J. Parallel Distrib. Comput. 93–94 (2016) 146–153

Contents lists available at ScienceDirect

J. Parallel Distrib. Comput.

journal homepage: www.elsevier.com/locate/jpdc

A failure index for HPC applications

Andrei Păun [a,b,*], Clayton Chandler [c], Chokchai Box Leangsuksun [c], Mihaela Păun [c,d]

[a] ICUB/Faculty of Mathematics and Computer Science, University of Bucharest, Bucharest 010014, Romania
[b] Escuela Técnica Superior de Ingenieros Informáticos, Universidad Politécnica de Madrid, Madrid 28660, Spain
[c] College of Engineering and Science, Louisiana Tech University, Ruston, LA, 71272, USA
[d] Bioinformatics Department, National Institute for R&D for Biological Sciences, Bucharest 060031, Romania

HIGHLIGHTS

- A novel metric called Failure Index (FI) that can be used in the evaluation of High Performance Computing failure or resilience information.
- Modeling resource allocation schemes leveraging batches and queues and a history of successful and unsuccessful jobs.
- Index estimates used to construct a reliability-aware metascheduler tasked with processing incoming jobs.

ARTICLE INFO

Article history:
Received 20 May 2015
Received in revised form
30 March 2016
Accepted 20 April 2016
Available online 28 April 2016

Keywords:
Failure Index (FI)
Inequality measures
High Performance Computing
Resilience
System volatility
Adequate level of performance

ABSTRACT

This paper conducts an examination of log files originating from High Performance Computing (HPC) applications with known reliability problems. The results of this study further the maturation and adoption of meaningful metrics representing HPC system and application failure characteristics. *Quantifiable metrics representing the reliability of HPC applications are foundational for building an application resilience methodology critical* in the realization of exascale supercomputing. In this examination, statistical inequality methods originating from the study of economics are applied to health and status information contained in HPC application log files. The main result is the derivation of a new failure index metric for HPC—a normalized representation of parallel application volatility and/or resiliency to complement existing reliability metrics such as mean time between failure (MTBF), which aims for a better presentation of HPC application resilience. This paper provides an introduction to a Failure Index (FI) for HPC reliability and takes the reader through a use-case wherein the FI is used to expose various run-time fluctuations in the failure rate of applications running on a collection of HPC platforms.

1. Introduction

The provision of resilient HPC applications running on innately unreliable hardware is a grand challenge within the exascale supercomputing community. Such development effort requires the coordination amongst computer scientists, mathematicians, application developers, production system operators and domain experts from various government laboratories, private organizations and academic institutions. One of the key enablers of such collaboration is the maturation of a consistent ontology for the metrics and data utilized within the HPC research and development community. This community must both re-examine the utility of existing metrics in anticipation of exascale's failure-rich computing environments as well as develop new statistics appropriate to the study of HPC application resilience in this regime [6,10,15,20,22,21]. The availability and performance of systems based on different reliability metrics have been considered in the literature in the last decade in various studies [8,11,14,18]. This paper suggests a new metric: the Failure Index (FI) as an interplay of various reliability indices and relates to the Gini and Atkinson indices from Econometrics. Other authors have previously considered the Gini/econometric indices in the area of HPC, for example in [4,12,13] and [23]. Gini is used for making better scheduling decisions at the level of cloud services or to model the aging of systems, but no other authors so far have used Atkinson or more sophisticated constructs for gathering insight in subsets of the system.

We use the FI index to study failure-rich HPC application log files from Los Alamos Laboratory, data described in [5,17]

* Corresponding author at: Escuela Técnica Superior de Ingenieros Informáticos, Universidad Politécnica de Madrid, Madrid 28660, Spain.

E-mail addresses: apaun@fi.upm.es, apaun@fmi.unibuc.ro (A. Păun), cfchandler@gmail.com (C. Chandler), box@coes.latech.edu (C.B. Leangsuksun), mihaela.paun@incdsb.ro (M. Păun).

http://dx.doi.org/10.1016/j.jpdc.2016.04.009
0743-7315/

and also in [20]. Previously, the data has been used differently from our approach: in [17] the authors are studying the data from a data mining and clustering perspective while in [20] the data is analyzed with basic statistical tools: mean, median, using the cumulative distribution function (CDF) to determine the distribution of failures, etc. The FI exposes fluctuations in the failure rate of an HPC system it operates. HPC platforms with substantial fluctuations in failure activity are defined to be more volatile than those with more consistent failure modes. Importantly, the FI is not intended to replace MTBF or related metrics as a measure of system availability. Rather, the FI should be used as a complement to these measures in expressing HPC failure modes, giving administrators and developers a more robust view of the health and status of their operational HPC environment.

The Gini index is a measure of dispersion frequently used to study inequality, which indicates total inequality in a sample. Both the traditional measures of central tendency, as well as the inequality indices (Gini, Atkinson, etc.) incorporate value judgment, hence not one of them should be considered as *the complete* measure. While using the measures, such as mean time to failure (MTTF), tends to ignore extreme or outlier values, if the two extremes cancel each other and do not change the value of the mean (as compared to a sample that has values around the mean), the Gini index is very sensitive to outliers, even in large samples.

In Table 1, we can see why the measures of inequality are better used complementarily, rather than using them as a single approach. One can see that the mean has the same value for the samples in the three scenarios, hence we could not identify if there is any difference between the scenarios. The coefficient of variation of Scenario 1 identifies the fact that there is less variation among the nodes reliability when compared to scenarios 2 and 3. Fact supported also by the value of the Gini index, which is smaller and closer to 0 when measuring the inequality in reliability for Scenario 1. However, when comparing Scenarios 2 and 3, we obtain the same mean, and an almost identical coefficient of variation. However, even by visual inspection of the sample, one cannot say that the two scenarios exhibit the same reliability characteristics. Inspecting the values of the Gini index, we can say that there is less inequality in the node's reliability for Scenario 3, as it is for Scenario 2. This is due to the fact that Gini investigates the global inequality in a system, however does not capture where in the distribution does the inequality occur. Consequently, we can have two very different distributions of failure with the same Gini index values.

On the other hand, Atkinson index offers the possibility to examine inequality in different areas of the failure distribution, and this is where it overcomes the Gini's index inability.

We have used in the above example the parameter 0.5, basically looking at how the inequality looks like in the upper part of the distribution, meaning that we look at the long term failures. Overall there is less inequality in Scenario 3 (fact captured by the Gini index), but there is more inequality captured in the upper failures (captured by the Atkinson index).

However, Atkinson's index is not enough by itself since it is much more sensitive to the lower part of the distribution (due to its nature: defined for poverty analysis). Due to this fact we need to define another index that is inspired by Atkinson's but it is designed for the failures of HPC systems. We call the newly defined index the Failure Index (FI).

Given the innately unreliable nature of exascale hardware platforms, we theorize that systems with lower FI's – those with more predictable failure modes – present better opportunities for preemptive error handling or fault prediction techniques, even if the severity of those individual failures are greater than those occurring in the more volatile environment.

We believe that by considering an ensemble of the FI and related inequality indices (such Gini and Atkinson indices) one would better highlight outliers (nodes) and their reliability characteristics in a way that will exhibit a clearer picture of the true resilience or reliability of such large scale systems.

Data required to compute the FI could be periodically pulled from the system using instrumentation tools and stored in the system's file system. The FI and associated statistics could be continuously updated and monitored to provide a near-real time view of application reliability trend to the scheduler, other management tools or the application itself.

For example, if one were to construct a reliability-aware metascheduler tasked with processing incoming jobs and scheduling those jobs to a farm of various heterogeneous computing resources, knowing the FI coefficients for each application/system permutation would yield more intelligent scheduling decisions. These metrics can be combined with additional information pertinent to the power footprint and raw failure rate of each permutation to maximize data center efficiency. The failure index of a given permutation can assist finding the appropriate trade space within the triangular computing constraints of performance, power and reliability.

The paper is structured as follows: the next part, Section 2 defines formally inequality indices and describes their previous application to HPC failure. Section 3 describes the proposed Failure Index and Section 4 gives the results of the paper applying the newly defined index on real data, the section finishes with discussions and comments. The conclusion and future work are discussed in Section 5.

2. Inequality metrics and their application to HPC failure data

2.1. Inequality metrics

The indices that prompted the creation of the Failure Index are well known econometric measures/metrics such as the Gini and Atkinson indices as well as the Lorenz curve. More specifically the idea of the new index defined in this paper is based on a modified version of the Gini index used to analyze aging or rejuvenating objects, which was defined in [12]. While these indices have been widely applied in areas outside of econometrics, it is a novel idea to apply them in the context of HPC *application reliability*. Inequality indices are based on the Lorenz curve. This curve, introduced in 1905 by Max Otto Lorenz as a graphical representation of a distribution's level of equality, compares observed events with distributions featuring perfect equality [16]. The Lorenz curve is based on a convex function and has been widely adopted by economists for use in comparing income distributions.

If the Lorenz curve is represented by the function $Y = L(X)$, where X is the population percentile defined by income, the Gini index can be formalized as

$$G = 1 - 2\int_0^1 L(x)dx. \tag{1}$$

From the definition, it is easy to show that the Gini index has values between 0 and 1. Both the Lorenz curve and Gini index can be applied to either continuous or discrete distributions. The Atkinson index is a normalized measure of the statistical inequality found in a discrete data set [1]. An Atkinson value of 0 indicates total equality while a value of 1 indicates total inequality. The Atkinson index is defined using the formula:

$$A_\epsilon(y_1, \ldots, y_n) = \begin{cases} 1 - \dfrac{1}{\bar{y}}\left(\dfrac{1}{N}\displaystyle\sum_{i=1}^{N} y_i^{1-\epsilon}\right)^{\frac{1}{1-\epsilon}}, & \epsilon \neq 1 \\ 1 - \dfrac{1}{\bar{y}}\left(\dfrac{1}{N}\displaystyle\prod_{i=1}^{N} y_i\right)^{\frac{1}{N}}, & \epsilon = 1 \end{cases} \tag{2}$$