The earliest Internet model at AS level was stochastic model, e.g. Waxman model [1], where a set of nodes are distributed in a plane uniformly at random, and then a link is added between each pair of nodes with a constant or a varying probability. Transit-Stub model [2] with three level domains are constructed to reflect the hierarchical domain structure and locality presented in the Internet. Later the studies reveal that real Internet topology is neither completely random nor completely regular (hierarchical), but with a prominent characteristic which the degree distribution of nodes follows a power-law form. Internet topology modeling steps into a quantified attributes modeling stage. Many grow and preferential attachment models have been proposed, such as BA model [3], GLP model [4], PFP model [5] and MLW model [6]. They closely mimic some of the characteristics of Internet.

Average node degree, average clustering coefficient, and average node distance are vary from time to time. Those models can only show the network structure at a given moment while it can't illustrate actual Internet dynamic growing process.

Besides, a local-world node deleting evolving network model [7] is proposed focusing on node adding, local preferential attachment, and node deleting. A model under IPv6 also has been studied [8].
To mimic the real Internet evolving, we propose a time varying evolving Internet topology model which takes advantage of the thoughts in multi-layer and preferential attachment models. By comparing the simulation network properties with statistical Internet ones in years, the model precisely describes Internet dynamic node growing process with adjustable parameters. Any simulated Internet network with nodes between 3000 and 22000 can be generated for other researches.

A. Thoughts and Settings for Modeling

At AS level, where each AS is approximately mapped to an Internet Service Provider, Internet has backbone connection and multi-level local connection. A new AS usually only connects to ASs in the same region which makes local preferential connection. This results in a tighter local connection. Moreover, a new AS intends to connect to a high-degree AS to minimize the communication delay instead of randomly connect to any AS. Instead of using static growth rate during the whole modeling process, we gradually adjust each operation's probability according to a growth rate function to simulate the real growing process of Internet. At any step of the modeling process, the network topology is a simulated Internet topology for the nodes at that moment. Let $G=(V, E)$ denote the network topology, which is an undirected graph with nodes set $V$ and link set $E$ (a set of links connecting all nodes in $V$). We define two types of networks, WAN and LAN, which represent two different layers. WAN is a network which constituted of geographical adjacent nodes. As usual, the links within WAN are more than those between WANs.

For any node $V_i \in V$, let $Q_i$ denote a node set for all adjacent nodes of $V_i$.

\[ Q_i = \{ v \mid w \in E, v \in V_i \} \]

For any node $v \in V$, let $Q_v$ denote a node set for all adjacent nodes of any node in $Q_i$.

\[ Q_v = \{ w \mid v \in E, v \in V_v, v \in Q_i \} \]

Both $Q_i$ and $Q_v$ are node set of LANs.

B. Modeling Process

The time-varying evolving model is generated by the following scheme. Start with no isolated WANs. Each WAN has mo nodes and eo links without isolated nodes. At each step, perform one of the following four operations at random:

1) With probability $p$, a new node is added to a randomly selected WAN $r$.

And then a node in $r$ is chosen with a probability given by to connect to the new node.

\[ \prod(k_i) = \sum_k k_i \]  

A. Properties for Evaluating Comparison

We provide a comparison of several properties. In particular, we consider the node degree distribution, average clustering coefficient, and average node distance. These quantities will provide a preliminary image of the network. 1) Node Degree Distribution

Node degree distribution function $P(k)$ is defined as the proportion of nodes with degree $k$ in all nodes. It shows the possibility of getting a node with degree $k$ from the network. Network average degree is the average degree of the sum of all node degree.

The scale-free nature of the Internet has been pointed out by inspecting the node degree probability distribution, and it implies that the fluctuations around the average connectivity are not bounded. Internet node degree distribution $P(k)$ matches power-law distribution, i.e. $P(k) \propto k^{-Y}$ The exponent is believed to between 2 and 3. In order to provide a full characterization of the scale-free properties of the Internet, we analyze the distributions for different time snapshots of the Internet topologies. These distributions exhibit an algebraic behavior and are characterized by exponent that is stationary in time.
2) **Average Clustering Coefficient**

Average clustering coefficient is a parameter for describing the network cluster characteristic from global view and the distances between a node's adjacent nodes from local view. Let $C$ denote clustering coefficient of a node $V_i$ with degree $k_i$: $C = \frac{2E_i}{k_i(k_i+1)}$ (3)

$E_i$ is the total links between adjacent nodes of $V_i$.

Let $C$ denote network average clustering coefficient, $C = \frac{\sum C_i}{N}$ (4) $N$ is the total nodes in the network.

Research [9] shows that Internet has fairly high average clustering coefficient and current router policies are the reason.

**B. Parameters Selection**

1) **Exponent selection**

The elegance and simplicity of the power laws provide a novel perspective of Internet structure. Node degree

**CASE B:**

distribution $P(k)$ is in proportion to $k^{-y}$

According to the statistical analysis [10], the evolution of the power-law exponents keeps stable for years which is fluctuating round 2.2. In our model, exponent $y$ is fixed at 2.2 during the simulation process for simplicity.

2) **Operations probability selection**

According to the statistical result (collected form [11] and related resources) of $(k, C, d)$ from 1997 to 2005, we compute the $(p, q, r, s)$ combination respectively for each year with fixed growth rate, i.e. without changing $(p, q, r, s)$ during the evolving process. A fitted curve for rate change during this period is generated for growth rate equation for $(p, q, r, s)$. 3) Initial values selection

For the selection of $no$, $mo$, and $eo$, any small numbers can be applied. At least a few thousand nodes in Internet simulated network, the probability selection during the modeling process will eliminate the effect of the original small number.

We compare the simulation results with $no = 4$, $no = 6$, and $no = 8$. There is no obvious difference between the final network topologies. For $mo= 4$, $mo= 6$, and $mo= 8$, we have similar conclusion.

For the selection of $eo$, it should not be too big since a link can't be added to a complete graph. Otherwise as $no$ and $mo$, any small $eo$ can do.

2) **Power-law in node degree distribution**

Power law characteristic. When we randomly select a number between 3000 and 22000 and take a look at the node degree distribution, the image looks alike.

Many low-degree nodes and a few high-degree nodes show that it's a scale-free network which is robust in random network breakdown and vulnerable in intentional attack.

3) **Hierarchy and locality**

In the way the model is building, WAN and LAN give the network a hierarchical structure while LAN reflects the local characteristic. The model is born with hierarchy and locality which fit the Internet characteristics well [13].

**2. CONCLUSION AND DISCUSSION**

We have shown a time-varying evolving Internet topology model at AS level and analyse its properties by comparing it with statistical data. This model takes dynamic growth rate, preferential attachment, hierarchical nodes and geographical location into account and captures scale-free node degree distribution, high clustering coefficient and small average node distance.
properties during the simulations. As we repeat the modeling process times and times again, the model stably shows similar topology properties. Given any node number between 3000 and 22000, we can build a simulated Internet topology by running this model until the specific nodes is reached. This help the researches on Internet communication protocols, transmission mechanism, virus spreading, and congestion control under a realistic network topology.

Most of the data we collected are based on NLANR Project which aimed to give a picture of global ISP interconnectivity by AS number. Since NLANR project had been suspended for some time, our test data are out of date. The rates we set in our model can only work for the node range as stated before. The dramatic increase in 2005's data leaves this model is not a good one for prediction. We need further research to obtain the new data to build a better model for predicting the future topology of Internet.

REFERENCES:


