Inter-Domain Route Diversity for the Internet

Xavier Misseri[1] misseri@telecom-paristech.fr,
Ivan Gojmerac[2] gojmerac@ftw.at
and Jean-Louis Rougier[1] rougier@telecom-paristech.fr

[1] TÉLÉCOM ParisTech, 46 Rue Barrault, 75013 Paris, France
[2] FTW – Telecommunications Research Center Vienna, Donau-City-Str. 1, 1220 Vienna, Austria

Abstract. The current inter-domain routing in the Internet, which is based on BGP-4, does not allow for the use of multiple paths, but rather restricts the routing to a single path for each destination prefix. This fact is especially unfortunate considering the vast route diversity which is inherently present in the global Internet graph. Therefore, we propose Inter-Domain Route Diversity (IDRD) as an overlay mechanism which enables efficient, backwards compatible and incrementally deployable introduction of route diversity in the Internet. Beyond presenting the architecture of IDRD, this paper also presents the conditions which ensure the stability of the proposed mechanism as a fundamental prerequisite for its deployment in real-world scenarios.

Keywords: Inter-domain routing, BGP, Path diversity, IDRD, Map-and-Encap

1 Introduction and Related Work

In order for the Internet to function as a set of individual networks belonging to different administrative domains, a mechanism is required which will provide for the global exchange of routing information. This role is currently fulfilled by the Border Gateway Protocol (BGP) [12], which propagates IP prefix reachability information by exchanging so-called *path vectors* between neighbor networks. In order to achieve scalability in such an approach, each domain selects only a single route (i.e., *path*) per IP prefix, and accordingly, only the selected path is propagated to the neighbors. Thereby, the inherent Internet-scale route diversity is not put into use, as effectively only single-path routes towards any global subnetwork are enabled. This impedes multi-path gains even at the level of Tier-1 networks, in spite of the vast diversity they are exposed to (cf. [15]).

Additionally, BGP route selection represents a stumbling block for the efficient and flexible operation of individual domains, as the *BGP decision process* (which determines the next hop neighbor network for each global prefix) is comprised of a number of successive, hard-coded and static rules based on comparisons of global or local path attributes, such as the *LOCAL_PREF*, *AS Path* length or the *MED* (Multi-Exit Discriminator), eventually always ending in a tie-break which determines the single path used.

Nevertheless, BGP does encompass some potential for traffic engineering, with several techniques having been proposed in literature. For instance, [11] proposes Local-Preference tweaking mechanisms in order for multihomed ASes to control their outbound traffic sent towards the different providers. However, BGP traffic engineering techniques are quite limited and coarse grained, as extensively laid out in [2].

On a similar note, some work has been performed in the field of inter-domain path diversity, however, a simple and effective solution for putting the multiple viable paths into use has not yet been formulated.

In [14], BGP add-path is presented as an extension of BGP that allows for the announcement of several paths per prefix, but which at the same time does not change the BGP decision process, effectively resulting in a good technique for fast failover in combination with BGP Route Reflectors (RRs).

In order to enable simultaneous use of multiple routes, the *Multipath BGP* extension [1, 7] slightly changes the BGP decision process. While *Multipath BGP* allows for balancing traffic in equal shares among the available paths, it does not propagate path diversity to its neighbors, and furthermore it is subject to tough constraints, as the different routes must be very similar concerning their metrics (i.e., Local-Preference, AS path length and MED).

[17] proposes Multi-path Interdomain ROuting (MIRO) as an architecture for the selection, export and enforcement of alternative inter-domain routes. Whereas MIRO does address some problems discussed in the present paper, at the same time it does not allow a router to utilize several paths simultaneously. Furthermore, [17] does not provide a stability analysis for the proposed scheme, which we consider to be a prerequisite for any practical deployment. And concerning message exchange, MIRO specifies that routes are *pulled* and not *pushed*; while we do understand this paradigm in the context of scalability issues, we believe that the solution should rather lie in the selection of routes which are to be pushed in the first place, than in abandoning the push concept altogether.

Last, but not least, [16] proposes D-BGP and B-BGP as two interdomain routing proposals that propagate path diversity. The goal of these proposals is to speed-up the recovery of BGP by propagating a maximally disjoint alternative path associated to the BGP best route. The propagated multi-paths are however only used for back-up purposes, and the proposed changes are supposed to be integrated into BGP, which we do not find to be realistic.

In the next section, we will advance the current state of the art by introducing Inter-Domain Route Diversity (IDRD) which allows ISPs to propagate more routes towards their customers, providers and peers. Moreover, we propose to relax the constraints on the BGP selection process, thereby mainly focusing on the *prefer client routes* conditions. Overall, the propagation of the multiple routes faces several challenges, foremostly in assuring the stability of the control plane. Indeed, the use and the propagation of diversity has substantial significance only if the domains can select policies that are different from the BGP decision process. E.g., a domain should be able to propagate routes that do not have the highest *LOCAL_PREF*. However, we recognize that the Gao & Rex-

ford conditions from [3], which ensure stability in the current Internet, are not sufficient in this case, but that additional mechanisms are required for ensuring the stability of the propagated diversity.

The rest of this paper is structured as follows: Section 2 presents the architecture of IDRD as our proposed solution, thereby equally focusing on control plane and data plane aspects, as well as on the prospective use cases, architecture deployability and its backward compatibility. Subsequently, Section 3 provides an analysis of IDRD stability before Section 4 concludes the paper with a summary of the main results.

2 Architecture Proposal for Route Diversity

In this section, we present Inter-Domain Route Diversity (IDRD), which enables the use of the inherent topological diversity present in today's Internet. More specifically, we describe the IDRD control and data planes, followed by a discussion of the most relevant IDRD use cases. Finally, we provide strong arguments for the real-world deployability of our solution due to several key properties of IDRD.

Before presenting IDRD in the next subsections, here we aim at providing clarity on the character of this solution, i.e., we wish to stress that IDRD by no means aims at replacing BGP. Instead, we see IDRD as an add-on to the present Internet, which can be deployed by some domains in parallel to BGP as a higher-layer overlay.

In order to make the propagation (control plane) and the use (data plane) of diversity possible, we base our architecture on the *map-and-encap* paradigm, which allows the traffic to be encapsulated according to the parameters provided by a Mapping System (cf. [9] for an existing example). The Mapping System (MS) can either be internal or external with respect to the encapsulating routers, and structurally it can either follow a centralized or a decentralized logic.

2.1 IDRD Control Plane

With IDRD, each domain stores the information on path diversity within its own Mapping System (MS). As a domain that has adopted IDRD may be connected to domains that have not adopted this architecture, the routing information coming from these neighboring domains (in the form of eBGP updates) can be redistributed into the MS. Conventional BGP is thus a source of diversity in such a case.

The propagation of multiple paths is performed at the MS-level for neighbors that have adopted the architecture, i.e., their MSes communicate directly in order to provide route diversity. This diversity information contains BGP metrics and may also contain other metrics, e.g., price, capacity, etc. Once the set of multiple paths has been received (either via BGP redistribution or via inter-MS communication), the domain can select a subset of those paths which it finds interesting. The MS can compute advanced selection policies based on price, stability, political relationships, etc.

Once a set of paths is selected, it is propagated by the domain to its neighbors. In this context we note that the paths selected by a domain are not necessarily

all put into use. Instead, the selected set of paths represents an assurance that neighbor domains can utilize them using the map-and-encap scheme.

The proposed approach deeply relaxes the 'prefer client' constraint of BGP due to the multiple potential paths which can be conveyed and used. However, it is important to note that all traffic which is received outside of the map-and-encap scheme (i.e., all *conventional* traffic) will still be forwarded via the standard BGP best routes for reasons of backward compatibility.

Once the diversity is propagated between ASes, the Autonomous System Border Routers (ASBRs) must be made aware of the corresponding mapping information. The MS can either directly push the mapping or await mapping requests from neighbor ASBRs. Each MS entry provided to an ASBR contains at least the following information:

- The association between the flow identifier and the *next hop* / ASBR,
- The association between incoming and outgoing flow identifiers (cf. Sec. 2.2).

2.2 IDRD Data Plane

Life-cycle of a packet: In order to implement the usage of alternative paths advertised by IDRD in the present Internet, we propose to apply packet encapsulation, similarly to the scheme presented in [6]. There are two different areas of path enforcement which must be taken into account:

- **Intra**-domain path enforcement: When a packet arrives at a domain entrance, the ASBR asks the mapping system about which exit ASBR the packet must be forwarded to. According to its *flow identifier*, the Mapping System (MS) specifies the exit ASBR. The entry ASBR then encapsulates the packet and forwards it to the correct exit ASBR. It is important to note that the encapsulation scheme is local and that it has got no impact on neighboring domains. Therefore, each AS can individually choose its encapsulation scheme (e.g., IPv4, MPLS, etc.). Finally, once the packet arrives at the exit ASBR, it gets decapsulated.
- **Inter**-domain path enforcement: When arriving at the exit ASBR, a packet gets encapsulated in order to enforce its path towards the next domain's ASBR. As in the case of intra-domain path enforcement, the mapping resolution can be either pushed or pulled. But in contrast to the intra-domain case, the inter-domain encapsulation scheme must be negotiated between neighboring ASes in order to be inter-operable.

Flow identification: The ASBRs must forward packets from one tunnel to next. As several paths are available in order to reach a destination IP prefix, the destination IP address in the inner IP header is no longer sufficient for making the forwarding decision. Therefore, in addition to the inner IP destination address (the real destination host), an identifier can be used to specify the route which is to be used. In order to be scalable, this identifier must be assigned and used locally in each domain or at each peering/transit link. Inter-AS and intra-AS identifiers of the same path must be aligned in order to be able to choose a coherent path. ASBRs must then be able to swap incoming identifiers with outgoing identifiers.

2.3 IDRD Use Cases

The use cases for inter-domain route/path diversity are well-known and we only briefly enumerate them here. Firstly, route diversity has the potential to increase the overall network capacity between two points in the Internet, i.e., it can be used for traffic engineering and load balancing (cf. [2, 8]). Secondly, having available a set of multiple disjoint paths can also be used for increasing resilience (cf. [18]). Both mentioned benefits could potentially also apply to end customers who employ layer-4 path diversity schemes, like e.g. MP-TCP [5]. And finally, flexible and explicit route enforcement represents an important tool for inter-domain Quality of Service (QoS) mechanisms, which will substantially gain importance if large-scale capacity overprovisioning in the Internet becomes unfeasible (cf. [**?**]).

2.4 Backward Compatibility and Deployability of IDRD

The design of IDRD repsects both successful protocol design requirements postulated by C. Dovrolis in [13]. Firstly, our architecture is *backward compatible*, as it is incrementaly deployable among only a subset of Internet Service Providers (ISPs) due to its seamless compatibility with the current Internet. Secondly, IDRD is *incrementally deployable* in the sense that it brings benefits to its early adopters even if not broadly deployed. Therefore, we believe that IDRD has the potential to achieve practical relevance in the mid-term future.

3 Discussion of IDRD Stability

3.1 Instability Example

Enabling the propagation of path diversity in the current Internet may lead to oscillations. Figures 1 and 2 provide a simple but stunning example for the instabilities which might occur: AS_A, AS_B and AS_C are peers, and AS_D is the client of the other three ASes. In order to be able to reach AS_D via multiple paths, each AS selects a second route according to local policies, in addition to the first (i.e., direct) BGP best route. Each AS uses a local decision process and ranks the potential paths for the alternative route choice. In our example, AS_A orders the alternatives by priority as ACD, AD, ABD. Figure 2 presents the stepwise change of path selection in each AS. Thereby, the selections which have changed since the last step are highlited in red, the selections which are being propagated to neighbors are underlined, and paths which are being unselected (withdrawn) are ~~crossed out~~. According to the previously listed priority list, if AS_A receives the path CD from C (as in Line 3 of Figure 2), it chooses the path ACD and withdraws the path AD. And concurrently, AS_A sends the withdrawal of AD to its neighbors, however, it does not propagate the newly selected path ACD to its peers (due to adherence to the Gao & Rexford conditions [3], as discussed in Section 3.2).

We can see from Figure 2 that Steps 3 and 9 are identical, which implies that the system will enter into oscillations (cf. also [4] for further examples of instabilities in inter-domain routing). The next subsection underlines the sufficient conditions to reliably avoid oscillations and it provides a pointer to a mathematical proof of IDRD stability.

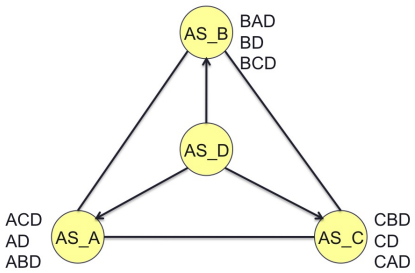

Fig. 1. Example topology.

Step	AS_A	AS_B	AS_C
1	AD	BD	CD
2	ACD(~~AD~~)	BD	CD
3	ACD	BD	CD
4	ACD	BD	CBD(~~CD~~)
5	AD	BD	CBD
6	AD	BAD(~~BD~~)	CBD
7	AD	BAD	CD
8	ACD(~~AD~~)	BAD	CD
9	ACD	BD	CD

Oscillation !!!

Fig. 2. Evolution of the routing decisions.

3.2 IDRD Stability Conditions

In the above example we have highlighted that the propagation of multiple routes can lead to oscillations. Therefore, in this part we present a criterion which – in addition to the the Gao & Rexford (G & R) conditions [3] – ensures the stability of IDRD. In the IDRD architecture, a domain d receives:

- A set of routes coming from clients (i.e., client diversity): $R_{d,c}$ (where d denotes a *domain* and c denotes the whole set of *clients*).
- A set of routes coming from peers and providers (i.e., peer/provider diversity): $R_{d,p}$ (where p denotes the whole set of *peers and providers*).

The domain d uses a decision process λ_d to select interesting routes among the two sets of route candidates: $S_d = S_{d,c} \cup S_{d,p} = \lambda_u(R_{d,c} \cup R_{d,p})$ where S_d denotes the set of routes selected by the domain d, $S_{d,c}$ stands for the set of selected client routes, and where $S_{d,p}$ denotes the set of selected peer/provider routes.

In the current BGP-based Internet, G & R policies already apply and they ensure its stability. In terms of IDRD, the well known G & R conditions translate to:

- Each d sends only the set of selected client routes ($S_{d,c}$) to its neighbor peers and/or providers.
- Each d sends all selected routes ($S_{d,c} \cup S_{d,p}$) to client neighbors.

In order to ensure the global stability of IDRD, the following stability criterion introduces a strong requirement in addition to the two G & R conditions stated above.

> IDRD Stability Criterion: Routes received from peers and providers must have no impact on the selection of routes received from clients. More formally, if we have $S_d = S_{d,c} \cup S_{d,p} = \lambda_d(R_{d,c} \cup R_{d,p})$, then $S_{d,c}$ must be independent from $R_{d,p}$.

We can analyse the impact of this criterion on the oscillation example given in Section 3.1. In Steps 2, 4, 6 and 8, the ASes have received a route from a peer, subsequently unselecting the route coming from their client. However, this is strongly prohibited by the stated stability criterion, which ensures that

the ASes select and propagate client routes independently of peer and provider routes. In our example, ASes must thereore select both peer and client routes. Nevertheless, for the traffic originated from within the domain, each AS can still opt for using only one of the advertised routes.

Due to limited space in this paper, for a comprehensive proof of IDRD stability we refer the reader to our technical report in [10]. There we prove that an IDRD system that respects the previously stated criterion and the G & R conditions is safe, meaning that it converges to a stable state from any initial state and that this stable state is unique. We prove this statement in a three step procedure. The first step proves that an IDRD system that respects the previous criterion and the G & R conditions has a stable state, and that this stable state is unique. The second step proves that it reaches the stable state for any initial state. Finally, the third step proves that this stable state is reached within a finite time interval.

4 Conclusions and Future Work

Due to its distributed nature, the global Internet displays vast potential path diversity. However, for stability reasons, BGP-4 as the current inter-domain routing protocol in the Internet does not enable its utilization, as it allows only for a single path towards each destination prefix.

In order to mitigate this restriction, in this paper we propose Inter-Domain Route Diversity (IDRD) which allows for the propagation (control plane) and the use (data plane) of the present Internet path diversity. In order to achieve optimal backward compatibility as well as incremental deployability, we have designed IDRD as an overlay mechanism which can operate in parallel to BGP, and which enables partial deployment at the AS-level. As far as use cases for IDRD are concerned, we identify substantial potential in the areas of traffic engineering and load balancing, which aim at optimal utilization of the available network capacities. Furthermore, we believe that explicit selection of multiple end-to-end paths using IDRD can play an important role in the provisioning of QoS-enabled Internet services, as well as in the improvement of Internet-wide resilience in the presence of anomalies and component failures. After presenting IDRD in detail, we have paid great attention to the issue of *stability*, which we consider to be the most important criterion when introducing any new protocol to the Internet. Accordingly, in Section 3 we provide an in-depth discussion of this topic, accompanied with pointers to our comprehensive proof of IDRD stability.

Concerning future research on IDRD, our work is far from coming to an end. Next, we aim at further detailing the IDRD architecture and devising advanced decision processes for route selection. Furthermore, we also intend to provide a quantitative analysis of the amount of path diversity in the present Internet.

Acknowledgments

This work has received funding from the European Unions's Seventh Framework Program (FP7/2007-2013) under grant agreement 248567 for the ETICS

project (https://www.ict-etics.eu/). FTW is funded within the COMET Program by the Austrian Government and the City of Vienna.

References

1. Cisco Systems: BGP Best Path Selection Algorithm, http://www.cisco.com/en/US/tech/tk365/technologies_tech_note09186a0080094431.shtml
2. Feamster, N., Borkenhagen, J., Rexford, J.: Guidelines for Interdomain Traffic Engineering. ACM SIGCOMM Computer Communication Review 33(5) (2003)
3. Gao, L., Rexford, J.: Stable Internet Routing Without Global Coordination. IEEE/ACM Transactions on Networking 9(6), 307–317 (2000)
4. Griffin, T.G., Shepherd, F.B., Wilfong, G.: The stable paths problem and interdomain routing. IEEE/ACM Transactions on Networking (TON) 10(2), 232–243 (Apr 2002)
5. Handley, M., Raiciu, C., Ford, A., Barre, S., Iyengar, J.: Architectural Guidelines for Multipath TCP Development. IETF RFC 6182 (Draft Standard) (2011)
6. Jiayue, H., Rexford, J.: Toward Internet-Wide Multipath Routing. IEEE Network 22(2), 16–21 (2008)
7. Juniper: Configure BGP to Select Multiple BGP Paths (2012), http://www.juniper.net/techpubs/software/junos/junos53/swconfig53-ipv6/html/ipv6-bgp-config29.html
8. Kostopoulos, A., Warma, H., Leva, T., Heinrich, B., Ford, A., Eggert, L.: Towards Multipath TCP Adoption: Challenges and Opportunities. In: Next Generation Internet (NGI), 2010 6th EURO-NF Conference on. pp. 1–8 (June 2010)
9. Lewis, D., Fuller, V., Farinacci, D., Meyer, D.: Locator/ID Separation Protocol (LISP), IETF Internet Draft (Work in Progress) (January 2012)
10. Misseri, X., Gojmerac, I., Rougier, J.L.: Technical report – Stability of Global Diversity Propagation (2012), http://perso.telecom-paristech.fr/misseri/Files/10-Stability-Diversity-Propagation-Technical-report.pdf
11. Quoitin, B., Pelsser, C., Swinnen, L., Bonaventure, O., Uhlig, S.: Interdomain Traffic Engineering with BGP. IEEE Communications Magazine 41(5), 122–128 (2003)
12. Rekhter, Y., Li, T., Hares, S.: A Border Gateway Protocol 4 (BGP-4). IETF RFC 4271 (Draft Standard) (2006)
13. Rexford, J., Dovrolis, C.: Future Internet Architecture: Clean-Slate Versus Evolutionary Research. Communications of the ACM 53(9), 36–40 (2010)
14. Scudder, J., Retana, A., Walton, D., Chen, E.: Advertisement of Multiple Paths in BGP, IETF Internet Draft (Work in Progress) (September 2011)
15. Uhlig, S., Tandel, S.: Quantifying the BGP Routes Diversity Inside a Tier-1 Network. Lecture Notes in Computer Science 3976, 1002–1013 (2006)
16. Wang, F., Gao, L.: Path Diversity Aware Interdomain Routing. In: IEEE INFOCOM 2009. pp. 307–315 (April 2009)
17. Xu, W., Rexford, J.: MIRO: Multi-path Interdomain ROuting. In: Proc. ACM SIGCOMM '06. pp. 171–182 (September 2006)
18. Yannuzzi, M., Masip-Bruin, X., Sanchez, S., Domingo-Pascual, J., Orda, A., Sprintson, A.: On the Challenges of Establishing Disjoint QoS IP/MPLS Paths Across Multiple Domains. IEEE Communications Magazine 44(12), 60–66 (2006)