NETWORK-AWARE LOSSLESS SOURCE CODING OF SPATIO-TEMPORALLY CORRELATED SENSOR DATA

Paul G. Flikkema   Brent W. West   Bill Ruggeri

Department of Electrical Engineering
Northern Arizona University
Flagstaff, Arizona 86011 USA

ABSTRACT

Wireless sensor networks for environmental monitoring have unusual characteristics and must meet a unique set of requirements. Data rates are extremely low, but limits on energy usage and total deployment cost dominate. Probably the most important metric is the average energy required to deliver a bit of information about the sensed environment to the destination. These networks also differ from the general case of ad-hoc wireless nets in that connectivity between two arbitrary nodes is not required, nor even desired; all that is needed is connectivity from each sensor to a gateway or "home" node. This paper describes a methodology for source coding of environmental sensor network data that exploits correlation in temporal and spatial domains. Our approach is an adaptive two-level scheme. The first level is model-based in that the time series from a sensor is approximated by a polynomial function. The metric for this approximation is the minimum number of bits required to represent the approximation error for the worst-case data point. The second level exploits spatial correlation by allowing nodes along paths from more distant nodes to adopt model parameters from farther-flung nodes, called code servers, to further reduce redundancy.

1. INTRODUCTION

The widespread deployment of large collections of wirelessly networked intelligent devices may be the next technology boom. One example of these systems that has received increasing attention is wireless sensor networks. To be successful, the design of wireless sensor networks must carefully orchestrate a number of technologies previously viewed as distinct, including wireless communication, ad hoc networking, and deeply embedded real-time computing systems. Driving interest in this area are a number of compelling problems: distributed surveillance and tracking, monitoring of building and transportation systems, hazard warning, and environmental sensing.

The particular application that we are targeting is the minimally-invasive monitoring of microclimate variables, such as temperature, humidity and insolation (light intensity). We are constructing both an architecture and a proof-of-concept implementation in collaboration with a plant physiologist and a community ecologist. Our overall goal is to provide an order-of-magnitude increase in spatio-temporal sampling performance over traditional non-networked sensing technologies, enabling significant improvements in models that will ultimately forecast ecosystem dynamics. Our primary objectives are (i) to demonstrate the utility and efficiency of wireless sensor networks for scientific data acquisition in support of microclimate-related ecological and biological research, and (ii) to develop algorithms and techniques that can enable the widespread application of wireless sensor networking to this, as well as other, problems.

2. WISARDNET ARCHITECTURE

In this study, the benchmark sensor technology that we are using is standalone datalogger units that have a low initial cost but require wired *in situ* data uploads that are invasive and labor-intensive. (Another existing solution is wired arrays of sensors which have even greater environmental impact and scale poorly.) Our wireless sensor network is composed of low-cost integrated sensing and communication nodes called WISARDs (Wireless Sensing and Relay Devices) [1]. They share several features with the Berkeley motes [2, 3], including the use of an energy-saving 8-bit programmable microcontroller, and communication in the 902-928 MHz ISM band. Our units consist of a two-board stack of a processor/analog module ("brains" board) and a radio module.

Our WISARD brains board is built around a Microchip PIC16LF877 microcontroller clocked at 10 MHz, to which we have interfaced external memory. Our boards can be populated with a different complement of chips depending on the application. For the sensor configuration, we in-

This work was supported in part by the National Science Foundation under Grant EIA-0131691. The work of B. West was supported in part by the Merriam Powell Center for Environmental Research.

stall 128Kb of SRAM with a multiplexed address/data bus to conserve I/O pins. For the gateway (or network controller) configuration, we install 16Mb of FLASH memory to support buffering and interfaces to long-haul links. The PIC16LF877 processor also has a self-writable program store that enables reconfigurability without an external auxiliary processor as used in [2].

We have integrated both dedicated and general-purpose analog and digital interface hardware on our brains board for a variety of sensors. For the proof-of-concept network, we support two temperature and two photodiode-based insolation (light intensity) sensors. Our sensors have been designed in close collaboration with our scientist partners: for example, we have designed thermocouples for temperature monitoring that include low-cost analog circuitry for real-time cold-junction calibration. Unlike thermistor-based sensors as used in [3], thermocouples have a minimum of mass and resultant thermal inertia. This technology is well-suited for the micro-met application, since it will be able to track transients that may be a driver for behavior of a variety of species. In addition, our sensors are deployed outside the WISARD package, eliminating temporal artifacts (such as limitations on transient response) that could result from internal mounting. The brains board also supports other sensors via a general-purpose analog input, a switched power source, and the 1-Wire bus.

The WISARD radio board uses a single-chip RF transceiver (Texas Instruments TRF9600A) with an FSK modulator, and an RF amplifier with user-controlled output power. It also has an integrated synthesizer providing 230 Hz frequency resolution that enables slow frequency hopping. Because the radio chip provides only downconversion and quadrature FM pre-detection (essentially a frequency detector), we perform real-time sampling, message detection, frame/bit timing acquisition, and noncoherent detection using software resident in the microcontroller.

A major focus of our effort has been the study of energy-conserving MAC protocols for this application. Our approach [4] is based on cross-layer interaction as part of the design. We exploit routing information to enable the establishment of one-hop cliques of nodes who agree on pseudo-random meeting times. This enables extremely high energy efficiency: a node awakens for communication (and hence consumes energy) during a given slot only if it is time for the clique to meet. We also exploit the capabilities of our PHY layer, allowing cliques to use pseudo-randomized frequency hopping patterns, thus mitigating multiple-access interference as well as frequency selectivity and external interference.

In general, the energy cost of a single instruction of computation is orders of magnitude lower than the cost of transmitting and receiving a bit of information: in our hardware implementation, these figures are (on average) 2.6 nJ and as much as 170 μJ, a communication vs. computation energy cost ratio of over 6×10^4. Hence, relative to communication, computational costs are almost negligible, meaning that the minimization of total energy cost is essentially tantamount to minimizing communication cost only. For example, we are using a joint bit detection/synchronization algorithm that performs within 1 dB of maximum-likelihood detection.

At the outset of this project, we argued that an integrated design approach at the circuit, system, and network levels of abstraction was required [1]. In that spirit, our work at the PHY layer builds on the extremely mature discipline of communication theory to minimize the energy required to successfully transfer a bit of information. At the MAC layer, our design ensures that nodes' radios are activated only when necessary with very high probability. We now turn to the task of minimizing the number of bits required to reconstruct at a destination node the information acquired by a networked array of sensors.

3. RELATED WORK

Research in distributed source coding has re-awakened in the last few years, motivated by two factors: the wireless sensor network application and the celebrated Slepian-Wolf existence theorem [5] for powerful distributed source coding. The Slepian-Wolf result shows that it is (asymptotically) possible to encode two sources X and Y with joint entropy $H(X,Y)$ in such a way that Y needs to send $H(Y)$ bits and X needs to send $H(X|Y)$ bits, even if X does not know Y for the purposes of coding.

The work of Ramchandran and colleagues [6] tackles the problem via a framework based on channel coding. The essential idea is the design of group codes with a coset structure that takes advantage of a limited amount of prior knowledge about the correlation between the data streams. Servetto [7] considers lattice codes in the context of a spatial "division-of-labor" model in which a network of routing nodes connects an array of sensor nodes to an array of destination nodes. In this model, collaboration between sensor nodes of any kind, and collaboration between routing nodes with respect to coding, is forbidden.

As in [7], we consider sampling a spatio-temporally dynamical system in this paper. However, all of our nodes can sense, transmit, and receive, and the goal is to forward the sensed information about the system to a common gateway node for uploading (most likely over satellite or cellular terrestrial infrastructure) to the ultimate user, e.g., an internet-enabled database . The pooling of information at intermediate nodes forbidden in [7] is exploited in our scheme: a node that both senses and routes explicitly determines the similarity between its data and data from nodes whose data it forwards. The use of this knowledge comes at very little cost: as shown earlier, the energy cost ratio of communication vs. computation is enormous, and the additional delay may be small relative to the temporal dynamics of the sensed data. Indeed, while we have no proof, we conjecture that exploiting known "side" information can yield benefits in terms of the complexity of the spatio-temporal coding algorithm at each node.

A critical aspect of source coding for sensor networks is the trade-off between delay and the effectiveness of coding. In our target application, we can expect on the order of 10's of readings within an interval of 1-3 hours. Our investigations have shown that, whether the algorithm is adaptive or not, the ratio of alphabet size (i.e., the number distinct quantized sensor values), to the file size is large enough to render naive approaches (e.g., Huffman or dictionary-based codes) ineffective. In this application, short sequences are common, and asymptotically optimum approaches can falter as a result.

In any sensor net where the sensed information must be forwarded to a much smaller number of gateways, it is well-known that nodes closer to the gateways are more vulnerable to energy depletion due the the higher information rates they must handle [8, 9]. For example, in a linear array of sensors with one gateway and without source coding, the amount of traffic each node must handle increases *linearly* (in the number of nodes) with proximity to the gateway. The objective of this work is to define a pragmatic strategy for network-aware source coding that can significantly reduce this, taking advantage of the spatio-temporal correlation of the sensed data to improve energy efficiency and network lifetime.

4. THE ALGORITHM

In our application, transmissions from the gateway to the internet will typically occur at a lower rate than the sampling at each sensor node. A typical scenario would involve sampling the environmental variables every 5 minutes and uploading to the internet every hour or two. However, this can vary: in our test data, we consider soil temperature time series 5 cm below the surface, which normally have a bandwidth of no more than 5×10^{-4} Hz.

Our algorithm is model-based in that we explicitly recognize the temporal correlation at a particular location. While this correlation could be modeled using a high-order Markov chain, our model is not based on a prior probability distribution, but rather the dynamics of the process. It is adaptive at the second level, since nodes test the similarity of their data with a certain set of neighboring nodes' data, and find the description that minimizes the number of additional bits required to represent their data. This per-realization approach is essentially sample-starved: attempts to estimate a (probabilistic) distribution are circumvented in favor of directly minimizing the representation using a (hopefully good) model [10].

Each node begins by fitting a low-order polynomial to its data. As we will see, the data's relatively low dynamics imply that a zeroth- or first-order polynomial is often sufficient. The idea here is to capture the trend of the data at that location: it can be viewed as a form of higher-order delta modulation, in which differences are encoded. However, here we are looking at a larger window of data, enabling a more compact representation.

In this paper, we limit ourselves to the lossless case. It can be strongly argued that the measurements alone are noisy and thus amenable to further, limited loss using networked compression. However, our scientist collaborators would take little comfort in knowing that the wireless connectivity that so eases the logistics of data collection also can result in a loss of accuracy.

Our criterion for choosing the polynomial coefficients is not the usual sum of squared errors criterion. In view of the short time series to be transmitted, we elect to encode differences between the data points and the polynomial model in fixed-length fields, since the alternative involves significant overhead. Hence our objective in choosing the polynomial coefficients is to minimize the maximum number of bits required to represent differences between the data points and the polynomial. This is currently being implemented using an exhaustive search initialized using the least-squares coefficients. The final temporally encoded package of data consists of the polynomial coefficients, the field width (in bits) of the difference data, and the difference data.

We posit that the differences of the data points from the polynomial representation arise from effects that are varying in their locality. Specific to each node are errors due to intrinsic sensor noise and micro-met variations at scales finer than the spatial sampling rate of the sensor grid. In general, variations occur across scales, so that some variations are global, causing network-wide common differences due to the low-order fit. In between are dynamics that are local to physical neighborhoods of sensor nodes. For this reason, each node's coefficients represent a local view of the global scene. This view's validity normally increases with proximity to its location; on the other hand, two nodes may be in similar local environments (e.g., shady), but separated by multiple hops. Hence correlation may not be a function of distance (either physical or number of hops), and this should be accommodated in a spatio-temporal source coding algorithm.

With this in mind, we turn to the spatial portion of our coding scheme. In our scheme, terminal nodes—those that need only sense and transmit their data, and don't need to route data from other nodes—do not execute this phase. However, nodes that route in addition to sensing *should* use their computational power to reduce the number of bits required to forward a package that includes its data as well as that of farther-flung (from the gateway) nodes that route through it to the gateway. We invert first-glance intuition and refer to these nodes as its *code servers*.

Each router node first uses the same temporal coding algorithm as used by terminal nodes, with the additional step of computing the number of bits required to represent its data. Together, these two steps comprise its view of the scene independent of other nodes and a measure of the communication cost of that view. The node then finds the number of bits required to represent its data as differences from the data of each of its code servers. (This requires decoding of its code servers' data, which we discuss shortly.) It then simply chooses the representation that minimizes the number of additional bits required to represent its data along with the data of its code servers. Normally, the node will often select one of the servers as a reference; on the other hand, if there are significant spatial variations, it can fall back to its independent view. In this way, the algorithm adapts to spatial correlation or decorrelation on a per-realization basis.

Implementation of the algorithm for a router node can be described by an example. Suppose the node under consideration is a router node n, $1 < n < N$, in a linear array of N sensors where node 1 is the gateway. Then, its code servers consist of all nodes $m > n$ farther away from the gateway. The decoding of its code servers' data is recursive: node $(n + 1)$'s data may be encoded as differences from node $n + 2$, etc. However, the recursive decoding is straightforward, and the complete coding algorithm can be accomplished well within the bounds dictated by the communication/computation energy cost ratio.

We mentioned earlier how per-realization similarity between time series may not be a function of either the physical distance or the number of hops. Within the limits of network routing, the algorithm allows a node to reference any of its farther-flung (i.e., outbound from the gateway nodes), thus allowing nodes to exploit pockets of similarity in data from code servers that may be physically dispersed.

5. EXPERIMENTAL RESULTS

To evaluate our algorithm, we used archived data acquired using conventional dataloggers by the NAU Piñon Ecology Research Group (http://pinyon.bio.nau.edu) at Sunset Crater National Monument approximately 20 miles northeast of Flagstaff, AZ. The ecosystem is piñon pine (pinus edulis) and juniper (juniperus monosperma) woodland, at approximately 6000 ft elevation. Some of the piñon pines have been attacked by scale and moth insects. The scale can cause significant defoliation, and the moth changes the tree structure; both of these effects may cause changes in the soil temperature. This, in turn, may effect soil moisture content and nutrient cycling processes.

As described earlier, no coding implies that the number of bits to be transmitted and received increases at least linearly with proximity (in number of hops) to the gateway. If we assume for the moment that the time series from all sensors are identical, then the terminal node is the only code server, and a performance upper bound for any spatio-temporal coding scheme would be a constant number of bits equal to the entropy of the terminal node (this equality is extremely crude given the non-probabilistic viewpoint driven by the short time series length). In practice, this implies a slight linear increase with number of hops due to the necessary overhead required for each node to point to the terminal node's information.

From the Sunset Crater data we selected soil temperature time series from sensors scattered approximately 15 m apart and about 5 cm below the surface; the raw time series showing the diurnal variation over a three day period are plotted in Figure 1. We chose a relatively linear portion of the series over an approximately 12 hour period for further study as seen in Figure 2. Compression results for the Sunset Crater data are shown in Figure 3, where we have plotted the cumulative number of bits that are required to be transmitted at each sensor as a function of the number of hops. In this experiment, we used first-order models. Because of the distances between the sensors and resulting dissimilarity of the time series, the gains are modest. In fact, in this dataset each sensor chooses its own polynomial model because it provides better compression than using any other's model. Hence Figure 3 provides an indication of the temporal coding gain that can be achieved.

In order to estimate the potential spatial-temporal coding gain, we created a synthetic data set from the Sunset Crater data by defining virtual nodes whose data was interpolated between adjacent nodes in the actual data. Then we modeled sensor-specific noise and micro-meteorological variations by adding a small Gaussian-distributed random variate to each interpolated value. This data is shown in Figure 4. In this case, a spatial coding gain occurs since nodes choose code servers other than themselves. As seen in Figure 5, the combined spatio-temporal coding gain is approximately 33%, as compared to 20% for data in Figures 2 and 3.

6. CONCLUSION

This paper considers the source coding of spatio-temporally correlated environmental sensor data for transmission to a sensor network gateway. In this application, timeliness constraints often imply the reporting of short time series, limiting the effectiveness of probabilistic and dictionary-based approaches. On the other hand, the energy efficiency of computation relative to that of wireless communication ar-

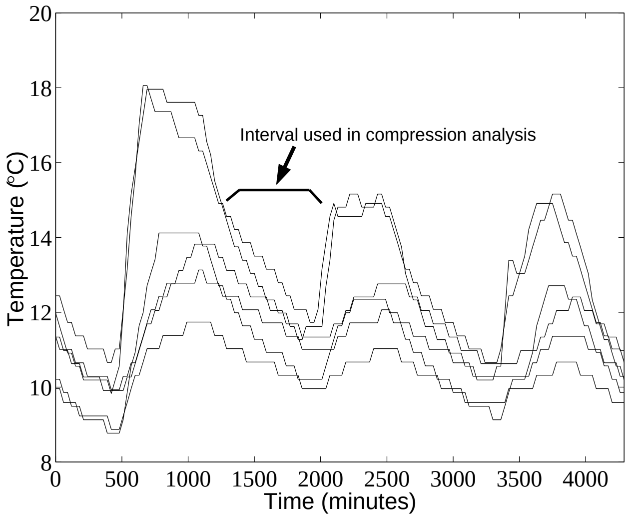

Figure 1. Raw sensor data; soil temperature as a function of time for six locations over a three day period.

gues strongly for approaches that can provide some compression. This paper proposes a lossless coding technique that exploits the joint sensing/routing duties of sensor nodes. It is model-based in the temporal domain, using a simple polynomial approximation. It is adaptive in the spatial domain in that a node can consider any of its code servers (including itself) and choose the one that yields the best compression for its time series. Our results indicate that the algorithm yields reasonable results that improve as the data similarity increases.

7. ACKNOWLEDGEMENT

We thank Mario Montes-Helu for making available spatio-temporal temperature data for use in this study.

8. REFERENCES

[1] B. West, P. Flikkema, T. Sisk, and G. Koch, “Wireless sensor networks for dense spatio-temporal monitoring of the environment: A case for integrated circuit, system, and network design,” in *Proc. 2001 IEEE CAS Workshop on Wireless Communications and Networking*, Aug. 2001.

[2] J. Hill *et al.*, “System architecture directions for network sensors,” in *Proc. 9th International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS-IX)*, pp. 93–104, Nov. 2000.

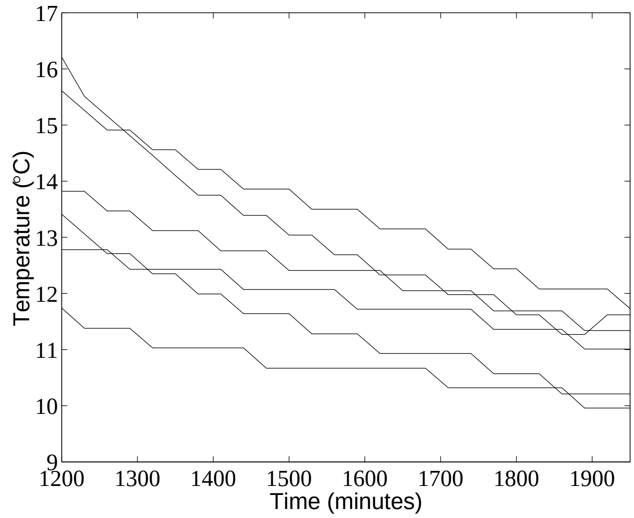

Figure 2. Raw sensor data; soil temperature as a function of time for six locations.

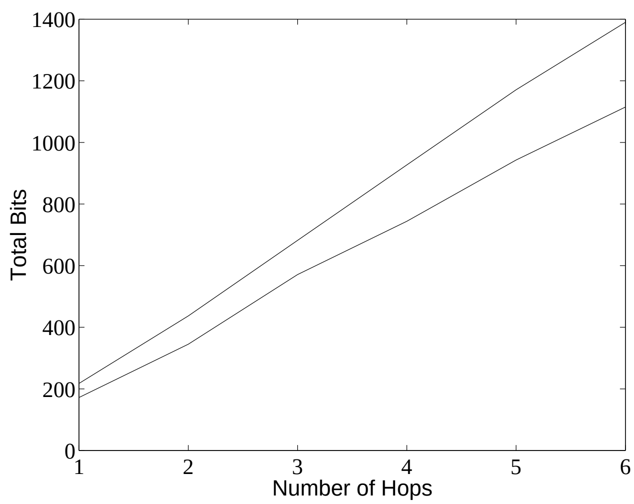

Figure 3. Compression performance for spatio-temporal coding of the raw sensor data; the upper line is for the uncoded data.

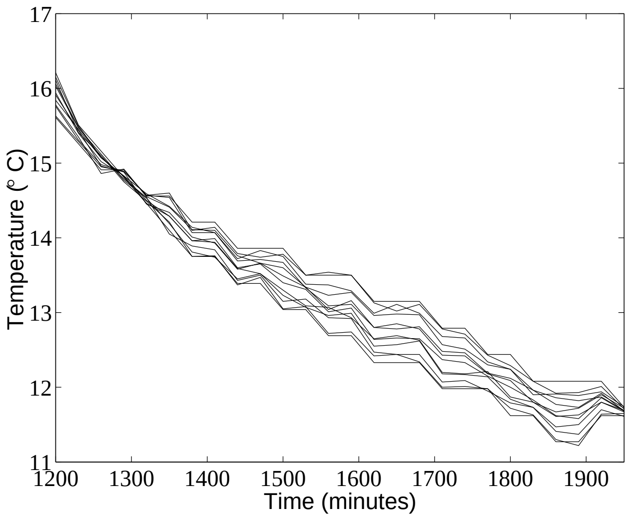

Figure 4. Synthesized sensor data; soil temperature as a function of time for twelve locations.

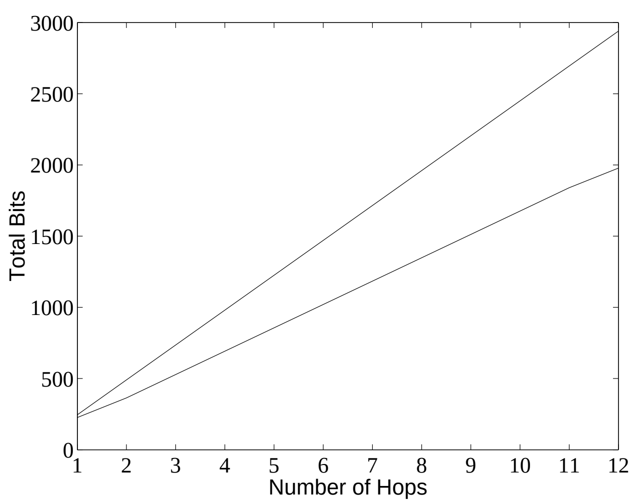

Figure 5. Compression performance for spatio-temporal coding of the synthetic sensor data; the upper line is for the uncoded data.

[3] A. Mainwaring *et al.*, "Wireless sensor networks for habitat monitoring," in *First ACM International Workshop on Wireless Sensor Networks and Applications*, Sept. 2002. (To appear.).

[4] P. Flikkema and B. West, "Clique-based randomized multiple access for energy-efficient wireless ad hoc networks." Submitted for publication.

[5] D. Slepian and J. K. Wolf, "Noiseless coidng of correlated information sources," *IEEE Transactions on Information Theory*, vol. IT-20, pp. 471–480, July 1973.

[6] S. S. Pradhan and K. Ramchandran, "Distributed source coding: Symmetric rates and applications to sensor networks," in *Proc. Data Compression Conference*, pp. 363–372, 2000.

[7] S. D. Servetto, "Quantization with side information: lattice codes, asymptotics, and applications in wireless networks." http://people.ece.cornell.edu/servetto/publications/papers/20020308/, 2002.

[8] B. Chen, K. Jamieson, H. Balakrishnan, and R. Morris, "SPAN: an energy-efficient coordination algorithm for topology maintenance in ad hoc wireless networks," in *Proc. 7th ACM International Conference on Mobile Computing and Networking*, pp. 85–96, July 2001.

[9] C.-K. Toh, "Maximum battery life routing to support ubiquitous mobile compuing in wireless ad-hoc networks," *IEEE Communications Magazine*, pp. 138–147, June 2001.

[10] V. S. Cherkassky and F. M. Mulier, *Learning from Data: Concepts, Theory, and Methods*. Wiley-Interscience, 1998.