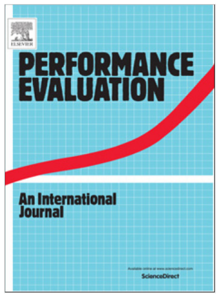
Accepted Manuscript



Parallel algorithms for fitting Markov arrival processes

Mindaugas Bražėnas, Gábor Horváth, Miklós Telek

PII: S0166-5316(17)30385-1
DOI: https://doi.org/10.1016/j.peva.2018.05.001
Reference: PEVA 1958

To appear in: *Performance Evaluation*

Received date : 15 December 2017
Revised date : 23 March 2018
Accepted date : 7 May 2018

Parallel algorithms for fitting Markov Arrival Processes

Mindaugas Bražėnas[a], Gábor Horváth[b], Miklós Telek[b,c]

[a]Department of Applied Mathematics, Kaunas University of Technology, Lithuania
[b]Department of Networked Systems and Services, Budapest University of Technology and Economics, Hungary
[c]MTA-BME Information Systems Research Group, Hungary

Abstract

The fitting of Markov arrival processes (MAPs) with the expectation-maximization (EM) algorithm is a computationally demanding task. There are attempts in the literature to reduce the computational complexity by introducing special MAP structures instead of the general representation. Another possibility to improve the efficiency of MAP fitting is to reformulate the inherently serial classical EM algorithm to exploit modern, massively parallel hardware architectures.

In this paper we present three different EM-based fitting procedures that can take advantage of the parallel hardware (like Graphics Processing Units, GPUs) and apply a special MAP structure, the Erlang distributed - continuous-time hidden Markov chain (ER-CHMM) structure for reducing the computational complexity.

All the proposed parallel algorithms have their strengths: the first one traverses the samples only once per iteration, the second one is memory efficient (far more than the classical serial algorithm), and the third one has exceptionally low execution times.

These procedures are compared with the standard serial forward-backward procedure for performance comparison. The new algorithms are orders of magnitudes faster than the standard serial procedure, while (depending on the variant) using less memory.

Keywords: Markov arrival process, traffic model fitting, EM algorithm, parallel computation, GPU

1. Introduction

Markov arrival processes (MAPs) are being used for modeling correlated workload for traffic, performance and reliability analysis in several fields for many decades [8]. However, for the successful application of MAP-based models, efficient fitting procedures are needed to approximate the real traffic behavior as accurately as possible.

The MAP fitting approaches published so far can be divided into matching, distance minimization, and combined methods, where inter-arrival time fitting and correlation structure fitting are performed with different approaches. The matching algorithms aim to match certain statistical quantities of the traffic such as moments and auto-correlation. Fitting methods belonging to the second group aim to minimize a measure of distance between samples and the model. In case of MAP traffic models, the dominant distance measure is the likelihood, and the dominant optimization procedure is the expectation maximization (EM) method aiming to maximize the likelihood.

Explicit results for MAP matching methods exist only for the second-order case [2]. For larger models, a combined two-step procedure for the MAP fitting problem has been developed, e.g., in [4]. In the first step a phase-type (PH) distribution is created to fit the marginal distribution, and in the second step, a lag-k auto-correlation fitting is performed to capture the correlated nature of the traffic. Another two-step

Email addresses: mindaugas.brazenas@yahoo.com (Mindaugas Bražėnas), ghorvath@hit.bme.hu (Gábor Horváth), telek@hit.bme.hu (Miklós Telek)