Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 35 (2014) 290 – 298

Procedia Computer Science

18th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems - KES2014

Spatial internet traffic load forecasting with using estimation method

Anna Kamińska-Chuchmała*

Wrocław University of Technology, Wybrzeże Wyspiańskiego 27, Wrocław 50-370, Poland

Abstract

Internet traffic is one of the most unpredictable and fluctuating phenomenon to forecast, although accurate prediction is difficult challenge. Many research about measurement experiments are dedicated to predict the performance of Internet network. Especially, during last years this issue is important, when growing demand on reliable access to the Internet is desired by users. In this paper spatial (temporal-area) Internet traffic load forecasting is proposed. Data are obtained from conducted active measurement experiment. Period of time from which is contained database amounts three weeks of October 2013 and each day at the same time at: 06:00 am, 12:00 pm, 06:00 pm and 12:00 am the data were collected. This experiment relies on download a copy of the same resource from servers located in Europe by Wrocław agent. One of the most interesting variable obtained from this experiment is total download time of indicated resource. On basis of this experiment, the Internet traffic forecasts with one week ahead are performed. Spatial forecasting is made by using geostatistical estimation method - ordinary kriging. Paper contains description of ordinary kriging method and preliminary measurement data analysis. Next, model of forecast with discussion of results are given. The final view of performance considered the Internet network in Europe ending the paper.

Peer-review under responsibility of KES International.
Keywords: Internet traffic loads, network performance, spatial forecast, kriging method, active measurement

1. Introduction

Since many years the issue of Internet traffic has been researched e.g. in terms of performance, scalability, and changeability. There exist organizations, such as the Cooperative Association for Internet Data Analysis (CAIDA) which is a collaborative undertaking among organizations in the commercial, government, and research sectors aimed at promoting greater cooperation in the engineering and maintenance of a robust, scalable global Internet infrastructure[1]. CAIDA researchers conduct regular measurements of Internet traffic at various networks, develop analysis tools, analyze available traffic samples, and indicate future directions in traffic classification[2]. The another example is integrated project funding by European Union, in which European Traffic Observatory Measurement InfrastruCture (ETOMIC) was created[3]. During these investigations, active experiment ETOMIC is distributed throughout Europe and allows users to infer network topology and discover its specific characteristics, such as delays and available band-

* Corresponding author. Tel.: +48 71 320 40 20; fax: +48 71 321 10 18.
E-mail address: anna.kaminska-chuchmala@pwr.edu.pl

Peer-review under responsibility of KES International.
doi:10.1016/j.procs.2014.08.109

widths[4]. Moreover a lot of researchers are focused on active measurement for a long time, because it is claimed that the passive measurements could not give any conclusive answer[5] and try to examine performance of Internet on different ways, for example by analyzing packet delay dynamics[6]. Additionally on basis of analysis of experiments could be concluded that traffic associated with the monitoring of individual Web servers is characterized by self-similarity, burstiness, long tail and seasonality[7].

In this paper author using active measurement experiment as database to forecast research, which could be helpful to picture how the traffic in the Internet and network performance are presented at the moment. The gain of this paper is to show new approach to - not in 2D but 3D spatial (temporal-area) - forecast Internet traffic loads with using geostatistical estimation method, which was not used in this domain by now. In the next section related works with Internet traffic load prediction are presented.

2. Related work

Over the past decade the researches were conducted many Internet traffic loads predictions. For example in[8] was proposed Generalized Autoregressive Conditional Heteroscedastic (GARCH) model to forecast traffic load and practical techniques for model fitting. The proposed simulation model Markov Chain Monte Carlo (MCMC) provides approach for simulating internet data traffic patterns. Moreover authors compared their approach with Seasonal Auotoregressive Moving Average (SARIMA) model.

In other paper[9] : ARM, ARIMA, FARIMA (Fractional ARIMA) models, and Fractional Gaussian Noise (FGN) were proposed to Web traffic modeling. Authors presented their approach in six steps. ARMA (Autoregressive Moving Average) and FARIMA are very popular and useful models for network traffic prediction. Paper[10] shows using these two mentioned models for conducted performance tests and comparisons with also Gaussian predictor. Authors conducted comparisons on basis on the mean packet delay, the variance of the packet delay, and the buffer requirements.

The different approach to performance prediction of network is in[11], where authors design and validation of a system that can be used by an autonomic manager to predict the response times of transaction-oriented applications.

However, in those mentioned forecasted methods there is no possibility to obtain temporal-area forecast, it can be only temporal. Moreover, additional parameters such as input data are required to prepare forecast for better accuracy. These disadvantages affects the thinking about better solution as a forecast method, because nowadays network operators would like to know how will appear network traffic in future on whole considered area, also in place where this information is not given.

Interesting response on this request could be geostatistical methods, where values are estimated on whole considered area on the created 3D grid. Moreover, minimum of input parameters is enough to prepare forecast with good accuracy. Geostatistical estimation method was use for example to forecast traffic cars basis on the floating car speed data in Beijing[12]. Having regard to all of advantages of geostatistical estimation method author decided to applied it to spatial Internet traffic forecast.

Currently, to the best of author knowledge the spatial kriging methods approach to Internet loads prediction as presented in this paper is unique, leaving no similar problem statement in the literature. Till now, there is not any research about spatial Internet network loads forecast with using geostatistical methods except author's works (for example[13,14]).

Next section described geostatistical estimation method in details.

3. Estimation method

One of the main geostatistical estimation methods is Ordinary Kriging (OK). The kriging term was coined by Georges Matheron in 1963 in honor of Danie Krige. OK is used to estimate a value at point of an area for which a variogram is known with using data in the neighborhood of the estimation location. It could be defined kriging estimate, named Z^* as a linear combination of the neighboring information Z with weights ω_α:

$$Z^*(x_0) = \sum_{\alpha=1}^{n} \omega_\alpha Z(x_\alpha), \tag{1}$$