Fault Tolerant ARIMA-based Aggregation of Data in Sensor Networks

Douglas Herbert, Gaspar Modelo Howard, Carlos Perez-Toro, Saurabh Bagchi
School of Electrical & Computer Engineering, Purdue University
{herbertd, gmodeloh, perezcr, sbagchi} @purdue.edu

1. Introduction

Sensor networks collect data using robust, energy efficient, low bandwidth protocols due to the few resources available to them. Among the many methods explored to improve the efficiency of sensor networks for data collection, researchers have proposed statistical methods to predict when each sensor should send data. Two shortcomings to this approach are: (1) the use of a static model, assuming unchanging environmental parameters to the sensor network and (2) the lack of an error detection mechanism to compensate for failures during the transmission of data.

Suppressing data transmissions by using the fact that data follows a statistical model has been proposed before [2], with the assumption that the environmental parameters do not change. This assumption does not fit several scenarios where sensor networks can be deployed. Additionally, recent advancements in sensor networks allow for more powerful (CPU and memory) on sensing nodes. This represents an opportunity to extend the role of end nodes, by allowing them to perform statistical analysis on the collected data at runtime and updating model parameters in real time. In this manner, we can create a self adapting model.

When creating such a model, we must remember that a failure-free environment is non-realistic for sensor networks. Errors come from many sources, such as the wireless medium. It is necessary to propose a system architecture in which to deploy our statistical model that will provide redundancy and compensate for failures in the network.

In this paper we address the dynamic nature of time series sensed data by using the ARIMA forecasting model and re-computing its parameters when environment values change. We also introduce error detection capabilities to allow the sensor network to transmit sensed values using an energy-efficient mechanism.

The contributions of the paper can be summarized as follows: (1) Enhancing the energy savings in a data aggregation sensor network running the LEACH protocol by using ARIMA forecasting even under changing environmental conditions, and (2) Adding robustness to ARIMA forecasting by employing redundant fitting of data among different nodes.

2. Clustering ARIMA Forecasting

Our design approach uses the clustering mechanism of LEACH to obtain accurate parameters for the ARIMA model. The main assumption is that the data sensed by different nodes within the same cluster will follow a similar ARIMA model. This allows the cluster head to verify the correctness of the ARIMA models by using redundant information gathered between the cluster nodes.

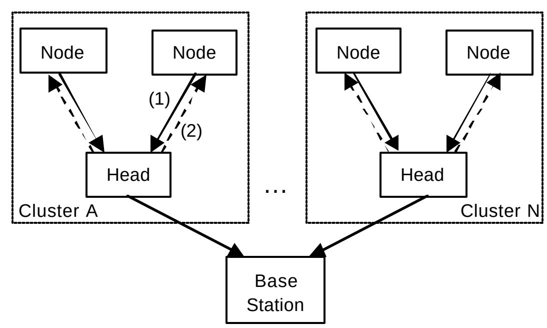

Figure 1 – Schematic of the LEACH-based sensor network showing the communication path between nodes to transmit (1) sensed data or ARIMA model parameters and (2) control data to request recalculation of parameters.

The ARIMA model was implemented into the Low-Energy Adaptive Clustering Hierarchy (LEACH) [2], a protocol architecture that provides energy-efficient routing and data aggregation by using hierarchical network architecture. Nodes organize themselves into clusters in a distributed fashion, and each cluster has a cluster head that will coordinate communication between the rest of the nodes in the cluster and a base station. The cluster head aggregates the data from the cluster nodes and sends this summary to the base station in a single high energy transmission.

The Auto-Regressive Integrated Moving Average (ARIMA) [1] non-seasonal model was chosen to provide energy efficiency to the data collection scheme in sensor networks by reducing the amount of data that must be transmitted. The ARIMA models are a class of models for forecasting a time series. The model is fully described by the numerical values of each of the terms in the following forecasting equation plus an initial set of time series data.

$$X_t^d = f_1 \times X_{t-1}^d + \ldots + f_p \times X_{t-p}^d + X_t^d + q_1 \times Y_{t-1}^d + \ldots + q_q \times Y_{t-q}^d$$

The p value is the number of auto-regressive terms, the q value is the number of moving average terms, and the d value is the number of non-seasonal differences. Fitting the ARIMA model consists of finding the appropriate values of the parameters, which will then be transmitted in place of an entire data set.

We extend the LEACH protocol by adding a verification step at the cluster head. After each node transmits the ARIMA parameters, the cluster head verifies the accuracy of the model by generating a time series with each set of parameters and calculating the mean square error between them:

$$\hat{X}_{MMS} = \frac{1}{N}\sqrt{\sum_{i=0}^{N}\left[x_i - \hat{x}\right]^2}$$

If the mean square error is above a fixed tolerance value, the cluster head can request all cluster nodes to recalculate their respective parameters repeatedly until all models are within tolerance. If this does not occur within a fixed number of retries, the cluster head labels this as a fault, and the cluster head orders the cluster nodes to transmit all sensed data instead of model parameters. In addition to the cluster head verifying model parameters, sensing nodes periodically calculate the mean square error between the actual observations and the response predicted by the ARIMA model. If the error exceeds a fixed threshold, the cluster node will calculate new parameters for the ARIMA model. This ensures the scheme can adapt to changing environmental conditions. Choosing the thresholds for recalculation is a tradeoff between energy consumption and the accuracy of the data aggregation.

3. Experiments and Results

This system was implemented in TinyOS and simulated using TOSSIM. Simulation experiments consisted of a network of nodes running an implementation of LEACH with the added ARIMA forecasting features. The nodes sense data at a constant rate and fit it to an ARIMA model in a round-robin fashion.

Simulations were run using temperature data gathered from real sensor motes, as well as computer generated data using an ARIMA (1,1,1) model. Figure 2 shows a plot of the number of packets transmitted from cluster nodes to a cluster head as time progresses for the case of plain LEACH data aggregation versus LEACH with ARIMA forecasting. The plot clearly illustrates a significant reduction in the amount of data transmitted using ARIMA forecasting. Plain LEACH mode also corresponds to the case when the cluster data cannot be fit into a single ARIMA model.

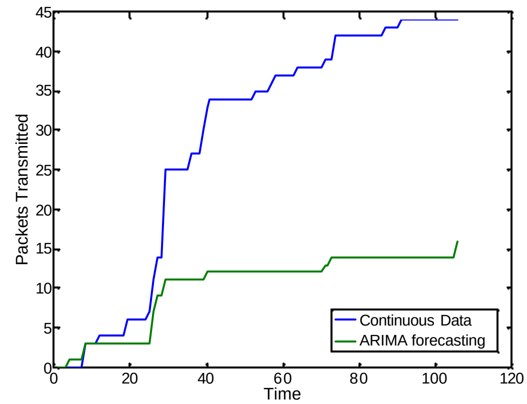

Figure 2 - Packet Transmitted for continuous transmission mode and ARIMA forecasting mode

4. Conclusions and Future Work

In this paper we proposed an efficient and fault-tolerant approach to using statistical prediction models to save energy in sensor networks. The approach was based on a coupling of the LEACH protocol with ARIMA forecasting techniques. Our approach allows environmental parameters to change over time by re-calculating them according to the mean square error technique, and provides robustness against data that does not fit an ARIMA model. It improves the accuracy of ARIMA forecasting and avoids repeated transmissions by checking the parameters before forwarding to the base station. Results show a marked reduction in packets transmitted between sensing nodes and cluster head in our approach.

Future work will be focused on finding better ways to fit sensor data to an ARIMA model taking advantage of feedback provided by the cluster head about how the parameters compare to those of other sensing nodes. An implementation on actual hardware will also be done in order to further verify our results.

5. References

[1] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, “An Application-Specific Protocol Architecture for Wireless Microsensor Networks”, *IEEE Transactions on Wireless Communications*, 1(4), 660-670.

[2] C. Liu, K. Wu, and M. Tsao, “Energy Efficient Information Collection with the ARIMA Model in Wireless Sensor Networks”, In Proceedings from IEEE Globecom’05: Global Communication Conference, 2005.