Scalability of iBGP Path Diversity Concepts

Uli Bornhauser, Peter Martini, Martin Horneffer

Scalability of iBGP Path Diversity Concepts

Uli Bornhauser[1], Peter Martini[1], and Martin Horneffer[2]

[1] University of Bonn – Institute of Computer Science 4
Roemerstr. 164 – D - 53117 Bonn – ub@cs.uni-bonn.de, martini@cs.uni-bonn.de
[2] Deutsche Telekom Netzproduktion GmbH - Fixed Mobile Engineering Deutschland
Hammer Str. 216 – D - 48153 Muenster – Martin.Horneffer@telekom.de

Abstract. Improving the path diversity seems to be the next fundamental step in the iBGP evolution. Focusing the advantages an improvement of the path diversity implies, network protocol designers have disregarded the most critical drawback so far: The effect on the *scalability* of the iBGP routing, a fundamental requirement for production usage. This aspect is examined by the analyses discussed in our paper.
In this paper, we provide the theoretical groundwork for scalability analyses of four highly relevant path diversity schemes. Based on this groundwork, we exemplarily predict the information load the schemes induce in a system of a large ISP. Generalizing the system-specific results, we give an outlook on the load that can be expected in comparable ASs. We found that for two schemes currently in the standardization process, scalability problems in large ASs as they are operated by ISPs seem likely.

1 Introduction

From a global perspective, the Internet we know today is a network of interconnected Autonomous Systems (ASs). Global connectivity across these systems is realized by means of the Border Gateway Protocol (BGP) [1]. BGP distinguishes two operational modes, representing its basic tasks: *External BGP* (eBGP) terms the inter-AS mode of BGP, used to exchange routing information between ASs. To spread this information within ASs, *internal BGP* (iBGP), the intra-AS mode, is used. For the following discussions, being familiar with BGP [1] is most helpful.

1.1 Motivation

Today, iBGP is the de-facto standard to spread global routing information in ASs that route default-free. Since ASs may cover up to thousand routers, maybe even more, scalability is an important aspect. As a result of this fact, large ASs usually implement iBGP via Route Reflection [2] or AS Confederations [3]. However, the information reduction these architectures induce causes significant disadvantages in real life: Suboptimal or inconsistent routing decisions may be forced [4], routing processes may behave undesirable [5], and convergence may be slowed down [6]. A clear improvement with respect to these aspects can be achieved if the information exchange scheme improves *path diversity*: Instead of advertising

only the best path for every address prefix, cf. [1], routers provide information on several known paths according to a certain scheme.

An improvement of path diversity seems to be the next important step in the iBGP evolution [7]. The path diversity is determined by the amount and kind of information speakers advertise to their peers. A classical tradeoff is given by the fact that providing more routing information principally improves the system behavior while it worsens scalability. However, even if the latter aspect is of high relevance in practice, proposals for new iBGP schemes only focus on the former aspect. Scalability issues are not adequately studied. This motivated our research.

1.2 Related Work

BGP as used today was specified as a consequence of scalability problems caused by class-based IP routing [8] in the mid 90ies. As scalability was a critical aspect at this time, the observed and expected behavior of BGP was thoroughly studied and documented [9, 10]. In essence, it followed "that BGP should have no scaling problems in the area of link bandwidth and router CPU utilization". However, concerns rise with respect to iBGP, as its full-mesh design ties up many resources. This deficiency was remedied by Route Reflection and AS Confederations [2, 3].

In the following years, several studies analyzed the expectable growth of the global routing table and average update rates per path [11, 12]. But even if they led to highly important results, conclusions on iBGP were never drawn: As routers performed well in practice and an abrupt significant load growth did not have to be expected, this did not seem to be necessary. In 2002, the situation changed drastically with the publication of the first draft of *Add-path* [13]. Applying Add-path, routers may provide their internal peers with clearly more information than common iBGP allows. The ability to advertise several paths per prefix in parallel led to the development of several path diversity schemes [14–17]: schemes, that would cause a significant load growth. Analyses for the Route Server architecture showed that scalability can be evaluated efficiently [16]. However, even if several path diversity schemes are already in the standardization process [14, 15], their effect on the scalability is still unknown. This is the starting point for our analyses.

1.3 Main Contribution and Paper Outline

Scalability is an essential linchpin for using an iBGP scheme productively. Nevertheless, the impact of increasing path diversity according to the proposed schemes is not well understood yet. Closing this gap is the main contribution of our paper. We focus on BGP Route Reflection, because this is the most common iBGP architecture in large ASs.

The rest of the paper is organized as follows: At first, in section 2, we give a basic outline on iBGP path diversity. After that, in section 3, we provide the theoretical groundwork to estimate the information load on routers in arbitrary ASs. In section 4, we exemplary evaluate the information load path diversity schemes induce in a large AS. In section 5, we generalize the system-specific basic results. Finally, in section 6, we close with a short conclusion and sketch future work.

2 About Add-path and iBGP Path Diversity

A core property of the today's iBGP routing information exchange is that routers only advertise their best path for a prefix. Since this limitation has proven to be a fundamental restriction in practice, Walton et al. started 2002 to specify a draft that relaxes this limitation, cf. [13]. The proposal is well-known as *Add-path*.

2.1 The Add-path Concept

Even if technical aspects are beyond the scope of the paper, basically understanding the Add-path proposal seems helpful. Using classical iBGP, paths are internally keyed by the destination address-prefix they belong to and the peer router they are learned from. Consequently, as shown in figure 1.a), each peer can only provide one path per prefix that is stored and taken into account in the best path selection. As depicted in figure 1.b), the basic idea behind Add-path is to extend the key of a path by another attribute, the *path ID*. This ID is arbitrarily set by the router that announces the path. As known paths are only replaced if they have the same key, Add-path allows routers to advertise several valid paths in parallel.

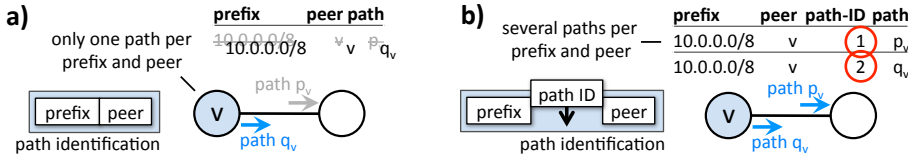

Fig. 1. Using classical BGP, routers can only store one path per prefix and peer speaker (a). The path ID removes this restriction (b).

The Add-path extension defines the semantics of the path ID and the required message exchange formats. However, while this defines the technical precondition for path diversity, Add-path does not specify which paths are to be announced. IBGP *routing information exchange schemes* are not defined. These schemes are specified in separate working documents [14, 15] and iBGP architecture proposals, cf. [16] or [17], for example.

2.2 Routing Information Exchange Schemes

Motivated by different use-case scenarios, protocol designers have defined several different sets of paths routers shall be provided with. Due to the space limitation, we focus our studies on the most reasonable and from the scalability perspective most interesting proposals. This covers the following four concepts:

Advertisement of All Paths The advertisement of all paths is one of the simplest schemes. It claims that routers advertise all known paths to all iBGP peers, provided that the export filters are passed. Usually, only paths are rejected which are already known at the receiver. Further details may be found in [15]. The basic idea is illustrated in figure 2.a).

Advertisement of n Global Neighbor-ASs Group Best Paths Advertising global neighbor-AS group best paths is a more complex, but indeed also more scalable concept. It specifies to advertise the first n best paths with respect to every peer-AS group. Knowing these paths solves certain correctness problems [16]. The concept is illustrated in figure 2.b). See [15] for details.

Advertisement of Optimal and Local Neighbor-AS Group Best Paths Another promising scheme is to provide every router with its optimal path and its AS-group best path for every AS it is neighbored to. Under adequate constraints, this scheme leads to consistent and optimal routing decisions. Even if this scheme is more complex than the second one, it further reduces the information routers are provided with. The idea is shown in figure 2.c). Details may be found in [16].

Advertisement of n Best Paths The advertisement of n best paths generalizes the common iBGP information exchange scheme. Instead of advertising the best path for each prefix, routers advertise the best n paths, cf. figure 2.d), [15].

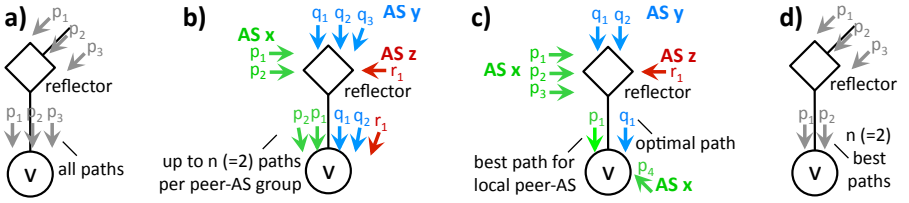

Fig. 2. The number of paths advertised to v may be unlimited (a), limited per prefix and peer-AS group (b), fitted to the receiver's view (c), or limited to n per prefix (d).

3 A Theoretical Framework for Scalability Analyses

As the schemes outlined in section 2.2 and proposed in [14–17] show, path diversity concepts may differ significantly. In this section, we derive an upper bound for the amount of ingoing routing information routers experience if the outlined concepts are applied.

3.1 Basic Model

Large ASs in the default-free zone where scalability is a critical aspect usually use Route Reflection to spread BGP data internally. As shown in figure 3.a), common routers in such architectures receive three different classes of paths: Firstly, there are those paths known due to external announcements. For router v, we refer to these paths as $\mathcal{P}^v_{ext}$. Secondly, paths may be known due to local configuration [1]. For v, we label these paths as $\mathcal{P}^v_{loc}$. The number of paths covered by both sets can be assumed as basically independent of the applied iBGP scheme. Thirdly,

there is the group of internally learned paths advertised by the reflectors, labeled as set $\mathcal{P}^v_{int}$. The elements and size of this set depend on the AS-internally used iBGP scheme. IBGP architectures ensure that all three sets are disjoint [1]. Thus, the *information load* on a common router v using iBGP scheme i is given by

$$|\mathcal{P}^v|_i := |\mathcal{P}^v_{ext}| + |\mathcal{P}^v_{loc}| + |\mathcal{P}^v_{int}|_i =: |\mathcal{P}^v_{ext}| + |\mathcal{P}^v_{\neg ext}|_i. \quad (1)$$

Problematic (CPU and memory) resource requirements can be expected, if the scheme increases the amount of ingoing information drastically in comparison to common iBGP. We derive the expectable load by predicting $|\mathcal{P}^v_{ext}|$ and $|\mathcal{P}^v_{\neg ext}|_i$.

The number of external paths a router maintains is determined by its external peering. For a fixed BGP session to a peer-AS, it receives one path for every prefix the peer-AS advertises. Generalizing this observation to all sessions a router $v \in V$ maintains, cf. figure 3.b), it holds that

$$|\mathcal{P}^v_{ext}| \leq \sum_{x \in peerAS(v)} |prefix(x, V)| \times |session(x, v)|, \quad (2)$$

where V denotes the set of all routers in the AS, $peerAS(\mathcal{X})$ the neighbor ASs of a router or a set of routers $\mathcal{X}$, $prefix(x, V)$ the prefixes routers in AS x advertise into the own AS, and $session(x, \mathcal{X})$ the set of sessions kept between x and $\mathcal{X}$.

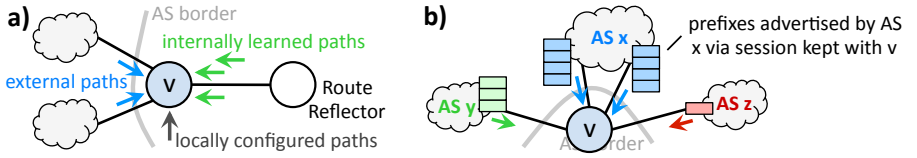

Fig. 3. Routers know paths due to local configuration, external, and internal announcements (a). The number of known external paths depends on the eBGP peer sessions (b).

The set of prefixes locally configured in the AS is called *internal routing table* IT in what follows. The vast majority of AS-internal prefixes is usually imported only once into the AS-internal BGP routing. Thus, for simplicity we assume that $\sum_{u \in V} |\mathcal{P}^u_{loc}| \approx |IT|$. Based on this assumption, we can now estimate $|\mathcal{P}^v_{\neg ext}|_i$.

3.2 Scheme-specific Estimations for Non-external Paths $|\mathcal{P}^v_{\neg ext}|_i$

The number of non-external paths a router keeps depends on the iBGP scheme i used in the system. We now explain how to efficiently predict $|\mathcal{P}^v_{\neg ext}|_i$ for the path diversity proposals sketched in section 2.2. We assume that router v is connected to m redundant Route Reflectors.

Implementing the advertisement of all paths, a reflector passes on all available paths v does not already know. These are all paths external or local on any AS-internal router $u \in V \setminus \{v\}$. Thus, including the own local paths, v knows around

$$|\mathcal{P}^v_{\neg ext}|_1 \leq m \times \Big(|IT| + \big(\sum_{u \in V \setminus \{v\}} \sum_{x \in peerAS(u)} |prefix(x, V)| \times |session(x, u)| \big) \Big) \quad (3)$$

non-external paths. Note that multiplying $|IT|$ by $m > 1$ is not quite accurate, as routers learn their own local paths $\mathcal{P}^v_{loc}$ only once. However, since routers usually do not import more than a few hundreds of paths locally, this simplification can be neglected. We proceed analogously for the following estimations.

The main disadvantage of advertising all paths is that keeping several sessions to one neighbor AS generally bloats the amount of information reflectors provide. This problem is addressed if not more than n paths per prefix for every peer-AS group are advertised (scheme II). For each peer-AS x, Route Reflectors advertise n paths per prefix, provided that n or more eBGP sessions are kept between x and the border routers $u \in V \setminus \{v\}$. Also taking into account local paths, it holds that v knows around

$$|\mathcal{P}^v_{\neg ext}|_2 \leq m \times \Big(|IT| + \sum_{x \in peerAS(V)} |prefix(x, V)| \times \min(n, |session(x, V \setminus \{v\})|) \Big) \quad (4)$$

non-external paths. Recall that $peerAS(V)$ labels all neighbor ASs of the system.

Providing a router per prefix with its optimal path and n best paths for each of its local peer-ASs, the number of received paths becomes topology-dependent. Unknown optimal paths of v must all be provided internally. In the worst case, this covers one path for each globally routable and AS-internal prefix. The former prefixes are well-known and covered by the *global routing table* GT [18]. Latter ones are covered by IT. The unknown AS-group best paths for a router's peer-ASs have to be provided internally, too. For every peer-AS $x \in peerAS(v)$, in the worst case, up to n paths per prefix AS x advertises to v must be provided. Thus, applying this scheme, router v may learn up to

$$|\mathcal{P}^v_{\neg ext}|_3 \leq m \times \Big(|GT| + |IT| + \sum_{x \in peerAS(v)} |prefix(x, V)| \times \min(n, |session(x, V \setminus \{v\})|) \Big) \quad (5)$$

non-external paths via its reflectors and local configuration. Usually, $n = 1$ is the most reasonable configuration, cf. [16]. We assume this in what follows.

Generalizing the classical iBGP scheme, reflectors can simply advertise their n best paths per prefix. Applying this scheme, a reflector advertises the $|IT|$ AS-internal paths and up to n paths for every globally routable address-prefix. This results in up to

$$|\mathcal{P}^v_{\neg ext}|_4 \leq m \times \Big(n \times |GT| + |IT| \Big) \quad (6)$$

non-external paths v receives. If $n = 1$, we estimate the non-external information load the common iBGP information exchange scheme induces.

3.3 Required Performance Data

The methodology developed in the sections 3.1 and 3.2 reveals how system-specific estimations for the information load can be determined. To apply the estimations to a production system, information on the required basic AS-performance data must be available. Understanding how to gather this information properly finally completes the theoretical framework.

As already mentioned, the number of prefixes the global routing table covers, parameter $|GT|$, is well known [18]. The number of system-internal prefixes $|IT|$ should be known by the AS operator. This also holds for the number of reflectors in the system, parameter m, and the desired path diversity factor n many proposals support. Information on the peer-ASs $x \in peerAS(u)$ and the number of kept sessions $session(x, u)$ for all routers $u \in V$ can be derived from the routers' BGP configurations. Based on this information, also $peerAS(\mathcal{X})$ and $session(x, \mathcal{X})$ for all sets of routers $\mathcal{X} \subseteq V$ and peer-ASs x can be determined. The router configurations are usually available for internal use.

The most complex task is to determine the number of prefixes $|prefix(x, V)|$ each peer-AS x advertises paths for into the system. Gathering exact data for a peer-AS x requires analyzing the routing tables of all internal routers that peer with x. Consequently, to gather exact data for all peer-ASs, the routing tables of all internal eBGP speakers would have to be analyzed. This is hardly realizable in large ASs. However, ASs typically advertise paths for nearly the same prefixes to routers located in the same peer-AS. Consequently, a good approximation for $prefix(x, V)$ is usually already given by $prefix(x, \epsilon)$ for any $\epsilon \subseteq V \,|\, x \in peerAS(\epsilon)$. In practice, analyzing the routing tables of a small number of those routers that peer with the most ASs allows determining approximations for a large proportion of the peer-ASs. For the neighbor ASs not covered by this method, the number of advertised prefixes must be estimated. A simple estimation can be determined on the basis of the business relationship that is kept with the system: Provider ASs advertise paths for nearly all prefixes in the global routing table. For other ASs, the approximations derived for peer-ASs keeping the same business relationship can be averaged. A systematic underestimation is avoided at this point, if ϵ is chosen such that the large peer-ASs of each class are covered. They are typically known by the operator. This avoids that scalability problems remain undetected.

4 Performance Evaluation for a Reference System

The basic framework derived in section 3 allows operators to perform AS-specific scalability analyses for their systems. However, basis for such analyses should be a general understanding of the impact of the different schemes on the information load. For that purpose, we apply the framework to a reference system. Based on the achieved results, we identify the generalizable implications in section 5.

4.1 The Reference System AS3320

The reference system we study in this paper is the AS with the registered number 3320, considered at a snapshot in time of August 2009. The system is the Internet Backbone of Deutsche Telekom AG, labeled as AS3320 in what follows. The AS is a comparatively large transit system. At the time the data have been taken, it maintained one customer relationship to a provider AS. All in all, the system peered with 588 neighbor ASs via 1198 sessions, kept by 238 AS-internal eBGP speakers. The relationships to these systems are well-known, too. Besides eBGP

speakers, the system also covered 642 exclusive iBGP speakers. At the point in time of data acquisition, the system used around 63,000 internal prefixes. The global table covered around 300,000 prefixes. Most routers learned internal BGP routing information from two redundant Route Reflectors, meaning that $m = 2$.

4.2 Prefix Advertisement by Neighbor-ASs

To gather information on the number of prefixes peer-ASs advertise, we analyzed the routing tables of the ten routers keeping the most eBGP sessions. The basic results are shown in figure 4: The tables allowed us to approximate $|prefix(x, V)|$ for the only provider and over 70% of the peer ASs. The blue bar chart shows that the provider and, according to the number of advertised prefixes, large peer ASs are connected comparatively often. In addition to provider and peer ASs, we also covered around 20% of the customer ASs. They provide fewer paths and are less often connected. All in all, analyzing the BGP routing tables of ten of 238 eBGP speakers (around 4%) allowed us to cover over 28% of the peer-ASs.

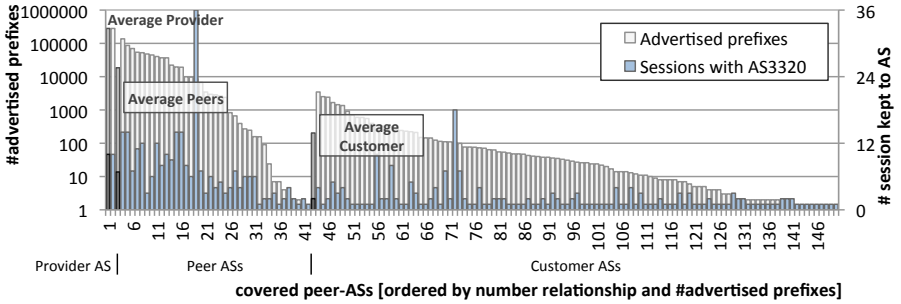

Fig. 4. Approximated number of prefixes neighbor-ASs of AS3320 advertise paths for.

The provider AS advertises paths for around 281,000 prefixes and is connected via ten eBGP sessions. These are around 15 times more prefixes than a peer AS of AS3320 advertises on average (around 18,500). With around six eBGP sessions on average, peer ASs are also significantly less frequently connected than the provider AS. A customer AS in turn only provides paths for around 200 prefixes on average, which is only about 1.1% of an average peer AS. Customer ASs are also clearly less often connected via eBGP sessions: On average, the covered customer ASs keep around two sessions with AS3320. There also exist small customer and peer ASs which advertise paths only for a single prefix.

4.3 AS-specific Information Load Predictions

The expectable information load on common routers in AS3320 using the schemes outlined in section 2.2 is shown in figure 5. To keep the figure simple, we list only routers receiving more than $0.05 \times (|GT| + |IT|)$ paths externally. The predictions

for unlisted routers are similar to those devices that receive little external information. The routers are ordered by the number of paths they receive externally. All values are normalized to multiplies of the routable prefixes $|GT|+|IT|$.

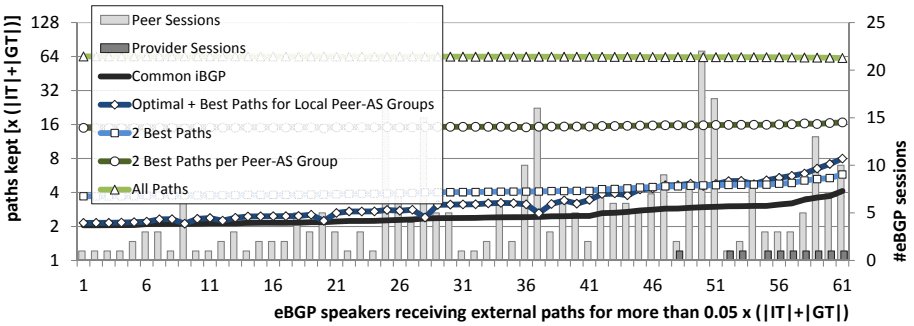

Fig. 5. Information load predictions for eBGP speakers receiving paths for more than 5% of the internal and global prefixes $|IT| + |GT|$.

Applying classical iBGP, routers have to manage two to four times more paths than prefixes are routed (black curve). The exact value depends on the externally learned information. Internally, around $2 \times (|GT| + |IT|)$ paths are received, cf. equation 6 with $m=2$, $n=1$. A significantly higher load (up to the factor of two) arises on routers that learn many paths externally. Such routers keep a session to the provider AS or (several) large peer ASs, cf. the bars on the secondary y-axis.

Exchanging information on all available paths via iBGP, routers must manage 16 to 32 times more paths than today. As it can be expected, the information load would nearly be the same for all routers. Speakers which learn little information from external peers have to maintain approximately 64.1 times more paths than routable prefixes are exchanged. Routers that learn a lot of external information keep around 4% less information, since external paths are only learned once (and not twice via the reflectors). In practice, this load level would for sure increase the required resources significantly. Scalability problems seem likely.

Limiting the number of paths routers learn via an internal session to two per prefix and peer-AS group decreases the load significantly. In case of AS3320, the load could be reduced by a factor of 3.7 to 4.3 in comparison to the advertisement of all paths. The reason for this reduction becomes clear if we have a closer look at the number of sessions large peer-ASs keep, cf. figure 4: ASs advertising paths for many prefixes (i.e. the provider and the peer-ASs listed on the left hand) are mostly connected via clearly more than two eBGP sessions. This results in a lot of paths reflectors provide if the number of advertised paths per prefix and peer-AS group is not limited. Nevertheless, a load increase by a factor of four to eight compared to classical iBGP still defines a massive growth in practice.

Reflecting the two best paths, the information load in comparison to classical iBGP is only increased by a constant factor of $m \times |GT|$, cf. equation 6. As peer-

ASs advertise many paths for the same prefixes, this scheme results in a further significant information reduction. Compared to the advertisement of two paths per peer-AS group and prefix, we observed a further reduction by a factor of two to four. Providing the optimal path and the best path for the local peer-ASs, the load is further reduced on the most speakers. For routers that receive only few external data, the expectable load is comparable to classical BGP. This is due to the fact that if router v keeps only sessions with few, mostly small ASs, equation 5 can be estimated by $m \times (|GT| + |IT|)$, which is equal to $|\mathcal{P}^v_{\neg ext}|_4$ for $n = 1$. For routers that keep eBGP sessions with the provider or (several) large peer-ASs, a load growth has to be expected. Here, besides the optimal path for every prefix, a wide range of AS-group best paths must be provided internally in the worst case. In case of AS3320, this increases the information load by a factor of up to two on some routers, cf. the routers on the right side of figure 5. Discussions with network engineers of Deutsche Telekom AG showed that for the latter two schemes, the expectable information load should be manageable in practice: Some additionally meshed routers in the reference AS manage the same amount of information today and scalability problems do not appear.

5 Generalization

Even if the absolute results obtained in section 4 are highly system-specific, they allow drawing general conclusions on the scalability of the studied path diversity schemes. We summarize the main aspects in what follows.

5.1 General Results

Large transit ASs in the default-free zone usually keep several sessions with large peer or provider ASs. In system-global terms, this from a forwarding perspective highly desirable property leads to a high number of paths available at the border of the AS. If all these paths are spread across a system, significant load increases in comparison to common iBGP must be expected in general. If the system peers with large-ASs clearly more often than n, exchanging only n paths per prefix and peer-AS group significantly lowers the expectable load. However, as long as paths for a comparatively high number of equal prefixes is received from different peer-ASs, a significant load growth in comparison to common iBGP must be expected. This effect is avoided if not more than a constant number of paths is advertised per prefix. The property that advertising n best paths per prefix leads to a constant load growth of $(n-1)\times(|IT|+|GT|)$ can be generalized. It is independent of AS- or other router-specific parameters, cf. equation 6. This makes the expectable load easily, reliably, and precisely predictable. The drawback of this performance behavior is that many routers may need hardware upgrades in the worst case.

Advertising the optimal path and best paths for the local peer-AS groups, the expectable load is heavily dependent on the router's peering. As long as routers peer only with few small ASs (as often the vast majority of routers in an AS do), the load is comparable to common iBGP. Routers that peer with the provider,

large peer, or several mid-size peer ASs are provided with significantly more data. Even if the load increase on such routers may exceed the load induced by advertising n best paths, the limitation to few routers may be a significant advantage: Hardware updates on few devices are much easier to realize in practice. Finally, it shall be noted that equation 5 is a worst case estimation: Externally learned optimal and group best paths are not provided by the reflectors. For AS-border routers that peer with large ASs, this may relatively often be the case.

5.2 ASs without Route Reflection

A basic assumption for our load predictions is that the analyzed AS realizes the iBGP information exchange by means of Route Reflection. To finish our analyses, we sketch what should be observed if this precondition is relaxed.

Applying full-meshed iBGP (either natively or as part of AS Confederations), routers receive paths from more than m iBGP peers. Exchanging information on all paths, this has no direct influence on the information load, since routers learn the same paths. If internal peer routers limit the number of paths they announce, having more internal peers reduces the effectiveness of the information reduction. For example, if advertising n best paths is implemented, every internal peer may advertise n best paths for every prefix. Having more than m internal iBGP peers that peer with large neighbor ASs, a router may receive significantly more paths than specified by equation 6. Implementing the advertisement of the n best paths per peer-AS group in a full-meshed AS, a significant load reduction compared to exchanging all paths cannot be expected: Generally, eBGP speakers do not keep a high number of sessions to the same peer-AS. Similar reflections can be made for the other proposed path diversity schemes.

6 Conclusion and Future Work

In this paper, we provided the first in-depth analyses of scalability aspects important iBGP path diversity involves. For four schemes proposed by protocol designers and researchers, we developed the theoretical basis for load predictions. The results allow operators to estimate the expectable load on the basis of few basic parameters. Reference studies for a large carrier AS and their generalization led to an understanding for the pros and cons of the different schemes with respect to scalability. We found that a significant load growth has to be expected if routers spread all or the n best known paths per AS-group and prefix. Limiting the number of advertised paths on a per prefix basis or realizing a receiver-based information reduction improves the scalability usually drastically. However, it should be kept in mind that scalability is not the only relevant aspect: Every scheme comes along with specific (convergence and correctness) properties. Finally, we want to remark that we also studied the accuracy of the predictions in detail. Even if we could not present the data here, we can affirm that the results seem accurate.

Our analyses gave a first valuable insight into the scalability issues improving iBGP path diversity comes along with. However, due to the space limitation, several important and deeply interesting aspects could not be covered by our studies.

Analyzing and discussing these aspects are important aspects for future research. Examples are the composition of the load or details on the effect on the resulting CPU and memory requirements. Another area we could not discuss is iBGP schemes where the information reduction depends on topology independent path attributes as the Local Preferences or the AS-path length. Examples are *AS-wide* and *Best Local Pref.* [15], specified by Uttaro et al.. Understanding why certain path diversity schemes may cause scalability problems, concepts could be revised and improved. Complementary work covering these aspects is reasonable, too.

References

1. Rekhter, Y., Li, T., Hares, S.: A Border Gateway Protocol 4 (BGP-4) (January 2006) RFC 4271.
2. Bates, T., Chen, E., Chandra, R.: BGP Route Reflection - An Alternative to Full Mesh IBGP (April 2006) RFC 4456.
3. Traina, P., McPherson, D., Scudder, J.: Autonomous System Confederations for BGP (August 2007) RFC 5065.
4. Bornhauser, U., Martini, P., Horneffer, M.: The Scope of the iBGP Routing Anomaly Problem. In: Proceedings of the 17th Conference on Communication in Distributed Systems (KIVS 2011). (March 2011)
5. Griffin, T., Wilfong, G.: On the Correctness of IBGP Configuration. SIGCOMM Comput. Commun. Rev. **32**(4) (August 2002) 17–29
6. Walton, D., Retana, A., Chen, E., Scudder, J.: Advertisement of Multiple Paths in BGP (August 2010) Internet Draft.
7. Raszuk, R., Cassar, C., Aman, E., Decraene, B.: BGP Optimal Route Reflection (BGP-ORR) (October 2010) Internet Draft.
8. Rekhter, Y., Gross, P.: Application of the Border Gateway Protocol in the Internet (March 1995) RFC 1772.
9. Traina, P.: Experience with the BGP-4 Protocol (March 1995) RFC 1773.
10. Traina, P.: BGP-4 Protocol Analysis (March 1995) RFC 1774.
11. Bu, T., Gao, L., Towsley, D.: On Routing Table Growth. SIGCOMM Comput. Commun. Rev. **32** (January 2002) 77–87
12. Elmokashfi, A., Kvalbein, A., Dovrolis, C.: On the scalability of bgp: the roles of topology growth and update rate-limiting. (2008)
13. Walton, D., Retana, A., Chen, E., Scudder, J.: Advertisement of Multiple Paths in BGP (May 2002) IETF IDR Internet Draft.
14. Walton, D., Retana, A., Chen, E., Scudder, J.: BGP Persistent Route Oscillation Solutions (May 2010) Internet Draft.
15. Uttaro, J., Van den Schrieck, V., Francois, P., Fragassi, R., Simpson, A., Mohapatra, P.: Best Practices for Advertisement of Multiple Paths in BGP (October 2010) Internet Draft.
16. Bornhauser, U., Martini, P., Horneffer, M.: An Inherently Anomaly-free iBGP Architecture. In: Proceedings of the 34th IEEE Conference on Local Computer Networks (LCN 2009), IEEE Computer Society (October 2009) 145–152
17. Basu, A., Ong, C.H.L., Rasala, A., Shepherd, F.B., Wilfong, G.: Route Oscillations in I-BGP with Route Reflection. In: SIGCOMM '02: Proceedings of the 2002 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications, New York, NY, USA, ACM (2002) 235–247
18. Huston, G.: BGP Routing Table Analysis Reports (Website) (November 2010) Online: http://bgp.potaroo.net.