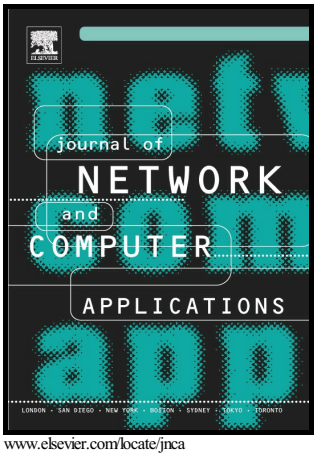
An Adaptive Prediction Approach Based on Workload Pattern Discrimination in the Cloud

Chunhong Liu, Chuanchang Liu, Yanlei Shang, Shiping Chen, Bo Cheng, Junliang Chen



www.elsevier.com/locate/jnca

PII: S1084-8045(16)30319-8
DOI: http://dx.doi.org/10.1016/j.jnca.2016.12.017
Reference: YJNCA1800

To appear in: *Journal of Network and Computer Applications*

Received date: 6 June 2016
Revised date: 13 November 2016
Accepted date: 6 December 2016

Cite this article as: Chunhong Liu, Chuanchang Liu, Yanlei Shang, Shiping Chen, Bo Cheng and Junliang Chen, An Adaptive Prediction Approach Based on Workload Pattern Discrimination in the Cloud, *Journal of Network and Computer Applications,* http://dx.doi.org/10.1016/j.jnca.2016.12.017

An Adaptive Prediction Approach Based on Workload Pattern Discrimination in the Cloud

Chunhong Liu[a,*], Chuanchang Liu[a], Yanlei Shang[a], Shiping Chen[b], Bo Cheng[a], Junliang Chen[a]

[a]*State Key Laboratory of Networking and Switching Technology,Beijing University of Posts and Telecommunications, Beijing, 100876, China*
[b]*Data61, Commonwealth Scientific and Industrial Research Organization, Australia*

Abstract

Generally speaking, the workloads are changing rapidly on the Internet, but there is still regularity of changing patterns. Currently, workload prediction has become a promising tool to facilitate automatic scaling of resource management, and thus reducing the cost and improving resource utilization in the cloud. Most current predication methods of workload are based on a single model. However, because the network traffics are usually mixed and inseparable, it is hard to get the satisfactory prediction performance by means of a single model. To solve this problem, an adaptive approach for work load prediction is proposed in this paper. This approach firstly categorizes the workloads into different classes which are automatically assigned for different prediction models according to workload features. Furthermore, the workload classification problem is transformed into a task assignment one by establishing a mixed 0-1 integer programming model, and an online solution is provided. We used Google Cluster trace to evaluate the proposed approach. The experimental results demonstrate that the proposed approach improves the platform cumulative relative prediction errors by 29.06%, 8.42% and 40.86% respectively in comparison with the time-series prediction methods (Autoregressive Integrated Moving Average(ARIMA), Support Vector Machines(SVM) and Linear Regression (LR).

Keywords: Cloud computing, Workload prediction, Workload category, 0-1 Programming

*Corresponding author
Email addresses: liuchunhong2012@bupt.Xedu.cn (Chunhong Liu), lcc3265@bupt.edu.cn (Chuanchang Liu), shangyl@bupt.edu.cn (Yanlei Shang), Shiping.Chen@csiro.au (Shiping Chen), chengbo@bupt.edu.cn (Bo Cheng), chjl@bupt.edu.cn (Junliang Chen)