Neighbor-Cooperative Measurement of Network Path Quality

Rocky K. C. Chang, Waiting W. T. Fok, Weichao Li, Edmond W. W. Chan, and Xiapu Luo
Department of Computing, The Hong Kong Polytechnic University, Kowloon, Hong Kong
Email: {csrchang|cswtfok|csweicli|cswwchan|csxluo}@comp.polyu.edu.hk

***Abstract*—In the current Internet landscape, a stub autonomous system (AS) could choose from a number of providers and peers to advertise its routes. However, the route selection may not always result in a best choice in terms of end-to-end path performance. Instead of having an AS to monitor all possible paths, we argue that it is much more effective and beneficial for a number of neighboring ASes to cooperate in the path measurement. In this paper, we present a neighbor-cooperative measurement system in which each participating AS conducts measurement using their current routes for the same set of remote endpoints. A collation of the measurement results can help identify and correct poor routes, compare different providers' network services, and diagnose network performance problems. We report measurement results from an actual deployment involving eight neighboring universities for over a year.**

I. Introduction

Many stub autonomous systems (ASes) today use multiple providers to increase the reliability and performance of their Internet connection performance [1]. Moreover, they are peered with other ASes either directly or through local exchanges. As a result, they could choose from a number of peers and providers to send their outgoing traffic. The outbound traffic policies are often based on monetary and performance considerations. While it is not difficult to estimate the cost for implementing an outbound traffic policy, improving the performance of the forward routes (subjected to a cost constraint) is not straightforward.

A common approach to evaluating the forward routes' performance is to monitor them through passive and active measurement methods. However, the main disadvantage with this approach is that it does not scale well to a large number of possible routes and destinations. The process of monitoring the route's performance, analyzing the measurement data, and coming up a set of route recommendations is also a daunting task. Moreover, an AS can evaluate only the routes that it is using, but not other possible routes.

In this paper, we argue that it is more effective and scalable for a number of neighboring stub ASes to cooperate on finding better forward routes. This neighbor-cooperative approach is not entirely new. For example, Winick et al. [11] have proposed that neighboring ASes should cooperate for interdomain traffic engineering. This loose cooperation could allow the ASes involved to predict the impact of any change on the BGP configuration before effecting it. Our focus in this paper, however, is not on BGP-based traffic engineering, but on path-quality monitoring. Moreover, we use the term "neighboring" more loosely. Two ASes, which are not immediate neighbors, may also cooperate with each other in our case.

To facilitate the cooperation among a number of neighboring ASes, we propose a distributed measurement system that performs active path measurement for a set of remote endpoints from each participating AS. By comparing the measurement results collected from the ASes, the system could identify and improve poor routes. Moreover, by collating the measurement results, many path performance problems observed by a single AS could be more accurately diagnosed by analyzing the measurement data obtained from all participating ASes. We have deployed a neighbor-cooperative measurement system at eight universities in Hong Kong for over a year. We report several cases where the system plays a pivotal role in enhancing the quality of the network paths used by the member institutions, evaluating providers' performance, and diagnosing path-quality problems.

Sections II and III, the two main sections in this paper, describe the neighbor-cooperative measurement system and present measurement results from an actual deployment, respectively. After that, section IV discusses previous works that are related to this paper, and section V concludes this paper.

II. A neighbor-cooperative measurement system

We propose a neighbor-cooperative measurement system to facilitate neighboring ASes to conduct path measurement cooperatively. Figure 1 depicts the system that consists of n measurement nodes, m measured networks, and a measurement management system. The management server dispatches measurement tasks to the measurement nodes, monitors their resources usages, and retrieves measurement data from them. Each measurement node executes the measurement tasks deployed by the management server and measures the quality of the path from itself to the selected nodes in the m measured networks.

At least one measurement node is installed in each participating AS. Therefore, the system measures the paths to each remote destination from n vantage points which are in close proximity geographically. However, each AS manages its own routes independently. As a result, they generally select different routes for a given destination. This route diversity allows them to find the best route and to diagnose various path performance problems.

To ensure accurate and reliable path-quality measurement, we have paid attention to quite a few important issues in

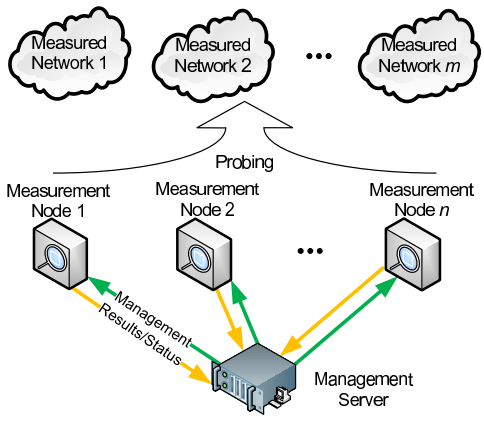


Fig. 1: The architecture of a neighbor-cooperative measurement system.

the design and implementation of the measurement system, including measurement methods, mitigating self-induced measurement bias, scheduling measurement tasks, and many others (such as, software design, performance tuning, and security).

1) Path-quality measurement method: We employ OneProbe [7] and Tcptraceroute [10] as the main measurement tools which are installed in each measurement node. OneProbe is an active measurement method which measures the quality of a network path to a web server. We choose OneProbe over ICMP ping, HTTPping, or other similar tools for its accurate, reliable, and metric-rich measurement. To augment OneProbe measurement with IP route information, we employ Tcptraceroute to obtain the IP forward path.

2) Mitigating self-induced measurement bias: Since a general concern with active measurement methods is self-induced measurement bias, we have taken several measures to mitigate it. First, we configure the measurement at a low frequency (such as 2 Hz). Second, we use two network interfaces to separate the measurement channel from the channel of receiving the measurement data, so that the latter will not affect the former. Third, the measurement node is installed as far out as possible in an AS. This setup is to prevent the measurement results from being biased by routers and middleboxes existing inside the AS. As we shall see in the next section, some middleboxes may mis-order the probe and response packets, thus biasing all measurement results.

3) Measurement scheduling and synchronization: In order to yield accurate correlation of the path measurement conducted from the n ASes, the measurement for a given destination is performed around the same time. To achieve this, each measurement node is synchronized to the closest time server every day. Moreover, to reduce the self-induced congestion, we divide the m remote destinations into several groups. The measurement nodes measure the paths to each group of destinations at the same time and measure the groups in a round-robin fashion.

III. Experience from a Deployment

We have been operating the system at eight universities in Hong Kong since 1 January 2009. The eight networks are peered with one another via a network called HARNET. We installed a measurement node at each university. Although the measurement nodes are all installed just behind their border routers, it is inevitable that they are still located behind some middleboxes, such as firewalls and load balancers. Each node measures paths to the web servers in the other seven universities and 36 other web servers in Hong Kong (via HKIX, a local exchange), Europe, US, Australia, China, Japan, Korea, and Taiwan.

Since all the measurement nodes measure the same destination around the same time, we minimize the measurement traffic by separating the destinations into five groups. Each group consists of eight to nine destinations. Therefore, the aggregated probing traffic from all measurement nodes towards a destination is less than 400 Kbits/second, assuming a sampling frequency of 2 Hz and a packet size of 1500 bytes. The measurement for each destination group lasts for one minute, and another minute is used for processing the measurement data. After that, the nodes measure the next destination group. As a result, each group is measured every ten minutes.

We highlight below three sets of measurement results to illustrate that a cooperative effort of measuring network paths could bring a greater benefit to all the participants in terms of identifying and correcting poor routes, evaluating providers' network services, and diagnosing network problems.

A. Identifying and correcting poor routes

Our system observed that the eight universities had very diverse RTT performance for the destinations in KREONET, CERNET, and TANET, which are research/academic networks in Korea, China, and Taiwan, respectively. Figure 2 shows the RTTs and packet loss rates of the routes from UB, UC, UE, and UF to the destinations in KREONET, CERNET, and TANET. The forward-path (Fw) and reverse-path (Rv) loss rates are plotted below the RTT. To differentiate the two types of loss rates, we plot the Fw and Rv loss rates above and below the 0% line, respectively. The routes UB→KREONET, UB→CERNET, and UF→TANET were quite good, because their RTTs were well below 50ms. Although the Fw loss rate was high for UF→TANET initially, it dissipated later on. On the other hand, the routes UC→KREONET, UE→CERNET, and UE→TANET experienced exceedingly high RTTs before 20 March 2009.

The Tcptraceroute results show that the universities employed different routes to reach these destinations.

1) KREONET: Figure 3(a) shows that the universities (which peered via HARNET) used three different routes to reach the destinations in KREONET. The most "direct" route (from HARNET to KREONET), however, yielded the highest RTT, which was about six times of the RTTs in the other two routes which went through the peering of HKIX and KREONET or through ASCC which is another AS in Taiwan.

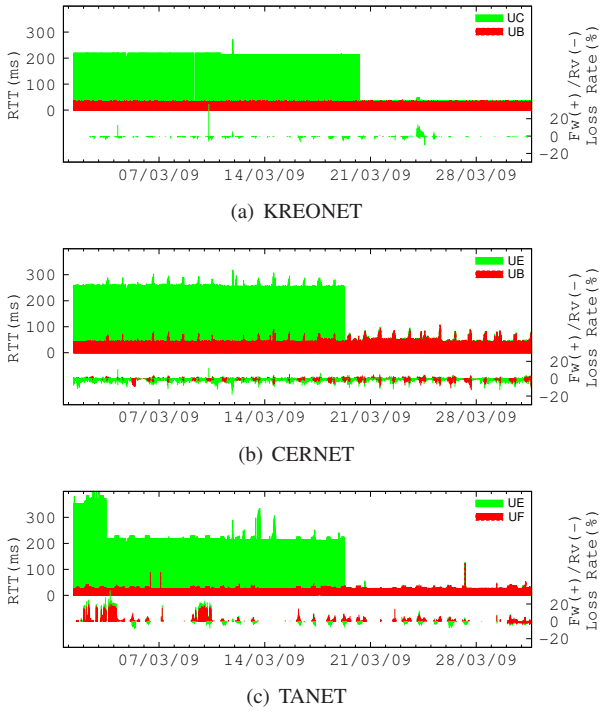


Fig. 2: RTTs of good and poor routes to KREONET, CERNET, and TANET.

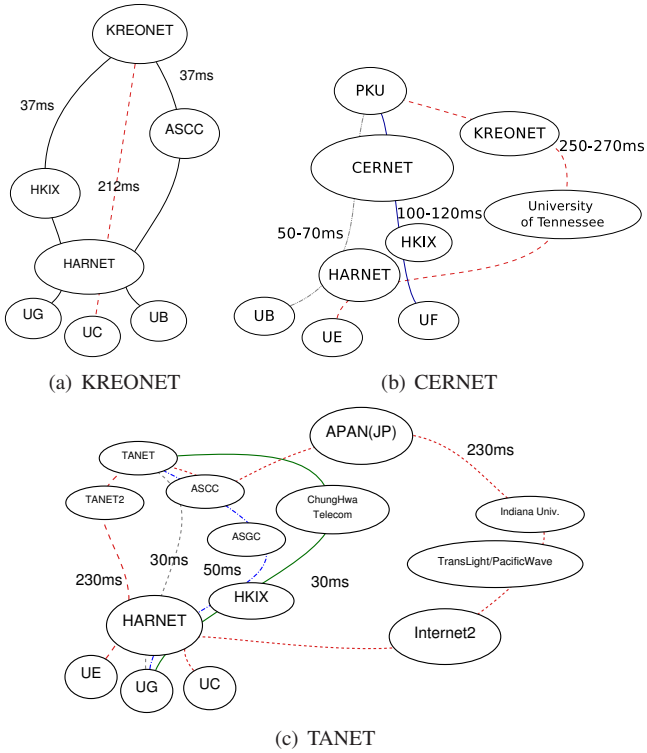


Fig. 3: A comparison of forward paths from the universities to destinations in KREONET, CERNET, and TANET.

A further analysis of the direct route reveals that the route went through a Trans-Pacific link.

2) CERNET: Figure 3(b) shows that the universities used three different routes to reach PKU in China. Two of them (which peered via HARNET) used CERNET and HKIX, and the third, for unknown reasons to us, used a route that went through the University of Tennessee in the US and then back to KREONET. It is clear that the third route incurs a much higher RTT than the other two routes.

3) TANET: Figure 3(c) shows that the universities (which peered via HARNET) used five different routes to reach the destinations in TANET. Altogether, 11 ASes were involved in the routes. The leftmost route (via TANET2) and the rightmost route (via Internet2) suffered from the highest RTTs. On the other hand, the other three routes traversed some ASes in Hong Kong and Taiwan, resulting in a low RTT.

After discovering these results, the universities which used the poor routes switched to use the good routes. Therefore, the poor route's RTT sees an abrupt decrease in each subfigure of Figure 2. Subsequently, the RTTs of both routes for each destination become very similar after 20 March 2009 (the green RTT is hidden behind the red RTT). It is also worth noting that the university which experienced a high RTT with the "direct" link to KREONET just needed to lower the preference setting of the Trans-Pacific link.

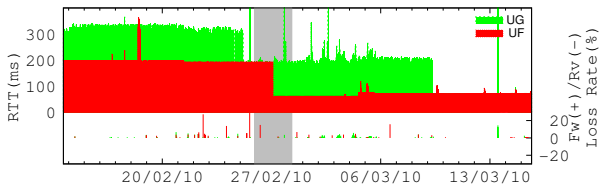
B. Evaluating provider's network services

Comparing different providers' network services is not a simple task. For example, Netdiff [8] was developed to compare different backbone ISPs' performance from a number of vantage points. The neighbor-cooperative measurement system also facilitates such evaluation based on the actual service subscriptions. Before February 2010, HARNET as a whole procured Internet transit services from a service provider (referred to as ISP1). Additionally, other universities also procured services from other providers for reliability and performance purposes. As a result, the system is able to observe the performance provided by different providers.

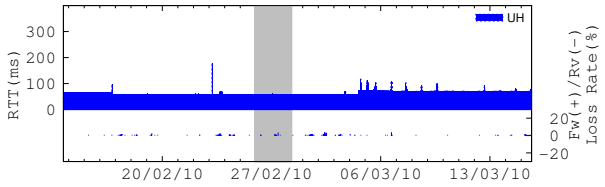
In particular, the system has noted that ISP1's network service in terms of RTT for the destinations in Japan, Europe, and US was actually poorer than that of another ISP (referred to as ISP2) used by university UH. Moreover, starting from late February 2010, the HARNET chose ISP2 to be the provider, and the system has observed that ISP2 was able to produce comparable performance as observed from the measurement at UH for destinations in Japan and Europe. However, the overall performance for the destinations in the US deteriorated.

1) Japan: Figure 4 shows one-month RTT and loss measurement for the paths to a web server in Japan obtained by UF, UG, and UH. The time period covers the services offered by ISP1 and ISP2. The shaded part is the period of switching from ISP1 to ISP2 during which the service for each university was switched one by one. Figure 4(a) shows that the RTTs for UF and UG were drastically reduced during the switch-over period, but we do not observe significant change in the packet loss rates. On the other hand, Figure 4(b) shows that UH's RTT was much better than UF's and UG's before the switch-over. However, after switching to ISP2, UF's, UG's, and UH's RTTs

all converged to approximately 100ms. Before the switch-over, UH already used ISP2 for destinations in Japan, whereas UF and UG used ISP1. Using different providers to reach the same destination therefore allows the system to compare their performance. Switching to ISP2 results in having UH, UF, and UG use the same paths to reach the destinations, therefore yielding similar RTT performance.



(a) UF's and UG's measurement



(b) UH's measurement

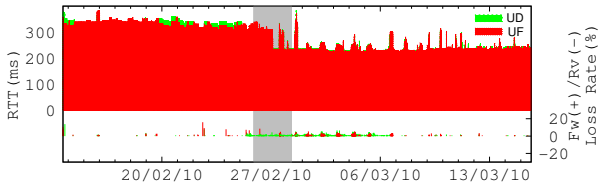
Fig. 4: RTTs and loss rates of network paths to a destination in Japan.

2) Europe: Figure 5 shows one-month RTT and loss measurement for the paths to a web server in Europe obtained by UD, UF, and UH. Similar to the last case, the time period covers the services offered by ISP1 and ISP2. A comparison of Figure 5(a) and Figure 5(b) shows that before the switch-over UH's RTT was lower than UD's and UF's by about 40ms, whereas UD's and UF's were very similar. During the switch-over, the RTTs for all three nodes improved, and their RTT measurements after the switch-over were very similar. Moreover, we have noted that the packet loss rates for UD and UF were higher in the first week after the switch-over, but this loss pattern did not appear in UH's measurement. Except for the first few hops, the forward paths for all three nodes were the same. Therefore, we suspect (but need to verify) that most of these packet losses came from the networks of UF and UH.

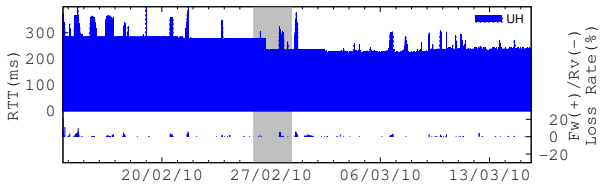
3) US: Figure 6 shows one-month RTT and loss measurement for the paths to a web server in the US obtained by UC, UG, and UH. A comparison of Figure 6(a) and Figure 6(b) shows that before the switch-over, except for a period of RTT surge, UH's RTT was slightly lower than UC's and UG's, whereas UC's and UG's RTTs were very similar. Unlike the two cases above, the overall path performance deteriorated after switching to ISP2: the RTTs for all three nodes increased, and there was also a higher delay variation for UC and UG.

C. Diagnosing network performance problems

The system observed various network performance problems from one or more measurement nodes. By collating the measurement from all the nodes, we are able to pinpoint the sources of the problems and characterize the impact of the

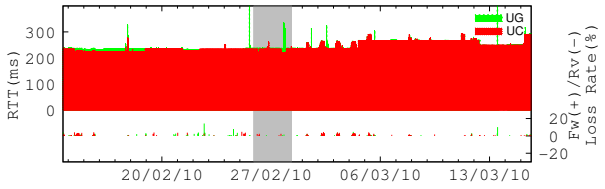


(a) UD's and UF's measurement

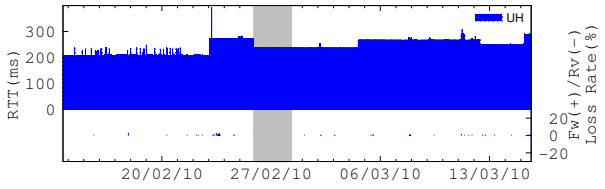


(b) UH's measurement

Fig. 5: RTTs and loss rates of network paths to a destination in Europe.



(a) UC's and UG's measurement



(b) UH's measurement

Fig. 6: RTTs and loss rates of network paths to a destination in the US.

problem on the network performance. We describe some of them below.

1) Anomalously high packet reordering rates: The system measured anomalously high (10%-20%) packet reordering rates for six paths from UB in May-July 2009. It is well known that a high packet reordering rate could seriously degrade TCP performance, because reordered packets are mistaken for packet losses [5]. A Tcptraceroute analysis identified an IP router common to these six paths, but this router did not appear in other paths for which the high reordering rate was absent. Subsequently, UB modified the routes to bypass this IP router, and the packet reordering became negligible.

2) Measuring the impact of a network device failure: The system observed the impacts of a network interface failure in HARNET, which is a fiber-optic ring, on 22 November 2009. The failure incident interrupted the HARNET service for around one hour. However, as observed from the path-quality measurement, the impacts were not uniformly felt across the eight universities. Four universities (UA, UH, UE, and UF)

were not affected by the interruption. The system observed that UA and UH switched to ISP2 and another provider to bypass the problem area. For the other four universities, UB and UD partially mitigated the problem by changing the forward routes for some measured networks, but other measured networks were unreachable during the period. In particular, UG was totally blacked out during the period, whereas UC only maintained connectivity to most local networks and some overseas networks. The cooperative measurement of this *unplanned* network failure provided us with an opportunity to evaluate the route resilience of each university network.

3) Discovering self-induced path performance problem: Another important advantage of the measurement cooperation is to discover self-induced path performance problem. The first step to discovering them is to identify consistent pattern of poor path performance. For example, referring to Table I for the forward-path packet reordering measurement for January 2010, UE's forward-path reordering measurement was significantly higher than other universities' measurement for most of the destinations. Moreover, the UE's routes to these destinations were basically the same as others that did not experience such high reordering rates. The system therefore concluded that the packet reordering was very likely occurred in the middleboxes in front of UE's measurement node. Consequently, it excluded UE's results for forward-path reordering measurement. There are also other similar cases for reverse-path packet reordering and packet losses.

Average%	UA	UB	UC	UD	UE	UF	UG	UH
APAN-JP	0.00	0.00	0.00	0.00	**0.03**	0.00	0.00	0.01
Local	0.00	0.00	0.00	0.00	**0.02**	0.00	0.00	0.01
TWGRID	0.00	0.00	0.00	0.00	**0.03**	0.00	0.00	0.01
KREONET	0.00	0.00	0.00	0.00	**1.00**	0.00	0.00	0.00
EU	0.00	0.00	0.00	0.00	**0.02**	0.00	0.00	0.01
US	0.00	0.00	0.00	0.00	**0.02**	0.00	0.00	0.01
JP	0.01	0.01	0.01	0.01	**0.03**	0.01	0.01	0.02
AU	0.00	0.00	0.00	0.00	**0.03**	0.00	0.00	0.01
I2	0.00	0.00	0.00	0.00	**0.03**	0.00	0.00	0.01
TANET	0.00	0.00	0.00	0.00	**0.03**	0.00	0.00	0.01
TEIN2	1.68	0.00	0.00	0.00	**2.54**	0.00	0.00	0.01

TABLE I: Relatively high forward-path reordering measurement obtained by UE in January 2010.

IV. Related work

A few previous works employed vantage points located in different ASes to detect network faults. For example, Zhang et al. used traceroute results from multiple ASes to diagnose routing disruptions [12]. Katz-Bassett et al. employed ping results from diverse vantage points to detect black holes in the Internet [6]. These works are related to ours in the sense that they also conducted cooperative measurement from different ASes. In addition to fault detection, our system also helps improve route performance, compare providers' networks services, and discover self-induced measurement biases.

In a broader sense, our measurement system uses network tomography [3] to diagnose path performance problems. However, the traditional network tomography problems mostly came from a network operator's point of view, such as characterizing loss distribution [4], locating congested IP link [9], and inferring temporal delay properties [2]. In contrast, our work is motivated by enabling stub ASes (i.e., the end systems) to improve their route performance by performing cooperative measurement. As a result, our vantage points are located close to each other, whereas the vantage points for traditional network tomography are usually distributed on a core network's boundary.

V. Conclusions

In this paper, we proposed a neighbor-cooperative approach to measuring and enhancing network path quality for participating ASes. We showed through our experience of operating such a system for eight universities that the cooperation is beneficial to all. By comparing path measurement for the same destination, some ASes discovered poor routes which were subsequently improved. The cooperative path measurement also provided an accurate and comprehensive comparison of different providers' network services and revealed problems which would not be discovered easily in the lack of such cooperation. While it is relatively easy for universities or NGOs to cooperate on path measurement, it is not clear to us whether there is enough incentive to cooperate in the commercial sector.

Acknowledgments

We thank the reviewers for their useful comments. This work is partially supported by a grant (ref. no. ITS/355/09) from the Innovation Technology Fund in Hong Kong and a grant (ref. no. H-ZL17) from the Joint Universities Computer Centre of Hong Kong.

References

[1] A. Akella, B. Maggs, S. Seshan, A. Shaikh, and R. Sitaraman. A measurement-based analysis of multihoming. In *Proc. ACM SIGCOMM*, 2003.

[2] V. Arya, N. Duffield, and D. Veitch. Temporal delay tomography. In *Proc. IEEE INFOCOM*, 2008.

[3] R. Castro, M. Coates, G. Liang, R. Nowak, and B. Yu. Network tomography: recent developments. *Statistical Science*, 19, 2004.

[4] N. Duffield, F. Presti, V. Paxson, and D. Towsley. Network loss tomography using striped unicast probes. *IEEE/ACM Trans. on Networking*, 14(4), 2006.

[5] L. Gharai, C. Perkins, and T. Lehman. Packet reordering, high speed networks and transport protocol performance. In *Proc. IEEE ICCCN*, 2004.

[6] E. Katz-Bassett, H. Madhyastha, J. John, and A. Krishnamurthy. Studying blackholes in the internet with hubble. In *Proc. USENIX NSDI*, 2008.

[7] X. Luo, E. Chan, and R. Chang. Design and implementation of TCP data probes for reliable and metric-rich network path monitoring. In *Proc. USENIX Annual Tech. Conf.*, 2009.

[8] R. Mahajan, M. Zhang, L. Poole, and V. Pai. Uncovering performance differences among backbone ISPs with Netdiff. In *Proc. USENIX NSDI*, 2008.

[9] H. Nguyen and P. Thiran. The boolean solution to the congested IP link location problem: Theory and practice. In *Proc. IEEE INFOCOM*, 2007.

[10] M. Toren. tcptraceroute. http://michael.toren.net/code/tcptraceroute/.

[11] J. Winick, S. Jamin, and J. Rexford. Traffic engineering between neighboring domains. http://www.cs.princeton.edu/~jrex/papers/interAS.pdf, July 2002.

[12] Y. Zhang, Z. Mao, and M. Zhang. Effective diagnosis of routing disruptions from end systems. In *Proc. USENIX NSDI*, 2008.