

Accepted Manuscript

Efficient clustering for ultra-scale application tracing

Amir Bahmani, Frank Mueller

PII: S0743-7315(16)30093-4

DOI: <http://dx.doi.org/10.1016/j.jpdc.2016.08.001>

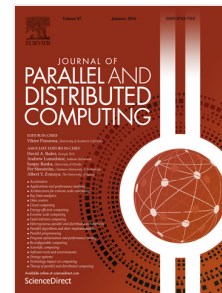
Reference: YJPDC 3533

To appear in: *J. Parallel Distrib. Comput.*

Received date: 15 June 2015

Revised date: 31 May 2016

Accepted date: 2 August 2016



Please cite this article as: A. Bahmani, F. Mueller, Efficient clustering for ultra-scale application tracing, *J. Parallel Distrib. Comput.* (2016), <http://dx.doi.org/10.1016/j.jpdc.2016.08.001>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

*Manuscript

[Click here to view linked References](#)

Efficient Clustering for Ultra-Scale Application Tracing ^{*}

Amir Bahmani ^a Frank Mueller ^{a,*}

^a North Carolina State University Department of Computer Science Raleigh, NC
27695-7534

Abstract

Extreme-scale computing poses a number of challenges to application performance. Developers need to study application behavior by collecting detailed information with the help of tracing toolsets to determine shortcomings. But not only applications are “scalability challenged”, current tracing toolsets also fall short of exascale requirements for low background overhead since trace collection for each execution entity is becoming infeasible. One effective solution is to cluster processes with the same behavior into groups. Instead of collecting performance information from each individual node, this information can be collected from just a set of representative nodes. This work contributes a fast, scalable, signature-based clustering algorithm that clusters processes exhibiting similar execution behavior. Instead of prior work based on statistical clustering, our approach produces precise results nearly without loss of events or accuracy. The proposed algorithm combines low overhead at the clustering level with $\log(P)$ time complexity, and it splits the merge process to make tracing suitable for extreme-scale computing. Overall, this multi-level precise clustering based on signatures further generalizes to a novel multi-metric clustering technique with unprecedented low overhead.

Key words: Clustering Algorithms, Programming Techniques, Concurrent Programming, Performance Measurement
PACS: 07.05.Bx

1 Introduction

Scientific computing applications continue to push the envelope on ever increasing demand for computational power. This trend is driven by a need to increase model

^{*} An earlier version of this paper appeared at ICS’14 [3]. This journal version extends the earlier paper by proposing a new method at the trace creation level called Filtering Function. The paper also compares and reports the performance of clustering under different trace creation approaches. This work was supported in part by NSF grants CCF-1217748 and CNS-0958311.

^{*} Corresponding author.

Email addresses: abahman@ncsu.edu (Amir Bahmani),
mueller@cs.ncsu.edu (Frank Mueller).

Download English Version:

<https://daneshyari.com/en/article/6875097>

Download Persian Version:

<https://daneshyari.com/article/6875097>

[Daneshyari.com](https://daneshyari.com)