Understanding Future Internet Routing: a Transit-Edge Separation Perspective

Kunpeng Liu, Bijan Jabbari, Stefano Secci

***Abstract*—With the significant growth of the Internet traffic, the uncoordinated routing practices are limiting the natural Internet evolution. It is becoming urgent to rethink the principles underlying the Internet infrastructure as well as the design of its major protocols, especially those related to Internet routing and traffic engineering. From our standpoint of view, appropriate characterization of the current Internet properties seems necessary in the future Internet research, as it may provide valuable information for the design of the future Internet protocols. In this paper, we analyze Internet routing maps of the last two years within a Transit-Edge (T-E) routing separation perspective, a promising direction to improve Internet scalability and resiliency by allowing explicit forwarding through routing locators on the way toward the destination network. Though separating the routing locator from the terminal identifier, it is also possible to achieve better user mobility and mitigate important routing security issues. In this paper, we focus on a statistical and analytical characterization of the behaviors of edge and transit ASes in terms of interconnection, routing and traffic engineering practices, highlighting their similarities and differences.**

I. Introduction

The Internet has been evolving from an academic network managed and operated by researchers, to a worldwide and ubiquitous network interconnecting devices of multiple natures. At its inception, many technology choices had to be taken, such as on the forwarding nature of the Internet Protocol, its addressing and the inter-domain routing principle. The history tells us that the Internet Protocol (IP) relies on packet switching with statistical multiplexing, that its addressing is based on a 32-bit space and is now migrating to a 128-bit space, and that the Border Gateway Protocol (BGP) [2] is the single inter-domain routing protocol used by Autonomous Systems (AS) to exchange routing information. BGP relies on a flat routing mode using path vectors for each IP network prefix, announced independently and in a totally uncoordinated fashion.

The lack of coordination amongst AS networks appears strategically reasonable as each AS needs to first follow its own interests and objectives. However, the flat routing mode of Internet routing is unable to scale with such a behavior for a very large number of networks. Meanwhile, the number of ASes as well as the announced network prefixes are increasing extremely fast (currently, about 36000 ASes and 400000 network prefixes). Such a large and increasing number of prefixes, even if dictated by reasonable traffic engineering and multi-homing practices, are posing many issues from a network management standpoint. Coupled with other aspects such as BGP routing convergence, instability and weak resiliency, they are undermining the healthy development of the Internet.

A direction recently evaluated to tackle the Internet routing scalability and resiliency issue is to adopt transit-edge (T-E) routing separation schemes [3]. With such a mechanism, one can significantly reduce the transit routing table sizes since a very large majority of the Internet networks are at the edges and do not transit traffic.

In this paper, we measure the Internet topology from a T-E routing separation standpoint. By analyzing the recent routing BGP tables on a two-year period, we aim at characterizing the properties of edge and transit networks from interconnection, routing and traffic engineering perspectives. The paper is organized as follows. Section II describes the technical background. Section III and IV analyze the T-E separation characteristics from interconnection and routing perspectives, respectively. Section V summarizes the paper with final conclusions.

II. Background

The Internet interconnection graph can be partially inferred via BGP routing tables, which contain the best routes chosen by a single router. Routeviews' public routing tables [4] aggregate the daily view of multiple backbone routers, which represents a very detailed mirror on the Internet ecosystem evolution. After a rapid analysis, we find that at present around 84% of the total ASes act as pure destination networks, only appearing at the last position of the AS paths. They are commonly considered as "stub ASes". In practice, some large stub ASes (content providers and delivery networks) functionally fragment their networks into multiple ASes for management reasons, and they may also appear in the penultimate or in the third from last position in AS paths. Nearly 13% additional ASes appear up to the third from last position of BGP AS paths, among which are certainly also some regional Internet Service Providers (ISPs). The sub-network composed of these 97% ASes can be treated as the edge of the Internet that given its interconnection behavior has different traffic engineering requirements and routing purposes than transit networks. In fact, the remaining 3% ASes do transit the global Internet traffic as their principal purpose, and they can be treated as the transit part of the Internet. Even though the Internet has grown significantly, these transit and edge network ratios have been rather stable. As of our observation, the total number of ASes has grown almost linearly in the last two years (since January 2009), and the increasing rate is roughly 250 ASes per month. Among those increases, 97% are edge ASes.

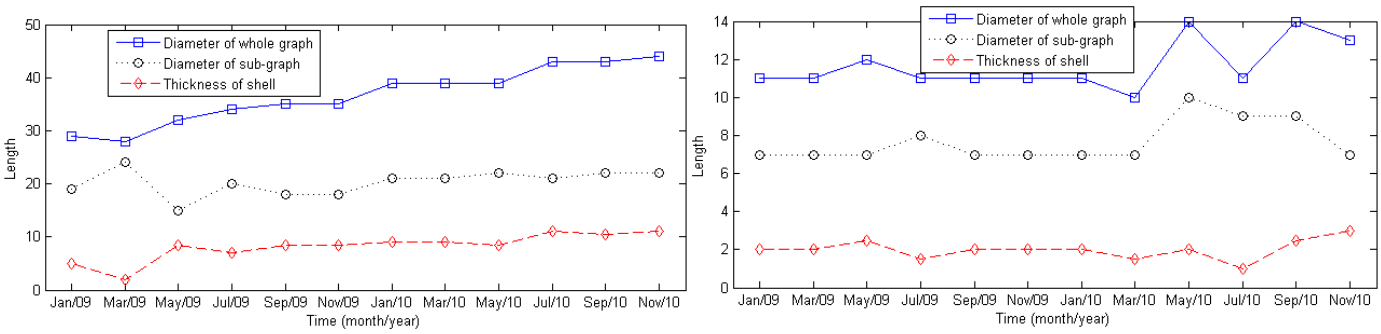
The T-E routing separation paradigm suggests to insert routing locators at the frontier between transit and edge networks.

K. Liu and B. Jabbari are with the ECE Dept., George Mason University, VA, USA. E-mail: {kliu3,bjabbari}@gmu.edu

S. Secci is with the LIP6, University Pierre et Marie Curie, Paris, France. E-mail: stefano.secci@lip6.fr

This work was funded by the Office of Naval Research (ONR) and performed at GMU Communications and Networking Lab (CNL) under US project "Secure Protocols and Services for Resilient Internetworking".

A preliminary version of this paper will be presented at the 2011 Network of the Future Conference (NoF 2011) [1]



(a) With AS path prepending

(b) Without AS path prepending

Fig. 1: The diameters of AS graphs as functions of time

Different protocols can be conceived to manage identifier-to-locator mappings and to encapsulate or aggregating (tunneling) packets in the transit sub-path, such as the Locator-Identifier separation protocol (LISP) [5] which is currently under standardization (which somehow supersedes other host-based approaches such as SHIM6 [6] or HIP [7] that appear as less scalable mechanisms).

Besides allowing a very important reduction of the Internet routing table, as discussed in [8], T-E separation can lead to important improvements in terms of routing resiliency. Indeed, the introduction of many routing locators for the same destination drastically increases the Internet path diversity. If adequately managed by traffic engineering procedures, the enlarged path diversity can lead to significant improvements of the Internet resiliency, as explained in [3] where a framework for coordinated edge-to-edge load-balancing and Internet-wide multipath routing is presented.

Therefore, new tools for Internet traffic engineering - currently limited to BGP tweaking practices such as prefix de-aggregation and transient announcements that are increasing the routing table size and are decreasing the Internet service reliability - could arise from T-E separation. At present, the potential achievable performance improvements for edge networks are attracting attention from content providers and content delivery networks, especially with the emergence of Cloud Computing applications that require high connection resiliency and persistent reachability [3]. In the following, we focus on the characteristics and properties of edge and transit networks presented by a measurement of BGP routing tables.

III. Interconnection Topology Analysis

BGP Routeviews' routing tables are captured from ASes that peer with many large transit carriers, so they represent a transit view on the Internet routes. Meanwhile, the AS interconnection information from the directional perspective of edge ASes is difficult to get. Therefore, it appears appropriate to use the routing tables to build an undirected graph. Through Studying the undirected graph, we first dissect the AS path properties of the Internet, and then we apply several graph theory measurements to characterize the different properties of edge and transit ASes.

A. Path Properties

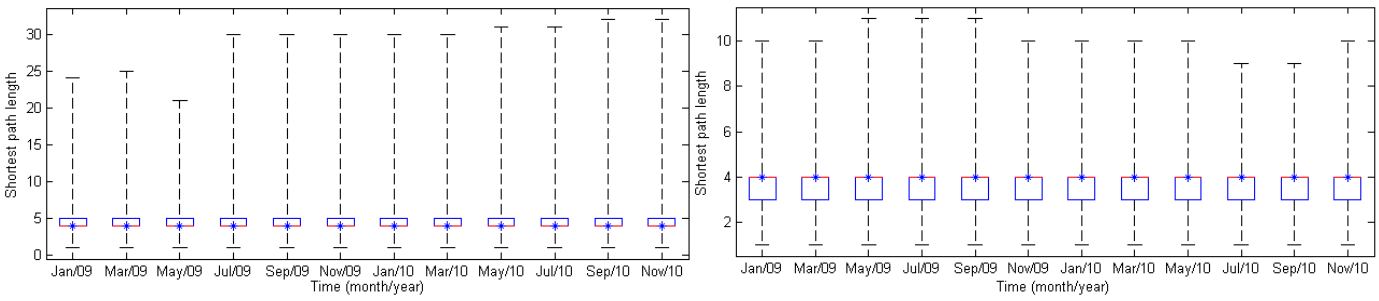
The path represents the path between two ASes in the AS graph, while the path length is counted as the AS path hop number. In practice, any AS can increase the AS path length artificially by repeating its own AS number, which is so-called AS path prepending [9]. To have a comprehensive view about the path properties, we approach our studies under two scenarios: without AS path prepending and with AS path prepending. We characterize the path properties in terms of the following aspects.

1) Diameter diagnosis: Diameter is a summary statistics that reflects the scope as well as the connection situation of the graph. The diameter of a graph is defined as the maximum shortest path length between any pair of nodes in the graph. If there is no path connecting two nodes, the diameter is set to infinity. In our studies, we probe the scope and connection situation of the Internet by diameters, and Fig. 1a and Fig. 1b are the results with and without considering the AS path prepending, respectively. The whole graph represents the whole Internet, the sub-graph is the transit networks and the thickness of shell is the half of the difference between the diameter of the whole graph and that of the sub-graph, which can represent the expected hop number from an edge AS to a transit AS.

We find that all the diameters in the two figures are finite, which shows that the Internet is a whole interconnected network, and there is no two different connected components that do not connect with each other. Under the scenario of not considering AS path prepending, the diameter of the whole graph increases slowly with an oscillation behavior from 11 to 14 and back to 13, while that of sub-graph increases in a similar way from 7 to 10 and back to 7. With AS path prepending, the two diameters as well as the thickness of the shell vary in a bigger scope with randomness behaviors. The statistic results show us that:

- Without AS path prepending, the diameters of whole graph and sub-graph are relatively stable, and share the same trend.
- Without AS path prepending, the thickness of the edge shell did not changed significantly in last two years, and the expected hop number from an edge AS to a transit AS is around 2.
- With AS path prepending, the two diameters as well as the thickness of the shell reflect a certain degree of randomness.
- AS path prepending does alter the AS graphs significantly.

2) Shortest paths diagnosis: We use the shortest paths between two edge ASes to analyze the potential inter-AS level



(a) With AS path prepending

(b) Without AS path prepending

Fig. 2: Edge pairs shortest paths as functions of time

routing efficiency from the perspective of edge networks. We choose 10% of the edge ASes that consistently act as edge ASes since January 2009. Then we measure and monitor the shortest paths between each pairs of the chosen ASes in last two years. We use boxplots to depict the results (each box, between the min. and the max., displays the first quartile, the median with a '*', third quartile). The medians can be treated as the expected values of the shortest paths. Fig. 2a and Fig. 2b are the results with and without considering the AS path prepending, respectively. Though the maxima of the shortest paths change from time to time, the medians, the first and the third quartiles remain constant within each figures. When comparing the two figures, we find that the medians within the two figures are the same, while the first quartiles and third quartiles only increase 1 with AS path prepending. From the observations we can infer that:

- The potential performance of inter-AS level routing remain at the same level from the standpoint of edge networks, although the Internet grows perpetually.
- For some edge network, the usage of AS path prepending enlarges the distance between them and some other edge networks, and their global routing performance can be partially impaired.
- For most edge networks, AS path prepending actually does not degrade the potential efficiency of their global routing as long as a proper routing scheme can be designed and deployed.

B. Edge and Transit ASes Interconnection Properties Comparison

From the standpoint of interconnection topology, edge and transit ASes hold dramatically different properties in the undirected graph. Next, we characterize the properties of the two types of ASes in the following aspects.

1) Degree analysis: The AS degree is defined as the total number of AS neighbors; it somehow reflects the importance of an AS in the Internet interconnection. In Fig. 3 we plot the complementary cumulative distribution function (CCDF) of the AS degree for edge and transit ASes.

Let x_e and x_t denote the degree of edge and transit ASes, respectively. The CCDFs in Fig. 3 are obtained by analyzing the routing tables of Jan. 2009, but the same profile is approximately maintained for successive routing tables. Note that Fig. 3(a) and Fig. 3(c) use a log-log scale, while Fig. 3(b) uses a log-linear scale. We can see that the x_e CCDF linearly decreases in a log-log scale, and so does the x_t CCDF when the degree is bigger than a relative large threshold, e.g., 40. When x_t is smaller than the threshold, the CCDF decreases almost linearly in a log-linear scale. It is worth recalling that the CCDF of a nonnegative random variable that follows truncated discrete power law distribution[1] can be calculated as $F_c(x) \sim ax^{-\alpha}$, while the CCDF of a random variable that has truncated probability density function (pdf) as $f(x) = b/x$ can be calculated as $F_c(x) \sim -b\ln(x)$. In the following, we define the distribution with pdf $f(x) = b/x$ as inverse distribution; note that the CCDF of power law distribution becomes to linear function in a log-log scale, while that of inverse distribution shows linear characteristic in a log-linear scale. When combining the above results, we find that:

- The degree of edge ASes can be well fit with a power law distribution.
- When the degree of a transit AS is relatively small, it approximately follows a truncated inverse distribution.
- When the degree of a transit AS is larger than a certain threshold, it approximately follows a power law distribution.

To simplify the following analysis, we treat x_e and x_t as continuous random variables. Let the CCDFs for the degree of edge and transit ASes be F_{ce} and F_{ct}, respectively. We investigate the following relations:

$$F_{ce}(x_e) \sim a_e {x_e}^{-\alpha_e} \tag{1}$$

$$F_{ct}(x_t | 2 \le x_t \le d) \sim -b\ln(x_t) \tag{2}$$

$$F_{ct}(x_t | x_t > d) \sim a_t {x_t}^{-\alpha_t} \tag{3}$$

Please note that in (1) and (3) the CCDFs have right hand side cutoffs C_e and C_t, respectively.

From (2), we find $f_t(x_t|2 \le x_t \le d) \sim b/x$. As $\int_2^d f_t(x_t|2 \le x_t \le d)dx = 1$, we get:

$$b \sim \ln^{-1}(\frac{d}{2}) \tag{4}$$

Hence, as long as d is a constant, b as well as the statistics of x_t given $2 \le x_t \le d$ will also be deterministic. Through a similar derivation, the relationship between a and α can also be found.

In order to inspect the parameter trends, we choose $d = 40$, and apply the least square error (LSE) as the model estimator

[1]Power law distribution have been observed in many fields for some time, especially in a wide variety of natural and man-made phenomena, and some physicists even have the idea that these correspond to certain "universal laws" [10].

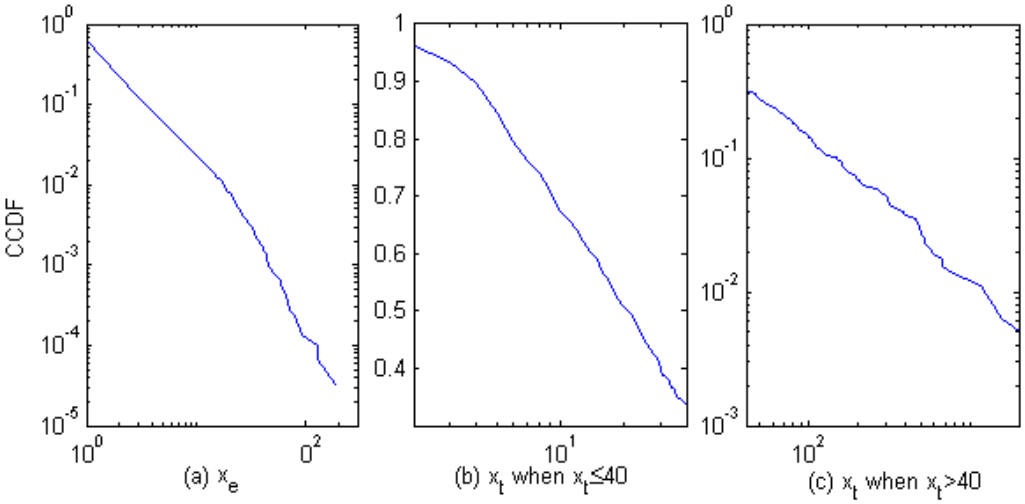


Fig. 3: The degree CCDF of edge and transit ASes

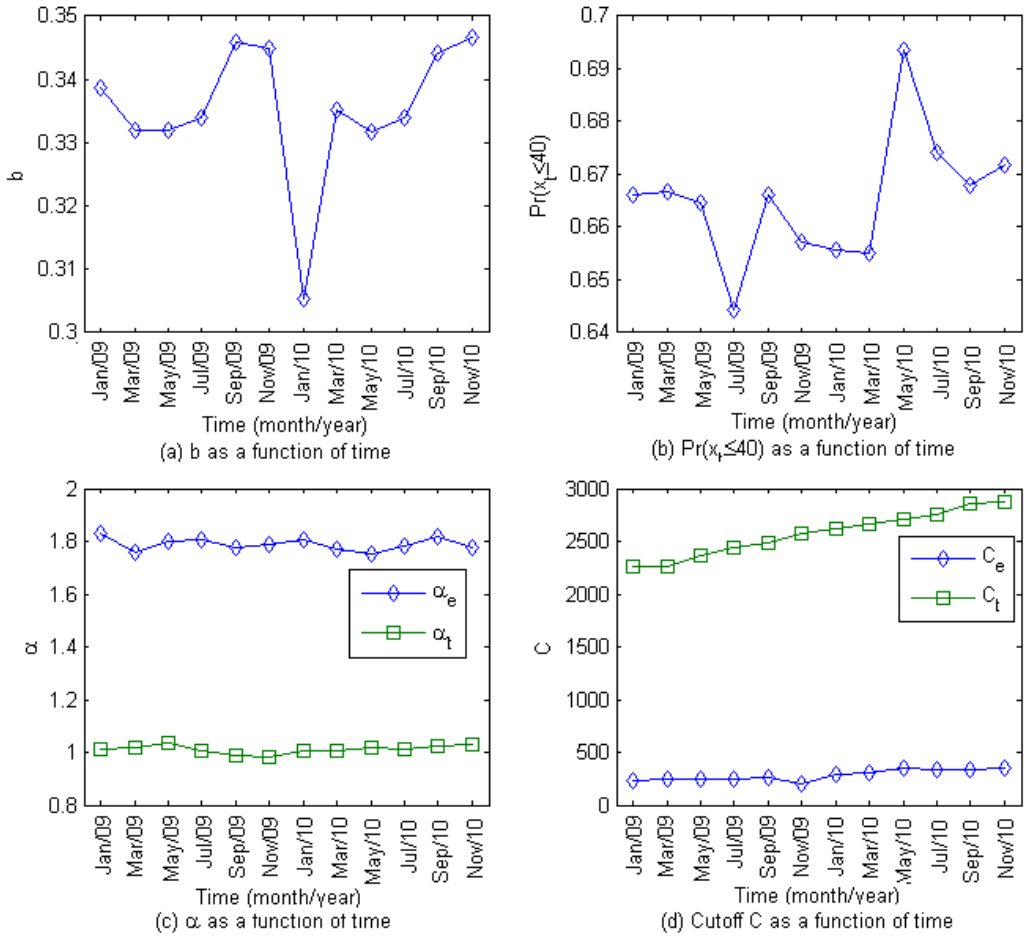


Fig. 4: Model parameters as function of time

to the last two years routing tables. We first examine the trend of b to validate our previous analysis. From (4), we know that b should be around 0.33 given $d = 40$. The theoretical analysis perfectly fits our measurements reported in Fig. 4(a).

Next, we are interested in the trends of $Pr(x_t \leq 40)$, α_e, α_t, as well as the cutoffs C_e and C_t. In Fig. 4(b), we find that $Pr(x_t \leq 40)$ is very stable, which represents the probability for the degree of an transit ASes to follow power law distribution or inverse distribution is very stable. Fig. 4(c) shows that α_e is larger than 1.5 and smaller than 2, while α_t is very close to 1. Fig. 4(d) shows than the cutoff of x_t is much larger than that of x_e, and C_t as well as C_e has a clear increase trend in the last two years. Before further analyzing the results, let us discuss the properties of truncated power law

distribution with pdf $f(x) \sim r x^{-\alpha-1}$ and two cutoffs c_1 and c_2 (c_1 is the left hand side cutoff, and c_2 is the right hand side cutoff). We only consider the case that $c_2 >> c_1$ and c_1 is 1 or 2. It is easy to show that:

$$E(x) \sim r \frac{c_2^{1-\alpha} - c_1^{1-\alpha}}{1-\alpha} \tag{5}$$

$$E(x^2) \sim r \frac{c_2^{2-\alpha} - c_1^{2-\alpha}}{2-\alpha} \tag{6}$$

When α is extremely close to 1, based on (5), we can get the equation

$$\lim_{\alpha \to 1} E(x) \sim r \ln(c_2) \tag{7}$$

Combining the observations and properties, we can assert that:

- The expectation of x_t is increasing in last two years, as α_t is very closer to 1 and the cutoff C_t is always raising. This shows the interconnection of transit ASes evolves permanently, by which a lot of new shortest paths can be created to improve the performance of the Internet.
- Following the raise of cutoff C_e, the expectation of x_e is also increasing in the last two years. This reflects the fact that more and more edge networks apply multi-homing to improve the interconnection situation of their networks.
- Based on (5)∼(7) and some simple calculations, we can find that the standard deviations of x_e and x_t are also increasing in last years. It indicates that the distributions for the degree of edge and transit ASes are stretching constantly.

2) Betweenness diagnosis: The centrality of a node within a graph can be measured by its betweenness [11], which is calculated by counting the number of all the possible shortest paths passing the corresponding node. In practice, researchers usually normalize the betweenness with the total number of the shortest paths to get so-called normalized betweenness. The normalized betweenness of an absolute center, through which almost every shortest path would go, should equal or very close to 1. In order to minimize the impact of the Internet growth to our analysis, we apply normalized betweenness to gauge the centrality of each AS. We still utilize boxplots to depict the data and use star sign to emphasize the medians of the data for each box.

Fig. 5a and Fig. 5b are the boxplots for the betweenness of edge and transit ASes, respectively. In Fig. 5a, only the third quartile and the maximum value can be seen, as other statistics are too small to be shown. Fig. 5a shows that at least 75% of edge ASes have extremely small betweenness. In Fig. 5b, the first quartiles are around $2 * 10^{-4}$, while the maxima changes around 0.3. These statistics show that

- Compared with edge ASes, transit ASes usually hold much bigger degree of centrality.
- Most transit ASes do not have high centrality, and they do not play the role of central ASes in the Internet presently.
- Certain transit ASes do have very large betweenness, and their betweenness are constantly around 0.3.
- From the standpoint of graph theory, some transit ASes are of very importance and serve as partial centers to the Internet, and the misbehavior of these ASes may affect 30% of the inter-AS routing decisions.

3) T-E Separation Properties: According to the position of each AS in the routing entries, the Internet can be artificially separated into edge and transit networks; obviously, an AS holds a single role (edge or transit) in the context of T-E separation. However, the role of a particular AS may change abruptly, due to interconnection evolution or routing fluctuations; this phenomenon is shown in Fig. 6[2]. The X axis represents the time difference, and the Y axis represents the percentage of a kind of ASes that still hold their original ranking after the time interval (defined as AS role immutability). From Fig. 6, we can see the immutability of edge networks drops almost linearly from 98% to 90% when the time difference increase from 2 months to 22 months, while at the same time the immutability of transit networks drops in a more dramatic way from 81% to 59%. Given these observations, we can state that:

- The roles of ASes are quite immutable in a short relative period, like 1 or 2 months.
- Not only the immutability of edge ASes is higher than that of transit ASes, but the role change rate of edge ASes is also much smaller than that of transit ASes.
- T-E separation should not rely on an automated detection of current roles, but should be set statically by transit ASes with little or no coordination with edge ASes.

Such measured role changes indicate that edge ASes rarely evolve adopting transit behaviors, but rather the inverse is more frequent, i.e., ASes in the transit core are pushed towards the edges as the time passes.

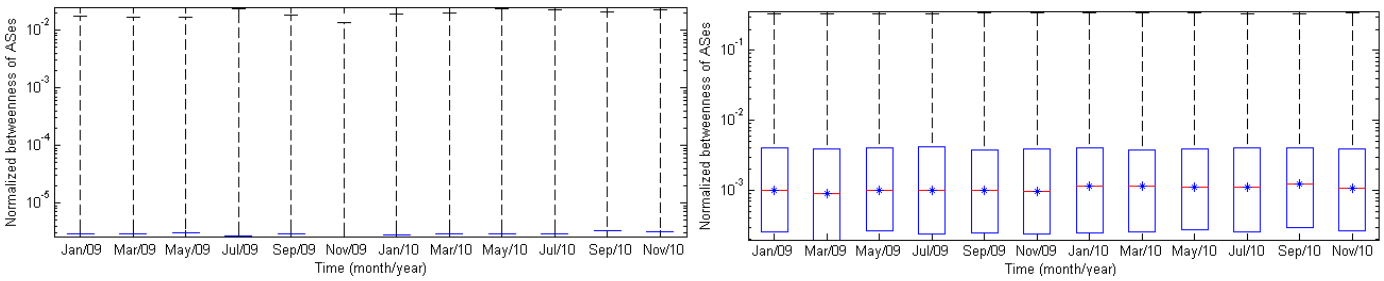
IV. Routing and Traffic Engineering Analysis

In this section, we characterize edge and transit networks from a routing and traffic engineering standpoint. Among all the available traffic engineering techniques in BGP routing, we can mention local preferences for outbound traffic engineering, AS path prepending for inbound traffic engineering, and IP de-aggregation for multi-homing traffic engineering. While the first cannot be inferred with adequate precision from routing table analysis, path prepending and IP de-aggregation can, as reported in the following. Such practices coupled with the BGP convergence issue indirectly affects the BGP routing instability, which is an aspect also analyzed in this section.

A. AS path prepending analysis

With AS path prepending, artificially repeating its own AS number to increase the length of certain AS paths passing through it, an AS can meet inbound traffic engineering goals, i.e., distracting incoming traffic toward more available or preferred entry points. We are interested in the occurrence of path prepending, including the probabilities for an AS applies path prepending as well as for an AS link is affected by path prepending. We categorize the AS links into three types: links inside edge networks, links between edge and transit networks and links inside transit networks. Fig. 7a shows the probabilities that edge and transit ASes use path prepending, while Fig. 7b shows the probabilities that the three types of AS links are affected by path prepending. In Fig. 7a, we find that not only are the probabilities for edge and transit ASes to employ AS path prepending very close to each other, they but also share the same time profile. In Fig. 7b, we find

[2]In our studies, we filter out all the AS path prepending information before positioning each AS, and AS path prepending does not impact this analysis



(a) Edge ASes (b) Transit ASes

Fig. 5: The normalized ASes betweenness as functions of time

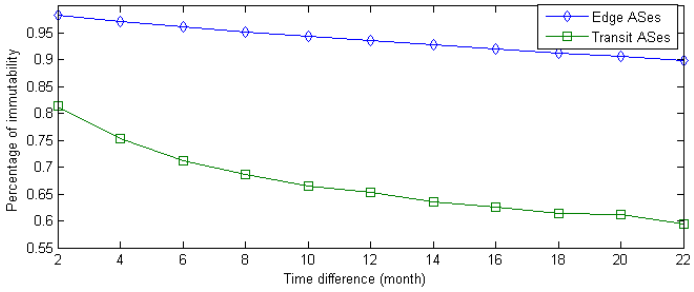


Fig. 6: The roles immutability of ASes as function of time difference

that the AS links inside transit networks are affected by path prepending with the highest probability while the links inside edge networks with the lowest probability. All in all, we can assert that:

- The probabilities for edge and transit ASes to employ AS path prepending are relatively low, as they are both below 0.1.
- The probabilities for edge and transit ASes to apply AS path prepending are very similar with each other.
- The transit networks have the highest degree of requirement for inbound traffic engineering.

Edge ASes apply path prepending essentially for inbound load balancing, while transit ASes perform path prepending as a second-level routing rule for provider transit vs. client transit and transit links vs. peering links load-balancing (the first-level rule for such operations typically is the local-preference).

B. IP de-aggregation probability diagnosis

For security, resiliency as well as load balancing purposes, ASes can artificially fragment large IP prefixes into several smaller prefixes and announce them separately [12], [13]. This behavior is usually known as IP prefix de-aggregation. Although both transit and edge networks may employ this technique to meet certain goals, due to the difference between their functions in inter-AS routing, the probabilities of their usage as well as the specific using behaviors may be different. In our analysis, we gather all the IP prefixes that announced by the same AS, and use seamless and precise IP aggregating rule to check if the AS utilize IP de-aggregation or not. For instance, suppose an AS announce 1.2.3.128/25 and 1.2.3.0/25 separately. As 1.2.3.128/25 and 1.2.3.0/25 can be aggregated into 1.2.3.0/24, we deem that the AS applies IP de-aggregation.

Fig. 8a and Fig. 8b show the probabilities for edge and transit ASes to apply IP de-aggregation, respectively. We find that the de-aggregation probability of edge ASes has a clear increase trend, while that of transit ASes oscillate between 0.734 and 0.758. These properties tell us that:

- IP de-aggregation is very a popular technique among edge and transit ASes in these two years.
- More and more edge ASes are trying to apply IP de-aggregation currently, and that imposes more pressure to the scalability and efficiency of the global routing.
- Compared with edge ASes, transit ASes are more active in utilizing IP de-aggregation to meet the goal of traffic engineering.

C. Prefix de-aggregation impairment analysis

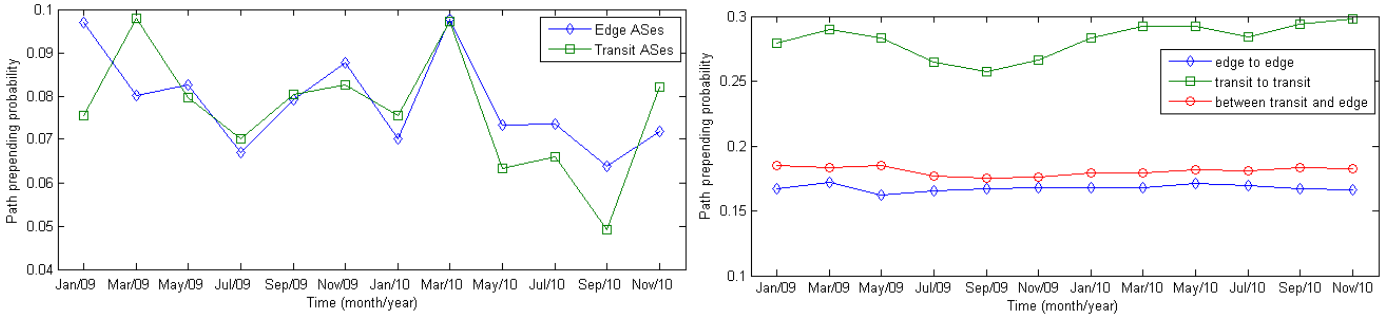
For security, resiliency as well as load balancing purposes, ASes can artificially fragment large IP prefixes into several smaller prefixes and announce them separately [12], [13]. This behavior is usually known as IP prefix de-aggregation. IP prefix de-aggregation inevitably enlarges the size of BGP routing table, thus impairs the efficiency of BGP routing. We analyze the impairment of IP prefix de-aggregation to BGP routing tables in the following way: first, we gather all the IP prefixes announced by a given AS x, noting the total number of prefixes as d_x; next, we recursively apply a seamless and precise IP aggregating rule to obtain the size of the IP prefixes before IP de-aggregation, which is noted as a_x; then the IP de-aggregation rate r_x of the AS x can be expressed as:

$$r_x = \frac{d_x - a_x}{a_x} \tag{8}$$

For instance, suppose an AS announces 1.2.3.128/25, 1.2.3.0/25 and 128.1.1.0/24, separately. As 1.2.3.128/25 and 1.2.3.0/25 can be aggregated with 1.2.3.0/24, the de-aggregation rate of the AS is (3-2)/2=0.5. Therefore, any AS that does not employ IP de-aggregation should have a zero IP de-aggregation rate.

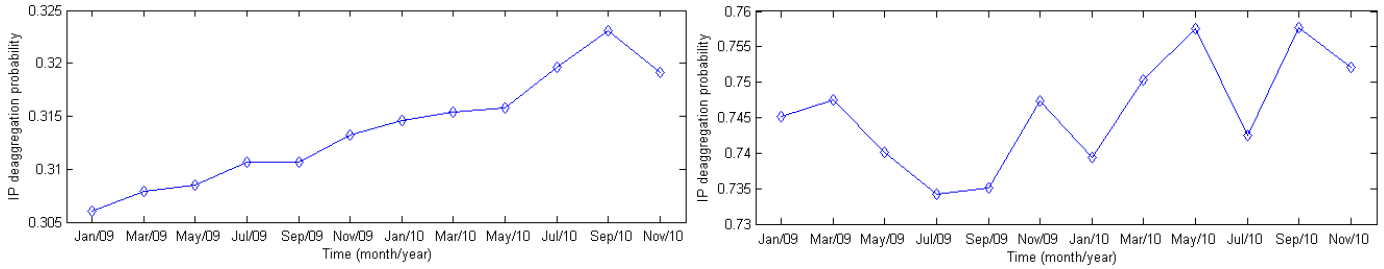
Fixing the total number of ASes to N, an AS that can communicate with every announced IP prefix should have a BGP routing table size close to $\sum_{i=1}^{N}(a_i r_i + a_i) = \sum_{i=1}^{N} a_i r_i + \sum_{i=1}^{N} a_i$. Nevertheless, in an ideal scenario, if there is no IP prefix de-aggregation, its BGP routing table size should only be $\sum_{i=1}^{N} a_i$. Due to IP prefix de-aggregation, the routing table size gets indeed significantly enlarged. Let R be the impact ratio, then:

$$R = \frac{\sum_{i=1}^{N} a_i r_i}{\sum_{i=1}^{N} a_i} \tag{9}$$



(a) AS node (b) AS link

Fig. 7: AS path prepending probabilities as functions of time



(a) Edge ASes (b) Transit ASes

Fig. 8: ASes IP de-aggregation probabilities as functions of time

where, $i \in [1, N]$, a_i are unknown constants and r_i are random variables.

From (9), we know that:

$$E(R) = E(r_i) \tag{10}$$

Therefore, if we could find an alternative routing mode with some form of hierarchical routing more natively supporting IP prefix de-aggregation – such as a T-E routing separation protocol – while allowing at least the same level of traffic engineering capabilities, the BGP routing table size could shrink dramatically. Let the shrink rate be S and the current BGP routing table size be Y. After shrinking, the routing table size becomes to $Y - S \cdot Y$. Comparing with (9), we can get $R = S \cdot Y/(Y - S \cdot Y)$, which yields:

$$S = \frac{R}{R+1} \tag{11}$$

Combining (10) and (11), we get:

$$E(S) = \frac{E(r_i)}{E(r_i)+1} \tag{12}$$

The expectation values of prefix de-aggregation rate for edge and transit ASes are shown in Fig. 9. We find that the expectation for edge ASes has a very clear raise trend, while the expectation for transit ASes has an obvious oscillation pattern. As the overall expectation of IP prefix de-aggregation rate mainly depends on the expectation rate of edge ASes, it has grown from 0.81 to 0.87 in last two years, which further stresses the Internet scalability. From the studies, we can assert that:

- Transit ASes are more used to prefix de-aggregation than edge ASes, which is roughly 3-times more often than edge ASes, and its de-aggregation usage can vary significantly in time and not necessarily increases, while edge ASes usage de-aggregation raises constantly.
- The IP de-aggregation rates of edge and transit ASes directly impair the scalability and efficiency of the Internet, and the expected impact ratio R is decided by the expectation of de-aggregation rate r_i.
- Following the growth of the overall prefix de-aggregation rate, the impairment of prefix de-aggregation also increases in these two years.
- From (12) and the prefix rate expectation, we find out that if an alternative traffic engineering technique for de-aggregation could be provided, the expectation of BGP routing table size could shrink around 45%.

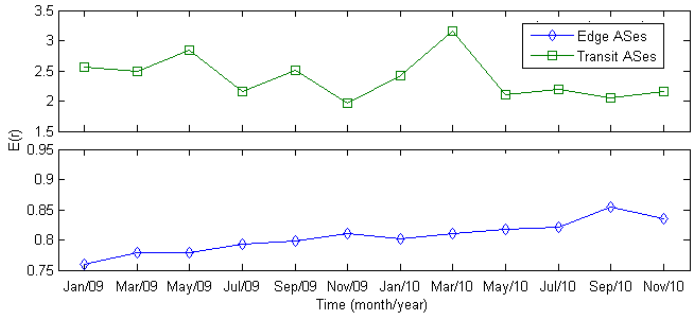
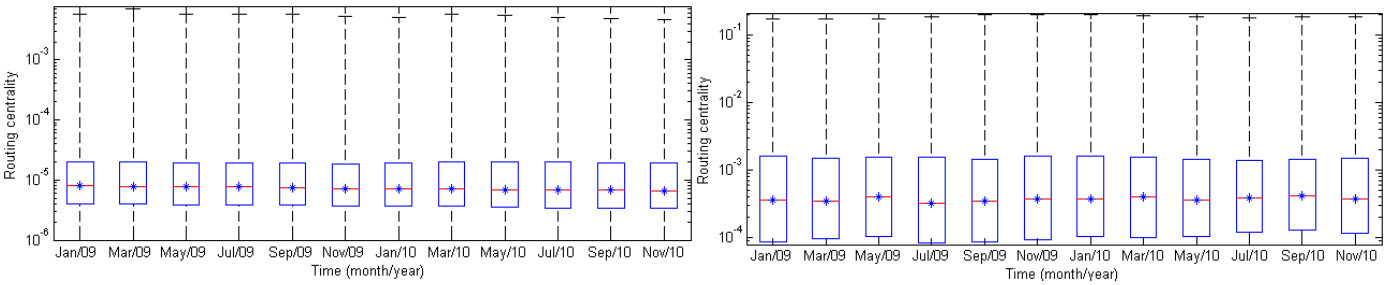


Fig. 9: The expectation of ASes prefix de-aggregation rates as function of time

D. Routing Centrality Comparison

Sparked by the definition of betweenness in graph theory, we use the appearance time of an AS in the routing table



(a) Edge ASes (b) Transit ASes

Fig. 10: The normalized ASes routing centrality as functions of time

to measure the routing centrality of the AS. For each AS, we count the number of times the AS appears in the routing table, and normalize the final count by the table size to get the normalized routing centrality. So the normalized routing centrality of an absolute routing central AS, which almost appears in every routing entry, should equal or very close to 1. We use boxplots to depict the normalized routing centrality statistics of edge and transit ASes in Fig. 10a and Fig. 10b, respectively. In Fig. 10a, the third quartiles change around $2*10^{-5}$, while the medians are $7*10^{-6}$. In Fig. 10b, the first quartiles changes around 10^{-4}, the medians are $4*10^{-4}$, and the maxima are round 0.2. All in all, we can assert that:

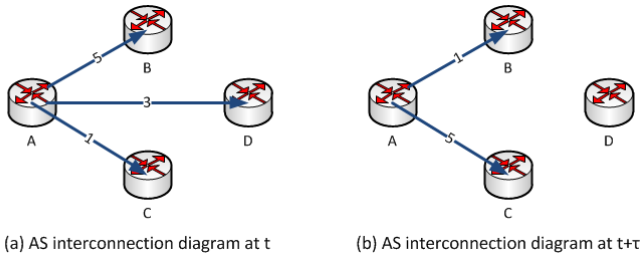
- The expected normalized routing centrality of a transit AS is almost 50 times larger than that of an edge AS.
- The normalized routing centralities of all the edge and transit ASes are far below 1, which reflects that there is no absolute routing central AS nowadays.
- The normalized routing centralities of some transit ASes are constantly around 0.2. Hence, some ASes hold a particular large normalized routing centrality in the Internet currently.

Here, we achieve a similar conclusion with what we got from the analysis of betweenness, which tells us that some transit ASes are very vital in the global routing. The failure of those ASes may result in a series of severe impairments, e.g., consuming enormous routers resource to converge new routing tables, evoking huge time delay, degrading the routing efficiency of the Internet, etc.

E. Routing Instability Analysis

Internet routing instability represents the fluctuation of routing information towards networks reachability. Many reasons are behind this phenomenon, including the change of infrastructure, the impact of traffic engineering, the employment of multi-homing, etc. However, high levels of routing instability can lead to serious impairments, e.g., packet loss, increase of network latency and time to convergence, and even the loss of interconnection availability in wide-area or national networks [14].

In inter-domain routing, the Internet routing instability can be observed from the fluctuation of the BGP routing table. In the following, we define the appearance time of an AS-level link i in a routing table as the occurrence count of the link, also define the average of the overall change rate as the routing instability rate, noted as RI. We consider RI as an adequate metric to quantify the routing instability. If we represent an undirected graph at time t with $\mathcal{G}_t = (\mathcal{V}_t, \mathcal{E}_t)$, where $\mathcal{V}_t$ is the set of the nodes and $\mathcal{E}_t$ is the set of links, the RI after time τ can be calculated as follows:



(a) AS interconnection diagram at t (b) AS interconnection diagram at t+τ

Fig. 11: AS interconnection diagrams

$$RI = \frac{1}{|\mathcal{E}_t|} \sum_{i \in \mathcal{E}_t} \frac{|n_i^t - n_i^{t+\tau}|}{\max(n_i^t, n_i^{t+\tau})} \tag{13}$$

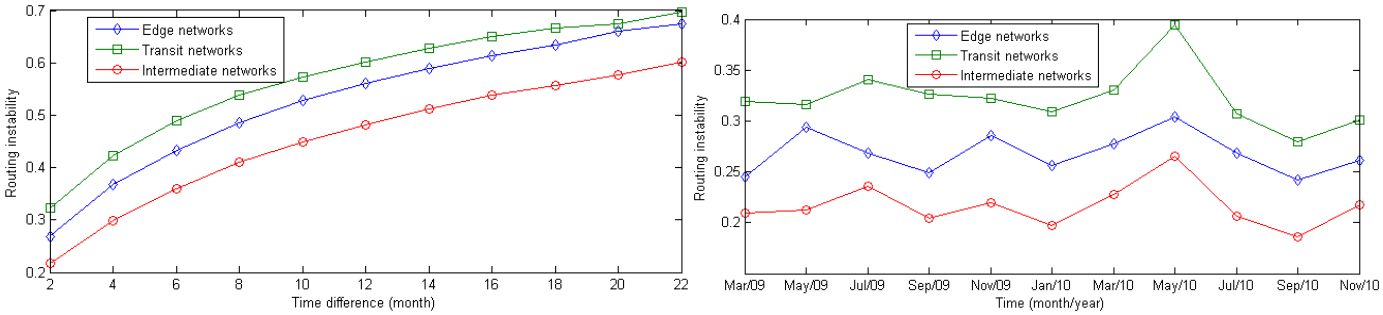
where, $|\mathcal{E}_t|$ is the size of the link set, n_i^t is the occurrence count of link i in the routing table at time t, and $n_i^{t+\tau}$ is the occurrence count of link i in the routing table at time $t+\tau$. If link i cannot be found in the routing table at time $t+\tau$, we set $n_i^{t+\tau} = 0$.

A demonstration of how to use (13) is shown here. Suppose we want to calculate the RI between Fig. 11(a) and Fig. 11(b), then $RI = 1/3 * (|5-1|/5 + (3-0)/3 + |1-5|/5) \simeq 0.87$. As there is considerable difference between Fig. 11(a) and Fig. 11(b), we get a very big RI, which represents the routing instability between the two graphs is in a significantly high degree.

We artificially partition the AS graph into three parts: edge networks constituted by edge ASes, transit networks constituted by transit ASes, and the intermediate networks connecting edge and transit ASes. Then we use (13) to measure the routing instability status of these three networks along the last two years, which are shown in Fig. 12a and Fig. 12b. In Fig. 12a, the X axis is the time difference τ and the Y axis is the routing instability given the time difference τ. In Fig. 12b, the X axis is the time t, and the Y axis is the routing instability between the routing table at time $t-\tau$ and the routing table at time t on a fixed time different $\tau = 2$ months. We find that the routing instabilities of the three networks all raise gradually in a similar way when the time difference increases. When the time difference is fixed at 2 months, the routing instabilities of the three networks also vary with a similar pattern.

From the two figures, we can assert that:

- The routing instabilities of the three networks have similar behaviors, and them all raise as long as the time



(a) Time difference (b) Time

Fig. 12: The diagrams of normalized ASes routing centrality

difference increases.

- Among the three networks, the intermediate networks have the least routing instability, while the transit networks have the largest routing instability.
- When time difference is fixed at 2 months, the routing instabilities of the three networks also share the similar pattern as time changes.
- The routing instability phenomenon is relatively serious presently, as the minimum value in the two figures is still around 0.2.

Two main factors can be behind such a routing instability: the inner convergence and oscillation problems of BGP, and the incentive of edge and transit networks in performing inbound and outbound traffic engineering operations.

V. Conclusion

Transit-edge routing separation functionally proposes to create a two-level hierarchical routing between networks that have different routing behavior. In this paper, we measure real inter-domain routing information to characterize the behavior and properties of edge and transit AS networks with a transit-edge routing separation perspective.

From an interconnection standpoint, we first analyze the diameters and the shortest paths between ASes pairs. We unravel that although the Internet grows constantly and AS path prepending impact the structure of the Internet significantly, the Internet service performance for most edge network would not degrade as long as a proper routing scheme can be deployed. Next, we found that the interconnection degree of an edge AS can be well fit with truncated power law distribution, while that of a transit AS can be fit by the combination of power law and inverse distribution, and we analytically and experimentally identified the different regimes of edge AS and transit AS degree distributions. From a routing and traffic engineering standpoint, we discovered that edge and transit ASes have similar probabilities of applying AS path prepending. We categorized the AS links into three types, and unraveled that they are affected by path prepending with different probabilities. We also discovered come up with the facts that edge and transit ASes have similar probabilities of applying AS path prepending, while transit ASes are more possible to utilize IP de-aggregation. We recognized that the impact ratios of BGP routing tables are directly determined by the IP prefix de-aggregation rate of edge and transit ASes, discovering that transit ASes do de-aggregate their own prefixes 3-times more often than edge ASes, which may appear surprising and counter-intuitive. Moreover, we described a mechanism to measure the routing instability phenomenon, recognizing that the transit networks have the largest routing instability while the intermediate networks have the least routing instability[3].

Kunpeng Liu received his Bach. degree in electrical engineering from Tsinghua University, Beijing, China, in 2005, and his M.Sc. degree from George Mason University in 2009, where he is now Ph.D. student at the Communications and Networking Lab. of GMU. His research interests are about future Internet routing and switching architectures.

Bijan Jabbari (F'09) received his Ph.D. degree from Stanford University in electrical engineering. He is a professor of electrical engineering at George Mason University, Fairfax, Virginia, and an affiliated faculty member with Telecom ParisTech (ENST), Paris, France. He is the past chairman of the IEEE Communications Society Technical Committee on Communications Switching and Routing. He is a recipient of the IEEE Millennium Medal (2000) and the Washington Metropolitan Area Engineer of the Year Award (2003). He continues research on multi-access communications and high-performance networking.

Stefano Secci (S'05-M'10) received the M.Sc. degree in communications engineering from Politecnico di Milano, Milan, Italy, in 2005, and Ph.D. degrees in computer science and networks from Politecnico di Milano and Telecom ParisTech, Paris, France, in 2009. He is an Associate Professor at the LIP6, Université Pierre et Marie Curie (UPMC - Paris VI - Sorbonne Universités), Paris, France, since Oct. 2010. He worked as Post-Doctoral Fellow with Telecom ParisTech, NTNU, Trondheim, Norway, and George Mason University, Fairfax, VA. His current research interests are about future Internet routing resiliency, mobility, and policy.

[3]Implementation and codes are given in [15].

REFERENCES

[1] K. Liu, B. Jabbari, and S. Secci, "Understanding transit-edge routing separation: Analysis and characterization," in *2011 International Conference on the Network of the Future (NoF'11)*, Paris, France, Nov. 2011, pp. 102–107.

[2] Y. Rekhter, T. Li, and S. Hares, "A Border Gateway Protocol 4 (BGP-4)," *RFC 4271*, 2006.

[3] S. Secci, K. Liu, G. K. Rao, and B. Jabbari, "Resilient traffic engineering in a Transit-Edge separated internet routing," in *ICC 2011 Communications QoS, Reliability and Modeling Symposium (ICC'11 CQRM)*, Kyoto, Japan, Jun. 2011.

[4] D. Meyer, "University of oregon route views archive project," *at http://archive.routeviews.org*.

[5] D. Farinacci, V. Fuller, D. Meyer, and D. Lewis, "Locator/ID separation protocol (LISP)," *draft-ietf-lisp-15*, 2011.

[6] E. Nordmark and M. Bagnulo, "Shim6: Level 3 multihoming shim protocol for IPv6," *RFC 5533*, 2009.

[7] R. Moskowitz and P. Nikander, "Host identity protocol (HIP) architecture," *RFC 4423*, 2006.

[8] Y. Wang, J. Bi, and J. Wu, "Empirical analysis of core-edge separation by decomposing Internet topology graph," in *Proc. of IEEE GLOBECOM*, 2010.

[9] Z. Mao, L. Qiu, J. Wang, and Y. Zhang, "On AS-level path inference," in *Proc. of ACM SIGMETRICS*, 2005, pp. 339–349.

[10] D. Cohen, "All the world's a net," *New Scientist*, vol. 174, no. 2338, pp. 24–29, 2002.

[11] T. Opsahl, F. Agneessens, and J. Skvoretz, "Node centrality in weighted networks: Generalizing degree and shortest paths," *Social Networks*, 2010.

[12] X. Meng, Z. Xu, B. Zhang, G. Huston, S. Lu, and L. Zhang, "IPv4 address allocation and the BGP routing table evolution," *ACM SIGCOMM Computer Communication Review*, vol. 35, no. 1, pp. 71–80, 2005.

[13] R. Gagliano, E. Grampin, J. Baliosian, X. Masip-Bruin, and M. Yannuzzi, "Understanding IPv4 prefix de-aggregation: challenges for routing scalability," in *Integrated Network Management-Workshops, 2009. IM '09. IFIP/IEEE International Symposium on*, 2009, pp. 107–112.

[14] C. Labovitz, G. Malan, and F. Jahanian, "Internet routing instability," *Networking, IEEE/ACM Transactions on*, vol. 6, no. 5, pp. 515–528, 1998.

[15] Details and codes, *at http://cnl.gmu.edu/TAVRI/research/*.