ELSEVIER

Physica A 289 (2001) 595–605

PHYSICA A

www.elsevier.com/locate/physa

Information transfer and phase transitions in a model of internet traffic

Ricard V. Solé[a,b,*], Sergi Valverde[a]

[a] Complex Systems Research Group, Departament of Physics, FEN Universitat Politècnica de Catalunya, Campus Nord B4, 08034 Barcelona, Spain

[b] Santa Fe Institute, 1399 Hyde Park Road, New Mexico 87501, USA

Received 4 October 2000

Abstract

In a recent study, Ohira and Sawatari presented a simple model of computer network traffic dynamics. These authors showed that a phase transition point is present separating the low-traffic phase with no congestion from the congestion phase as the packet creation rate increases. We further investigated this model by relaxing the network topology using a random location of routers. It is shown that the model exhibits nontrivial scaling properties close to the critical point, which reproduce some of the observed real Internet features. At criticality, the net shows maximum information transfer and efficiency. It is shown that some of the key properties of this model are shared by highway traffic models, as previously conjectured by some authors. The relevance to Internet dynamics and to the performance of parallel arrays of processors is discussed. © 2001 Published by Elsevier Science B.V. All rights reserved.

1. Introduction

The exchange of information in complex networks and how these networks evolve in time has been receiving increasing attention by physicists over the last few years [1,2]. Two main lines of research have been developed: (a) the analysis of the structural properties displayed by traffic networks [3] and (b) the analysis of dynamical patterns of information exchange. Recent studies have revealed that phase transition phenomena arises in Internet traffic and are allowed to quantitative analysis by means of appropriate tools from statistical physics [4,5].

The WWW is a virtual graph connecting nodes containing different amounts of information. This information flows through a physical support which also displays

* Corresponding author. Fax: +34-93-4017100.

E-mail address: ricard@complex.upc.es (R.V. Solé).

0378-4371/01/$ - see front matter
PII: S0378-4371(00)00536-7

scale-free behavior [3]. The network of computers is itself a complex system, and complex dynamics has been detected suggesting that self-similar patterns are also at work [6].

Some previous studies have shown evidence for critical-like dynamics in Computer Networks [7] in terms of fractal, $1/f$ noise spectrum as well as long-tail distributions of some characteristic quantities. Some authors have even speculated about the possibility that the traffic of information through computer networks (such as Internet) can display the critical features already reported in cellular automata models of traffic flow, such as the Nagel–Schreckenberg (NS) model [8–10]. The NS model shows that as one increases the density of cars ρ, a well-defined transition occurs at a critical density ρ_c. This transition separates a fluid phase showing no jams from the jammed phase were traffic jams emerge. At the critical boundary, the first jams are observed as back-propagating waves with fractal properties.

A number of both quantitative and qualitative observations of real computer network dynamics reveals some features of interest:

1. Extensive data mining from Internet/Ethernet traffic shows that it displays long-range correlations [6] with well-defined persistence, as measured by means of the Hurst exponent. This analysis totally rejects the previous theoretical approach to Poisson-based (Markovian) models assuming statistical independence of the arrival process of information.
2. Fluctuations in density of packets show well-defined self-similar behavior over long time scales. This has been measured by several authors [7,11]. The power spectrum is typically a power law, although local (spatial) differences have been shown to be involved.
3. The statistical properties of Internet congestion reveal long-tailed (lognormal) distributions of latencies. Here latency times T_L are thus given by

$$P(T_L) = \frac{1}{T_L \sigma \sqrt{2\pi}} \exp\left(- \frac{\ln T_L}{2\sigma^2} \right) .$$

   Latencies are measured by performing series of experiments in which the round-trip times of ping packets is averaged over many sent messages between two given nodes.
4. There is a clear feedback between the bottom-level where users send their messages through the net and increase network activity (and congestion) and the top-level described by the overall network activity. Users are responsible for the global behavior (since packets are generated by users) and the later modifies the individual decisions (users will tend to leave the net if it becomes too congested).

On the other hand, previous studies on highway traffic dynamics revealed that the phase transition point presented by the models as the density of cars increased was linked with a high degree of unpredictability [12]. Interestingly, this is maximum at criticality [13] as well as the flow rate. In other words, efficiency and unpredictability are connected by the phase transition. In this paper, the previous conjecture linking

Internet dynamics with critical points in highway traffic is further explored. By considering a generalization of the Ohira–Sawatari (OS) model, we show that *all* the previously reported features of real traffic dynamics are recovered by the model. The paper is organized as follows. In Section 2, the basic model and its phase transition is presented. In Section 3 the self-similar character of the time dynamics is shown by means of the calculation of the latency times and queue distributions as well as by means of spectral and Hurst analysis. In Section 4, the efficiency and information transfer are calculated for different network sizes. In Section 5 our main conclusions and a discussion of its implications is presented.

2. Model of computer network traffic

Following the work by Ohira and Sawatari, let us consider a two-dimensional network with a square lattice topology with four nearest neighbors [14]. The network involved two types of nodes: *hosts* and *routers*. The first are nodes that can generate and receive messages and the second can only store and forward messages. Our square, $L \times L$ lattice will be indicated as $\mathscr{L}(L)$, following previous notation [15]. All our simulations are performed using periodic boundary conditions. In previous papers, either the hosts were distributed through the boundary [14] (and thus the inner nodes were routers) or *all* nodes were both hosts and routers [15]. Here we consider a more realistic situation, where only a fraction ρ of the nodes are hosts and the rest are routers [14] (Fig. 1).

The location of each object, $\mathbf{r} \in \mathscr{L}(L)$, will be given by $\mathbf{r} = i\mathbf{c}_x + j\mathbf{c}_y$, where $\mathbf{c}_x$, $\mathbf{c}_y$ are Cartesian unit vectors. So the set of nearest neighbors $C(\mathbf{r})$ is given by

$$C(\mathbf{r}) = \{\mathbf{r} - \mathbf{c}_x, \mathbf{r} + \mathbf{c}_x, \mathbf{r} - \mathbf{c}_y, \mathbf{r} + \mathbf{c}_y\} . \quad (1)$$

Each node maintains a queue of unlimited length where the packets arriving are stored. The local number of packets will be indicated as $n(\mathbf{r}, t)$ and thus the total number of packets in the system will be

$$N(t) = \sum_{\mathbf{r} \in \mathscr{L}(L)} n(\mathbf{r}, t) . \quad (2)$$

The rules are defined as in the OS deterministic model (the stochastic version only shows the differences already reported by these authors [14]). The rules are defined as follows:

- *Creation*: The hosts create packets following a random uniform distribution with probability λ. Only another host can be the destination of a packet, which is also selected randomly. Finally, this new packet is appended at the end of the host tail.
- *Routing*: Each node picks up the packet at the head of its queue and decides which outgoing link is better suited to the packet destination. Here, the objective is to minimize the communication time for any single message, taking into account only shortest paths and also avoiding congested links. First, the selected link is the one that points to a neighbour node that is nearer to the packet destination. Second,

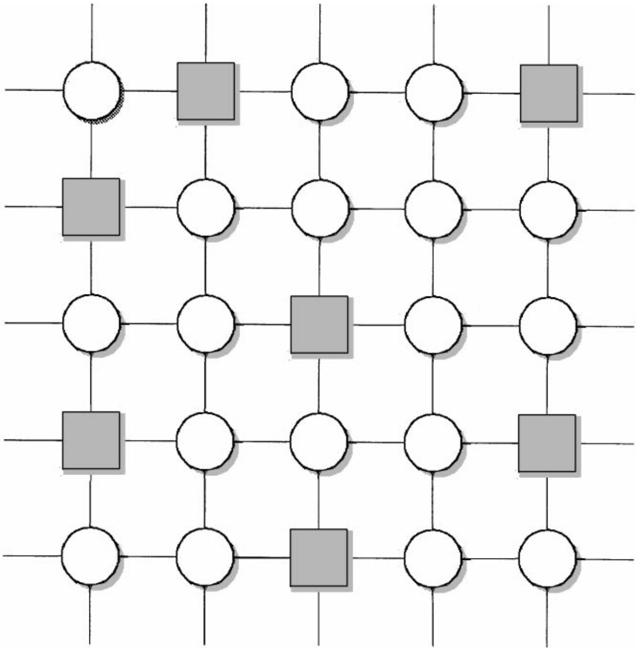

Fig. 1. Model network architecture (two-dimensional lattice, periodic boundary conditions). Two types of nodes are considered: hosts (gray squares) which can generate and receive messages, and routers (open circles) which can store and forward messages.

when two choices are possible, the less congested link is selected. The measure of congestion of a link is simply defined as the amount of packets forwarded through that link. Once the node has made the routing decision, the packet is inserted at the end of the queue of the node selected and the counter of the outgoing link is incremented by one.

These rules are applied to each site and each $L \times L$ updatings define our time step.

This model exhibits a similar phase transition than the one reported in previous studies [14,15]. It is shown in Fig. 2 for a $L = 32$ system with $\rho = 0.08$ (the same density is used in all our simulations). We can see that the transition occurs at a given $\lambda_c \approx 0.2$. As it occurs with models of highway traffic, the flow of packets is maximized at criticality, as shown in Fig. 2B, where the number of delivered packets (indicated as NDP) is plotted.

3. Scaling and self-similarity

An example of the time series at criticality for the previous system is shown in Figs. 3Aand B. It confirms our expectations and previous observation from real computer traffic: the local fluctuations in the number of packets $n(\mathbf{r}, t)$ are self-affine, as we can appreciate from an enlargement of the first plot. This is confirmed by the

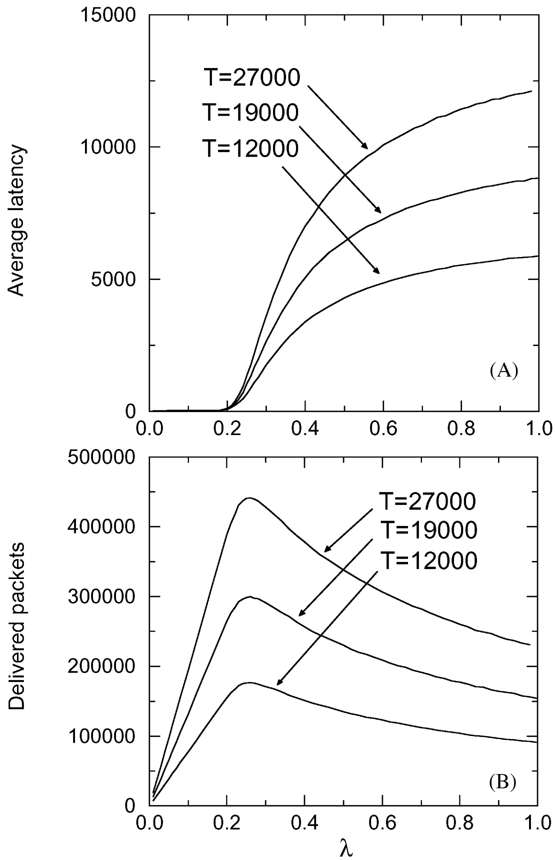

Fig. 2. (A) Phase transition in network traffic. Here $L = 32$ lattice has been used and the average latency has been computed over different, increasing intervals of time T steps, as indicated. The density of hosts is $\rho = 0.08$. (B) As a measure of efficiency, the number of delivered packets N_{dp} has been measured under the same conditions. We can see the optimum at the critical point $\lambda_c \approx 0.2$. For $\lambda < \lambda_c$ we have a linear increase $N_{dp} = \gamma\lambda$ with $\gamma = \rho L^2 T$, corresponding to the number of released packets.

calculation of the power spectrum $P(f)$. It is shown in Fig. 4 and scales as

$$P(f) \approx f^{-\beta} \quad (\beta = 0.97 \pm 0.06)\,. \tag{3}$$

There is some local variability in the value of the scaling exponent through space, but it is typically inside the interval $-0.75 < \beta < 1.0$, in agreement with data analysis [7,11].

The statistics of latencies and queue lengths leads to long tails close to criticality. Some examples of the results obtained are shown in Fig. 5. Here a $L = 256$ lattice has been used. Latencies are measured as the number of steps needed to travel from emitting hosts to their destinations. The distribution of latencies close to λ_c is a log normal, in agreement with the study of Huberman and Adamic for round-trip times of ping packets [12]. This means that there is a characteristic latency time but also very

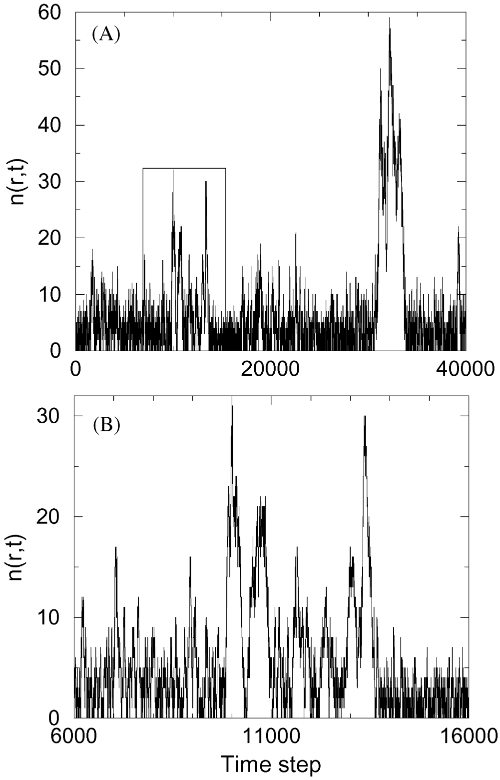

Fig. 3. An example of the time-series dynamics of the number of packets $n(r,t)$ at a given arbitrary node. Here $L = 256$, $\mu = 0.08$ and $\lambda = \lambda_c = 0.055$. We can see in (A) fluctuations of many sizes, which display self-affinity, as we can see from (B) where the fraction of the previous time series indicated by means of a window has been enlarged.

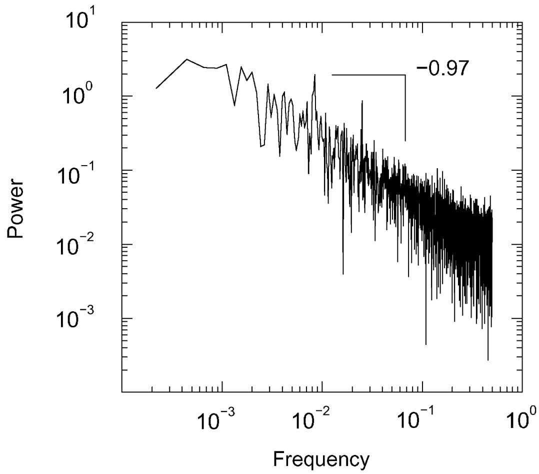

Fig. 4. Power spectrum $P(f)$ computed from the time series shown in Fig. 3A. A well-defined scaling is at work over four decades, with $\beta \approx 1$.

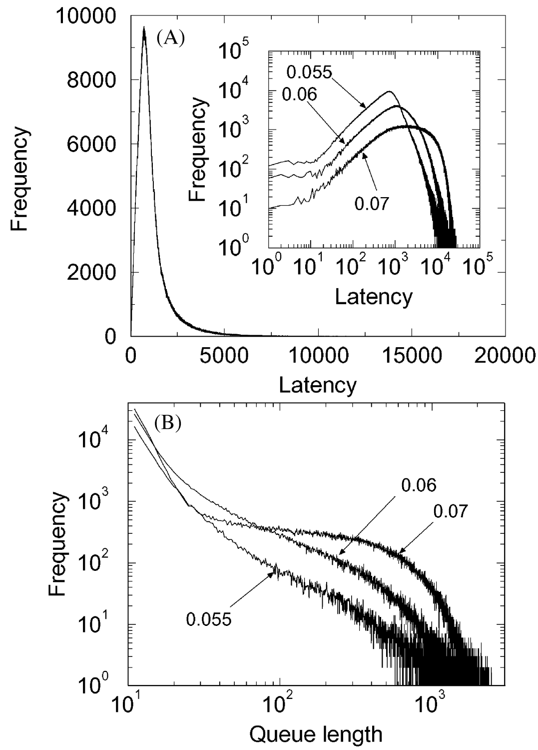

Fig. 5. (A) Log-normal distribution of latency times at criticality for a $L = 256$ system. Here $\lambda_c \approx 0.055$. Inset: three examples of these distributions in log–log scale for three different λ values (as indicated); (B) Distributions of queue lengths for the same system at different λ rates. Scaling is observable at intermediate values close to criticality.

long tails: a high fluctuation regime is present. As λ goes into the congestion phase, longer times are present but also long tails. This is due to the fact that at this phase the number of packets is always increasing with time.

The distribution of queue lengths is equivalent to the distribution of jam sizes in the highway traffic model. As with the Nagel–Schreckenberg model, the distribution approaches a power law for $\lambda \approx \lambda_c$ but it also displays some bending at small values (and a characteristic cutoff at large values). This is probably the result of the presence of spatial structures, which propagate as waves of congestion and will be analyzed elsewhere (Valverde and Solé, in preparation).

4. Efficiency, uncertainty and information transfer

As mentioned in the introduction, models of highway traffic flow revealed that the flow of cars (and thus the system's efficiency) is maximal at the critical point, but that the unpredictability is also maximal.

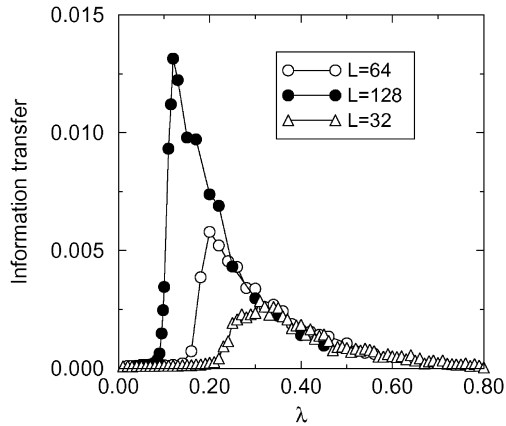

Fig. 6. Information transfer for three different lattice sizes (as indicated). Information transfer grows rapidly close to criticality but reaches a maximum at some point λ^* close to λ_c.

Efficiency can be measured in several ways. One is close to our model properties: efficiency is directly linked to information transfer and thus information-based measures can be used. Here we consider an information-based characterization of the different phases by means of the Markov partition Π. Specifically the following binary choice is performed:

$$\Pi = \{n(\mathbf{r}) = 0 \Rightarrow S(\mathbf{r}) = 0;\ n(\mathbf{r}) > 0 \Rightarrow S(\mathbf{r}) = 1\}$$

which essentially separates *non-jammed* from *jammed* nodes.

Information transfer is maximized close to second-order phase transitions [16,17] and should be maximum at λ_c. In order to compute this quantity we will make use of the previous partition Π. Let $S(\mathbf{r})$ and $S(\mathbf{k})$ the binary states associated with two given hosts in $\mathscr{L}$. The Π-entropy for each host is given by

$$H(\mathbf{r}) = -\sum_{S(\mathbf{r})=0,1} P(S(\mathbf{r})) \log P(S(\mathbf{r})) \tag{4}$$

and the joint entropy for each pair of hosts,

$$H(\mathbf{r}, \mathbf{k}) = -\sum_{S(\mathbf{r}),\, S(\mathbf{k})=0,1} P(\mathbf{r}, \mathbf{k}) \log P(\mathbf{r}, \mathbf{k}) \,, \tag{5}$$

where for simplicity we use $P(\mathbf{r},\mathbf{k}) \equiv P(S(\mathbf{r}),S(\mathbf{k}))$ to indicate the joint probability.

From the previous quantities, we can compute the information transfer between two given hosts (Fig. 6). It will be given by

$$M(\mathbf{r}, \mathbf{k}) = H(\mathbf{r}) + H(\mathbf{k}) - H(\mathbf{r}, \mathbf{k}) \,. \tag{6}$$

The average information transfer will be computed from $\bar{M}_q = \langle M(\mathbf{r}, \mathbf{k}) \rangle$ where the brackets indicate average over a sample of q hosts randomly chosen from the whole set (here $q = 100$).

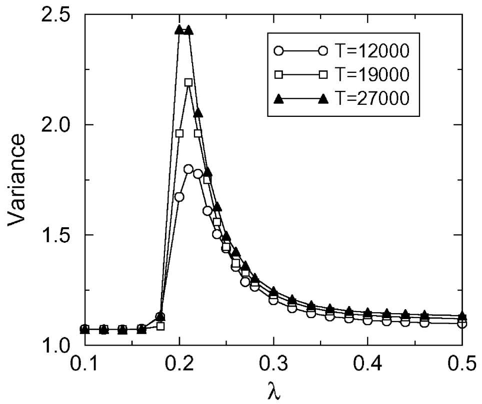

Fig. 7. The variance plot for the $L = 32$ system. The critical point λ_c is perfectly indicated by this measure with a sharp maximum. Three different times have been used in the averages.

At the sub-critical domain, in terms of information transfer under the Markov partition, all pairs of nodes will be typically in the non-congested (free state) and $P(i, j) \approx \delta_{00}$ so it is easy to see that in this phase we have vanishing entropies and the mutual information is small. The information is totally defined by the entropy of the single nodes, as far as the correlations are trivial. A similar situation holds at the congestion phase, where nodes are typically congested. At intermediate values, the fluctuations inherent to the system lead to a diversity of states that gives a maximum information transfer at some λ_c^*. It should be noted however that this measure is not very good for small systems, where $\lambda_c^* > \lambda_c$, but we can see that $\lambda_c^* \to \lambda_c$ as L increases.

Unpredictability will be measured, following Nagel and Rasmussen [13] by means of the normalized variance of latencies:

$$\sigma(T_L) = \frac{[(T_L - \langle T_L \rangle)^2]^{1/2}}{\langle T_L \rangle}, \tag{7}$$

where $\langle T_L \rangle$ is the average over a given number of steps.

The unpredictable nature of the critical point is sharply revealed by the plot of the variance $\sigma(T_L)$ (Fig. 7). We can see that, as it was shown by Nagel and co-workers for highway traffic, the system shows the highest unpredictability close to the critical point. At the subcritical regime $\lambda < \lambda_c$, the packets reach their destinations in a time close to the characteristic, average time of traveling. This situation sharply changes in the neighborhood of λ_c where the fluctuations (experienced as local congestion) lead to a rapid increase in the variance. As λ grows beyond the transition, these fluctuations are damped and $\sigma(T_L)$ decays slowly. This result also confirms the study by Nagel and co-workers who analyzed the behavior of the variance of travel times [13] for a closed-loop system. They found that there was a nontrivial implication for this result: increasing efficiency (i.e., traffic flow) tunes the system to criticality and as a consequence to unpredictable behavior.

5. Discussion

In this paper we have analyzed the statistical properties of a computer network traffic model. This model is a simple extension of the OS system, but with a random distribution of hosts scattered through $\mathscr{L}$ with a density ρ. One of the goals of our study was to see if the reported regularities from real networks of computers (such as the Internet) were similar to those observed in highway traffic and reproducible by our model. The second was to explore the possibility that the observed features correspond to those expected from a near-to-critical system. We have presented evidence that real Internet traffic takes place close to a phase transition point, although further work is needed involving the scale-free network topology of the real web. However, our result provide a preliminary support to the presence of critical phenomena in parallel arrays of computers.

The model has been shown to match some basic properties of Internet dynamics: (i) it shows self-affine patterns of activity close to criticality, consistent with the fractal nature of computer traffic; (ii) the observed time series display $1/f$ behavior and the corresponding Hurst exponents reveal the presence of persistence and long-range correlations in congestion dynamics, as reported from real data; (iii) the distribution of latency times close to the transition point is a lognormal, and the distribution of queue lengths approaches power laws with some bending for small lengths (as in highway traffic models).

The model confirms the previous conjecture [7] suggesting some deep links between the NS model and the dynamics displayed by computer networks close to critical points. In this sense, the previous measures and other quantitative characterizations support the idea that the two type of traffic share some generic features. The model also exhibits the same kind of variance plot shown by the NS and related models: it is almost zero at the subcritical (free) regime and it abruptly grows close to λ_c. This leads to the same conclusion pointed by Nagel and co-workers: maximum efficiency leads to complex dynamics and unpredictable behavior.

Some authors have discussed the origins of Internet congestion in terms of the interactions among users [12]. Huberman and Luckose suggested that this is a particularly interesting illustration of a social dilemma. Our study suggests a somewhat complementary view: there is a feedback between the system's activity and the user's behavior. Users introduce new packets into the system, thus enhancing the congestion of the net. As congestion increases, users tend to leave the net, thus reducing local activity. This type of feedback is similar to the dynamics characteristic of self-organized critical systems (such as sandpiles) [18]. The main difference arises from the driving. Activity is being introduced into the system without a complete temporal separation between two scales. In this sense, this is not a self-organized critical system but it is close enough to be the appropriate theoretical framework. An immediate extension of this model should contain a self-tuning of λ: users might increase their levels of activity if congestion is low and decrease it (or leave the system) in a very congested situation. In this way, the system might self-organize into the critical state.

Our previous results can also be applied to other, similar networks. This is the case of large, parallel arrays of processors. In this sense, some previous studies [19] have shown the validity of the OS model in describing the overall dynamics of small arrays of processors with simple topologies. They also found some additional phenomena such as the presence of hot spots, which we have also observed in our model. This work can be extended to high-dimensional parallel systems such as the connection machine [20] in order to test the presence of phase transitions and their dependence on dimensionality (in particular they will help to determine the upper critical dimension for this system).

Internet dynamics and the WWW growth provide an extremely interesting, real evolution experiment of a complex adaptive system [12]. In the near future, we are likely to see new types of behavior in the web. As Daniel Hillis predicts, as the information available on the Internet becomes richer, and the types of interactions among computers become more complex, we should expect to see new emergent phenomena going beyond any that has been explicitly programmed into the system [21]. Models based on phase transitions in far from equilibrium systems will be of great help in providing an appropriate theoretical framework.

Acknowledgements

We thank B. Luque for many useful discussions and his earlier participation in this work. This work has been supported by a grant PB97-0693 and by the Santa Fe Institute (RVS).

References

[1] B. Huberman (Ed.), The Ecology of Computation, North-Holland, Amsterdam, 1989.
[2] J.O. Kepart, T. Hogg, B. Huberman, Phys. Rev. A 40 (1989) 404.
[3] R. Albert, H. Jeong, A.-L. Barabási, Nature 406 (2000) 378.
[4] M. Takayasu, K. Fukuda, H. Takayasu, Physica A 274 (1999) 140.
[5] M. Takayasu, H. Takayasu, K. Fukuda, Physica A 277 (2000) 248.
[6] W.E. Leland, M.S. Taqqu, W. Willinger, IEEE Trans. Networking 2 (1994) 1.
[7] I. Csabai, J. Phys. A: Math. Gen. 27 (1994) L417.
[8] K. Nagel, M. Schreckenberg, J. Phys. I France 2 (1992) 2221.
[9] K. Nagel, M. Schreckenberg, J. Phys. A 26 (1993) L679.
[10] K. Nagel, M. Paczuski, Phys. Rev. E 51 (1995) 2909.
[11] M. Takayasu, H. Takayasu, T. Sato, Physica A 233 (1996) 824.
[12] B.A. Huberman, R.M. Luckose, Science 277 (1997) 535.
[13] K. Nagel, S. Rasmussen, in: R.A. Brooks, P. Maes (Eds.), Artificial Life IV, MIT Press, Cambridge, MA, 1994, p. 222.
[14] T. Ohira, R. Sawatari, Phys. Rev. E 58 (1998) 193.
[15] H. Fucks̀, A.T. Lawniczak, preprint adap-org/9909006.
[16] R.V. Solé, S.C. Manrubia, B. Luque, J. Delgado, J. Bascompte, Complexity 1(4) 1996.
[17] R.V. Solé, O. Miramontes, Physica D 80 (1995) 171.
[18] P. Bak, C. Tang, K. Wiesenfeld, Phys. Rev. Lett. 59 (1987) 381.
[19] K. Bolding, M.L. Fulgham, L. Snyden, Technical Report CSE-94-02-04.
[20] W.D. Hillis, The Connection Machine, MIT Press, Cambridge, MA, 1985.
[21] W.D. Hillis, The Pattern on the Stone, Weidenfeld and Nicolson, London, 1998.