EVALUATION OF TRAFFIC CONTROL STRATEGIES IN SCALE FREE NETWORKS USING A PARALLEL PROCESSING SIMULATOR

Radu DOBRESCU, Matei DOBRESCU, Stefan MOCANU, Sebastian TARALUNGA
POLITEHNICA University of Bucharest, ROMANIA

EXTENDED ABSTRACT

1. Introduction

A large number of complex networks, both natural and artificial, share the presence of highly heterogeneous, scale-free degree distributions. In this paper we present the evidence for the emergence of scaling and of the self-similarity properties in two important software procedures for Internet traffic optimization: congestion avoidance and epidemic spreading prevention. Although the rules that define the strategies involved in software design should lead to a tree-like structure, the final net is scale-free, perhaps reflecting the presence of conflicting constraints unavoidable in a multidimensional optimization process. Two basic features are common to many complex networks, including Internet: their scale-free (SF) topology [1] and a small-world (SW) structure [2], both being introduced in the proposed model . The first states that the proportion of nodes $P(k)$ having k links decays as a power law $P(k)=k^{-\lambda}$ (with $2<\lambda<3$). The second refers to a web exhibiting very small average path lengths between nodes along with a large clustering [3]. Due to the high number of nodes and links the simulations have been made on a cluster with six processing units using the principles of the Linux operating NS simulator [4].

2. Using the scale free approach in complex networks modeling

2.1 The topology of the Internet and the dynamics of Internet traffic

The Internet is a prime example of a self-organizing complex system, having grown mostly in the absence of centralized control or direction. It has been reported that Internet traffic fluctuations are statistically self-similar [5] and that the traffic displays two separate phases, congested and non-congested. The Internet flow is strongly localized: most of the traffic takes place on a spanning network connecting a small number of routers which can be classified either as “active centers,” which are gathering information, or “databases,” which provide information. So far, simulations and analytical studies have shown that it may have a considerable impact on network performance that could not be predicted by the traditional short-range-dependent models. In our simulations with the new proposed SFN model we have shown that the Internet displays a number of properties that distinguishes it from random graphs: wiring redundancy and clustering, non-trivial eigenvalue spectra of the connectivity matrix and a scale-free degree distribution.

2.2 Spread of epidemics in complex networks

The propagation of errors occurring on routers and servers that are physically linked in a large network is a typical example of epidemic process, in which the corruption (virus) is transmitted from infected to healthy individuals. The epidemic process is usually described in terms of *individuals* and their *interactions*. In this simplified formalism, individuals can only exist in a discrete set of states, such as susceptible (or healthy), infected (and ready to spread the disease), immune, dead (or removed), etc. Within this formalism, the system can be described as a *network* or graph, in which the nodes represent the individuals and the links are the connections along which the epidemics propagates. Standard epidemiological models usually consider *homogeneous* networks, which are those that have a connectivity distribution peaked at an average connectivity $\langle k \rangle$, and decaying exponentially fast for $k<<\langle k \rangle$ and $k>>\langle k \rangle$ [6]. On the other hand computer viruses and worms spread in environments characterized by scale-free connectivity. The proposed model allows studying the effect of the special nature of scale-free distribution on the properties of random network models, including methods for the study of the layer structure of the graph, the percolation threshold and the critical exponents.

3. Using self-similarity to infer parameters values in Internet traffic model

To model a distributed network environment like the Internet, it is necessary to integrate data collected from multiple points in a network in order to get a complete picture of network-wide view of the traffic. Knowledge of dynamic characteristics is essential to network management (e.g., detection of failures/congestion, provisioning, and traffic engineering like QoS routing or server selections). However, because of a huge scale and access rights, it is expensive (sometime impossible) to measure such characteristics directly. To solve this, methods and tools for inference of unobservable network performance characteristics are used in large scale networking environment. Our approach assume that network traffic (*T)* can be modeled as a function of three sets of parameters: number of user (*N*), user behavior (*U*) and application-specific parameters (*A*). In other words $T = f(N,U,A)$. The *user-behavior parameters* could be the distributions of user "think" time or the number of pages requested by the users in web traffic, while the *application-specific parameters* could be the distributions of object size or the number of objects in a web page.

Our study shows that the distributions of user behavior parameters (*U*) tend to be correlated over time on the same network. Such an observation suggests that one can model the distributions of user-behavior parameters at time t_2 based on measurement taken at time t_1 on the same network. Additionally, we find that the distributions of application-specific parameters (*A*) are likely to be correlated between two networks with "similar" user populations. Specifically, we find that the distributions of application-specific parameters between two similar networks tend to be correlated when traffic is highly aggregated. Based on the above observations, we propose the following approaches to infer traffic from network n_1 by utilizing measurements taken from network n_2, provided networks n_1 and n_2 have *similar* user populations:

- The first step is to collect some period of traffic on both network n_1 and n_2. Such initial measurements are used to derive the three sets of traffic parameters (*N, U, A*) of n_1 and n_2.
- Once the similarity between n_1 and n_2 is confirmed and the spatial correlations between them are computed, one then are allowed to model all three set parameters (*N, U, A*) of n_1's traffic at any future time based on *only* the measurements taken from n_2.

4 Efficient immunization strategies

The simplest immunization procedure one can consider consists in the random introduction of immune individuals in the population, in order to get a uniform immunization density. In this case, for a fixed spreading rate *R*, the relevant control parameter is the density of immune nodes present in the network, the immunity *g*. The presence of a uniform immunity will have the effect of reducing the spreading rate *R* by a factor *1-g*, *i.e*. the probability of finding and infecting a susceptible and no immune node will be *R(1-g)*. For homogeneous networks we can easily see that, for a constant *R*, the stationary prevalence R_g is 0 if $g>g_c$ and respectively $(g_c-g)/(1-g)$ if $g \leq g_c$, where g_c is the critical immunization value above which the density of infected individuals in the stationary state is null and depends on *R* as $g_c = 1 - R_c/R$. Thus, for a uniform immunization level larger than g_c, the network is completely protected and no large epidemic outbreaks are possible. On the contrary, uniform immunization strategies on SFN are totally ineffective. The presence of uniform immunization is able to locally depress the infection's prevalence for any value of *R* , but it does so too slowly, and it is impossible to find any critical fraction of immunized individuals that ensures the infection eradication

However, we can take advantage of the heterogeneity of SFN, by devising an immunization strategy that takes into account the inherent hierarchy in the network nodes. SFN posses a noticeable resilience to *random* connection failures, which implies that the network can resist a high level of damage (disconnected links), without loosing its global connectivity properties; *i.e*. the possibility to find a connected path between almost any two nodes in the system. At the same time, SFN are strongly affected by *selective* damage; if a few of the most connected nodes are removed, the network suffers a dramatic reduction of its ability to carry information. Applying this argument to the case of epidemic spreading, one can devise a *targeted* immunization scheme in which the most highly connected nodes are progressively make immune. While this strategy is the simplest solution to the optimal immunization problem in heterogeneous populations, its efficiency is comparable to the uniform strategies in homogeneous networks with finite connectivity variance. In SFN, on the contrary, it produces an arresting increase of the network tolerance to infections at the price of a tiny fraction of immune individuals. A most effective strategy seems to be a selective one, based on the immunization of a small fraction of *random acquaintances* of randomly selected

individuals, which prevents epidemics without requiring global knowledge of the network. Simulations on the SFN model shown that a large fraction f_c of the nodes need to be removed (immunized) before the integrity of the network is compromised. This is particularly true for scale-free networks with $2<\lambda<3$ where the percolation threshold $f_c \rightarrow 1$, and the network remains connected (contagious) even after immunization of most of its nodes. The fraction f_c needed to be immunized in order to stop the epidemic can be computed analytically.

5. Experiments and results

In order to do an Internet like network simulation we used a topology generation algorithm which follows the scale free networks laws. The algorithm is based on *preferential attachment* strategy which means that a new node will have a higher probability of attachment to those nodes which already have many connections (with other existing nodes). Fewer connections nodes have smaller chances to receive new connections. Along with this strategy we imposed rules for avoiding cyclic connections, very long chains connections and other restrictions. Therefore, we do not have isolated components, the chains are shorter than 30 nodes and the topology does not look like a tree. Using the topology generator we can control how cyclic the network will be, in our case aprox. 4% of the nodes build a cycle. There are 3 types of nodes: *Routers*, defined as nods with two or more connections; *Servers,* defined as nodes with a single connection; *Clients* (end-users), defined as nodes with a single connection that initiate connections to servers at random time moments. For the proposed model we have one server for each 20 to 100 clients. In simulations heavily loaded lines (such as those between routers) are faster, for example server-router 1 Gbps, client-router 10 Mbps, router-router 10, 100 Mbps or 1Gbps depending on the router type. Simulation can be implemented using a distributed network with 6 nodes by running *PDNS (Parallel/Distributed NS***)** which represents a number of extensions for *NS* simulator. This facilitates parallel execution of a network simulation in a distributed manner over a 6 nodes network. The primarily objective is to reduce the processing effort when simulating huge networks, in which case the demands (for memory and processor) may become to big to run the application on a single machine.

6 Conclusions

The main goal of this paper has been to study the effect of the special nature of scale-free distribution and of the self-similar behavior on the properties of Internet network models. Regarding the traffic optimization, the proposed methodology aims to infer traffic's QoS at places where taking measurements are infeasible and at the level of the whole network, to prevent congestion and blocking. The paper analyses also an epidemiological framework obtained in population networks characterized by a scale-free connectivity pattern. SFN are very weak in face of infections, and its susceptibility to epidemic spreading is reflected also in an intrinsic difficulty in protecting them with uniform immunization policies, but targeted or selective immunization procedures achieve the desired lowering of epidemic outbreaks and prevalence.

References:

[1] Barabasi, A.-L. and R. Albert (1999). Emergence of scaling in random networks. *Science* **286**, pp. 509–512

[2] Watts D. J. and S. H. Strogatz (1998). Collective dynamics of 'small-world' networks. *Nature,* **393**, pp.440–442

[3] R. Dobrescu, M. Rothenberg - A method for inferencing of network characteristics based on traffic similarities, *Bul. UPB, vol.67*, seria C Inginerie Electrica, nr.1, 2005, pag.23-32

[4] R. Dobrescu, M. Dobrescu, S. Mocanu, S. Taralunga – Client-Server Architecture for Parallel Image Processing, *WSEAS Transactions on Signal Processing,* Issue 9, vol.2, September 2006, p. 1181-1188, ISSN 1790-5022

[5] Dobrescu, R., M. Dobrescu and St. Mocanu (2004). Using Self Similarity To Model Network Traffic, WSEAS Transactions on Computers, Issue 6, **3**, pp. 1752-1757

[6] Pastor-Satorras R. and A. Vespignani (2001). Epidemic spreading in scale-free networks. *Phys.Rev. Lett.* **86**, pp. 3200–3203.