

# Describing and Simulating Internet Routes

J r mie Leguay, Timur Friedman, Kav  Salamatian

LIP6 – CNRS and Universit  Pierre et Marie Curie  
8, rue du Capitaine Scott, 75015 Paris, France  
Tel. +33 1 44 27 71 34, Fax +33 1 44 27 53 53  
{jeremie.leguay,timur.friedman,kave.salamatian}@lip6.fr

---

This paper introduces relevant statistics for the description of routes in the internet, seen as a graph at the interface level. Based on the observed properties, we propose and evaluate methods for generating artificial routes suitable for simulation purposes. The work in this paper is based upon a study of over seven million route traces produced by CAIDA’s *skitter* infrastructure.

**Keywords:** Network measurements, graphs, statistical analysis, modeling, simulation.

---

## 1 Introduction

Realistic modeling of routes in the internet is a challenge for network simulation. Until now, one has had to choose one of the three following approaches to simulate routes: (1) use the shortest path model, (2) explicitly model the internet hierarchy, and separately simulate inter- and intra-domain routing, or (3) replay routes that have been recorded with a tool like *traceroute*. All of these methods have serious drawbacks. The first method does not reflect reality since routes have not the same properties as shortest paths [Pax97], mainly because of routing policies [TGS01]. The second method is limited by our ability to explicitly simulate the internet hierarchy [TGJ<sup>+</sup>02, LAWD04]. Finally, the third method is not suitable if routes from a large number of sources are to be simulated. Today’s route tracing systems employ at most a few hundred sources. CAIDA’s *skitter* [HPMc02, CAI] infrastructure, for instance, produces an extensive graph suitable for simulations, but it based on routes from just thirty sources.

This paper’s principal contribution is a new approach to modeling routes in the internet, one that does not share the drawbacks just described. We suggest using an actual measured graph of the internet topology, such as the graph generated by *skitter*. Between random chosen sources and destinations, we suggest generating artificial routes with a model chosen to reflect statistical properties of actual routes.

The remainder of this paper is organized as follows. Sec. 2 describes the data set that we have used. Sec. 3 proposes the set of statistical properties to describe routes in the internet. Sec. 4 proposes the models we use to simulate routes based on these properties. Sec. 5 evaluates those models, and Sec. 6 concludes the paper.

## 2 The data set

This study uses *skitter* data from July 2<sup>nd</sup> 2003. The data was collected from 23 servers targeting 594,262 destinations. We obtained the corresponding IP graph by merging the results of the 7,075,189 *traceroutes* conducted on that day. This graph captures the small-world, clustered, and scale-free nature of the internet already pointed out for instance in numerous publications [JB02, FFF99]. In particular, the average distance is approximately 12.54 hops, and the degree distribution is well fitted by a power law of exponent 1.97.

## 3 Statistical properties of routes

This section presents a set of properties for statistical description of internet routes. These properties motivate the models of Sec. 4. Several properties have already been studied in previous work, and the work here

serves to evaluate and update them.

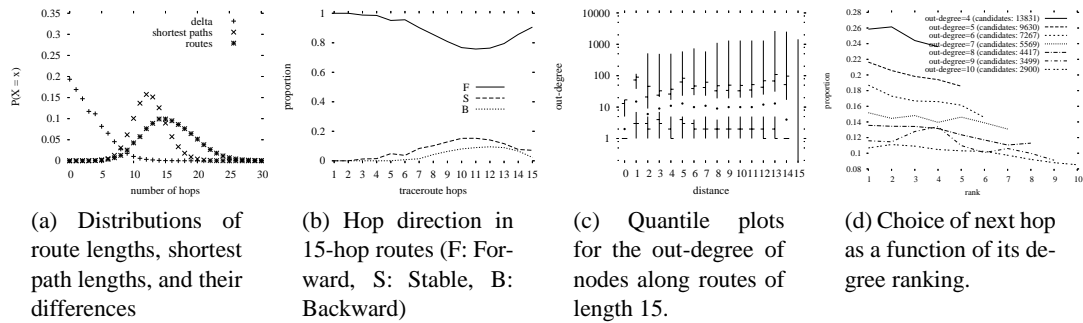


Figure 1: Statistical properties of internet routes.

### 3.1 Route lengths

It is well known that routes are not shortest paths: they are not optimal in general. Fig. 1(a) shows the length distributions of the routes in our data set, and of the corresponding shortest paths. It also shows the distribution of the difference (*delta*) between the length of a route and the corresponding shortest path. The mean length of 15.57 hops for routes in this data set fits closely Paxson’s observations [Pax97, Pax96] on a data set from nine years prior. The shortest paths have a mean length of 12.55 hops (11.4 hops if the graph is considered to be undirected).

### 3.2 Hop direction

When a packet travels from one router to another, it may move closer to its destination, but also it may move farther, or it may move to an interface that is at the same distance from the destination as one it just left. Likewise, the distance from the source may increase, decrease, or stay constant. We will call these behaviors the *hop direction*. Note that in shortest paths hops always increase the distance from the source and decrease the distance to the destination. Fig. 1(b) shows the portion of forward, backward, and stable hops at each hop distance for routes of 15 hops (the most numerous ones). Note that, as one would expect, the first and last few hops are generally forward because there are few alternatives. On the contrary, in the core of the network a significant proportion of the hops do not go closer to the destination. This type of behavior has already been described as the product of policy-based routing in the core of the internet [TGS01].

### 3.3 Degree evolution along a route

Recent work has shown that in many real-world complex networks, most of the short paths between pairs of nodes tend to pass through the highest degree nodes [KYHJ02, GLM04]. These observations lead us to ask how the node degree evolves along a route. Fig. 1(c) shows<sup>†</sup> how node degree evolves for routes of length 15. It reveals that a typical route does not pass through the highest degree nodes, though a certain number of routes do pass through some very high degree nodes. There is a peak in median out-degree observable at distance 1. The median falls at distance 2, rises again, and then stays fairly flat out to distance 13, with a median degree of about 10. This leads us to the following interpretation: the hosts have low degree, they are connected at their first hop router to relatively high degree nodes which play the role of access points, and then packets are routed in a core network where the degree does not depend much on the distance from the source or from the destination.

Furthermore, one may wonder if there is a simple local rule that can be observed for the degree evolution. In particular, if there is a choice of next hop interface along a route, is there a correlation between the degree rank of an interface and its probability of being chosen? Fig. 1(d) plots the probability that a packet travels

<sup>†</sup> In Fig. 1(c), dots indicate the median. Vertical lines run from the min to Q1 and from Q3 to the max. Tick marks indicate the 5<sup>th</sup>, 10<sup>th</sup>, 90<sup>th</sup> and 95<sup>th</sup> percentiles.

to an interface's  $i$ -th ranked neighbor, where the neighbors are ranked from highest out-degree to lowest. This figure highlight such a correlation.

## 4 Route models

The previous section provides a set of simple statistical tools to capture some properties of routes in the internet. We now propose two models designed to capture these features.

### 4.1 Random deviation model

The random deviation model is based upon the idea that a route usually follows a shortest path, but might occasionally deviate from it. We modeled this using one single parameter,  $p$ , the probability at any point of deviating from the current shortest path to the destination, if such a deviation is possible. We found  $p = 0.2$  to work well. A random deviation route from source  $s$  to destination  $d$  is therefore based upon a shortest path  $u$  from  $s$  to  $d$ . At each hop, with probability  $1 - p$ , the route continues along  $u$ . But with probability  $p$  it will, if possible, deviate off  $u$  to another path.

### 4.2 Node degree model

The basic idea is that a path which goes preferentially towards high degree nodes tends to see most nodes very rapidly [KYHJ02]. The node degree model is based upon a similar approach, as follows. Two paths are computed, one starting from the source and the other from the destination. The next node on the path is always the highest degree neighbor of the current node. The computation terminates when we reach a situation where a node is the highest degree neighbor of its own highest degree neighbor. One can show that this is the only kind of loop can occur. Then, one of two cases applies: either the two paths have met at a node, or they have not. In the first case, the route produced by the model is the discovered path (both paths are truncated at the meet up node, and are merged). In the second case, we compute a shortest path between the two loops, and then obtain the route by merging the two paths and this shortest path, removing any loops. This method has already been proposed [BLP04] as an efficient way to compute short paths in complex networks.

## 5 Evaluation

This section compares the performance of the random deviation and node degree models. For each model, we chose at least 60,000 (source, destination) pairs at random from amongst the nodes of the graph and generated an artificial route for each of them. We compute the same statistics on these routes as we had computed for actual routes in Sec. 3. Fig. 2 shows the statistics for each model.

Comparing the route length distributions, we find that both models generate distributions that are symmetric, average somewhat higher than the shortest path distribution, and have tails similar to the actual route length distribution shown in Fig. 1(a). Mean route length is 15.15 for the random deviation model and it is 14.96 for the node degree model. Lengths of paths generated with the node degree model tail off somewhat quicker than in reality, but the degree of fidelity is nonetheless remarkable given that the length distributions are not explicitly part of the model.

Looking at the hop directions for the most frequent route length, we found that the curves for the random deviation model better match the shapes of the curves for real routes shown in Fig. 1(b). Hops are mostly forward near the source, but dip to around 80% roughly ten hops out (whereas in reality the portion of forward hops dips to around 80% at eleven or twelve hops out). This is in marked contrast to hop directions produced by the node degree model because forward hops dip much sooner and a bit less steadily. But overall portions of forward, stable, and backward hops closely match reality for both models.

The node degree model shines compared to the random deviation model in capturing the evolution of the out-degree close to a route's source. Routes generated with this model show the peak in the out-degree before settling down to a median value. The peak is reached at distance 2 rather than at the first hop router.

Based upon this comparison to real routes, we can state that the random deviation and node degree models do a reasonable job of emulation, though each model captures some aspects better than others, and their strengths are different. Both models clearly out-perform the shortest path model.

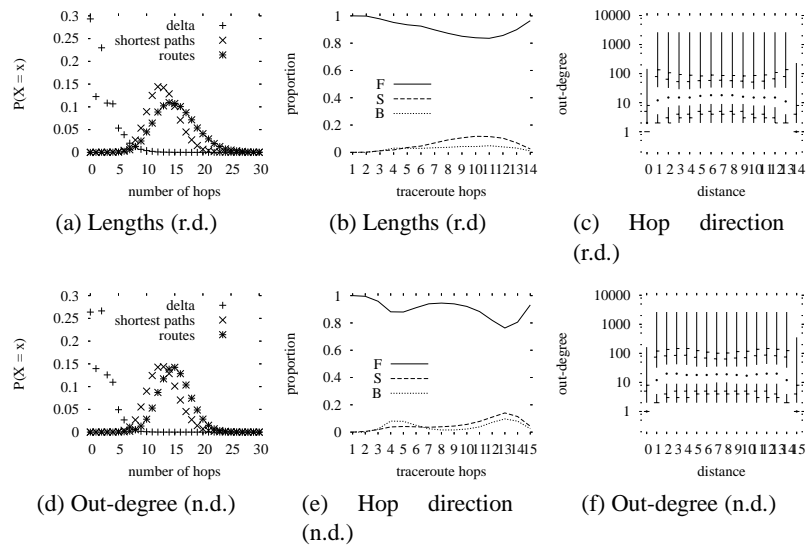


Figure 2: Experiments using the *random deviation model* (top) and the *node degree model* (bottom) on the undirected skitter graph using sources and destinations chosen at random from amongst all the nodes in the graph.

## 6 Conclusion and future work

The main contribution of this paper has been to propose new alternatives for the simulation of routes in the internet: the use of simple models that capture non-trivial statistical properties of routes. Future work along these lines might include the development of models that explicitly incorporate some additional characteristics, such as the clustering coefficient, or that captures something of the dynamics of internet routes.

## References

- [BLP04] E. Bampis, M. Latapy, and F. Pascual. Computing short paths in scale-free networks. 2004. preprint.
- [CAI] CAIDA. skitter. a tool for actively probing the Internet, <http://www.caida.org/tools/measurement/skitter/>.
- [FFF99] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. In *Proc. ACM SIGCOMM*, 1999.
- [GLM04] J.-L. Guillaume, M. Latapy, and C. Magnien. Comparison of failures and attacks on random and scale-free networks. In *Proceedings of the 8th International Conference on Principles of Distributed Systems (OPODIS)*, 2004. <http://www.liafa.jussieu.fr/~latapy/Publis/>.
- [HPMc02] B. Huffaker, D. Plummer, D. Moore, and k claffy. Topology discovery by active probing. In *Proc. Symposium on Applications and the Internet (SAINT)*, January 2002.
- [JB02] S. Jin and A. Bestavros. Small-world internet topologies: Possible causes and implications on scalability of end-system multicast. Tech. Report BUCS-2002-004, Boston Univ. Computer Sci., 2002.
- [KYHJ02] B. J. Kim, C. N. Yoon, S. K. Han, and H. Jeong. Path finding strategies in scale-free networks. *Phys. Rev. E* 65, 027103, 2002.
- [LAWD04] L. Li, D. Alderson, W. Willinger, and J. Doyle. A first-principles approach to understanding the internet's router-level topology. In *Proc. ACM SIGCOMM*, 2004.
- [Pax96] V. Paxson. End-to-end routing behavior in the Internet. In *Proc. ACM SIGCOMM*, 1996.
- [Pax97] V. Paxson. End-to-end routing behavior in the Internet. *IEEE/ACM Trans. on Networking*, 5(5):601–615, October 1997.
- [TGJ<sup>+</sup>02] H. Tangmunarunkit, R. Govindan, S. Jamin, S. Shenker, and W. Willinger. Network topology generators: Degree-based vs. structural. In *Proc. ACM SIGCOMM*, 2002.
- [TGS01] H. Tangmunarunkit, R. Govindan, and S. Shenker. Internet path inflation due to policy routing. In *Proc. SPIE ITCOM*, 2001.