A Simple Conceptual Model for the Internet Topology

Sudhir L. Tauro, U.C. Riverside, Dept. of Comp. Science
Christopher Palmer, C.M.U., Dept. of Comp. Science
Georgos Siganos, U.C.Riverside, Dept. of Comp. Science
Michalis Faloutsos, U.C.Riverside, Dept. of Comp. Science

***Abstract* - In this paper, we develop a conceptual visual model for the Internet inter-domain topology. Recently, power-laws were used to describe the topology concisely. Despite their success, the power-laws do not help us visualize the topology i.e. draw the topology on paper by hand. In this paper, we deal with the following questions:**

- **Can we identify a hierarchy in the Internet?**
- **How can I represent the network in an abstract graphical way?**

The focus of this paper is threefold. First, we characterize nodes using three metrics of topological "importance"', which we later use to identify a sense of hierarchy. Second, we identify some new topological properties. We then find that the Internet has a highly connected core and identify layers of nodes in decreasing importance surrounding the core. Finally, we show that our observations suggest an intuitive model. The topology can be seen as a jellyfish, where the core is in the middle of the cap, and one-degree nodes form its legs. *

1. Introduction

In this paper, we examine several topological properties of the Internet topology at the Autonomous system (AS level) and synthesize them in an intuitive conceptual model. Our goal is to facilitate researchers in visualizing the topology. We want a model that a human can draw on paper. We believe that such a conceptual representation can help researchers approach the complexity of the topology and lead to a better more intuitive understanding. Note that we will use topology to refer to the AS graph unless otherwise specified.

The networking community does not have a simple conceptual model of the Internet topology despite the recent attention that topology modeling has attracted. First, the topology is large and complex. Despite the recent measurement studies, we do not know which properties to look for and how to quantify them [10][5]. Second, we cannot define hierarchy in a straightforward way, although the Internet is assumed to be hierarchical by construction. It is too densely connected for an obvious hierarchy. Third, several efforts to visualize the topology have been made [11][9], but they attempt to show all the available information. We find that these visual models are overwhelming for a human. Therefore, there is a need for a high-level simple to understand model that will hide the overwhelming details.

The contribution of this paper is threefold. First, we suggest three metrics for the importance of a node. We later use these metrics to identify a sense of hierarchy in the network. Second, we identify several properties that we use to create our model. Third, we integrate our observations in a conceptual topological model. The main results of our work can be summarized in the following points:

- The Internet has a core of nodes that form a clique and this clique is located in the "middle" of the network.
- The topological importance of the nodes decreases as we move away from the center.
- The distribution of the one-degree nodes across the network follows a power-law.
- The network is very sensitive to failures of the important nodes, while it is insensitive to random node failures
- The Internet topology can be visualized as a jellyfish. The value of the model lies in its simplicity and its ability to represent graphically important topological properties.

The rest of this paper is structured as follows. In section 2, we present definitions and previous work. Section 3 explains the node metrics used to classify a node according to importance. In section 4 we present three important topological properties of the Internet. In section 5 we develop and present a conceptual model for the Internet topology. We conclude our work in section 6.

2. Background

We study the topology of the Internet at the inter-domain or Autonomous Systems level. The network is represented by a graph with each node representing a domain and each edge representing an inter-domain interconnection.

Metrics: We use the following standard graph definitions. The degree of a node is defined as the number of edges incident on the node. The distance between two nodes is the number of edges on a shortest path between the two nodes. The rank of a node is its index in the order of decreasing degree.

Recall that a power-law is an expression of the form $y \propto x^a$, where a is a constant, x and y are measures of interest and $\propto$ stands for "proportional to".

Previous Work: Faloutsos et al. [1] studied the Internet topology and identified several power laws that concisely describe skewed distributions of graph properties such as the node degree. Govindan and Reddy [3] study the growth of the inter-domain topology of the Internet. They however classify nodes into four classes based on degree and not according to importance of the node. Gao [14] classifies nodes according to their AS relationships, we however focus on topology. Pansiot and Grad [5] study the topology of the Internet in 1995 at the router level. Barabasi et al. [4] explore the fault tolerance of the network using the diameter as a metric. Topology generators have been developed for simulation purposes, which create topologies from scratch [6][7][8].

* This material is based upon work supported by the National Science foundation under Grant No. 9985195, DARPA award N660001-00-1-8936, TCS Inc., and DIMI matching fund DIM00-10071.

Some of the more recent generators make graphs that obey the observed power-laws. However, all this previous work does not help us visualize the topology in an abstract high-level way.

Real Graphs: We use three instances of the inter-domain Internet topology from the end of 1997 until the middle of 2000, which correspond to approximately three yearly intervals. The National Laboratory for Applied Network Research [9] provided the data.

1. Int-11-97: 3015 nodes and 5156 edges.
2. Int-10-98: 5896 nodes and 11424 edges.
3. Int-10-99: 7864 nodes and 15713 edges.

3. The Topological Importance of a Node

In this section, we present three metrics that capture the topological importance of a node. We will later use these metrics to define a hierarchy.

Degree of a node is the number of incident edges of a node as we have already mentioned. The higher the degree, the higher the importance of the node.

Effective Eccentricity ecc (v) of node v is the minimum number of hops required to reach at least 90% of the nodes that are reachable from that node. For connected graphs each node reaches all other nodes. The lower the eccentricity, the higher the importance of the node. The eccentricity has already been used successfully to analyze graphs [12].

Significance of a node intuitively, captures not only how many but how important are the neighbors of a given node. The definition is recursive, and can be calculated by a recursive algorithm. Initially, all nodes have equal significance. At each step, the significance of each node is set to the sum of the significance of its neighbors. Then all values are normalized so that their sum is one. We stop when the significance converges. In our experiments, the convergence was quite fast.

A similar definition of significance is used by Kleinberg [2] for graphs representing the connectivity of WebPages.

Note that in this paper we use the term Significance as defined above, while we use the term importance to refer to all three-node metrics.

In our effort to explore the meaning of these metrics we make the following observations. First we observe that the effective eccentricity of adjacent nodes cannot differ by more than one.

Lemma 1. Let G=(V, E) be a connected undirected graph and (u, v) an edge in E, then the effective eccentricity of node u ecc (u), is bounded by:

$ecc(u) \leq ecc(v) + 1$ or $|ecc(u) - ecc(v)| \leq 1$

Intuitively this lemma tells us that the difference of one corresponds to the case where we have a “center” and the node having lower eccentricity is closer to the center than the adjacent node. For example all one-degree nodes have an eccentricity that is one more than their adjacent nodes. This observation will help us evaluate the model we develop.

Second we compare effective eccentricity and Significance. Figure 1 shows a plot between Significance (y-axis) and effective eccentricity (x-axis). We observe that significant nodes have low effective eccentricity indicating very good correlation between Significance and effective eccentricity.

4. Topological properties of the Internet

In this section, we study novel topological properties of the Internet. First we study nodes of degree one and find that their distribution follows a power-law. Second, we look at alternate paths between a pair of nodes. We find that we can approximate the relationship between number of paths for a pair of nodes and the length of the paths by a power-law. Third we study the robustness of the network and we find that the network is vulnerable to targeted failures but is robust to random failures.

4.1 Distribution of One-Degree Nodes

Here we give a power-law that states that the number of one-degree nodes directly connected to a particular node is related to the rank of the number of one-degree nodes hanging from that node. We sort the nodes in decreasing order of one-degree d_v and plot the (r_v, o_v) pairs on log scale (r_v is the rank of the node, o_v stands for the number of one degree nodes directly connected to a node v). The results are shown in figure 2. The scale is double logarithmic with the y-axis showing the number of one-degree nodes connected to a particular node and the x-axis showing the rank of that node. This leads us to the following power-law.

Power Law 1 : The number of one degree nodes (o_v) connected to a node v is proportional to the rank of the node r_v (in order of decreasing one-degree nodes connected to a node) to the power of a constant R

$$o_v \;\alpha\; r_v^{R}$$

The correlation coefficients are good ranging from 97.69 to 98.32. Intuitively this tells us that there does not exist a connectivity scheme via which the highest degree nodes only connect to other high degree nodes and so on with the one-degree nodes connecting to the boundary of the network. Rather, the one-degree nodes connect to all types of nodes. Note that the rank we use here is different from the rank used in the power law for degree [1] i.e. the number of one-degree nodes is not a straightforward percentage of the degree of the node [13].

4.2 The Length of the Alternate Paths

In this section, we give a power-law that approximates the relationship between the number of node disjoint paths and path length between a pair of nodes. Figure 3 shows the relationship between the RCDF (Reverse Cumulative Distribution Function) distribution of node disjoint paths v/s the path length for a pair of nodes. We ignore edges (paths of length 1). Note that this approximation is more interesting for pairs of nodes that have several distinct path lengths.

Approximation Power Law 2: The RCDF distribution of the number of paths Rn_v of length l_v between a pair of nodes is inversely proportional to the length of that path l_v to the power of a constant b.

$$Rn_v \;\alpha\; l_v^{-b}$$

As we will see later, this observation can be attributed to the existence of a super-concentrated center.

4.3 Robustness

We show that the Internet is vulnerable to targeted failures but is robust to random failures. Previous work used diameter as a measure of robustness, however we don't find diameter to be an accurate metric. We use two metrics pairs of nodes and largest connected component as a more accurate measure of robustness. Here we present results using the largest connected component. Figure 4 shows the graph of random v/s targeted removal of nodes. We observe that the connectivity is good when nodes are removed at random but connectivity suffers when the nodes of higher degree are removed in order. This behavior occurs because the heterogeneous distribution of nodes in the network. A significant percentage of nodes of nodes have a degree of one or two and therefore there is a high probability that a random selection can choose low degree nodes. Therefore, the connectivity of the network is not affected. The network is however held together by a few highly connected nodes. When these nodes are removed the connectivity is affected.

5. A Conceptual Model for the Internet Topology

In this section, we develop a conceptual model for the Internet topology. First, we use the metrics for the importance of a node, and we define a sense of loose hierarchy. Then, we use this hierarchy and our other topological observations to develop a simple conceptual model for the Internet topology.

5.1 Hierarchy through node Classification

Our first goal is to examine whether there exists a central point in the network. We observe that the highest degree nodes are strongly connected forming a clique. We define this clique to be our core. In other words, the core is the maximal clique that contains the highest-degree node. Then, we define the first layer to contain all the nodes that are neighbors of the core. Similarly, we define layer two to be the neighbors of layer one except for the core. By repeating this procedure, we identify five layers. Table 1 shows the distribution of the nodes for the three Internet instances. We can trivially refer to the core as layer-0.

	Int-11-97	Int-10-98	Int-10-99
Core/Layer-0	8	9	13
Layer-1	1354	2491	3628
Layer-2	1202	2440	3055
Layer-3	396	843	1077
Layer-4	43	108	81
Layer-5	12	5	10

Table 1: Distribution of nodes in layers.

We now show that this classification is meaningful. We use our metrics to show that each layer differs in importance. Figure 5 shows the natural logarithm of the average degree distribution, effective eccentricity and Significance (*100 for ease of viewing) for all the layers. All metrics suggest that the importance of the nodes of each layer decrease as we move away from the core. Note that for the average degree distribution the scale is logarithmic so the drop from 5.5 to 1.2 from core to layer-1 is fairly significant. We also see that the average effective eccentricity increases as we go away from the core and this increase is nearly linear (with slope 1). The increase in effective eccentricity of approximately one per layer indicates that each layer is approximately one link further away from the "center" of the network as suggested by lemma 1. Intuitively, nodes at the layers need to go through the core for the majority of their shortest path connections. This strongly suggests that our selection of the core is successful. Next we find that the average Significance of the layers decreases rapidly as we go away from the core which has an average significance of 21.4 followed by layer 1 with only 1.58 and big decreases after that for the other layers. This indicates that our layers manage to cluster the nodes according to their Significance.

5.2 The Internet Topology as a jellyfish

In this section, we integrate all previous observations to create a simple conceptual model of the topology. We visualize the Internet topology as a jellyfish.

Intuitively, the core is the center of the cap of the jellyfish surrounded by layers of nodes that we call shells. Figure 6 shows a graphical illustration of this model. The one–degree nodes connected to each such shell is shown hanging forming the legs of the jellyfish. We make the length of the legs longer to graphically represent the concentration of one-degree nodes for each shell (number of one-degree nodes per node of the shell). We can color each shell according to its importance, and this way add more topological information to the model.

More formally, we define the core or Shell-0 as before. Then, we define the one-degree nodes that are attached to the core as hanging nodes of level zero or Hang-0. We define the neighbors of the core except the Hang-0 nodes to be the next shell Shell-1. Hang-1 has the one-degree nodes that are attached to Shell-1 and so on. Table 2 shows the size of each group of nodes in our new classification.

	Int-11-97	Int-10-98	Int-10-99
Core/Shell-0	8	9	13
Hang-0	465	514	808
Shell-1	889	1977	2820
Hang-1	623	1022	1243
Shell-2	579	1418	1812
Hang-2	299	526	683
Shell-3	97	317	394
Hang-3	41	95	67
Shell-4	2	13	14
Hang-4	12	5	10

Table 2: Distribution of nodes in Hanging layers and shells

It is easy to see that there is a clear correspondence between this classification and the previous one. Namely:

Layer-k = Shell-k + Hang- (k-1)

From table 2 we can calculate the average number of one-degree nodes per node for each shell. (For e.g. Int-11-97 has an average number of 58.125 for shell-0, 1.42 for shell-1, 1.93-for shell-2, 2.36 for shell-3 and 6 for shell-4).

From this we can make the following two observations.

- One-degree nodes are 40-45% of the network.
- The concentration of hanging one-degree nodes is the highest in the core, which is something we expected from Power-law-1.

In this representation, we separate the one-degree nodes and classify them as "hanging". This discrimination is justified if we think that the one-degree nodes are useless in terms of connectivity. They are "dead-ends" and do not provide any value to the rest of the network. In contrast, even a two-degree node may be useful as it can reduce the distances between other nodes.

6. Conclusion

The goal of this paper is to develop a simple and intuitive topological model for the inter-domain Internet topology. Our effort had three components. First, we introduce metrics to quantify the importance of nodes. Second, we identify some new topological properties. Third, we analyze and model the topology.

- We define a notion of loose hierarchy in the network. We show that the Internet has a highly connected core and identify layers of nodes in decreasing importance surrounding the core.
- We show that our observations suggest an intuitive model. The topology can be seen as a jellyfish, where the core is in the middle of the cap, surrounded by layers of nodes of decreasing importance. Finally, the one-degree nodes form the legs of the jellyfish.

Why is the "jellyfish" a good model? Apart from its apparent cuteness, our model provides a useful visual representation of the topology. It illustrates several topological properties and can help us explain empirical observations.

- The topology has a core, which is represented, by the center of the jellyfish cap.
- There is a gradual reduction in node importance as we move further away from the core. We can illustrate this by using appropriate coloring of each shell i.e. from lighter to darker.
- The middle long legs and the decreasing length of the subsequent legs (of the hanging nodes) represent the observed concentration of one-degree nodes i.e. our power-law 1.
- The network is robust to random failures and the model provides an intuitive explanation for this. There is a high probability that we select pick one of the hanging nodes, as they account for approximately 40% or some node in the outside shells. Therefore, connectivity is not affected.
- Focused failures are devastating for the network connectivity. This case corresponds to the removal of nodes starting from the core and then each shell in order of importance.
- Our model provides an intuitive explanation for the approximation power-law 2. Connectivity exists in or towards the center, which is densely connected, so we do not find many long wondering paths.

We would like to stress again the intent of our Internet model. It is supposed to provide an intuitive visual representation and not an accurate mathematical model. In addition, the topological properties presented in this paper have independent stand-alone value.

References

[1] Michalis Faloutsos, Petros Faloutsos, Christos Faloutsos. On Power-Law Relationships of the Internet Topology. ACM SIGCOMM'99.

[2] David Gibson, Jon Kleinberg, Prabhakar Raghavan, Inferring Web Communities from Link Topology, ACM Conference on Hypertext and Hypermedia, 1998.

[3] R. Govindan, A. Reddy. An analysis of Internet inter-domain topology and route stability. IEEE INFOCOM, April 7-11 1997.

[4] R. Albert, H. Jeong, and A. -L. Barabási, Attack and error tolerance in complex networks, Nature, 406, 387-482, 2000.

[5] J. -J. Pansiot and D.Grad. On routes and multicast trees in the Internet. ACM Computer Communication Review, 28(1): 41-50,January 1998.

[6] S. Jamin and C. Jin and Y. Jin and D. Raz and Y. Shavitt and L. Zhang. On the Placement of Internet Instrumentation. In *Proceedings of IEEE INFOCOM,* Tel Aviv, Israel, March 2000.

[7] M. Doar. A Better Model for Generating Test Networks. In *Proceedings of Global Internet*, November 1996.

[8] A. Barabasi and R. Albert. Emergence of Scaling in Random Networks. In *Science*, vol. 286, 509-512, October 1999.

[9] National Laboratory for Applied Network Research. Routing data. Supported by NSF, http://moat.nlanr.net/Routing/rawdata/.

[10] R. Govindan and H. Tangmunarunkit. Heuristics for Internet Map Discovery. INFOCOM, March 2000.

[11] Bill Cheswick, Hal Burch. Wired Magazine, December 1998.

[12] C. Palmer, P. Gibbons and C. Faloutsos. A fast approximation for the Neighborhood function for massive graphs (under review)

[13] Sudhir L. Tauro, Masters thesis, U.C.Riverside 2001.

[14] Lixin Gao, On inferring Autonomous system relationships in the Internet, IEEE Globe Internet, Nov 2000.

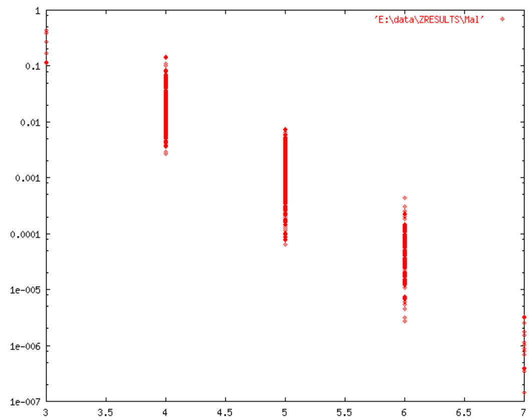

Figure 1: Plot of Significance (y-axis) v/s Effective Eccentricity (x-axis). (Int-11-97)

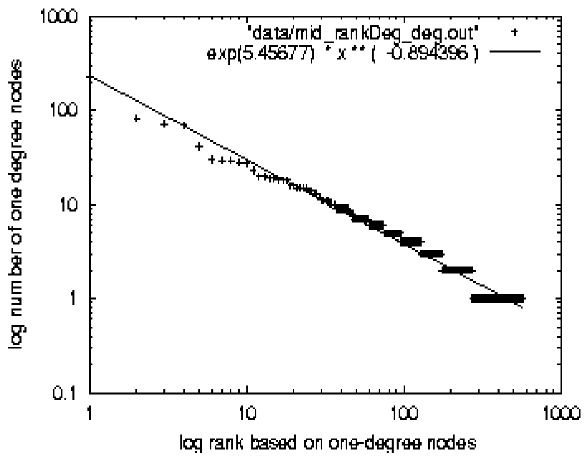

Figure 2: Log-Log plot of one-degree nodes connected to a node v/s the rank of that node based on one-degree. (Int-10-99,Correlation coefficient = 98.07%)

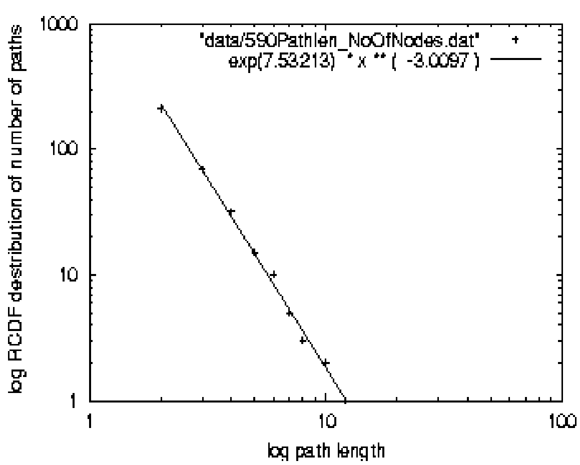

Figure 3 : Log-Log plot of the RCDF distribution of the number of paths v/s path length for a node (Int-11-97, degree 590-524, Correlation coefficient = 99.8%)

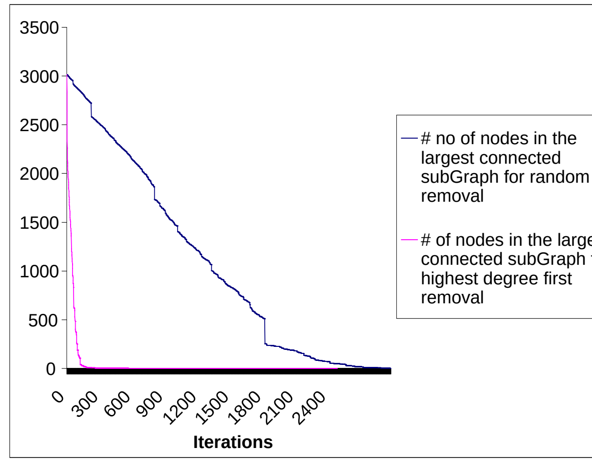

Figure 4: Random v/s Targeted (Int-11-97)

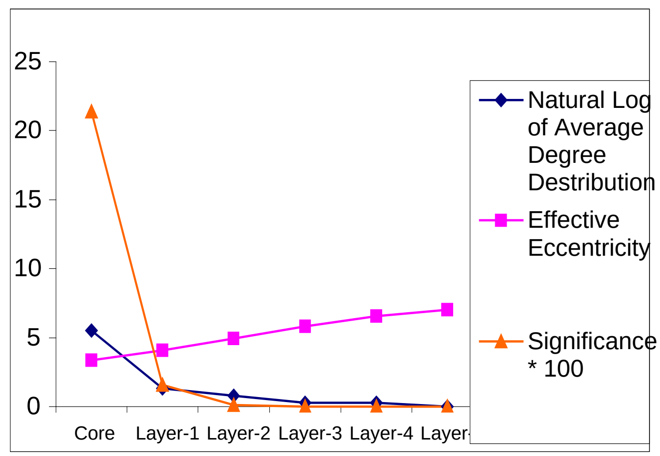

Figure 5 Plot of the average importance of each layer: natural log of the average degree, average effective eccentricity and average Significance (*100 for ease of viewing)

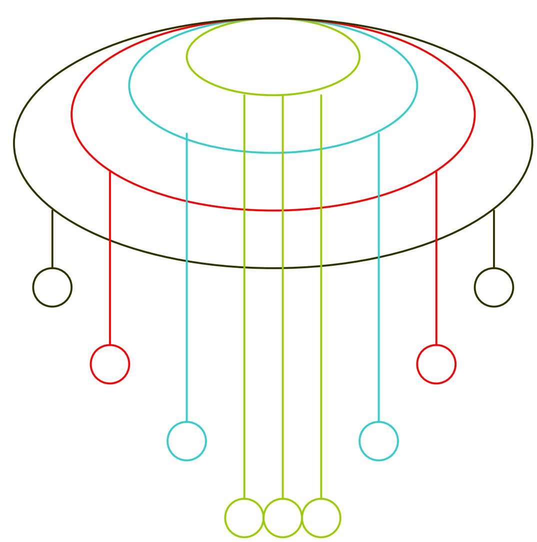

Figure 6: The jellyfish as a model for the AS Internet topology.