Delta Compression in Wireless Sensor Networks

Tommy Szalapski
Department of Computer Science
Missouri S&T, USA
tmsytc@mst.edu

1. Introduction

Wireless sensor networks and other data stream-based systems may be subject to higher input rates than their resources can handle. When overloaded, the system must utilize some scheme to overcome this difficulty while maintaining quality, throughput, and latency requirements. One common way of dealing with this problem is the use of compression to reduce the amount of data flowing through the system.

Due to the limited storage and processing power of wireless sensors, the methods for compression within such networks must be have low associated costs for performing the compression. Also to achieve savings within the network compression must be done without global knowledge of the data.

Many current methods sacrifice latency [1], scalability [2], or quality [3] to achieve improvements in bandwidth utilization and throughput.

We propose a two-layer compression scheme that reduces the amount of data flowing through the network while avoiding the above sacrifices. In the first layer, the data is transformed using delta compression then the second layer compresses the delta values. No delay is needed to collect similar values and the data can be decompressed at the sinks or base stations without any loss in granularity or accuracy. Also the compression can be done on any size data set with any range of possible values. Preliminary simulations show promising results for several real-life data sets.

2. Delta compression

In delta compression data is stored in relation to differences between similar pieces of data. Delta compression is most often used to compress similar files [4] or web pages [5].

In a wireless sensor network, sensor readings and other streamed values can be expressed as the change in the value from some expected value. In many data sets, with well chosen expected values, these changes tend to be small relative to the range of possible values and lead to a high frequency of values within a small range.

Relative frequencies of change values tend to fit "bell-shaped" statistical distributions. These change values can then be compressed using standard compression techniques such as Huffman coding [6]. In addition to data size reduction the bell-shaped nature of the frequency histogram allows for compression of the lookup tables for the Huffman codes.

3. Obtaining expected values

The higher the variance in the frequency histogram, the smaller the compression ratios can be. The expected values can be based on historical data or determined dynamically over some subset of the current data.

3.1 Mean

The most basic method for obtaining an expected value would be to use the mean value of the data. However, while the typical bell-shaped nature of the data centered about the mean would allow for the reduced lookup table, the compression would be no better than simply applying Huffman coding to the original data.

3.2 Previous value

In many sensor network applications the data tends to increase or decrease over time and often the values differ little from the previous value. In data sets analyzed the frequency histograms based on a previous value expectation often were tightly clustered around the expected value with a sharp spike at no change. This method also tends to give the data a more symmetrical distribution making the lookup table reduction more effective.

3.3 Prediction models

More sophisticated prediction models for the data would lead to even tighter clustering around the expected values. Prediction models also allow for more scalability to a wider range of data sets and applications. However, this reduces the ability to form generalized solutions. Currently we are researching prediction models for sensor networks to attempt to find a small set of models that will effectively represent a significant percentage of data sets.

4. Limitations and Conclusions

In order to achieve data reduction by compressing the data sacrifices must be made to processing time on the individual sensing nodes. Also, storage requirements would increase to accommodate the lookup tables and prediction models. However, in wireless networks the limiting resource is often bandwidth and the reduction in data and buffer sizes could cause a net gain in storage.

If the data in the network is uniformly distributed, delta compression offers no savings at all. However, uniform data is fairly rare in wireless sensor networks and no compression scheme to date is effective on such data.

5. References

[1] T. Arici, B. Gedik, Y. Altunbasak, and L. Liu, "PINCO: a Pipelined In-Network Compression Scheme for Data Collection in Wireless Sensor Networks," In Proceedings of 12th International Conference on Computer Communications and Networks, October 2003.

[2] D. Petrovic, R. C. Shah, K. Ramchandran, J. Rabaey, "Data Funneling: Routing with Aggregation and Compression for Wireless Sensor Networks," In Proceedings of First IEEE International Workshop on Sensor Network Protocols and Applications, May 2003.

[3] M. Li and D. Kotz, "Towards collaborative data reduction in stream-processing systems," International Journal of Communication Networks and Distributed Systems (IJCNDS), February 2008.

[4] R. C. Burns and D. D. E. Long, "Efficient distributed backup with delta compression," In Proceedings of the 1997 I/O in Parallel and Distributed Systems (IOPADS'97), 17, 1997, pp. 26-36.

[5] J. C. Mogul, F. Douglis, A. Feldmann, and B. Krishnamurthy, "Potential benefits of delta encoding and data compression for http," in SIGCOMM, 1997, pp. 181-194.

[6] Mikio Kawahara, Yi-Jen Chiu, Toby Berger: High-Speed Software Implementation of Huffman Coding. Data Compression Conference 1998:553