The Spanish Inquisition Protocol—Model Based Transmission Reduction for Wireless Sensor Networks

Daniel Goldsmith and James Brusey
Cogent Computing Applied Research Centre
Coventry University, Coventry, UK
Email: (goldsmid,j.brusey)@coventry.ac.uk

***Abstract*— The Spanish Inquisition Protocol (SIP) reduces Wireless Sensor Network (WSN) energy cost by transmitting only unexpected information and is so-named because "nobody expects the Spanish Inquisition!" SIP extends prior Dual Prediction Scheme (DPS) algorithms that model phenomena at both node and sink. SIP's key advancement is that it transmits a state vector estimate rather than individual readings. SIP can be tuned according to the desired estimate accuracy, with lower desired accuracy typically leading to fewer transmitted packets. In simulation with real data, less than 5% of the samples needed to be transmitted to provide the sink with an accurate estimate of the sensor value (within 0.5 °C, in the case of temperature). SIP also significantly outperforms prior DPS results when using the same data sets. In deployment on Telos motes, SIP shows similar performance to the simulations.**

I. Introduction

Wireless Sensor Networks (WSNs) are deployed to monitor, observe, and potentially improve understanding of environments and phenomena. Domain scientists and industrial technicians need robust and reliable raw data to allow them to find patterns and confirm hypotheses about how the monitored phenomena evolves over time. When monitoring is being used to ensure that the condition of a system remains within certain bounds, then the timeliness of the information is also important so that corrective action can be taken quickly. The critical factor in the design of WSN systems is often the energy cost of communicating the data [1]. Although other components, such as sensing and processing, play a part in the energy budget, these are typically much less than the cost of communication. For example, a comparison of power requirements for a range of WSN motes and components by Polastre [2] shows that, for the commonly used Telos platform, the power required to operate the radio is approximately ten times greater than that required to operate the CPU. Therefore, reducing the energy cost associated with communication will substantially reduce the overall energy usage, lengthen the time that a system can be left unattended (without battery changes, for example), and thus enable many WSN applications that would otherwise be infeasible.

Periodic sensing, the most common WSN functional mode, tends to produce much *data* but little *information*. This suggests that data encoding or compressing at the source might help to reduce the energy cost of transmission. However, compression on a packet-by-packet basis is not applicable since, by and large, packets tend to be quite small and thus compressing individual packets would yield only a small (if any) saving. Furthermore, aggregation of several packets into one may help but at the cost of reducing the timeliness of data. Again the energy saving will be minor. Finally, reducing the sensing frequency may stretch the energy budget at the cost of potentially missing important phenomenological events.

The approach described in this paper is based around encoding the data using a simple, approximate model of the phenomena. This model is shared by both node (transmitter) and sink (receiver) while the parameters for the model (or state vector) are only transmitted from node to sink when needed. Specifically, the node assumes that the sink can apply the model and predict the current state. By keeping track of what the sink knows, the node can identify when the error in the sink's prediction will exceed some predefined threshold ε, triggering an update message to be sent to the sink.

The paper is organized as follows. The following section reviews some prior approaches in this area. Section III describes the Spanish Inquisition Protocol (SIP) in detail. Experimental results for SIP are given in Section IV followed by conclusions and future work (Sections V and VI).

TABLE I

SUMMARY OF PRIOR WORK

Approach	Dataset	Error threshold	% Data Transmitted
PMC [3]	Sea Surface Temperature [4]	1% Range of Data (ε=0.06)	50 %
PMC [3]	Salinity [4]	1% Range of Data (ε=0.0187)	45 %
PMC [3]	Shortwave Radiation [4]	1% Range of Data (ε=13.513)	45 %
DPS [5]	Intel Node 13 [6]	ε=0.5 °C	10%
DKF [7]	Electric Power Load [8]	ε=150 MW	45%
DKF [7]	Network Monitoring Dataset [9]	ε=5 pkts	5%

II. RELATED WORK

Several methods have been proposed for reducing the amount of data transmitted in a WSN at the cost of losing some measurement accuracy. One such method is event-based transmission where messages are transmitted only when a predefined event is detected by a node. This approach is well suited to applications where events are both sparse and easily detectable (such as sniper localization [10] and intrusion detection [11]). The difficulty with event-triggered delivery is in specifying event thresholds. Since a system's state tends to evolve over time (e.g., due to seasonal variations), predefined triggers may lose relevance to real events in the underlying phenomena, thus making the approach less useful.

Olston *et al.* [12] performed a study into reducing the amount of data transmitted in stream-based monitoring systems, examining the trade off between precision and performance when using approximations of the original data stream. The method that they propose is an event-based transmission scheme that dynamically adjusts the event trigger according to desired accuracy requirements and changes in the measurand over time. SQL-like queries are registered with a central stream processor located at the sink, along with the maximum error permissible for that query. Based on the set of registered queries, the stream processor adjusts the allowable error associated with each sensor, with the aim of optimising the communication for the entire network while maintaining bounded accuracy. The node only sends an update to the sink when the sensor value leaves the error bounds, which are based on the allowable error and the last transmitted value.

Lazaridis and Mehrotra [3] define Poor Man's Compression (PMC), which is a bounded transmission suppression scheme and which uses a form of Run Length Encoding (RLE) to reduce the amount of data transmitted. PMC divides the time-series into series of segments such that the range of values within a segment does not exceed some threshold. Instead of transmitting the whole segment, the count of values and midpoint are transmitted. While this approach offers an attractive level of compression, the reported values suffer from quantisation, so much of the short term detail within the data is lost. Lazaridis and Mehrotra also define a method for reconstruction of sensor readings from the segment summaries. This allows factors such as network and compression latency to be addressed since the current reading at a given time can be estimated, rather than waiting for an update from the sensor.

Jain *et al.* [13], [7] introduce the Dual Kalman Filter (DKF) approach as a general solution to reducing the amount of data transmitted. The node uses a separate Kalman Filter (KF) per sensor to perform prediction of sensor readings. If the value predicted by the KF differs from the actual sensor reading by more than the user specified error threshold, the information required to update the KF is transmitted to the sink. The sink maintains a KF for each sensor and is able to use a similar process to replicate the data predicted by the node. This method requires detailed prior knowledge of the system under consideration in order to estimate the KF parameters correctly.

Santini and Römer [5] propose a transmission-based approach that uses the Least Mean Squares Adaptive Filter (LMS) [14] for filtering and prediction of sensor values. As with DKF, the LMS filter is used at both node and sink. In comparison with DKF, LMS is model-free. Nevertheless, LMS is sensitive to the step-size parameter μ and a poor initial choice can lead to the filter either not converging or becoming unstable. To address this Santini and Römer give a scheme to calculate μ during an initialisation stage. The difficulty with this approach is that the filter typically takes some time to converge. Thus during the initial stages, all samples must be transmitted. Normalised Least Mean Squares Adaptive Filters (NLMSs) are a variant of LMS that avoid the problem of instability depending on μ by normalising depending on the input.

An overall summary of past results of comparable work is given in Table I, where the error threshold is denoted ε.

III. The Spanish Inquisition Protocol

The SIP uses a simple, approximate model of the phenomena that is shared between node and sink. A state vector forms the parameter for this model, allowing a forward prediction of the state of the phenomena to be made. Rather than report the last received sensor reading, the sink predicts the state of the phenomena based on the last received state vector. Using knowledge of the last state vector transmitted to the sink, the node can identify when the error in such a prediction will exceed some threshold ε. Different applications will have different requirements for the error threshold. For example, household temperature monitoring might only need $\varepsilon = 0.5$ °C.

The SIP approach is model agnostic. A simple model that works well in many cases, however, is piece-wise linear approximation of the measurand. In this case, the state vector $\mathbf{x}_t = (x_t, \Delta x_t)^T$ consists of a measurand *estimate* x_t and an *estimate* of the rate of change Δx_t.

The overall approach assumes that a guaranteed delivery scheme is used at lower protocol layers.

SIP does not require that clocks are synchronized between node and sink. However, a local clock at the node is generally required by the predictive model.

Figure 1 shows the algorithm for node and sink. The node, which is typically a remote, battery powered mote, begins each sensing cycle by querying its sensor. Next, a new state estimate is obtained by combining the sensor reading with the last state estimate $\mathbf{x}_{\text{old}}$ and its associated time t_{old}. Typically a simple filter is used to update the state estimate. Filters will be discussed further below. A prediction is then made of the state that would be estimated by the sink based on the last transmitted state $\mathbf{x}_{\text{sink}}$ and the time that it corresponds to t_{sink}. If the difference exceeds some threshold ε, then the new state estimate is transmitted to the sink. Note that the threshold is, in the generic case, a vector but treated as a scalar here as typically only one component of the state vector needs to be tested.

When the sink receives a new state estimate $\mathbf{x}$, it only needs to store it in a database of past state vectors $\mathbf{x}_{\text{sink}}(t)$ along with its associated timestamp t. When a high-level application queries the sink, if the current state is being requested, then the sink needs to predict the state estimate based on the last received vector. In the case where the sink is queried for past readings, it must interpolate from temporally neighbouring state estimates.

A. State Estimation and Filtering

As described previously, a basic predictive model that can be used with SIP is piece-wise linear. This requires an estimate of the gradient (or rate of change) as well as an estimate of the point in time value of the measurand. For some sensing problems it is not possible to make even this assumption and instead one must assume that the state is piece-wise constant in time.

To improve the estimate of the gradient, some form of filtering is generally needed. Selection of a filter depends partly on the model used and partly on the requirements of the application. Basic filtering can be performed using an Exponentially Weighted Moving Average (EWMA) filter. This recursive filter returns an estimate of the measurand that combines the current reading with past readings. Apart from removing some of the noise in the signal, the filter also smooths over quantisation introduced by Analogue Digital Converter (ADC). LMS or NLMS is a more sophisticated filter that has been tried by some authors, as previously discussed. KF or Extended Kalman Filter (EKF) are more sophisticated approaches that use linear or non-linear, respectively, models of the state of the environment. However, these are more computationally costly and may be more difficult to tune without prior knowledge of the phenomena.

IV. Results

Key metrics for evaluating the performance of SIP algorithm are % data transmitted and Root Mean Squared Error (RMSE) in the reconstructed signal. In principle, reducing the % data transmitted will produce a corresponding reduction in energy use. Many authors ignore the question of the RMSE of the reconstructed signal. Nonetheless, RMSE is important as it is an indicator of the quality of the reconstructed signal at the sink and thus the lossiness of the protocol.

A summary of results for SIP is given in Table II. The table shows results for 6 data sets, the last being results from a deployed rather than simulated use of the protocol. These results consistently improve over past results shown in Table I. In particular, the performance for Intel temperature data shows an improvement by a factor of 10. Also, the performance for the Network Monitoring Dataset again demonstrates improvement over prior work. The SIP algorithm is scale invariant and so if the scale of the range is increased along with the error threshold, performance will be maintained. The rate of fluctuation of the data also plays a part in the performance of the algorithm but whether this has a dramatic effect depends on the filter and its parameters.

As shown in Figure 2(a), the amount of data transmitted decreases exponentially as the allowable error ε increases. The data set used here is the HomeREACT [15] temperature data but these results are typical. Figure 2(b)

Node:

$s \leftarrow$ query sensor()
$\mathbf{x}' \leftarrow$ estimate new state $(s, \mathbf{x}_{\text{old}}, t_{\text{old}})$
$\mathbf{x}_{\text{sink}} \leftarrow$ predict sink state $(\mathbf{x}_{\text{sink}}, t_{\text{sink}})$

if $|\mathbf{x}' - \mathbf{x}_{\text{sink}}| > \varepsilon$:
  transmit $(\mathbf{x}')$
  $\mathbf{x}_{\text{sink}} \leftarrow \mathbf{x}'$
  $t_{\text{sink}} \leftarrow t$
$\mathbf{x}_{\text{old}} \leftarrow \mathbf{x}'$
$t_{\text{old}} \leftarrow t$

Sink:

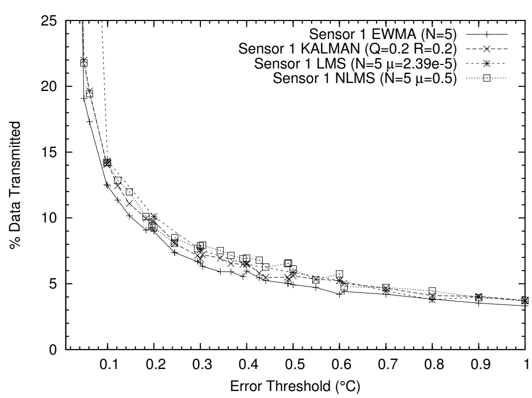
[On receipt of new state estimate$(\mathbf{x})$]
  $\mathbf{x}_{\text{sink}}(t) \leftarrow \mathbf{x}; t_{\text{last}} \leftarrow t$

[Estimate value for time(t)]
  if $t \geq t_{\text{last}}$
    predict from $\mathbf{x}_{\text{sink}}(t_{\text{last}})$
  else
    interpolate from neighbouring $\mathbf{x}_{\text{sink}}$

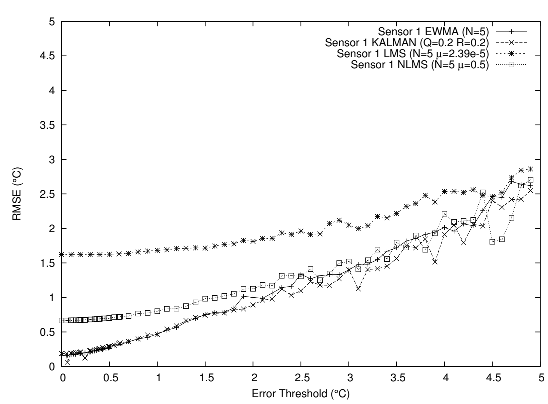
Fig. 1. Pseudocode for node and sink for the Spanish Inquisition Protocol

TABLE II
SUMMARY OF SIP PERFORMANCE FOR VARIOUS DATA SETS

Data-set	Error Threshold (ε)	Filter	RMSE	Transmitted (%)
HomeREACT Temperature [15]	0.5 °C	EWMA	0.24 °C	4.1
(sensor 1)	0.5 °C	NLMS	0.75 °C	4.0
	0.5 °C	KF	0.25 °C	3.9
HomeREACT Humidity [15]	0.5 %RH	EWMA	0.46 %RH	13.3
	0.5 %RH	NLMS	2.22 %RH	12.7
	0.5 %RH	KF	0.58 %RH	11.3
HomeREACT Light [15]	5 lux	Pass through	2.2 lux	4.4
	9 lux	Pass through	2.5 lux	2.4
	5 lux	EWMA	2.7 lux	1.4
	9 lux	EWMA	5.8 lux	0.37
Intel (Node 13) [6]	0.5 °C	EWMA	0.24 °C	1.0
	0.5 °C	NLMS	0.41 °C	1.1
	0.5 °C	KF	0.26 °C	1.4
	0.05 °C	EWMA	0.06 °C	5.3
Network Monitoring Dataset [9]	5 pkts	EWMA	2.33 pkts	0.64
Telos Deployment	0.5 °C	EWMA	0.22 °C	1.7



(a) Percentage of data transmitted



(b) RMSE of predicted data

Fig. 2. SIP performance on temperature sensor 1 (HomeREACT data set)

shows that there is a linear relationship between the RMSE of the predicted data and the error threshold. While the choice of filter has a minor effect on the amount of data transmitted, there is a more significant effect on the RMSE of the predicted data stream. In particular, LMS and NLMS produce a non-zero RMSE when the error threshold is set to zero. This is due to their initialisation phase. In general, the RMSE is roughly half of the error threshold for the other filters. Overall, EWMA usually outperforms the more specialised filters, in terms of both the amount of data transmitted and the RMSE of the reconstructed data stream.

V. Conclusions

SIP provides a mechanism for reducing the amount of data transmitted by a WSN using a shared model between node and sink and by only transmitting data from node to sink when the node identifies that the error at the sink would exceed some threshold.

SIP was evaluated over several data sets, featuring different sensing modalities. Evaluation also took place on mote hardware, with comparable reductions in the amount of data transmitted to that experienced during simulation.

SIP has been shown to outperform similar algorithms, allowing greater reductions in the amount of data transmitted while maintaining a similar level of accuracy. In the case of the Intel Lab data set, the amount of data requiring transmission was 10 times less than when using a comparable algorithm.

Filtering of the raw sensor values is used within the state estimation process to improve the estimation of the rate of change. While filter selection is application dependent, experimentation has shown that in many cases, a simple EWMA filter provides comparable performance (in terms of both transmission reduction and RMSE) to more complex filter types. As the computational cost of the EWMA is minimal, it would be a good choice when processing and memory overheads need to be minimised.

VI. Future Work

This paper focused on the application of SIP over individual sensor readings. However, it is possible to extend the algorithm to support multiple sensors (and sensor types) per node by increasing the number of dimensions encoded in the state-vector. A preliminary investigation has shown encouraging results, and further work will include a through analysis of this mode of operation.

It may be possible improve the accuracy of interpolation when reconstructing past data, by using a more sophisticated approach than piece-wise linear approximation. Future work will investigate using a spline-based method. As well as allowing data to be reconstructed, SIP offers the opportunity to reduce the amount of persistent storage required or each time series, as only the state vector rather than each predicted value needs to be stored.

Further investigation into how data characteristics affect the amount of data transmitted will aim to better understand the relationship between those characteristics and the data reduction provided by SIP.

Finally, it is planned to investigate the effective energy savings for SIP when deployed on mote hardware with a variety of sensor modalities.

References

[1] G. J. Pottie and W. J. Kaiser, "Wireless integrated network sensors," *Commun. ACM*, vol. 43, no. 5, pp. 51–58, 2000.

[2] J. Polastre, R. Szewczyk, and D. Culler, "Telos: enabling ultra-low power wireless research," in *Proceedings of the 4th international symposium on Information processing in sensor networks*, ser. IPSN '05. Piscataway, NJ, USA: IEEE Press, 2005. [Online]. Available: http://portal.acm.org/citation.cfm?id=1147685.1147744

[3] I. Lazaridis and S. Mehrotra, "Capturing sensor-generated time series with quality guarantees," in *ICDE*, U. Dayal, K. Ramamritham, and T. M. Vijayaraman, Eds. IEEE Computer Society, 2003.

[4] M. McPhaden. (2010, 7) Tropical atmosphere ocean project, pacific marine environmental laboratory. [Online]. Available: http://www.pmel.noaa.gov/tao/

[5] S. Santini and K. Römer, "An adaptive strategy for quality-based data reduction in wireless sensor networks," in *Proceedings of the 3rd International InProceedings on Networked Sensing Systems (INSS 2006)*. Citeseer, 2006, pp. 29–36.

[6] S. Madden. (2010, 7) Intel lab data. [Online]. Available: http://db.lcs.mit.edu/labdata/labdata.html

[7] A. Jain, E. Y. Chang, and Y.-F. Wang, "Adaptive stream resource management using kalman filters," in *SIGMOD Conference*, G. Weikum, A. C. König, and S. Deßloch, Eds. ACM, 2004, pp. 11–22.

[8] Anon. (2010, 7) Basic generation services data room. [Online]. Available: http://www.bgs-auction.com/bgs-dataroom.asp

[9] ——. (2000, 3) The internet traffic archive. [Online]. Available: http://www.ita.ee.lbl.gov/

[10] G. Simon, M. Maróti, A. Lédeczi, G. Balogh, B. Kusy, A. Nádas, G. Pap, J. Sallai, and K. Frampton, "Sensor network-based countersniper system," in *Proceedings of the 2nd international conference on Embedded networked sensor systems*, ser. SenSys '04. ACM, 2004, pp. 1–12. [Online]. Available: http://doi.acm.org/10.1145/1031495.1031497

[11] A. Arora, P. Dutta, S. Bapat, V. Kulathumani, H. Zhang, V. Naik, V. Mittal, H. Cao, M. Demirbas, M. Gouda, Y. Choi, T. Herman, S. Kulkarni, U. Arumugam, M. Nesterenko, A. Vora, and M. Miyashita, "A line in the sand: A wireless sensor network for target detection, classification, and tracking," *Computer Networks*, vol. 46, pp. 605–634, 2004.

[12] C. Olston, J. Jiang, and J. Widom, "Adaptive filters for continuous queries over distributed data streams," in *SIGMOD Conference*, A. Y. Halevy, Z. G. Ives, and A. Doan, Eds. ACM, 2003, pp. 563–574.

[13] A. Jain and E. Y. Chang, "Adaptive sampling for sensor networks," in *DMSN*, ser. ACM International Conference Proceeding Series, A. Labrinidis and S. Madden, Eds., vol. 72. ACM, 2004, pp. 10–16.
[14] S. Haykin, *Adaptive filter theory*. Prentice Hall, 2002.
[15] T. Daniel, E. I. Gaura, and J. Brusey, "Wireless sensor networks to enable the passive house - deployment experiences," in *EuroSSC*, ser. Lecture Notes in Computer Science, P. M. Barnaghi, K. Moessner, M. Presser, and S. Meissner, Eds., vol. 5741. Springer, 2009, pp. 177–192.