Journal of Southeast University (English Edition) Vol. 20 No. 4 Dec. 2004 ISSN 1003—7985

On approximating multifractal traffic burstiness with Markov modulated Poisson processes

Ji Qijin

(Key Laboratory of Computer Network and Information Integration of Ministry of Education, Southeast University, Nanjing 210096, China)

Abstract: We investigate the approximating capability of Markov modulated Poisson processes (MMPP) for modeling multifractal Internet traffic. The choice of MMPP is motivated by its ability to capture the variability and correlation in moderate time scales while being analytically tractable. Important statistics of traffic burstiness are described and a customized moment-based fitting procedure of MMPP to traffic traces is presented. Our methodology of doing this is to examine whether the MMPP can be used to predict the performance of a queue to which MMPP sample paths and measured traffic traces are fed for comparison respectively, in addition to the goodness-of-fit test of MMPP. Numerical results and simulations show that the fitted MMPP can approximate multifractal traffic quite well, i.e. accurately predict the queueing performance.

Key words: multifractal traffic; Markov modulated Poisson processes; queueing delay; packet loss rate

Since the seminal work of Leland, et al. [1], there has been numerous evidence solidly confirming that traffic in packet communication networks, particularly the Internet, is better characterized by statistical self-similar or fractal-like processes which present surprising long-range dependence (LRD) in second-order statistics (see Ref. [2] and references therein). Self-similarity is appealing due to its concise characterization of large-time scale burstiness in network traffic. However, when looking into finer time scales, from a few hundreds milliseconds downwards, there are more complex singularities beyond self-similarity. A multifractal model, such as multifractal cascades[3,4] has been proposed to capture the burstiness of these small time scale behaviors. Self-similarity in large time scales and multiscaling in small time scales provide a complete description of packet network traffic[5,6].

A number of performance studies[7-9] have shown that self-similarity can have a detrimental effect on network performance leading to increased queueing delay and packet loss rate. Fractional Brownian motion (FBM) is considered proper for performance analysis of self-similar traffic[7]. Although Internet traffic is statistically characterized by LRD with sound evidences, for multimedia applications, whose QoS requirement, such as packet delay and loss rate is very stringent, only correlations under critical time scale[10] or correlation horizontal[11] are relevant to the queueing performance due to the reset effect[12]. This insight makes modeling long-range dependence traffic with short-range dependence processes possible.

Performance impact of multifractal traffic characteristics was investigated in Ref. [13], which showed that the fine timescale features can affect performance substantially at low and intermediate link utilizations, while the coarse timescale self-similarity is important only at intermediate and high utilizations. This discovery further confirms that burstiness in small time scales is dominant when a small buffer/large bandwidth resource provisioning policy[9] is employed, yet till now no analytical model has been suggested and examined for multifractal scaling behavior of network traffic.

Performance analysis based on an analytical model is flexible and inexpensive compared with simulations. A Markov modulated Poisson process (MMPP) is a versatile traffic model that has been extensively studied in Refs. [14 - 17]. While it is powerful in modeling several types of traffic, it is analytically tractable with matrix geometric method, or more recently the invariant subspace approach[18]. In this paper we investigate whether multifractal traffic burstiness can be approximated with MMPP. The choice of MMPP is motivated by its ability to capture both the variability and correlations in moderate time scales. Particularly, we emphasize two significant statistics, the index of dispersion for counts (IDC)[15,19] and coefficient of skewness[20] for modeling burstiness in different time scales. We compare the mean queue length and the survival

Received 2004-04-26.

Foundation items: The National Natural Science Foundation of China (No. 90104009), the National Bacic Research Program of China (973 Program) (No. 2003CB314801).

Biography: Ji Qijin(1974—), male, graduate, andyji@seu.edu.cn.

function or packet loss rate (PLR) in finite buffer situations of the queueing model between real traffic trace and MMPP samples via extensive simulations. Numerical results show that MMPP is an accurate performance model for characterizing multifractal traffic burstiness in the interesting time scales.

1 Multifractal Scaling of Network Traffic

Let $A(t)$, $0 \leqslant t < \infty$ denote the traffic arrival process representing the total number of packets or bytes sent over a link up to time t, then the traffic rate process is the increment process of $A(t)$, which represents the number of packets or bytes in a time interval, e.g. $[t_0, t_0 + \delta(t)]$. The traffic is said to be local scaling with exponent $\alpha(t_0)$ if the traffic rate process behaves like $(\delta(t))^{\alpha(t_0)}$ as $\delta(t) \to 0$. Informally, signals with $\alpha(t_0) = H$ at all instants t_0 are called monofractal (including exactly self-similar processes such as FBM) while signals with non-constant scaling exponent $\alpha(t_0)$ are called multifractal. While self-similarity or mono-fractality describes the consistent scaling behavior in a global time scale, multifractality focuses on the local time irregularity of a signal. Multifractal processes have two properties: spiky increment processes and non-Gaussian marginals that are described with multifractal spectra and partition function respectively[4]. For packet networks, in which the traffic is self-similar in the time scales from hundreds of milliseconds upwords, locally multifractal behavior of network traffic means extreme burstiness in small time scales.

The Auckland Ⅱ data traces[21], which are collected on a link at the University of Auckland by WAND group, are examined in this paper. Since multifractality of these traces has been checked in several previous works[6,22], we omit the multifractal test and scaling analysis here.

2 Modeling Multifractal Burstiness with MMPP

Multifractal traffic is very bursty due to both correlations in wide range time scales and variability in local time. Consider a discrete time increment process of the traffic arrival process $A(t)$, a wide-sense stationary, non-negative random sequence, $\{X_i\}_{i=0}^{\infty}$ as a traffic rate process, and X_i as the packet number or bytes in the i-th time slot. In this section we first propose an index of dispersion for counts and third-order moments or related coefficient of skewness as burstiness measures of network traffic, we then present a fitting procedure of MMPP to match these measures of real traffic traces.

2.1 Measures of traffic burstiness

Burstiness has been recognized as a key characteristic of broadband traffic that plays critical roles in determining network performance, such as queueing length distribution and packet loss rate. Two regimes of scaling are discovered in Internet traffic. Fractal scaling, like self-similarity, means LRD, i.e. correlation in large time scales, while multifractal scaling results in extreme variability in local time with non-Gaussian distribution. To characterize the traffic burstiness, both the first order and the second order statistics are indispensable.

A simple class of burstiness measures takes only the first order properties into account. These measures can be considered as different characteristics of the marginal distribution of the arrival processes. Geist, et al.[20] demonstrated that ignoring the distribution character would result in optimistic estimation of queueing performance. The third order moments, or the normalized version, coefficient of skewness is an efficient candidate for non-Gaussian distribution. An estimation of the third order central moment is $\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^3$, and the coefficient of skewness is the third order moment normalized by σ^3. Here μ and σ are the mean and standard variance of X_i. For Gaussian process, the coefficient of skewness is expected to be zero.

Correlation structure of the input processes is another significant factor influencing queueing performance. Measures expressing second order properties of the traffic are more complex. We consider the aggregation variance or its normalized version, the index of dispersion, as a useful tool for characterizing the correlation properties of the traffic. For traffic arrival process $A(t)$, the index of dispersion for counts at time t is the variance of the number of arrivals in the time divided by the mean number of arrivals in the same time[19]:

$$\mathrm{IDC}(t) = \frac{\mathrm{var}[A(t)]}{E[A(t)]}$$

Assume the arrival process is denoted as the previous random sequence $\{X_i\}_{i=0}^{\infty}$, then $A(t)$ is the sum of n sequential random variables, the variance of which is given by

$$\mathrm{var}[A(t)] = \mathrm{var}(X_{i+1} + \cdots + X_{i+n}) =$$

$$n\mathrm{var}(X) + 2\sum_{j=1}^{n-1}\sum_{k=1}^{j} \mathrm{cov}(X_j, X_{j+k})$$

where $\mathrm{cov}(X_j, X_{j+k})$ is the autocorrelation coefficient of X_j at lag k. The contribution to aggregation variance of correlations in the arrival process is clearly shown in the above expressions.

2.2 Fitting procedure of the MMPP model

MMPP is a doubly stochastic process where the intensity of a Poisson process is defined by the state of a continuous-time Markov chain. This modulation of Markov chain introduces correlations between successive inter-arrival times in the process. The main advantage that springs from using MMPPs as traffic models is that they lend themselves to analysis better than some alternatives. Important properties of queueing systems like MMPP/G/1 can, albeit not very easily, be derived. For a two state MMPP (MMPP-2), when the underlying chain is in state j ($j = 1, 2$), the arrival process is a Poisson process with rate λ_j and mean sojourn time r_j^{-1}.

The fitting procedure of MMPP-2 is to obtain the four parameters (λ_1, λ_2, r_1 and r_2) according to the information of given samples, which is stationary, to be modeled. Several fitting procedures of MMPP have been proposed in Refs. [14 − 17]. We consider two special requirements while choosing one of them for multifractal traffic here: ① The fitted MMPP can capture the most important statistics of the traffic: mean, second order moments and third order moments. ② Feasibility. Some fitting procedures try to characterize the statistics of the inter-arrival time. Although Auckland Ⅱ is collected with enough precise time resolution, these procedures are infeasible in most other cases because of time resolution limitation.

A fitting procedure based on the moments of the increment of counting process was presented in Ref. [14], which can be outlined as follows:

1) Matching the mean arrival rate;

2) Matching IDC of the arrival process in $(0, t_1)$;

3) Matching the "limit" of IDC of the arrival process, i.e. IDC(t) as $t\to\infty$;

4) Matching the third order moment of the arrival process in $(0, t_2)$.

Since MMPP-2 is a special case of versatile Markov point process, its moments can be derived analytically and expressed with the four parameters of MMPP. With estimated statistics and derived ones in hand we can obtain the four parameters of MMPP by solving the four equations. Please refer to Ref. [14] for details.

Although our fitting procedure is similar to the above one, more issues have to be considered. Firstly, while the moments of arrival process in Ref. [14] are analytical based on renewal theory, we have to estimate the moments for our traffic traces. Secondly, three time scales in the fitting procedure should be chosen properly for our modeling interest. We set t_1 for a good fitting of IDC(t). The IDC(t) will keep constant as t approaches infinity. In practice, infinite time is impossible. What's more, the fitted trace is LRD and IDC will increase monotonically for a very large time scale. We choose a large enough time scale denoted as t_∞ with practical QoS requirements under consideration. Time scale of t_2 is chosen for matching burstiness in a small time scale well. Fortunately, t_2 is insensitive to IDC[14].

3 Numerical Results

In this section we check the approximating capacity of MMPP to multifractal traffic by comparing the queueing performance when the queue is loaded with Auckland Ⅱ traffic trace and the fitted MMPP samples respectively.

3.1 Goodness of fit tests

We get a set of parameters of MMPP from the fitting procedure in section 2, which is λ_1 = 1 536.443 8, λ_2 = 319.804 6, r_1 = 5.012 4 and r_2 = 0.044 80. A powerful goodness-of-fit test is the Q-Q plot, which plots the quantiles of the trace data versus the quantiles of the fitted distribution. The Q-Q plot of Auckland Ⅱ trace and the fitted MMPP are shown in Fig.1. The Q-Q plot shows that the fit is very good except for the right-hand tail, where the fitted MMPP has too little probability. This means that the fitted

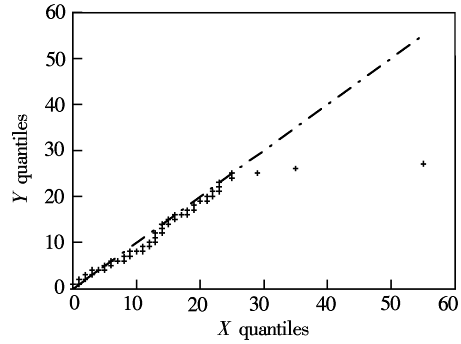

Fig.1 Q-Q plot of Auckland Ⅱ trace and fitted MMPP

MMPP is less bursty than the real trace in the time scale of the fixed interval of the arrival process.

Matching IDC of the multifractal traffic in proper time scale is key to the MMPP model. The matched IDC is influenced by the choice of two time scales, i.e. t_1 and the time corresponding to the limit of IDC, which levels off with time increasing. Fig. 2 demonstrates the estimated IDC of Auckland Ⅱ trace and two fitted IDCs with different time scales. We find from Fig.2 that while the IDC curve fitted with $t_\infty = 2$ s is accurate in time period between 500 ms and 1 s, the IDC curve fitted with $t_\infty = 1$ s is more accurate in a small time scale of less than 500 ms. For moderate link utilization, which is interesting in practice, our succeeding fitting procedure is based on the later. We set $t_2 = 10$ ms to capture the burstiness in a small time scale.

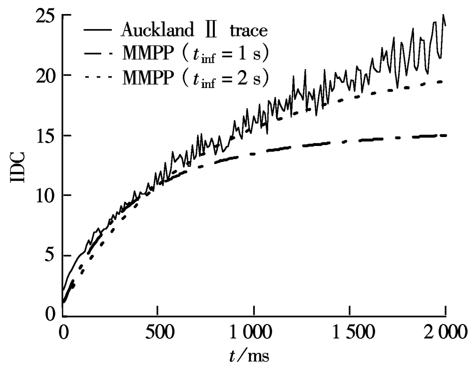

Fig.2 IDC curves of the Auckland Ⅱ trace and fitted MMPP

3.2 Performance simulations

The performance measures, especially queueing delay and packet loss rate, are obtained by simulation with NS2[23]. A simple simulation model is illustrated in Fig.3 (The unit of bandwidth is bit/s). We focus on the queueing performance at the bottleneck link. Although MMPP/G/1 is an analytically tractable queueing model, we generate samples of MMPP and also obtain the queueing statistics by simulations. The mean arrival rate of the trace is about 1.324 Mbit/s and we set different bottleneck capacities according to the required link utilizations.

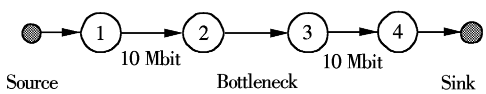

Fig.3 Simulation model

The queueing delay is the most significant component of the end-to-end packet delay, which is stringent for real time applications such as streaming media. In practice, end-to-end delay allowing for multimedia applications is about 200 ms or so and queueing delay in a single node of a path is at most 20 to 30 ms, which means a very limited queueing buffer size. Fig.4 shows the mean queueing length vs. link utilization obtained in simulation.

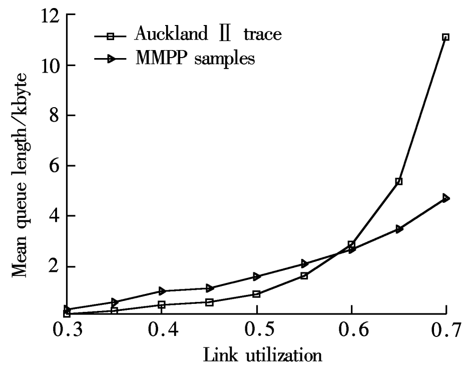

Fig.4 Mean queue length vs. link utilization

From Fig.4 we can see that mean queue length from MMPP sample paths is very close to that generated from the measured traffic trace when the link utilization is less than 65%, resulting in a mean queueing delay of less than 24 ms. When link utilization reaches 70%, the deviation of mean queue length resulting from the traffic trace and MMPP model is large, but in this situation, the queueing delay of the traffic trace simulation is about 50 ms, which is unacceptable in practice. We also see that estimation of mean queueing delay with MMPP is conservative as the link utilization is less than 60%.

End-to-end packet loss rate is another interesting performance measure for a network path. The complementary queue length distribution or survival function of the queueing model is a reasonable approximation of PLR. Some simulation results of complementary queue length distribution in log scale are presented in Fig. 5. From Fig. 5 we find similar results like the mean queue length situation, i.e. the MMPP model can capture the PLR of multifractal traffic trace fairly accurately when the link utilization is less than 65%. Again, MMPPs underestimate the PLR at high link utilization (>65%). An interesting crossover is shown in the figure. When the buffer size is very small, MMPPs underestimate the PLR but they overestimate the PLR when the buffer size is moderate. When the buffer size is large, MMPPs result in comparable PLR with the traffic trace. We conjecture that the crossover both in mean queue length and PLR is caused by the simple correlation structure in a single time scale of MMPP. In actual traffic, the correlation is weaker, even independent for smaller time scales, but stronger in larger time scales.

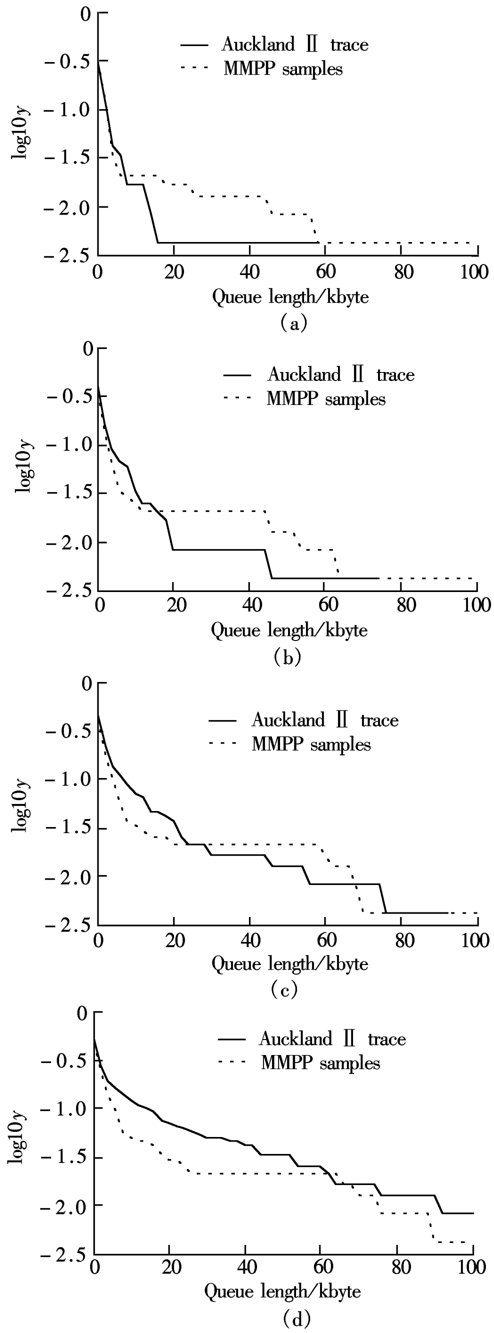

Fig. 5 Complementary queue length distributions (PLR) with different link utilizations. (a) Link utilization is 0.50; (b) Link utilization is 0.55; (c) Link utilization is 0.60; (d) Link utilization is 0.65

4 Conclusion

We have examined whether MMPP, a versatile traffic model extensively investigated in literatures, can also be employed for modeling multifractal traffic, e.g. Auckland Ⅱ. We do not try to capture the exact statistical character of multifractal traffic such as multifractal spectrum with MMPP, but the most important statistics that influence queueing performance, concretely; IDC for correlation of packet inter-arrivals and coefficient of skewness for non-Gaussinity in small time scales. We present a procedure for estimating the parameter of MMPP-2 customized for our modeling objectives. The simulation results show that in practically interesting time scales, MMPP is competent for modeling the multifractal traffic, i.e. predicting queueing performance loaded with such traffic, which bring us an analytically tractable model for multifractal network traffic. Extensions of this work to other traffic traces presenting multifractal behavior are under investigation and the refinement of time scale choices in the fitting procedure is left for future research.

Acknowledgments The author would like to particularly thank professor Gu Guanqun for his supervision and the anonymous reviewers for their suggestions on the earlier draft of this paper.

References

[1] Leland W E, Taqqu M S, Willinger W, et al. On the self-similar nature of Ethernet traffic (extended version) [J]. *IEEE/ACM Trans on Networking*, **1994, 2**(1): 1 - 16.

[2] Cappe O, Moulines E, Pesquet J C, et al. Long-range dependence and heavy-tail modeling for teletraffic data [J]. *IEEE Signal Processing Magazine*, **2002, 19**(3): 14 - 27.

[3] Feldmann A, Gilbert A C, Willinger W. Data networks as cascades: investigating the multifractal nature of Internet WAN traffic [A]. In: *Proc ACM SIGCOMM'98* [C]. Vancouver, Canada, 1998. 42 - 66.

[4] Riedi R H, Crouse M S, Ribeiro V, et al. A multifractal wavelet model with application to network traffic [J]. *IEEE Trans on Information Theory*, **1999, 46**(3): 992 - 1018.

[5] Riedi R H, Willinger W. Toward an improved understanding of network traffic dynamics [A]. In: Park K, Willinger W, eds. *Self-Similar Network Traffic and Performance Evaluation* [C]. New York: Wiley, 2000.507 - 530.

[6] Abry P, Baraniuk R, Flandrin P, et al. Multiscale nature of network traffic [J]. *IEEE Signal Processing Magazine*, **2002, 19**(3): 28 - 46.

[7] Norros I. A storage model with self-similar input [J]. *Queueing Systems*, **1994, 16**(2): 382 - 396.

[8] Erramilli A, Narayan O, Willinger W. Experimental queueing analysis with long-range dependent packet traffic [J]. *IEEE/ACM Trans on Networking*, **1996, 4**(2): 209 - 223.

[9] Park K, Kim G, Crovella M. On the effect of traffic self-similarity on network performance [A]. In: *Proc SPIE Int'l Conf Perf and Control of Network System* [C]. Dallas, USA, 1997. 296 - 310.

[10] Grossglauser M, Bolot J C. On the relevance of long-

range dependence in network traffic [A]. In: *Proc ACM SIGCOM'96* [C]. Stanford University, 1996. 15 - 24.

[11] Ryu B K, Elwalid A, The importance of long-range dependence of VBR video traffic in ATM traffic engineering: myths and realities [A]. In: *Proc ACM SIGCOM'96* [C]. Stanford University, 1996. 3 - 14.

[12] Heyman D P, Lakshman T V. What are the implications of long-range dependence for VBR-video traffic engineering? [J]. *IEEE/ACM Trans on Networking*, **1996, 4**(3): 301 - 317.

[13] Erramilli A, Narayan O, Neidhardt A, et al. Performance impacts of multi-scaling in wide area TCP/IP traffic [A]. In: *Proc of IEEE INFOCOM'00* [C]. **Israel, 2000, 1**: 362 - 369.

[14] Heffes H, Lucantoni D. A Markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance [J]. *IEEE Journal on Selected Areas in Comm*, **1986, 4**(6): 856 - 868.

[15] Gusella R. Characterizing the variability of arrival processes with indexes of dispersion [J]. *IEEE Journal on Selected Areas in Comm*, **1991, 9**(2): 203 - 211.

[16] Fischer W, Meier-Hellstern K S. The Markov-modulated Poisson process (MMPP) cookbook [J]. *Performance Evaluation*, **1993, 18**(2): 149 - 171.

[17] Kang S H, Kim Y H, Sung D K, et al. An application of Markovian arrival process (MAP) to modeling superposed ATM cell streams [J]. *IEEE Trans on Communications*, **2002, 50**(4): 633 - 642.

[18] Akar N, Oguz N C, Sohraby K. TELPACK: an advanced teletraffic analysis package [J]. *IEEE Comm Magazine*, **1998, 36**(8): 84 - 87.

[19] Cox D R, Lewis P A. *Statistical analysis of series of events* [M]. London: Methuen, 1966.

[20] Geist R M, Westall J M. Correlational and distributional effects in network traffic models [J]. *Performance Evaluation*, **2001, 44**(1): 121 - 138.

[21] Auckland Ⅱ trace 20000125-143640 [EB/OL].http: //pma.nlanr.net/Traces/long/auck2.html. 2004-04-26.

[22] Ribeiro V J, Riedi R H, Crouse M S, et al. Multiscale queueing analysis of long-range-dependent network traffic [A]. In: *Proc IEEE INFOCOM'00* [C]. Israel, **2000, 2**: 1026 - 1035.

[23] NS2, network simulation 2 [EB/OL].http: //www.isi.edu/nsnam/ns/. 2004-04-26.

用马尔可夫调制的泊松过程近似多分形突发流量研究

纪其进

(东南大学计算机网络和信息集成技术教育部重点实验室, 南京 210096)

摘要: 研究了用马尔可夫调制的泊松过程(MMPP)对 Internet 多分形流量突发行为进行近似建模的能力. MMPP 可用于描述适当时间尺度范围内流量的变化以及相关性, 而且它可作为排队系统输入过程得到分析结果. 描述了刻画突发流量行为的重要统计量, 在此基础上给出了一个基于矩的 MMPP 参数估计方法. 除了对 MMPP 进行拟合优度检测以外, 本文通过将 MMPP 的样本过程和实际流量记录输入到排队系统模型中比较其输出结果来研究 MMPP 对排队性能的预测能力. 数值和仿真实验表明, MMPP 能够较好地用于对多分形流量近似建模, 即可以准确地预测网络结点的排队性能.

关键词: 多分形流量; 马尔可夫调制的泊松过程(MMPP); 排队时延; 分组丢失率

中图分类号: TP393