ISSN 1392 – 1215 *ELECTRONICS AND ELECTRICAL ENGINEERING* / *ELEKTRONIKA IR ELEKTROTECHNIKA* 2011. No. 6(112)

T 180 TELECOMMUNICATIONS ENGINEERING / *TELEKOMUNIKACIJŲ INŽINERIJA*

Testing the Null Hypothesis of Stationarity of Internet Traffic

G. Lauks, A. Skrastins, J. Jelinskis
Institute of Telecommunication, Faculty of Electronics and Telecommunication, Riga Technical University, Āzenes str. 12, LV-51368 Riga, Latvia, phone: +371 7089201, e-mail: gunars.lauks@rtu.lv

crossref http://dx.doi.org/10.5755/j01.eee.112.6.439

Introduction

Network management techniques have long been of interest to the networking research community. Networks can be viewed as a distributed system in which coordinated and informed decision making is crucial for optimal resource allocation. In this paper we study the problem of finding an optimal policy for Network resource allocation as a Partially Observable Markov Decision Process (POMDP). Testing the stationarity of Network traffic is one of the keystone problem. Partially Observable Markov Decision Process (POMDP) is a basic framework for Multi-Agent planning, when the traffic model is not perfectly known and may change over time and is a well-studied framework for sequential decision-making in partially observable domains. This paper restricts itself to two network management techniques: admission control and the partitioning of transmission and buffer resources among two or more classes of traffic using a common transmission path. The Decision Policy Agent (DPA) model and Network model are presented in fig.1. For Network resource allocation, we are interested in the performance of a queue which represents the bottleneck of a network. In this paper we will use a simple hierarchical MMPP traffic model from [1] and queue model MMPP/GI/1/m. The aim of this paper is to estimate the various stationarity testing procedures for intergation into Network resource allocation agents.

Problem statement

The problem of the testing of the stationarity hypothesis for real traffic measurements is caused by the fact the mean of traffic with LRD does not exist. But, statistical analyses of measured traffic traces often contain non-stationary effects like level shifts or polynomial trends. The testing of the stationarity hypothesis is particularly difficult in the presence of LRD, where many classical statistical approaches cease to hold [2]. In these cases several popular tests for long-range dependence can result in wrong conclusions and unreliable estimate of the Hurst parameter. On the other hand for decision making in CP MDP it is significant to detect the level shifts and/or polynomial trends with reasonable computational complexity. On longer time scales we can observe also a regular character of the traffic due to daily or weekly variations. Three types of trend models are used in our experiments, e.g. linear trend, parabolic trend, and level shift model. Level shift model can be observed when during our traffic measurements suddenly a new source starts to generate the traffic to the network nodes and the *linear and parabolic trends*, which can be observed in daily traffic variations. For example, when people start to work in their office between 8 and 10 am a monotonic increase of the total load of the aggregated traffic can be observed. These traffic trends should be identified by Decision Policy Agent.

MMPP traffic model

In this paper for traffic generation we will use a MMPP traffic model proposed by [1]. This is hierarchical MMPP traffic model, capable of generating traffic that accurately emulates the aggregate Network traffic measured at an edge router. The model is based on a layered architecture of sessions, that generate flows, that finally generate packets.
MMPP model is completely described by these five parameters:

- λ_s – the arrival rate of new sessions,
- λ_f – the flow arrival rate per active session,
- λ_p – packet arrival rate per active flow,
- N_f –average numbers of flow per session,
- N_p – average numbers of packets per flow.

Then:

- $\beta = 1 - 1/N_f$ – probability that a flow is not a last of session,
- $\mu_f = \lambda_p/(N_p - 1)$ – the average duration of a flow,
- $n_f \lambda_p$ –global packet generation rate,

- $n_s \lambda_f + \lambda_s$ – flow generation rate (n_f – number of active flows and n_s – number of active sessions).

For network planning and dimensioning, we are typically interested in the performance of a queue which represents the bottleneck of a network. Besides the advantages of being simple to implement and efficient, a synthetic Markovian source as the one we propose has the additional advantage of allowing a Markovian model of a queue. In general, the buffer model can be described by a MMPP/GI/1/m queue, where the service time represents the transmission time of a packet, and can be easily derived from the capacity of the link and the distribution of the packet length. The general service time distribution can be approximated by a phase type distribution. By adopting an exponential service time distribution, we obtain an MMPP/M/1/m queuing system. [1]

The infinitesimal generator (IG) of such a (Continuos-Time Markov Chain) CTMC is matrix A

$$A = \begin{pmatrix} A_{00} & A_0 & 0 & \dots & \dots & 0 \\ A_{10} & A_1 & A_0 & \ddots & \dots & \vdots \\ 0 & A_2 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & \dots & \ddots & A_2 & A_1 & A_0 \\ 0 & \dots & \dots & 0 & A_2 & A_{mm} \end{pmatrix}, \quad (1)$$

where $A_{00} = Q - \Lambda$, $A_0 = \Lambda$, $A_1 = Q - \mu \cdot I_n - \Lambda$, $A_{mm} = Q - \mu \cdot I_n$, $A_2 = \mu \cdot I_n$ and Λ is the rate matrix $\Lambda \in R^{n \times n}$, $A_{00}, A_{10}, A_0, A_1, A_2, A_{mm}, I_n \in R^{n \times n}$, and I_n is the identity matrix. More detailed description of this IG can find in [1].

Decision Policy Agent model

Decision policy agent (DPA) and Network model are presented in Figure 1.

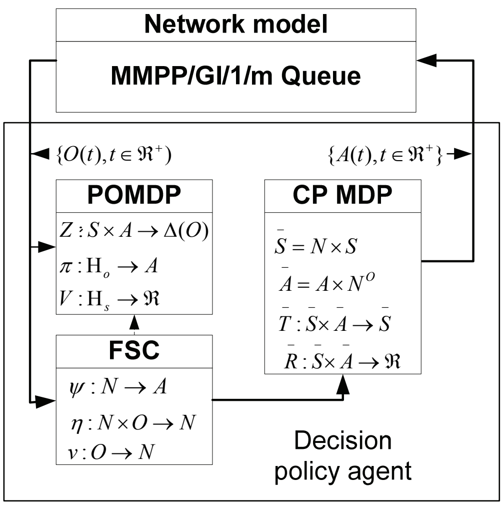


Fig. 1. Decision policy agent (DPA) and Network model

DPA model consists of three objects: Partially Observable Markov Decision Process (POMDP), Finite state Controller (FSC) and Cross Product Markov Decision Process (CP MDP). For POMDP we use the following notations. A POMDP is a tuple $\langle S, A, T, R, O, Z \rangle$, where:

- a set of states $S = \{s_1, s_2, \dots s_{|s|}\}$,
- a set of actions $A = \{a_1, a_2, \dots a_{|A|}\}$,
- a observation space O,
- a set of transition probabilities $T(s_i, a, s_j) = p(s_j | s_i, a)$,
- a observation function $Z : S \times A \to \Delta(O)$,
- a rewards function $R : S \times A \to \Re$,
- a set of system trajectories or histories H,
- a decision policy $\pi : \mathrm{H}_o \to A$,
- a value function $V : H_s \to \Re$,
- a time steps $t \in \Re^+$.

We use a following definition: policies that do not dependent on stages are called stationary policies. We are study over the infinite horizon policies and agent`s goal is to find a policy π by exacuting at each step (state) the actions that would maximize value function (*cumulative reward over the horizon).*

Finding the optimal policies traditionally needs significant computational resources and is limited in time. Our idea is to reduce these resources by computing a new decision policy only when traffic stationarity has changed. For this we needed proportionate stationarity testing procedure of incoming traffic. Fully observable part of the system is process $\{O(t), t \in \Re^+\}$, where *O(t)* is number of packets in queue. In this paper we are testing several known stationarity test and to look for an appropriate policy π are not resesrch object of this paper.

FSC: Finite state controller model as a deterministic policy graph π is a triple $\langle N, \psi, \eta \rangle$, where:

- N is set of controller nodes n ,
- $\psi : N \to A$ - action selection function,
- $\eta : N \times O \to N$ is the internal node transition function.

We can define the following FSC controller nodes:

- n1 – state with stationary observations in POMDP,
- n2 – state of testing of stationarity in POMDP,
- n3 – state of finding (learning) new decision policy in POMDP.

Cross Product MDP: POMDP with $\langle S, A, T, R, O, Z \rangle$ and policy graph with the node set N cross-product MDP $\langle \overline{S}, \overline{A}, \overline{T}, \overline{R} \rangle$ can be described [7]:

- $\overline{S} = N \times S$ – state space as the Cartesian product of external system states and internal memory nodes, which consists of pairs $\langle n, s \rangle, n \in N, s \in S$,
- $v : O \to N$ – conditional observation strategy for each state pairs,
- $\overline{A} = A \times N^O$ – common action space as the cross product between A and space of observation mappings N^O,

- $\langle a, v \rangle, a \in A, v \in N^O$ – conditional plan, where $v \in N^O$ – is a deterministic observation strategy,
- $\overline{T} : \overline{S} \times \overline{A} \to \overline{S}$ – transition function,
- $\overline{R} : \overline{S} \times \overline{A} \to \Re$ – reward function, such as $R(\langle n, s \rangle, \langle a, v \rangle) = R(s, a)$.

Goal of this article is to experimentally test the following observation strategy:

***IF** (traffic flow is stationary) **THEN** (decision policy is stationary), **ELSE** (Change decision policy).*

Used data set and stationarity tests

General MMPP data set is generated using NS2 network simulator with implemented MMPP traffic model. Parameters were chosen the same as proposed by author of MMPP traffic model. [1]: The arrival rate of new sessions $\lambda_s = 1.42$, the flow arrival rate per session $\lambda_f = 0.71$, packet arrival rate per flow $\lambda_p = 30$, average numbers of flow per session $N_f = 10$, average numbers of packets per flow $N_p = 10$. Generated MMPP data set consists of 3600 samples and is depicted on Fig. 2.

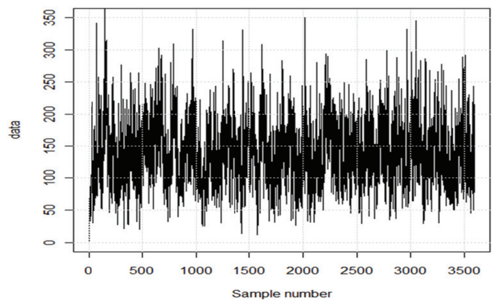
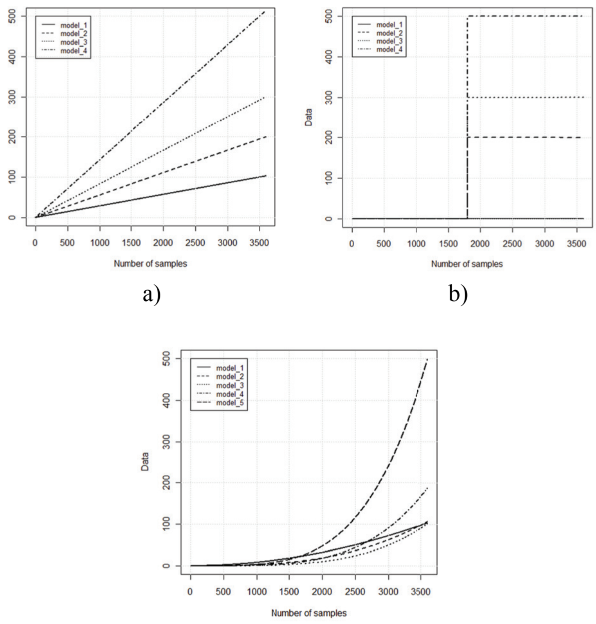


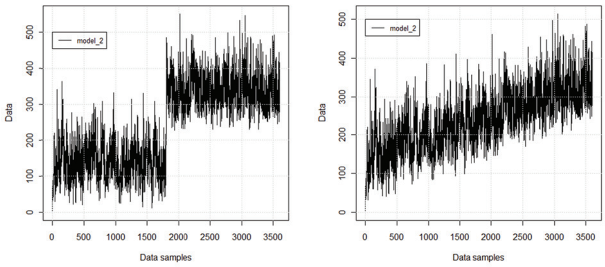
Fig. 2. General MMPP data set generated using NS2 simulator



a) b) c)

Fig. 3. Types of pattern of non-stationarity with different degrees of trend: linear trends (a), level-shift (b) and parabolic trends (c)

This general MMPP data set was simply modified to three types of non-stationary models. In our simulation we only concerned three non-stationarity patterns: *linear trend, level-shift with two states and parabolic trend*. MMPP traffic modification to some of these non-stationary models is obtained by adding general MMPP data set with specific non-stationary pattern (see Fig. 3. for more information of these non-stationarities patterns). Each pattern includes several modification (*degrees of trend*) e. g. linear trend with four different slopes. Example of resultant traffic data set with linear trend and level-shift can be seen in Fig. 4.



a) b)

Fig. 4. Examples of investigated traffic data in cases with level-shift (a) and linear trend (b)

Stationarity testing problem of traffic with Long Range Dependence (LRD) is a hard problem and is a subject of many researches: [2, 4, 5, 6] and others. Contribution of authors is to develop the testing procedures based on statistically correct approaches. Complexity of tests is problem for application these tests in real time regime. There are three important types of time series in our research: stationary, trend stationary and non-stationary (level shifts). A weakly stationary time series has a constant mean and variance. A trend stationary variable is a variable whose mean grows around a fixed trend. In our experiments we are focusing on Kwiatkowski-Phillips-Schmidt-Shin tests (KPSS test) [7]. In our paper we use the following definitions of hypothesis. Let $\{x_t\}, t = 1,2,..., N$ be the observed series for which we wish to test stationarity. Definition of null hypothesis H_0 - (stationarity): the random variables $\{x_t\}, t = 1,2,..., N$ satisfy the null hypothesis, if $x_t = \mu + \xi_t$ where μ is real number and $\{\xi_t\}$ - stationary sequence (noise) with zero mean. The departure from stationarity is modelled by the alternative hypothesis H_1 which includes the deterministic trend plus noise hypothesis – H_T and the unit root hypothesis H_U. Definition of H_T - (deterministic trend plus noise): the random variables $\{x_t\}, t = 1,2,..., N$ satisfy the hypothesis H_T, if $x_t = \mu + g_n(t) + \xi_t$, where $g_n(t)$ - deterministic trend function. Alterantive H_u- (unit root): the random variables $\{x_t\}, t = 1,2,..., N$ satisfy the hypothesis H_U, if $x_t - x_{t-1} = \mu + g_n(t) + \xi_t$, where $g_n(t)$ and μ - are the same as under hypothesis H_T.

In ADF and PP test the null hypothesis H_0 that data are non-stationary against the alternative H_1 that its are stationary.

Statistical analysis was performed using the R statistics software with *tseries* package used for stationarity tests. The test results are listed in Table 1.

Table 1. Test of stationarity using three different kinds of tests by three different traffic models (*Accepted = stationary*)

Linear trend	**Stationarity test**		
	KPSS.test	***PP.test***	***adf.test***
Model_1	Accepted	Accepted	Accepted
Model_2	Accepted	Accepted	Accepted
Model_3	Accepted	Accepted	Accepted
Model_4	Accepted	Accepted	Accepted
Level-shift	**Stationarity test**		
	KPSS.test	***PP.test***	***adf.test***
Model_1	Accepted	Accepted	Accepted
Model_2	Rejected	Accepted	Accepted
Model_3	Rejected	Accepted	Rejected
Model_4	Rejected	Accepted	Rejected
Parabolic trend	**Stationarity test**		
	KPSS.test	***PP.test***	***adf.test***
Model_1	Rejected	Accepted	Accepted
Model_2	Rejected	Accepted	Accepted
Model_3	Rejected	Accepted	Accepted
Model_4	Rejected	Accepted	Accepted
Model_5	Rejected	Accepted	Rejected

Note: *Level of significance* in all cases are chosen 1% (α=0.01)

Conclusions

The tests used in our experiments does not enable to decide between non-stationarities and LRD. We have shown that the presence of different non-stationarities such as level shifts, linear and polynomial trends (parabolic in such case) in the observations can deceive classical LRD methods. These simulation results confirm that short-range dependent (SRD) process with non-stationarities can produce the same variance-time plot as LRD processes. In the case of LRD processes trends can significantly destroy the accuracy of the estimation of the H parameter. These results show that granular computing methods could be more acceptable for selection of conditional observation strategy.

References

1. **Muscariello L., Mellia M., Meo M., Ajmone Marsan M., Lo Cigno R.** An MMPP–Based Hierarchical Model of Internet Traffic // Proceedings of IEEE International Conference on Communication, 2004. Vol. 4 – P. 2143–2147.
2. **Vaton S.** A new test of stationarity and its application to teletraffic data // Proceedings of IEEE International Conference USA, 1998. – Vol. 6 – P. 3449–3452.
3. **Meuleau N., Kim K. E., Kaelbling L. P., Cassandra A.**.Solvinmg POMDPs by searching the space of finite policies // In Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence, 1999. – P. 417–426.
4. **Cappuccio N., Lubian D.** The Fragility of the KPSS Stationarity Test // Statistical Methods and Applications – Springer, 2010. – Vol. 19(2). – P. 237–253.
5. **Dang D. T., Molnar S.** On the effects of non–stationarity in long–range dependence tests // Periodica Polytechnica Ser. El. Eng., 1999. – Vol. 43. – P. 227–250.
6. **Ming Li, Yun–Yun Zhang, Wei Zhao**. A Practical Method for Weak Stationarity Test of Network with Long_Range Dependence // International Journal of mathematics and computers in simulation, 2007. – Vol. 1. – Iss. 4. – P. 307–311.
7. **Kwiatkowski D., Phillips P. C. B., Schmidt P., Shin Y.** Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root: How sure are we that economic time series have a unit root? // Journal of Econometrics, – Elsevier, 1992. – Vol. 54(13). – P. 159–178.

Received 2011 02 15

G. Lauks, A. Skrastins, J. Jelinskis. Testing the Null Hypothesis of Stationarity of Internet Traffic // Electronics and Electrical Engineering. – Kaunas: Technologija, 2011. – No. 6(112). – P. 29–32.

The statistical analyses indicate that the measured traffic traces from the packet networks often contain non-stationary effects. In these cases several popular tests for long-range dependence and/or stationarity can result in wrong conclusions and unreliable estimate of the Hurst parameter. In this paper non-stationarities are modeled as the trends and/or level shifts in Internet traffic data. MMPP-Based Hierarchical Model simulation data are used for stationarity tests. Application of testing results are integrated into network resource allocation function as a Partially Observable Markov Decision process. Ill. 4, bibl. 7, tabl. 1 (in English; abstracts in English and Lithuanian).

G. Lauks, A. Skrastins, J. Jelinskis. Nulinės hipotezės tyrimas pastoviuose interneto srautuose // Elektronika ir elektrotechnika. – Kaunas: Technologija, 2011. – Nr. 6(112). – P. 29–32.

Duomenų srautų tinkle statistinė analizė rodo, kad duomenų paketai dažnai yra veikiami tam tikrų nepastovių veiksnių. Būtina atsižvelgti į Hursto parametrą. Darbe atliktas nepastovių veiksnių įtakos interneto duomenų srautams modeliavimas. Taikant hierarchinį modelį, atlikti keli pastovumo testai. Testavimo rezultatai buvo integruoti į duomenų perdavimo tinklą. Il. 4, bibl. 7, lent. 1 (anglų kalba; santraukos anglų ir lietuvių k.).