data compression; the base station finally recovers the original data with the error tolerance. We have proposed an approach to obtain the desired optimal step size in LMS by overcoming the difficulty of then on-regular matrix with the aid of the generalized inverse of matrix, which is also very interesting in the field of prediction methods. We propose a couple of algorithms that can be implemented at sensor nodes, CHs and the BS. We evaluate the communication cost and analyse the means square error in the Wiener algorithm. .The simulation results by using the real world data demonstrate that our frame work is efficient and workable. First, our prediction algorithm provides better performance in terms of prediction accuracy, convergence speed and communication reduction than the traditional one which is solely based on LMS. Second, our data compression algorithm achieves the communication cost saving when sensor data is spatially correlated. We also investigate the mean square error during the data processing. The result further verifies the feasibility of such a frame work by guaranteeing the error limitation. Future work will mainly look at the scheme that is integrated closely with the clustering algorithms to better reduce the communication amount and improve the error accuracy of data prediction and recovery .Additionally, in this paper, we ignored the outliers a rise from sudden change in local environment and unreliable measurement. Therefore, we will explore an efficient outlier detection mechanism to improve robustness of our frame work for this specific case. How to use our frame work in a lossy wireless network with all sensor nodes and links suffering from non-synchronization, failure and data loss would be an interesting extension for further research as well.

## References

Mou Wua, Lian shengTana, Naixue Xiong, Data prediction, compression, and recovery in clustered wireless sensor networks for environmental monitoring applications, ELSEVIER ,Information Sciences 329 (2016) 800–818

I.F.Akyildiz,M.C.Vuran,Wireless Sensor Networks, 4, JohnWiley &Sons, 2010.

H.Alemdar,C.Ersoy,Wireless sensor networks for health care: a survey, Comput. Netw. 54 (15) (2010) 2688–2710

W.Dargie,C.Poellabauer, Fundamentals of Wireless Sensor Networks: Theory and Practice, JohnWiley &Sons, 2010.

J.Zheng, P.Wang, C.Li, Distributed data aggregation using Slepian–Wolf coding in cluster-based wireless sensor networks, IEEE Trans.Veh.Technol.59(5)(2010)2564–2574.

Z.He, B.S.Lee, X.S.Wang, Aggregation in sensor networks with a user-provided quality of service goal, Inf.Sci.178(9)(2008)2128–2149.

H.Jiang,S.Jin,C.Wang, Prediction or not? An energy-efficient frame work for clustering-based data collection in wireless sensor networks, IEEETrans. Parallel Distrib.Syst.22(6)(2011)1064–1071.

H.Li,K.Lin,K.Li,Energy-efficient and high-accuracy secure data aggregation in wireless sensor networks,Comput.Commun.34(4)(2011)591–597.

C.Caione, D.Brunelli, L.Benini, Distributed compressive sampling for life time optimization in dense wireless sensor networks,IEEE Trans.Ind.Inf.8(1)(2012)30–40.

Y.Quer,R.Masiero,M.Rossi,M.Zorzi,Sensing,compression and recovery for wireless sensor networks: monitoring frame work design,IEEE Trans.Wirel.Commun.11(2012)3447–3461.

D.Tulone, S.Madden, Paq: Time series forecasting for approximate query answering in sensor networks, in: Proceedings of the 3rd European Conference on Wireless Sensor Networks (EWSN),2006,pp.21–37.

Y.-A.LeBorgne, S.Santini, G.Bontempi, Adaptive model selection for time series prediction in wireless sensor networks, Signal Process. 87(12)(2007)3010–3020.

D.Tulone,S.Madden,An energy-efficient querying framework in sensor networks for detecting node similarities, in:Proceedings of the 9th ACM Interna-tional Symposium on Modeling Analysis and Simulation of Wireless and Mobile Systems,2006,pp.191–300.

S.S.Haykin, Adaptive filter theory,Pearson Education India,2005.

S.Santini, K.Romer, An adaptive strategy for quality-based data reduction in wireless sensor networks, in: Proceedings of the 3rd International Conference on Networked Sensing Systems (INSS), 2006, pp.29–36.

H.Zhou, Y.Wu, Y.Hu, G.Xie, A novel stable selection and reliable transmission protocol for clustered heterogeneous wireless sensor networks ,Comput.Commun.33(15)(2010)1843–1849.

V.Rajendran, K.Obraczka, J.J.Garcia-Luna-Aceves, Energy-efficient, collision-free medium access control for wireless sensor networks, Wirel. Netw. 12(1)(2006)63–78.

B.Deb, S.Bhatnagar, B.Nath, Information assurance in sensor networks, in: Proceedings of the 2nd ACM International Conference on Wireless Sensor Networks and Applications, 2003, pp.160–168.