Incentives for BGP Guided IP-Level Topology Discovery

Benoit Donnet*

Université catholique de Louvain, CSE Deparment, Belgium

Abstract. Internet topology discovery has been an attractive research field during the past decade. In particular, the research community was interested in modeling the network as well as providing efficient tools, mostly based on traceroute, for collecting data. In this paper, we follow this track of rendering traceroute-based exploration more efficient. We discuss incentives for coupling passive monitoring and active measurements. In particular, we show that high-level information, such as BGP updates, might be used to trigger targeted traceroutes. As a result, the network dynamics might be better capture. We also provide a freely available tool for listening to BGP feeds and triggering dedicated traceroutes.

1 Introduction

The past ten years have seen a growing body of important research work on the topology of the Internet [1]. Since Faloutsos et al. seminal paper on the power-law relationships in the Internet [2], researchers strongly investigated the Internet topology at the IP, router, and AS level. The IP level considers routers and end-systems IP interfaces. The basic idea for collecting data is to probe the Internet from multiple vantage points using the technique of *traceroute*. The router level considers each router as being a single node in the topology. This is done by aggregating a router IP interfaces under a single identifier using *alias resolution*. Finally, the AS level provides information about autonomous systems (ASes) connectivity. The past research efforts were done on Internet modeling and techniques for efficiently collecting data. In this paper, we push further techniques for gathering data for the IP level Internet topology by providing incentives for using high-level information for triggering traceroute-like exploration.

The traceroute-based exploration works as follows: probes, basically UDP packets, are sent with increasing TTL values. When the TTL expires, an intermediate router is supposed to reply with an ICMP 'Time Exceeded" message to the sender. By looking at the IP source address of this ICMP message, the measurement point can learn one of the IP address of the router. When the probe

* This work has been partially supported by the European Commission-funded 034819 OneLab project. Benoit Donnet is funded by the Fonds National de la Recherche Scientifique (FNRS – Rue d'Egmont 5, 1000 Brussels).

M. Papadopouli, P. Owezarski, and A. Pras (Eds.): TMA 2009, LNCS 5537, pp. 101–108, 2009.

reaches the destination, the destination is supposed to reply with an ICMP "Destination Unreachable" message with code "port unreachable". This works if the specified port in the UDP probe is presumably unused. Extensions to traceroute have been proposed to use ICMP and TCP probes.

Unfortunately, probing this way from multiple vantage points towards a large set of destinations is somewhat inefficient. First some routers along the path are repeatedly discovered for each traceroute [3]. Second, it is time consuming. For instance, the recent *Archipelago* [4] infrastructure takes roughly three days to complete its destination list. In such a context, it is very difficult to capture the network dynamics. Efforts have been made for rendering traceroute exploration less redundant [3,5,6], allowing also to speed up the exploration process. However, this does not entirely solve the network dynamic capture issue.

In this paper, we follow this track of rendering traceroute-based exploration more efficient. As recently mentioned by Eriksson et al. "passive measurements of packet traffic offer the possibility of a greatly expanded perspective of Internet structure with much lower impact and management overhead" [7]. We echo this call by proposing a way to discover the Internet topology at the IP level by using passively collected information for triggering (and guiding) traceroute.

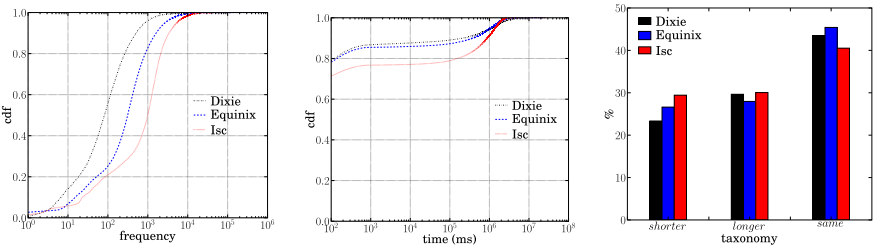
We propose to consider BGP information to guide probing and trigger specific targeted traceroute. In particular, we focus on updates that modify two given BGP attributes: the AS_PATH and the communities. We argue that a change in one of these attributes might be a route change indication and, thus, be considered as a trigger event for launching a traceroute towards a specific prefix. By acting so, a traceroute system might better capture network dynamics information. This is thus complementary to existing tools. In addition to this, we provide a tool for listening to BGP feed and deciding whether a traceroute must be launched or not.[1]

The remainder of this paper is organized as follows: Sec. 2 explains how BGP information might be used for guided probing; Sec. 3 discusses our implementation; Sec. 4 positions our work regarding the state of the art; finally, Sec. 5 concludes this paper by summarizing its main contributions and discussing further research directions.

2 BGP as Trigger Event

In this section, we study how some BGP events might be used to trigger targeted traceroutes. We base our evaluation on Routeviews [8] data, starting from October 1^{st}, 2007 to September 30^{th}, 2008. The Routeviews project aims at frequently collecting BGP table dumps and BGP update messages from the perspective of several locations. For our study, we considered three BGP routers: Dixie (Japan), Equinix (United States of America), and Isc (United States of America). Finally, we only took into account IPv4 routes.

[1] The code is freely available, under a BSD-like license at `http://gforge.info.ucl.ac.be/projects/bgpprobing/`



(a) Frequency (b) Time interval (c) Taxonomy

Fig. 1. As_Path modification

2.1 AS_PATH

The As_Path is a standard BGP attribute that is used to list the ASes a route advertisement has traversed. For a given prefix, if the As_Path is modified between two BGP updates or between a BGP update and the current record in the routing table, it means that the path has changed. This can be seen as a trigger event for a traceroute exploration towards the source prefix advertised in the BGP update.

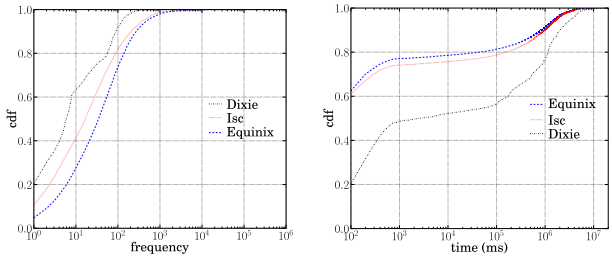
Fig. 1 shows statistics on the As_Path modification over time. In particular, Fig. 1(a) shows the cumulative distribution of modifications frequency (horizontal axis in log-scale), i.e., how many times, for each prefix, the As_Path has changed over the considered period. Fig. 1(b) shows the time interval (in ms – horizontal axis in log-scale) between two As_Path modifications for a given prefix.

We see that in 50% of the cases, an As_Path is modified more than 1000 times for the Isc router (Fig. 1(a)). However, the time interval between two modifications is extremely short (less than 100ms) in 80% of the case (Fig. 1(b)), probably due to a path exploration process. Nevertheless, there is a kind of plateau between 1.000 and 1.000.000ms in the remaining 20%, suggesting so that As_Path changes might be somewhat "persistent".

Fig 1(c) gives, for each Routeviews router, the taxonomy of the BGP As_Path attribute modification. An As_Path can be *shorter* (the new As_Path counts less intermediate ASes than the recorded one), *longer* (the new As_Path counts more intermediate ASes than the recorded one), or *same length* (the new As_Path counts the same number of ASes than the recorded one but at least, one of them is different). It is interesting to notice that, in most of the cases, the modified As_Path has the same length that the previous As_Path.

2.2 BGP Communities

The BGP communities attribute provides a way of grouping destinations into a single entity, named *community*, to which similar routing decisions might be



(a) Frequency (b) Time interval

Fig. 2. BGP communities attribute modification

applied. A BGP communities attribute is composed of one or more 32 bits numbers. These numbers are structured as follows: the high-order 16 bits represent an AS number, while the low-order 16 bits define the semantic of the value. Each AS can use the 2^{16} communities whose high-order 16 bits are equal to its own AS number.

Donnet and Bonaventure recently showed that the BGP communities attribute is more and more used [9]. They further proposed a classification of BGP communities usage. They identified three classes:

- *inbound* communities refer to communities added or used when a route is received by a router on an eBGP session. It is typically used for setting a particular value to the Local_Pref attribute (i.e., the degree of preference for an external route) or for tagging route with the location where it was received from an external peer.
- *outbound* communities are used by a router to filter BGP announcements for traffic engineering purposes. A community is inserted by the originator of the route in order to influence its redistribution by downstream routers.
- *blackhole* communities refers to a particular BGP community used by an ISP to block packets. These communities are used only inside ISPs and should not be distributed on the global Internet.

It is clear that a change in inbound communities (in particular those tagging the received route) might indicate a change in the path a packet follows and, thus, be considered as traceroute trigger-event.

In the fashion of Fig. 1, Fig. 2 presents statistics on the BGP communities attribute modification. We see that modifications are much less frequent than for the As_Path attribute, while we observe the same kind of behavior for the time interval between changes.

In the fashion of the As_Path, the time interval between two BGP communities attribute modification is quite short. Except for Dixie, in 60% of the cases, the time interval is less or equal to 100ms.

Up to now, we have seen that BGP communities might change over time. If we are able to identify to which class (see Donnet and Bonaventure for details [9])

Table 1. Classification of BGP communities changes

Router	Inbound				Outbound	
	IXP	Type of Peer	Geographic	AS	Announcement	prepending
Dixie	0.27%	7.14%	1.01%	1.33%	4.96%	0.11%
Equinix	16.75%	52.52%	30.01%	0%	0.93%	0.51%
Isc	0.06%	20.78%	43.55%	0.08%	2.88%	0.59%

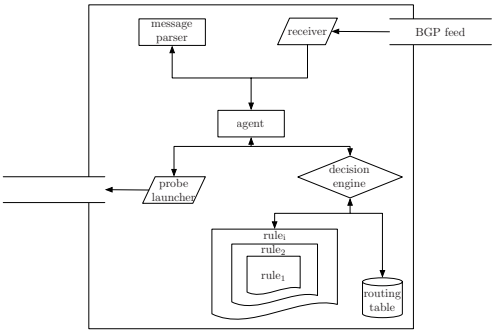


Fig. 3. Interactions between modules

belong the modified communities attribute, we can potentially trigger traceroute. Using the database provided by Donnet and Bonaventure, we tried to perform this classification on our six months dataset. Results are shown in Table 1. It provides a proportion of modified BGP communities we were able to classified. Due to the lack of standardization and documentation of the BGP communities attribute, we were not able to classify all the BGP communities (in particular for Dixie). We however identified an interesting proportion of modifications in "Geographic" BGP communities attribute (for instance, 43.55% for Isc). This means that, for the Isc router, in 43.55% of the cases, a modification of the BGP communities attribute concerns the geographic location of a route received from an external peer.

Such an observation is of keen interest of us as it clearly indicates a route change and is thus a good trigger-event for a traceroute exploration.

3 Implementation

We implemented a tool for listening to BGP updates and determine whether a traceroute must be triggered or not towards a particular prefix. Fig. 3 shows a high-level view of our implementation.

The *Receiver* module aims at listening to BGP incoming BGP messages. These BGP messages can directly come from a BGP feed provided by local operator (byte streams as defined in RFC 1771 [10]) or from the BGPMon project [11] (XML files as defined by Cheng et al. [12] - a particular message parser then be

implemented). The *Decision* module is in charge of deciding whether the received message can trigger a traceroute or not. This decision is based on existing information (the routing table - the system uses an existing routing table as input and this routing table is updated with incoming messages) and the applications of *rules*. Currently, four rules have been implemented:

- Withdraw rule. An existing route is suppressed from the routing table.
- Add rule. A non-existing route is added to the routing table.
- As_Path rule. The As_Path attribute of an existing route changes (as discussed in Sec. 2.1).
- BGP communities rule. The BGP communities attribute of an existing route changes (as explained in Sec. 2.2).

The system has been implemented so that a new rule can be easily implemented and added to the system.

Nevertheless, even if one of the rules above is matched, it does not necessarily trigger a traceroute. Several conditions must be checked before. Indeed, some prefixes might generate route flapping [14] or be in a path exploration process. In such a case, traceroute should not be launched. A traceroute will be triggered at the following conditions:

- The prefix contained in the message did not trigger a traceroute recently. A timed-cache (i.e., a timer is associated to each entry in the cache), system has been implemented to avoid to constantly probing the same prefix. If the prefix is in the cache, the traceroute is not trigger and the associated timer in the cache is reset. At the timer expiration, the corresponding entry is removed from the cache.
- The received BGP message is not considered as noise. A received prefix is considered as noise if it belongs to the top 20 of unstable prefix (according to Geof Huston weekly report [13]) or if the route is flapping (route flap damping algorithms have been implemented [14,15]).
- The token bucket is not full. In order to avoid flooding the traceroute server and the network, traceroute are triggered at a certain rate.

4 Related Work

Systems, such as Ripe Ncc *TTM* [16] and Nlanr *AMP* [17], consider a larger set of monitor, several hundreds, but avoid to trace outside their own network. A more recent tool, *DIMES* [18], is publicly released as a daemon. *Rocketfuel* [5] focuses on the topology of a given ISP and not on the whole Internet topology as skitter does, for instance. *Scriptroute* [6] is a system that allows an ordinary Internet user to perform network measurements from several distributed vantage points. Finally, the recent *iPlane* constructs an annotated map of the Internet and evaluates end-to-end performances (latency, bandwidth, capacity, etc). Finally, the recently deployed *Archipelago* [4] probes all routed /24 from several locations. Others have proposed improvements to traceroute for reducing measurement redundancy [3,19] or for avoiding anomalies [20]. None of these

aforementioned works provide a link with higher level information, such as BGP, to guide probing.

Finally, topology discovery might be done through a deployment facility. Examples of such a system are *m-coop* [21], *pMeasure* [22], and *DipZoom* [23]. These solutions are complementary to our tool as they can be used to dispatch the traceroute trigger to several vantage points.

5 Conclusion

The Internet topology at the IP interface level has attracted the attention of the research community for a long time now. People are interested in modeling the network as well as in traceroute-based tools for efficiently collecting data.

In this paper, we made a step towards a more network-friendly traceroute-based system. Indeed, we discussed incentives for considering high-level information, such as BGP data, as a trigger event for targeted traceroutes. In particular, we focused on two BGP attributes, the As_Path and the communities. We believe that a tracing system using this kind of information can increase its coverage capabilities by better capturing network dynamics. In addition, we provide a freely available implementation of a tool for listening to BGP feed and deciding whether a traceroute must be sent or not.

A deployment of our tool using for instance BGPMon [11] should reveal, in the near future, to what extend we are able to capture network dynamics.

References

1. Donnet, B., Friedman, T.: Internet topology discovery: a survey. IEEE Communications Surveys and Tutorials 9(4), 2–15 (2007)
2. Faloutsos, M., Faloutsos, P., Faloutsos, C.: On power-law relationships of the internet topology. In: Proc. ACM SIGCOMM (September 1999)
3. Donnet, B., Raoult, P., Friedman, T., Crovella, M.: Efficient algorithms for large-scale topology discovery. In: Proc. ACM SIGMETRICS (June 2005)
4. claffy, k., Hyun, Y., Keys, K., Fomenkov, M.: Internet mapping: from art to science. In: Proc. IEEE Cybersecurity Applications and Technologies Conference for Homeland Security (CATCH) (March 2009)
5. Spring, N., Mahajan, R., Wetherall, D.: Measuring ISP topologies with Rocketfuel. In: Proc. ACM SIGCOMM (August 2002)
6. Spring, N., Wetherall, D., Anderson, T.: Scriptroute: A public internet measurement facility. In: Proc. USENIX Symposium on Internet Technologies and Systems (USITS) (March 2002)
7. Eriksson, B., Barford, P., Nowak, R.: Network discovery from passive measurements. In: Proc. ACM SIGCOMM (August 2008)
8. University of Oregon: Route views, University of Oregon Route Views project See, `http://www.routeviews.org/`
9. Donnet, B., Bonaventure, O.: On BGP communities. ACM SIGCOMM Computer Communication Review 38(2), 55–59 (2008)
10. Rekhter, Y., Watson, T.J.: A border gateway protocol 4 (BGP-4). RFC 1771, Internet Engineering Task Force (March 1995)

11. Yan, H., Matthews, D., Burnett, K., Massey, D., Oliveira, R., Zhang, L.: BGPmon: a real-time, scalable, extensible monitoring system. In: Proc. IEEE Cybersecurity Applications and Technologies Conference for Homeland Security (CATCH) (March 2009)
12. Cheng, P., Yan, H., Brunett, K., Massey, D., Zhang, L.: BGP routing information in XML format. Internet Draft (Work in Progress) draft-cheng-grow-bgp-xml-00, Internet Engineering Task Force (February 2009)
13. Huston, G.: BGP update report (2008), See: `http://www.potaroo.net`
14. Villamizar, C., Chandra, R., Govindan, R.: BGP route flap damping. RFC 2439, Internet Engineering Task Force (November 1998)
15. Mao, Z.M., Govindan, R., Varghese, G., Katz, R.H.: Route flap damping exacerbates internet routing convergence. In: Proc. ACM SIGCOMM (August 2002)
16. Georgatos, F., Gruber, F., Karrenberg, D., Santcroos, M., Susanj, A., Uijterwaal, H., Wilhelm, R.: Providing active measurements as a regular service for ISPs. In: Proc. Passive and Active Measurement Workshop (PAM) (April 2001)
17. McGregor, A., Braun, H.W., Brown, J.: The NLANR network analysis infrastructure. IEEE Communications Magazine 38(5) (2000)
18. Shavitt, Y., Shir, E.: DIMES: Let the internet measure itself. ACM SIGCOMM Computer Communication Review 35(5) (October 2005)
19. Donnet, B., Friedman, T., Crovella, M.: Improved algorithms for network topology discovery. In: Dovrolis, C. (ed.) PAM 2005. LNCS, vol. 3431, pp. 149–162. Springer, Heidelberg (2005)
20. Augustin, B., Cuvellier, X., Orgogozo, B., Viger, F., Friedman, T., Latapy, M., Magnien, C., Teixeira, R.: Avoiding anomalies with paris traceroute. In: Proc. ACM USENIX Internet Measurement Conference (IMC) (October 2006)
21. Srinivasan, S., Zegura, E.W.: Network measurement as a cooperative enterprise. In: Druschel, P., Kaashoek, M.F., Rowstron, A. (eds.) IPTPS 2002. LNCS, vol. 2429, pp. 166–177. Springer, Heidelberg (2002)
22. Liu, W., Boutaba, R., Won-Ki Hong, J.: pMeasure: a tool for measuring the internet. In: Proc. 2nd Workshop on End-to-End Monitoring Techniques and Services (E2EMON) (October 2004)
23. Wen, Z., Triukose, S., Rabinovich, M.: Facilitatiing focused Internet measurements. In: Proc. ACM SIGMETRICS (June 2007)