Evaluating Potential Routing Diversity for Internet Failure Recovery

Chengchen Hu[1,2], Kai Chen[3], Yan Chen[3], Bin Liu[1]
[1]CST Department, Tsinghua University, {huc, liub}@tsinghua.edu.cn
[2]SKLNST, Beijing University of Post and Telecommunication
[3]EECS Department, Northwestern University, {kchen, ychen}@northwestern.edu

***Abstract*—As the Internet becomes a critical infrastructure component of our global information-based society, any interruption to its availability can have significant economical and societal impacts. Although many researches tried to improve the resilience through the BGP policy-compliant paths, it has been demonstrated that the Internet is still highly vulnerable when major failures happen. In this paper, we aim to overcome the inherent constraint of the existing BGP-compliant recovery schemes and propose to seek additional potential routing diversity by relaxing BGP peering links and through Internet eXchange Points (IXPs). The focus of this paper is to evaluate the potentiality of these two schemes, rather than on their implementations. By collecting most complete AS link map up-to-date with 31K nodes and 142K links, we demonstrate that the proposed potential routing diversity can recover 40% to 80% of the disconnected paths on average beyond BGP-compliant paths. This work suggests a promising venue to address the Internet failures.**

I. INTRODUCTION

Despite of the remarkable availability and responsiveness demonstrated by the Internet in most cases, they still need a substantial improvement when a disaster/emergency strikes, such as 911 terrorists' attack [9] and Taiwan earth quake [13]. Recently, there are several literatures investigating the Internet resilience enhancement on the BGP routing layer [6, 8, 19], which only focus on exploration and utilization of the policy-compliant paths satisfying the inter-domain policy of BGP. Therefore, they are inherently constrained by the underlying resilience of BGP-based Internet routing structure. In [17], the authors disclose that 32% of all the ASes are vulnerable to the failure like a single critical customer-provider link cut, and up to 93.7% of the reachability of Tier-1 ISP's single-homed customers has been lost due to the failure like Tier-1 network depeering.

The BGP-based Internet routing protocol is a policy-driven paradigm under which the physical connectivity does not necessarily translate into network reachability, causing many physical paths not visible to the victim network in a failure emergency. To further make use of these physical paths, in this paper, we exploit and evaluate the *potential routing diversities*, which are one step beyond the scope of BGP policies. Specifically, at least two kinds of inter-domain links can be used for emergency recovery via simple configurations. They are,

- **Policy relaxation between neighboring ASes.** By relaxing the policy restrictions between victim AS and its neighboring survival ASes in an emergency, the previously policy-prohibited routing can be reused and the outage could be potentially recovered.
- **Setting up BGP sessions between IXP participants.** Internet eXchange Point (IXP) is a physical infrastructure that allows multiple ASes to exchange Internet traffic. Since the co-located ASes in a same IXP have routers geographically nearby and physically interconnected [10], new BGP sessions in the IXP can be easily set up.

Please note that while the idea of policy relaxation has been mentioned in [17], its potentiality to failure recovery is unknown. Furthermore, setting up new BGP sessions between IXP participants to rescue the failure emergencies is a novel idea proposed in this paper which is *the first contribution*.

To discover and utilize these *potential routing diversities* is an ambitious endeavor. As a first step towards this goal, in this paper, we examine the potentiality of these resources through a series of detailed measurements and simulation experiments. We manage to obtain the most complete AS graph to date. With such data, we evaluate the effectiveness of the proposed recovery schemes under various failures. This is *the second contribution* and also the major contribution of this paper.

The rest of this paper is organized as follows. Section II describes the two potential routing diversities. Section III introduces the evaluation methodologies. Section IV presents the evaluation results. Section V summarizes the paper.

II. POTENTIAL ROUTING DIVERSITY

A. Policy relaxation

The inter-domain routing between ASes is policy-driven, so the physical connectivity does not necessarily lead to network reachability. The exporting policies [4] of BGP routing allows: 1) each AS exports to its providers/peers its own routes and those learned from its customers or siblings; 2) each AS exports to its customers/siblings its own routes and any learned routes. It is shown in [4] that valid AS paths should follow the "valley-free"rule: AS paths start with a sequence of consecutive customer-provider links, followed by zero or one peer-peer link, ending with a sequence of consecutive provider-customer links. As a result, by relaxing peering links or customer-to-provider links into provider-to-customer links in a failure emergency, restricted physical paths would be visible to victim networks. Since peer-peer links usually have more bandwidth than the links to its customers, we focus on the potentiality of relaxing peering links in this paper.

B. Reconnection of IXP participants

An IXP is a physical infrastructure that allows any two ISPs to exchange Internet traffic between their networks by means of mutual peering agreements. Currently, most IXPs connect their participants through a layer-2 switches [10]. In an IXP, although the co-located ASes have their routers geographically nearby and physically interconnected by switches, whether to establish BGP peering sessions on top of the physical connectivity is up to individual networks. Two physically connected routers do not necessarily run a BGP session if the corresponding ASes do not have business relationship.

In case of Internet failure emergencies, we propose to use the potential resources in the IXPs to recover lost network connectivity. Our insight is that a new BGP session between two ASes (which are physically connected through switch previously) in an IXP can be easily established. In order to setup a BGP connection between the routers of victim AS and survival AS, each router only needs to input a command of "neighbor *ip-address* remote-as *AS-number*" to set a BGP session with a neighbor router on a Cisco router. We envision that the IP addresses and the AS numbers of the participants in an IXP can be obtained through the third party who manages it. Even an automated mechanism over the lay-2 switching fabric is a feasible alternative.

III. EVALUATION METHODOLOGY

A. AS topology

We use both BGP data and traceroute data to build a most complete AS graph to date as shown in Table I.

BGP data: The BGP data used in this study include a collection of BGP routing tables from 790 BGP speaking routers in 438 unique ASes. Specifically, we combine several BGP feeds: Routeviews [15], RIPE/RIS [14] and UCLA IRL lab [7]. We use 10 months of data gathered between Dec 1, 2007 and Sep 30, 2008. According to [10], ten months of the public view data should be enough to cover all the hidden links in the Internet AS graph. However, this graph still misses a lot of low-tier peering links [10].

Traceroute data: The traceroute data are collected by P2P users located in 580,000 hosts across the 5,500 ASes at the same time as BGP data. This is the largest scale measurement to date that consists of 541,023,742 measurements containing over 6.2 billion hops. Using a series of sophisticated heuristics, we have extracted about 23,000 AS links that are complementary to the BGP AS links [2].

Network classification: The Internet structure is considered to be loosely hierarchical. We apply the state-of-the-art method proposed in [10] which uses the number of downstream customer ASes to classify AS hierarchy. Besides 9 well-known Tier-1 ASes, we have 281 "large AS" (that has more than 50 customers), 1644 "media AS" (that has 5-50 customers), 4642 "small AS" (that has less than 50 customers), and 25269 "stub AS" (that does not provide transit service to any other AS).

B. Policy inference

We infer the business relationships between ASes based on the PTE algorithm proposed by Xia [18], which is considered to outperform most other approaches [5]. Most AS links are classified as one of three kinds of relationships: *customer-provider* links, *peering* links, and *sibling* links. We assume that the AS relationships did not change significantly within our ten-month measurement period. To justify this, we sample the AS relationships from CAIDA [1] for the past five years. With ten-month intervals, we find that more than 98.5% of AS pairs do not change their relationships. Since peering links are used for peer relaxation, we analyze the distribution of peering links. Without stub AS, the average number of peering links for an AS is 5.35 and an AS can peer with 503 ASes at most. Among all the non-stub ASes, 47.1% have at least one peering link and 18.4% have more than 10 peering links.

C. IXP dataset

There are several sources, such as Packet Clearing House [11], Peeringdb [12], and Euro-IX [3] that maintain a list of IXPs, as well as their participants. While there are more than 200 IXPs worldwide, we intentionally select 100 of them from Europe, South/North America, Asia Pacific and Africa according to their scale. Although there are more IXPs in the Internet, the number of participant ASes in these IXPs is quite small and the their contributions are minor. So we exclude the minor ones from the experiments in this paper. There are 3468 unique ASes (all of these ASes are not stub ASes) presented in our IXP data set. The average number of ASes in each IXP is about 59, with minimum number of 4 and maximum number of 321. One IXP member can be in 1.70 IXPs in average and be in 33 IXPs at most. 52.7% of the unique IXP members are only in one IXP and only 6.3% of the non-stub ASes can be in more than three IXPs.

D. Failure models

We measure the proposed routing diversities to understand their potentiality on recovering the Internet emergencies. Please note that we study the logical link outages, which correspond to failures of one or more physical links. In what follows, unless otherwise specified, a link denotes a logical link, and a node refers to an AS. We classify the failures according to the types of the failed links as the following.

- **Tier-1 depeering.** A depeering could be caused by contractual reasons, mis-configurations or physical damages. As pinpointed in [17] and as evidenced by contractual disputes between Cogent and Level 3 [17], Internet connectivity can be significantly affected by the depeering over a peer-to-peer link between two tier-1 ASes. To exam the failure of peering links teardown, we simulate the tier-1 depeering cases in our evaluation.
- **Access links teardown.** Provider-customer links connect the networks in different tiers of the Internet, which contribute to the major connectivity. Tier-1 ASes construct the core of today's Internet and are the top service providers. We study the failure that a number of tier-1 provider-custom links tear down which may badly affect the network connectivity.
- **Regional failures.** Regional failure breaks several peer links and provider-customer links, which can be caused by a regional emergency. We simulate the Taiwan earthquake failure and check how much potential routing diversity could be found in such a failure type.

	# nodes	# links	# customer-provider	# peering	# sibling
with stub AS	31845	142970	94500 (66.1%)	43709 (30.6%)	4761 (3.3%)
without stub AS	6576	78090	38580(49.4%)	35151(45.0%)	4359(5.6%)

TABLE I
STATISTICS OF THE AS TOPOLOGY IN OUR STUDY.

ASN	174	209	701	1239	2914	3356	3549	3561	7018
AS	PSINET	Quest	UUNET	SPRINT	Verio	Level 3	Global Crossing	Cable & Wireless	AT&T

TABLE II
LIST OF TIER-1 ASES.

E. *Potential routing diversity evaluation*

In the evaluation, we first simulate the failure models as mentioned above and recover the lost source-destination pairs with BGP itself. For the ones that can not be recovered, we explore the potential routing diversity by the following method.

- **Peer Relaxation (PR).** We validate this by investigating the reachability of each peer of the victim. As long as at least one of its peers reaches the recovering destination, we say that the connectivity of the victim to the destination could be recovered by relaxing the peer relationship.
- **IXP reconnection (IXP).** We first check whether a victim is in any of the 100 IXPs as mentioned before. If it is in one or more IXPs, the reachability of the co-located ASes in a same IXP to the recovering destination is examined. Again, the reachability could be recovered if at least one of the participants in the same IXP is able to access the destination.

IV. EVALUATION RESULTS

A. *Evaluation metrics*

To quantify the potentiality, we define the following metrics:

- **Recovery ratio.** This is the ratio between the number of non-reachable source-destination ($< Src, Dst >$) AS pairs[1] that can be recovered and the total number of non-reachable source-destination AS pairs.
- **Path diversity.** *Path diversity* of a source-destination pair stands for the number of parallel paths that do not share links and it can be formally defined as the minimal number of links (including the link itself) that must be removed in order to disconnect the two endpoints of the pair. This is used to show the path redundancies between the source AS and the destination AS.
- **Shifted path.** The traffic from the failed link will shift to a new link after recovery. Since it is impossible to get the traffic metrics over each two ASes, we count the number of the shortest valley-free paths to estimate the amount of traffic over a certain link as [17] did.

B. *Depeering*

In this section, we check the potential routing diversities during tier-1 depeering. Table II illustrates the nine well known tier-1 ASes. Totally 36 experiments are performed, and in each experiment, we: 1) assume one peering link between two tier-1 ASes to be down and then label the disconnected $< Src, Dst >$ pairs which are originally connected through this link; 2) seek alternative valley-free routes for the disconnected pairs and these do not have alternative routes (non-reachable pairs) are selected for evaluation of potential resources; 3) check the number of non-reachable pairs that can be recovered (*i.e.*, *recovery ratio via PR, IXP, and PR+IXP*).

1) Recovery ratio: Among the 36 experiments, with only peer relaxation, the minimum, mean and maximum recovery ratio are 0.44, 0.65 and 0.86; with only IXP participant reconnection, the minimum, mean and maximum recovery ratio are 0.28, 0.48 and 0.64; and with both peer relaxation and IXP participant reconnection, the minimum, mean and maximum recovery ratio are 0.63, 0.78 and 0.92. The absolute number of non-reachable pairs is 31456 in average (with a maximum of 144516 and with a minimum of 1996). The cumulative distribution of the percentage of experiments to recovery ratios via PR, IXP, and PR+IXP of the above experiment are depicted in Figure 1. The recovery ratio via PR is larger than those via IXP because the number of peering links that can be used for peer relaxation is much larger than the number of IXP participant reconnection. However, if a victim AS is involved in an IXP, the probability to get potential connectivity is much larger than the case it has peer links. This evidence is indicated from Figure 2. We look into the non-reachable pairs that are failed to be recovered by peer relaxation and IXP. The black part presents the proportion of the victim ASes that do not have peering links or do not appear in any IXP; while the white part is the proportion that no alternative routes are found by PR or IXP. Among all non-reachable pairs that cannot be recovered by peer relaxation, only 9.4% of them do not have peering links; In other words, 90.6% of them do have peering links but these links could not help during failures. Among all non-reachable pairs that cannot be recovered by IXP, 54.3% of them do not appear in any IXP; In other words, 45.7% of them is in IXPs but could not find IXP participant reconnection. The probability for a victim AS in an IXP but cannot be recovered is less than that for a victim AS having peering links but cannot be recovered.

2) Path diversity: Similar as the evaluations on recovery ratio, 36 experiments (each of which breaks two tier-1 ASes) are performed to study the improvement on the path diversity of the recovered AS pairs. After the introduction of potential connectivity, the path diversity of these pairs increases from zero. Among all the 36 experiments, the average path diversity is 3.6, the minimum path diversity is 1.9 and the maximum path diversity is 6.3. The distribution of the average path diversity in each experiment is depicted in Fig. 3. In most

---

[1] Note that in this section, the non-reachable source-destination pairs indicate these cannot be recovered by the BGP self-recovery

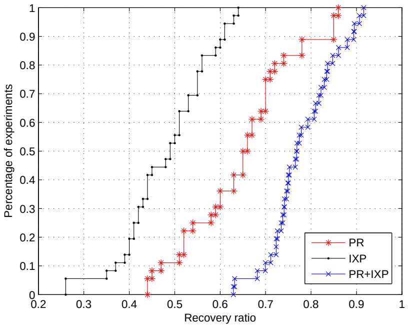

Fig. 1. CDF of the *recovery ratios via PR, IXP,* and *PR+IXP* for tier-1 depeering. It shows the proportion of experiments whose recovery ratio is less than a specific value.

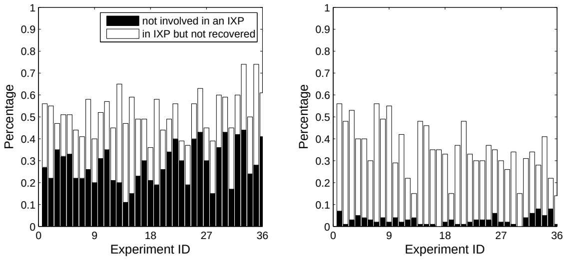

Fig. 2. Details of not recovered ASes through Peer Relaxation or IXP participant reconnection. The left one shows the results for PR and the right one indicates the results for IXP.

of the experiments, the path diversity is 2-3 (12 experiments, 33.3%) or 3-4 (14 experiments 38.9%). The existence of such multiple parallel paths increases the chance to avoid possible congestion caused by traffic moving.

3) Shifted path: We observe that the average increased number of paths traversing a link, *i.e.*, the average number of shifted path, varies from 3.75 to 17.2 in all the 36 possible scenarios of tier-1 AS depeering disaster. We also find that the number of the shifted paths to some links could be very large and the maximum number of the shifted paths to a link varies from 174 to 4217 in the 36 experiments. Fortunately, only a very small number of paths are shifted over most of the links. For example, when AS1239 and AS2914 depeers, the number of shifted path is no more than 4 over 86.3% of the links.

C. Access link teardown

Today's core Internet consists of a small number of tier-1 ASes acting as the top transit providers. Damages on several provider-customer links belonging to tier-1 ASes could cause the disconnections of a large number of its customers and grand-customers. With randomly picking 10-30 provider-customer links between tier-1 and its customers to be down, we mimic the failure of multiple provider-customer links.

Existing routing diversity by BGP valley free feature is first used to recover the disconnected AS pairs caused by the disconnection of multiple provider-customer links. And then the potential diversity is evaluated for the pairs that can not be recovered by BGP routing.

1) Recovery ratio: The number of non-reachable $< Src, Dst >$ pairs in this failure type is relatively smaller than the ones in tier-1 depeering example. Even there are 30 failed links for each tier-1 AS, the number of non-reachable $< Src, Dst >$ pairs is 1835 on average (with a maximum of 6515 and with a minimum of 240). The detailed recovery ratios via PR, IXP and PR+IXP are illustrated in Table III. Putting the two methods together, we can recover 1/3 to half of the non-reachable $< Src, Dst >$ pairs, even in the face of severe failures, *e.g.*, 30 customer-provider links down simultaneously. While PR and IXP provide similar potential routing diversities in this failure type, the number of victim ASes that are not in any IXP is larger than the victims that do not have peers. This indicates that IXPs are more efficient than peering links in failure recovery.

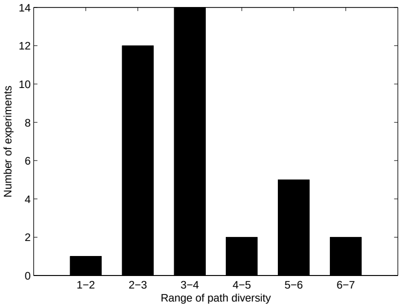

Fig. 3. Distribution of path diversity among 36 experiments, each of which breaks two tier-1 ASes. x-label indicates the range of path diversity and y-label counts the number of experiments.

2) Path Diversity: The path diversities of the recovered ASes pairs are also checked under this failure type. The average path diversity of non-reachable pairs is increased by 4.64 with peer relaxation and IXP participant reconnection when 10 provider-customer links are down. The increased path diversity does not show an obvious decrease if the number of broken links increases. The path diversity is 4.54 on average for the case that 20 provider-customer links fail.

3) Shifted path: The average number of shifted path is 3.4, 4.0 and 4.2 respectively when 10, 20 and 30 links are damaged. It is a slightly increasing trend with the increase of the number of the affected links. Although the number of shifted paths over some links could be as large as thousands, the number of shifted paths is less than 4 over about 87.9% to 98.3% of the links in different experiments.

D. Regional Failure

In this section, we check the potential resilience to regional failure occurs. Specifically, we investigate the improvement to Taiwan earthquake incident stricken in Dec. 2007.

1) Recovery ratio: It is reported in [16] that several ASes used newly built links for traffic recovery. To strengthen this fact, we checked the BGP paths for 9 of the most heavily affected ASes pointed in [16]. To simulate the disaster and study the above nine large victims, we randomly remove 50% of the links for each of the nine AS. By checking BGP AS paths, we collect the AS pairs that traverse these removed links as the lost $< Src, Dst >$ pairs. Then, we check whether there is an alternative path (*i.e.*, valley free) to recover the lost $< Src, Dst >$ pairs. The still non-reachable pairs are used to evaluate the potential connectivities. In the simulation, we observe that the number of non-reachable pairs caused by the failure of different victims varies from 644 to 103236. The recovery ratio via PR, IXP and PR+IXP are shown in Figure 4.

Links down	recovery ratio via PR	No peer	Recovery ratio via IXP	Not in IXP	Recovery ratio via PR+IXP
10	35.8%	10.7%	32.7%	26.2%	43.7%
20	27.8%	10.9%	27.6%	29.33%	38.1%
30	26.8%	9.4%	26.0%	29.8%	37.0%

TABLE III

POTENTIAL ROUTING DIVERSITY BY PR OR IXP WHEN SEVERAL PROVIDER-CUSTOMER LINKS ARE DAMAGED.

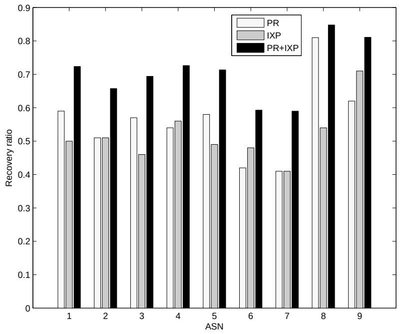

Fig. 4. The recovery ratio by potential connectivity diversity beyond BGP.

ASN	average	max.	less than 4
4134	6.2	413	89.6%
4755	6.0	790	82.4%
4761	7.6	529	84.4%
4795	6.9	261	78.7%
4837	7.2	259	77.7%
9498	7.9	652	79.3%
7473	9.1	3682	81.8%
9929	3.8	55	83,8%
24077	3.9	43	90.1%

TABLE IV

SHIFTED PATH FOR TAIWAN EARTHQUAKE FAILURE.

Each group of the three bars indicates the recovery ratio to the disconnections caused by the deleted links belonging to one of the victim AS. The average recovery ratios via PR and via IXP are 52% and 56%, respectively. Utilizing both of them, the average recovery ratio is increased to 70%. Among the not recovered non-reachable pairs through IXP participant reconnection, 63% of them are not involved in any IXP; and among the not recovered non-reachable pairs through peer relaxation, 25% of them do not have peer links.

2) Path diversity: The increased path diversity for the non-reachable pairs, which are caused by the broken links to the nine largest victim ASes as mentioned above, is evaluated. About 2.2 to 4.2 parallel paths that do not share any common link could be found in average through peer relaxation and reconnection in IXPs.

3) Shifted path: The number of shifted path is shown in Table IV. The first column indicates the AS number of the nine largest victim ASes. The second column is the average number of shifted paths, and the third column lists the maximum number of shifted paths. The fourth column demonstrates the percentage of links, over which the number of shift paths is less than 4. We observe that, a) the average number of shifted path is small; b) there are heavily used links which absorbs a large number of shifted paths; c) however, the numbers of such links are quite small (there are about less than 4 shifted paths over 77.7% to 90.1% links).

V. CONCLUSION

We have explored potential connectivity that can be utilized to recover Internet failures beyond existing BGP-based routing: 1) relaxing policy between neighboring ASes, and 2) setting up BGP sessions between IXP participants. We are the first one to propose the second scheme and the first one to evaluate the potential routing diversities for both schemes. With the largest AS map collected to date, we evaluate the potential routings that could be provided in various Internet disasters through peer relaxation and reconnections in a same IXP. The results shows that about 40% to 80% of disconnected $< Src, Dst >$ pairs can be restored beyond BGP with rich path diversity to avoid congested bottlenecks.

ACKNOWLEDGMENT

We want to thank Richard Yang for the insightful discussions with him at early stage of this work, and thank the anonymous reviewers for their valuable comments. This work is supported by NSF Award (CNS-0917233), NSFC (60625201, 60873250, 60903182), 863 project (2007AA01Z216), 973 project (2007CB310702), fund for the doctoral program of higher education (20060003058) and open project of SKLNST (SKLNST-2008-1-05), Tsinghua Research Plan Z02-2. Any opinions, findings, and conclusions expressed are those of the authors and do not reflect the views of the funding sources.

REFERENCES

[1] CAIDA. The CAIDA AS Relationships Dataset, 2004-2008. http://www.caida.org/data/active/as-relationships/.
[2] K. Chen, D. 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 *CoNEXT*, 2009.
[3] Euro-IX. Europe's leading internet exchange points. http://www.euro-ix.net/.
[4] L. Gao. On inferring Autonomous System relationships in the Internet. *IEEE/ACM Trans. Netw.*, 2001.
[5] Y. He, G. Siganos, M. Faloutsos, and S. V. Krishnamurthy. A systematic framework for unearthing the missing links: Measurements and Impact. In *USENIX/SIGCOMM NSDI'07*, 2007.
[6] N. Kushman, S. Kandula, D. Katabi, and B. M. Maggs. R-BGP: staying connected in a connected world. In *USENIX/SIGCOMM NSDI'07*, 2007.
[7] A. links. UCLA IRL. http://irl.cs.ucla.edu/topology/.
[8] M. Motiwala, M. Elmore, N. Feamster, and S. Vempala. Path splicing. In *ACM SIGCOMM'08*, pages 27–38, New York, NY, USA, 2008. ACM.
[9] A. Ogielski and J. Cowie. Internet Routing Behavior on 9/11 and in the following weeks. www.renesys.com/tech/presentations/pdf/renesys-030502-NRC-911.pdf.
[10] R. 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 *ACM SIGMETRICS'08*, 2008.
[11] Packet Clearing House. Internet exchange directory. http://www.pch.net/resources/data.php?dir=/exchange-points.
[12] Peeringdb. Peering database. http://www.peeringdb.com.
[13] A. Popescu, T. Underwood, and E. Zmijewski. Quaking tables: The taiwan earthquakes and the internet routing table. In *NANOG 39*, 2007.
[14] RIPE. Routing Information Service. http://www.ripe.net/projects/ris/.
[15] ROUTEVIEWS. Routeviews project. http://www.routeviews.org/.
[16] S. Wilcox. Quaking Tables: The Taiwan Earthquakes and the Internet Routing Table. http://www.thedogsbollocks.co.uk/tech/0705quakes/AMSIXMay07-Quakes.ppt, 2007.
[17] J. Wu, Y. Zhang, Z. M. Mao, and K. G. Shin. Internet routing resilience to failures: Analysis and implications. In *CoNEXT'07*, 2007.
[18] J. Xia and L. Gao. On the Evaluation of AS Relationship Inferences. In *IEEE GLOBECOM'04*, 2004.
[19] W. Xu and J. Rexford. MIRO: multi-path interdomain routing. In *ACM SIGCOMM'06*, pages 171–182, New York, NY, USA, 2006. ACM.