

Efficient Estimation of More Detailed Internet IP Maps

Sangmin Kim and Khaled Harfoush
Department of Computer Science
North Carolina State University
E-mail: {skim12, harfoush}@cs.ncsu.edu

Abstract—Router-level maps of the Internet implicate a large body of research on network management, overlay networks, performance evaluation, and security. The inaccuracies in these maps result in misleading conclusions. In this paper, we propose AROMA (Accurate Router-level MAP), a tool to infer router-level, layer-3 maps of the Internet. AROMA uncovers more routers and links in targeted (mapped) networks than existing tools with less probing overhead. For example, AROMA reveals the same number of routers and links as the Rocketfuel tool after sending less than 5.1% of the number of probes used by Rocketfuel, and reveals at least 100% more links and routers than Rocketfuel while using the same number of probe packets.

We use AROMA to draw the maps of four major ISP networks and revisit the conclusions drawn by earlier research on the Internet IP structure. Surprisingly, AROMA maps consistently reveal that core routers have a higher degree than edge routers in contrast to the recently suggested higher connectivity of routers at the network edge. The maps also reveal that routers' degree distribution follows a power-law in contrast to the recently suggested Weibull distribution.¹

I. INTRODUCTION

Starting from an experimental infrastructure with few hosts and routers, the Internet has evolved to an enormous network with multi-million nodes, and is directly affecting our lives in many ways. Understanding the structure and the characteristics of the Internet router-level maps (IRLM) is essential to: (1) improve network management capabilities and help better plan for network infrastructures, (2) optimize routing protocols and optimize the construction of overlay networks, (3) simulate new network protocols and evaluate their performance before actual Internet deployment, (4) promote the deployment of network-aware applications, and (5) protect the Internet by isolating security breaches and by identifying vulnerable portions of the Internet that need to be reinforced.

The huge size and complexity of the Internet, its anarchistic evolution, together with the fact that ISPs do not publish their router-level maps for security reasons contribute to the lack of an accurate IRLM. Despite the large body of research targeted at unveiling the IRLM [16], [23], [25], the task remains challenging and the conclusions drawn from these efforts are to-date a source of controversy and debate in the research community.

In general, techniques to uncover IRLM can be classified into two categories. The first relies on a set of k controlled hosts (sources) spread across the Internet sending *traceroute* packets to m arbitrary Internet destinations [3], [8], [14], [16], [20], [26] – Figure I (A,B). The topology incorporating *all* the identified routers, together with the links connecting them, is used to represent the IRLM. Donnet et al. have proposed techniques to make this approach more efficient by avoiding probing redundancy over already visited hops [12]. However, the k - m traceroute approach has been questioned in [18] and shown to be biased towards routers that are close to the source hosts. That is, links to routers that are close to the source hosts are revealed better than those close to the destinations. Therefore, the resulting graph is deemed unrepresentative of the IRLM. The second category does not traceroute to arbitrary destinations; instead, it focuses on an internet *area*, typically corresponding to an autonomous system (AS). Published BGP tables permit wise selection of the traceroute destination hosts, such that traceroute packets cross the target AS, thus avoiding probing redundancy – Figure I (C,D). The fact that it is targeting the complete router-level topologies of ASes permits the study of common factors driving the layout of AS infrastructures, the factors driving AS interconnectedness, and the structure of IRLM in general. The Rocketfuel [25], [27] tool belongs to this category. As we make the case throughout this paper, Rocketfuel unveils mostly *backbone* routers and links of the targeted AS. This is mainly due to the Border Gateway Protocol (BGP) [22], the mainstream inter-domain routing protocol, which typically leads transit probes to the closest exit router through the targeted AS's backbone links.

Contributions: In this paper, we propose AROMA, an IRLM mapping tool. AROMA tries to identify all interfaces/links associated with each router in a targeted AS by *directly* probing these interfaces from multiple vantage points – Figure I (E,F). As opposed to k - m traceroute, *only* interfaces that have been probed from the vantage points are included in the final map. As opposed to rocketfuel, AROMA does not target address prefixes that transit the targeted AS and thus can deeply penetrate inside an AS; and while Rocketfuel relies on *insider* probing servers, if available, to probe *useful* prefixes identified by BGP tables, traced paths still mainly follow backbone paths in the targeted AS towards their intended destinations, and mostly miss the AS details. Address space information of

⁰This work is partially supported by NSF grant CAREER ANIR-0347226.

¹The authors will make the collected data publicly available before the camera ready version time.

targeted ASes is typically available from authorized registry services like ARIN [1], APNIC [2], and RIPE [6]. The vantage points are filtered-out to avoid probing redundancy and the probed address space is filtered-out to avoid probing IP addresses assigned to end-hosts and IP addresses assigned to nodes in customer ASes, which partly contributes to AROMA’s efficiency. Furthermore, AROMA trades off its revealed AS details (*completeness*) for efficiency. These efficiency-aware enhancements result in significant efficiency. Specifically, our results indicate that AROMA reveals the same number of routers and links as Rocketfuel after sending less than 5.1% of the number of probes consumed by Rocketfuel in all investigated ASes, and reveals between 100% and 1700% more links and routers than Rocketfuel after consuming the same number of probes as Rocketfuel. We use AROMA to draw the maps of four major ISP networks (SprintLink, Level3, Verio, and Abovenet) and report on the structure of their networks. We also revisit the conclusions drawn by earlier research on the Internet structure and the degree distribution of the Internet routers.² Surprisingly, AROMA maps consistently reveal that core routers have a higher degree than edge routers in contrast to the conclusion drawn in [19]. The maps also reveal that routers’ degree distribution consistently follows a power-law. This is in contrast to the widely accepted Weibull distribution revealed in the Rocketfuel traces, and in agreement with the highly questioned k - m traceroute tools. These results highlight the need for more thorough investigations of the different factors driving the Internet structure and behavior.

The rest of the paper is organized as follows. In Section II we elaborate on the basic techniques used by AROMA to draw maps. We illustrate the techniques through our four case studies for the Sprintlink, Level3, Verio, and Abovenet networks. In Section III we highlight tradeoff between completeness and efficiency in the AROMA maps and the edge that it enjoys over current tools. In Section IV we analyze the AROMA maps for the four case studies and question common perceptions about the Internet router-level topology. We finally conclude in Section V.

II. DETAILED AROMA MAPS

The Aroma mapping process is sketched in Figure 2. Initially, AROMA is fed the IP address space corresponding to the targeted AS from an authorized registry service [1], [2], [6] and a list of potential probing servers (controlled machines). These lists are refined in Steps 1 and 2 and only k servers and m IP addresses are selected. The process then proceeds recursively in phases. In each phase, targeted interfaces are probed and a new list of target interfaces is identified, which is then fed to the next phase for further processing – steps 3 and 4. The process converges when no new targets are identified. Following convergence, a process for alias resolution is carried

²Our choice of the degree distribution simply reflects the Internet measurement community’s interest in this metric. Finding more interesting metrics that are capable of capturing the Internet structural and behavioral characteristics is an interesting research area on its own.

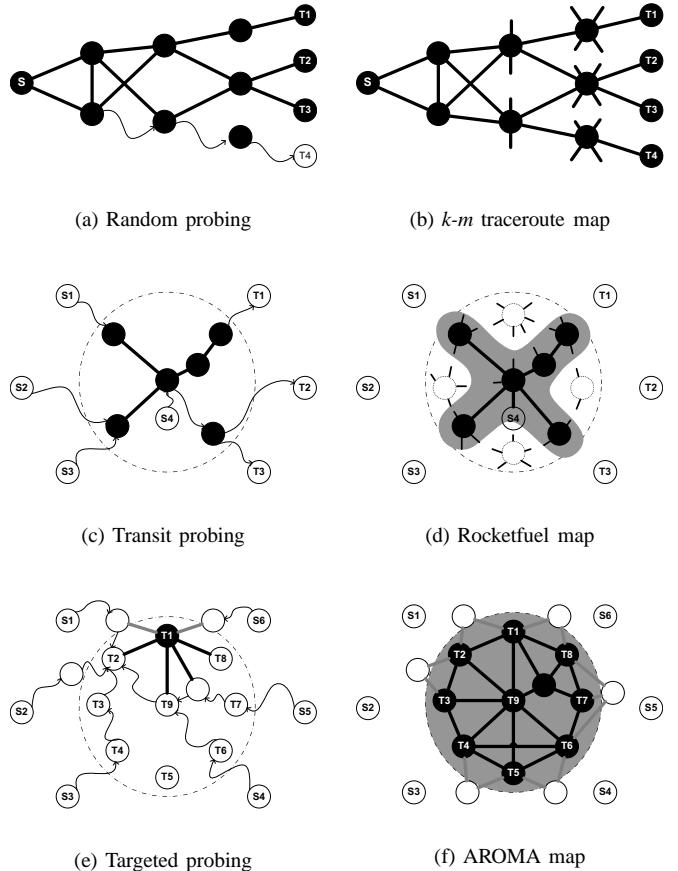


Fig. 1. Probing philosophies to construct IRLMs: (a) probing random destinations and (b) the resulting k - m traceroute- map; (c) probing destinations that lead to probes transiting a particular AS, and (d) the resulting Rocketfuel map; and, (e) targeted probing to a particular AS, and (f) the resulting AROMA map.

TABLE I
NUMBER OF IP ADDRESSES TARGETED BY AROMA.

AS	AS #	Addr. Space Size	Addr. with Names	Targeted Addr.
SprintLink	1239	11,615,500	397,920	397,890
Level3	3356	43,506,700	11,623,900	306,500
Verio	2914	6,895,000	788,800	788,800
AboveNet	6461	884,000	466,350	184,630

before the final map is generated – step 5. We next describe the details of each step.

A. Selecting Targeted IP Addresses

Not every IP address in the address space of the targeted AS needs to be probed. Specifically, IP addresses that belong to endpoints and unused IP addresses do not need to be probed since our objective is to draw a router level map. In order to distinguish routers’ IP addresses from the rest, we rely on querying the Domain Name Service (DNS) [4]. Typically, ISP administrators assign meaningful names to their routers for management purposes. Naming convention is different from ISP to ISP, but a name typically includes the location and role

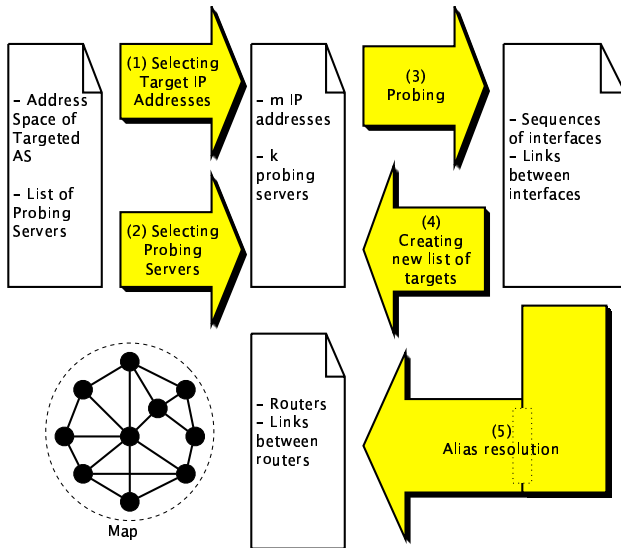


Fig. 2. AROMA Diagram

of the router such as “gw” for gateway routers and “bb” or “bbr” for backbone routers. Reverse DNS is used to convert IP addresses to their names. By selecting only the IP addresses that are assigned names referring to the target AS, then (1) the unused IP addresses and (2) IP addresses belonging to customer ASes of the target AS are left out, which serves our purpose. However, (1) end-hosts in the target AS that have registered DNS names are included and (2) routers in the target AS that are not assigned names are singled out. The former case is resolved by disregarding IP addresses with names including the substrings “dialup”, “DSL”, etc. While the latter case is quite uncommon, unnamed IP addresses of routers are revealed through AROMA’s recursive process, which we describe in the following sections. Table I shows the size of the address space corresponding to the four ISPs that we are mapping in this paper: SprintLink, Level3, Verio, and AboveNet. The table also shows the number of addresses that are assigned names, and the number of selected addresses for probing (neglecting IP addresses with names belonging to customers ASes and those including the “dialup” or “DSL” substrings).

B. Selecting Probing Servers

We have access to almost 280 probing servers that are geographically dispersed all over the globe on the Planetlab facility [5]. However, for each targeted AS, a small fraction of these servers is enough to reveal the routers/links that would be revealed if all servers were used for probing. The reason being that if traceroute probes from two servers to the same destination(s) enter the targeted AS through the same ingress router then these traceroutes will follow the same paths inside the AS and will reveal the exact same information. In this case, only one of these servers should be used for probing. AROMA applies this idea by randomly picking a small set of IP addresses in the targeted AS and probes this set from all 280 probing servers. Then, the largest set of servers for which

TABLE II
PERCENTAGES OF PROBES THAT SUCCESSFULLY REACHED THEIR INTENDED DESTINATIONS AND OF PROBES THAT WERE BLOCKED.

	Spintlink	Level3	Verio	Abovenet
Successful	32.3%	59.8%	21.6%	35.8%
Blocked	67.7%	40.2%	78.4%	64.2%

TABLE III
NUMBER OF TARGET IPs IN EACH PHASE.

AS	Phase I	Phase II	Phase III
Sprintlink	397,890	477	5
Level3	306,500	868	8
Verio	788,800	1,171	16
Abovenet	184,630	224	3

probe packets enter the targeted AS through different ingress routers are selected for probing. This selection process brings the number of probing servers for the Sprintlink network to 93, and for Level3, Verio, and Abovenet to 92, 105, and 51, respectively.

C. Probing

In the probing step, each IP address in the target list becomes the destination of traceroute packets (is probed) from multiple probing servers. In order to avoid overwhelming any single router, probing traffic was rate-limited and the order in which IP addresses were probed was randomized. Recall that the traceroute utility relies on a sequence of TTL-limited ICMP packets and a random destination port number. These packets will reveal interfaces of routers along the path to the destination through ICMP_TIME_EXCEEDED messages, and will reveal an interface of the destination router through an ICMP_PORT_UNREACHABLE message. The result of probing is thus sequences of interfaces and links connecting these interfaces. Throughout our mappings, we did not receive complaints from AS administrators. However, many of the probes were *blocked* at some point along the path towards the destination. This is mostly due to firewalls or other security measures. Table II shows the percentage of probes that successfully made it to their intended destination and the percentage of probes that were blocked.

D. Creating New List of targets

Any revealed interface from the probing step, that belongs to the targeted AS and that was not included in the target list is included, and steps 3 and 4 are repeated. Identifying whether an interface belongs to the targeted AS is done by checking the interface’s IP address against the pool of IP address of the targeted AS, obtained from the registry service. If no new interfaces are revealed, the mapping proceeds to the alias resolution step – step 5. Our results indicate that the number of newly revealed interfaces drops dramatically with each phase, and no new interfaces are revealed after the third phase. Table III shows the number of IP addresses that are probed in each phase for our four case studies.

E. Alias Resolution

Alias resolution refers to the process of clustering interfaces (IP addresses) belonging to the same router together. Several approaches have been proposed for alias resolution. They mostly rely on traceroute queries and can be distinguished based on how they process the replies to the traceroute packets. Typically, if a probed router does respond to traceroute queries, then the source IP address field in the header of the reply packets will either correspond to the probed router's (1) default interface, or (2) its outgoing interface towards the probing host. In order to identify whether two IP addresses are aliases for the same router, Mercator [16] compares the source address fields in the reply packets to traceroute queries to the two IP addresses, and if they match, then Mercator concludes that the two addresses are aliases; otherwise, they are not. This approach should work only if routers respond with their default interfaces. A different approach is used by Ally [24], which sends two back-to-back probe packets to the two investigated IP addresses, and inspects the sequence numbers in the reply packets. If the sequence numbers are in order and close enough, then Ally declares that two interfaces are aliases. Mercator and Ally have a couple of weaknesses. First, they only work if routers are responsive to probe packets. Second, their input is a couple of IP addresses. In a network of n interfaces, there are potentially $n!$ aliases to investigate, and exploring all the possibilities is prohibitive. A third popular approach for alias resolution relies on the DNS service. DNS names provide a wealth of useful information for alias resolution. By storing DNS names of routers in a database, aliases of an interface with some domain name are obtained by searching the database for names with common substrings, without the need to communicate with targeted interfaces. As a result, alias resolution can be done even if routers are configured not to respond to probing messages or if they are temporarily unreachable. Our results indicate that almost 30% of the routers are not responsive to probe packets, which highlights the importance of DNS names in resolving aliases. Also, not all $n!$ combinations of interfaces need to be tested, which improves alias resolution efficiency. Since Pansiot [21] introduced the reversed DNS method for alias resolution, virtually most of the Internet topology discovery techniques use the reversed DNS in a certain degree. Research studies have found that about 0.5% of IPs are misnamed [28].

AROMA mainly relies on reverse DNS for alias resolution [21] and complements it using Ally mechanisms by further probing the questioned IP addresses from multiple probing hosts. Interfaces that are suspected to be aliases from their DNS names are verified using Ally and by probing them from multiple PlanetLab servers and the source IP address fields in the reply packets are used to verify aliases. This improves the accuracy as aliases which might not be revealed from one vantage point may be revealed from another. To understand why this is the case, consider a router, which responds to probe packets with the outgoing interfaces towards the probing hosts. Also, consider two interfaces I_1 and I_2

belonging to this router. If I_1 and I_2 are probed from a single probing host, then the replies may be through two different interfaces, which does not indicate that I_1 and I_2 are aliases. However, as we probe from more probing hosts, it becomes more likely that a reply for a probe to I_1 from one of the probing hosts will contain the same outgoing interface as the reply for a probe to I_2 from a different probing host, which signals that I_1 and I_2 are aliases.

The result of the alias resolution step is the set of routers in the target AS map together with links connecting these routers. Table IV shows the number of routers and links revealed by AROMA maps as compared to those revealed by Rocketfuel for our four case studies. The Rocketfuel data can be found in [7]. Clearly, the number of routers/links revealed by AROMA is significantly larger than those revealed by Rocketfuel. However, this comes at the expense of more probing overhead. Furthermore, one can argue that the considered ASes have grown in size since the Rocketfuel experiments were conducted, almost three years ago. We resolve the probing inefficiency problems in Section III, and pinpoint the sources of bias in Rocketfuel in Section IV.

Admittedly, the AROMA maps may be less than perfect. AROMA relies on ICMP packets and its accuracy will be offset by routers not responding to traceroute packets, by routers/links that are not revealed in the probing process. While these problems may, more or less, impact the AROMA maps, they persist with much larger magnitude in current state-of-the-art techniques to uncover Internet maps.

III. EFFICIENT AROMA

Large ASes are assigned a large pool of IP addresses and the number of probing servers is potentially large as well. Probing every IP addresses in the targeted AS from every server is intolerable not only because it will put extensive stress on the AS infrastructure but also because it will take months if not years to get a router-level map. The same holds even after the wise selection of servers and destination IP addresses described in Sections II-A and II-B. In this section we make the case that AROMA (1) can achieve significant efficiency improvement at the cost of a tiny reduction in the details of the resulting map, and (2) can lead to much more detailed maps while being more efficient than existing tools.

We use the number of routers/links revealed in the targeted AS as a measure of **completeness** and the number of traceroute probes as a measure of the **efficiency** (also a measure of **overhead**).³ We intentionally ignore the overhead introduced by querying the DNS infrastructure. That is not because the overhead is small but because (1) the DNS infrastructure is basically doing what it is intended for, accommodating DNS queries. The DNS system deals with billions of requests per day and accommodating a couple of more million requests –

³Note that measuring *accuracy* is difficult due to the lack of a known topology. Furthermore, the sources of inaccuracy in AROMA, outlined at the end of Section II, are similar to those in Rocketfuel. The comparison in this sense, based on completeness, seems fair.

TABLE IV
NUMBER OF ROUTERS AND LINKS REVEALED BY AROMA AND ROCKETFUEL

	Router		Link	
	AROMA	Rocketfuel	AROMA	Rocketfuel
Sprintlink	35,757	10,332	51,314	25,841
Level3	5,831	1,786	27,144	13,838
Verio	74,114	6,523	176,213	19,289
Abovenet	21,619	654	52,590	2,675

refer to Table I – is not stressing especially if they are rate-limited and spread over a reasonable period of time. Most importantly, (2) DNS information is used for alias resolution, a major component of any router-level mapping tool. Without using DNS information, the alias resolution part would not scale as explained in Section II-E, and the load on the DNS system would have to shift to routers in targeted ASes in order to accommodate an accurate alias resolution technique like Ally [24]. For this reason, Rocketfuel also relies on the DNS system for alias resolution.

In Figure 3, we compare AROMA and Rocketfuel’s completeness and efficiency. From one server, we probe every IP address selected as in Section II-A. We then plot the ratio of the number of routers/links revealed to the overall number of routers/links (completeness) as we increase the number of probed IP addresses (efficiency). We also plot the ratio of the number of routers/links revealed by Rocketfuel to map the same ASes and the number of probes used to get these maps. The rocketfuel data is publicly available in [7]. The figure clearly shows that even with one probing server and with much less targeted IP addresses (efficiency) than those selected in Section II-A, the number of routers and links revealed by AROMA (completeness) is much larger than those revealed by Rocketfuel. Based on Figure 3, and as shown in Table V, AROMA requires at most 5.1% of the overhead incurred by Rocketfuel to attain similar completeness. Also, based on Figure 3, it is clear that it is not necessary to probe all selected IP addresses. A more efficiency-aware implementation would measure the *utility* of probing more IP addresses and would halt probing at any probing server once this utility drops below some threshold. This utility can be expressed as the ratio of the number of revealed routers/links to the number of probed IP addresses during some probing time interval.

In Figure 4 we plot the number of revealed routers/links as we increase the number of probing servers. The figure reveals an interesting trend: The rate at which more routers are revealed decreases sharply as we increase the number of probing servers; however, the rate at which more links are revealed does not. In other words, more servers are useful in revealing more links between routers since they access the targeted AS from different ingress points but are not that useful in revealing more routers since the AROMA’s initial list includes most routers.

To sum up, AROMA can be tuned to be very efficient reducing the number of IP addresses probed from any vantage point with a small degradation in completeness. However, reducing the number of probing servers will typically come

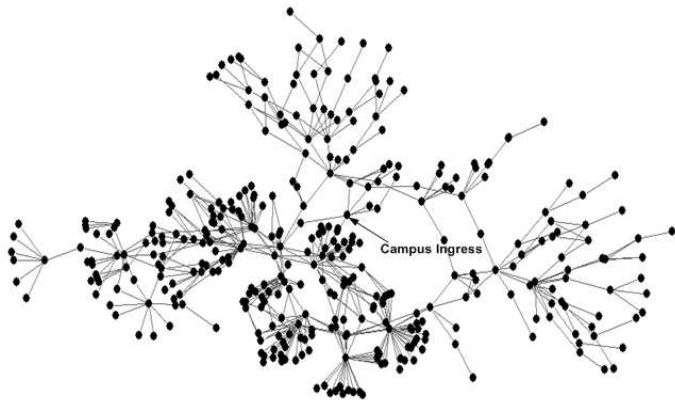


Fig. 5. AROMA map for NCSU campus network.

at the cost of hiding the connectivity information between the routers.

IV. RESULTS AND ANALYSIS

In this Section we introduce a simple validation of AROMA by mapping our campus map, and analyze in more detail the maps generated for the SprintLink, Level3, Verio, and AboveNet networks. We also use our results to revisit common perceptions about the structure of the Internet router level topology.

A. Validation

We used AROMA to map our campus network at NCSU (AS 11442) to validate its completeness and accuracy, and the map is plotted in Figure 5. While the network has several different egress points, all traffic from the Planetlab probing servers reach the same ingress router (following the Abilene network). Thus only one probing server was enough to probe our campus network. By targeting 40,785 destination addresses in the campus AS, we found 360 routers and 482 links between them. The map was verified with a campus network administrator, who confirmed that the map is very accurate, missing only a few peering connections.

B. Topology Structure

In order to get a better understanding of the structure of the revealed maps, we characterize routers based on how close they are from the backbone routers in their AS. Specifically, let L_0 be the set of backbone routers (identified from their DNS names), L_1 be the set of routers that are linked to routers in L_0 , L_i be the set of routers that are linked to routers in L_{i-1} ,

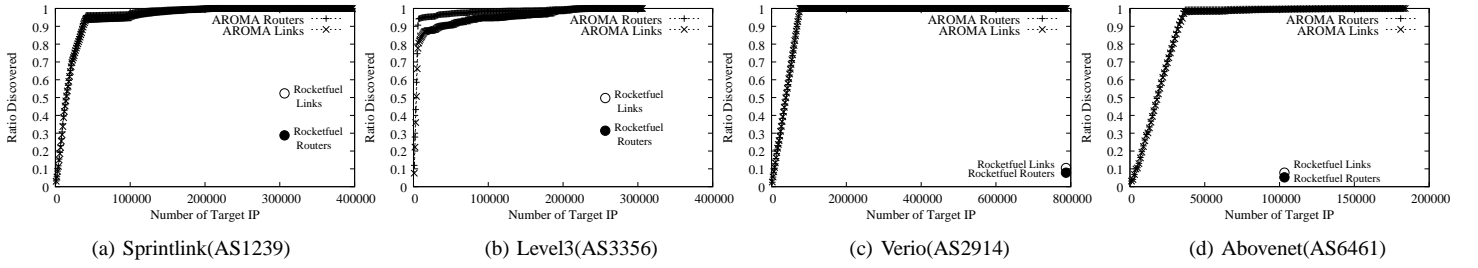


Fig. 3. Probing efficiency versus mapping completeness of AROMA using one probing server compared to Rocketfuel.

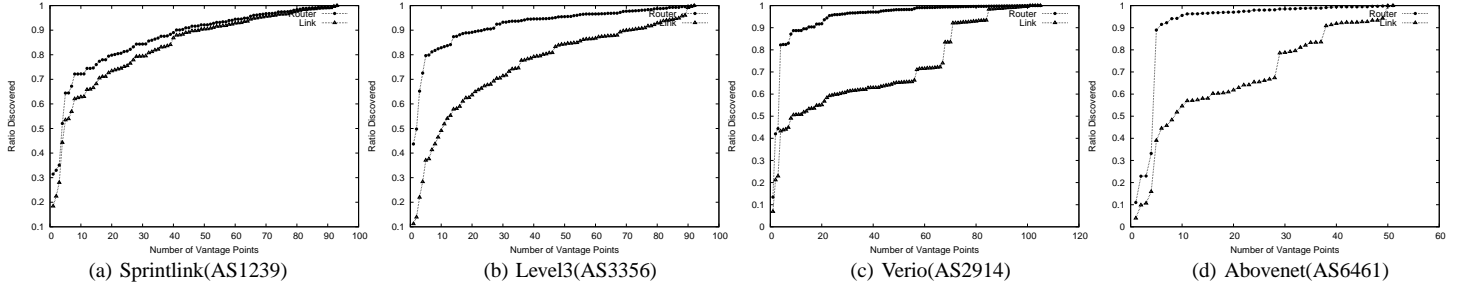


Fig. 4. Mapping completeness as we vary the number of probing servers.

TABLE V
NUMBER OF PROBES REQUIRED TO ATTAIN ROCKETFUEL COMPLETENESS

	Router			Link		
	AROMA	Rocketfuel	%	AROMA	Rocketfuel	%
Sprintlink	13,300	307,605	4.3%	15,580	307,605	5.1%
Level3	2,235	268,237	0.1%	4,790	268,237	1.8%
Verio	7,090	814,061	0.1%	10,273	814,061	1.3%
Abovenet	680	103,122	0.1%	2,180	103,122	2.1%

TABLE VI
NUMBERS (AND PERCENTAGES) OF ROUTERS IDENTIFIED BY AROMA COMPARED TO ROCKETFUEL

ISP	Method	L_0	L_1	L_2	L_3	L_4	L_5+	Total
Sprintlink	AROMA	1,225 (3.4%)	1,752 (4.9%)	7,132 (19.9%)	10,310 (28.8%)	10,091 (28.2%)	5,247 (14.7%)	35,757 (100.0%)
	Rocketfuel	700 (6.8%)	6,637 (64.2%)	2,566 (24.8%)	275 (2.7%)	35 (0.3%)	119 (1.2%)	10,332 (100.0%)
Level3	AROMA	459 (7.9%)	1,551 (26.6%)	1,832 (31.4%)	1,096 (18.8%)	498 (8.5%)	854 (14.6%)	5,831 (100.0%)
	Rocketfuel	625 (35.0%)	995 (55.7%)	152 (8.5%)	13 (0.7%)	1 (0.1%)	0 (0.0%)	1,786 (100.0%)
Verio	AROMA	3,217 (4.3%)	5,047 (6.8%)	32,231 (43.5%)	15,750 (21.3%)	4,145 (5.6%)	13,724 (18.5%)	74,114 (100.0%)
	Rocketfuel	1,013 (15.5%)	3,657 (56.0%)	1,269 (19.5%)	268 (4.1%)	112 (1.7%)	204 (3.1%)	6,523 (100.0%)
Abovenet	AROMA	1,472 (6.8%)	6,501 (30.1%)	6,006 (27.8%)	2,696 (12.5%)	3,318 (15.3%)	1,626 (7.5%)	21,619 (100.0%)
	Rocketfuel	358 (54.7%)	281 (43.0%)	36 (5.5%)	14 (2.3%)	10 (1.5%)	0 (0.0%)	654 (100.0%)

etc. In Table VI we compare the numbers of routers identified in the sets L_0 up-to L_5+ using both AROMA and Rocketfuel.⁴ The numbers reveal the deep hierarchy in these large ASes but also reveal that most of Rocketfuel’s revealed routers are at the core of the network, in the sets L_0 through L_3 . On the other

hand, AROMA finds a considerable number of routers in L_4 and L_5+ . This confirms our intuition that Rocketfuel is biased towards the backbone of the targeted AS due to the routing of transit traffic through the backbone links. AROMA does not suffer from this structural bias as it probes directly the targeted IPs regardless of where their associated routers are located in the network hierarchy.

⁴In this table, the column labeled L_5+ refers to the number of routers in the set $L_5 \cup L_6 \cup L_7$. Since the number of routers in L_6 and L_7 is relatively insignificant, we sum the numbers together.

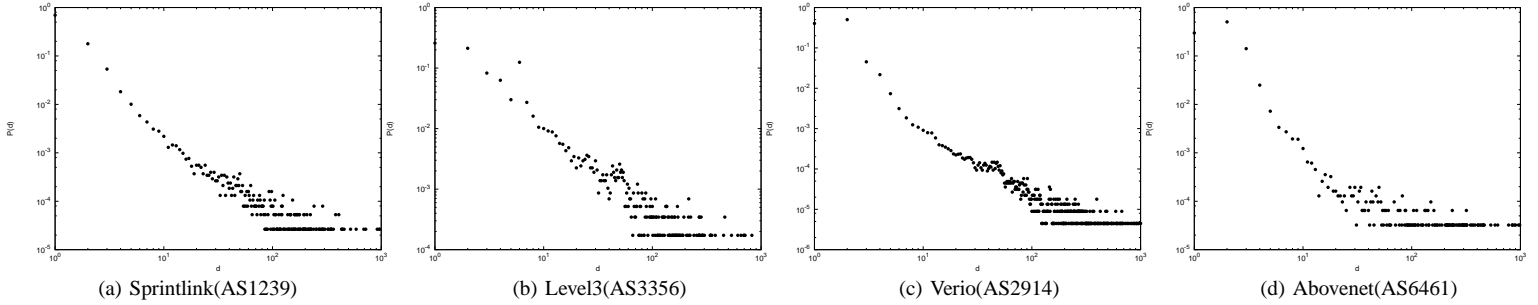


Fig. 6. Router degree distribution of ISP observed by AROMA.

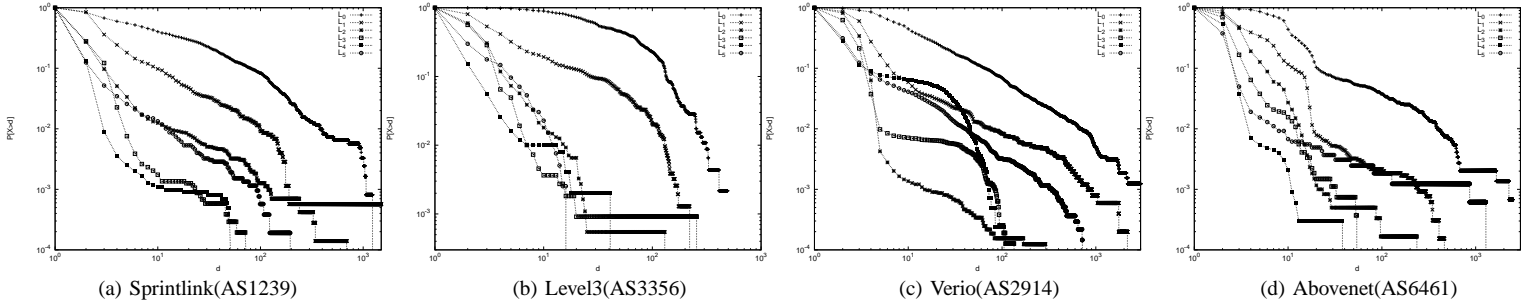


Fig. 7. Degree distribution by distance from backbone network: L_i represents the distance from backbone, L_0 are the routers at the backbone while L_3 are routers 3 hops away from the backbone.

C. Degree Distribution

The degree distribution of the Internet topology has been a source of controversy [9]–[11], [18], [19], [25] since Faloutsos et al. suggested that it follows a power-law distribution [13]. Some researchers advocate the power-law distribution based on their measurements [3], [16], and others question the power-law hypothesis and suggest biased measurements [9], [11], [18], [25].

In Figure 6, we use the AROMA maps of the four investigated ASes to plot the probability density function, $P(d)$, showing the fraction of routers with degree d . The figure clearly shows that the distribution follows a power-law, $P(d) \sim d^{-\gamma}$. The power-law exponent, γ , is 2.7, 1.7, 1.94, and 2.3 for Sprintlink, Level3, Verio, and Abovenet, respectively. The estimation of γ is done through least squares regression using the first 5 points of the log-log fitting [17]. This can be justified by that the first 5 points on a log-log scale contain most of the data of a power-law graph, and fitting the graph with all the points, instead of the first 5 points, will distort the results [15]. Previous studies have found power-law exponents of 2.57 with 3,800 routers [13], and of 2.66 with 150,000 routers [16].

The routers' degree distribution revealed by Rocketfuel is quite different from the power-law distribution revealed by AROMA. It has been shown in [25] that the routers' degrees follow a Weibull distribution, $P[X \geq d] \sim e^{-d^c}$, where c is the shape parameter. As opposed to AROMA's power-law distribution, the Weibull distribution is not long-tailed. Rocketfuel's Weibull distribution can be explained by

Rocketfuel's bias towards backbone routers. Notice that the edge part of a power-law graph hosts a large fraction of the low degree nodes. By being mostly blind to edge routers, Rocketfuel deflates the *head* of the power-law distribution, where low degree nodes are represented, which inflates the *tail* of the power-law distribution, where high degree nodes are represented. The result is a Weibull-like distribution even if the underlying topology has a strict power-law distribution. In order to prove our hypothesis, in Figure 7 we plot the degree distribution for routers belonging to the set L_0 as revealed by AROMA, and the distribution for routers belonging to L_1 , etc. Figure 7 reveals a consistent pattern in all investigated ASes: The number of low-degree routers is small in the core, while it is relatively large at the edges, which supports our argument that Rocketfuel's Weibull distribution is due to its bias. This observed pattern also contradicts conclusions drawn in [19], in which the authors argue that core routers have a lower degree than edge routers. We intend to investigate this issue in our future work.

V. CONCLUSIONS AND FUTURE WORK

In this paper, we introduced a probing tool to unveil detailed Internet IP topologies, and made a case for its efficiency compared to existing tools. We used this tool to map four major ISP topologies and revisited earlier conclusions about the Internet IP topology. Our results, contradicting earlier conclusions, raise more questions than answers. While we believe that AROMA provides a positive step towards unveiling Internet IP maps, understanding these maps and the factors affecting their structural and behavioral characteristics is a challenging task.

Understanding the reasons behind the contradicting results and Identifying other Internet features, like the location of layer-3 tunnels and generating layer-2 maps will be part of our future research.

REFERENCES

- [1] American registry for internet numbers (arin). <http://www.arin.net/>.
- [2] Asia pacific network information centre (apnic). <http://www.apnic.net/>.
- [3] Cooperative association for internet data analysis(caida), the skitter project. <http://www.caida.org/Tools/Skitter>.
- [4] Domain name system. <http://www.dns.net/dnsrd/>.
- [5] Planetlab. <http://www.planet-lab.org/>.
- [6] Ripe network coordination centre (ripe). <http://www.ripe.net/>.
- [7] Rocketfuel: An isp topology mapping engine. <http://www.cs.washington.edu/research/networking/rocketfuel/>.
- [8] University of oregon route views project. <http://www.routeviews.org/>.
- [9] C.-C. Chang and K.-F. Hwang. Towards the forgery of a group signature without knowing the group center's secret. *Lecture Notes in Computer Science*, 2229:47–51, 2001.
- [10] H. Chang, S. Jamin, and W. Willinger. Inferring as-level internet topology from router-level path traces. In *In Proceeding of SPIE ITCOM*, Denver, CO, August 2001.
- [11] Q. Chen, H. Chang, R. Govindan, S. Jamin, S. Shenker, and W. Willinger. The origin of power laws in internet topologies revisited.
- [12] B. Donnet, P. Raoult, T. Friedman, and M. Crovella. Efficient algorithms for large-scale topology discovery. In *SIGMETRICS '05: Proceedings of the 2005 ACM SIGMETRICS international conference on Measurement and modeling of computer systems*, pages 327–338, 2005.
- [13] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. In *SIGCOMM*, pages 251–262, 1999.
- [14] F. Georgatos, F. Gruber, D. Karrenberg, M. Santcroos, A. Susanj, H. Uijterwaal, and R. Wilhelm. Providing active measurements as a regular service for isps. In *In Proc. PAM 2001*.
- [15] M. L. Goldstein, S. A. Morris, and G. G. Yen. Problems with fitting to the power-law distribution. pages 255–258, 2004.
- [16] R. Govindan and H. Tangmunarunkit. Heuristics for internet map discovery. In *IEEE INFOCOM 2000*, pages 1371–1380, Tel Aviv, Israel, March 2000.
- [17] J. H. Jones and M. S. Handcock. An assessment of preferential attachment as a mechanism for human sexual network formation. (1520):1123–1128, June 2003.
- [18] A. Lakhina, J. Byers, M. Crovella, and P. Xie. Sampling biases in ip topology measurements. In *IEEE INFOCOM*, San Francisco, CA, March 2003.
- [19] L. Li, D. Alderson, W. Willinger, and J. Doyle. A first-principles approach to understanding the internet's router-level topology. Portland, OR, 2004. ACM SIGCOMM.
- [20] T. McGregor, H.-W. Braun, and J. Brown. The nlanr: Network analysis infrastructure. May 2000.
- [21] J. Pansiot and D. Grad. On routes and multicast trees in the internet. pages 41–50, 1998.
- [22] Y. Rekhter and T. Li. A border gateway protocol 4 (bgp-4), request for comments: 1771, March 1995.
- [23] R. Siamwalla, R. Sharma, and S. Keshav. Discovering internet topology.
- [24] N. Spring, M. Dontcheva, M. Rodrig, and D. Wetherall. How to resolve ip aliases. Technical Report UW-CSE-TR 04-05-04, May 2004.
- [25] N. Spring, R. Mahajan, and D. Wetherall. Measuring isp topologies with rocketfuel, 2002.
- [26] L. Subramanian, S. Agarwal, J. Rexford, and R. H. Katz. Characterizing the internet hierarchy from multiple vantage points. In *Proc. of IEEE INFOCOM 2002, New York, NY*, Jun 2002.
- [27] R. Teixeira, K. Marzullo, S. Savage, and G. Voelker. In search of path diversity in isp networks, 2003.
- [28] M. Zhang, Y. Ruan, V. S. Pai, and J. Rexford. How dns misnaming distorts internet topology mapping. In *Proceedings of the 2006 Usenix Annual Technical Conference*, Boston, MA, June 2006.