EPJ manuscript No.
(will be inserted by the editor)

A model to study the scaling of traffic fluctuations on complex networks

Jordi Duch[1] and Alex Arenas[2]

[1] Universitat Rovira i Virgili, Spain, e-mail: jordi.duch@urv.cat
[2] Universitat Rovira i Virgili, Spain, e-mail: alexandre.arenas@urv.cat

Received: date / Revised version: date

Abstract. In a recent article, we studied the dynamics of traffic in complex networks [1]. In particular, we computed how the fluctuations scale with the mean, $\sigma \sim \langle f \rangle^{\alpha}$. Using a general model which includes nodes with finite capacity we found a continuous range of α values between 1/2 and 1. Here we resume the results, adding a brief analysis about the self-similarity of the traffic dynamics in our model.

PACS. 89.75.Hc Networks and genealogical trees – 89.75.Da Systems obeying scaling laws – 89.20.Hh World Wide Web, Internet

1 Introduction

The study of the dynamic processes over complex networks have emerged recently, complementing the extensive work about topological properties. Most of the work has been focused in the traffic flow, determining the bounds for this flow to become congested [4] or the origin of the traffic fluctuations [2].

Menezes and Barabási [2] propose a model to understand the origin of fluctuations in traffic, relating the average number of packets $\langle f \rangle_i$ processed by nodes during a certain time interval, and the standard deviation σ_i of this quantity. They find that there is an scaling relation $\sigma \sim \langle f \rangle^{\alpha}$, and propose two universal values of α depending on the characteristics of the dynamic process.

Using a more general framework [1] we show that the value of this exponent depends on some extra factors not considered previously, as the sampling process or the effects of limited capacity processing, obtaining a continuous range of values between $\alpha = 1/2$ and $\alpha = 1$. Experimental results also confirm that there are not two universal classes for the α exponent. For instance, the Abilene network (part of the Internet 2 core) has values between 0.7 and 0.85 [1], and the NYSE has values between 0.77 and 0.88 [3].

In this article we review the model and the analysis of some effects over the relation between the mean and the fluctuations. We also show that our model in its present form is unable to reproduce the self-similarity of traffic in time reported in many real systems. The reason is that we use a Poisson distribution for the injection of packets in the network that does not account for a bursting behavior of traffic, in accordance with the literature in computer science about this same problem in the Internet traffic [5, 6, ?].

2 The model

The underlying network topology used in this analysis is a BA Scale-Free with $N = 1000$ nodes and $\gamma = 3$, where each node has an associated M/M/1 queue. The dynamic process works in continuous time with the following rules: Data packets enter the system at a randomly selected node following a Poisson distribution with parameter λ. The time needed to process one packet in each node is controlled by an exponential distribution with parameter μ. Every packet performs S random steps, and then is deleted from the network.

Each individual node has an arrival ratio controlled by a Poisson distribution with parameter $\lambda_i^{ef} = B_i \lambda$ where B_i is the algorithmic betweenness of node i (the relative number of paths in the network that go through node i given a specific routing algorithm) [3]. The system starts collapsing if $\lambda_*^{ef} > \mu$ (being B_* the highest algorithmic betweenness).

Every P time units we collect data of how many packets had passed through each node.

3 Transitions of the exponent

In [1] we examined the effects of three parameters in the relation between $\sigma \sim \langle f \rangle^{\alpha}$, observing different transitions of the α exponent.

The first transition appears changing the size of the time window P used to gather the data from the nodes. Selecting a value of $P \ll 1/\lambda_*^{ef}$, we always observe the scaling $\sigma \sim \langle f \rangle^{1/2}$, regardless of other parameters. This scaling exponent is easy to understand. In this situation, the nodes will deliver either one packet or none, at each time interval. Therefore the average and the standard deviation can be expressed as:

$$\langle f \rangle = n_1/n \tag{1}$$

$$\sigma = [\frac{1}{n}\left[n_1(1-\langle f \rangle)^2 + n_0\langle f \rangle^2\right]]^{1/2} \tag{2}$$

where n_1 is the number of intervals a packet is delivered and n_0 the intervals where no packet is delivered. Eq. 2 can be simplified to

$$\sigma = [(1-\langle f \rangle)\langle f \rangle]^{1/2} \tag{3}$$

So in the case where the average flow is $\langle f \rangle \ll 1$ we always recover the $\sigma \sim \langle f \rangle^{1/2}$ scaling law. Otherwise, the exponent value depends on the other parameters.

Assuming that the sampling of the data is performed at intervals of length $P \ll 1/\lambda_*^{ef}$, we can produce a second transition between $\alpha = 1/2$ and $\alpha = 1$ changing the behavior of the traffic in the system. In this case, the total traffic T (number of packets flowing through the network per unit time) is determined by a Poisson process with mean $\langle T \rangle = \lambda S$. Keeping the total traffic mean $\langle T \rangle$ fixed, we can control the variability (fluctuations) of the incoming traffic to a node by varying the values of λ and S proportionally.

The transition is obtained here simply by increasing the number of steps S the packet performs on the network while maintaining the mean value of the total traffic (decreasing the number of new packets λ). The $\alpha = 1/2$ turns up because the number of steps each packet performs is small, acting like a random deposition and being independent of the topology of the network, $\lambda_i^{ef} \sim \lambda$. Otherwise, when the number of steps in the network grows, the topology induces dynamical correlations that affect the scaling of fluctuations via the algorithmic betweenness, $\lambda_i^{ef} \sim \lambda B_i$, producing the transition to $\alpha = 1$, and therefore reproducing the results obtained in [2].

Finally we extend the simple model adding persistent queues controlled by the parameter μ (rate of service) and observing the interactions between packets. When congestion occurs, the queues corresponding to those nodes with the highest B_i will always have more packets than those that can be delivered in a period P. This means that the number of packets delivered by these nodes will be controlled exclusively by the service rate and will be again fitted by $\alpha = 1/2$ (corresponding to the exponential service distribution), independently of the other parameters.

4 Self-similarity of traffic in the model

The next step when using the proposed model to analyze traffic consist into respond to the following question: Is the proposed traffic model able to reproduce the self-similarity of traffic in time observed in some real systems, as for example the Internet? It has been discussed that Poisson models aren't realistic [5] because do not reproduce some characteristics of the real dynamics like 'burstinness' that Internet exhibit. Other authors [7] still defend that in certain cases, Internet traffic can still be modeled using Poisson models when we are near the edge of congestion. To test the validity of the simple model used in the simulations, we have studied one statistical property of the dynamics, the self-similarity of the traffic that flows in the model.

There are several methods to determine the self-similarity of traffic flow. Here we will use the relation between the time and the variability of the traffic, $\sigma \sim P^\beta$. It is expected that in Poisson models we will find a decay with exponent $\beta = -1$, due to the lack of autocorrelations in the traffic [6], as we find in the analysis of our model. Instead, real Internet router traffic data used in [1] exhibit values between $\beta = -0.4$ and $\beta = -0.1$. See Figure 1.

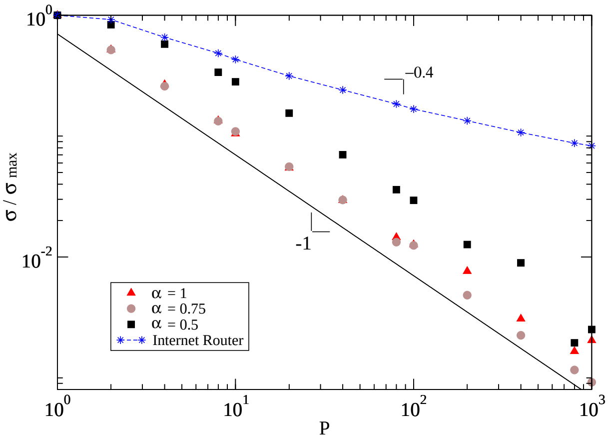

Fig. 1. Plot σ versus P. The symbols corresponds to different realizations of our model varying the parameters. The dotted line corresponds to the analysis of the Washington Abilene router.

We conclude that our model in its present form can not reproduce the self-similarity expected, independently on the parameters that control the scaling of fluctuations of the mean of traffic. We guess that the self-similarity will be reproduced if the injection of packets into the system follows a heavy-tailed distribution instead of a Poisson distribution, however we can still not prove this conjecture.

References

1. J. Duch and A. Arenas, Phys. Rev. Lett. **96**, (2006) 218702.
2. M. A. de Menezes and A.-L. Barabási, Phys. Rev. Lett. **92**, (2004) 028701.
3. Z. Eisler, J. Kertész, S.-H. Yook and A.-L. Barabási, Europhys. Lett. **69**, (2005) 664-670.
4. R. Guimerà, A. Díaz-Guilera, F. Vega-Redondo, A. Cabrales, and A. Arenas, Phys. Rev. Lett. **89**, (2002) 248701.
5. W.-E. Leland, M.-S. Taquu, W. Willinger and D.-V. Wilson, ACM/SIGCOMM Comp. Comm. Rev. **23**, (1993) 183-193.
6. V. Paxson and S. Floyd, Networking, IEEE/ACM Transactions on, **3**, (1995) 226-244.
7. T. Karagiannis, M. Molle, M. Faloutsos and A. Broido, INFOCOM 2004 Proceedings, **3**,(2004) 1558-1569.

This work has been supported by DGES of the Spanish Government Grant No. FIS2006-13321-C02-02.