# China's Internet: Topology Mapping and Geolocating

Ye Tian[‡], Ratan Dey[§], Yong Liu[§], Keith W. Ross[§]
[‡]University of Science and Technology of China, Hefei, Anhui 230026, China
[§]Polytechnic Institute of New York University, Brooklyn, NY 11201, USA
yetian@ustc.edu.cn, ratan@cis.poly.edu, yongliu@poly.edu, ross@poly.edu

## ABSTRACT

We perform a large-scale topology mapping and geolocation study for China's Internet. To overcome the limited number of Chinese PlanetLab nodes and looking glass servers, we leverage unique features in China's Internet, including the hierarchical structure of the major ISPs and the abundance of IDC datacenters. Using only 15 vantage points, we design a traceroute scheme that finds significantly more interfaces and links than iPlane with significantly fewer traceroute probes.

We then consider the problem of geolocating router interfaces and end hosts in China. When examining three well-known Chinese geoIP databases, we observe frequent occurrences of null replies and erroneous entries, suggesting that there is significant room for improvement. We develop a heuristic for clustering the interface topology of a hierarchical ISP, and then apply the heuristic to the major Chinese ISPs. We show that the clustering heuristic can geolocate router interfaces with significantly more detail and accuracy than can the existing geoIP databases in isolation. We show that the resulting clusters expose several characteristics of the Chinese Internet, including the major ISPs' provincial structure and the centralized inter-connections among the ISPs. Finally, using the clustering heuristic, we propose a methodology for improving commercial geoIP databases and evaluate using IDC datacenter landmarks.

## 1. INTRODUCTION

China[1] is the country with the largest number of Internet users and the second largest IP address space [5]. Nevertheless, China's Internet has received relatively little attention in the measurement community to date. This is perhaps because China's Internet lacks the infrastructure and resources that are essential for large-scale Internet measurement studies such as those carried out in Rocketfuel [26] and iPlane [14]. For example, China has few PlanetLab nodes and looking glass servers, which are important infrastructure components for large-scale Internet measurement studies. Moreover, whereas many routers outside of China have names from which geolocation can be inferred, few router interfaces have names in China.

Nevertheless, China's Internet is complex and has its unique structural features, which makes it very different from the Internet in US and Europe. China has a very simple AS-topology with few Chinese ASes. However, both of two major ISPs in China each have one giant AS that not only includes a national backbone network, but also includes regional networks in many provinces as well as residential networks. As China's Internet is dominated by few major ISPs, it is therefore largely shaped by the internal structure of these giant ASes rather than the AS-topology.

Of particular interest is geolocation services for China's Internet. More and more online businesses and services – including targeted advertising, spam filtering, and fraud prevention – are based on geolocation of IP addresses. Commercial geoIP databases for China and elsewhere typically incorporate multiple information sources, including information directly from ISPs, DNS reverse lookups, and end user inputs. As we will show in this paper, existing commercial geoIP databases for Chinese IP addresses have many incomplete and erroneous entries, particularly for router interfaces.

In this paper, we carry out a large-scale topology mapping and geolocation study for China's Internet. To overcome the insufficient number of Chinese PlanetLab nodes, looking glass servers, and router interfaces with geographical names, we leverage unique features in China's Internet, including the hierarchical structure of the major ISPs and the abundance of IDC datacenters. The contributions of this paper are as follows:

- We find that existing measurement practices do not adequately cover China's Internet. We develop two techniques, namely *nested IP block partitioning* and *collaborative tracerouting*, which allow us to perform a comprehensive and efficient traceroute measurement study of China's Internet using only 15 internal vantage points. In particular, our approach discovers significantly more interfaces and links than iPlane with significantly fewer traceroute probes.

[1]By China we mean Mainland China.

- Using the IP addresses obtained from our traceroute measurements, we examine three well-known Chinese geoIP databases and MaxMind. We find that the three Chinese geoIP databases are only moderately accurate for end host geolocating, and substantially less accurate for router interfaces. In particular, we observe frequent occurrences of null replies and erroneous entries, suggesting that there is significant room for improvement.

- With the goal of accurately geolocating routers in China, we develop a heuristic for clustering the interface topology of a hierarchical ISP, so that each cluster is a connected component within a city. We then apply the heuristic to the major Chinese ISPs, leveraging the interface topologies derived from our traceroute measurements as well as the existing Chinese geolocation services. We show that this clustering heuristic can geolocate router interfaces with significantly more detailed location information than the existing geoIP databases in isolation.

- We analyze the clusters generated by our clustering heuristic. We show that they expose several characteristics of the Chinese Internet, including recent mergers of ISPs. We observe the provincial capital cities are not only government centers but are also hubs in the ISPs' networks, and inter-ISP traffic is concentrated to a few routers across China.

- Using the geo-clustering heuristic, we propose a methodology for improving commercial geoIP databases. By evaluating with datacenter landmarks, we show that our approach is able to provide more detailed and accurate location information as compared with the original geoIP database, and the methodology can also differentiate the results from geoIP databases with different confidence levels. By improving on the best geoIP databases in China, we are currently providing the most accurate geolocation service for China's Internet.

The reminder part of this paper is organized as follows. Section 2 presents a brief overview of China's Internet. We present our traceroute measurement methodology in Section 3 and compare our results with iPlane. Section 4 evaluates the geolocation services of popular geoIP databases. We propose the geo-clustering heuristic in Section 5, and use the resulting geo-clusters to analyze China's Internet. In Section 6 we develop a methodology for improving geoIP databases. Section 7 discusses related work and we conclude in Section 8.

## 2. OVERVIEW OF CHINA'S INTERNET

Before presenting our methodologies for mapping and geolocating China's Internet, it is useful to briefly overview China's Internet. The two largest ISPs in China are China Telecom (a.k.a. ChinaNet and henceforth referred to as Telecom) and China Unicom (henceforth referred to as Unicom)[2]. Both Telecom and Unicom have high-performance national backbone networks, connecting regional and residential networks in China's provinces and major cities; and both also provide high-performance connections to the Internet outside of China [5]. Both Telecom and Unicom also have their own networks in many provinces and cities in China, and also provide access directly to end users. Unicom holds the dominant market share in the northern provinces, whereas Telecom dominates southern China [28]. The other commercial ISPs in China are much smaller than Telecom and Unicom; they generally rely on Telecom/Unicom's backbone networks for accessing services connected to Telecom/Unicom, and for accessing services on the international Internet.

In addition to Telecom and Unicom, CERNET[3] is also a major ISP in China. As an academic network that connects the universities and research institutes all over China (analogous to Internet2 in the USA), CERNET is largely independent with its own national backbone and peers with many international commercial and academic networks.

Since Telecom and Unicom are government-owned companies, it is not surprising that their topological structure is hierarchical and mimics the provincial organization of the Chinese government. Using Telecom as an example, Telecom Corporation is in charge of Telecom's national backbone network. For each province, Telecom has a provincial subsidiary that manages its provincial network; and for most cities in the province, there is a city company under the provincial company that is responsible for constructing and maintaining residential networks and providing Internet services to end users in that city. In addition, Telecom assigns blocks of IP addresses along this three-level structure, that is, from its address space, it allocates blocks to each of the provincial subsidiaries, and each provincial subsidiary allocates blocks of addresses to cities. In fact, many Chinese geoIP databases exploit this hierarchical IP address assignment to provide basic geolocation services.

Internet Datacenters (IDCs) are widely used in China. An IDC datacenter can provide many services to enterprises and individual customers, including server hosting, server leasing, and virtual servers. Most of the datacenters connect directly to either the Telecom or Unicom backbones, and some connect to both (and di-

[2]Here we are referring to the current China Unicom, which merged with China Netcom (a.k.a. CNCGroup) in 2008.
[3]China Eduction and Research Network

rectly to other Chinese ISPs). Because the datacenters have high-speed connections directly to the backbones, they are attractive for hosting websites and other online services for various customers. In fact, many Chinese university web sites are not hosted on the university campuses, but instead in IDC datacenters; similarly, many government web sites are hosted in IDC datacenters. Furthermore, many popular Web sites in China are mirrored across several datacenters.

## 3. TRACEROUTE MEASUREMENT

Traceroute is one of the most fundamental measurement tools for studying the Internet. Unfortunately, existing large-scale traceroute measurement practices, such as iPlane [14] and CAIDA/Ark [2], do not satisfactorily cover China's Internet. These projects use very few vantage points within China: only two PlanetLab nodes from China are used in iPlane and only one Chinese monitor is used in Ark. As a result, these two projects use vantage points from outside China to collect most of their Chinese traceroute path segments. Moreover, a recent study shows that there are only 24 ASes in China's Internet that peer with foreign ISPs. In fact, it is well known that Telecom and Unicom have most of the international Internet connections in China [5]; therefore most of the traceroute probes originating from outside of China will enter China through a small number of ASes in Telecom and Unicom. *Thus, for traceroutes originating from outside of China, they are likely to follow similar paths when traversing China's Internet, thereby not revealing many diverse interfaces and links.* For comprehensively mapping China's Internet, we must therefore use vantage points located in China.

**Table 1: Distribution of IP block prefix length**

| ⩽18 | 19 | 20 | 21 | 22 | 23 | 24 | ⩾25 |
|---|---|---|---|---|---|---|---|
| 2,793 | 1,197 | 1,987 | 1,210 | 1,175 | 1,133 | 4,404 | 104 |

We face two challenges when attempting to map China's Internet with traceroute. The first is to identify a set of target IP addresses that is sufficiently, but not overly, dense within the Chinese Internet. Large-scale traceroute measurement studies (e.g., [14] and [2]) often use CIDR IP blocks from public BGP snapshots (e.g., from Oregon Routeviews [23] and RIPE RIS [22]); the blocks are used to partition the IP space, and then one address is selected from each block in the partition as the traceroute targets. However, there is no operational public BGP router in China's Internet [27]; therefore, we can only gather Chinese blocks from routers that are outside of China. Because these blocks are likely to have been aggregated by the border routers in China's Internet, they are generally too coarse for topology mapping. To establish this claim, we have downloaded eight BGP snapshots from different routers in Oregon Routeviews and RIPE RIS. (The routers are located in USA, Europe, and Japan.) Table 1 lists the numbers of distinct IP blocks in China with their prefix lengths. We can see that there are many large blocks (e.g., blocks with prefix lengths smaller than 20, 18, and so on).

The other challenge is efficiency, that is, devising a traceroute strategy that sufficiently covers the Chinese Internet without overly burdening the traceroute sources (vantage points). iPlane and Ark spread their workload over hundreds of vantage points. In our traceroute measurements, we only use stable vantage points from within China, for which we have only identified 15 (7 PlanetLab nodes and 8 web-based traceroute servers). If we use iPlane's or Ark's probing strategy, we would overload our 15 vantage points with too many tasks. To address these two challenges, we devise two techniques, namely, *nested IP block partitioning* and *collaborative tracerouting.*

### 3.1 Nested IP Block Partitioning

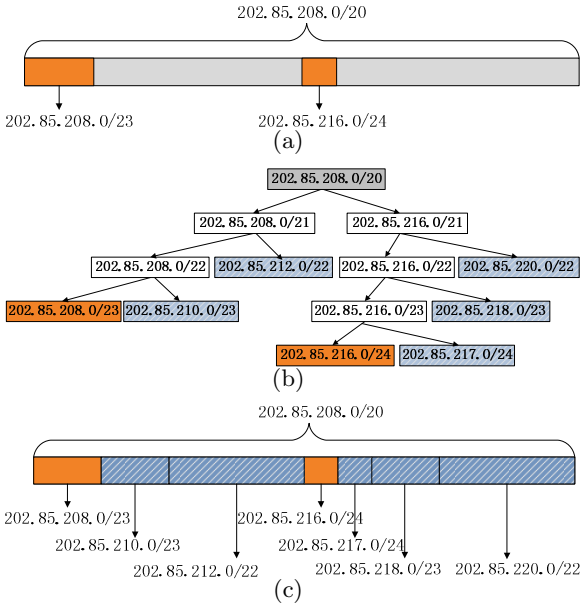



**Figure 1: Nested-block Partitioning**

We need to partition the large Chinese IP address space, and then choose one IP address from each set in the partition as a traceroute target. For the partitioning, a simple approach is to evenly divide the large blocks obtained from the public BGP tables. However, taking a close look at these blocks, we find that *block nesting* [30], where a block from one BGP routing table entry resides in another block from a different entry, is very common; moreover, there are often several levels of nesting. An example of nested IP blocks is shown in Figure 1(a). In the Figure, three blocks are obtained from BGP tables, i.e., 202.85.208.0/20, 202.85.208.0/23, and

202.85.216.0/24, where the latter two blocks are nested in the first one. Clearly, the smaller nested blocks suggest the existence of different subnets, as they appear as separate entries in the routing tables. If we set the granularity of the traceroute probing up to prefix /22, then for the block 202.85.208.0/20, we would obtain four equal-sized /22 blocks, but the smaller nested blocks would be masked. On the other hand, evenly dividing 202.85.208.0/20 into /24 blocks results in 16 blocks, which may overly increase the workload of the measurement.

We design a tree-based method to partition the Chinese IP address space with a minimal number of blocks while preserving the nested blocks obtained from the BGP tables. The blocks from the BGP tables are nodes in trees. We consider a block encompassing other blocks as the root of a binary tree, and all the nested blocks as leaves. With this tree the problem becomes: given the root node and a number of leaf nodes, construct a binary tree with the fewest leaves. After the tree is obtained, we use all the blocks corresponding to the leaf nodes (including the original nested blocks) to replace the root block. For example, for the case in Figure 1(a), the block binary tree is shown in Figure 1(b), and we use seven blocks to replace the original large block 202.85.208.0/20, as shown in Figure 1(c). After partitioning the nested IP blocks, we further evenly divide any blocks that are larger than our granularity, while reserving the smaller blocks for traceroute probing. For the example in Figure 1, if the granularity is prefix /22, then 7 blocks are probed instead of the 4 or 16 blocks that would be generated by evenly dividing. Thus, with nested-block partitioning, we can fully exploit the small nested blocks, suggesting different subnets, without naively dividing all the large blocks into smaller ones, which would geometrically increase the probing workload.

## 3.2 Collaborative Tracerouting

Ark and iPlane apply different strategies to reduce the workload (when probing the entire Internet). In Ark, each /24 block is probed in one measurement round, but Ark groups its vantage points into teams, with each team having a geographically distributed set of members. Each team only probes a subset of the targets. Although a target is only probed by one team, the number and the geographical distribution of the team's vantage points ensure the diversity of the traceroutes. In iPlane, IP blocks from BGP snapshots with similar AS paths are further combined to reduce the workload [18].

We, however, cannot apply either Ark's or iPlane's strategies for two reasons: (*i*) we have only 15 vantage points to spread the workload over; and (*ii*) we need to divide IP blocks from BGP snapshots rather than cluster them. Even after the nested-block partitioning, as described in Section 3.1, there are still 223,714 CIDR blocks in China to be tracerouted. It is impractical to probe each block from each of the vantage points. A recent study [4] shows that there are many redundant probes in Ark and iPlane. We propose a mechanism for having the vantage points collaboratively and dynamically determine their traceroute targets, thereby avoiding redundant probes.

In our measurement, the IP blocks obtained in Section 3.1 (which partition the Chinese IP space) are the basic probe units. When a vantage point probes a block, we always use the second IP address of that block (i.e., a.b.c.1) as the target, as such addresses are usually used for gateways and are, thus, more likely to respond to a probe than other addresses. In our collaborative tracerouting scheme, a vantage point actively uses the results of its previous probes and other vantage points' probes to avoid redundant probes. Specifically, each vantage point keeps a set, *reach_set*, of all the addresses the vantage point has observed during its previous probes; and each IP block keeps a set, *source_set*, containing all the IP addresses that lead to this block from previous traceroutes from all the vantage points. When a vantage point $v$ encounters an IP block $B$ it has not probed before, it examines $v$'s *reach_set* and $B$'s *source_set*; if the two sets overlap, then an interface path can be found from $v$ to the block $B$ from previous traceroutes, so the vantage point $v$ doesn't probe the block $B$.

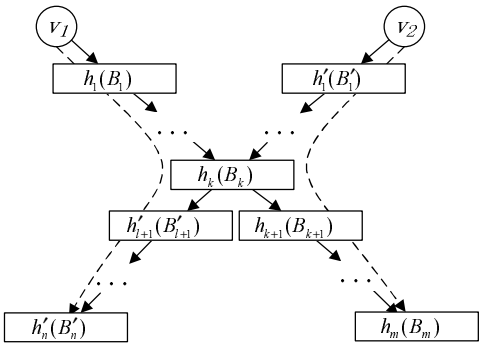



**Figure 2: An example of collaborative tracerouting**

As an example, suppose a vantage point $v_1$ probes a target with the traceroute path

$$h_1, ..., h_k, h_{k+1}, ..., h_m$$

where the interfaces are in the blocks $B_1$, ..., $B_k$, $B_{k+1}$, ..., $B_m$, as shown in Figure 2. $v_1$ inserts all the interface IP addresses it has reached, i.e., $h_1$, ..., $h_m$, into its *reach_set*. For each interface IP, the corresponding block inserts all the IPs preceding it along the path into its *source_set*. For example, $h_1$, ..., $h_{m-1}$ are inserted into $B_m$'s *source_set*. Clearly, after this probing, $v_1$ can skip $B_1$, ..., $B_{m-1}$ in future measurements, as $v_1$'s

*reach_set* overlaps with the *source_set* for each of these blocks. Moreover, suppose another vantage point $v_2$ has a traceroute path

$$h'_1, ..., h_k, h'_{l+1}, ..., h'_n$$

that traverses the blocks of $B'_1$, ..., $B_k$, $B'_{l+1}$, ..., $B'_n$. As a result of this probe, $h_k$ will be included in *source_set*s of $B'_{l+1}$, ..., $B'_n$, which means that $v_1$ can skip these blocks as an interface path has already been found from $v_1$ to them via $h_k$, as shown in Figure 2. (Shown as the dotted line in the left of the figure.) Similarly, $v_2$ can also skip the blocks of $B_{k+1}$, ..., $B_m$. (Shown as the dotted line in the right of the figure.)

### 3.3 Measurement Results

**Table 2: Traceroute measurement results**

| | iPlane (one day) | iPlane (two days) | cTrace | Both |
|---|---|---|---|---|
| Traceroutes | 1,244,667 | 2,381,482 | 106,580 | |
| Interfaces | 17,308 | 17,761 | 71,047 | 10,023 |
| Links | 76,120 | 82,791 | 146,542 | 27,735 |

Using nested-block partitioning and collaborative traceroutouting, we perform a traceroute measurement on China's Internet with 15 vantage points (from 9 different cities and in 4 different ISPs) in China. We applied the nested-block partitioning algorithm on the IP blocks from 8 BGP snapshots and further divided them to prefix /22 blocks for obtaining the target addresses. The measurement was performed from 12 December 2010 to 2 January 2011. We also downloaded iPlane's traceroute data on Dec. 19 and Dec. 20, 2010 for comparison. For each path in iPlane, we extract the segment that is on China's Internet.

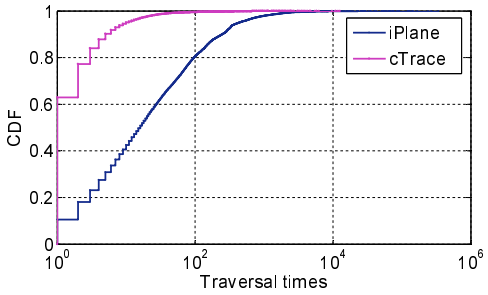



**Figure 3: Number of times the links are visited**

Table 2 compares the iPlane data with our measurement results (referred to as *cTrace*). For iPlane, we present the results for both one and two days of measurement. As compared with iPlane, our approach employs only 5% of the number of traceroute probes but finds four times as many interfaces and twice as many interface links. This experiment therefore shows that using vantage points in China is much more efficient in exposing China's Internet, and collaborative traceroutouting can effectively eliminate redundant probes.

To further demonstrate our point, we plot the distributions of the number of times the links are visited in iPlane (over 2 days) and in cTrace in Figure 3. In two days, iPlane visited some links thousands of times, even though most of its vantage points are outside of China and, thus, are far away from these links.

Finally, from Table 2 we note that although cTrace contains many more interfaces and links than iPlane, there are still over 7,000 interfaces and 50,000 links in iPlane that are not discovered by cTrace. We believe these interfaces and links are located on the border of China's Internet that connects to the international Internet. These links are thus unlikely to be traversed by cTrace, which uses vantage points within China. For this reason, we combine cTrace with the 2-day iPlane data, and use the combined data for further study in this paper.

*In summary, we perform a traceroute measurement with as few as 15 vantage points on China's Internet. As compared to existing large-scale traceroute measurements, our scheme not only reveals a much larger number of Chinese links and interfaces, but also uses significantly fewer traceroute probes.*

## 4. GEOLOCATION SERVICES ON CHINA'S INTERNET

One goal of this paper is to develop a methodology for accurately geolocating Chinese IP addresses for both end hosts and router interfaces. In this Section, we briefly examine the geolocation services currently available for China's Internet. In the subsequent section, we will develop methodologies to improve these services.

We consider four geoIP databases in this study, namely, IP138 [12], QQWry [21], IPcn [13], and MaxMind [19]. The first three are Chinese databases that are well-known in the Chinese Internet community, whereas MaxMind is a leading global geolocation service provider. The locations returned by these databases generally have two levels: the province level and the city level. For the directly-controlled municipalities of Beijing, Shanghai, Tianjin and Chongqing, we consider them as both provinces and cities. For cases when bogus locations are returned (e.g., a non-exist location name), we consider the corresponding level location information as null.

**Table 3: Null reply ratios for the addresses from traceroute**

| | IP138 | QQWry | IPcn | MaxMind |
|---|---|---|---|---|
| Province | 0.105 | 0.074 | 0.108 | 0.186 |
| City | 0.240 | 0.212 | 0.280 | 0.227 |

**Table 4: Null reply ratios for Xunlei peers**

| | IP138 | QQWry | IPcn | MaxMind |
|---|---|---|---|---|
| Province | 0.011 | 0.004 | 0.021 | 0.161 |
| City | 0.153 | 0.137 | 0.178 | 0.224 |

We first consider null-reply ratios for each database. A database's null-reply ratio is defined as the fraction of the cases for which the database fails to provide location information [25]. We use the 78,229 IP addresses from the combined traceroute data to examine the geoIP databases. Table 3 shows the null-reply ratios for the four databases at the province and city levels. We can see that each database frequently returns null replies, particularly for the city-level location information.

Two types of IP addresses are included in our traceroute data: router interface addresses and end host addresses. To gain further insight into the databases' performance for different types of addresses, we randomly selected 2,000 IP addresses from peers collected by crawling the Xunlei DHT [6] (a popular P2P download acceleration application in China) and fed these end host addresses to the geoIP databases. Table 4 shows the null-reply ratios on Xunlei peers. Comparing with Table 3, we can see that except for MaxMind, the three Chinese databases have fewer null replies for Xunlei peers, suggesting that the three Chinese databases cover better end host IP addresses than router interface addresses.

*In summary, we find that the three Chinese geoIP databases are moderately accurate for end host geolocating, and substantially less accurate for router interfaces. In particular, we observe frequent occurrences of null replies and erroneous entries, suggesting that there is significant room for improvement.*

# 5. GEOLOCATING THE INTERFACE TOPOLOGY

With the combined traceroute data obtained in Section 3, we have obtained a separate interface topology for Telecom, Unicom and CERNET. Each of these interface topologies can be viewed as a directed graph: Each interface (IP address) forms a vertex, and each pair of successive interfaces from the traceroutes forms a directed edge. In this section, we seek to geolocate the three interface topologies. In many countries, router interfaces are often assigned names that indicate the interface's location. In such cases, the location of an interface can be determined by simply performing a reverse DNS lookup on the corresponding IP address. In China, however, very few router interfaces have names. We therefore must develop an alternative approach for geolocating the router interfaces. We develop a clustering approach, as described subsequently.

For a given interface topology $T$, we say a set of router interfaces $S$ forms a *cluster* if (a) all the interfaces in $S$ belong to the same city, and (b) the subgraph of $T$ induced by $S$ is weakly connected. We further say that a cluster $S$ is a *maximal cluster* if it is not possible to create a larger cluster by adding more interfaces to it. Our goal is to determine the maximal clusters in each of three interface topologies. Note that a city could have more than one maximal cluster, for example, it could have two maximal clusters which do not have a direct link between them, but which have an indirect path between them via another city.

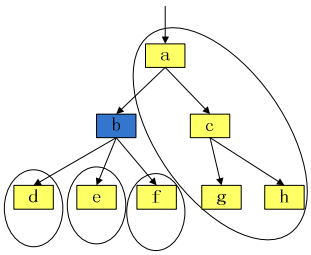



**Figure 4: Erroneous clusters example**

A naive method to create the clusters is to simply use the city information provided by the geoIP databases on face value. However, this naive approach leads to a large number of small and disconnected erroneous clusters due to missing and erroneous entries in the geoIP databases. Figure 4 provides an example, with boxes representing the interfaces and arrows representing the links. All the interfaces on the graph are at the same location, and should be included in one cluster. However, if interface $b$'s location from geoIP database is wrong or missing, four instead of one cluster is formed, as shown in the figure. This example shows that a few errors in geoIP databases will cause many clusters to be erroneously formed. On the other hand, by combining the information in the geoIP databases with the topological information obtained from the traceroutes, it may be possible for us to identify the errors in geoIP databases and determine the interfaces' real locations. For example, for interface $b$ in Figure 4, as all the interfaces adjacent to it are at the same location, we can conclude that $b$'s database location is likely incorrect and $b$ is likely located at the same location as all the other interfaces on the graph. Inspired by this observation, we propose a heuristic for accurately determining the maximal clusters in each of the three interface topologies.

## 5.1 Geo-Clustering Heuristic

Geolocating an interface network using a partially accurate geoIP database is a challenging problem for an arbitrary interface topology. Fortunately, the ma-

jor Chinese ISPs have a hierarchical structure, which makes the problem more tractable. The heuristic we present here could be used for any ISP with a hierarchical structure (not just Chinese ISPs).

For each of these ISPs, using the traceroute data, we first obtain an interface topology that expands from the ISP's backbone network to the traceroute targets in that ISP. After obtaining the interface topology, the clustering algorithm starts from the interfaces at the edge of the topology, then gradually moves towards the backbone interfaces located at the core. The heuristic algorithm consists of four steps. In the first step, we form singleton clusters using the interfaces at the edge of the interface topology. In the second step, we repeatedly select the interfaces that are one step closer to the backbone network and, based on their inferred locations, group them into existing clusters. In this step, we cluster router interfaces in the residential and provincial networks. In the third step, we cluster the router interfaces in the backbone network using a method similar to step two, but we apply different rules for inferring the interfaces' locations. Finally, in step four, we merge the singletons and small clusters that remain after step 3 to create the maximal clusters. In the following we describe each of the steps in detail.

#### 5.1.1 Step Zero: Preprocessing

Before clustering the interfaces, we first filter out the influence of the vantage points and anonymous routers in the interface topologies. For a typical traceroute path traversing Telecom, Unicom, or CERNET, it contains three subpaths: the subpath from the vantage point to the first backbone router, the subpath inside the backbone network, and the subpath from the last backbone router to the target. We only include the last two subpaths in our interface topology. By filtering out all the "up backbone" links in the first subpath, the resulting interface topology can be viewed as expanding from the backbone network to the traceroute targets. In addition, if an anonymous router is found in the third subpath, all the interfaces after the anonymous interface are removed.

In this preprocessing step, we need to identify the backbone routers. Although most routers in China are nameless, and thus cannot be reverse DNS queried, the three Chinese databases do return backbone network information, indicating whether an address being queried belongs to the Telecom, Unicom, or CERNET backbone. By filtering we remove only a small fraction of the addresses from the traceroute data. For example, only 2.5%, 5.9%, and 15.3% of the interfaces on the interface topologies of Telecom, Unicom, and CERNET are removed using the geoIP database of IP138.

#### 5.1.2 Step One: Startup Clusters from the Edge

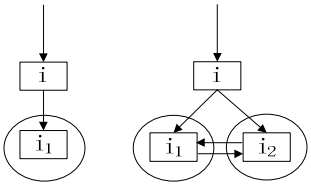



**Figure 5: Example of step one clustering**

In the first step, we select the addresses at the edge of the interface topology to form startup singleton clusters. Specifically, for an interface in the topology, say $i_1$, if it has no outgoing links, or for each interface $i_2$ it links to, there exists a path from $i_2$ to $i_1$ in the interface topology, a singleton cluster containing only $i_1$ is formed. We use $i_1$'s location from the geoIP database (referred to as $i_1$'s *DB location*) as the cluster's location. A simple example is shown in Figure 5. In the left figure, $i_1$ has no outgoing links, and a singleton cluster is formed. In the right figure, $i_1$ and $i_2$ link to each other, so there is a return path for both of them; thus two singleton clusters, one containing $i_1$ and the other containing $i_2$, are formed.

#### 5.1.3 Step Two: Clustering Residential and Provincial Networks

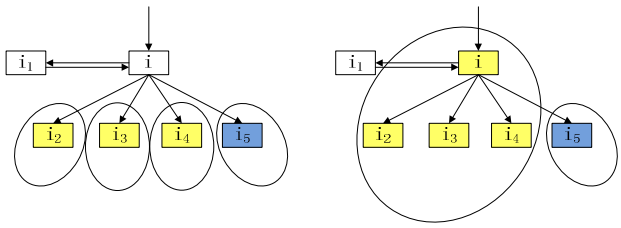



**Figure 6: Example of step two clustering**

After obtaining the startup singleton clusters, we continue to cluster more router interfaces. The heuristic works in rounds. In each round, we select some of the unclustered addresses in the interface topology as candidates for clustering. An unclustered interface $i$ is selected as a candidate if each of $i$'s out-linked interfaces is (a) either clustered (as are interfaces $i_2$, $i_3$, $i_4$, and $i_5$ in Figure 6) or (b) there exists a path in the interface topology from the out-linked interface back to $i$ (as is $i_1$ in Figure 6).

For each candidate interface, we use its out-linked interfaces to infer its location. Suppose a candidate interface $i$ links to interfaces $i_1$, $i_2$, ..., $i_n$, which belong to clusters $c_1$, $c_2$, ..., $c_m$. We then have $i_1$, $i_2$, ..., $i_n$ *vote* to infer $i$'s location. For each out-linked interface, say $i_k$, if its DB location contains a city-level location, it uses its DB location to vote; otherwise it uses the location of the cluster it belongs to (referred to as $i_k$'s *cluster location*). After voting, if there exists a city-level loca-

tion $x$ that wins the voting by exceeding a threshold of $c\_vote$, $x$ is assigned as $i$'s cluster location (but $i$ still keeps its DB location). We further merge all clusters among $c_1$, $c_2$, ..., $c_m$ that have cluster location $x$ into one larger cluster, and also put $i$ into this newly merged cluster. When there is no winner from the voting, if $i$'s DB location has a city-level location, say $y$, we merge all clusters among $c_1$, $c_2$, ..., $c_m$ that have cluster location $y$ into one larger cluster, and put interface $i$ into this newly merged cluster. Otherwise, $i$ cannot be put into any cluster and will form a singleton cluster containing only itself. A simple example of voting and cluster merging is shown in Figure 6: in the left graph, the candidate interface $i$'s DB entry does not include city-level information; however, its cluster location is inferred by the voting among its out-linked interfaces $i_1$, ..., $i_5$; after the voting, $i$ is assigned the same location as the clusters of $i_2$, $i_3$ and $i_4$, and they join $i$ to form a larger cluster, as shown in the right graph.

For a candidate interface, if more than one province appears in the voting, it is likely that this interface belongs to a backbone network. In this case, we abort the voting-based inference without forming or merging any clusters, and move on to the next candidate. After all the candidate interfaces are processed, the heuristic finishes a round and selects new candidates for the next round. Step 2 stops when we can't form or merge any clusters during a round.

#### 5.1.4 Step Three: Clustering Interfaces in the Backbone

Step 3 works similarly as Step 2 by first selecting a set of candidate interfaces, inferring their cluster locations, and merging the clusters with the same cluster location. We use the same method as in Step 2 to select a candidate. However, unlike Step 2, where candidate interfaces are on routers in residential or provincial networks, in Step 3, nearly all the candidate interfaces are on backbone routers, which usually connect many routers at different locations. In addition, the links that connect backbone interfaces are usually traversed many times during the traceroute measurement. This makes it possible to accurately estimate delays on those links. For a candidate interface $i$, we sequentially apply the following rules to infer its cluster location.

- We first examine each link incident from $i$ to a clustered interface, say $i_1$, that has been traversed more than $t\_times$ times. For such a link, we use the median of the delays estimated from different vantage points, as proposed in [9], as its delay. If the delay is smaller than the threshold of $l\_delay$, we assign $i_1$'s cluster location $x$ to $i$, and merge $i$ with all the clusters that connect to $i$ with location $x$. However, if more than one link is observed, suggesting different cluster locations, we apply the next rule.
- We use the same voting-based method as in Step 2 to infer candidate $i$'s city-level location. But here we don't apply the stop condition when more than one province appears. We use the location that wins the voting as $i$'s cluster location and merge its out-linked clusters if possible. If there is no winner from the voting, apply the next rule.
- We conduct a province-level voting using a method similar to Step 2. For each of $i$'s out-linked interfaces, only its province-level cluster location information is used. We use a threshold of $p\_vote$ to find the winner. After the voting, if a province-level location wins, we assign the capital city of that province as the candidate's city-level location. This is because a provincial network usually accesses the backbone network at the capital city of that province (see Section 5.3). We then perform cluster merging with the assigned city-level location if possible.

Finally, if none of above rules can be applied to $i$, a singleton cluster is formed for it.

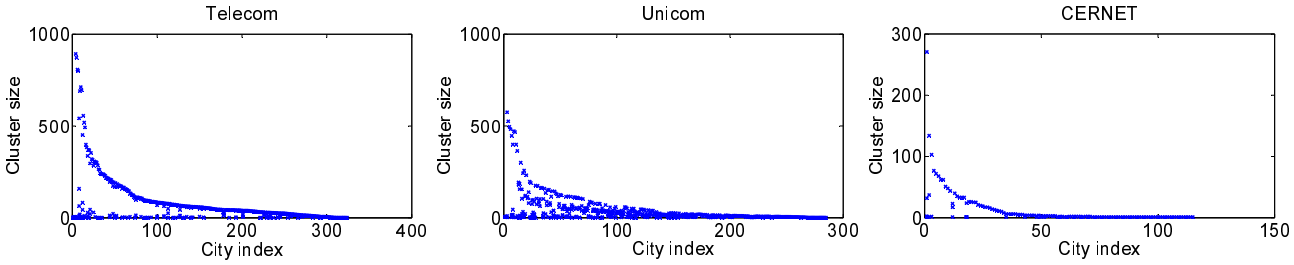
#### 5.1.5 Step Four: Merging Singleton and Small Clusters

After applying Steps 2 and 3, all the interfaces in the topology are clustered. Careful examination on the resulting clusters shows that for nearly all the cities, there are one or two large clusters containing most of the interfaces, as well as a number of singleton or small clusters. The objective of Step 4 is to merge these singleton and small clusters into a large one. We consider the clusters containing less than $c\_size$ interfaces as mergeable *small clusters*, and the others are regarded as *large clusters*. For a small cluster, if it is only connected to one large cluster, then the location information given in the database for the small cluster is likely to be wrong; we therefore merge it into the large cluster, regardless of its original cluster location. By repeatedly merging small clusters, we can eliminate most of them.

We refer to this four-step heuristic as the *geo-clustering* heuristic on the interface topology.

### 5.2 Geo-Clusters

We applied the geo-clustering heuristic on the Telecom, Unicom, and CERNET interface topologies using each of three geoIP databases. For the parameters, we use $c\_vote = 0.5$, $l\_delay = 1ms$, $t\_times = 5$, $p\_vote = 0.5$ and $c\_size = 5$. For example, for the Telecom's interface topology using the geoIP database of IP138, there are $38,181$ interfaces. After Steps 1, 2, 3, and 4, we get $26,518$, $7,326$, $7,467$, and $1,125$ clusters, respectively. 532 of the final clusters contain $37,488$ interfaces and have been assigned city level locations.




(a) Telecom (b) Unicom (c) CERNET

**Figure 7: Distribution of the sizes of geo-clusters across cities in three major ISPs**

(The remaining clusters are singleton clusters for which the heuristic did not assign to a city since there was no clear majority winner in the voting.) The observation indicates that by geo-clustering, we can group most of the interfaces into clusters with detailed city-level location information. We refer to a cluster with a city-level location as a *geo-cluster*. Similar results are observed using the two other geoIP databases and for the two other backbone ISPs. We omit them due to lack of space.

By examining the 532 geo-clusters obtained on Telecom's interface topology, we find they are located in 324 different cities, which are nearly all the cities in China. We show the sizes of the geo-clusters for each city for Telecom, Unicom, and CERNET in Figure 7, where the x-axis is the city index, the y-axis is the cluster size, and each point on the figure corresponds to a geo-cluster. For each ISP, the cities are indexed according to the total number of IP addresses across all geo-clusters in the city. From Telecom and CERNET's figures, we can see that for many cities, there is only one geo-cluster. For a small fraction of the cities, multiple clusters are found, with one cluster containing the majority of the interfaces. There are two possible reasons for multiple clusters in a city: (*i*) the ISP has multiple networks serving different purposes in that city; and more likely (*ii*) some of the singleton and small clusters cannot be merged into large clusters in step four. Note that the Unicom's geo-cluster distribution is distinctly different from those of Telecom and CERNET. In particular, for Unicom in many cities there are two large geo-clusters of comparable size, as shown in Figure 7(b). Our heuristic is consistent with the fact that in 2008 Unicom merged with China Netcom, which used to be the second largest ISP in China. As a result, in many cities we can observe one large geo-cluster for the former Unicom network, and another large geo-cluster for the former Netcom network.

Figure 8 shows the geo-clusters of the top 10 cities. For clarity we remove the singleton clusters. From the figure we an see that each top-10 city has only one major cluster per ISP (including Unicom for these cities);

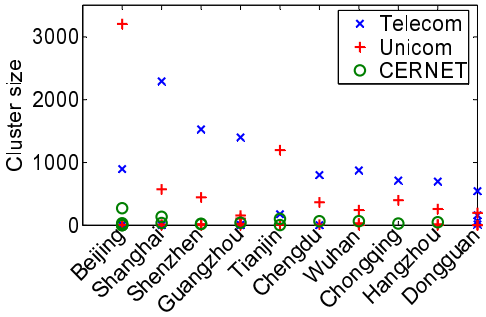



**Figure 8: Geo-clusters in the top-10 cities**

moreover, Unicom has much larger geo-clusters than Telecom in Beijing and Tianjin, located northern China, while Telecom has larger geo-clusters in other cities in southern and western China.

## 5.3 The Hierarchical Structure

**Table 5: Statistics of inter-cluster links**

| | Same province | | Different province | | |
|---|---|---|---|---|---|
| | Cap. | Other | 2Cap. | Cap. | Other |
| Telecom | 3,236 | 2,097 | 169 | 283 | 42 |
| Unicom | 1,504 | 1,281 | 199 | 25 | 0 |
| CERNET | 69 | 1 | 181 | 21 | 0 |

Using our clustering heuristic, we now study the internal structure of each ISP. Table 5 categorizes inter-cluster links based on the locations of the two endpoints of the links. In this table we have removed the links with both endpoints on the backbone. The first and the second columns are for the intra-province links, where the first column is for links between the capital city and another non-capital city in that province, and the second column is for links between two non-capital cities. The third through fifth columns are for inter-province links: links between the capital cities of two different provinces (column 3), links with only one endpoint at a capital city (column 4), and links between two non-capital cities in different provinces (column 5). From the table we can see that for Telecom and Unicom, there are many

intra-province links, and more than half of them are between capital and non-capital cities. There are relatively few inter-province links, and the majority of them connect to at least one capital city. *We can therefore conclude that the major Chinese ISPs are highly hierarchical following China's provincial organization, and that the provincial capital cities are not only government centers but also hubs in the ISPs' networks. This strikingly contrasts with flattening trends in the international Internet [3] [16].* Finally, we observe relatively few intra-province links in CERNET, as CERNET only reaches cities that have many universities and research institutes, and such cities are usually the provincial capital cities in China.

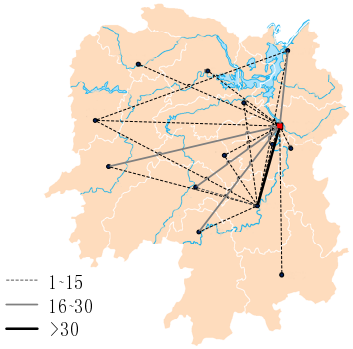



**Figure 9: Cluster topology for Hunan province**

As an example, Figure 9 shows the topology of the Telecom geo-clusters of all the cities in Hunan province, where the width of the edge between two cities represents the number of distinct interface links between geo-clusters located at the two cities. We can see that the topology is strongly centered around the capital city of Changsha, as shown by the red square on the graph.

## 5.4 Inter-Connectivity among Major ISPs

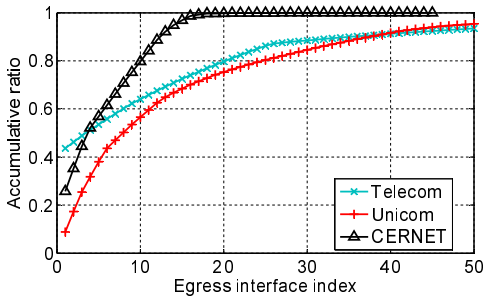



**Figure 10: Ratio distribution for paths traversing egress interfaces**

We investigate the inter-connectivity among the major Chinese ISPs in this section. For a major ISP, we focus on its backbone interfaces through which traceroute paths exit the ISP to enter another major ISP's backbone network. We refer to such an interface as the ISP's *egress interface*. From the combined traceroute data, we collected 150, 98, and 45 distinct egress interfaces for Telecom, Unicom, and CERNET, respectively.

We rank egress interfaces by the number of inter-ISP traceroute paths traversing them. Figure 10 shows the accumulative ratios for the three ISPs. For each of the three ISPs, the paths are distributed unevenly among the egress interfaces: 99.9%, 86.2% and 80.3% of the paths departing CERNET's, Telecom's and Unicom's backbone, respectively, are through the first 25 egress interfaces.

**Table 6: Shared interfaces and traceroutes across distant vantage points in same ISP**

| | Vantage point | Interface | Traceroute |
|---|---|---|---|
| Telecom | Chengdu | 76.1% | 53.6% |
| | Xiamen | 67.3% | 50.1% |
| Unicom | Tianjin | 76.2% | 95.9% |
| | Shenzhen | 84.3% | 92.7% |
| CERNET | Beijing | 78.6% | 64.7% |
| | Changsha | 68.8% | 79.7% |

We now investigate how the egress interfaces are traversed. For each major ISP, we select two vantage points on its network that are geographically far away from each other. For each vantage point, we examine the egress interfaces traversed by its traceroutes, identifying the egress interfaces that are shared by the two vantage points in a same ISP. Table 6 lists the percentages of the shared egress interfaces among all the egress interfaces of each vantage point, as well as the percentages of the paths through these shared egress interfaces among all the inter-ISP traceroute paths from each vantage point. From the table we can see the two distant vantage points in the same ISP share an astonishing percentage of interfaces when accessing other ISPs' backbone networks. Note that the inter-ISP traffic concentration is more severe in Unicom than in the other two ISPs. *Although we only use two vantage points for each ISP, their geographical locations suggest that traffic from all over China within a major ISP will concentrate to a few routers for accessing other ISPs' networks, potentially making these routers bottlenecks for inter-ISP traffic in China.*

## 5.5 Locating Interfaces with Null Replies

**Table 7: Null reply ratios**

| | IP138 | QQWry | IPcn |
|---|---|---|---|
| DB province | 7.7% | 6.2% | 8.0% |
| Cluster province | 0.99% | 1.00% | 0.93% |
| DB city | 21.7% | 18.7% | 26.4% |
| Cluster city | 1.51% | 1.64% | 1.66% |

Each interface in an ISP's interface topology has now been assigned two locations: the geoIP database location and its cluster location (with the clusters derived from the same database). In this section, we show that the cluster locations are significantly more complete and accurate.

We first examine the completeness by comparing the null reply ratios. In this comparison, all the IP addresses of the interfaces on Telecom, Unicom, and CERNET's interface topologies are included. Table 7 shows the null reply ratios at the province and the city levels for both DB and cluster locations. Observe that the ratios for cluster locations are much smaller than those for the DB locations. The geoIP services give a high-level of null replies because many router addresses do not have city-level or province-level locations in the database. However, the cluster locations for many of these router interfaces have been inferred at the city level (by the voting in Steps 2 and 3 and by the merging in Step 4).

**Table 8: Number of the interfaces that have consistent locations**

| | Total | 3DB identical | 3Cluster identical |
|---|---|---|---|
| Telecom | 38,181 | 25,625 (67.1%) | 35,376 (92.7%) |
| Unicom | 24,781 | 15,794 (63.7%) | 21,938 (88.5%) |
| CERNET | 1,798 | 1,343 (74.7%) | 1,602 (89.1%) |
| Total | 64,760 | 42,762 (66.0%) | 58,916 (91.0%) |

We now examine the accuracies of the DB and cluster locations. Unfortunately, given the lack of landmarks for router interfaces, it is not possible to say with 100% certainty whether a geoIP database location or a cluster location is correct. (However, we will be able to use landmarks in Section 6 when we study end host geolocation.) Instead, here we use cross validation to support our claim that clustering approach is substantially more accurate than the geoIP databases for router interfaces.

For an interface, if the locations from the three databases are the same, it is likely that the location is correct; if, however, all three databases do not give the same location, then we have a low level of confidence on the location information. Similarly, using the three sets of geo-clusters based on the three different geoIP databases, we can cross-validate the cluster locations. Table 8 shows for each of the three ISPs, the number of the addresses that have consistent locations for the two approaches. We see that the three geoIP databases agree only for 66.0% of the interfaces (average across the three ISPs), but after applying the geo-clustering heuristic, as many as 91.0% interfaces have the same cluster locations.

*In summary, for a hierarchical interface topology, we propose a heuristic to geolocate the interfaces from traceroute measurements by forming geo-clusters. We apply the heuristic to China's Internet and provide evidence that resulting large geo-clusters are essentially the maximal clusters. The geo-clusters clearly expose China's hierarchical structure down to the city level. We also observe a concentration of inter-ISP traffic at a relatively small number of interfaces. In addition, we show that our heuristic can geolocate router interface addresses with more detailed and accurate location information than can existing geoIP databases.*

# 6. IMPROVING GEOLOCATION SERVICES WITH GEO-CLUSTERS

In the previous section, we showed how our methodology can geolocate router interface addresses that have null or erroneous entries in the geoIP databases. In this section, we develop a methodology for accurately geolocating arbitrary Chinese IP addresses, including host interfaces. Our goal here is to provide a significant improvement over the existing Chinese geoIP databases.

## 6.1 Geolocating an Arbitrary IP Address

Our methodology relies on the geo-clustering heuristic described in Section 5.1. For a given IP address $p$ that we wish to geolocate, we first determine the ISP to which it belongs (e.g., by first determining the AS to which it belongs from BGP tables). This ISP has an interface topology, say $T$, which we obtained from our traceroute data.

To apply the geolocating algorithm in Section 5 to an arbitrary IP address $p$, we need to first augment $T$ to reach $p$. This requires us to conduct additional traceroute probes. We choose a subset of existing vantage points, each of which keeps a queue of targets to be probed. For initialization, we put $p$ into the target queue of each vantage point. Then vantage points conduct traceroute probes by working through their target queues: at each step, each vantage point dequeues a target $t$ and performs a traceroute to $t$. Along the traceroute path, if there exists an interface $i$ between $T$ and $t$ for which there is no anonymous router between $T$ and $i$, we insert $i$ into the target queues of all the vantage points (except for the one that just returned this path). This process continues until the queues of all the vantage points become empty.

We then use the new traceroutes to augment the topology $T$ to create a new interface topology $T'$ (using Step 0 in the heuristic, as described in Section 5.1). Applying the geo-clustering heuristic to the new augmented topology $T'$, we obtain a new set of geo-clusters. The location of $p$ is then determined from these new geo-clusters using one of the following three cases:

- Case 1: $p$ is in the topology $T'$ and therefore is included in one of the geo-clusters. In this case, we simply set $p$'s location to the location of the cluster that encompasses it.

- Case 2: $p$ can be reached by at least one traceroute path, but $p$ is not in $T'$ (due to the occurrence of anonymous routers in the traceroute paths). In this case, we find the geo-cluster that is closest to $p$ among all the traceroute paths, which we refer to as the *last-hop geo-cluster*. If the distance between the last-hop geo-cluster and $p$ is no larger than a threshold (2 hops in our evaluation), we set $p$'s location to the location of the last-hop geo-cluster. However, if there are multiple last-hop geo-clusters with different cluster locations for $p$, then Step 2 is inconclusive, and we proceed to Step 3.

- Case 3: If we don't set $p$'s location in Case 1 and 2, the location from the geoIP database is used.

## 6.2 Evaluation

### 6.2.1 Collecting Landmarks

We use a number of landmarks as the ground truth for evaluating the accuracies of the geoIP databases and of our methodology. Generally it is difficult to obtain ground truth landmarks, particularly for China's Internet, as many websites are hosted by IDCs, including university and government websites (which are often used as landmarks in other studies [20] [29]). In this paper, we leverage the numerous IDC datacenters located in many cities in China, for collecting the landmarks. For a datacenter, we find one or more of its IP addresses, and associate the datacenter's location with the IP addresses to get landmarks.

We combine the IDC datacenter IP addresses we have found from the websites listed on *www.IDCquan.net* and the mirror servers of the popular software downloading site *onlinedown.net* to collect IDC datacenter landmarks. We have successfully collected 305 landmarks – 199 on Telecom and 106 on Unicom – with their ground-truth locations detailed to the city level.

### 6.2.2 Evaluation Results

**Table 9: Evaluation using Telecom landmarks**

| | | Case 1 | Case 2 | Case 3 | Total |
|---|---|---|---|---|---|
| IP138 | DB | 105/115 | 11/15 | 56/69 | 172/199 |
| | Improve | 110/115 | 15/15 | 56/69 | 181/199 |
| QQWry | DB | 107/117 | 11/15 | 54/67 | 172/199 |
| | Improve | 111/117 | 14/15 | 54/67 | 179/199 |
| IPcn | DB | 102/117 | 11/15 | 57/67 | 170/199 |
| | Improve | 111/117 | 14/15 | 57/67 | 182/199 |
| MaxMind | DB | N/A | N/A | N/A | 85/199 |

We use ten vantage points located in seven different cities to geolocate the 305 landmarks. Our methodology requires us to probe a few additional addresses for each landmark to extend the interface topology. For each landmark, 4 additional probes from each vantage point were required on average.

**Table 10: Evaluation using Unicom landmarks**

| | | Case 1 | Case 2 | Case 3 | Total |
|---|---|---|---|---|---|
| IP138 | DB | 46/55 | 9/10 | 34/41 | 89/106 |
| | Improve | 52/55 | 9/10 | 34/41 | 95/106 |
| QQWry | DB | 48/55 | 8/8 | 33/43 | 89/106 |
| | Improve | 53/55 | 8/8 | 33/43 | 94/106 |
| IPcn | DB | 44/55 | 8/10 | 28/41 | 80/106 |
| | Improve | 52/55 | 9/10 | 28/41 | 89/106 |
| MaxMind | DB | N/A | N/A | N/A | 57/106 |

For each landmark, we compare the location determined by our geo-clustering methodology and the location from the corresponding geoIP database with the landmark's ground truth location. The number of the landmarks that are accurately located by the different methods are shown in Table 9 for Telecom landmarks and in Table 10 for Unicom landmarks. We further classify the landmarks into three cases based on how their locations are determined by our methodology. As an example, consider the case of Telecom and the IP138 geoIP database. Of these 199 Telecom landmarks, 115 fall into Case 1. Of these 115 landmarks, the IP138 database correctly located 105; whereas our methodology (using the same location database) correctly located 110. We also evaluate the MaxMind database, and find that MaxMind is inaccurate comparing with the three Chinese databases.

From Table 9 and Table 10, we see that for both ISPs, our geo-clustering methodology can accurately geolocate more landmarks than can the geoIP databases. For the landmarks in Case 1 and Case 2, we are able to accurately geolocate over 7% more Telecom landmarks and over 10% more Unicom landmarks on average. In addition, more than 60% of the landmarks under evaluation fall into case 1 and case 2, suggesting that our methodology can improve the geolocation services for many IP addresses in the Chinese Internet. Although these improvements for locating end host IPs are not as dramatic as our results for locating router interface IPs, we believe that the improvements are nevertheless significant and useful.

In addition to improved accuracy, a less obvious benefit of our methodology is that it provides a means for users to assess the quality of the results returned from geoIP databases. Specifically, each geoIP database is significantly more accurate for targets falling into Case 1 or 2 than those falling into Case 3. Thus, when using a geoIP database, if the target falls into Case 1 or 2, the user can be relatively confident about the result, but less confident when the target falls into Case 3. Furthermore, we find that when a geoIP database gives an accurate result, our methodology always provides the

same result, with only one exception of a Unicom landmark using the databases of IP138 and QQWry.

For the landmarks belonging to Case 3, by examining the traceroute paths to them, we find that the distances between their last-hop geo-clusters and the traceroute targets are larger than 2 hops, and many paths never reach the landmarks. We remark that for an IP address that is unreachable with traceroute, or is far behind anonymous devices, it becomes difficult to geolocate for any traceroute-based mechanism.

*In summary, we have designed a traceroute-based methodology for improving the Chinese geoIP databases. Our evaluation with ground-truth landmarks shows that the methodology provides more detailed and accurate location information, and also allows users to assign levels of confidence to the results returned from the geoIP databases. Finally, we point out that by improving the results from IP138, QQWry, and IPcn, which are currently considered as the best geoIP databases in China, we are indeed providing the (currently) best geolocation service for China's Internet.*

## 7. RELATED WORK

Spring *et al.* [26] propose Rocketfuel to probe ISPs' networks with public traceroute servers. Rocketfuel improves measurement efficiency by avoiding traceroutes through the same ingress and egress interfaces of the target ISP. The iPlane project [14, 18] also uses a public platform composed of PlanetLab nodes and traceroute servers; it reduces the probing workload by reducing the targets with BGP atoms [1]. On the other hand, CAIDA/Ark [2] works with dedicated monitors, dividing its monitors into teams for workload reduction. Several algorithms are proposed for improving the measurement efficiency on dedicated platforms: Donnet *et al.* [7] propose the Double Tree algorithm to avoid redundant probing packets by heuristically probing forward and backward to a vantage point and a target from the mid-point. Beverly *et al.* [4] propose a scheme that optimally selects the probing packets that can fully cover the entire interface topology obtained in the previous measurement cycle. Our collaborative tracerouting scheme differs from these works in that it is more suitable for a public platform of PlanetLab nodes and traceroute servers, and is more effective than the approaches used in Rocketfuel and iPlane; in particular, our approach avoids probes if an earlier interface path was found from the vantage point to the target. In addition, we leverage the block nesting in BGP snapshots [30] to improve the measurement coverage without introducing many unnecessary targets.

For mapping the Internet, interfaces are typically clustered to routers and PoPs in order to reveal the Internet structure [26, 14, 2, 24]. However, router and PoP clusterings typically rely on having numerous vantage points and on the ability to reverse DNS router interface IPs, both of which are unavailable in China's Internet. In this work, we instead group the interfaces into geo-clusters, which reveals the internal structure of the major Chinese ISPs.

Many automatic IP address geolocation techniques based on landmarks and active delay measurement have been proposed in recent years [20, 10, 15, 8, 29]. However, Li *et al.* [17] show that the delay-distance correlation, which is a foundation for many delay measurement based geolocation techniques, is weak in China's Internet. Shavitt *et al.* [25] propose to use PoP-level topologies, which are derived from delay measurements [24], to compare and evaluate geoIP database services. In this paper, we exploit the location information in geoIP databases to cluster the interfaces from a traceroute measurement; we then use the resulting geo-clusters to improve the completeness and accuracy of these databases.

There have only been a few studies focused on China's Internet. Guo *et al.* propose Structon [11], which mines and extracts location information from web pages in order to provide a geolocation service within China. Our approach differs from Structon in that we first obtain interface topologies from traceroute measurements, then combine the interface topology with the partially correct locations from commercial geoIP databases. Structon uses a prefix partitioning rule and location voting. We instead use traceroutes, which directly reveal the underlying network structure, to infer IP addresses' locations; also, instead of using locations found on webpages, we use commercial geoIP databases that provide richer and more accurate location information to drive our heuristic.

## 8. CONCLUSION

China's Internet has received relatively little attention in the measurement community to date. In this paper, we carried out a large-scale topology mapping and geolocation study for China's Internet. We first developed two traceroute techniques, namely, nested-block partitioning and collaborative tracerouting, to comprehensively and efficiently probe China's Internet from a small number of vantage points inside China. Our approach is able to discover many more interfaces with significantly fewer traceroute probes than the existing traceroute schemes. By further exploiting the hierarchical structure of China's Internet, we proposed a geo-clustering heuristic that clusters interfaces within the same city. We show that the clustering heuristic can geolocate IP addresses with significantly more detail and accuracy than can the existing geoIP databases in isolation.

## 9. REFERENCES

[1] Y. Afek, O. Ben-Shalom, and A. Bremler-Barr.

On the structure and application of BGP policy atoms. In *Proc. of the 2nd SIGCOMM Workshop on Internet Measurment*, Marseille, France, Nov. 2002.

[2] Archipelago measurement infrastructure. http://www.caida.org/projects/ark/.

[3] B. Augustin, B. Krishnamurthy, and W. Willinger. IXPs: Mapped? In *Proc. of IMC'09*, Chicago, IL, USA, Nov. 2009.

[4] R. Beverly, A. Berger, and G. G. Xie. Primitives for active internet topology mapping: Toward high-frequency characterization. In *Proc. of IMC'10*, Melbourne, Australia, Nov. 2010.

[5] China Internet Network Information Center. Statistical report on Internet development in China, Jan. 2011.

[6] P. Dhungel, K. W. Ross, M. Steiner, Y. Tian, and X. Hei. Xunlei: Peer-assisted download acceleration on a massive scale. Technical report, Polytechnic Institute of NYU, Apr. 2011.

[7] B. Donnet, P. Raoult, T. Friedman, and M. Crovella. Efficient algorithms for large-scale topology discovery. In *Proc. of SIGMETRICS'05*, Banff, Alberta, Canada, Jun. 2005.

[8] B. Eriksson, P. Barford, J. Sommersy, and R. Nowak. A learning-based approach for IP geolocation. In *Proc. of PAM'10*, Zurich, Switzerland, Apr. 2010.

[9] D. Feldman and Y. Shavitt. An optimal median calculation algorithm for estimating Internet link delays from active measurements. In *Proc. of Workshop on End-to-End Monitoring Techniques and Services*, Munich, Germany, May 2007.

[10] B. Gueye, A. Ziviani, M. Crovella, and S. Fdida. Constraint-based geolocation of Internet hosts. *IEEE/ACM Trans. Net.*, 14(6):1219 – 1232, 2006.

[11] C. Guo, Y. Liu, W. Shen, H. J. Wang, Q. Yu, and Y. Zhang. Mining the web and the Internet for accurate IP address geolocations. In *Proc. of INFOCOM'09*, Rio de Janeiro, Brazil, Apr. 2009.

[12] IP138. http://www.ip138.com/.

[13] IPcn. http://www.ip.cn/.

[14] iPlane: An information plane for distributed services. http://iplane.cs.washington.edu/.

[15] E. Katz-Bassett, J. P. John, A. Krishnamurthy, D. Wetherall, T. Anderson, and Y. Chawathe. Towards IP geolocation using delay and topology measurements. In *Proc. of IMC'06*, Rio de Janeiro, Brazil, Oct. 2006.

[16] C. Labovitz, S. Iekel-Johnson, D. McPherson, J. Oberheide, and F. Jahanian. Internet inter-domain traffic. In *Proc. of SIGCOMM'10*, New Delhi, India, Aug. 2010.

[17] D. Li, J. Chen, C. Guo, Y. Liu, J. Zhang, Z. Zhang, and Y. Zhang. IP-geolocation mapping for involving moderately-connected Internet regions. Technical report, Microsoft, 2009.

[18] H. V. Madhyastha, T. Isdal, M. Piatek, C. Dixon, T. Anderson, A. Krishnamurthy, and A. Venkataramani. iPlane: An information plane for distributed services. In *Proc. of OSDI'06*, Seattle, WA, USA, Nov. 2006.

[19] MaxMind. http://www.maxmind.com/.

[20] V. N. Padmanabhan and L. Subramanian. An investigation of geographic mapping techniques for Internet host. In *Proc. of SIGCOMM'01*, San Diego, CA, USA, Aug. 2001.

[21] QQWry. http://www.cz88.net/.

[22] Routing information service. http://www.ripe.net/data-tools/stats/ris/routing-information-service.

[23] University of Oregon route views project. http://www.routeviews.org/.

[24] Y. Shavitt and N. Zilberman. A structural approach for PoP geo-location. In *Proc. of INFOCOM Workshop on Network Science for Communications (NetSciCom)*, San Diego, CA, USA, Mar. 2010.

[25] Y. Shavitt and N. Zilberman. A study of geolocation databases. Preprint, arXiv:1005.5674v3 [cs.NI], Jul. 2010.

[26] N. Spring, R. Mahajan, and D. Wetherall. Measuring ISP topologies with rocketfuel. In *Proc. of SIGCOMM'02*, Pittsburgh, PA, USA, Aug. 2002.

[27] traceroute.org. http://www.traceroute.org/.

[28] P. Uria-Recio. China telecommunications panorama, 2006. http://globthink.com/2009/08/12/china-telecommunications-panorama/.

[29] Y. Wang, D. Burgener, M. Flores, A. Kuzmanovic, and C. Huang. Towards street-level client-independent IP geolocation. In *Proc. of NSDI'11*, Boston, MA, USA, Mar. 2011.

[30] Y. Zhu, J. Rexford, S. Sen, and A. Shaikh. Impact of prefix-match changes on IP reachability. In *Proc. of IMC'09*, Chicago, IL, USA, Nov. 2009.