Journal of Physics: Conference Series

PAPER • **OPEN ACCESS**

The Basic Principles and Methods of the System Approach to Compression of Telemetry Data

To cite this article: A V Levenets 2018 *J. Phys.: Conf. Ser.* **944** 012074

IOP Conf. Series: Journal of Physics: Conf. Series **944** (2018) 012074 doi:10.1088/1742-6596/944/1/012074

The Basic Principles and Methods of the System Approach to Compression of Telemetry Data

A V Levenets

Pacific National University, 136, Tihookeanskaya St., Khabarovsk, 680035, Russia

levalvi@bk.ru

Abstract. The task of data compressing of measurement data is still urgent for information-measurement systems. In paper the basic principles necessary for designing of highly effective systems of compression of telemetric information are offered. A basis of the offered principles is representation of a telemetric frame as whole information space where we can find of existing correlation. The methods of data transformation and compressing algorithms realizing the offered principles are described. The compression ratio for offered compression algorithm is about 1.8 times higher, than for a classic algorithm. Thus, results of a research of methods and algorithms showing their good perspectives.

1. Modern Data Compression Techniques

The problem of efficient data coding with minimum redundancy is still urgent due to prevalence of information-measurement systems in different areas of activity. The problem is enhanced due to the wide use of data in audio and video systems, particularly, high quality ones, which results in increasing volumes of transmitted data [1–4]. So, the development of data compressing methods and devices has not lost its urgency.

An amount of work has been done in this area. The majority of the proposed compression methods are applicable to media data compression [5, 6], which can be explained by their prevalence over other types of data in the telecommunications traffic. The peculiarity of such data compression consists in reducing their information content by reference to specific features of human perceptive organs. It should be noted that media data for information-measuring systems are not always considered high priority.

For telemetry systems the most important methods for measuring data compression are those ones which are characterized by reciprocity. Nowadays there is a little information in literature available on the subject, and the decisions offered are oriented to specific fields, e.g. aerospace engineering, sensor networks, etc. At the basis of the majority of such methods there lies the modification of already existing methods for data compression. For example, discrete cosine transform for integer data compression, different realizations of the Burrows-Wheeler transform and the peculiarities of the LZW algorithm is offered. The paper [7] is of special interest since it offers to use the procedure of lossless data compression incorporating two stages: decorrelation and entropy coding. Decorrelation of data was carried out with the use of a few variants of linear prediction and bitwise operation XOR. Entropy coding comprised the well-known Huffman method, Rice coding, arithmetic coding and the Deflate algorithm. The data received via telemetry compression in IRIG-106 format showed, in particular, that the frame with a complex structure built on the ground of the two-stage commutation undergoes more substantial compression if decorrelated in each communication channel.

A some authors for data compression offer to use a neural nets [8] and fuzzy logic [9], although there is no significant advance in this line of research, and the methods offered can be applied to a restricted tasks. It should be noted that the suggested methods for data compression are usually oriented to text-based information.

Alongside with compression methods proper a number of scholars offer data preprocessing methods which lead to a better structured mode, thus increasing the probability of their efficient compression. For example, in [10] the authors offer to consider the frames of telemechanical data as one-dimensional bit sequences which, after correlation dependence assessment, form structures of higher dimensions having a better ordering as compared to initial data. The disadvantage of the suggested method is its focus on quasi-steady zones of the telemetry system. It is necessary to mention that in some cases preprocessing is closely connected with the compressing procedure, which makes it impossible to set them apart.

Alongside with the development of new methods and modification of already existing ones there was developed hardware enabling to increase high-speed performance of data compression and decompression subsystems. Also, some of scholars tried to formalize data compression and preprocessing tasks and view them together with general tasks of the information-measuring system. So, article [11] views the questions of task formalization of adaptive compression and provides a classification of compression algorithm selection criteria. Paper [12] is devoted to the methodology for building distributed information-measuring subsystems, collecting and processing information of complex technical systems whose characteristics can vary over time. In paper [13] there was made an attempt to develop an adaptive compression system whose peculiarity consists in the effective automatic choice of data compression algorithms which results in enhancing data-exchange network characteristics. The offered program system is aimed at improving Internet efficiency by use of three compression algorithms (LZO, Bzip, and Zlib). It uses a number of methods for assessing the transmission time of compressed and uncompressed data and takes into consideration free resources. A decision about the use of a compression algorithm based upon the assessment of received data is made.

However, the overwhelming majority of the developed procedures for data compression are oriented toward the processing of a restricted data class. It should also be noted that there has not been developed a general approach to the building of data compression systems which solves the problems of data preprocessing and compression together with the problems of data adequacy, the efficient use of the communication channel, etc. This integrated approach would make it possible to enhance efficiency of data compression systems as well as the telemetry system on the whole.

A significant point is that the majority of the suggested methods and techniques for data preprocessing and compression are aimed at processing the data of each source (sensor) as a separate object, and do not consider the telemetry frame as a whole. It should be noted that parameter values transmitted through the telemetry frame, as a rule, refer to one complex technical system so the data inside the telemetry frame are correlated even if belonging to one measurement object. Thus, viewing the telemetry frame as a whole and using both existing and new methods, we can achieve a higher level of data compression efficiency. We can assume that the application of this approach can be efficient under certain conditions, which makes the case for conducting research in this area.

2. Basic Principles of System Approach to Compression of Telemetry Data

As it follows from the current status of research, to date, the overwhelming majority of methods for data preprocessing and compression are being developed within the framework of the traditional approach to the tasks, when each data source is processed independently from other sources, without searching for cross-correlation between them. An exception to this are visual data processing methods characterized by lossy compression, which makes direct application of such methods to compressing measuring data of little promise.

On the other hand, telemetry information most commonly comes from one complex technical object whose measurement parameters can be cross-correlated, at times implicitly. Thus, we can argue that when the compression method considers the frame as a whole, its efficiency is higher than if otherwise. We can also assume that the compressed frame can be viewed as a single object (sample) whose substructure can be changed inconvertibly.

The most demonstrable transformation of this type is the presentation of all signal frames as a bit sequence which can later be transformed by given parameters into more complex geometrical patterns. Such patterns, provided they are easily described, can undergo compression more efficiently than the source data frame. It should be noted that we can expect better compression of these patterns using traditional compression methods. However, the maximum effect can be achieved when developing application-specific compression methods and algorithms. In general, we can assume that the dimension of the data structure depends significantly both on the number of data sources, stored in the frame, and their correlation degree.

It should also be noted that telemetry systems are characterized by the status change of the communication channel which links all constituent parts of the system. In this case adaptive control of the communication channel is required, particularly the regulation of the power of the error correcting code. In this instance in order to prevent data loss the communication channel bandwidth is chosen wide enough to provide normal operation under worst conditions.

To lower requirements to the communication channel affected by a short-term impairment, I can offer word length reduction of the transmitted data with a following transmission of the cut-off bits into periods of "normal" channel status. Thus, both the compression subsystem and the data transmission system should be considered as a unified control system, which is to be reflected in the principles of constructing such subsystems. Moreover, the widespread occurrence of distributed information-measuring systems, e.g. sensor networks, makes it necessary to develop such methods for data processing that would require lower computational costs and, consequently, low energy costs, which has a substantial impact on weight-size parameters and self-contained operation time of the elements of such systems.

Thus, considering the aforesaid, I offer the following principles of designing of systems of compression of telemetry information:

1. The telemetry frame is proposed to be considered as a single information space which has its internal cross-correlation relationships including the ones on the bit level.
2. The source data frame is offered to be considered as a bit sequence with a possibility of constructing structures of a better ordering and higher dimensions. To achieve this, we offer to use autocorrelation computation modified with allowance for the binarity of the processed data. The application of this approach will make it possible to develop data preprocessing methods to enhance compression efficiency, including that of the existing methods.
3. The codeword bound violation principle is understood as the absence of the rigid support of initial word length. This approach is supposed to build n-dimensional structures having a better ordering as compared to the structures built with the consideration of the length of the source word.
4. Data preprocessing methods are supposed to be divided into two classes: those for the stationary (quasi-stationary) operation condition and those for the non-stationary operation condition of the system. The first case is based on singling out trends including harmonic ones, dynamic range compression (differential method) and the construction of homogeneous structures (compression method). In the second case may be used the "reduction to stationary" principle, the basis of which is the description of the source data with the use of quasi-random sequences.
5. The development of the methods and algorithms on the basis of the above mentioned principles implies compressing the whole frame or the whole series of sequential frames. By doing so, we exclude the correlation of traditional compression methods with a single data source, which supposedly gives advantages in compression efficiency, at least, for stationary modes.
6. The structures for data preprocessing and compression subsystems should provide minimum equipment, computational and energy costs aiming at ensuring the possibility of their application in different areas of modern technology, including sensor networks and information-measuring systems of different purposes.

For full realization of possibilities of offered approaches, the compression subsystem should be organized in the form of the multilevel system presented on Figure 1. The first level is formed by the well-known compression algorithms providing the guaranteed compression ratio. They work when you turn on the system, and also in case of an inefficiency of algorithms of higher levels. In process of accumulation of the statistical data, the compression subsystem starts to work with algorithms of higher level which consider an implicit correlation both in a frame of the data, and in the data flow from one source.

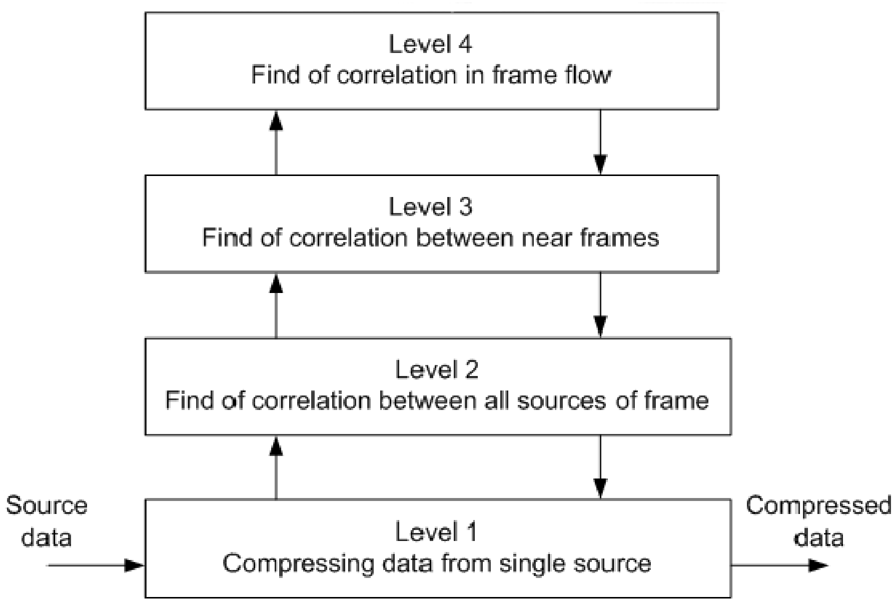

Figure 1. Multilevel organisation of data compressing system.

In case of compression of telemetry it is necessary to consider that they are transferred by frames of the fixed size with invariable (as a rule) structure. Besides, wide using of the distributed information measuring systems, in particular, of sensor networks, leads to the need to developing of methods of data handling with low computing expenses and small energetic expenses.

3. Methods and Algorithms

In this paper a stream of frame, for which the mean of an arithmetical difference between the same samples in two next frames will make no more than four units (without a sign) we will designate as stationary. It is obvious that for such a flow as a method of compression of the first level will be effective differencing algorithm, the essence of which is the element by element calculation of the arithmetic difference between the same counts in the current and the previous (next) frames.

An important task is the decomposition of the source data frame to two-dimensional structure and highlight in it the regions that are maximally correlated in the time. The application of difference algorithm to these blocks will provide a high compression ratio. Thus if we have the completely identical blocks in neighboring frames, to describe them we need only a few bits (a flag of the immutability of the data block and a relative address in the frame). An example of such algorithm and results of a study of its effectiveness for the problem of compression of telemetry data is describing in [14]. This algorithm provides an average compression ratio of telemetry data about 4.0 and can works in real-time mode. Some methods for higher levels of compression are described below.

3.1. Word Length Convertion

This method belongs to methods of preprocessing of measuring data before holding a procedure of compression and is necessary for transformation of statistical characteristics of the compressed data for compressing them with maximum efficiency. It should be noted that the measuring data from the object operating in normal mode have a stationary parameters. These data are well compressed by delta-encoding algorithm. However, when we need a compression of non-stationary data, the compression efficiency may reduce to unacceptable values.

As a method of data preprocessing, which will allow avoiding (reducing) the loss of efficiency of the compression algorithm, it is proposed to change the word length of the compressed data. It is based on a representation of the source data by one-dimensional bits sequence that is further breaking onto words of a predetermined bit width. An example of a procedure for a test sequence of 8-bit unsigned samples is shown on Figure 2.

Here the most optimal division of source sequence of the data was by 5-bit words. However, the offered method will not be so effective for all possible data sets.

It is obvious that high tetrads of words of the source sequence are changed less than low tetrads. Therefore in case of such break it is possible to expect the maximal quantity of repetitions of one value. Therefore, it is possible to expect increase in efficiency of compression.

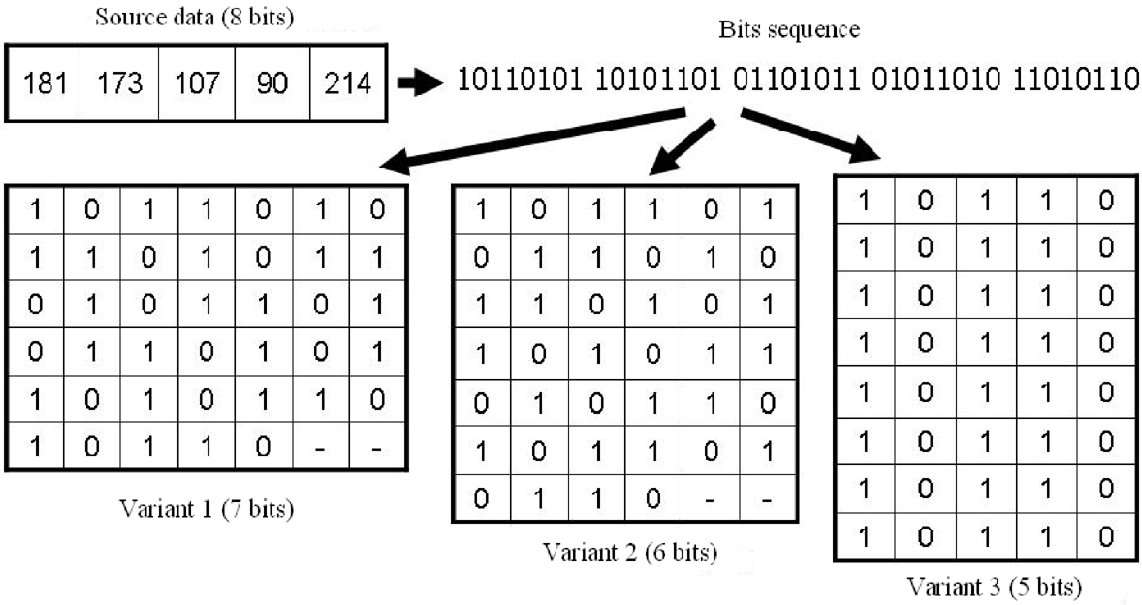

Figure 2. Example of word-length convertion

For verification of such assumption the data from system of telemechanics of energetic object were studied. Word length of data is 8, and the volume of data is 10 000 – 20 000 samples. The typical histograms received before and after preprocessing of 19 285 samples are shown on Figure 3.

The estimation of transform efficiency was carried out using a compression algorithm with codes of variable length. It was assumed that the most frequent value is encoded by bit "0", a three less frequent values - bit sequence "1xx", and other values – a bit sequence "111xx".

The histogram value "1" corresponds to a maximum repetition rate of 14 046, three less frequently occurring values "0", "2" and "3" correspond to the frequency 7 594, 621 1 559 and 1, respectively. Thus, the total volume of the transformed data stream is 14 046 (7 594 1 621 1 559)·3 13 750·5 = 115 118 bit, the volume of the original flow is 154 280 bits, so the value of compression ratio is approximately 1.33. Compressing of the source sequence (not transformed) in the same way provide of value of compression ratio is approximately 0.2.

In the process of the study were reviewed of 150 sets of data, and for 87% of the sets were obtained similar results. The average value of the compression ratio is approximately 1.4. Thus we can say what the proposed method is useable for the tasks of compression of the measuring data. In my opinion, a further increase of the compression ratio we can obtain only with a specialized compression algorithm.

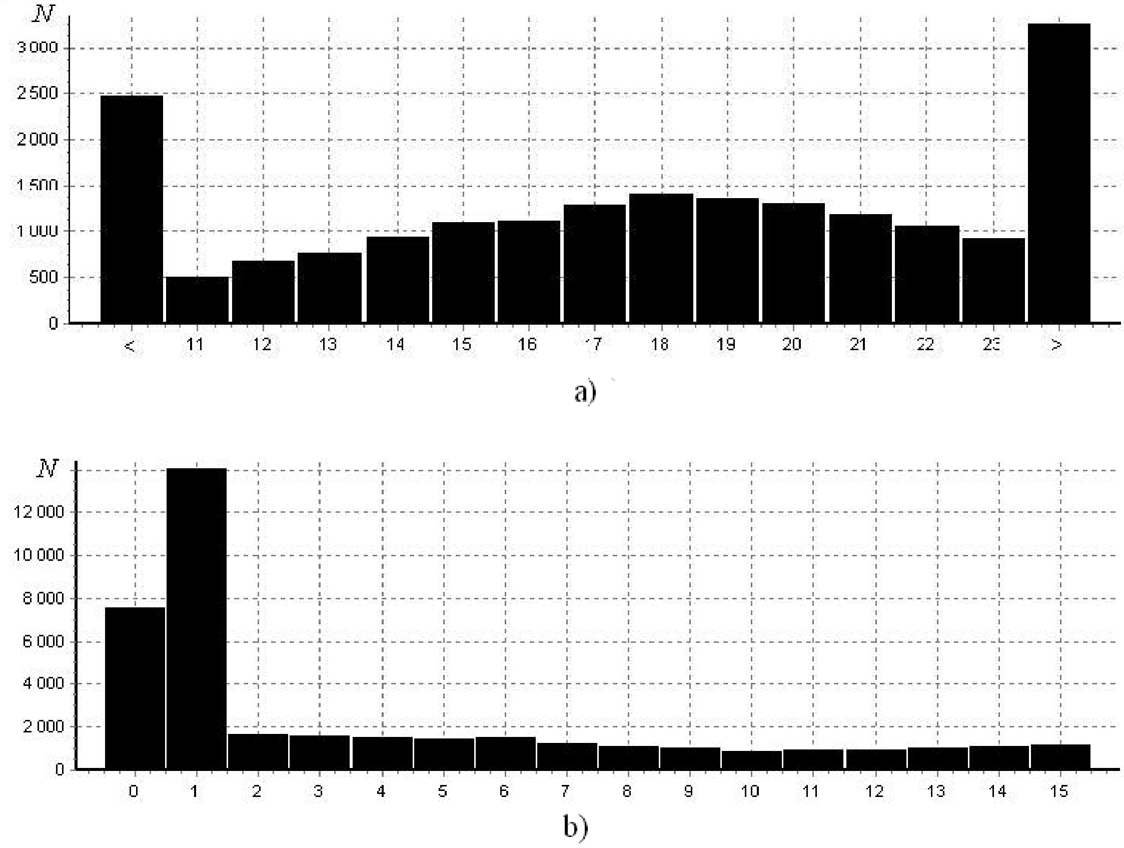

Figure 3. Example of histogram before (a) and after (b) data preprocessing

3.2. Representation of Input Data

An important task is to find the optimal representation (for compression) of the source sequence of digital data. As the source data is used a random data with uniform distribution. The optimal transformation can be estimated by the number of trees. In this case the term "tree" refers to the set of closely spaced identical elements.

Search of optimum representation happens as follows:

- The initial sequence of measuring data will be transformed to the bit sequence;
- The bit's autocorrelation function of a data is calculated. Function $R(n)$ is defined as number of coincidence of values of the current bit of the signal and bit with given distance n. For bit sequence S with length N this function defines as:

$$R(n) = \frac{1}{N}\left(\sum_{i=0}^{N-1} S[i] \leftrightarrow S[i+n] - \sum_{i=0}^{N-1} \overline{S[i] \leftrightarrow S[i+n]}\right),$$

  where $\leftrightarrow$ is a logic operation of equivalence.
- The autocorrelation function breaks into blocks on m bits. In each block the searching of the maximum value is run. Maxima collect in a special array. Width of two-dimensional representation is calculated as the ratio of an index of the current maximum to an autocorrelation function of number of its block;
- Matrixes of values of width of the unit are built and for each matrix the number of trees is counted. To optimum width of the block there corresponds the smallest number of trees.

The analysis of obtained data shows that mathematical expectation of optimum width of the block for different amount of samples of a measuring signal (from 250 to 4000) lies in the range 7–13 while boundary values of optimum length of block lies in the range 5-30.

The average length of a tree can be used for estimating of compression ratio of the processed data. The conducted researches showed that this value is about 6.0 and practically doesn't changing. It allows speaking on the one hand, about good potentiality of compression, and with another – about need of search of more effective methods of determination of optimum parameters of data representation.

It is necessary to note that the offered data transformation isn’t restricted by two-dimensional representation, i.e. its representation by objects of higher dimensionality is possible that can lead to even more essential abbreviation of information redundancy.

3.3. Compressing Algorithm

Considering features of the telemetry information, it is possible to offer the following compressing algorithm of the second level. The source data frame is presented as the two-dimensional structure which breaks onto the rectangular blocks which are slightly changing in time. Then a block-by-block compression by the delta-encoding algorithm is executing.

For the offered algorithm it is necessary to note the two main operations: a choice of the amount of two-dimensional representation of a frame and its break onto rectangular blocks with the maximum correlation. Estimation of correlation of blocks happens by computation of a difference between adjacent blocks with a choice of the minimum value. Estimation was made for some given period.

The algorithm was tested by means of telemetry data from energetic objects. Mean value of a compression ratio lies in the range 7.5–8.0 that is about 1.8 times higher, than for a test algorithm. In case of stationary data mean value of a compression ratio makes about 9.5.

Thus, the offered algorithm in simple variant, without of optimization of parameters is more effective, than the algorithms with classical technology of compression. Shortcoming of this algorithm is that it doesn't consider dependence of more than for two adjacent frames. This is especially noticeable for data with the monotonic increase and decrease of values. It is also necessary to consider that the algorithm potentially allows working in real time, possessing rather low computing expenses.

3.4. Conclusion

Methods for preprocessing before the compression procedure based on the geometrical approach to data structuring and resulting in improving the existing compression algorithms. The result corresponds to the international research level.

Methods for telemetry data compression considering cross-correlation in the telemetry frame among data coming from different sources (sensors). The result corresponds to the international research level.

Thus, the offered approaches potentially can be used by developers of efficient information-measuring multi-purpose systems, including telemetry systems, where combine the requirements of operation in real time and necessity of transmissions of large volumes of information by communication channels with limited throughput.

References

[1] Wootton C 2005 *A practical guide to video and audio compression* New York Springer 787.
[2] Salomon D 2006 *Data compression: the complete reference* New York Springer 900
[3] Candes E J and Wakin M B 2008 *IEEE Signal Processing Magazine* **25** 21-30
[4] Krintz C and Sucu S 2006 *IEEE Trans. Parall. and Distrib. Syst.* **17** 15-24
[5] Miano J 2002 *Compressed Image File Formats* New York ACM Press 336
[6] Pujar J H and Kadlaskar L M 2010 *J. of Theoretical and Applied Information Technology* **15** 18–23
[7] El`shafei M A, Sidyakin I M, Haritonov S V and Vornychev D S 2014 *Vestnik MGTU im. N.E. Baumana. Ser. “Priborostroenie”* **3** 92-104
[8] Talalaev A A, Tishchenko I P, Fralenko V P and Hachumov V M 2008 *Iskusstvennyy intellekt i prinyatie resheniy* **2** 24-33

IOP Conf. Series: Journal of Physics: Conf. Series **944** (2018) 012074 doi:10.1088/1742-6596/944/1/012074

[9] Ibrahim Wael R Anis and Morcos Medhat M 2005 *IEEE Trans. Power. Deliv.* **20** 2136-43
[10] Ch'e En Un, Levenets A V and Nilga V V 2011 *Informatsionno-upravlyayushchie sistemy (Information and Control Systems)* **6** 7-10.
[11] Filippov A K and Rufitskiy V A 2006 *Proektirovanie i tekhnologiya elektronnykh sredstv* **3** 14-19
[12] Vasil`ev A M 2014 *Vestnik Moskovskogo gosudarstvennogo universiteta priborostroeniya i informatiki. Seriya: Priborostroenie i informatsionnye tekhnologii* **50** 5-16.
[13] Krintz C and Sucu S 2006 *IEEE Trans. Parall. and Distrib. Syst.* **17** 15-24
[14] Levenets A V and Chye En Un 2012 *Proc. Int. Conf. European Science and Technology* (Munich) 181-4