Computer Networks 57 (2013) 954–975

Contents lists available at SciVerse ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

BGP-XM: BGP eXtended Multipath for transit Autonomous Systems

Jose M. Camacho *, Alberto García-Martínez, Marcelo Bagnulo, Francisco Valera

Universidad Carlos III de Madrid, Av. Universidad 30, 28911 Leganes, Madrid, Spain

ARTICLE INFO

Article history:
Received 28 February 2012
Received in revised form 15 November 2012
Accepted 21 November 2012
Available online 29 November 2012

Keywords:
BGP
Multipath
Interdomain
Routing
Protocol

ABSTRACT

Multipath interdomain routing has been proposed to enable flexible traffic engineering for transit Autonomos Systems (ASes). Yet, there is a lack of solutions providing maximal path diversity and backwards compatibility at the same time. The BGP-XM (Border Gateway Protocol-eXtended Multipath) extension presented in this paper is a complete and flexible approach to solve many of the limitations of previous BGP multipath solutions. ASes can benefit from multipath capabilities starting with a single upgraded router, and without any coordination with other ASes. BGP-XM defines an algorithm to merge into regular BGP updates information from paths which may even traverse different ASes. This algorithm can be combined with different multipath selection algorithms, such as the K-BESTRO (K-Best Route Optimizer) tunable selection algorithm proposed in this paper. A stability analysis and stable policy guidelines are provided. The performance evaluation of BGP-XM, running over an Internet-like topology, shows that high path diversity can be achieved even for limited deployments of the multipath mechanism. Further results for large-scale deployments reveal that the extension is suitable for large deployment since it shows a low impact in the AS path length and in the routing table size.

1. Introduction

The Internet topology is inherently redundant. Intra-site topologies show high redundancy in their interconnection [1,2]. Autonomous Systems (ASes) are usually connected to multiple provider ASes [3] to leverage fault tolerance. Internet's increasingly richer connection degree is the result of the quest for improved performance and lower transit costs. Even after removing all the paths which are not usable according to the business relationships established among the connected ASes, a large number of paths traversing different sequences of ASes exists between most of the Internet sites [4]. Additional redundancy comes from the use of multiple links between pairs of neighboring ASes [5].

For years, the networking community has sought for flexible and simple ways to use the largest number of available paths in order to improve availability and perform traffic engineering. In the same way as technologies such as MPLS (MultiProtocol Label Switching) achieve that flexibility in the intra-domain scope, the management of traffic exchanged between domains could be further improved if inter-domain routing policies are changed dynamically.

As reported in [6], this occurs in today's Internet for the particular case of egress traffic in stub ASes. Another example are the mechanisms to coordinate traffic engineering across multiple links between two peering ASes proposed in [7,8]. These solutions share in common that there is no need to advertise to other domains the routing changes due to the traffic engineering.

Nevertheless, transit ASes must advertise their selected paths to other ASes using BGP (Border Gateway Protocol). In the extreme case, a change in a route selected by a transit AS, which wishes to modify its outgoing traffic pattern, may result in undesired changes for its incoming traffic,

* Corresponding author. Tel.: +34 6248795; fax: +34 6248749.
E-mail addresses: jcamacho@it.uc3m.es (J.M. Camacho), alberto@it.uc3m.es (A. García-Martínez), marcelo@it.uc3m.es (M. Bagnulo), fvalera@it.uc3m.es (F. Valera).

1389-1286/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.comnet.2012.11.011

because of subsequent routing decisions made by other ASes [9]. Moreover, even if the effects for the incoming traffic are negligible, the deployment of this strategy is likely to stress BGP routers all over the Internet, since they will have to cope with frequent routing changes. In addition to the harm caused by route recomputation, more undesired interactions may appear if ASes perform active path monitoring as in *route flap damping* [10].

Multipath routing has been proposed to achieve flexible and frequent load balancing without path recomputation [11,12]. Should the routing protocol make available multiple paths to a destination, a router is able to balance traffic across any combination of available paths without being forced to send additional routing messages. This should provide transit ASes with finer control over their outgoing traffic without incurring in the aforementioned problems.

Nevertheless, upgrading the current inter-domain routing to support multipath is far from trivial, especially due to the need of incremental deployments. Current proposals for multipath inter-domain routing are based on the advertisement of additional paths, others than the BGP best path, by means of a parallel negotiation between routers [13] or new BGP capabilities [14]. Unfortunately, their deployment requires the support of the mechanism in two or more neighboring ASes.

Other proposals [15,16] advocate the utilization of multiple paths while maintaining the BGP advertising scheme, thus announcing only one of them as in-use. Yet, withholding the advertisement of in-use paths requires additional mechanisms to ensure loop-freeness.

The taxonomy is completed with some commercial router implementations, which already allow the use of multiple paths as long as all of them share most of their BGP attributes with the BGP best path [17]. In particular, all the routes to a destination can only differ in their IGP NEXT_HOP attribute. The obtained path diversity mostly result from the use of different links between consecutive ASes.

We claim that a multipath routing mechanism enabling inter-domain traffic engineering must fulfill the following requirements:

Allow high path diversity. The mechanism should impose as few restrictions as possible to the selection of multiple routes. In particular, it must not be restricted to select paths with the same sequence of traversed ASes or paths received from the same neighboring AS, since transit ASes may require moving traffic freely from one neighboring AS to another, e.g., to offload traffic to another provider or to obtain better performance.

Be backwards compatible. The mechanism must allow the selection of multiple routes being advertised by (unipath) BGP routers. Additionally, current BGP routers must be able to receive and use any set of routes created by a multipath router. To fulfill the latter, any set of routes selected by the multipath router must be expressed in a BGP-compatible format. The mechanism must not require any data plane modification in devices other than the multipath-upgraded routers. Hosts and routers must be able to exploit multipath routing without including multipath-specific information in data packets. Therefore, the resulting multipath service can be described as a multipath *next-hop routing*, i.e., every router takes forwarding decisions independently of the decisions taken by other routers.

Be incrementally deployable. In order to start providing multiple routes, the mechanism must not require being simultaneously deployed in different ASes, or in all the routers of an AS. Enabling multipath operation in a single router should be enough to disclose the multiple paths available to the router.

Be highly configurable. The selection of multiple routes must be tunable. For example, it must allow controlling the size of routing tables, the number of alternative paths per prefix or the prefixes for which multipath routing is enabled. In addition, the multipath mechanism must be configurable to limit the *route stretch*, i.e., the length difference between the longest and the shortest path to a destination, measured in the number of traversed ASes.

Seamlessly preserve usual business relationships. The current business model for inter-domain connectivity results in some widely used relationships such as *transit*, *peering*, *siblings* or *partial transit* [18]. The relationships define preference, import and export filtering rules for route advertisements. Despite the increment in the number of routes, the effort put in the configuration of a multipath mechanism should be close to that for configuring BGP.

Preserve the effects of usual traffic engineering techniques. An AS administrator may choose from a set of techniques to determine how its traffic exits from the AS, and to influence the path followed by the ingress traffic. Domain's inclinations for outbound traffic are usually controlled by associating explicit preferences to routes. The most popular tools to influence the path for the incoming traffic, besides the injection of more specific prefixes, are the use of pre-agreed COMMUNITY values to inform other ASes about local preferences, the artificial increase of the length of the AS_PATH attribute to make the path less attractive, and the use of inter-AS metrics (MED attribute) [19]. Any multipath solution must allow both the domain deploying multipath and the rest of the domains to continue using these tools in a similar way as they currently do.

Generate loop-free paths. Resulting routes must be loop-free. Note that, in order to be backwards compatible, data packets are not required to carry any path selector, so loop-freeness is only assured if none of the possible combinations of the routes selected independently by different routers generate a cycle.

Stable under non-conflicting routing policies. ASes can choose their policies on their own, so the absence of conflicting policies in inter-domain routing cannot be guaranteed [20]. It has been shown that BGP is stable when transit and peering, the most common business relationships, are the only ones deployed [21]. Any multipath routing mechanism must assure that, at least, it is stable in the scenarios described in [21]. Note that this is not a trivial statement, as the work of Chau and Griffin [22] states that multipath can be unstable for configurations in which unipath is stable.

In this paper we present *BGP eXtended Multipath*, BGP-XM, a BGP extension that allows routers to use multiple paths across different ASes, including paths with different next ASes. BGP-XM defines a router architecture tailored