Studying the impact of measurement frequency on the IP-level routing topology dynamics

Sergey Kirgizov, Clémence Magnien, Fabien Tarissan, Azhu Liu

Laboratoire d'informatique de Paris 6
Boîte courrier 169, 4 place Jussieu, 75252 Paris Cedex 05
sergey.kirgizov@lip6.fr,clemence.magnien@lip6.fr,fabien.tarissan@lip6.fr, reddylao@gmail.com

Résumé – De nombreux travaux ont étudié la topologie de l'internet, mais peu d'entre eux se sont intéressés à comment elle évolue. Cet article se concentre la dynamique de la topologie de routage au niveau IP, et nous étudions plus particulièrement l'impact de la fréquence de mesure sur les observations de la dynamique. Pour cela, nous étudions tant des données issues de mesures périodiques des arbres de routage à partir d'un moniteur vers un ensemble de destinations, que le comportement d'un modèle de la dynamique de la topologie que nous avons présenté précédemment. Analyses précédentes ont montré que, après une augmentation initiale rapide, le nombre de nouveaux liens observés soutient une croissance linéaire pour une longue période de temps. Bien que la pente de cette partie linéaire puisse être considérée comme un indicateur de la vitesse de la dynamique, Nous montrons en fait que cette pente dépend intrinsèquement de la fréquence de mesure et qu'il est très difficile, sinon impossible, de quantifier la vitesse réelle de l'évolution de la topologie de l'internet.

Abstract – Many works have studied the Internet topology, but few have investigated the question of how it evolves over time. This paper focuses on the Internet routing IP-level topology dynamics, and in particular on the impact of the measurement frequency on the observed dynamics. For this end, we study both data from periodic measurements of routing trees from a single monitor to a fixed destination set, and the behavior of a model of the topology dynamics that we previously introduced. Previous analyses showed that after an initial fast increase, the number of new observed links sustains a linear growth for extended periods of time. The slope of this linear part can be considered as an indicator of the speed of the observed dynamics. We show that this speed depends intrinsically on the measurement frequency and that it is very difficult, if not impossible, to quantify the actual speed of the internet topology evolution

1 Introduction

Studying the structure of the internet topology is an important and difficult question. No official map being available, researchers have to conduct costly measurement campaigns, and deal with the fact that the obtained data can be biased [3, 1]. Studying the dynamics of this topology is therefore an equally hard, if not harder, problem.

Here we study the dynamics of *ego-centered view* [4] of the internet topology at the IP-level. Each *ego-centered view* is a tree of routing paths from the monitor to the destinations. A routing tree represents a snapshot of the network around the monitor at a given time. One *ego-centered view* can be measured quickly and with low network load with the `tracetree` tool [4]. Repeating those measurements periodically therefore allows to study the dynamics of this view.

Previous work has shown that ego-centered views exhibit strong dynamics, and in particular that the set of observed IP addresses and links evolves much more quickly than what was previously expected [6]. In all our measurements, the number of links observed since the measurement beginning displays a linear progression after a fast initial growth. Fig. 1(a) illustrates this. Our main goal in this paper is to show how measurement frequency affects the observed behavior. We will focus on the study of the slope α of the linear part of the plot. We analyze both the real data and the behavior of a model of the topology dynamics that we previously introduced [5, 7].

2 Model

We introduced in previous works [7, 5] a simple model that reproduces the observed behavior. This model incorporates two factors that have been shown to play an important role in our observations: load-balancing, i.e. the fact that at any given time several routes may exist from the monitor to a destination, and routing dynamics, i.e. the fact that routes can change with time. Here we briefly describe the model and show it captures the observed behavior.

The model incorporates four ingredients: the routing topology, the routes from the monitor to the destinations, load-balancing, and routing changes. The goal of the

The model

Topology:
random graph (Erdős–Rényi or power-law)
n – number of nodes
m – number of links
d – number of destinations
r – number of measurement rounds
Measurements:
random breadth-first search towards d destinations
Dynamics:
s – number of link swaps.

Table 1: Model parameters

model was to obtain a simple baseline model which makes it possible to investigate the role of each component.

The network topology is modeled by a random graph with n nodes and m links, obtained with the Erdős–Rényi model [2]. Given a generated topology, we assume that the route between the monitor and a destination is a shortest path. To simulate load-balancing we implement a *random breadth-first search*: the neighbors of each node are considered in a random order. In this way, two consecutive measurements of shortest path trees will not be identical, even if the underlying graph does not change inbetween. Next, we model routing changes by using a simple approach based on link rewiring, or *swap*. It consists in choosing uniformly at random two links and swapping their extremities. Finally, the simulation is performed in the following way: we generate a random graph with n nodes and m links, randomly select one node as the monitor and d nodes as the destinations, then we simulate r measurement rounds with random breadth-first searches. Each obtained tree simulates one ego-centered view. Between two consecutive rounds we modify the topology by performing s random swaps, where n, m, s and r are parameters of the model. These parameters are summarized in Table 1.

Simulations show that the model captures the main characteristics of the dynamics of the ego-centered views: (1) initial fast growth at the beginning; (2) linear behavior which can be described by the slope α (see Fig. 1). Note however that there is a quantitative difference between empirical data and simulation results: in particular the size of a routing tree, given by the first point in the plot, is approximately equal to 2 500 in practice whereas it is close to 1 000 in the model; the slope for empirical data is approximately equal to 0.1, while it is closer to 1 for the model. Note however that our goal is to reproduce the *qualitative* behavior observed in real data, in order to understand fundamental questions such as the one we are studying here. Quantitavely matching the model with our empirical observations is an interesting, but much less fundamental question, which we leave for future work.

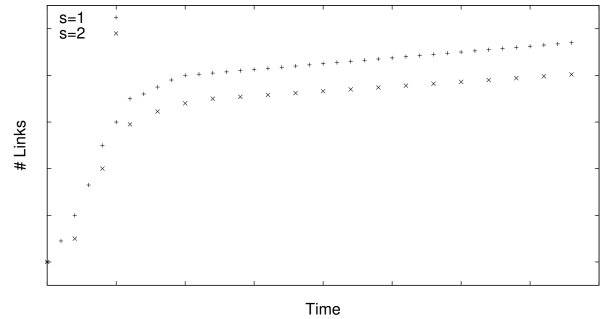


Figure 2: Number of links observed since measurement beginning with two different frequencies. Simulations.

3 Impact of the time interval between measurements

The observed slope depends both on the rate of routing changes and load balancing: the more changes happen, the more new links will be observed over time; and the more routes exist between two nodes, the more consecutive measurements will observe new routes.

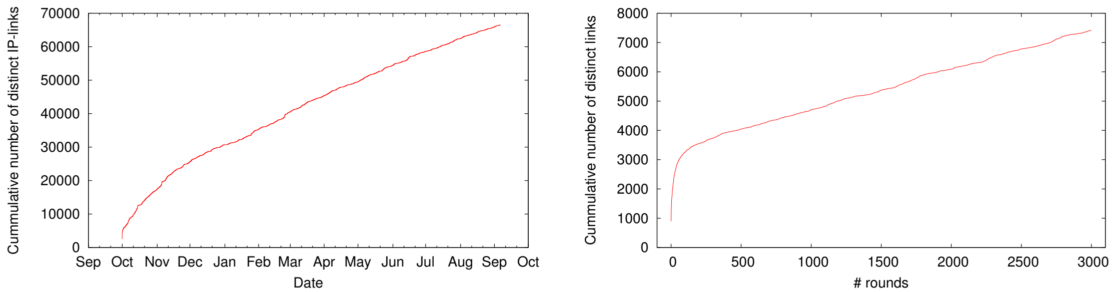
On the one hand, if the measurement frequency is too low, we will fail to observe some links, because they will disappear before the corresponding route is explored, and the value of the slope will be lower than α_m. Figure 2 illustrates this with model simulations performed with one and two swaps per round.

On the other hand, assuming that routing changes happen at a constant rate, if measurements are performed fast enough, all routes from the monitor to the destinations will be observed. More formally, let Δ be the time interval between two consecutive rounds. If Δ_i and Δ_j are two such intervals, let α_i and α_j be the corresponding slopes. We therefore expect that there exists Δ_m such that for all $\Delta_k \leq \Delta_m$ we should observe the same slope:

$$\Delta_k \leq \Delta_m \Rightarrow \alpha_k = \alpha_m \, .$$

Figure 3(a)) illustrates this.

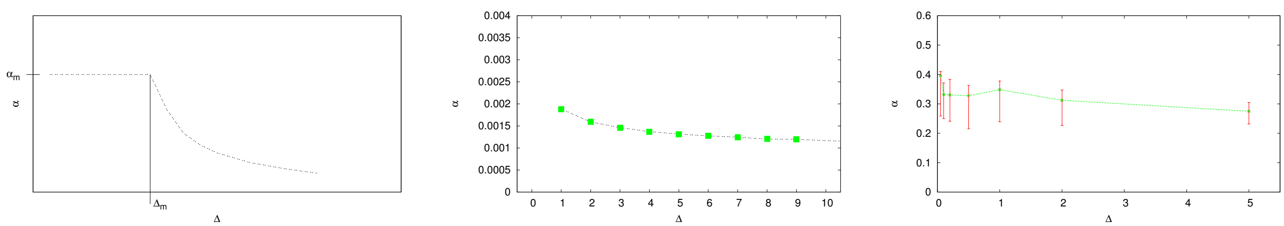
In order to test whether we are measuring with sufficient frequency or not, we compute the slopes for different values of Δ. For performing a rigorous analysis with real data, one should ideally perform several measurements at different frequencies. For these measurements to be comparable, they should be performed from the same monitor towards the same destination set, and at the same time. As this is not feasible in practice, we simulate from a real measurement performed with some $\Delta_{original}$ other measurements with different, lower, frequencies. We do so by taking into account only every n-th measurement, so that the simulated interval Δ_n will be equal to $n \cdot \Delta_{original}$. We are then able to compute the corresponding slope α_n. We use measurements performed at a high frequency ($\Delta_{original}$ is equal to 1m 25s) and for a long time, so that we are able to simulate measurements using a wide fre-



(a) Real measurements

(b) Model

Figure 1: Number of distinct IP links observed since measurement beginning. Each curve can be decomposed into two parts: 1) initial fast growth at the beginning; 2) linear behavior which can be described by the slope α. The model exhibits a qualitatively similar behavior to the empirical data.



(a) Expected behavior

(b) Real measurements, $d = 500$

(c) Erdős–Rényi model ($n = 10^5$, $m = 2 \cdot 10^5$), $d = 300$.

Figure 3: Impact of the measurement frequency. Each figure represents α (the slope of the plot of the number of observed links as a function of time) as a function of the time interval between consecutive rounds Δ. For the model simulations, we present results for 100 experiments. Each point is the average of the slope, and the errorbars represent the 25- and 75-percentiles.

quency interval from them.

The model has no such parameter as Δ. However we can assume that the topology changes, i.e. swaps, happen at a constant rate. Let Δ stand for the elapsed period of time between two consecutive rounds if only one swap between them is performed. Then we can simulate a lower frequency in the model by performing n swaps between each two consecutive rounds, which will represent a period Δ_n equal to $n \cdot \Delta$. Conversely, we can simulate a higher frequency by performing only one swap every n rounds, which will represent a period Δ_n equal to Δ/n.

Figure 3 shows the observed results. For the simulation results, we observe that there is a high variability between results (the plot shows the average observed values, as well as the 25- and 75-percentiles). This makes it difficult to draw a rigorous conclusion, but the observations are compatible with the presence of a plateau for $\Delta \leq 1$. The median of the observed values, not presented here, also strongly suggests the presence of a plateau. Then for $\Delta \geq 1$, the slope decreases with Δ, showing that for these parameters, a frequency smaller than one round each swap is too slow to observe all changes. In the case of empirical data, we again observe that, the longer the interval between measurements, the smaller the slope α. This confirms our expectations in the case of $\Delta \geq \Delta_m$. However, in this case, there is no plateau at the beginning of the curve, therefore we don't know if the highest frequency is optimal or not.

4 Conclusion

We studied the dynamics of the internet topology. After an initial fast increase, the plot of the number of observed links sustains a linear growth for extended periods of time. Although the slope of this linear part could be considered as an indicator of the speed of the *observed* dynamics, we show that this slope depends intrinsically on the measurement frequency. We have introduced a methodology for testing whether a measurement is performed fast enough to observe all changes that happen in the underlying topology. Unfortunately it seems that it is impossible to obtain this in real measurements (we used measurements performed with a very high frequency and it seems very difficult to improve this).

When comparing empirical observations to simulations, we found that the variability between different experiments performed with the same parameters is quite high, making it difficult to understand the model's behavior. Future work should improve this, either by performing massive numbers of simulations in order to obtain significant results, or by developing analytical results for the model behavior.

Another important direction would be to find out if it is possible to reproduce quantitatvely with the model the observations made empirically. This would help in understanding the structure of the internet topology better.

Finally, another interesting direction consists in considering our problem as some kind of sampling problem. Given a function $f : \mathcal{TIME} \rightarrow \mathcal{S}$ such that $f(t)$ is the set of all routes existing from the monitor to the destinations at time t, represented in some space $\mathcal{S}$, our measurements only discover one route to each destination, and can be represented by a function $m(t) \subset f(t)$. Moreover, we can only have access to discrete values of $m(t)$, corresponding to the moments where we make a measurement roudn, which happens with a finite frequency. The question then becomes: how to estimate $f(t)$ from our partial sampling, both in time and in the space $\mathcal{S}$?

References

[1] D. Achlioptas, A. Clauset, D. Kempe, and C. Moore, *On the bias of traceroute sampling*, Journal of the ACM, 56 (2009).

[2] P. Erdös and A. Rényi, *On random graphs*, Publicationes Mathematicae Debrecen, 6 (1959), p. 290.

[3] A. Lakhina, J. Byers, M. Crovella, and P. Xie, *Sampling biases in IP topology measurements*, in INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies, vol. 1, march 2003, pp. 332 – 341 vol.1.

[4] M. Latapy, C. Magnien, and F. Ouédraogo, *A radar for the internet*, Complex Systems, 20 (2011), pp. 23–30.

[5] C. Magnien, A. Medem, S. Kirgizov, and F. Tarissan, *Towards realistic modeling of IP-level routing topology dynamics*. Submitted, 2012.

[6] C. Magnien, F. Ouédraogo, G. Valadon, and M. Latapy, *Fast dynamics in internet topology: Observations and first explanations*, in Proceedings of the Fourth IEEE International Conference on Internet Monitoring and Protection, 2009, pp. 137–142.

[7] A. Medem, C. Magnien, and F. Tarissan, *Impact of power-law topology on IP-level routing dynamics: simulation results*, in Proceedings of the Fourth International Workshop on Network Science for Communication Networks (NetSciCom), 2012.