Collection and Analysis of data for Inter-domain Traffic Engineering

Juan Camilo Cardona[1], **Pierre Francois**[1], **Paolo Lucente**[2]

[1]IMDEA Networks Institute – Av. Mar Mediterraneo 22 – Leganes – Madrid – Spain

[2]Cisco Systems

{juancamilo.cardona, pierre.francois}@imdea.org, plucente@cisco.com

Abstract. *Inter-domain IP traffic engineering requires the collection and analysis of data that is distributed among multiple network platforms. This process is usually considered complex for network operators and limits the efficiency that traffic management methods can achieve. The introduction of Software Define Networks may change this rigid context, as it is pushing operators to request flexible routing system architectures. In the future, the network operation team might have the resources to obtain, maintain, and analyze a rich variety of data from within and outside their networks. Without the need to change many of their routing devices, such data can be leveraged to implement more complex inter-domain TE applications. In this paper, we review the process of data collection and management required to benefit from such techniques. Additionally, we describe various enhanced TE applications, providing details of one of them in a case study.*

1. Introduction

Network operators must control the distribution of traffic crossing their network, in order to provide highly available and efficient services. Traffic engineering (TE) is referred to as the set of techniques and tools that operators utilize to this end [Awduche et al. 2002]. For the case of Internet Service Providers (ISPs), TE processes can be divided in two types: the management of how traffic flows within their own network infrastructure (intra-domain TE), and the control on where traffic enters and where traffic exits onto neighboring networks (inter-domain TE). These two types are very different from an operational point of view. For the intra-domain case, operators have complete control of the protocols and configuration of the devices of the network. Inter-domain methods, on the other hand, must consider the policies of external, sometimes remote, Autonomous Systems (ASes). In *inter-domain TE*, operators should not only be concerned by *how to route* traffic, but also by *how to assess* the efficiency of their strategies and the impact of the policies of external ASes on them.

The implementation of enhanced inter-domain TE applications is difficult due to the challenges that network operators face in relation to the collection and analysis of network data. Network operators must gather and correlate multiple types of data, such as traffic statistics, device configurations, and BGP feeds, to build proper strategies for their inter-domain traffic [Feamster et al. 2003]. To obtain this data, management systems must connect with devices of different classes and vendors. A few protocols have been proposed to serve as single monitoring interfaces to heterogeneous devices [Stallings 1998];

however, these protocols still suffer from considerable rigidness and lack of consistency across vendors. Furthermore, many data sources are located in information systems that only support other types of interfaces, such as SQL databases, XML, or MRT files. These obstacles have obliged operators to either build ad-hoc monitoring platforms, or resort to management systems that only support basic levels of data.

The ability of network operators to perform inter-domain TE can be boosted by the introduction of more powerful information analysis systems and flexible interfaces for network data collection. The former can be possible thanks to the proliferation of architectures and technologies designed for big data analysis [Agrawal et al. 2011]. The latter can be the result of the recent demand for networks supporting the Software Defined Networking (SDN) model. SDN promotes the decoupling of forwarding and control planes from the physical infrastructure, therefore facilitating the creation of programmable networking systems [Casado et al. 2012]. The demand for SDN has pressured manufacturers to implement into their systems more flexible protocols and Application Programming Interfaces (APIs), such as NETCONF/YANG [Bjorklund 2010]. Additionally, the introduction of SDN-like features in the network design cycle can create an environment prone for the analysis of network data. Therefore, network operators will have in the near future the necessary resources to analyze inter-domain traffic data.

The objective of this paper is to describe how operators can put into effect different inter-domain TE by exploiting multiple sources of data. We provide a description on the inter-domain related data and the methods that ISPs can use to collect them in Section 2. We then revisit some applications that use this data for inter-domain TE in Section 3. We provide detail of one of these applications in Section 4, where we describe how unexpected traffic flows can be generated by overlapping prefix filtering, and how an AS can detect them by correlating multiple sources of data. Finally, we conclude in Section 5.

2. Data collection

In this Section, we provide a summary of the data that ISPs can collect to analyze their network requirements and perform efficient inter-domain TE. We describe different methods to obtain this data, their limitations, and briefly provide recent techniques and protocols, proposed in the last few years, to facilitate data collection. In Sections 3 and 4, we describe different applications that leverage this data for inter-domain TE purposes.

2.1. BGP Feeds

BGP feeds describing inter-domain paths reaching the border of an ISP are necessary for analyzing the different route alternatives available at each router in the network. In order to perform a complete network analysis, ISPs need to collect every path received from external neighbors. The simplest architecture for BGP data collection consists of a centralized collector, which communicates with all edge BGP devices. The completeness of the data collection depends on the approach / protocol followed to convey the data to the collector. We describe some of the options next.

IBGP An iBGP session is one of the options to communicate the centralized collectors and the edge devices. The main problem of using iBGP is the limitation of the

protocol to send a single path per prefix [Rekhter et al. 2006]. This can potentially hide many of the paths sent by external ASes. ADD-PATH, best external, or other path diversity techniques can be deployed to overcome this limitation [Walton et al. 2012][Marques et al. 2012][Raszuk et al. 2012]. From these techniques, ADD-PATH provides the option of delivering the largest amount of paths with minimum configuration overhead. The problem with ADD-PATH and the other techniques improving path diversity is that they do not signal to the BGP neighbor the paths that are actually selected as best. Therefore, operators would need to simulate the BGP decision process or complement the analysis with other data sources to obtain this data. In the future, the information of the state of the paths in the local-RIB of each devices, such as whether the path was selected as best, one of the best, or a valid alternative could, for example, be included in BGP using communities [Cardona et al. 2013].

BGP Monitoring Protocol (BMP) BMP is a simple protocol that provides access to the information stored in the Adj-RIB-in tables of BGP devices [Scudder et al. 2012]. Similar to ADD-PATH, BMP can currently not signal information of the state of the path after the BGP selection process.

I2RS The IETF is currently working in a standardized interface to the routing system of network devices denominated I2RS [Hares and White 2013]. I2RS should provide the necessary interface for a centralized controller to fetch all paths received by edge devices from external ASes, including information of the state of the path in the RIB and FIB of the routers. The time of availability of such technology is however not yet known.

2.2. Traffic data

For inter-domain TE purposes, an ISP should be able to monitor the traffic per prefix at every interface connecting the network to external ASes. More granular information, such as traffic per application or transport protocol, can also be very useful in specific cases. Netflow and sFlow are the two most popular technologies for collecting this data [Claise 2004]. The traffic data generated by ISP networks can be quite large. To cope with this overhead, operators can decide to focus only on the traffic observed in peak networking hour times. Likewise, operators could simplify many traffic management techniques by considering the prefixes that drive most of the network traffic [Feamster et al. 2003].

2.3. External AS policies

Every Autonomous system in the Internet has the freedom, under certain constrains, of filtering, redistributing, and altering the paths it receives from neighbor ASes. As we will describe in Section 3, for some advanced inter-domain applications, ISPs might need to evaluate the impact of the policies of external ASes in their networks. The estimation of external polices is extremely complex, but, some data can be used to obtain a reasonable estimation of them. We summarize some of them in the following:

AS relationships The knowledge of the relationships among the AS conforming the close AS neighborhood of an ISP can be useful for different inter-domain traffic management applications (Sections 3.2 and 3.3). The estimation of the relationship is difficult, and requires a high level of analysis of BGP data

[Luckie et al. 2013]. An ISP can obtain this information through some commercial companies, through public data [Luckie et al. 2013], or thanks to social interaction, since operators usually know the relationships of many of their peers.

External BGP data Some public Internet projects, such as route views or RIPE, release the BGP data of devices located in different points of the Internet [Meyer et al. 2005][RIPE 2013]. This information can be used to obtain a partial view of the policies of external AS. For instance, it can help determine if some ASes filter prefixes or modify their path attributes [Lutu et al. 2014][Donnet and Bonaventure 2008].

The described data is a good starting point for the analysis of policies of external ASes. Nonetheless, ISPs must understand that it is almost impossible to obtain a complete view of the policies of the ASes in the Internet. Hence, this data can include some inaccurate information, and operators must consider this fact at the time of the analysis. In general, network operators still require tools for the conversion of large amounts of data (routing, traffic, etc.) into exploitable data that can be easily understood by their management, design, and architecture teams.

2.4. Network infrastructure and Shared Risk Link Groups

An inventory of the physical devices of the network is useful for troubleshooting and network management. For network design, a summary of the physical location and the groups of devices that can fail simultaneously, also referred to as Shared Risk Link Group (SRLG), can provide engineers with information necessary for network failure analysis. The automatic discovery of SRLGs has been extensively explored in optical networks [Sebos et al. 2001]. For the case of IP networks, operators must still rely on proprietary applications or protocols to perform network inventory and SRLG knowledge construction.

2.5. Corporate BGP Policy

Each autonomous system decides how to manage the routing information received from external ASes. The *routing policy* of each AS defines what to do with the paths it receives from their neighboring ASes: to which other ASes it can propagate the routes, how to modify the paths, whether to filter them or not, etc. This policy is reflected in the configuration of edge devices. Maintaining the policy information not only at the edge devices, but also in a centralized location can help operators to match which network states and events conflict with the local policy. New network applications and devices are increasingly supporting protocols such as NETCONF/YANG [Bjorklund 2010], which provide a flexible interface for network configuration. These protocols could be leveraged to build centralized network management solutions, which would allow operators to maintain policy related information in a single database. We describe examples of applications that would benefit from local policy data in Sections 3 and 4.

2.6. Path performance details

Content provider networks or other ASes that offer real-time applications might need to obtain data on the performance characteristics of the paths received by their neighboring ASes. Delay, packet loss, or other performance information can be used by these companies to select the path that their packets should take to reach an end user. This

type of companies might prefer good performance paths over paths traversing cheaper links. This can push them, for example, to send packets through a transit provider, instead of a peering link, when the latter suffers from high delay or packet loss due to link congestion [Savage et al. 1999][Duffield et al. 2007]. The collection of performance data requires the measurement of packet statistics, which can be obtained using probes, or taken directly from user applications. Operators still face the challenge of correlating this data with control-plane information (BGP paths). These can be facilitated if the respective data is stored in information systems with flexible and standard interfaces [Lucente and Jasinska 2014].

3. Applications

In the previous section, we described different sources of data that can be useful for inter-domain traffic management and how network operators could obtain them. In this section, we describe some applications that can be implemented using this data.

3.1. Traffic control and load balancing

The control of network traffic is one of the main goals of traffic engineering applications. For the inter-domain case, the strategies deployed to control outbound traffic differ from the ones used for inbound traffic:

Outbound traffic An autonomous system can decide, from all available paths, the ones that the network can use for sending their traffic. Transit costs are normally the first characteristic used for path selection. However, operators might decide to select paths based on performance metrics, such as delay or packet loss, which are usually not reflected in the BGP attributes of the paths (see Section 2.6). For a better path selection, it is beneficial to have large path diversity. In Section 2.1, we discussed about the problem of path-hiding that can occur within a network that does not implement methods improving the diversity of disseminated path. Another path-hiding problem occurs in Internet Exchange Points using Route Servers [Jasinska et al. 2014]. Similar to any BGP speaker, standard IXP route servers cannot send more than one path to their peers. Even if a route server receives different paths for the same destination, it will select only one to send to each client. Since the policies of companies differ, the route server might not select the path that a client had chosen if it had received all paths. The IXP could implement path diversity techniques, such as diversity paths [Raszuk et al. 2012] or ADD-PATH [Francois et al. 2014], to counter this problem.

Inbound traffic Since each AS is free to choose the paths it uses for its outbound traffic, an ISP can only try to influence the path selection of external ASes. Operators can use different techniques for this end, such as selective advertisement, path prepending, or MED tuning [Feamster et al. 2003]. Network operators must understand that other ASes might implement policies that limit the efficiency of their inbound TE strategies. Operators could adjust their inbound traffic engineering using information of the policies of external ASes (Section 2.3). For instance, operators could estimate how effective would be to announce a prefix through a single peer, by detecting the external ASes that would tend to select a path via this AS.

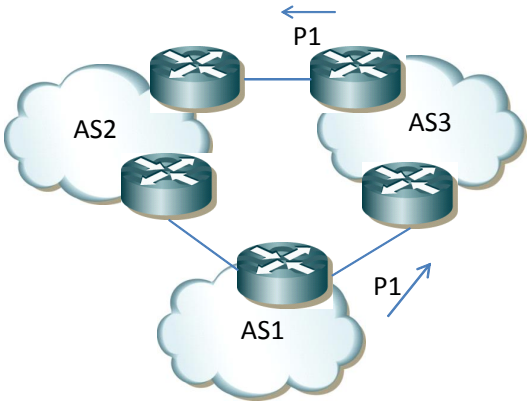

Figure 1. Interconnections among three ASes.

3.2. Network failure Analysis

Network operators are not always aware of the traffic distribution that would occur in a network after failure events. By using the data described in Section 2, operators can study the impact of failures of single hardware elements. If the ISP has safe knowledge of groups of hardware that could fail simultaneously (Section 2.4), a more complete analysis of network failure can be performed [Kiese et al. 2009]. For Outbound traffic, network operators might need to simulate the BGP decision process. For inbound traffic, an AS should also consider the relationships and policies of external ASes to improve their traffic estimations. Since the simulation of the network under different failures is challenging, ISPs should use historical data to assess the quality of their estimations. Operators could obtain this data, for instance, by storing routing and traffic data when the network is under maintenance.

3.3. External policy impacts

The Internet routing table is the result of the interplay of the policies of all autonomous systems. Some ASes might have conflicting business interests, which would be projected into their routing policies. We illustrate an example of conflicting business interests among ASes in Figure 1. The figure depicts an AS (AS1) with two transit providers (AS2 and AS3). AS1 might prefer to exchange the traffic for prefix P1 through only one of them (AS3). Thanks to the policy of AS1, AS2 would need to route the packets heading for P1 to AS3, even if it has a direct connection with AS1.

Conflicting business interests among ASes do not have a simple solution that satisfies all involved ASes. By analyzing the internal and external data, ASes can better understand the conflicts and then proceed to deal with some of these situations, by, in most cases, contacting directly the other network operators playing a role in the situation. In our example, AS2 could use BGP and policy data to detect that P1 belongs to one of its customers (AS1), but that it is sending traffic to it via one of its peer (AS3). AS2 could then query traffic data to measure the impact of this specific configuration from AS1 to its inter-domain traffic policy. Depending of the level of traffic to P1, management personnel from AS2 could attempt to convince their counterparts in AS1 to announce P1 to them. The Internet is full of cases as the one just described. Other more complex policy

conflicts might lead to cases in which ASes are unexpectedly harmed. A scenario where this situation occurs is described in Section 4.

4. Case Study

In this section, we illustrate a potential motivation for overlapping prefix filtering and how this practice could create unexpected traffic flows in other, distant, ASes. We then discuss how by correlating different sources of data, ASes can detect and manage these cases.

Figure 2 depicts the scenario used for this case study. AS1 is a customer of AS2 and AS3. AS4 is the transit provider of AS5, AS2 and AS3. AS5 and AS2 are establishing a free-settlement peering. AS1 is announcing a covering prefix (10.0.0.0/22) to both its providers. Additionally, AS1 announces another prefix (10.0.0.0/24), overlapping the first one, to AS3 only. AS2 and AS3 announce the prefix 10.0.0.0/22 to their peers and transit providers. The overlapping prefix is announced by AS3 to AS4. AS5 receives the overlapping prefix only by its transit provider (AS4).

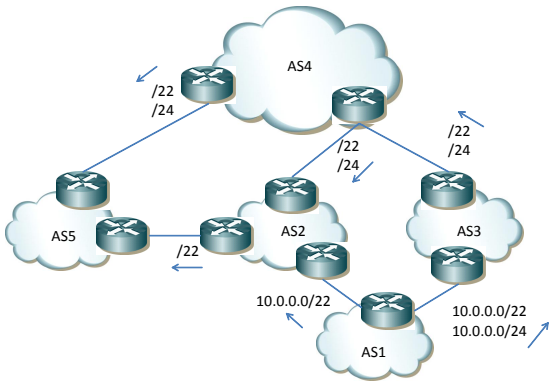

Figure 2. Case study scenario.

AS5 finds itself in a situation in which traffic flowing to prefix 10.0.0.0/24 is sent to its transit provider, even when it receives a valid path to this range of hosts via a peer (AS2). It does not matter how much preference AS4 places to the path through AS2, as its routers would always choose the path to the more specific prefix [Baker 1995]. AS5 could identify this issue by analyzing their data. For instance, they could use their internal BGP data and find cases in which an overlapping prefix is not advertised by a peer and received through a transit provider. They could also use their traffic data to detect traffic heading to hosts covered by a prefix received from a peer that is actually sent to a transit provider [Lucente 2011]. AS5 could then decide to filter the incoming overlapping prefix (10.0.0.0/22) from AS4[1]. Subsequently, traffic from AS5 to prefix 10.0.0.0/24 would be forwarded towards AS2.

After AS5 filters the route to 10.0.0.0/24, an inconvenient situation occurs for AS2. Due to the existence of a route to prefix 10.0.0.0/24 through AS4, AS2 receives the traffic heading to this prefix from AS5 and sends it to AS4. From the perspective of AS2, a traffic flow between a peer (AS5) and a transit provider (AS4) is created. This unexpected traffic flow contradicts AS2's policy [Cardona et al. 2014]. AS2 should be

[1] An automatic approach for this situation has been proposed in [White and Retana 2013].

able to detect the unexpected flows correlating their local policy data (i.e. which links belong to providers, peers, and customers) with traffic flow data. AS2 could attempt to pinpoint the exact source of the problem by using external BGP data. AS2, for instance, could use this data to find that AS5 is also a customer of AS4 and then infer that AS5 is filtering prefix 10.0.0.0/24.

There might be no feasible solution that satisfies all actors in this scenario. AS2 can decide to stop announcing the covering prefix (10.0.0.0/22) at the peering session with AS5. If this happens, all traffic heading to the prefix 10.0.0.0/22 from AS5 would no longer traverse AS2. However, AS2 would lose the traffic share that is not covered by the affected overlapping prefix. Alternatively, A2 could decide to filter the overlapping prefix 10.0.0.0/24 from the session with AS4. As a result, the traffic destined to 10.0.0.0/24 would be forwarded by AS2 directly through its link with AS1, despite the actions performed by AS1 to have this traffic coming in through its link with AS3. However, as AS2 will no longer possess a route to the overlapping prefix, it risks losing the traffic share from customers different from AS1 to that prefix. Furthermore, this action can generate other types of conflicts between AS2 and AS1, since AS2 does not follow the policy expressed by AS1 in its BGP announcements.

As aforementioned, any of these solutions could be considered the wrong one. Network operators and peering coordinators from AS2 and AS4 should *understand the situation and its impact* into their networks using different sources of data. They can then decide which solutions is the best for their case.

5. Conclusions

In the last ten years, the research community has studied different techniques to perform better inter-domain TE. The data required to implement these methods, however, might not be always available to service providers. In the near future, thanks to the data-driven environment born from SDN-like applications, many operators could be able to perform them. We revisited some of these applications and described the data required for them. Additionally, we provide a case study for an inter-domain TE application, which detects unexpected traffic flows created by conflicts in business interests and illustrate the difficulty of finding solutions for these scenarios.

6. References

References

Agrawal, D., Das, S., and El Abbadi, A. (2011). Big data and cloud computing: current state and future opportunities. In *Proceedings of the 14th International Conference on Extending Database Technology*, pages 530–533. ACM.

Awduche, D., Chiu, A., Elwalid, A., Widjaja, I., and Xiao, X. (2002). Overview and principles of internet traffic engineering.

Baker, F. (1995). Requirements for ip version 4 routers. *IETF RFC 1812*.

Bjorklund, M. (2010). Yang-a data modeling language for the network configuration protocol (netconf). *RFC 6020*.

Cardona, J. C., Francois, P., and Lucente, P. (2014). Making bgp filtering a habit: Impact on policies. *draft-ietf-grow-filtering-threats-02. Work in Progress. IETF Draft.*

Cardona, J. C., Francois, P., Ray, S., Patel, K., Lucente, P., and Mohapatra, P. (2013). Bgp path marking. *draft-bgp-path-marking-00. Work in Progress. IETF Draft.*

Casado, M., Koponen, T., Shenker, S., and Tootoonchian, A. (2012). Fabric: a retrospective on evolving sdn. In *Proceedings of the first workshop on Hot topics in software defined networks*, pages 85–90. ACM.

Claise, B. (2004). Cisco systems netflow services export version 9. *IETF RFC 3954*.

Donnet, B. and Bonaventure, O. (2008). On bgp communities. *ACM SIGCOMM Computer Communication Review*, 38(2):55–59.

Duffield, N., Gopalan, K., Hines, M. R., Shaikh, A., and Van Der Merwe, J. E. (2007). Measurement informed route selection. *Passive and Active Network Measurement*, pages 250–254.

Feamster, N., Borkenhagen, J., and Rexford, J. (2003). Guidelines for interdomain traffic engineering. *ACM SIGCOMM Computer Communication Review*, 33(5):19–30.

Francois, P., Cardona, J. C.and Simpson, A., and Haas, J. (2014). Add-path for route servers. *draft-francois-idr-rs-addpaths-00. Work in Progress. IETF Draft.*

Hares, S. and White, R. (2013). Software-defined networks and the interface to the routing system (i2rs). *IEEE Internet Computing*, 17(4).

Jasinska, E., Hilliard, N., Raszuk, R., and Bakker, N. (2014). Internet exchange route server. *draft-ietf-idr-ix-bgp-route-server-04. Work in Progress. IETF Draft.*

Kiese, M., Marcheva, V., Eberspacher, J., and Schupke, D. (2009). Diverse routing based on shared risk link groups. In *Design of Reliable Communication Networks, 2009. DRCN 2009. 7th International Workshop on*, pages 153–159. IEEE.

Lucente, P. (2011). Detecting routing violations. `http://wiki.pmacct.net/DetectingRoutingViolations`.

Lucente, P. and Jasinska, E. (2014). Netflow and bgp multi-path: quo vadis? `https://www.nanog.org/sites/default/files/monday_general_lucente_netflow_32.pdf`.

Luckie, M., Huffaker, B., Dhamdhere, A., Giotsas, V., et al. (2013). As relationships, customer cones, and validation. In *Proceedings of the 2013 conference on Internet measurement conference*, pages 243–256. ACM.

Lutu, A., Bagnulo, M., Cid-Sueiro, J., and Maennel, O. (2014). Separating wheat from chaff: Winnowing unintended prefixes using machine learning. *Proceedings of 33rd IEEE International Conference on Computer Communications, IEEE INFOCOM.*

Marques, P., Fernando, R., Mohapatra, P., Gredler, H., and Chen, E. (2012). Advertisement of the best external route in bgp. *draft-ietf-idr-best-external-05. Work in Progress. IETF Draft.*

Meyer, D. et al. (2005). University of oregon route views project.

Raszuk, R., Fernando, R., Patel, K., McPherson, D., and Kumaki, K. (2012). Distribution of diverse bgp paths. *IETF RFC 6774*.

Rekhter, Y., Li, T., and Hares, S. (2006). Border gateway protocol 4. *IETF RFC 4271*.

RIPE (2013). Ripe ncc ris.

Savage, S., Collins, A., Hoffman, E., Snell, J., and Anderson, T. (1999). The end-to-end effects of internet path selection. In *ACM SIGCOMM Computer Communication Review*, volume 29, pages 289–299. ACM.

Scudder, J., Fernando, R., and Stuart, S. (2012). Bgp monitoring protocol. *draft-ietf-grow-bmp-07. Work in Progress. IETF Draft.*

Sebos, P., Yates, J., Hjalmtysson, G., and Greenberg, A. (2001). Auto-discovery of shared risk link groups. *Optical Fiber Communication Conference*, 4.

Stallings, W. (1998). *SNMP, SNMPv2, SNMPv3, and RMON 1 and 2*. Addison-Wesley Longman Publishing Co., Inc.

Walton, D., Retana, A., Scudder, J., and Chen, E. (2012). Advertisement of multiple paths in bgp. *draft-ietf-idr-add-paths-08. Work in Progress. IETF Draft.*

White, R. and Retana, A. (2013). Filtering of overlapping routes. *draft-white-grow-overlapping-routes-02. Work in Progress. IETF Draft.*