The Extended BGP4+ Algorithm for Multihoming

Jongwook Woo[1], Seunghyup Ryu[2], Jaiyong Lee[2]

Jaehoon Jeong[3], Jungsoo Park[3], Hyoungjun Kim[3]

[1]Computer Information Systems Department, California State University, Los Angeles
90032-8530, Los Angeles, CA, USA
e-mail : jwoo5@calstatela.edu
[2] Dept. of Electronic Engineering, Yonsei University,
120-749, Seoul, Korea
e-mail : {shr, jyl}@nasla.yonsei.ac.kr
[3] ETRI (Electronics and Telecommunications Research Institute),
305-350, Dae-Jeon, Korea
e-mail : {paul, pjs, khj}@etri.re.kr

Abstract: Multihoming is the one of issues to be addressed in IPv6. This paper presents the possible solution of multihoming methods among site exit routers. We also present the possible policies to select the proper router for load sharing. Besides, We propose the extended BGP algorithm for multihomed sites, as BGP has been the popular routing algorithm in IPv4 and IPv6. The extended BGP algorithm can enhance the load sharing and fault-tolerance of multihomed sites. And, it is not difficult to convert the existing routers with BGP to the extended BGP for the multihoming environment because the extended BGP algorithm slightly modifies the existing BGP.

1. Introduction

The address space of IPv6, the next generation of Internet Protocol, extends the existing 32-bit space of IPv4 to 128-bit space. With IPv6, various terminals such as cellular phone and many other electric devices can have their own Internet address, which is not possible for the lack of address space in IPv4. Besides, IPv6 has many good properties that fit well for those devices, for example, automatic address allocation. However, many issues to address in IPv4 still exist in IPv6. In this paper, we present the *Multihoming* issue of IPv6 and propose the possible solution.

A site, which might be a host, a server or an *AS* (Autonomous System), can have multiple network interfaces. If each interface is connected to an *ISP* or an *AS*, it may have multiple connections to the different *IS*Ps or *AS*s. Such a site is called *Multihomed*. Thus, if a connection to an *ISP* or an *AS* is failed, the multihomed site can access Internet via other *ISP*s or *AS*s. On the other way, a multihomed site can be accessed from the outside via one of the multiple interfaces. In a multihomed site, the site becomes fault-tolerant in case of a network failure. And, we can improve the network performance with the load-sharing of the multiple connections.

In this paper, we define *Multihoming* as a situation where a group of the exit routers of an *AS* is connected to the multiple external *AS*s or where an external *AS* is connected to the multiple routers of an external *AS*. Thus, we analyze the *BGP* routing algorithm that is for the routing of the external *AS*s. There are many studies on *Multihoming* of IPv6 in the various aspects [1, 2, 4, 5, 6, 7, 8, 9] and several drafts are presented to the IETF working group. Some of the drafts are about the requirements of IPv6 *Multihoming* in load sharing, address selection, routers, simplicity, and other requirements. The others present the default router selection in *Multihoming* and the backbone selection among the multiple backbones [1, 2, 4, 5, 6, 7, 8, 9].

At this moment, most of the studies in *Multihoming* are only for a network connection failure. Even the multihomed site is safe for the failure, it is inefficient in terms of the network resource utilization. For the efficient resource management, we need to consider not only the redundancy for a link failure but also the load sharing of the multiple interfaces for the utilization.

This paper introduces existing *Multihoming* mechanisms for redundancy and load sharing. And, it describes *BGP* (Border Gateway Protocol) routing algorithm between site exit routers of *AS*s. Then, we extend the existing *BGP* algorithm for *Multihoming* sites. In section 2, we describe the existing *Multihoming* mechanisms for redundancy and load sharing between site exit routers. And, for load sharing, the possible policies to select the router to distribute the load are presented. In section 3, we introduce existing *BGP4+* algorithm. In section 4, we propose our extended *BGP4+* algorithm for *Multihoming* and describe it. Finally, the conclusion is presented.

2. Multihoming Mechanism in IPv6

2.1 Multihoming Mechanism between Site Exit routers

In RFC2260 [1], two *Multihoming* mechanisms between Site Exit routers are proposed. One is *Auto Route Injection*, and the other is *Non-direct eBGP Peering*. These mechanisms are presented for multiple Site Exit routers connected to multiple *ISP*s. This paper describes these *Multihoming* mechanisms with *AS*s instead of *ISP*s because *AS* covers *ISP*.

2.1.1. Auto Route Injection

In Figure 1, each Site Exit routers of *AS A*, *B*, *C* is connected to its corresponding Site Exit routers of the neighbor *AS*'s. *AS C* can access Internet via each Site Exit routers of the *AS A* or *AS B* for the addresses which have *prefix A* and *B* respectively. The local Site Exit routers connected each other in the *AS C* share prefix information - *prefix A* and *B* - with *iBGP* (Interior BGP) Peering.

In the *AS C*, if the route R_b to *AS B* is disconnected because of the failure of the Site Exit router in the *AS B*, the router R_b will send information about *prefix B* to *AS A* through the router R_a with *iBGP*. This mechanism is called *Auto Route Injection*. That is, site exit routers R_a and R_b of the *AS C* connected to each *AS*s exchange each prefix information. If one of the routes fails, it sends its prefix information to the other router which has a link alive. Thus, multihomed site *AS C* maintains fault tolerance.

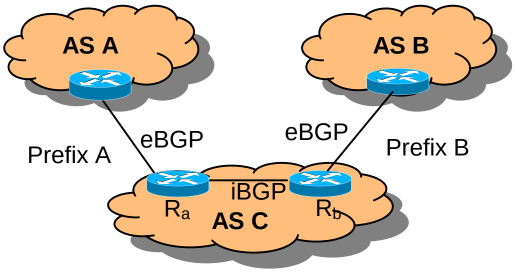

Figure 1. Auto Route Injection

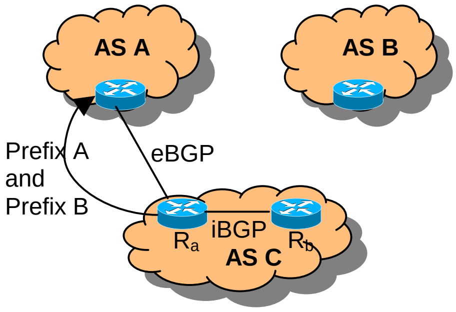

Figure 2. Redundancy in Auto Route Injection

2.1.2 Non-direct eBGP Peering

Figure 3 shows the *Non-direct eBGP Peering*. A Site Exit router has primary and secondary link integrated with *eBGP*; the router R_a's primary information is *prefix A* and secondary information is *prefix B*; the router R_b's primary information is *prefix B* and secondary information is *prefix A*. In general case, each Site Exit router is connected with the primary *AS* for the primary prefix information. As shown in Figure 4, if the link between *AS B* and *AS C* is failed, the traffic of *prefix B* can be forwarded to *AS B* via the site exit router R_a of *AS C* as a secondary link.

In *Auto Route Injection*, as shown in Figure 2, when a local site exit router R_a or R_b has a problem or if a link is disconnected, both *prefix A* and *B* traffics are forwarded to a *AS*. Therefore, the AS should handle both *prefix A* and *B* traffics. And, in case the traffics are heavy to the AS that takes care of both traffics *prefix A* and *B*, the *AS* will have huge load. However, in the *Non-direct eBGP Peering*, though either site exit router R_a or R_b has a problem, both *AS A* and *B* may not have heavy load because the site exit router alive can distribute traffics to both *AS A* and *B* via its primary and secondary links.

2.2. Load Sharing

BGP and *IGRP* (Interior Gateway Routing Protocol) protocols are examples of the existing static load sharing methods. These methods rely on the configuration of the router administrator. The administrator configures the router with various parameters such as bandwidth, link cost, or any related to the network condition. And, with these parameters, the router can find the best path to destination and process load sharing for multiple links. Cisco routers can configure load sharing between site exit routers with the static method for the existing *BGP4+* protocol. When setting up the router, the administrator divides the prefix information so that the traffic loads can be distributed to each site exit router. It is simple to share the load with the static configuration. However, if the network condition changes, the router may not perform optimally with the static parameters. As a result, it may bring the entire network the lower performance.

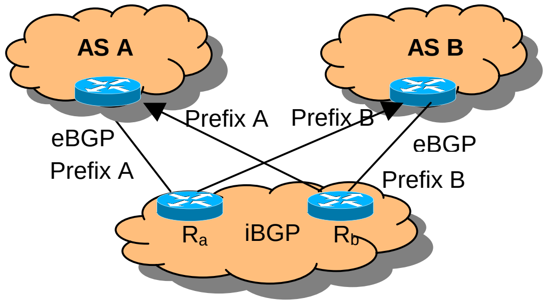

Figure3. Non-direct eBGP Peering

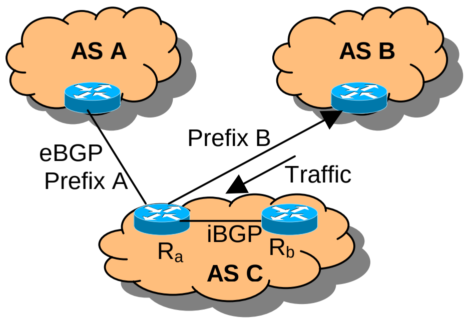

Figure 4. Redundancy in Non-Dorect eBGP Peering

To solve the problem, a router or server checks the network status automatically and sets the load sharing parameter that fits to the network status. This automatic load sharing is called dynamic load sharing. The dynamic load sharing is much faster than static load sharing and done automatically. Besides, it saves the time used for setting the parameters by the administrator. And, it reduces the packet losses and delays so that the entire network performance can be improved.

We need a policy to select a proper router when loads are distributed. As follows, we present the two possible policies that select the candidate router for load sharing among the site exit routers that are multihomed in a group of site exit routers in the *Auto Route Injection* (Figure 1) and *Non-direct eBGP Peering* (Figure 3) mechanisms: the random router selection and the best router selection.

- *Random Router Selection*

It selects each site exit router on round robin or randomly without regarding the status of the routers and distributes the load sequentially for load sharing. It is simple but it may have problems, such as a delay, by regardless of workload on each router.

- *Best Router Selection*

It selects the router that fits best on the status of routers or the network as follows.

a. It distributes the load to the router that is idle among the multihomed routers during run-time.
b. It distributes the load to the router that has the shortest response time for the multicast requests sent to the multihomed routers.
c. It distributes the load to the router with the weight that is assigned to each router based on workload information analyzed in the group of site exit routers during run-time.
d. It distributes the load to the router by other factors such as metrics, preference value, and prefix information, etc.

3. BGP4+ Algorithm

Among the *EGP*s (Exterior Gateway Protocol), *BGP4* is the most popular. *BGP* has been developed to *BGP4* and to other versions. The most important function of the *BGP* is to exchange the network reachability between *BGP* peer routers in the *AS*s. *BGP* uses TCP to guarantee the reliability. The mostly used *BGP4* works for *IPv4* only. RFC2283 [12] extends *BGP4* and defines many additional factors in order to exchange the information of the other protocols. For example, it presents how to exchange information in *IPX* or *IPv6* with *BGP4*. Thus, it defines *BGP4+* protocol to exchange the routing information between *AS*s using *BGP* in *IPv6*.

BGP4+ selects just one path as the best path to a destination [11]. Once the path is selected, it is stored into the routing table. And, the path is propagated to the neighbors. It is called *the best path selection algorithm* and the algorithm in *BGP4+* is the same as that of the *BGP4*. Therefore, in this paper, we regard as the same the solution of load sharing for *BGP4+* and *BGP* in the multihomed network. In the following, we introduce the **decision process** that is the part of the *best path selection algorithm* of the *BGP4* to select the optimal route for a destination.

BGP Decision Process for a BGP router [11]

Phase 1. Calculate Preference Degree

A router in a local *AS* calculates the preference degree of the each route given from a *BGP* speaker of the neighboring *AS*. Then, it advertises the highest degree preference route for each destination to other *BGP* speakers in the local *AS*.

Phase 2. Route Selection

It selects the best route to each destination in the local *AS* and installs it on the Loc-Rib. That is, for each possible route to a destination in the Adj-RIBS-In, the local *BGP* speaker selects the route as follows:

(a) The route that has the highest preference among the routes to the same destination
(b) The unique route to one destination
(c) The route chosen by the tie-breaking rule

The selected route is replaced with the route that has the same destination of the selected route in the Loc-Rib of each speaker

Phase 3. Route Propagation

By the policy of *PIB* (Policy Information Base), it propagates the routes in the Loc-Rib to each *BGP* peer of the neighboring *AS*s. Route Aggregation and Information Reduction are optionally performed.

4. The extended BGP4+ Algorithm

This section presents the solution for load sharing in *IPv4* and *IPv6* between site exit routers in the multihomed network by extending Routing Table and the *BGP best path selection algorithm*. First of all, if a site exit router *a* in one *AS* is connected to the *n* routers in one *AS* or multiple *AS*s, the router *a* for a destination *d* has the number of *n* next hops at most. Thus, the *BGP* routing table grows at least *n* times in ours for the multihomed site. Besides, the route selection algorithm is modified to select at most *n* next hop paths for a destination *d*.

Based on them, we propose the following algorithm that modifies and extends the Phase 2 of the existing *BGP* Decision Process for the mutihomed site in IPv4 and IPv6. Our algorithm compares the received routes and the existing optimal route and it selects the optimal route for each Next Hop to a destination instead of selecting one optimal route for the destination. Therefore, the routing table includes each optimal path for multiple next hops for the destination but the existing algorithm has the routing table that contains one optimal path for the destination even it has the multiple next hops.

BGP Decision Process for one Multihomed BGP router

Phase 2. Router Selection

The router selects the best route <u>of each next-hop</u> to each destination in the local *AS* and installs it on the Loc-Rib. That is, for each possible route to a destination in the Adj-RIBS-In, the local *BGP* speaker selects the route as follows:

(a) <u>Each</u> route that has the highest preference <u>for each next-hop AS path route</u> to the same destination – that is, <u>the best route per an interface</u>.
(b) The unique route to one destination.
(c) <u>Each</u> route chosen <u>among the different next-hop AS path routes to the same destination</u> by the Tie-breaking rule.

The selected routes are replaced with the routes that have the same destination in the Loc-Rib of each speaker – if each route is in the different next-hop AS path from the route in the table even for the same destination, the selected route is not replaced with the route in the table.

It is possible to apply our updated phase 2 to the existing best path selection algorithm of *BGP4* for the load sharing in the multihomed network. Even we do not show the solutions of dynamic traffic load sharing algorithm in this paper, we present the basis for the load sharing algorithms. Thus, based on our extended BGP algorithm, each router can select the best path for each Next-Hop to one destination among the multiple paths to the Next-Hop eBGP routers in the multihomed network.

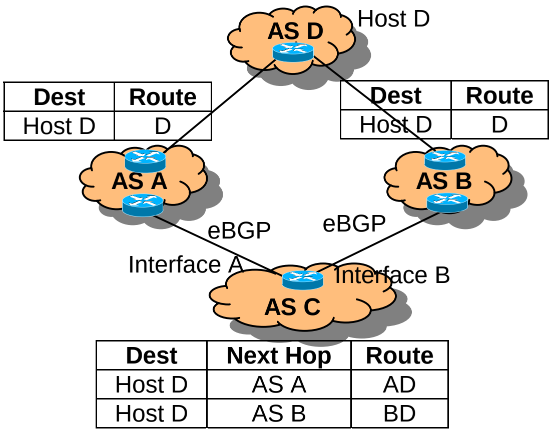

Figure. 5. The extended BGP4+ Example

For example, Figure 5 shows the routing table of *AS A*, AS *B*, and *AS C* for BGP algorithm when there is a destination *host D*. *AS D* for the destination host propagates the reaching information to the neighboring *AS*s *A* and *B*. Thus, both *AS A* and *AS B* store the reaching information for *D* as (*D*) in each routing table. And, *AS C* receives the reaching information to the destination *D* from *AS A* and *AS B*. With the existing algorithm, *AS C* will select either a *route A D* or a *route B D* as the best route. But, in our algorithm, *AS C* will select the *route A D* for the *interface A* and the *route B D* for the *interface B* as the best routes respectively. Therefore, with our algorithm, we can utilize the network interfaces and implement the load sharing for the interfaces.

To analyze the extended BGP algorithm, the Loc-Rib table of a *BGP* router that installs the extended algorithm should be analyzed. For each destination *D*, the space complexity of the Loc-Rib table is, for the existing BGP algorithm, $O(1 \times D)$ and for the extended algorithm, $O(n \times D)$; *n* is the number of the interfaces within one *AS*. Similarly, for each destination *D*, the time complexity of the Loc-Rib table is $O(1 \times D)$ for the existing BGP algorithm and is $O(n \times D)$ for the extended algorithm. Even though the time and space complexities of the extended algorithm proposed in this paper are not as good as them of the existing algorithm, we believe that resource and time efficiency of the extended algorithm may be better than that of the existing algorithm for the entire network.

5. Conclusion

This paper describes the basis for load sharing implementation between multihomed site exit routers in IPv4 and IPv6. We show the multihomed network frames based on *Auto Route Injection* and *Non-direct eBGP Peering* mechanism that are mentioned in the RFC 2260 [1] and RFC 3178 [2]. We also present the possible route selection policy and requirement for load sharing in multihomed sites. Finally, we propose the extended BGP algorithm that selects the best route of each next hop for a destination in site exit routers. Because the router has the optimal route information for each next hop for a destination, we can apply load-sharing algorithms. Our study has been on progress for measuring efficiency and performance with the SSFNet simulator [17]. As a future work, a load sharing in multihomed sites with the extended BGP algorithm will be presented.

References

[1] T. Bates and Y. Rekhter, Scalable Support for Multi-Homed Multi-Provider Connection, RFC 2260
[2] J. Hagino and H. Snyder, IPv6 Multihoming Support at Site Exit Routers, RFC 3178
[3] Yong Kim, M. Shin, J. Park, S. Lee, Next Generation Internet Protocol IPv6, 2002, Da-Sung publication
[4] Steven Deering, Overview of IP(v6) Multihoming Issues, IPng working , September 29, 1999 in Tokyo
[5] B. Black, V. Gill, J. Abley, Requirements for IP Multihoming Architectures, draft-ietf-multi6-multihoming-requirements-01, Network Working Group, June 25, 2001
[6] B. Black, V. Gill, J. Abley, IPv4 Multihoming Motivation, Practices and Limitations, draft-ietf-multi6-v4-multihoming-00, Network Working Group, June 25, 2001
[7] R. Hinden, IPv6 Host to Router Load Sharing, draft-ietf-ipv6-host-load-sharing-00.txt, Jan 4, 2002
[8] Kris Bubendorfer, Nomad: Towards an Architecture for Mobility in Large Scale Distributed Systems, PhD Thesis, Victoria University of Wellington, December 2001
[9] Syed Faraz Shamim, How Does Unequal Cost Path Load Balancing (Variance) Work in IGRP and EIGRP, http://www.cisco.com/warp/public/103/19.html, CISCO Systems
[10] Praveen Akkiraju, Kevin Delgadillo, Yakov Rekhter, Enabling Enterprise Multihoming with Cisco IOS Network Address Translation (NAT)
[11] Y. Rekhter, T. Li, A Border Gateway Protocol 4, RFC 1771
[12] T. Bates, R. Chandra, D. Katz, Y. Rekhter, Multiprotocol Extentions for BGP4, RFC 2283
[13] An Introduction to Interdomain Routing and BGP, Timothy G. Griffin, http://www.research.att.com/~griffin/interdomain.html
[14] ARIN, BGP Tutorial, Avi Freedman, http://avi.freedman.net
[15] An Application of the BGP Community Attribute in Multi-home Routing, E. Chen and T. Bates, RFC 1998
[16] Using the Border Gateway Protocol for Interdomain Routing, Cisco Systems, http://www.cisco.com/univercd/cc/td/doc/cisintwk/ics/icsbgp4.htm
[17] Scalable Simulation Framework Network Models, http://www.ssfnet.org/
[18] Jongwook Woo, Seunghyup Ryu, Jaiyong Lee, Jaehoon Chung, Jungsoo Park, "BGP Extension for Multihoming in IPv6", in *Proceedings of Eighteenth Korea Information Processing Society*, Nov. 2002