

# Study of Compression Mechanism for Reducing the Data Redundancy of Sensor

Huafeng Sun

College of Electronic and Information Engineering, Tianjin Vocational Institute  
Tianjin, China  
shf4211@163.com

**Abstract**—Data acquisition is the most fundamental task of the wireless sensor network. This sensor data to reduce redundancy of the compression mechanism, compression module is given a concrete realization of the cost of the LTC algorithm itself. Compression mechanism of the design decisions were discussed motives, given the specific description of the mechanism, combined with the concrete realization of the system compression module analyzes its performance.

**Keywords**—data acquisition; data redundancy; cost projections; LTC; compression judgments

## I. INTRODUCTION

Data acquisition is the most fundamental task of wireless sensor network. In order to analyze the characteristics of the target area, usually frequent data collection node is required, and the results are sent back to the base station. As sensor data with high redundancy, if the direct transmission untreated raw data, will lead to two issues:

- Node needs a large amount of data transmitted, not only energy-consuming node itself, will increase network congestion and packet loss;
- Even if the data transmission to the security, but the high data redundancy makes cryptanalysis easier, and thus seriously affect the security of the system.

In order to reduce energy costs and improve data security, commonly used mechanism is mainly divided into two kinds of data compression and data fusion, respectively, for the time and space sensor data redundancy for processing. Data compression schemes are usually collected on a single node, data processing, data redundancy is to reduce the most important part.<sup>[1]</sup>

Lightweight sensor network compression algorithm of the computational overhead is still large. Therefore, when the data characteristics and system parameters change, the compression algorithm does not necessarily significantly reduce data redundancy.<sup>[2]</sup> This paper proposes a decision for a single node, the compression mechanism to enhance the adaptive compression algorithm, which based on the different requirements of energy and improve system security, can be configurable.

## II. IMPLEMENTATION OF DATA COMPRESSION MODULE

In order to reduce the transmission of data redundancy, improve system security, data compression mechanism for this article will be directly introduced into the security framework for sensor networks, select the LTC (Lightweight

Temporal Compression) algorithm<sup>[3]</sup> as the underlying algorithm compression module system. This paper used MicaZ platform<sup>[4]</sup> of the algorithm to do a special optimized. Optimization measures include the following:

- by adjusting the algorithm structure, eliminating unnecessary floating point operations;
- Try to avoid spending a great division, allowing the use of shift operations instead of division;
- the use of assembly language, the core algorithm.

Using these optimization measures, LTC algorithm ultimately cost the 2290 code bytes, SRAM footprint is 186 bytes. LTC algorithm through simulation and analysis of computational overhead, the computation time can be obtained:

$$T = 2940 + \left(\frac{n}{2} - 2\right) T_1 \quad (1)$$

Type of unit on the clock cycle, where the initial cost of the algorithm 2940, n behalf of the amount of data needs to be compressed (in bytes), T1 is maintained according to the current linear trend of the data will be different, the possible values including 551,645 and 1843. When the target when the data set changes slowly, T1 values of 551 is equal to the probability of a large and rapid changes when the target when the data set, T1 the value is equal to 1843 the probability is great. Therefore, for specific data collection, the formula (1) to calculate the estimated cost range of compression.

Sensor data can change the speed (that is, the level of data redundancy) fall into three categories: rapid changes in type, medium-and slow change-type change.<sup>[5,6]</sup> the standard classification of the data in the same margin of error obtained compression ratio is similar. Figure 1 shows the three belong to different types of sensor data in the error margin - the compression ratio curve, which is the abscissa value of each data unit of the corresponding international standard units. It can be seen from Figure 1, the air temperature data is slowly changing type, data redundancy is high, so the low margin of error to get good under compression; are rapidly changing the relative humidity data type, data redundancy I is low, so a higher margin of error to get a better under compression; moderate changes in atmospheric pressure data type, and the compression somewhere in between.

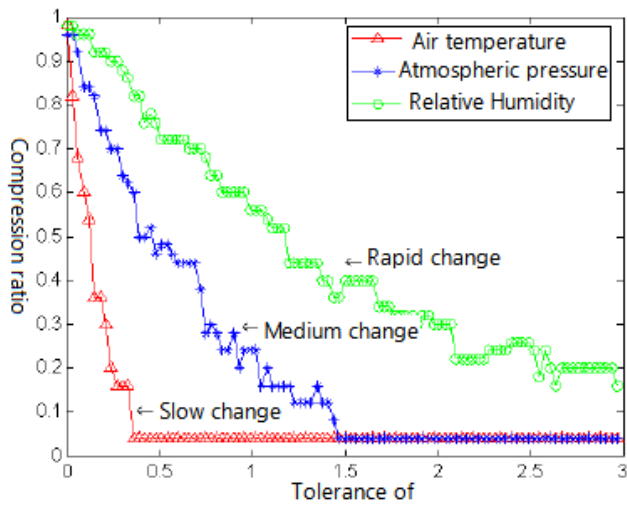


Figure 1. Three different types of data error margin - the compression ratio curve

### III. DECISIONS TO IMPROVE THE ENERGY EFFICIENCY OF THE COMPRESSION MECHANISM

In wireless sensor networks, data compression does not necessarily save energy, depending on the target data characteristics and application requirements. Therefore, it is necessary to introduce a dynamic compression algorithm decision mechanism, the compression effect on the actual forecast. To this end, we propose a decision for a single node compression mechanism to enhance the adaptive compression algorithm, to avoid possible unnecessary energy loss.

#### A. The Decision Mechanism of The Overall Framework of Compression

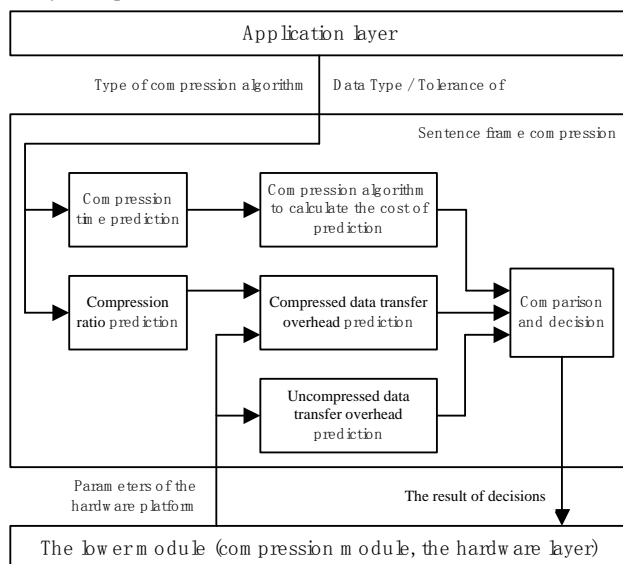


Figure 2. Compression mechanism of the overall framework of decision

Figure 2 shows a compression mechanism for the overall framework of decision. The decision mechanism can

be divided into 6 sub-modules. First of all, there are two sub-modules were used to predict the compression time and compression ratio. The two sub-modules according to different target data types and a specific compression algorithm to compress the execution time and to predict the optimal compression ratio. According to the size of the original data collection, data transmission rate, transmit power and radio unit forecast and the optimal compression ratio, compression time, the calculation can predict the energy cost of microcontroller and RF modules in compressed and uncompressed two cases transport energy costs. Finally, in comparison with the decision sub-module, and send the required compression energy cost and energy required to directly compare the cost in order to predict whether the implementation of compression can bring energy savings. Appropriate feedback to the hardware layer verdict, the behavior of data processing to adjust.

#### B. Mathematical Analysis of Compression Decision Mechanism

Compression mechanism is a decision to find the process of break-even point. The error in the application of tolerances required, the compression can indeed reduce the amount of data required. However, the implementation of data compression algorithm, the calculation inevitably introduces extra energy overhead. Compressed decision mechanism can predict the size of these two, resulting in compressed and sent directly between the two acts to make more choices to save energy.

Let  $E_{uncomp}$  direct transmission of uncompressed data

representative of the energy required for overhead;  $E_{comp}$  representative of the same data is compressed, and then send the results of the energy required to compress the overhead, then:

$$E_{uncomp} = P_{RF}(d) \times L \times T_{tran} \quad (2)$$

$$E_{comp} = E_{MCU} + E_{RF} \quad (3)$$

$$E_{MCU} = P_{MCU} \times L \times T_{MCU}(e) \quad (4)$$

$$E_{RF} = P_{RF}(d) \times L \times CR(e) \times T_{tran} \quad (5)$$

Which  $T_{tran}$  said the time required to transmit a byte, its value determined by the node's transmission rate;  $T_{MCU}$  said generated compressed time cost of a byte, and its range by the formula (1) is given, the specific size and tolerance of e-related; transmit power  $P_{RF}$  is closely related with the communication distance d for a typical node, its value is usually determined by the transmission power of some discrete values [4];  $P_{MCU}$  said the specific microprocessor power consumption; L represents the original data set length, in bytes; CR is predicted compression ratio, its value related with tolerance of e.

One said that the time required to transmit a byte, the value of the transmission rate determined by the node; that

compress the time cost of generating a byte, and its range by the formula (1) is given, the specific size and error tolerance limit of e-related; send power consumption is closely related with the communication distance d, for a typical node, its value is usually determined by the particular microprocessor power of some discrete values<sup>[4]</sup>; that particular microprocessor power consumption; L represents the original data collection length, in bytes; CR is predicted compression ratio, its value and the error tolerance e related.

That the formula (2) and (3) equal to the compression ratio can be expressed in the profit and loss balance point:

$$CR' = 1 - \alpha \frac{T_{MCU}(e)}{P_{RF}(d)} \quad (6)$$

Which

$$\alpha = \frac{P_{MCU}}{T_{tran}} \quad (7)$$

If  $CR < CR'$ , to show that the implementation of compression can save energy. If  $CR > CR'$ , it indicates that the implementation of compression does not save energy. According to comparison results, we can determine in the end is compressed transfer or direct transfer.

### C. Description of Algorithm

Decision mechanism to achieve compression process can be divided into two phases: offline training phase and online at the sentencing stage.

#### 1) The offline training phase

Compressed decision mechanism needs to create two prediction models, one can get used to predict the compression ratio, and the other used to predict the execution time overhead of compression. Time overhead due to compression ratio and the target data characteristics and requirements of the error tolerance is closely related to, so need to use statistical analysis and data fitting techniques to build two forecasting models.

In order to reduce the decision error rate (ie the error in the effective implementation of compressed directly selected data), will be different for each type of data can be obtained under the tolerance of the optimal compression ratio and the shortest time overhead as the corresponding parameters in the decision mechanism estimate. Same time, it is also a compression algorithm in order to maximize the probability of execution, thereby minimizing the transmission of data redundancy.

#### 2) The line at the sentencing stage

This stage is the need to process the practical implementation of sensor nodes, the algorithm flow chart shown in Figure 3.

By comparing the energy cost to get the compression is equivalent to the comparison of the predicted verdict compression ratio and break-even point. Break-even point CR 'actually representing a critical state point, at which point the compression algorithm itself introduces additional computational overhead is just equal to the implementation

of the transmission of compressed energy cost savings. If the compression ratio of the predictive value of CR is still in critical condition can not be met, there is no need to compress the data.

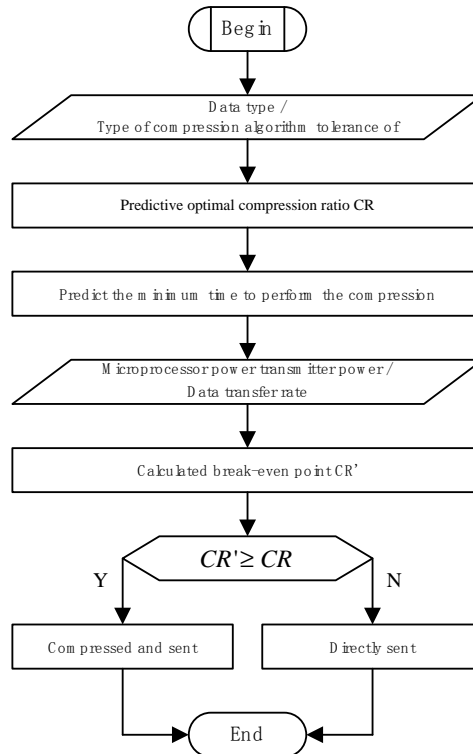


Figure 3. The line at the sentencing stage

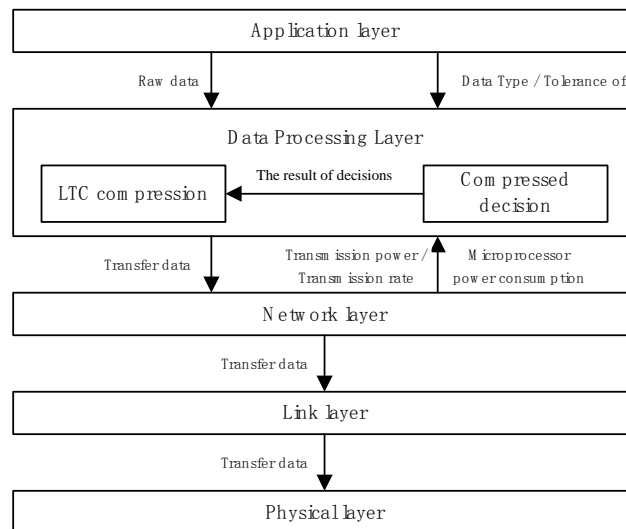


Figure 4. The embedded software architecture adaptive compression

Nodes to achieve compression in the MicaZ decision mechanism, enhance the adaptive capacity of data compression. Overall software structure shown in Figure 4. In this paper, and optimize the data compression module by

means of the same decision mechanism to achieve compression is optimized. According to the results achieved, the compression code decision mechanism itself is only 376 bytes of overhead, SRAM footprint is 16 bytes. Decision to perform a compression time overhead of 406 clock cycles, the overhead of one byte of data than the compression required minimum overhead even smaller, so the mechanism of the overhead introduced is negligible.

*D. Performance Analysis of Compressed Decision Mechanism*

In order to evaluate the energy savings of compression effect of decision mechanism, the need for the introduction of compressed energy cost after the decision mechanism and direct the energy cost of compression algorithm used for comparison. It should be noted that, here are the sensor data compression decision mechanism does not limit the type, select the three types of data compression decision mechanisms simply to illustrate the role of redundancy in the data with different types of effects. Similarly, the decision mechanism of compression algorithm used for data compression and no restrictions, select the LTC algorithm only because of its nodes to achieve low overhead. If you select more complex compression algorithms, compression, there must be more significant judgments mechanism of energy saving effect, since the data for a particular type of greater complexity leads to greater compression algorithms calculate the energy costs.

Each type of sensor data from the 20 groups were selected for analysis, each data length of 100 bytes. Consider the following energy cost in both cases: one is to directly call the data compression module for processing and transmission, and the other is called compression based on the result of decision module to decide whether to call the data compression module. 20 sets of data in both cases the average cost of energy, the results shown in Figure 5,6 and 7, which is defined as the ratio of the energy saving mechanism in the introduction of compressed sentence and the direct use of the energy cost of the energy cost of compression algorithms ratio. To ensure the fairness of the evaluation process, 20 sets of data used here to construct the corresponding set of prediction models with the use of the data set is completely different, they are just the same type of sensor data only.

Figure 5-7 shows the transmit power for the greater use of compressed decision on the mechanism and can not bring energy savings. The reasons for this phenomenon is very intuitive, because for large RF power of communication through the compression for energy savings greater than the compression algorithm into the probability calculation of energy cost is greater and therefore more likely to make a compression mechanism for the implementation of decisions compression the verdict. With the reduction of RF transmit

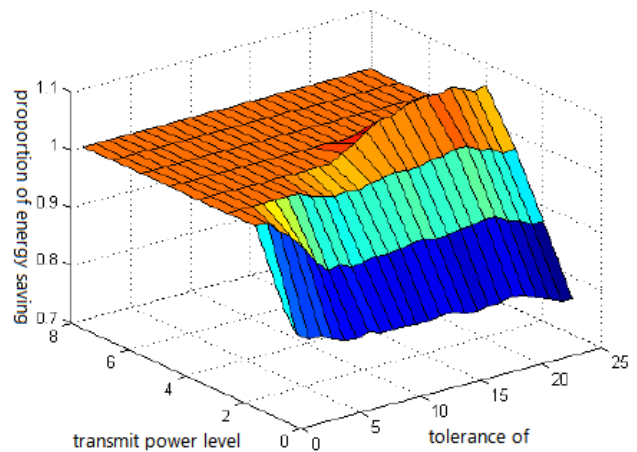


Figure 5. The role of decision mechanisms for rapid changes in the nature of the effect of relative humidity data

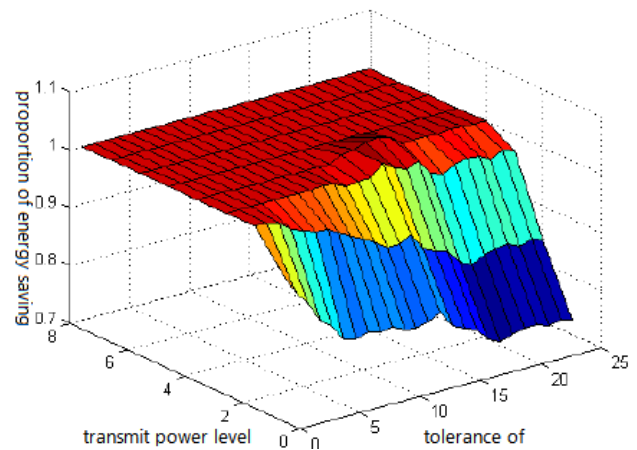


Figure 6. The role of decision mechanisms with moderate changes in the nature of the effect of atmospheric pressure data

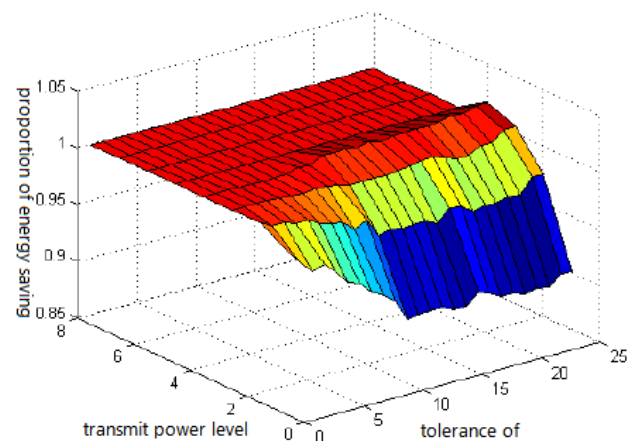


Figure 7. The role of decision mechanism with the slow change in the nature of the effect of air temperature data

power, compression is no longer so much of the communication energy savings, the trend began to waste energy, when the compression mechanism will intervene decision to avoid unnecessary energy loss.

For the type of rapid change, medium change and slow change of type data type, when the implementation of energy saving is no longer compressed, the compression mechanism can make the right decisions decisions. Compared with the direct implementation of compression, the compression decision mechanism, the type of data for the slow changes, the maximum can save nearly 10% of the energy change for the medium type data, the maximum energy savings of over 20% of the energy, and for rapid changes in type data, the energy savings close to 30%.

Moreover, even when the prediction model used in the establishment of reducing the rate of false judgments related measures, but the compression decision mechanism will still produce some false positives, as shown in Figure 5 to 7 in the upper part of Fig. These false positives were seen in the case of secondary transmission power. Because when the transmission power larger or smaller, the effect the implementation of compression (to save energy or waste of energy) is more obvious, more conducive to making the right decision decision mechanism. For slowly changing type of data, waste of energy less than the maximum 2% change for the medium type data, the highest energy wasted less than 3%, while for the fast-changing type of data, waste of energy, but also 7% of the maximum. Decision mechanism and compression can save energy compared to those false positives are acceptable. If you are prone to misjudge the transmission power for proper handling, such as the mechanisms involved in the decision to set compression threshold transmit power, can further reduce the rate of false judgments.

#### IV. CONCLUSION

Therefore, this article presents a decision for a single node compression mechanism to enhance the adaptive compression algorithm, which based on the different requirements of energy and improve system security can be configurable. Experimental results show that different types of sensors for data collection, the mechanism can save up 10% to 30% of the energy, and as energy waste caused by false judgments only 2% to 7%. In addition, the computational overhead introduced by the mechanism itself is only 406 clock cycles, the overhead of one byte of data than the compression required minimum overhead even smaller, so approximation can be ignored. The decision mechanism can be compressed with any compression algorithm for a single node to integrate, as compression algorithms improve the adaptability of the support measures.

#### REFERENCES

- [1] Estrin D. Tutorial for wireless sensor networks: Sensor networks protocols [EB/OL]. [2009-04-10]. [Http://nesl.ee.ucla.edu/tutorials/mobicom02/](http://nesl.ee.ucla.edu/tutorials/mobicom02/).
- [2] Raghunathan V, Schurgers C, Park S, et al. Energy-aware wireless microsensor networks. *IEEE Signal Processing Magazine*, 2002, 19 (2): 40-50.
- [3] Schoellhammer T, Greenstein B, Osterweil E, et al. Lightweight temporal compression of microclimate datasets. *Proceedings of the 29th Annual IEEE International Conference on Local Computer Networks*, 2004: 516-524.
- [4] Crossbow Technology, Inc. MicaZ Datasheet [EB / OL]. [2009-04-10]. [Http://www.xbow.com/Products/productdetails.aspx?sid=164](http://www.xbow.com/Products/productdetails.aspx?sid=164).
- [5] Sadler CM, Martonosi M. Data compression algorithms for energy-constrained devices in delay tolerant networks. *Proceedings of the 4th International Conference on Embedded Networked Sensor Systems*, 2006: 265-278.
- [6] Barr K, Asanovic K. Energy aware lossless data compression. *Proceedings of the 1st International Conference on Mobile Systems, Applications and Services*, 2003: 231-244.