Impact of Alias Resolution on *traceroute*-based Sample Network Topologies

Mehmet Hadi Gunes, Nicolas Sanchis Nielsen, and Kamil Sarac

University of Texas at Dallas, Richardson, TX 75083
{mgunes,nicolas,ksarac}@utdallas.edu

Abstract. Most Internet measurement studies utilize traceroute-collected path traces to build Internet maps. In this paper, we measure the impact of alias resolution problem on Internet topology measurement studies. Our analysis shows that the alias resolution process has a significant effect on the observed characteristics of the topology maps.

1 Introduction

Internet measurement studies require the availability of representative Internet maps. Most measurement studies utilize a well-known Internet debugging tool called *traceroute* to collect a router-level topology map from the Internet. After collecting the path traces, the information needs to be processed to build the corresponding network topology. This step requires identification of the IP addresses belonging to the same router, a task often referred to as *IP alias resolution*. Since routers have multiple interfaces with different IP addresses, it is likely that a router may appear on multiple paths with different interface IP addresses. The goal of IP alias resolution is to combine the IP addresses that belong to the same router into a single node in the resulting topology map.

Several tools have been proposed to resolve IP aliases in traceroute-based topology construction studies. The current best practice is to use the existing tools to resolve IP aliases to build a topology map. However, there is no practical way to measure the success rate of the alias resolution process. That is, it is extremely difficult to collect the underlying topology information for verification purposes. In fact, the lack of the underlying topology information is the main reason that leads the researchers to conduct topology collection studies.

At this point, one issue is to understand the impact of the alias resolution process on the observed topological characteristics of the collected topology maps. If the impact is acceptably small, then we could have confidence on the conclusions of the measurement study even without an alias resolution process. On the other hand, if the impact is high, the conclusions of the study may significantly depend on the accuracy of the alias resolution process. Although several recent studies pointed out the impact of inaccurate alias resolution in certain measurement study results, to the best of our knowledge, there is no systematic study that quantifies the impact of inaccurate alias resolution on traceroute-based Internet measurement studies.

In this paper, we present an experimental study to quantify the impact of alias resolution on Internet topology measurement studies. First we generate several synthetic network graphs using Barabasi-Albert (BA), Waxman (WA) and Transit-Stub (TS) network models. Then, we emulate traceroute functionality by collecting a number of path traces from the network graphs. During the sample topology construction, we use different success rates for the alias resolution process to obtain different sample topologies. Here, 0% indicates that alias resolution fails for all nodes in the network and 100% indicates that it succeeds for all nodes. We then study various topological characteristics of these sample topologies to quantify the impact of alias resolution on the observed results. We consider over 20 different graph characteristics including topology size, node degree, degree distribution, joint degree distribution, characteristic path length, betweenness, and clustering related characteristics. Due to size limitations, we present only a subset of the results to summarize our findings.

Our main conclusion in this study is that the accuracy of the alias resolution process has a significant impact on almost all topological characteristics that we consider in this study. Therefore, Internet measurement studies should employ all the means possible to increase the accuracy/completeness of the alias resolution process. Even in this cases, our confidence in the results of such measurement studies will be limited by the lack of a mechanism to verify the accuracy/completeness of the alias resolution process.

2 Impact of Alias Resolution on Degree Characteristics

In this section, we study changes in node degrees with improving alias resolution. In our experiments we observe that the accuracy of the alias resolution process has an important impact on the node degree-related characteristics of the sample topologies. Although one may intuitively expect an improvement on the accuracy of the degree-related characteristics with an increasing success rate of the alias resolution process, we may not necessarily observe such a trend all the time. Fig. 1 presents an example scenario where 'no-alias resolution' case (Fig. 1-b) results in a better approximation to (1) the degree of node a and (2) the average and the maximum degrees of the original subgraph (Fig. 1-a) compared to the 'partial alias resolution' case (Fig. 1-c) when we resolve aliases only for a.

Next, we study several sample topologies to observe the changes in node degrees as the success of the alias resolution process increases. This helps us gain more insight into the impact of the alias resolution process on the node degree characteristics. Fig. 2-a,-b,-c show the changes in node degrees for (100,100)-sample topology of BA graph for 0%, 40%, and 80% alias resolution success rates. In these figures, 'Observed Degree' indicates the degrees of the nodes in the sample topology with imperfect alias resolution and 'True Degree' indicates the degrees in the sample topology with perfect alias resolution. Each point in these figures may correspond to one or more nodes in the sample topology with the same 'Observed' and the same 'True' degrees. The number of nodes corresponding to each point is presented in the frequency distribution graph in Fig. 2-d. As an example, the '+' tick at location (99,1) in Fig. 2-d indicates

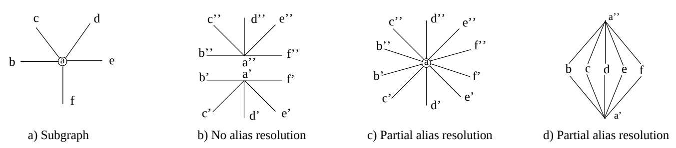
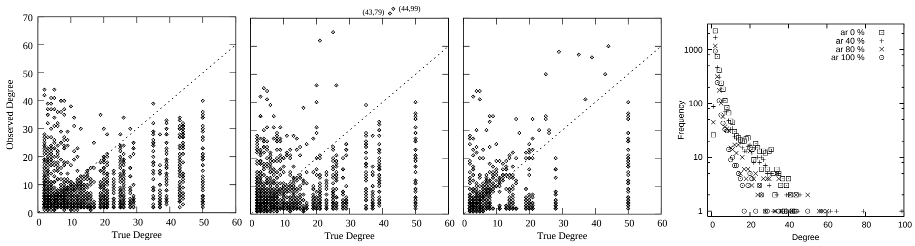


Fig. 1. Effect of partial alias resolution on a subgraph



(a) 0% Alias Resolution (b) 40% Alias Resolution (c) 80% Alias Resolution (d) Frequency distribution

Fig. 2. Degree comparison for (100,100)-sample topologies from Barabasi-Albert

that there exists only one node with an 'Observed Degree' of 99 under 40% alias resolution success rate and this node is presented in Fig. 2-b with the point marked at (44,99) label at the top of the figure. The label (44,99) indicates that the 'Real Degree' of this node is 44.

We now present several observations about the results presented in these figures. The points above the x=y line in Fig. 2-a,-b,-c correspond to overestimation of node degrees and the points below the x=y line correspond to underestimations of node degrees in the sample topologies. In general, overestimation is caused by alias resolution problems at the neighboring nodes of a given node. Fig. 1-c presents an example for this case. Similarly, underestimation is caused by alias resolution problems at the node itself. In addition, the comparison of Fig. 2-a,-b,-c show that the *observed* maximum degree of the graph first increases from 44 in Fig. 2-a to 99 in Fig. 2-b. It then goes down to 60 in Fig. 2-c (and down to 50 with 100% alias resolution success rate). Another observation from the figure is that alias resolution problems at a node may introduce a significantly large number of artificial nodes in the resulting sample topologies. As an example, according to Fig. 2-d, there is only one node with a *true degree* of 50 in the *real* sample graph (i.e., refer to (50,1) point in Fig. 2-d). On the other hand, Fig. 2-a,-b,-c show a large number of nodes with *observed degrees* less than 50 that correspond to a node with a degree of 50. Finally, we observe that as the alias resolution success rate increases some of the underestimation cases change to overestimation (compare Fig. 2-a vs. Fig. 2-b for x=43 and x=44 and Fig. 2-b vs. Fig. 2-c for x=35, x=37, and x=39). This indicates that although the alias resolution problems of the corresponding nodes are fixed, there exists some neighbors of these nodes with alias resolution problems causing overestimation.

3 Impact of Alias Resolution on Graph Characteristics

In this section we summarize the impact of poor alias resolution on topologies.
Topology Size: According to the experiment results, the success of alias resolution has a big impact on the topology size. Number of nodes and edges reduces 57% and 62%, respectively, on average for sample topologies as alias resolution improves from 0% to 100%. The impact of imperfect alias resolution increases as the size of the sample topology increases. The results also suggest that (n,n)-

samples are affected more by imperfect alias resolution (an average reduction of 68% in nodes and 74% in edges). (k,m)-samples, by nature, have fewer routers with alias resolution problem due to the tree-oriented nature of the traces.

Degree Distribution: Degree distribution has been used to characterize network topologies and several topology generators use this characteristic to generate synthetic topologies. In our experiments, we observe that degree distribution changes with the changing success rate of the alias resolution process, but different effects are observed with different samples. For the power-law based graph samples, i.e., BA-based samples, imperfect alias resolution distorts the power-law characteristic of the distributions. For TS- and WA-based samples, the alias resolution process has different types of impacts especially at low degree (3-13) or high degree (20 and up) ranges, respectively.

Joint Degree Distribution: Joint Degree Distribution (JDD) characterizes the degree relation of nodes, i.e., it reports the probability that a node of degree k and a node of degree k' are connected. *Assortative coefficient* is a summary statistic of JDD and it measures the tendency of a network to connect nodes of the same or different degrees. Positive values indicate assortativity, i.e., most of the links are between similar degree nodes and negative values indicate dis-assortativity, i.e., most of the links are between dissimilar degree nodes. In our experiments, assortativity of the topologies changes drastically with an increase in alias resolution success rate in most of the samples. For instance, (n,n)-samples from the BA graph seem to be assortative with 0% alias resolution. With 100% alias resolution, they appear to be non-assortative. In general, (k,m)-samples do not show such big differences except for TS-based samples.

Characteristic Path Length: Characteristic path length (CPL) measures the average of the shortest path lengths between all node pairs in a network. In all of the sample topologies, CPL values reduce with the increasing alias resolution success rate. The average reduction for BA, and WA-based sample topologies is about 30%. For TS-based samples, we do not observe much changes. This is possibly due to the fact that the TS graph is a hierarchical graph and the shortest path lengths are not affected by the alias resolution process.

Betweenness: Betweenness is a mesure of centrality. It reports the total number of shortest paths that pass through node v. Usually betweenness is normalized with the maximum possible value, i.e, $n(n-1)$ where n is the number of nodes. We analyze betweenness distribution and observe considerable changes with the increasing alias resolution success rate. The average betweenness reduces with an improvement in alias resolution success rate. On the other hand, the normalized betweenness increases as the alias resolution success rate increases.

Clustering: Clustering characterizes the density of the connections in the neighborhood of a node. We analyze clustering distribution with respect to node degree and observe an increase with increasing alias resolution success rate. Clustering coefficient, a summary metric of clustering, is the ratio of the number of triangles to the number of triplets. In experiments, all samples yield a clustering coefficient of 0 with 0% alias resolution success rate and always increases with the increasing alias resolution success rate except for a single case.