

The Importance of Data Compression for Energy Efficiency in Sensor Networks

Mo Chen and Mark L. Fowler

Department of Electrical and Computer Engineering
State University of New York at Binghamton

Binghamton, NY 13902

{mchen0,mfowler}@binghamton.edu

Abstract –First, we demonstrate that not only does data compression help reduce the communication latency in a sensor network (as one would expect) but that it can also be used to gain energy efficiency since data compression reduces the energy spent transmitting data – which is one of the most energy-hungry tasks in a sensor network. Through simulations we show that data compression can bring more energy efficiency to a network than does recently proposed combinations of routing and data aggregation. Second, we propose a new viewpoint for data compression in sensor networks that measures distortion based on the compression scheme’s impact on statistical inference tasks (e.g., detection, estimation, recognition, etc.) and makes trades in a Rate-Energy-Accuracy design space. The usefulness of this idea is demonstrated through a particular sensor network application: object location in sensor-gathered images.

I. INTRODUCTION

Advances in sensor technology have focused interest towards using networks of sensors to collect useful information from an environment [1]. Tasks given to sensor nodes include collecting signal data, sharing the data between themselves, making inferences (estimations and decisions) from the data, and communicating the collected data and/or the inference results to one or more information sinks. In general, sensor networks must be designed to satisfy constraints on metrics assessing energy efficiency, communication latency, and accuracy of the conveyed information and there is a fundamental tradeoff over these metrics [2]; there are also constraints on fault tolerance and scalability [2] but we don’t address those here.

There can be many different scenarios for sensor networks; in this paper we focus on the so-called “reach-back” issue: communicating the data collected within the network back to a single information sink (e.g., base station, central command, etc.) with minimal latency and energy use. Energy efficiency in reach-back has been previously addressed by many researchers including [3], where energy efficiency was meas-

ured in terms of network lifespan. A related study has been carried out in [4] to show the need for compression to address the latency issue. The usefulness of data compression for energy efficiency is less clear. Thus, our first contribution here is to show that data compression has potential for helping to achieve energy efficiency in sensor networks and is the tool through which the energy-latency-accuracy tradeoff is made. A second contribution here is to present a new viewpoint for data compression in sensor networks performing statistical inference tasks.

Transmission of data is one of the most energy-expensive tasks a node undertakes – using data compression to reduce the number of bits sent reduces energy expended for transmission. However, compression requires computation, which also expends energy. Fortunately, trading computation for transmission can save energy since a recent paper [1] asserts that typically on the order of 3000 instructions can be executed for the energy cost required to transmit one bit over a distance of 100 m by radio.

Before discussing our results, we put our study into perspective with recent related results in [3] and [4]. The results in [3] address energy efficiency for reach-back through use of a combination of routing and data fusion/aggregation (called “LEACH”). By using data fusion/aggregation to combine two or more collected data sets that become co-located during transmission of the data through the network towards the sink, LEACH significantly reduces the overall data needed to be transferred and increases network lifespan. However, it is not clear in these papers how data fusion/aggregation can be relied on in *general* in a sensor network. In particular, in [3] the data from sensors grouped into clusters get fused through processing, where a stated assumption is that this fusion specifically implements beamforming; therefore, data aggregation is valid in this application setting. However, it is not clear that fusion/aggregation is possible in a general sensor network setting. Thus, the studies in [3] spurred our interest in developing a general framework using data compression rather than fusion/aggregation that would give similar gains in energy efficiency.

The results in [4] don't consider reach-back but rather the task of conveying the network's total collected information to each and every sensor node. The results in [4] establish fundamental information theoretic limits on the rate of information transfer through the network and show that data compression is needed to transfer the data without latency under a channel rate constraint. But for us, the key idea established in [4] is the effectiveness of combining classical source codes with routing algorithms and that this is competitive with distributed compression methods such as in [5], which remove common information between two nodes without sharing any data between them.

Thus, our study was initially motivated by the lack of *general* methods in [3] for removing common information between nodes and our interest was later bolstered by the information theoretic promise put forth in [4] for the effectiveness of classical, non-distributed compression combined with routing. Our goals in this paper are: (i) to show the energy-saving potential of data compression and (ii) propose a new viewpoint for data compression in sensor networks that measures distortion based on the compression scheme's impact on statistical inference tasks.

II. COMPRESSION AND ENERGY EFFICIENCY

For ease of comparing results, we use the same radio model used in [3] the radio dissipates 50 nJ/bit in the transmitter circuitry, 50 nJ/bit in the receiver circuitry, and 100 pJ/bit/m² for the transmitter amplifier. Because we aren't using a *specific* compression algorithm in this part of our study it is hard to specify how much energy is spent compressing the data, so we use the same energy cost that is used for data fusion/aggregation via beamforming in LEACH, namely 5 nJ/bit/message. To compare our methods with LEACH we ran tests using the following scenario: 100 sensors, randomly placed uniformly inside a 50m × 50 m square of real estate.

LEACH randomly selects 5% of its nodes as cluster heads; data from all the nodes in a cluster are beamformed together (data aggregation). This gives LEACH an inherent "compression ratio" of 20:1 since at each cluster head, 20 signals get beamformed into one. One of the key published conclusions for LEACH is that sending *directly* to the sink is inferior to LEACH; however, this really is an unfair comparison since the direct transmission method did not use any form of compression in [3]. Therefore we performed a simple simulation to show that using general compression without any routing provides better network lifetime – thus, it is LEACH's beamforming-achieved compression, not the routing protocol, that achieves the energy efficiency. Granted, the routing does have the advantage that it uniformly spreads sensor deaths over the network.

A. Direct Transmission With Non-Distributed Compression

We simulated LEACH as well as direct transmission with compression, with the later using a compression of 6:1 and 10:1. As in [3], each "round" of network transmission consisted of each node receiving 2000 bits of data and the network transferring the data to the sink. In LEACH, cluster heads are randomly selected on each round and the remaining nodes are assigned to clusters. Each cluster head receives 2000 bits from each of its cluster nodes and beamforms them into a single 2000 bit signal, which is then transmitted from the cluster head to the sink. Alternatively, direct-with-compression compresses the 2000 bits received at each node and then transmits the resulting bits to the source. The lifetime of the network is assessed by noting the number of nodes still alive at each round, where a live node is taken to be a node that has energy remaining. As mentioned above, comparable amounts of computational energy are assumed to be spent for beamforming and compression.

Our study shows that if a compression ratio of 6:1 is achievable at each sensor prior to sending the data directly to the sink, then the time it takes for 50% of the nodes to die is comparable to LEACH, as seen in Figure 1. Obviously, higher compression ratios further improve the direct w/compression curve, as is shown in Figure 1 for the case of 10:1, where the direct-with-compression now clearly outperforms LEACH.

This result is important because previous results [3] indicated that direct transmission was to be avoided and that special routing schemes were the answer to the energy efficiency problem. In our results we see that it is not the routing in LEACH that makes the difference, it is the beamforming-achieved compression. In understanding our result it is important to keep in mind that the direct method has no energy cost for reception since sensor nodes don't receive any transmissions; that compensates for the excess compression ratio that LEACH is assumed to have here (20:1 vs. 6:1 or 10:1). However, it should also be mentioned that [3] points out that a problem with the direct method is that the death of nodes begins with the nodes farthest from the sink and sweeps through the network towards the sink – this is generally undesirable and one nice feature of LEACH is that node deaths are uniformly distributed. Clearly direct-with-compression isn't directly applicable but it does point out the importance of data compression.

B. Direct Transmission With Compression Exploiting Spatial Redundancy

The direct-transmission method discussed above does not exploit any redundancy between signals re-

ceived at closely located sensor nodes. As a *simple* demonstration of the effectiveness of exploiting spatial redundancy we ran simulations to characterize the effect of exploiting this spatial redundancy. The results are shown in Figure 2 where we have simulated the effect of signals within a radius of 10m of a randomly selected set of primary nodes (making up 10% of the total number of nodes) as having virtually the same information content. This is more like LEACH but with beamforming-based aggregation replaced by general data compression. The spatially similar data is compressed and then the remaining data sets are compressed using non-spatial methods having a compression ratio of 6:1, after which all compressed data is sent to the information sink using direct transmission. The results in Figure 2 show the potential of exploiting this spatial redundancy through routing and local compression (rather than distributed compression). By randomly rotating which sensors are used as the central-compression sites the “sweep-of-death” for direct transmission is eliminated as it is in LEACH. The reason that this scheme far outperforms LEACH even though its CR is only 6:1 compared to LEACH’s 20:1 is that we have also reaped the benefit of compression as each sensor sends its data to the cluster head.

It is important to realize that we are not proposing that either of these two direct-transmission approaches should be considered as viable real-world methods; rather we are simply using them to illustrate that data compression plays an important role not only in controlling latency in sensor networks (as in [4]) but plays an important role in addressing network energy efficiency.

C. Rate-Energy-Accuracy function for Sensor Network Compression

Classical data compression theory relies on trade-offs between rate (R) and distortion (D) in terms of a R-D function. Rate is usually measured in terms of bits/sample and distortion is often measured as a mean-square error between the original and reconstructed signal. In the classical view, rate impacts latency and distortion impacts the accuracy of the signal reconstruction.

As we explored above, in sensor networks the rate can also impact energy efficiency. Thus, for sensor networks we propose the use of a 3-D extended version of the R-D function: the Rate-Energy-Accuracy (R-E-A) function. The Energy axis assesses the amount of energy needed to move the collected data to the desired destination. Clearly, decreasing the rate decreases the amount of transmission energy spent but a decreased rate comes at the expense of computational energy. A simple characterization of this is

$$\Delta E(R) = E_C(R) + E_T(R), \quad (1)$$

where $\Delta E(R)$ is the change in energy due to compression to a rate of R , $E_C(R)$ is the computational energy used to compress to the rate R , and $E_T(R)$ is the energy needed to transmit at the rate R . As R decreases, $E_C(R)$ increases (more compression requires more computation) and $E_T(R)$ decreases (more compression requires less transmission). Clearly, these measures depend on the computational efficiency of the compression algorithm, the energy efficiency of the computational architecture, and the energy efficiency of the transmission hardware.

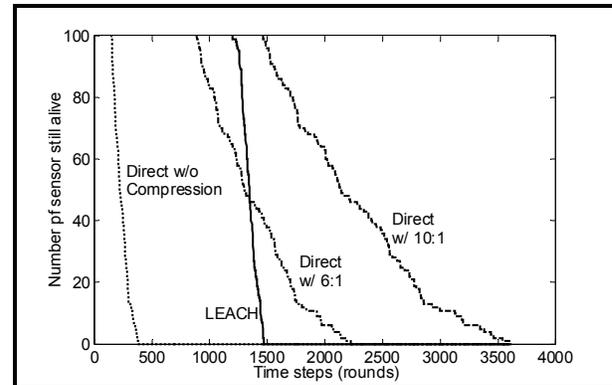


Figure 1: Non-distributed compression improves network lifespan when using direct transmission.

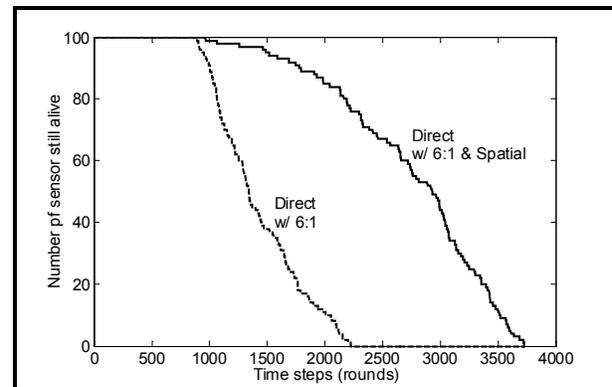


Figure 2: Results showing improvement using compression with a simple routing scheme.

Accuracy is related to distortion but is intended to better capture the effect of the compression error on the final use of the data – namely, the making of statistical inferences. Thus, if the inference task is estimation, then the accuracy measure should capture the impact of the compression on the estimation accuracy (see [6] for an example).

Thus we can specify a desired operating point in R-E-A space and develop compression algorithms (as well as low-power computing & transmitting architectures) that achieve it. In classical R-D function theory, there are two main dual R-D goals: (i) minimize distort-

tion while obeying a rate constraint, and (ii) minimize rate while obeying a distortion constraint. In the R-E-A viewpoint the goal of sensor network compression can be specified in many ways; for example: (i) minimize energy subject to constraints on rate and accuracy, (ii) maximize accuracy subject to constraints on energy and rate, (iii) jointly minimize energy and rate subject to a constraint on accuracy, etc.

III. COMPRESSION FOR STATISTICAL INFERENCE

The results outlined above show the importance of using data compression in sensor networks. Here we propose a new distortion measure for compression in sensor networks that is useful in the R-E-A view. Since the use of the data collected by sensor networks is ultimately intended to be used for making statistical inference, the distortion of the compression should be measured in terms of its impact on the performance of the inference algorithm. For example, if we wish to estimate parameters, we assert that a useful distortion measure should be based on the Cramer-Rao bound for the estimation problem. In this case one would perform a R-E-A optimization with respect to its effect on the Cramer-Rao bound.

An important sensor network task is image recognition in a distributed security surveillance system. In particular, we consider here the task of locating a template image of an object within a larger image containing the template object. Such a scenario may occur as follows: sensors in the network transmit collected images to a central location for comparison to a database of object images. The collected images are then cross-correlated with each object image to find the object's location in the image. In time-critical, low-bandwidth sensor networks data compression is needed to efficiently (time and energy) transmit the collected images to the central processor.

An important aspect of this kind of application is that locating the object in the collected image does not require visual-quality reconstruction but rather the preservation of information that is crucial to the inference task. Standard image compression algorithms (e.g., JPEG) are based on the fact that human vision is typically less sensitive to high frequency patterns than to low frequency variations. Therefore, the high-frequency components are usually considered unimportant in visual applications and are omitted in the image compression algorithm. But from the theoretical deduction below, we can see that, in this application, the components of high frequencies are more important than those of low frequencies.

To illustrate the idea here, we assume that there are only translations (along the x-axis, along the y-axis, or both) among the images and the database objects. Then the signal model can be formulated as

$$z(n,m) = s(n-x, m-y) + w(n,m) \quad \begin{matrix} n=0,1,\dots,N-1 \\ m=0,1,\dots,M-1 \end{matrix} \quad (2)$$

where translations x and y are to be estimated. The Cramer-Rao bounds for translation x and y are

$$\begin{aligned} \text{var}(x) &\geq \frac{1}{|I(\theta)|_{11}} = \frac{\sigma^2}{\sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \frac{\partial^2 s(n-x, m-y)}{\partial x^2}} \\ &= \frac{\sigma^2}{\sum_{v=0}^{M-1} \sum_{u=0}^{N-1} \frac{\partial^2 s(u,v)}{\partial u^2} \Big|_{u=n-x, v=m-y}} \end{aligned} \quad (3)$$

If $S(u,v)$ is the 2-D DFT of $s(n,m)$, and u and v are the frequencies along the x-axis and y-axis, these become

$$\text{var}(x) \geq \frac{\sigma^2}{\sum_{v=0}^{M-1} \sum_{u=0}^{N-1} (2\pi u)^2 S^2(u,v)} \quad (4)$$

and

$$\text{var}(y) \geq \frac{\sigma^2}{\sum_{v=0}^{M-1} \sum_{u=0}^{N-1} (2\pi v)^2 S^2(u,v)} \quad (5)$$

From (4) and (5) we see that the accuracy of estimating translation parameters x and y are governed by

$$\bar{F}_x^2 = \sum_{v=0}^{M-1} \sum_{u=0}^{N-1} (2\pi u)^2 S^2(u,v) \quad (6)$$

and

$$\bar{F}_y^2 = \sum_{v=0}^{M-1} \sum_{u=0}^{N-1} (2\pi v)^2 S^2(u,v). \quad (7)$$

Large \bar{F}_x^2, \bar{F}_y^2 reduce the Cramer-Rao bounds on $\text{var}(x)$ and $\text{var}(y)$, which result in more accurate results. Thus we desire to ensure that during compression the quantities in (6) and (7) remain as large as possible. Because of the quadratic frequency weighting term in (6) and (7) the high frequency components are especially important, but these are exactly the terms that a standard image compression algorithm (like JPEG) tends to truncate.

The JPEG standard consists of the following main steps: (1) pixel range shifting; (2) DCT on 8×8 pixel blocks; (3) DCT Quantization; (4) DCT “zig-zag” or-

Figure 4, where not only is the correct peak properly recognized but the peak is shaper leading to more accurate location.

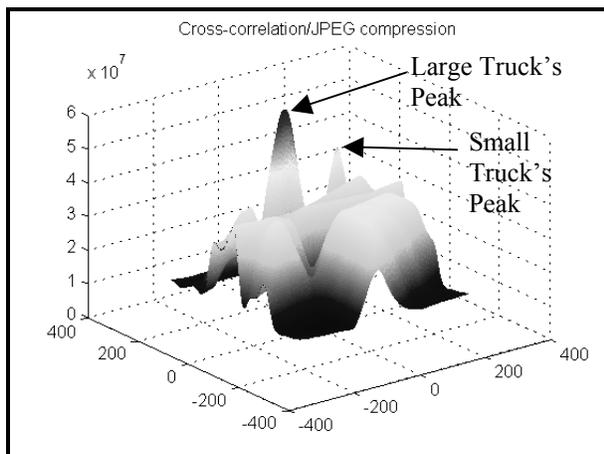


Figure 3: Cross-Correlation when using JPEG compression.

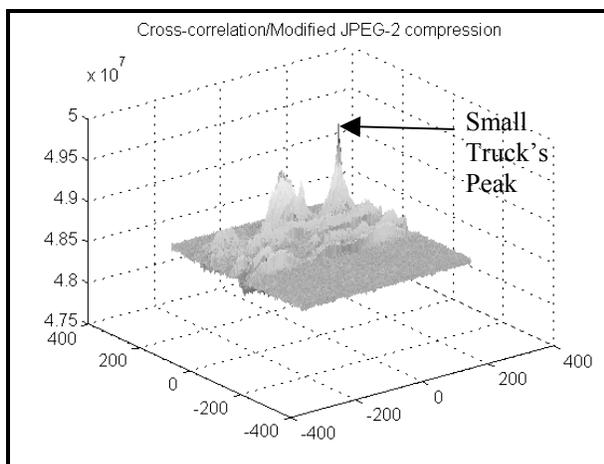


Figure 4: Cross-Correlation when using the proposed distortion measure with JPEG.

IV. SUMMARY

The main ideas put forward in this paper can be summarized as follows: (i) data compression is an use-

ful *general* tool for achieving the needed trade-offs between Rate, Energy, and Accuracy in sensor networks – and provides the benefits of data aggregation in a wider range of scenarios; (ii) the classical approach of using a MSE-based distortion measure should be replaced by inference-accuracy-based distortion measures; and (iii) the standard R-D function used in classical compression should be replaced by the R-E-A function to drive the development of sensor network data compression schemes that can provide operation at desired R-E-A operating points.

REFERENCES

- [1] J. Pottie and W. J. Kaiser, “Wireless integrated network sensors,” *Communications of the ACM*, vol. 43, pp. 51 – 58, May 2000.
- [2] S. Tilak, N. B. Abu-Ghazaleh, and W. Heinzelman, “A taxonomy of wireless micro-sensor network models,” *ACM Mobile Computing and Communications Review (MC2R)*, vol. 6, pp. 1 – 8, April 2002.
- [3] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, “Energy-efficient communication protocol for wireless microsensor networks,” *Proc. of the Hawaii Conf. on System Science*, Jan. 2000.
- [4] A. Scaglione and S. Servetto, “On the interdependence of routing and data compression in multi-hop sensor networks,” *MOBICOM’02*, Sept. 23 – 26, 2002, Atlanta, Ga.
- [5] S. S. Pradhan, J. Kusuma, and K. Ramchandran, “Distributed compression in a dense microsensor network,” *IEEE Signal Processing Magazine*, pp. 51– 60, March 2002.
- [6] M. L. Fowler, “New distortion measures for data compression for emitter location,” *Thirty-Fifth Asilomar Conference on Signals, Systems, and Computers*, November 5 – 7, 2001, pp. 658 – 662.
- [7] N. Nguyen, P. Milanfar, and G. Golub, “Efficient generalized cross-validation with applications to parametric image restoration and resolution enhancement,” *IEEE Trans. Image Proc.*, vol. 10, no. 9, Sept. 2001, pp. 1299 – 130