Improving Chinese Internet’s Resilience through Degree Rank Based Overlay Relays Placement

Bin Yuan$^{1,2}$, Guoqiang Zhang$^{1,2}$, Yanjun Li$^1$, Guoqing Zhang$^1$, Zhongcheng Li$^1$

$^1$Institute of Computing Technology, Chinese Academy of Sciences, Beijing, 100080, P.R.China
$^2$Graduate University of Chinese Academy of Science, Beijing, 100049, P.R.China
Email: {yuanbin, guoqiang, liyanjun, gqzhang, zcli}@ict.ac.cn

Abstract—The current interdomain routing protocol (BGP) used in the Internet tends to be restrictive limiting communication between source-destination pairs to one route, which may not fully utilize the potential Internet’s connection redundancy. Although overlay routing could utilize redundant communication paths between endpoints to improve reliability and performance of the Internet, the studies of overlay network still suffer two handicaps: (i) lack of accurate Internet topology model which can be used for the large-scale overlay simulation eligibly and (ii) most existing overlay systems are inefficient in choosing the relay nodes to create the disjoint overlay path.

In this paper, we first present an experimental study of resilience analysis on the Chinese Internet which we obtained through a large-scale active measurement recently and contains greater connectivity than what Skitter and Routeviews have discovered. Then we propose a heuristic strategy, called Degree Rank based Overlay relays Placement (DROP) to create the redundant disjoint overlay path between each source-destination pair. Simulations show that DROP can improve the entire Chinese Internet’s resilience efficiently with less topological information acquired and can be put into practice easily.

I. INTRODUCTION

Internet itself is composed of thousands of independent Autonomous Systems (ASes), which form a hierarchical structure, of which the core tier are well connected to each other. This structure results in good connectivity and high route diversity in the Internet topology, but the de facto inter-domain routing protocol BGP [1] cannot fully utilize it to provide satisfactory routing service.

The Internet seems to work most of the time but sometimes recovery from failures is painfully slow, because BGP is known to suffer from slow convergence and instability due to a wide-range of causes [2]. The development of overlay routing techniques was a systemic approach to counter these major shortcomings of BGP. Overlay networks attempt to leverage the inherent redundancy of the Internet’s underlying routing infrastructure to detour packets along an alternate overlay path when the given primary path becomes unavailable or suffers from congestion, however the studies of overlay network still suffer two important technical issues:

1) There is no accurate Internet topology model can be used as the simulation testbed eligibly. Accurate Internet topologies should be the basis for the large-scale overlay research activities, however, as a constantly evolving and distributedly administered system, obtaining an accurate view of it is not an easy task. Most prior works take the simulation on the basis of synthetic topologies, such as GT-ITM [3], BRITE [4], which are not based on the policy routing and not reflect the actual redundancy of Internet [5], or of limited realistic topology data getting from some public routing data site with poor connectivity due to the inadequate collecting points. These obstacles will inevitably decrease the simulation result’s credibility.

2) The most existing overlay systems are inefficient in choosing or placing relay nodes to create the disjoint overlay path. Although the effectiveness of overlay networks depends on the natural diversity of overlay paths between two endhosts in terms of physical links and routing infrastructure, there has been very little work done in placing proper relay nodes in overlay network design with suitable topological information.

To achieve a more accurate Internet topology data, we undertake a large-scale measurement of one subgraph of the Internet—the Chinese Internet, and obtain the Chinese Internet AS graph (Trace-CN) out of traceroute results, which contains more links per node as compared with the Chinese AS graph derived from the Skitter [6] and Routeviews [7]. We choose the Chinese subgraph because the Chinese AS graph itself is a representative regional subgraph of the Internet, and the peering ASes of Routeviews’ collectors are all out of China, which gives us a chance to focus on a representative snapshot of the real Chinese Internet AS graph. Due to limitations and resource constraints, we can only acquire the Chinese subgraph of Internet. However as our previous study points out, the Internet subgraph can show self-similarity with the global Internet [8]. Therefore we are quite sure the outcome of this study will be applied to the whole Internet scope.

In this paper, we consider path disjointness at the AS-level. To evaluate the Chinese Internet’s resilience, we introduce a metric, called Internet resilience degree. In order to create the redundant disjoint overlay paths between each source-destination pair through choosing a set of locations for placing overlay relays, we propose a Degree Rank based Overlay...
relays Placement (DROP) strategy, which is based on the intuition and attempts to create overlay paths that are as disjoint as possible from the default path at the AS-level. Simulations on Chinese Internet show that DROP can optimize the entire Chinese Internet's resilience with less topological information obtained and can be easily put into practice.

The rest of the paper is structured as follows. Section II gives a brief introduction to related works. Section III introduces the Chinese AS-level topologies obtained from active measurement and presents the topological discrepancies with other data sources. Section IV presents the Chinese Internet resilience analysis systematically. The heuristic method of DROP and it’s performance evaluation are provided in Section 3. Finally, we conclude our work in Section VI.

II. RELATED WORK

Two state-of-art methodologies to achieve Internet topology awareness are widely used within the research communities: one is collecting the BGP tables and updates, the other is mapping the actively measured traceroute IP paths to AS paths. However, these two methods both suffer from some inherent limitations and resource constraints. The most prestigious active measurement project Skitter [6], deploys dozens of dedicated servers worldwide to collect traceroute results, however, compared with the enormous Internet scale, these servers are apparently inadequate. The DIMES [9] project takes advantage of the end users’ location diversity to address this limitation, but the end users’ behaviors are out of their control. The Routeviews [7] project is a passive measurement project, which sets up several collectors, each peering with dozens of BGP speakers located in different ASes, to collect BGP tables and BGP updates. It also faces the problem of inadequate collecting points.

Redundancy of the Internet has been well-studied by a number of research groups. As mentioned above, overlay routing can leverage this redundancy, creating a redundant path in the overlay framework via overlay relay technology [10]. The relays are placed in a number of important locations in the infrastructure and each relay performs a simple forwarding operation. The basic idea is that the sender sends packets to a relay node, and the relay node then forwards packets to the receiver [11], [12], which forms a 2-hop overlay path. By selecting the relay node appropriately, the packets traveling through the relay node take a different underlying physical path than that of the Internet default path between the sender and receiver. Because a 2-hop overlay path provides nearly the same degree of path diversity for most destinations as the optimal multi-hop overlay path does [11], [13], most overlay researches as well as ours follow it for simplicity, creating overlay path through just one intermediate relay node.

There are two kinds of methodologies in selecting or placing relay node to create the overlay path which has the ability to find good detours in the Internet. One of them is using probing to find alternate paths between sender and receiver [11], [14]. For example, the prestigious RON [11] tries to find the best alternate path to the destination by aggressively sending probes and exchanging path performance information among overlay nodes at very short intervals of the order of a few seconds. The extreme bandwidth requirements result in RON’s scalability problem with a limited size overlay network. The other kind of past studies [13], [15], [16] aim at countering this scalability issue, focusing on topology aware overlay networks which try to leverage enough topological information to minimize the overlap of network elements between overlay path and default path in various level. The constraint is that these systems need to gather topological information accurately and in time, since the routing and topology of Internet are dynamically changing.

III. THE CHINESE AS SUBGRAPH

A. The Data Sources

Recently, we have undertaken a large-scale traceroute based active measurement of the Chinese Internet. We have deployed 18 probes, probing to 8757 IP addresses sampled from the routable IP address blocks and web sites of China. Of the 18 probes, one is located in U.S, the other 17 are distributed in geographically different regions of China, covering 10 ASes. The data was collected during the period of April 21st to May 8th in the year of 2007. We also collected the Routeviews data from routeviews2.oregon-ix.net, which uses 40 peers to collect BGP tables, and the Skitter AS link data available from http://sk-aslinks.caida.org. We collected ten days of data for both the Routeviews and Skitter, from April 21st to April 30th. We extracted the Chinese AS subgraph out from the global graph and merged the ten graphs into one. For convenience, we call these three graphs Trace-CN, Routeviews-CN and Skitter-CN respectively.

B. Elementary Graph Properties

To testify whether Trace-CN captures most part of the Chinese AS graph, we accumulate the AS graphs derived from each probes trace result, in the sequence of each probes residing ASs degree, i.e., we merge the AS graphs derived from 7 probes in AS4134 first, one by one, then the AS graph from AS4808, and etc. We measure the number of nodes and number of links after each merging procedure and plot it in Fig. 1. The diagram apparently shows a diminishing returns after 14 trace results are merged, after which the addition of new probes discover only a few links and contribute little to the final graph.

We present the elementary graph properties in Table I and plot the cumulative degree distribution and rich-club in Fig. 2. The prestigious Routeviews project has no data collection point in China and CAIDA’s Skitter has only one data collection point in China, which will inevitably decrease the data’s credibility. We have found Routeviews-CN is indeed unconnected. The relatively lower average degree, clustering coefficient and maximum coreness are all indications of their poorer connectivity. Although the maximum degrees are similar, the maximum coreness of Routeviews-CN is only half of Trace-CN. We also see that despite Routeviews-CN’s sampling bias, its degree distribution still approximates a power-law shown in Fig. 2(a). However, the rich-club phenomenon which is important because the connectivity between rich nodes can
be crucial for network redundancy and robustness [17] shown in Fig. 2(b) is quite different. The rich-club phenomenon shown by global AS graph is well maintained in the Trace-CN, but is not observed in the Routeviews-CN.

### Table I

BASIC GRAPH PROPERTIES OF THE THREE DATA SOURCES

<table>
<thead>
<tr>
<th></th>
<th>Skitter-CN</th>
<th>Routeviews-CN</th>
<th>Trace-CN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Nodes</td>
<td>75</td>
<td>161</td>
<td>135</td>
</tr>
<tr>
<td>Number of Edges</td>
<td>159</td>
<td>274</td>
<td>338</td>
</tr>
<tr>
<td>Average Degree</td>
<td>4.24</td>
<td>3.40</td>
<td>5.01</td>
</tr>
<tr>
<td>Maximum Degree</td>
<td>20</td>
<td>38</td>
<td>41</td>
</tr>
<tr>
<td>Characteristic Path Length</td>
<td>2.79</td>
<td>+∞</td>
<td>2.75</td>
</tr>
<tr>
<td>Clustering Coefficient</td>
<td>0.179</td>
<td>0.054</td>
<td>0.163</td>
</tr>
<tr>
<td>Mean Clustering</td>
<td>0.291</td>
<td>0.132</td>
<td>0.361</td>
</tr>
<tr>
<td>Maximum Coreness</td>
<td>4</td>
<td>3</td>
<td>6</td>
</tr>
</tbody>
</table>

Based on the above elementary graph properties analysis, we come to the conclusion that our measurement result can be considered as a relative complete snapshot of the real Chinese AS subgraph with greater connectivity than the other two graphs. Therefore, we use Trace-CN as the basis of our simulation, and analyze the Chinese Internet’s resilience on it.

### IV. INTERNET RESILIENCE ANALYSIS

To the best of our knowledge, no attempt has been made so far to quantify the redundancy of the Internet from the perspective of end-to-end. In this section, to evaluate the Chinese Internet’s resilience, we are motivated to introduce a metric of the topology resilience, called Internet Resilience Degree which can be used to quantitative analysis the path redundancy of Internet.

#### A. Definition of Internet Resilience Degree

In each Source-Destination (SD) pair, if there exists at least two disjoint paths connecting these two endpoints, the SD pair can be recognized resilient, and it has the potential ability to carry out path diversity transmission and failure recovery. Since we focus on the topology at AS-level, the SD pair should be considered as the SD AS-pair, and the path between these ASes should be AS path. The disjoint paths mean they do not experience any overlapping at AS-level, namely they do not have any overlapped ASes.

**Definition 1:** According to the above explanation, the Internet Resilience Degree (RD) is defined as the ratio of the number of Source-Destination (SD) AS-pairs in the network which have no less than two disjoint AS paths (which could be one direct path plus either another direct physical path or a 2-hop overlay path) connecting these two ASes to the total number of SD AS-pairs which have at least one existing path connecting them. The RD can be seen as the total path diversity of all SD AS-pairs in the Internet, and its formula is presented as follows:

\[
RD = \frac{\sum_{s \neq d} I\{P_{sd} \geq 2\}}{\sum_{s \neq d} I\{P_{sd} \geq 1\}}
\]

where \(I\{\cdot\}\) is an indicator function with \(I\{P_{sd} \geq 1\} = 1\) if there is at least one path from \(s\) to \(d\) and 0 otherwise, \(I\{P_{sd} \geq 2\} = 1\) if there is at least two disjoint paths from \(s\) to \(d\) and 0 otherwise.

Because of the existing technologies of multihoming and prefix based routing, some SD AS-pairs may have more than one physical direct path defined by the routing protocol. The proportion of these SD AS-pairs among the overall AS-pairs should be an important determinant to evaluate the Internet’s resilience intuitively.

#### B. Path Set

In order to ensure the accurate calculation of RD, we should guarantee the accuracy of AS paths which are used for the calculation. Therefore, we setup a path set which contains not only the observed AS paths but also the extended paths from other ASes along the path, except the source AS, to the destination AS. This is because routing is destination based, so any sub-path can be a real existing path.

Note that if a single-homed stub AS is a source or a destination of SD AS-pair, there is no another direct or 2-hop overlay path can be found with no overlapping with the direct path of this AS-pair, because they must be overlapped at its provider AS. Therefore it is necessary to recursively prune all the vertices of degree smaller than 2, until the degree of all remaining vertices is larger than or equal to 2. After the process of the pruning, we obtain a subgraph with 53 ASes being wiped out. The remaining 79 ASes have degrees no less than 2, and the number of reachable SD AS-pairs in the path set is 1062 (79 × 79 in total), which can be seen as the main part of the Chinese Internet routes.

#### C. Chinese Internet’s Resilience

We characterize the Chinese Internet’s resilience by using our metric presented above. Firstly, we consider the Chinese Internet resilience without taking overlay paths into account. Among 1062 Source-Destination (SD) AS-pairs which have at least one physical direct path, 351 SD AS-pairs have more than one direct path. Therefore according to our definition, the Chinese Internet RD should be 33.1%. It’s quit low because of the de facto BGP protocol as mentioned above.
Obviously, deploying the overlay systems can increase the RD efficiently by creating 2-hop overlay paths disjointing from the default paths for most SD AS-pairs. This raises the question that to what extent overlay systems can improve the resilience. Therefore, secondly we carry out a overlay nodes placing scheme to maximize the Chinese Internet RD, which is to place an intermediate node to the optional relay set of all 79 ASes at will for every SD AS-pair to create a 2-hop overlay path which has as little overlapping with the direct path as possible. For the sake of the efficiency of overlay path, these 2 hops of overlay path are required to completely separate from each other. This scheme causes 863 SD AS-pairs find 2-hop overlay path with no overlapping with the direct path, also called redundant overlay path or resilient overlay path. Accordingly, the RD raises to 81.3%, which is much higher than the RD without overlay deployment above, bringing a significant improvement in path diversity and fault tolerance.

When it is impossible to find a completely disjoint overlay path, we allow overlay path to have a few overlapped ASes with default path. Our experiment shows that there are another 141(13.3%) SD AS-pairs have 1 overlapping AS between direct path and overlay path, 16(1.5%) have 2 overlapping ASes. The uppermost line in Fig. 4 presents this cumulative distribution of overlapped ASes with the relay set of all ASes. Since the result come from the scheme of choosing the best relay ASes from the set of all ASes for every SD AS-pair, the result above can be regard as the standard of maximal Chinese Internet RD. In next section we will show the heuristic strategy of narrowing the scope of candidate relay ASes to achieve close result to the maximal RD.

V. DEGREE RANK BASED OVERLAY RELAYS PLACEMENT

Due to the constraints of overlay systems deployment costs in terms of hardware, network connectivity and human effort, it is critical to carefully place infrastructure overlay nodes to balance the trade-off between performance and resource constraints. Now that we analyze the maximal Chinese Internet RD with overlay systems, the key question is how to narrow the location scope of placing candidate intermediate nodes. In this section, we propose a scalable, heuristic method for ensuring suboptimal RD via choosing a subset of locations for overlay relay nodes placement.

A. Problem Description

The overlay relay node placing problem was described as follows. Given an Internet AS topology or path set, we need to determine the set of AS locations for placing overlay relay nodes to create the overlay paths disjointing from the default paths for as many SD AS-pairs as possible, so as to optimize the Internet RD. Therefore, the objective of the problem is to find a subset \( O \subseteq V \) as small as possible under the premise of approximate optimal RD, where \( V \) is a set of all ASes, a set \( O \) is a subset of ASes that are potential locations for intermediary relays placement.

B. Degree Rank based Overlay Relays Placement

We propose a intuitive heuristic strategy, called Degree Rank based Overlay relays Placement (DROP) to select a subset of ASes for relay nodes placement based on the ASes' degree rank. The degree of ASes represents it's network connectivity. If a AS has a high degree, it should be rich in links and densely connected with other ASes. Intuitively, the AS with more connections should have the priority of being the overlay relay. In our maximal Chinese Internet RD result we find most overlay relay ASes in the overlay paths are the ASes with high degree.

The degree rank based strategy is to sort all ASes by decreasing number of degree firstly. The degree rank denotes the position of a AS on this ordered list. Secondly, we incrementally add the ASes to the candidate relay subset by the degree order to seek the balance between the number ASes to place relay nodes to and derived Internet RD.

C. Experimental Analysis

We investigates this approach to perform optimized placement of infrastructure overlay nodes in Trace-CN, and evaluate it compared with betweenness rank based and random based approaches. The random based approach is a conventional wisdom to randomly place overlay nodes in various ASes.

The betweenness of a vertex \( v \) is defined as the number of shortest paths between pairs of other vertices which run through \( v \). Betweenness in Internet topology at AS-level also reflects the significant extent of ASes.
Overall, for obtaining the Chinese RD our comparison results among 3 approaches are presented by Fig. 3, which shows DROP and the betweenness based approaches outperforms the random approach significantly. And DROP is slightly better than the betweenness based, because the AS with higher betweenness only reflects its essentiality in network nor in possession of numerous connections to other ASes which is essential for it to become the relay AS preferentially. Fig. 3 also shows that random scheme enable RD increasing approximately linearly with the number of ASes in the candidate relay subset to place the overlay nodes to. In contrast, the other two approaches improve the RD dramatically when the number of ASes in the relay subset from 1 to 4, which means the approximate maximal RD can be reach by only choosing a subset of top 4 ASes of degree (which are the core ASes in Chinese) where to place the overlay nodes.

We plot the cumulative distribution function (CDF) of the number of overlapped ASes in different candidate relay subsets in Fig. 4, which shows the disjoint comparison among three kinds of approaches in choosing relay subset. Comparing with the optimal results, DROP and betweenness based approach with only 5 ASes as the relay subset have already achieved the approximate results, even better than random scheme with 30 relay ASes.

Simulations show that DROP can improve the entire Chinese Internet’s resilience efficiently, which remind us just to place overlay nodes in the regional core ASes can achieve enough disjoint overlay paths in regional Internet.

VI. CONCLUSION

In this work, we undertake a large-scale measurement of one subgraph of the Internet—the Chinese Internet firstly, and obtain the Chinese Internet AS graph out of traceroute results containing great connectivity which fits the demand of topology to be the basis of simulation for overlay routing. Then we presented a redundancy analysis of Chinese Internet based on the metric of RD we introduced and analyzed the resilience of Chinese Internet with overlay system. Another contribution of our work is to develop a DROP strategy to addresses the problem of choosing the subset of locations for overlay relay nodes placement based on the ASes degree order with the objective of reaching the approximately optimal resilience degree. Because the Chinese AS graph itself is a representative regional subgraph of the Internet and the Internet has the property of self-similarity, the results in this paper could be applicable in the whole scope of Internet.

REFERENCES