AIDR: Aggregation of BGP Routing Table with AS Path Stretch

Yangyang Wang, Jun Bi, Jianping Wu
Network Research Center, Department of Computer Science of Tsinghua University
Tsinghua National Laboratory for Information Science and Technology (TNList)

***Abstract*—As Internet growth, more and more prefix fragments are announced into the global routing system due to operational reasons of inconsecutive address allocation, multihoming, and traffic engineering. The BGP routing table size in Default Free Zone (DFZ) fast growth will consume more memory space and computational capacity. It has been known that Internet will face with routing scalability issue, especially in the large address space (e.g., IPv6) deployment. In this paper, we propose an innovation to BGP, named Aggregation-aware Inter-Domain Routing (AIDR). It will take the prefix aggregation into account to make tradeoff in the best route selection. We evaluate the effect of AIDR on global routing system using the BGP traces from RouteViews and RIPE. It shows that, averagely, AIDR-based aggregation can reduce to roughly 15%∼35% of original routing table size under the 2.0 AS path stretch constraint, and to 25%∼40% with no AS path stretch.**

I. Introduction

The Internet is composed of a large number of Autonomous Systems (ASes). BGP has been the de facto standard protocol for today's inter-domain routing. Due to many factors, such as practical multihoming and traffic engineering, lots of prefix fragments are announced in the global routing system which leads to rapid growth of the BGP routing table size in Default Free Zone (DFZ). It has been recognized that internet routing will face with scaling challenge on the road of IPv6 wide deployment in the future.

Many solutions have been proposed for routing scalability issue. Solutions like core-edge separation and core-edge elimination [1], and compact routing [2] are revolutionary approaches, and could be take a long-term way. On the other hand, some work take an evolutionary approach to design practical methods to routing scalability. Ballani et al. proposed a prefixes virtual aggregation (VA) method [3] that is to reduce the size of the Forwarding Information Base (FIB) in ISP networks, and can be deployed ISP-individually. B. Zhang et al. [4] presents an evolutionary approach from the intra-domain VA isolands to the inter-domain VA by merging the existing VA isolands, which can head to the situation of core-edge separation finally. Some studies [5][6][7] is to construct optimal FIB with minimum size, which indicate route space can be aggregatable. FIB aggregation can be deployed individually on local routers without any impact on global routing.

This work is supported by National Science Foundation of China under Grant 61073172, Program for New Century Excellent Talents in University, Specialized Research Fund for the Doctoral Program of Higher Education of China under Grant 200800030034.

TABLE I
AGGREGATION-AWARE ROUTE DECISION STEPS

rank	route selection steps
1	highest local pref
2	shortest AS path length **- replaced with aggregation consideration with constraint on AS path stretch**
3	lowest origin value
4	lower MED
5	eBGP-learned over iBGP-learned
6	lowest IGP cost/distance
7	lowest Router ID

However, reducing FIB locally cannot decrease routes deaggregation and prevent prefixes fragments global propagation and process cost. In this paper, we propose an innovation to BGP. named Aggregation-aware Inter-Domain Routing (AIDR). Its basic idea is to take the route aggregation into account in the best route selection. As the dense connectivity of Internet topology, BGP routers inside one ASes may have multiple available routes with different nexthops to reach the same destination. We can select suboptimal routes as the "best" routes to enable the prefixes with consecutive address space have the same nexthop, such that they are aggregatable in local forwarding table and routing table. It may incur path stretch. Actually, BGP doesn't always prefer shortest AS paths due to routing policies among ASes [8]. Based on routing path diversity, it is reasonable to save the overhead in routers memory and computation by routes aggregation with an acceptable cost of AS path stretch, especially in the critical situation of routing table inflation in the future. We make an evaluation and analysis for AIDR based on the BGP routing table traces from RouteViews and RIPE RIS Project [9]. The results shows that it could get a reduction of roughly 25%∼40% of original routing table size with no AS path stretch, and 15%∼35% under the 2.0 stretch constraint.

II. Architecture and algorithms of AIDR

AIDR modifies the BGP route decision process as showed in the Table I, We relax the shortest path rule, and take aggregation into account with compromising path stretch when select best route from the routes received from its BGP neighbors. After applying the second aggregation-aware rule, there could be multiple equal routes left and they need to go through the rest tie-break decision rules based on other BGP path attributes, such as ORIGIN, Multi-exit Discriminator (MED), etc. We describe our route selection method considering route

aggregation. All prefix nodes construct a tree in terms of the immediate covering relation between prefixes. The aggregation and decision process is performed from bottom to top, and include two levels: the horizontal and vertical. Our vertical aggregation is based on the method in [10] which computes the optimal aggregation of FIB in the case of each prefix having multiple selectable nexthop by dynamic programming. Different from that method, our method has a horizontal aggregation of sibling prefixes under the same parent. It could generate new less-specific prefixes in the prefix tree, and get a more aggregation. The detailed steps is described as follows. *It does not introduce extra routable space*.

Horizontal level aggregation: This aggregation is applied on the sibling prefix nodes that have the same immediate parent node. **type 1)** If the address block of two prefixes a and b are consecutive in address space, and they have the common set of available nexthops, they will be aggregated into a less-specific prefix node c with intersection iH_c of their individual available nexthop set H_a and H_b. Suppose that the number of prefixes of optimal aggregation is known before (it is set to 1 initially for each of prefix nodes), denoted as $v(a,i), i \in H_a$ and $v(b,j), j \in H_b$, the optimal value $v(c,k)$ for the newly generated prefix c is set to $v(a,k) + v(b,k) - 1$ for each nexthop k in the intersection set iH_c. **type 2)** If the address block of two prefixes a and b are consecutive in address space, but they have no common available nexthops, they will be aggregated into a less-specific prefix node c with union uH_c of their individual available nexthops set H_a and H_b. And the optimal value $v(c,k)$ is set to $v(a,i) + v(b,j), (i \in H_a, j \in H_b, k \in uH_c)$. Because there always is one of a and b could not be aggregated into the new routes c whichever of nexthops is selected. Note that, $v(a,i) = v(a,j), (i \neq j; i,j \in H_a)$, the same to b and c.

Vertical level aggregation: Suppose that the prefix p has a set of immediate child $c_1, c_2, ..., c_n$, and the optimal value of each subtree root at these children is denoted by $v(c_i,l), (l \in H_{c_i}, 0 \leq i \leq n)$, H_{c_i} represent the available nexthop set of each child c_i. Then, we compute $v(p,k)$ by traversing each available nexthop k of the available nexthop set H_p of p. For each k, compute the cumulative sum iteratively across each child $c_i : v(p,k) = v(p,k) + v(c_i,k) - 1$ if $k \in H_{c_i}$, otherwise $v(p,k) = v(p,k) + v(c_i,l), l \in H_{c_i}$ if $k \notin H_{c_i}$. And the last is to obtain all the minimum value of $v(p,k), k \in H_p$ as the optimal value, and update the original nexthop set H_p to the set of nexthops associated with optimal value.

Iterate the two steps across each parent-child hierarchy in the tree by postorder from bottom to top, we can obtain the number of prefixes after optimal aggregation, and the available nexthops (i.e., route paths) associated with each prefix. And then, we can traverse the tree from top to bottom to make a route selection decision for each prefix and get the aggregate routes. For more details refer to [11].

III. Evaluation on cases

We estimate the feasibility and effect of AIDR by collecting BGP routing data from RouteViews and RIPE [9] on April 22, 2010. Based on all the collected BGP data, we can partly conclude the received routes of a give AS from its neighbors. The monitored and concluded BGP routes has no detailed nexthop IP for each route. We use the AS-level nexthop as an approximate estimation of IP nexthop as in [7]. Figure 1(a) shows the aggregation rate with path stretch growth. It could get a reduction of 15%∼35%, and the average AS path stretch is about 1.3 under the 2.0 stretch constraint.

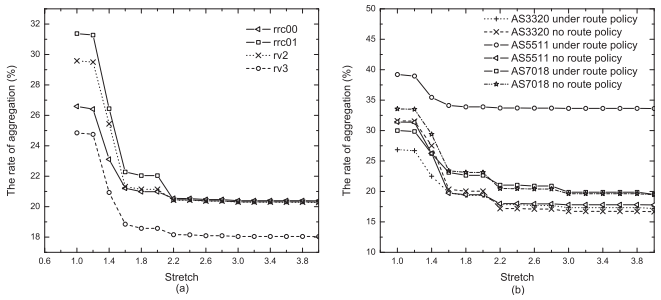

Fig. 1. (a) Reduction vs. stretch of AIDR; (b) AIDR route aggregation with and without route preference policy constraints

Figure 1(b) shows the practical impact with and without considering route preference policy (i.e., prefer customer routes *over* peer *over* provider) of Gao-Rexford condition [12]. The extent of reduction with no route policy is larger than that with route policy. This is because route policy decrease the prefixes and available paths to be aggregated. Based on conclued BGP routes, we calculate the aggregate rate of local RIB is 51% for AS5511, 49% for AS3320, and 46% for AS7018 without AIDR selection, which will get a more reduction about 10%∼20% by AIDR selection with no stretch.

References

[1] D. Jen, M. Meisel, H. Yan, D. Massey, L. Wang, B. Zhang, and L. Zhang, "Towards a new internet routing architecture: Arguments for separating edges from transit core," in *Proc. HotNets-VII*, 2008.

[2] D. V. Krioukov, K. C. Claffy, K. R. Fall, and A. Brady, "On compact routing for the internet," *CoRR*, vol. abs/0708.2309, 2007.

[3] T. C. Hitesh Ballani, Paul Francis and J. Wang, "Making routers last longer with viaggre," in *USENIX NSDI 2009, Boston, MA*, April 2009.

[4] B. Zhang and L. Zhang, "Evolution towards global routing scalability," in *Internet Draft*, 2009.

[5] B. Zhang, L. Wang, and L. Zhang, "Fib aggregation, draft-zhang-fibaggregation-02.txt," in *Internet Draft*. [Online]. Available: http://tools.ietf.org/html/draft-zhang-fibaggregation-02

[6] R. Draves, C. King, V. Srinivasan, and B. Zill, "Constructing optimal ip routing tables," in *INFOCOM*, 1999, pp. 88–97.

[7] X. Zhao, Y. Liu, L. Wang, and B. Zhang, "On the aggregatability of router forwarding tables," in *INFOCOM*, 2010, pp. 848–856.

[8] L. Gao and F. Wang, "The extent of as path inflation by routing policies," in *IEEE Global Internet Symposium*, 2002.

[9] "Routeviews: http://www.routeviews.org/; ripe: http://www.ripe.net/data-tools/stats/ris/ris-raw-data."

[10] Q. Li, D. Wang, M. Xu, and J. Yang, "On the scalability of router forwarding tables: Nexthop-selectable fib aggregation," in *INFOCOM*, 2011, pp. 321–325.

[11] Y. Wang, J. Bi, and J. Wu, "Aidr: Towards an aggregation-aware inter-domain routing," Tech. Rep., 2011. [Online]. Available: http://netarchlab.tsinghua.edu.cn/˜wyy/paper/tr-aggre.pdf

[12] L. Gao and J. Rexford, "Stable internet routing without global coordination," *IEEE/ACM Trans. Netw.*, vol. 9, no. 6, pp. 681–692, 2001.