

Surprise Event Detection of the Supercomputer Execution Queues*

Zoltan Gal^a, Tibor Tajti^b, Gyorgy Terdik^a

^aUniversity of Debrecen, Debrecen, Hungary
zgal@unideb.hu, terdik.gyorgy@inf.unideb.hu

^bEszterházy Károly College, Eger, Hungary
tajti@aries.ektf.hu

Abstract

Huge amount of data is generated by and collected from the IoT (Internet of Things) physical and virtual devices. These sets of data series reflect in complex form the state of a given system in multidimensional space. Healthiness evaluation of a given system implies state analysis with enhanced methods. Statistical analysis of the stationary systems can be done favourable with Fourier transform. Time invariant global parameters of such time series can be derived based on this transform. Special events can appear during the execution of jobs in a supercomputer (HPC – High Performance Computer) system. Depending on the HPC architecture hundreds or thousands of computation nodes are working in parallel. The scheduler of the HPC front-end node manages different queues (parallel, serial, test, etc.) of the job execution. The multitude of data series captured periodically with several tens of thousands of samples creates a set of several dozen variables for each computation node. Because of the existence of several thousand or tens of thousands of CPU cores, the healthiness of the whole HPC system is a temporal concept in the term of 2D or 4D multidimensional spaces. Transient phenomenon appearance during the HPC job execution time interval implies development of extreme event detection mechanism based on non-stationary statistical analysis. In this paper we propose a healthiness evaluation method for each execution queue of a HPC system with ~ 20 TFLOP/s computation capacities. Time independent community structure is determined and controlled based on multiple similarity measures (i.e. cosine similarity, Sorensen-Dice coefficient, parameterised correlation index, Hamming distance, Jaccard index, Tanimoto similarity) and ANN (Artificial Neural Network) based SOM (Self-Organized Map) algorithm. For each cluster of variables is determined a representing variable, including time specific and global characteristics of the own cluster. The resulting set of representing variables contains less than ten dissimilar time series. Short Time Fourier Transform and Wavelet methods are used for extreme event detection in time of each representing variable. The surprise event detection in time of the HPC execution queues is based on the simultaneity of extreme event fingerprints.

* This work was partially supported by the TÁMOP-4.2.2.C-11/1/KONV-2012-0001 (FIRST – Future Internet Research, Services and Technology) project. This project has been supported by the European Union, co-financed by the European Social Fund. This work was also partially supported by the European Union and the European Social Fund through project Supercomputer, the national virtual lab (grant no.: TAMOP-4.2.2.C-11/1/KONV-2012-0010).

Keywords: High Performance Computer, sensors/actuators, IoT, Complex Event Processing, Event Stream Processing, Artificial Neural Networks, SOM, FFT, STFT, Wavelets.

MSC: 91A28, 65C60, 60G35

References

- [1] Daxin Jiang, Jian Pei, Aidong Zhang, "DHC: A Density-Based Hierarchical Clustering Method for Time Series Gene Expression Data", *BIBE*, pp. 393-400, 2003.
- [2] Jain, AK and Murty, MN and Flynn, PJ "Data Clustering: A Review" In: *ACM Computing Surveys*, 31 (3). pp. 264-323.
- [3] Gy. Terdik, Z. Gal, "Advances and practice in Internet of Things: A case study", *Proceedings of 4th IEEE International Conference on Cognitive Infocommunications (CogInfoCom 2013)*,, Budapest, Hungary, December 2-5, 2013.
- [4] Zoltan Gal, Tibor Tajti (2013): "Complex Event Processing in Supercomputer Environment: Sensor and Neural Network Based Analysis", *Proceedings of IEEE 4th International Conference on Cognitive Infocommunications (CogInfoCom 2013)*, Budapest, Hungary, December 2-5, 2013, ISBN: 978-1-4799-1544-6, pp 735-740.
- [5] S. Sadek, A. Al-Hamadi, B. Michaelis, and U. Sayed, "A statistical framework for real-time traffic accident recognition," *Journal of Signal and Information Processing*, vol. 1, pp. 70–81, 2010.
- [6] Special Issue on Cognitive Infocommunications (2012), *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 16:2.