Quantifying the Importance of Vantage Point Distribution in Internet Topology Mapping

Yuval Shavitt and Udi Weinsberg
School of Electrical Engineering
Tel-Aviv University, Israel
Email: {shavitt,udiw}@eng.tau.ac.il

***Abstract*—The topology of the Internet has been extensively studied in recent years, driving a need for increasingly complex measurement infrastructures. These measurements have produced detailed topologies with steadily increasing temporal resolution, but concerns exist about the ability of active measurements to measure the true Internet topology. Difficulties in ensuring the accuracy of every individual measurement when millions of measurements are made daily, and concerns about the bias that might result from measurements along the tree of routes from each vantage point to the wider reaches of the Internet must be addressed. However, early discussions of these concerns were based mostly on synthetic data, oversimplified models or data with limited or biased observer distributions.**

In this paper, we show the importance that extensive sampling from a broad and well spread set of vantage points has on the resulting topology and bias. The majority of this paper is devoted to a first look at the importance of the distribution quality. We show that diversity in the locations and types of vantage points is required for obtaining an unbiased topology. We analyze the effect that broad distribution has over the convergence of various autonomous systems topology characteristics, and show that although diverse and broad distribution is not required for all inspected properties, it is required for some. Finally, claims against bias in active traceroute sampling are revisited, and we empirically show that diverse and broad distribution can question their conclusions.

I. Introduction

The study of the topological structure of the Internet, ranging from the finest IP-level to the coarsest Autonomous-Systems (AS) level, is the driving force of several measurements effort in recent years.

Internet topology mapping is commonly performed by using either passive or active measurements. RouteViews [1] is the major passive measurement project; it relies on the collection of BGP announcements and updates from a few tens of vantage points (VPs). Most other topology measurement projects rely on active probing, mostly using dedicated instrumentation boxes (e.g., Archipelago [2]) or utilize PlanetLab servers [3] (e.g., iPlane [4] and RocketFuel [5]). A third approach is to use software agents. DIMES [6] deploys a large number of software agents and maintains an active community of participants. Ono [7] uses the BitTorrent P2P network for performing active measurements amongst peers that installed their plugin.

Using dedicated hardware boxes as a measurement infrastructure often limits the number of possible VPs. Using PlanetLab allows an increase in the number of VPs, but PlanetLab servers are mainly located in academic networks. Both approaches create a relatively stable and consistent output that is more easily analyzed. Community-based projects benefit from contributions of a large and widespread community, but often produce intermittent results that are more challenging to analyze. Currently, there are three major operational distributed active topology discovery infrastructures, namely Archipelago (Ark), iPlane, DIMES. Ono, which is not a pure topology discovery infrastructure, was shown to be very useful in finding hidden links of the AS-level topology. The data used in this paper is obtained from DIMES and iPlane. Both are highly distributed active measurement infrastructures with hundreds of measurement points.

This paper studies the effects that a broad set of VPs have on the quality of the observed topology, showing both the importance of the number of VPs and of the diversity in the location and the type of the ASes that host VPs for topology discovery. While some question the possibility of cleaning up data from a community-based infrastructure [8], we describe several simple filtering techniques and show that given a sufficiently diverse and broad distribution of VPs (in terms of geography, type, and quantity), it is possible to obtain data of comparable quality to infrastructures that have been deployed in a controlled manner. We then use the filtered data to explore the benefits of having a broad distribution in order to reevaluate some recent bias claims. Moreover, we analyze various properties of the Autonomous Systems (AS) graph, and show that broad distribution can further assist in reducing the bias of the results. We employ convergence-testing techniques [9], and show that some graph properties require more than 40 different VPs in order to converge to a value that represents the measured topology. Such a high number of VPs is more than most existing work uses as dataset.

II. Related Work

There is much research devoted to the analysis of the Internet topology measurement data, whereas only a few papers perform an in-depth analysis of the measurement infrastructures themselves. Barford *et al.* [10] studied the utility of adding VP for topology discovery, and showed that beyond the second VP, the utility quickly diminishes. However, Shavitt and Shir [6] later showed that although the utility indeed diminishes, the data from adding hundreds and thousands of VPs have a substantial effect on the resulting topology.

Following this observation, it became well accepted that attempting to infer the Internet topology from a few VPs leads to incomplete [6], [11], [12] and, even more important, biased topologies [6], [8], [13]. However, a common problem with previous work is their usage of either synthetic networks or real data that is either inaccurate or insufficiently understood for the tasks it is used [14]. As such, the commonly used power-law model for generating synthetic AS-level graphs [15] has been shown to be attributed both to the measurement process itself [16] and to the incorrect analysis of the data used [14].

Creating AS-level topologies from BGP data was shown [17], [18] to miss a substantial amount of AS-links if data is taken from a few VPs or for insufficiently long time. For example, Oliveira *et al.* [18] showed that BGP data can miss 10–20% of the tier-1 and tier-2 AS-links, and 85% or more AS-links of large content provider networks.

Mahadevan *et al.* [19] performed a comparative analysis of the AS topology using three different data collection methods – traceroutes (using Skitter), BGP (RouteViews) and IRR (WHOIS). The authors showed that topologies created from active traceroutes and passively collected BGP announcements are similar but differ substantially from the user-maintained WHOIS topology.

The ability of active topology measurements to map the Internet topology in general and the AS-level topology in particular was also shown to raise some difficulties in uncovering missing links [7], performing frequent probing [12] and mapping IP-level traceroutes to AS-level topology [20], [21], [7], [22].

Chen *et al.* [7] claimed that it is possible to extend the known AS topology by deploying VPs in P2P networks, and find links that were unseen by BGP data. However, the authors aggregate data collected for almost a year, making the assumption that AS-links are only added to the topology during this time frame. Although the authors apply various methods for cleaning their data from possibly false links, small measurement mistakes get accumulated. Even when using the extensive heuristics presented by the authors for removing false links, such false data is hard to identify when the number of measurement per VP is small. Although this very broad distribution (with VPs in over 3,700 different ASes) contributes many previously unseen links, when compared with an aggregated DIMES AS-level topology over the same time frame, but with a much narrower set of 300 VPs, we found that the overall number of links is similar (roughly 140,000 AS-links), but each topology misses roughly 60,000 links that the other topology finds. This shows that a simple increase in the number of VPs does not guarantee better coverage.

Beverly *et al.* [12] studied methods for enabling high frequency probing of the Internet topology by reducing the number of probes per VP and by changing the way destinations are assigned to VPs. Although the authors show that the utility of adding VPs slowly decreases, their focus is strictly on the interface-level and AS-level topology size, i.e., the number of discovered entities. Furthermore, the paper uses data from Ark, making it limited in both the number of VPs (reaching only 38 VPs), and in the types of ASes that host these VPs. In this paper we perform a much larger-scale study of the effect that the number of VPs and the broadness of their locations has on the AS-level topology and its properties.

Oliveira *et al.* [23] created an evolutional model of the AS topology and provided an evaluation of the proposed model using different data sources. The authors used BGP as the basis for their analysis and extended it with Internet Routing Registry (IRR) data and active probing including Skitter, DIMES and iPlane. They conclude that although active traceroute probing is an important source for topology information, it has a problem of broadness (covering all sampled topology) and freshness (updating the destinations list). A longer ten year study of the Internet evolutions was later performed [24], however it uses only BGP data.

Krishnamurthy and Willinger [14] recently raised several concerns about the quality of measurement-based research, focusing on a fundamental question of whether the measurements and their analysis actually support the resulting claims. Following some of their insights, this paper attempts to improve the understanding of large-scale measurement data, its quality, and its limitations, mainly in the context of the AS-level topology analysis.

III. Measurement Setup

As previously noted, increasing the number of VPs is a challenging task for all Internet measurement projects, either due to the need of purchasing and deploying new specialized machines, or convincing users to install an agent on their PCs. We wish to study empirically if and to what extent using a large number of VPs affects the observed topology.

To this end, we use traceroute data from DIMES collected during the month of August 2009. Due to its community-based design, DIMES may exhibit changing behaviors, depending on the activity of its community. However, some of the analysis here appears in the conference version of this paper [25], where we used only one week of data from early 2008, and the results of both datasets are quite similar. The following section briefly highlights some of the important aspects of DIMES and provides understanding of how to detect and filter out unwanted behavior, leading towards a more accurate data analysis.

A. Infrastructure Overview

DIMES performs measurements using hundreds of software agents installed on users' PCs. Agents perform measurements by following a script that is sent to them from a central server. An agent can perform traceroute and ping measurements using either ICMP or UDP packets, with a default of two measurements per minute, inducing minimal bandwidth overhead. Upon completion of a script, an agent submits the results and requests a new script to perform. By default, an active agent performs approximately 10,000 weekly traceroute measurements.

The measurement scripts aim to cover the entire IP prefix space, mostly focusing in AS and Point-of-Presence (PoP) topologies. DIMES collects the list of prefixes from the RouteViews project, providing roughly 400k prefixes. Using each prefix, a set of IP destinations is constructed. A typical script includes traceroute and ping commands to 60 destination IP addresses.

We note that although DIMES agents run on Windows, Linux and Mac, the measurement algorithms are implemented using raw-socket API, and do not change between different OSes. Although this method makes installation a bit complicated, it is essential in order to allow uniform execution of the measurements, regardless of the OS the agent operates on and without relying on a specific OS behavior.

Raw measurement data that is reported back to the server is filtered in order to remove trivial measurement artifacts that can later cause analysis mistakes. Traceroutes that exhibit some known problems [26], namely routing loops and the appearance of the destination address in the middle of the traceroute, are discarded from analysis. These measurements account for less than 0.1% of the total number of traceroutes performed. However, there still remains a substantial amount of non-trivial measurement artifacts that can lead to biased results. Therefore, additional filtering is applied as described next.

B. Data Filtering

In the context of DIMES measurements, we define a VP as an AS that homes one or more agents. The set of VPs (denoted by V) changes over time due to the churn in the agent population, and laptop-based agent mobility. Therefore, there is a need to correctly identify from where measurements are performed at a given point of time, and filter out agents that exhibit some abnormal behavior and may contribute bias. This can be achieved by using the AS from which the agent reports the results of the measurements. However, a mobile agent can perform measurements from one AS, and report them later from a different AS. This contributes a mis-identification mistake which is difficult to quantify.

Therefore, the identification of the AS that hosts the measurement is done by following each of the traceroutes until reaching a hop with a routable IP address that can be resolved into a valid AS. This method is not error proof since the routers in the hosting AS might be non-responsive and the first routable IP address might belong to a peering AS. In this case, the agent will be assigned the peering AS instead of the hosting AS. When multiple peering is used by the hosting AS, the agent might be assigned to several ASes, depends on the peering policies of its hosting AS, e.g., per-destination selection of egress point and load-balancing strategies.

To reduce VP identification mistakes, we limit the search to the first four hops in the path, hence ASes with multiple non-responsive hops will not induce mistake.

Approximately 52% of the agents that are resolved to an AS appear to be homed in more than one AS and 20% of the agents are homed in more than 5 ASes. However, most of the measurements of an agent are performed from one VP and the rest of the VPs appear to have only a few measurements, usually less than ten out of several thousands. This can either be a sign of mis-identification, or alternatively, the result of a laptop-based agent that perform measurements from multiple ASes. The latter, however, commonly results in a larger number of measurements per VP, since even laptop-based agent are usually homed in only a few ASes.

Accounting for these possibly mis-identified VPs, we filter out all the measurements for which $tr(a_i, vp_j) < T$, i.e., agent a_i measured from VP vp_j less than T traceroutes. Applying this filter with $T = 200$, reduces the number of agents that have more than 5 different VPs to less than 0.55% of the agents and leaves no agent that have more than 10 different VPs. Filtering with $T = 500$ yields less than 0.4% of the agents with more than 5 VPs. Since there is a trade-off between the accuracy of VP identification and loss of data due to over-filtering, we select $T = 200$ and estimate the extent of mis-identification errors.

The VP identification error of an agent is estimated as the percentage of measurements it performed from VPs suspected as mis-identified out of the overall number of its measurements. More formally, consider $tr(a_i, vp_j)$ to be the number of traceroutes performed by agent a_i from VP $vp_j \in V$, and the indicator x_j which marks a possibly mis-identified VP:

$$\forall j \in \{1..|V|\},\ x_j = \begin{cases} 1 & tr\,(a_i, vp_j) < T \\ 0 & otherwise \end{cases}$$

The VP identification error is thus given by:

$$err\,(a_j) = 100 \cdot \frac{\sum_j x_j \cdot tr(a_i, vp_j)}{\sum_j tr(a_i, vp_j)}$$

Using $T = 200$, Fig. 1 plots the cumulative distribution of the error estimation (right) and the mis-identification error as a function of the total number of measurements per agent. Most agents (73%) have no mis-identification error and only 7 agents (less than 1%) have 100% error estimation. Out of these, only 2 have more than 5,000 measurements, making these an indication of error since their measurements span across multiple (mis-identified) ASes, for each at most 200 measurements are performed. The overall error, i.e., the percentage of all measurements that are suspected as mis-identified out of the total measurements performed by all agents, is 0.23% (marked by the red line in the right plot), indicating that only a few agents (less than 4%) contribute most of the mis-identification, as can also be seen by the few marks above the zero line in the left plot. In the remainder of the paper, we use only measurements for which $tr(a_i, vp_j) \leq T,\ T = 200$, and refer to it as the *filtered* data.

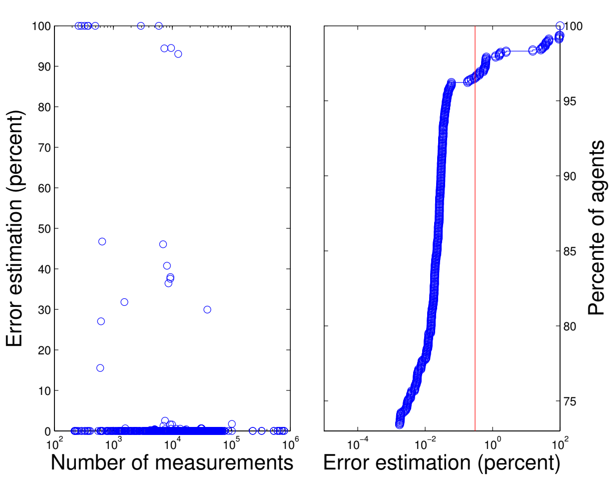


Fig. 1: Estimation of error in the identification of VPs

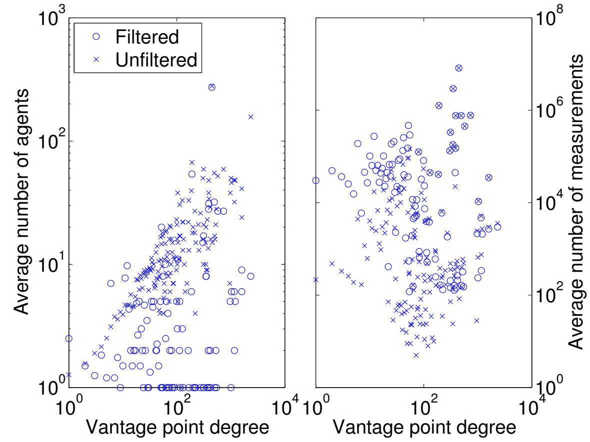


Fig. 2: Average number of agents and measurements per VP degree. Filtered data includes only agents from VPs where $tr(a_i, vp_j) \geq 200$

For each vp_i we extract the number of measuring agents, $agents(vp_i)$ and the number of total traceroute measurements performed from it, $tr(vp_i)$. Using the degree k_i of its hosting AS in the undirected AS graph, i.e., number of links to neighboring ASes, we calculate the average number of traceroute measurements per degree:

$$tr(k) = \frac{\sum\limits_{i \in \{k_i = k\}} tr(vp_i)}{|\{i | k_i = k\}|}$$

and average number of agents per AS degree:

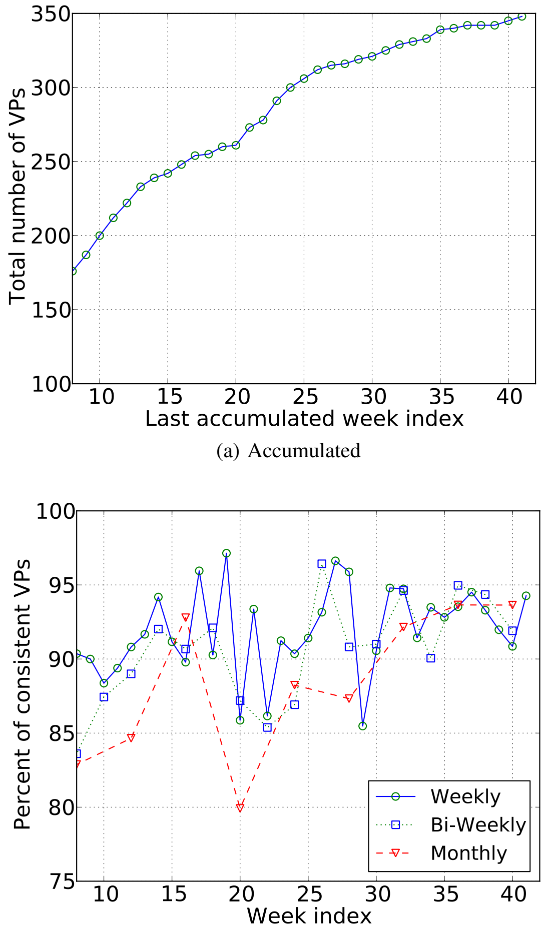
$$agents(k) = \frac{\sum\limits_{i \in \{k_i = k\}} agents(vp_i)}{|\{i | k_i = k\}|}$$

We repeat this calculation for the filtered set, and find $tr^f(k)$ and $agents^f(k)$.

Fig. 2 (left) shows that VPs with higher degree tend to have more agents. Using degrees extracted from RouteViews produced the same results, indicating that this is not a sampling bias caused by measurement artifacts [27], i.e., the high degree is not the result of having more agents but rather an indication that large ASes tend to "host" more agents. Additionally, Fig. 2 (right) shows no direct relationship between the number of measurements and the VP degree, showing that there is a good distribution of measuring agents across the different ASes.

C. Measurement Time-frame

Using a community-based project mandates the existence of a certain churn in VPs [28]. Therefore, we need to find the best time-period length during which measurements are collected. We therefore quantify the churn in the VPs during consecutive time-frames. First, Fig. 3a shows the total number VPs using increasing number of accumulated weeks from January until October 2009. The figure exhibits a steady increase in the number of VPs, showing that overall the number of new VPs suppresses the number of lost VPs. Fig. 3b shows the percentage of consistent VPs during consecutive periods of weeks, fortnights, and months. The figure shows that (a) for most time-frames, there are over 85% consistent VPs, (b) the set of VPs are less consistent during the first months of 2009 and more stable after July, and (c) weekly analysis exhibits a more consist set of VPs than monthly. This is since most of the inconsistency of the VP group over time comes from agents that register for a short period of time (or laptops taken on trips to uncommon locations) and not due to agents that stop working for a week or two, which is also seen from the constant increase in Fig. 3a. Thus, looking at longer time period does the opposite of smoothening the data.



(a) Accumulated

(b) Consistency

Fig. 3: VP churn during 2009 showing that (a) accumulating weeks results in a steady increase of the the VP count, and (b) VPs exhibit less churn on shorter time frames

These observations lead us to use data collected for a complete month during August 2009, where the churn is low. This helps obtain a set of relatively consistent VPs, allowing easier analysis of the large dataset. Moreover, using a complete month, increases the probability that there is a good coverage of the underlying topology, meaning that most ASes and AS-links are observed by more than one agent.

IV. Vantage Points Distribution Analysis

This section provides an in depth analysis of the various aspects related to the distribution of VPs and its effect on the resulting topology features.

A. Diminishing Returns

We now have the data needed to revisit the diminishing returns claim by examining how the observed topology changes as data from more VPs is added. Evaluating the effect of adding VPs is done by first building the *local* AS topology as observed from each of the VPs, i.e., the AS topology measured by agents when they are hosted in each VP.

TABLE I: The top-10 VPs sorted by ascending number of discovered AS-links per VP

AS-links	Degree	ASN	Name
36236	243	25229	Volia Ukraine
29552	166	286	KPN Netherlands
25565	4	33660	Comcast
24570	959	209	Qwest
23902	69	2116	Ventelo Networks
23009	695	3320	Deutsche Telekom
20416	12	25521	Industrial Media Network
18974	102	15435	Kabelfoon
18823	383	7922	Comcast IBONE
17816	39	19262	Verizon

Creating the AS-level topology from IP-level traceroutes provided from DIMES is achieved by performing AS resolution for each hop in all paths. We perform AS resolution by first applying longest-prefix-matching on BGP tables obtained from the RouteViews archive (using the same week being studied). This resolves approximately 98% of the IP addresses. The remaining 2% are queried against two WhoIs databases, namely RIPE and RADB. This resolves additional 1.5% of the IP addresses. The remaining 0.5% unresolved IP addresses are discarded and do not contribute ASes and AS-links to the topology.

In order to create an accurate topology, we use the list of Internet exchange points (IXPs) provided by PCH [29]. Whenever a traceroute contains an IP address that is resolved to an IXP prefix we create a direct link between the preceding and following ASes.

To further reduce the probability of including false links in our inferred topology, we do not include IP addresses that are resolved to AS sets or Multiple Origin ASes (MOAS) [20], [21], and treat them as unresolved. Furthermore, when constructing the complete DIMES AS topology, we only include ASes and links that exist in the local topologies of the filtered set of VPs, V^f (we drop the f notation since all further analysis uses only the filtered data). We believe that these steps help cope with most of the issues with translating IP traceroute measurements to AS level topologies [20], [21], [7], [22], and the possibly accumulated error does not significantly change the topology nor the results of our analysis.

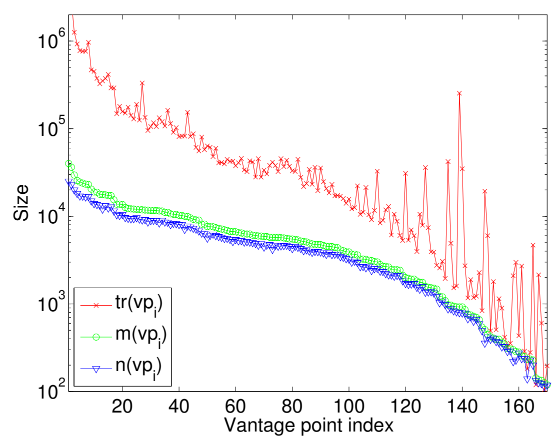
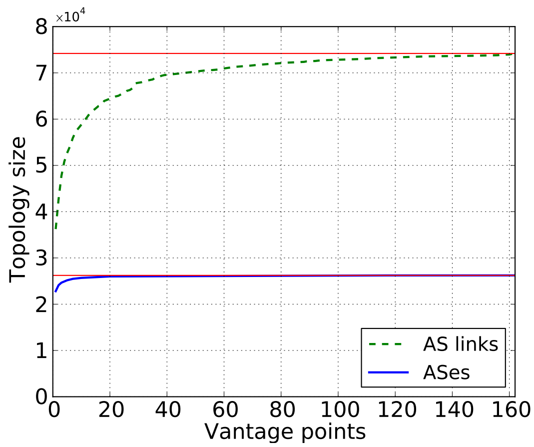


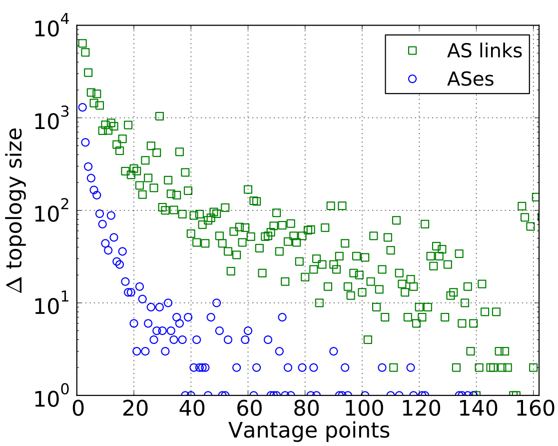
Fig. 4: Vantage point statistics sorted by ascending number of discovered AS-links per VP

We denote by $tr(vp_i)$, $n(vp_i)$ and $m(vp_i)$ the number of traceroutes performed, ASes and links in the local topology of vp_i, respectively. We then sort the set V in a non-increasing order of $m(vp_i)$. Fig. 4 depicts these values for each vp_i in the sort order, showing a high correlation between the number of measurements and the size of the topology. The top 10 ASes in this sort order are provided in Table I. This correlation, however, breaks at the tail of the ordered list.

We use this sort order in the following sections in order to quantify the effect of VP aggregation on various topology parameters. We point out that VPs that discover the largest local topologies are mostly located in Europe and the United States. Out of the top ten VPs, 7 are in the Europe and 3 are in USA (Qwest, Comcast and Verizon). Only the 13th VP is the first in a different geographical region (Israel) and the 30th is in Australia. This means that "remote" regions that are outside the areas used by other instrumented active measurement projects (e.g., Ark), and are likely to introduce new contributions to the AS topology, are considered only later in this sorted VP list. Table I shows that ASes with large degrees as well as ASes with medium and small degrees are represented in the top 10 largest topologies, indicating that our order is not slanted towards high degree ASes.



(a) Topology



(b) Δ Topology

Fig. 5: Number of ASes and AS-links in the aggregated topology

B. Aggregation of Local Topologies

Once we have the sorted set V we build a set of aggregated AS topologies, Agg. An aggregated topology agg_i includes the ASes and links from the set of local topologies, $agg_i = \cup_{j=1}^{i} vp_j$. Fig. 5a depicts the size of the aggregated topology as a function of the number of VPs and Fig. 5b depicts the contribution of each VP to the aggregate growth. The number of ASes almost reaches its final value after aggregating only a few VPs. We note that the full set of ASes

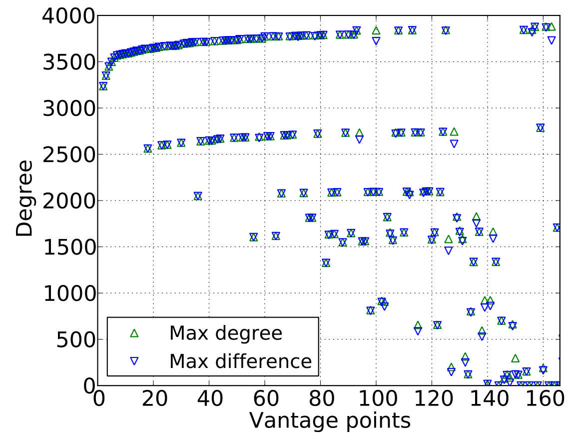


Fig. 6: Maximal AS degree and maximal neighboring AS degree difference of newly discovered AS links

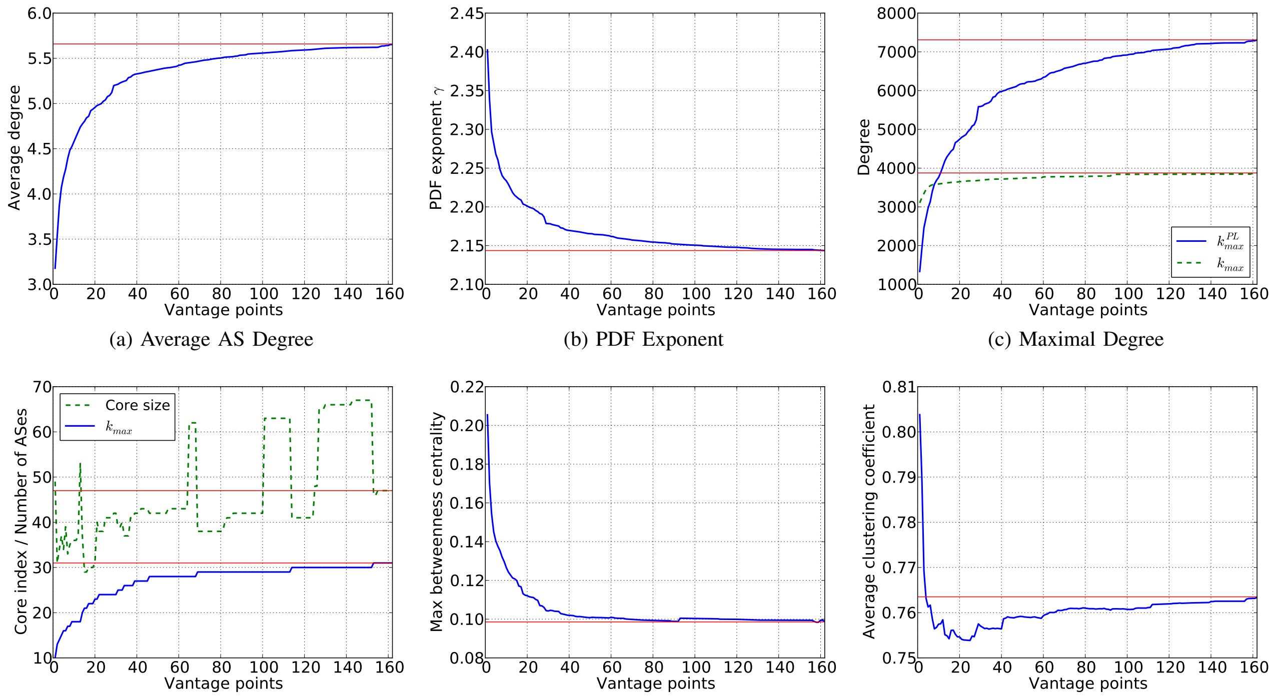
is actually quite easy to obtain from BGP messages without the need for large-scale measurement effort. The number of AS-links, on the other hand, which is much harder to obtain using passive collection efforts [17], gradually increases even when going over 80 VPs. Note that some VPs do not contribute new ASes or links and have thus no mark in Fig. 5b due to the semi-log scale.

The figure questions the diminishing returns argument in two ways. First, for AS-links, the return diminishes much slower than was observed in the past. Second, the tail of the distribution is indeed thick. This is especially interesting given the fact that in general as we add more VPs we get those with fewer measurements (see Fig. 4) due to the sort order. Therefore, the tail of the distribution in Fig. 5, where the number of measurements is in the few thousands, is actually a lower bound on the possible VP contribution.

To further understand the contribution of VPs in the tail, we collect all AS links that were added in each step of the aggregation and the set of ASes adjacent to these links. In this set, we find the maximal AS degree, and the maximal degree difference between ASes of an added link. The maximal degree shows whether the VP detects links in or to the core of the Internet. High degree difference indicates that the VP manages to detect "radial" links (meaning customer-provider [30], [31] links towards the core) and low degree difference indicates that the VP manages to detect only new "tangential" links (meaning peer-to-peer links). Fig. 6 shows that even VPs in the tail manage to discover new links towards the core. However, starting from roughly the 110th VP, it is possible to see VPs that contribute only "tangential" links.

We further study how merging observed topologies from distributed VPs affects several graph characteristics that are commonly used in graphs analysis. Following Latapy and Magnien [9], we examine the convergence trend of each analyzed property to the value of the overall AS graph by analyzing the graph characteristics of each AS graph $agg_i, 1 \leq i \leq |VP|$. Each graph property is plotted to show how the values extracted from the AS graphs agg_i converge to the final value of the property which is calculated on the complete AS graph using V. We show that while some properties converge to the overall value using only a few VPs, others converge slowly and require many VPs to reach the vicinity of the value of the complete topology.

We start by analyzing the **node degree distribution** which is the probability that a randomly selected node is of degree k. Let $n(k)$ be the number of nodes with degree k in a graph containing n nodes, the node degree distribution is: $P(k) = n(k)/n$. The degree distribution



(a) Average AS Degree (b) PDF Exponent (c) Maximal Degree

(d) k-core (e) Betweenness Centrality (f) Clustering Coefficient

Fig. 7: Topology characteristics analysis of aggregated AS graphs. The horizontal line marks the value calculated on the complete AS graph

has became one of the most frequently analyzed Internet topology characteristics [19] since the work of Faloutsos *et al.* [15] that showed that the degree distribution of Internet topologies follows a power-law, meaning $P(k) \sim k^{-\gamma}$, where γ is a positive exponent. We use the closely related Zipf [32] distribution, $n(k) \sim k^{-\alpha}$, and calculate $\gamma = 1/\alpha + 1$.

Fig. 7a and Fig. 7b show that the average degree and PDF exponent monotonically converge, reaching the vicinity (within 10%) of the overall value after roughly 40 VPs, and 120 VPs to reach within 1% of the overall value. Interestingly, the exponent value we get using all VPs ($\gamma = 2.14$) is quite similar to the one reported by Faloutsos *et al.* [15] who used a single VP ten years ago ($\gamma \simeq 2.20$).

Fig. 7c shows that the maximum degree converges even faster, indicating that the first few VPs accurately map the highest degree AS, namely Level3 (AS-3356). We also plot a theoretical maximum node degree in a power-law degree distribution, $k_{max}^{PL} = n^{1/(\gamma-1)}$ [33]. The theoretical maximum degree starts by converging nicely to the true maximal degree. If we had only about 10 VPs we would believe that the formula works well for the Internet AS graph. However, as we keep adding more VPs, k_{max}^{PL} climb very fast and eventually reaches a value, which is almost double the true maximal degree. A similar deviation from the strict power-law model, observed in [34], was attributed to the mixture of customer-provider links and peering links, having the first follow the model whereas the latter do not. Indeed, as is seen in Fig. 6, the first VPs detect mostly radial customer-provider links whereas VPs farther in the list detect more peer-to-peer links, causing the observed deviation from the power-law model. Overall, these findings further strengthen the recent claims against the accuracy of the power-law model for Internet topology [14], [16].

k-Pruning [35] is a method for decomposing graphs into *shells*, having each node being mapped to a shell based on its connectivity. Nodes in the first shell are those who have only one link leading to the 'center' of the graph. Nodes in the kth shell have k-connectivity towards the center. The *nucleus* (or *core*) is the shell with the highest index, k_{max}. Fig. 7d plots the value of k_{max} and the number of ASes in the core, when applying k-shell analysis on the aggregated AS graphs.

The nucleus index, shown in Fig. 7d, converges to 10% of its overall value ($k_{max} = 31$) after over 60 VPs. Additionally, the number of ASes in the nucleus is dynamic as we add more VPs. The drops we see in the number of ASes in the core is due to the separation of the core into two shells at the points where k_{max} increases, as more links are added. These changes occur even when the number of VPs is well over 100.

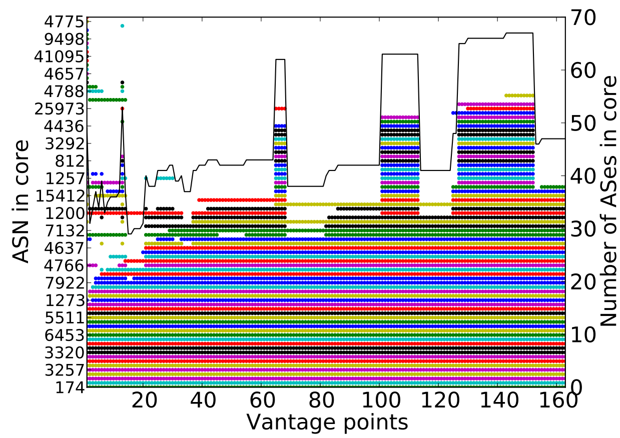


Fig. 8: AS membership in the nucleus of the aggregated AS graphs

In order to further understand the dynamics of the core, we looked at the ASes that are in the core and their connectivity. Fig. 8 depicts the core ASes in each aggregated topology. Each horizontal line represent the membership of a specific AS (not all AS numbers are shown for brevity of the plot) in the aggregations by a dot where the AS was a core member of agg_i. The right y-axis shows the overall number of ASes (number of dots in each agg_i) in the core, which is illustrated by the line. The figure shows that the true core is comprised of about 43 ASes that are present in all the aggregations after roughly 90 VPs, and additional 20 or so join and leave the core almost simultaneously. The simultaneous departure is the result of detecting more links in the core, which in turn increases the connectivity between the real core ASes and causes them to separate into a new higher shell. This can also be seen in Fig. 7d, by observing that each drop in the number of ASes in the core after the 60th aggregation, is perfectly aligned with the increase in the core index (k_{max}). We also observed that the average degree of the ASes in the true core graph (i.e., not their overall degree, but their degree after the pruning process) is roughly 40, indicating that the core is dense, with almost a full mesh structure. However, when the core becomes larger with more than 60 ASes, the average degree does change, making the core sparse. Another important observation is that in the first 16 aggregations, there are about 20 VPsthat are not seen afterwards, indicating that they are mistakingly included in the core due to insufficient aggregation.

Given that the core extracted using k-pruning captures a global view of the Internet top-level providers [35], it holds both tier-1 and tier-2 ASes, hence even if finding most tier-1 links can be done only BGP data [18], it takes a much more comprehensive probing to find the complete Internet core. This indicates that finding links in the Internet core, which captures a broader concept than tier-1 ASes, is not as easy as is believed.

Betweenness centrality (bc) is commonly used for measuring the centrality of a node or a link. Node betweenness measures the number of shortest paths passing through a node as an estimate to the potential traffic load on this node assuming uniformly distributed traffic which follows shortest paths [19]. We calculate the maximal betweenness over all ASes in each Agg_j as a measure for possible congested nodes in the graph. In order to compare topologies with different sizes, we normalize the average betweenness by the maximal possible betweenness value, $n_j(n_j - 1)$. Fig. 7e shows that bc converges within 10% of the overall value after roughly 20 VPs. Since bc is a measure of load, adding links almost always decreases the average load on ASes (except pathologies like the Braess paradox), thus the monotonic descent. Since high-degree ASes are more "central" than low-degree ASes, we expect that tier-1 ASes will have the maximal betweenness values. Indeed, Level3 (AS3356), a tier-1 AS, is the AS with the maximal bc for all resulting graphs.

Clustering Coefficient (cc) of a graph measures the local cliquishness of a node neighborhood [36]. Simply put, clustering coefficient estimates how close a given node and its immediate neighbors are from being a clique. A graph average clustering coefficient is used to estimate how close a graph is to a small-world network, such that graphs with higher average clustering coefficient can be better modeled by a small-world network. Fig. 7f shows that the cc is surprisingly estimated to within 2% of the final value after only 3 VPs and stays within this narrow distance from the final value while jugging up and down. The value reported here, which is roughly 0.76, is quite different from what was reported for topologies collected in the Skitter project (0.46) and WHOIS repositories (0.49) [19].

In the conference version of this paper [25] we presented most of the analysis of this subsection but for a single week in early 2008. While the topologies (both per VP and the aggregates) are now larger by almost 60%, the convergence properties are similar.

C. Sampling Bias

Several studies [27], [37], [38] analyze the bias that the commonly used traceroute sampling method potentially introduces into the inferred topology. In [27] the router-level topology inferred using traceroute sampling was shown to be biased by the distance between the measuring VP and the probed interface. This claim was partially confronted by showing [37] that that various traceroute exploration strategies can produce topologies with minimal bias.

We expect that achieving a broad distribution of VPs, alongside with the relatively low diameter of the AS-level topology, should result in a good sampling process of the underlying topology so that it will exhibit less bias that result from the distance between VPs and observed ASes.

Measuring the distance between VPs to ASes is done by searching for the shortest valley-free path between each VP and the ASes it observes. A valley-free path follows a strict hierarchical structure – an uphill segment of zero or more customer-provider or sibling links, followed by zero or one peer-to-peer link, followed by a downhill segment of zero or more provider-customer or sibling links. For this end, we customized the shortest-path Dijkstra algorithm to obey the valley-free routing rules. Calculating a valley-free path requires the inference of the type-of-relationship between adjacent ASes (customer-provider, peers or siblings) in the AS graph. The relationships are inferred using the near-deterministic type-of-relationship algorithm [31].

First, for each AS we find all the VPs that include it in their observed topology (referred to as "observing VPs"), and calculate the number of hops from the AS to each of them. Fig. 9 shows the average number of observing VPs calculated over all ASes with a given degree obtained from DIMES and RouteViews AS-level topologies. As expected, the figure shows that low-degree ASes are observed from much fewer VPs than high-degree ASes. This is attributed to the fact that small degree ASes are harder to detect and probe. However, there are a few ASes that have high degree (mostly in RouteViews) and are observed by only few VPs, but these are quite rare.

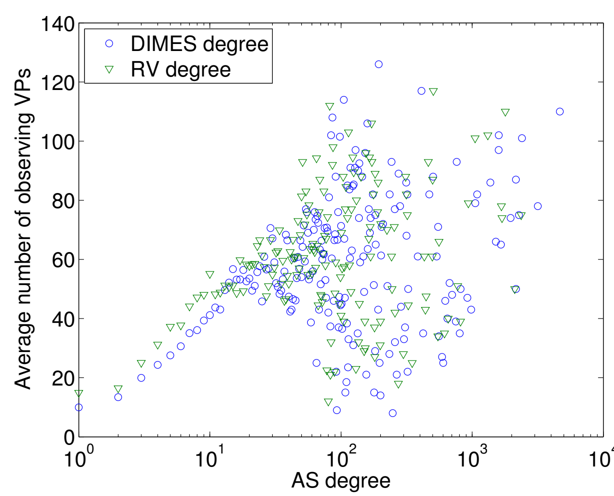


Fig. 9: Number of observing VPs per AS degree

Table II shows the distribution of the number of ASes per distance to the nearest observing VP. The table shows that most ASes are 1 to 3 hops away from the nearest VP, 191 ASes serve as VPs (thus are zero hops away), and a small fraction of ASes (just over a hundred) have VPs that are 4 or 5 hops away. The average distance is 1.9 hops with a standard deviation of 0.62, and the median is 2 hops, showing that the set of VPs are well spread.

TABLE II: Distances from AS to nearest vantage point

Distance to nearest VP	0	1	2	3	4	5
Number of ASes	191	4727	17482	4695	106	2

Fig. 10 shows the average distance over all ASes with a given degree (using DIMES and RouteViews) to the nearest VP. The standard deviation surrounding the average is less than 1 hop. While it is possible to see a correlation between the distance and AS degrees, the reason for this is not sampling bias but rather probabilistic distribution of agents. As seen in Fig. 2 the probability that an agent is located in a given AS is roughly proportional to the degree of the AS. When using degrees from the RouteViews AS graph, the bias is even less noticed, since there are high-degree ASes that are relatively far away (2 hops) from the nearest VP.

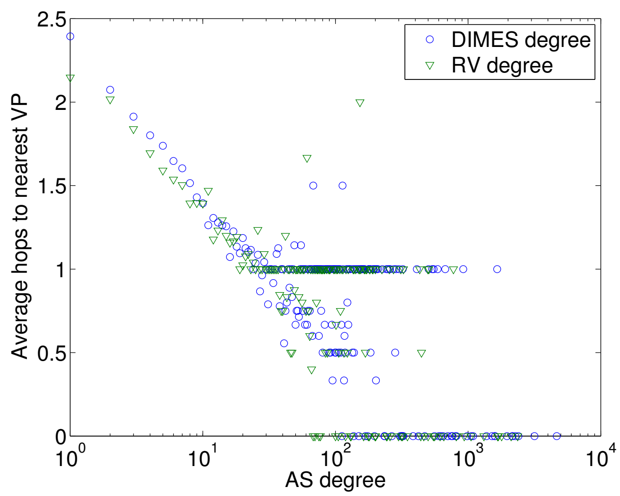


Fig. 10: Number of hops to nearest VP per AS degree

D. Diversity Bias

We further wish to study the effect of VPs distribution by examining how the diversity of the types of ASes from which measurements are performed affects the observed topology.

For this analysis we use the AS-level topology obtained from the iPlane Atlas [39]. Since iPlane uses PlanetLab and traceroute servers, the identification of VPs is precise, and was provided to us from the iPlane team.

The iPlane Atlas is a compact way of representing the AS-level topology measured by iPlane, so that it can be easily pushed to end-users for path prediction applications in peer-to-peer networks. The Atlas contains a large set of 31,611 ASes, which is almost all of the 32,326 ASes seen by RouteViews, as opposed to the 27,203 seen by DIMES. However, the connectivity is measured by collecting traceroutes of only a few days, hence is expected to be somewhat lower than in DIMES. However, as we later show, the smaller topology is attributed mostly to bias in the type of measuring VP rather than the duration of measurements.

We first study the geographical distribution of DIMES and iPlane's VPs. Finding the country of a VP was done by querying the commercial MaxMind database with the IP addresses of the DIMES agents and iPlane's sources. A few VPs were resolved to more than one geographic location and were removed from this analysis. We note that the technique for constructing MaxMind database is not publicly known. However, although it was shown [40] to have some mistakes in city resolution, it is very accurate in the country level, which we use in the following analysis.

In DIMES, the largest number of VPs are in the USA (40.8%), followed by UK (8%), Russian Federation (7.3%), Germany (6.5%) and Canada (5.8%). Other VPs are spread over the entire globe, including most of the European countries, some countries in the Middle East, the Far East, Australia, South America and South Africa. There are no VPs in Central Africa and the Arab countries in the middle east. iPlane VPs are also mostly located in the USA (43.3%). Like DIMES, there are no VPs in Central Africa and the Arab countries. This analysis shows that both projects cover a relatively large geographical area, which should assist in locating ASes in remote locations.

We compare the VPs of iPlane and DIMES by examining their AS types. The AS type is determined using data from Dimitropoulos *et al.* [41] that uses a machine learning approach to classify an AS as a large ISP (t1), a small ISP (t2), an academic network (edu), an Internet exchange point (ix), a network information center (nic) which holds important network infrastructure, or a customer (comp) of either the small or large ISPs. We manually classified additional ASes, such as those having a description that contains the word "university" that were missing from the inferred set. In total we used 18,639 classified ASes, which do not cover all the ASes in the datasets but are sufficient for this analysis. ASes with unknown types are ignored, and both infrastructures have a similar number of unclassified ASes.

TABLE III: Vantage point types of iPlane and DIMES

	t1	t2	edu	comp	ix	nic	unknown
iPlane	17	104	117	22	3	5	46
DIMES	29	106	10	11	2	1	47

Table III provides the number of VPs for each AS type. It shows that iPlane uses much more academic VPs than DIMES. This is mainly the contribution of the PlantLab servers that are used for iPlane distribution, whereas only a few were used in DIMES, mainly to achieve high geographical diversity, such as the Far East, Africa and South America. DIMES has more VPs in tier-1 ISPs and the majority of VPs reside in tier-2 ASes, which align with its community-based distribution.

TABLE IV: Number of ASes per type for which DIMES degrees are larger than iPlane ($D > P$), smaller ($D < P$) and equal ($D = P$)

	t1	t2	edu	comp	ix	nic	unknown
$D > P$	26	2392	309	2436	20	70	4239
$D < P$	17	1198	153	1679	4	49	2410
$D = P$	1	974	283	3760	4	54	5349

For each AS that exists in both AS-level topologies (26,261 ASes out of 31,611 ASes in iPlane and 27,203 ASes in DIMES), we calculated the ratio between the iPlane measured degree and DIMES measured degree. Table IV shows the number of ASes for each type when DIMES degrees are larger than iPlane's ($D > P$), smaller ($D < P$) and equal ($D = P$).

The table shows that although the projects agree on the degrees of almost 10,500 ASes, DIMES exhibits overall higher degrees, especially for tier-2 networks, IXPs, and most surprisingly academic (edu) networks. However, looking at Table V, which shows the average AS degree per type, reveals that iPlane has a higher average degree for academic networks. Further examining these ASes, we find that for the academic networks that haver higher degree in DIMES, the average ratio between DIMES and iPlane degrees is 2.4. However, for the fewer academic networks that have higher iPlane degree, the average ratio between iPlane and DIMES degrees is 4.2. This shows that while DIMES has an overall better probing of the academic networks, as it has for the majority of the observed ASes, iPlane has a significantly better view of some of these networks.

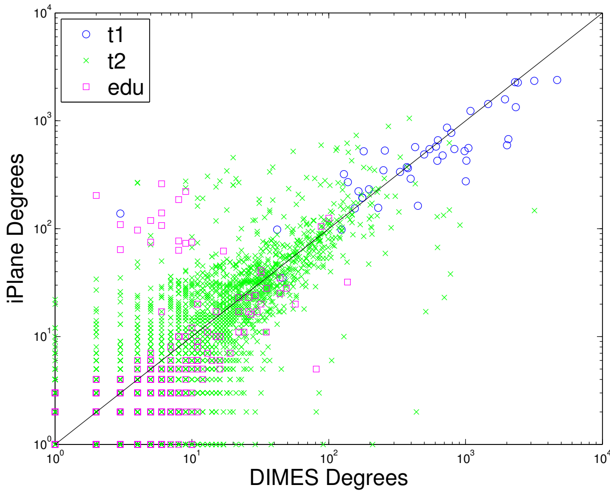


Fig. 11: Comparison of iPlane and DIMES degrees for tier-1, tier-2 and academic networks (edu)

These observations are further depicted in Fig. 11, which shows a scatter plot of iPlane degree vs. DIMES degree for tier-1, tier-2 and academic ASes. The figure shows that for the vast majority of tier-1 and tier-2 ASes both topologies degrees are similar, while for some academic networks iPlane measures significantly higher degrees than DIMES.

The two infrastructures agree on the degree of only a single tier-1 AS (AS372 Korea Telecom, degree 372). However, the average degree of tier-1 ASes is very high (more than 600 for both projects), hence the chances to agree are low. Moreover, although the average ratio of degrees (DIMES/iPlane) for tier-1 ASes is 2.04, the standard deviation is 6.7, showing that the degrees are relatively similar but the variance is high.

TABLE V: Average AS degrees using RouteViews, DIMES and iPlane per AS type

	t1	t2	edu	comp	ix	nic	unknown
RouteViews	606.2	9.9	3.0	2.3	6.4	4.7	5.1
DIMES	864.5	15.0	4.3	3.0	75.2	5.5	9.0
iPlane	656.0	15.1	5.6	2.7	18.2	8.1	6.3

Table V compares average AS degrees per AS type of the DIMES, iPlane, and RouteViews AS graphs. RouteViews passively collects BGP messages and is considered a reliable source for the AS topology, although it can miss links between ASes that do not participate in BGP, which is more common in tier-2 ASes [17]. Table V provides the average AS degree per type using data collected during August 2009 and the iPlane Atlas. It shows that for all AS types RouteViews has the lowest average degrees. As stated before, iPlane observes higher average degree in academic networks networks, whereas DIMES has higher degrees in tier-1 and IXPs. These differences are attributed to bias that each project holds towards its VP types. The observation that DIMES has much higher degree for IXPs is mostly attributed to incorrect inference of their connectivity, whereas the iPlane Atlas employs more accurate connectivity analysis [39] for IXPs. Interestingly, the average degree of tier-2 ASes is almost identical in DIMES and iPlane, while both are 50% more than RouteViews. This can possibly be attributed to the difference in the probing method (active vs. passive) and placement of VPs.

TABLE VI: Average VP degree per type, calculated over origin ASes of iPlane, DIMES and both

VPs	Avg. Degree	t1	t2	edu	comp	ix	nic
iPlane	iPlane	1004	51.1	19	6	20	3.6
	DIMES	1510	88.4	9.5	8.7	324.6	5.8
DIMES	iPlane	834.4	84.8	7.6	10	39.5	37
	DIMES	1155.5	141.3	21.3	13	787	44
Both	iPlane	1068.1	84.5	8.2	-	38	-
	DIMES	1637.7	292.8	23	-	841	-

Finally, we evaluate how the presence of a VP within an AS affects its observed connectivity to other ASes. Table VI provides the average VP degree per type, calculated over VPs of iPlane, DIMES and both. Strengthening the above observations, DIMES has higher average degrees for almost all VPs except academic networks, which are the primary distribution method of iPlane. Interestingly, DIMES has higher degrees in tier-1 ASes that are not part of its VP set. We attribute this mostly to the relatively easy probing of tier-1 ASes, as many traceroutes pass through them [35], therefore enabling easy probing even from the “outside”.

E. Discussion

The above analysis illustrates several important results. First, it shows that although increasing the number of VPs can help reducing sampling bias, it still does not guarantee unbiased results. Although both iPlane and DIMES have a very broad distribution, the types of ASes in which their VPs are located generate a topology that is biased towards these types. Most obvious are the academic networks which are probed significantly better in iPlane than in DIMES. Overcoming this bias cannot be achieved by simply increasing the number of VPs but rather a broad diversity in types is required.

Second, the analysis strengthens the assumption that measuring from within a network is important for discovering more of its links, mainly for low-tier ASes, as these are harder to probe than tier-1 ASes. However, it also exposes that using a large spread of VPs, that perform many measurements for a sufficient period, can still result in extensive coverage of networks, even from the “outside” of the network.

Finally, we found that in tier-2 ASes both iPlane and DIMES have significantly higher degrees than RouteViews. We attribute this to several key features of Internet measurement efforts. First, DIMES and iPlane have significantly more VPs than RouteViews, especially in these lower level ASes, enabling better probing. Second, RouteViews mostly measures from the core of the Internet while iPlane and even to a greater extent DIMES, measure from the far reaches of the Internet. Due to the valley-free routing policy, certain types of peer-to-peer links are only visible from the customers of the peering ASes [11], [34]. Thus, the location of VPs in these ASes is required in order to gain a good coverage and have a more complete view of the Internet topology. Finally, RouteViews uses passive collection of BGP messages, hence it is only capable of capturing links that are published in BGP. In recent years, there is an increasing usage of private peering between ASes [42], [43], making these links visible only to active probing. Although active measurements are more challenging to analyze, the benefits of improved probing of networks is significant.

V. Conclusion

This paper presents an analysis of the significance of the distribution of vantage points (VPs) in an active Internet measurement infrastructure. We showed that diverse and broad distribution can help overcome sampling bias and uncover hidden parts of the Internet topology. However, even broadly distributed infrastructures still exhibit some bias towards the type of ASes from which measurements are performed, further stressing the need for a diversity of VP types and geographic locations. To this end, community based infrastructures are well suited since their growth potential is theoretically unlimited. Looking at various commonly analyzed graph properties, we showed that some require more than 40 VPs to converge, but surprisingly several others converge with only a few VPs.

REFERENCES

[1] University of Oregon Advanced Network Technology Center, "Route Views Project, http://www.routeviews.org/."

[2] CAIDA Macroscopic Topology Project Team, "Caida archipelago, next generation active measurement infrastructure (http://imdc.datcat.org/collection/)."

[3] B. Chun, D. Culler, T. Roscoe, A. Bavier, L. Peterson, M. Wawrzoniak, and M. Bowman, "PlanetLab: An overlay testbed for broad-coverage services," *ACM SIGCOMM Computer Communications Review*, vol. 33, no. 3, Jul. 2003.

[4] H. V. Madhyastha, T. Isdal, M. Piatek, C. Dixon, T. E. Anderson, A. Krishnamurthy, and A. Venkataramani, "iPlane: An information plane for distributed services," in *OSDI*, 2006, pp. 367–380.

[5] N. T. Spring, R. Mahajan, and D. Wetherall, "Measuring ISP topologies with Rocketfuel," in *SIGCOMM*, 2002, pp. 133–145.

[6] Y. Shavitt and E. Shir, "DIMES: Let the Internet measure itself," *ACM SIGCOMM Computer Communications Review*, vol. 35, no. 5, 2005.

[7] K. Chen, D. R. Choffnes, R. Potharaju, Y. Chen, F. E. Bustamante, D. Pei, and Y. Zhao, "Where the sidewalk ends: Extending the internet AS graph using traceroutes from P2P users," in *ACM CoNEXT*, Dec. 2009.

[8] kc claffy, M. Crovella, T. Friedman, C. Shannon, and N. Spring, "Community-oriented network measurement infrastructure (conmi) workshop report," *SIGCOMM Computer Communication Review*, vol. 36, no. 2, pp. 41–48, 2006.

[9] M. Latapy and C. Magnien, "Complex network measurements: Estimating the relevance of observed properties," in *IEEE Infocom*, 2008.

[10] P. Barford, A. Bestavros, J. Byers, and M. Crovella, "On the marginal utility of network topology measurements," in *IMW'01: the 1st ACM SIGCOMM Workshop on Internet Measurement*, 2001, pp. 5–17.

[11] R. Cohen and D. Raz, "The Internet dark matter - on the missing links in the AS connectivity map," in *IEEE INFOCOM*, Barcelona, Spain, Apr. 2006.

[12] R. Beverly, A. Berger, and G. G. Xie, "Primitives for active Internet topology mapping: Towards high-frequency characterization," in *IMC*, 2010.

[13] D. Achlioptas, A. Clauset, D. Kempe, and C. Moore, "On the bias of traceroute sampling: or, power-law degree distributions in regular graphs," in *STOC*. Baltimore, MD, USA: ACM, 2005, pp. 694–703.

[14] B. Krishnamurthy and W. Willinger, "What are our standards for validation of measurement-based networking research?" *SIGMETRICS Perform. Eval. Rev.*, vol. 36, pp. 64–69, Aug. 2008.

[15] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On power-law relationships of the Internet topology," *ACM SIGCOMM Computer Communications Review*, vol. 29, no. 4, pp. 251–262, 1999.

[16] Q. Chen, H. Chang, R. Govindan, S. Jamin, S. J. Shenker, and W. Willinger, "The origin of power laws in Internet topologies revisited," in *INFOCOM*, 2002, pp. 608–617 vol.2.

[17] R. V. Oliveira, D. Pei, W. Willinger, B. Zhang, and L. Zhang, "In search of the elusive ground truth: the Internet's AS-level connectivity structure," in *SIGMETRICS*. ACM, 2008, pp. 217–228.

[18] R. Oliveira, D. Pei, W. Willinger, B. Zhang, and L. Zhang, "The (in)completeness of the observed Internet AS-level structure," *IEEE/ACM Transactions on Networking*, vol. 18, pp. 109–122, February 2010.

[19] P. Mahadevan, D. Krioukov, M. Fomenkov, B. Huffaker, X. Dimitropoulos, kc claffy, and A. Vahdat, "The Internet AS-level topology: Three data sources and one definitive metric," *ACM SIGCOMM Computer Communications Review*, vol. 36, p. 17, 2006.

[20] Z. Mao, J. Rexford, J. Wang, and R. H. Katz, "Towards an accurate AS-level traceroute tool," in *SIGCOMM*, 2003.

[21] Z. Mao, D. Johnson, J. Rexford, J. Wang, and R. H. Katz, "Scalable and accurate identification of AS-level forwarding paths," in *INFOCOM*, 2004.

[22] "TraceNET: an Internet topology data collector," in *IMC*. ACM, 2010.

[23] R. V. Oliveira, B. Zhang, and L. Zhang, "Observing the evolution of Internet AS topology," *ACM SIGCOMM Computer Communications Review*, vol. 37, no. 4, pp. 313–324, 2007.

[24] A. Dhamdhere and C. Dovrolis, "Ten years in the evolution of the Internet ecosystem," in *IMC*. ACM, 2008, pp. 183–196.

[25] Y. Shavitt and U. Weinsberg, "Quantifying the importance of vantage points distribution in internet topology measurements," in *Infocom*, Rio de Janeiro, Brazil, Apr. 2009.

[26] B. Augustin, X. Cuvellier, B. Orgogozo, F. Viger, T. Friedman, M. Latapy, C. Magnien, and R. Teixeira, "Avoiding traceroute anomalies with Paris traceroute," in *IMC*, 2006, pp. 153–158.

[27] A. Lakhina, J. W. Byers, M. Crovella, and P. Xie, "Sampling biases in IP topology measurements," in *INFOCOM*, Apr. 2003.

[28] D. Stutzbach and R. Rejaie, "Understanding churn in Peer-to-Peer networks," in *IMC*. ACM, 2006, pp. 189–202.

[29] "Packet clearing house, http: //www.pch.net/resources/data/exchange-points." [Online]. Available: http: //www.pch.net/resources/data/exchange-points

[30] L. Gao, "On inferring autonomous system relationships in the Internet," *IEEE/ACM Trans. on Networking*, vol. 9, no. 6, pp. 733–745, 2001.

[31] Y. Shavitt, E. Shir, and U. Weinsberg, "Near-deterministic inference of AS relationships," in *ConTEL*, 2009.

[32] G. K. Zipf, *Human Behavior and the Principle of Least-Effort*. Addison-Wesley, 1949.

[33] S. Dorogovtsev and J. Mendes, *Evolution of Networks: From Biological Nets to the Internet and WWW*. Oxford University Press, Jan. 2003.

[34] Y. He, G. Siganos, M. Faloutsos, and S. Krishnamurthy, "Lord of the links: a framework for discovering missing links in the Internet topology," *IEEE/ACM Trans. on Networking*, vol. 17, pp. 391–404, Apr. 2009.

[35] S. Carmi, S. Havlin, S. Kirkpatrick, Y. Shavitt, and E. Shir, "A model of Internet topology using k-shell decomposition," *Proceedings of the National Academy of Sciences USA (PNAS)*, vol. 104, no. 27, pp. 11 150–11 154, Jul. 3 2007.

[36] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks." *Nature*, vol. 393, no. 6684, pp. 440–442, June 1998.

[37] L. Dall'asta, I. Alvarez-Hamelin, A. Barrat, A. Vazquez, and A. Vespignani, "A statistical approach to the traceroute-like exploration of networks: theory and simulations," Jun. 2004.

[38] R. Cohen, M. Gonen, and A. Wool, "Bounding the bias of tree-like sampling in IP topologies," *Networks and Heterogeneous Media*, vol. 3, no. 2, pp. 323–332, Jun. 2008.

[39] H. V. Madhyastha, E. Katz-Bassett, T. Anderson, A. Krishnamurthy, and A. Venkataramani, "iPlane Nano: path prediction for peer-to-peer applications," in *NSDI'09*, 2009, pp. 137–152.

[40] S. S. Siwpersad, B. Gueye, and S. Uhlig, "Assessing the geographic resolution of exhaustive tabulation for geolocating Internet hosts," in *PAM*. Springer-Verlag, 2008.

[41] X. Dimitropoulos, D. Krioukov, G. Riley, and kc Claffy, "Revealing the AS taxonomy: The machine learning approach," in *PAM*, 2006.

[42] C. Labovitz, S. Iekel-Johnson, D. McPherson, J. Oberheide, and F. Jahanian, "Internet inter-domain traffic," in *SIGCOMM*. ACM, 2010, pp. 75–86.

[43] P. Gill, M. Arlitt, Z. Li, and A. Mahanti, "The flattening Internet topology: natural evolution, unsightly barnacles or contrived collapse?" in *PAM*. Springer-Verlag, 2008, pp. 1–10.