A Design of Architecture for Federating between NRNs and Determination Optimal Path

Jinhyung Park[1], Hyunhun Cho[1], Wonhyuk Lee[1], Seunghae Kim[1], Byoung-Ju Yun[2]

[1] KREONET Center, Korea Institute of Science and Technology Information
245 Daehakro, Yuseonggu, Daejeon, Republic of Korea
[e-mail: {ntoskr, hhcho, livezone, shkim}@kisti.re.kr]

[2] School of Electronics Engineering, IT College, Kyungpook National University
80 Daehakro, Bukgu, Deagu, Republic of Korea
[e-mail: bjisyun@ee.knu.ac.kr]

*Corresponding author: Byoung-Ju Yun

Received January 5, 2014; accepted January 21, 2014; published February 28, 2014

Abstract

The current networks do not disclose information about a management domain due to scalability, manageability and commercial reasons. Therefore, it is very hard to calculate an optimal path to the destination. Also, due to poor information sharing, if an error occurs in the intermediate path, it is very difficult to re-search the path and find the best path. Hence, to manage each domain more efficiently, an architecture with top-level path computation node which can obtain information of separate nodes are highly needed

This study aims to investigate a federation of a united network around NRN(National Research Network) that could allow resource sharing between countries and also independent resource management for each country. Considering first the aspects that can be accessed from the perspective of a national research network, ICE(Information Control Element) and GFO(Global Federation Organizer)-based architecture is designed as a top-level path computation element to support traffic engineering and applied to the multi-domain network. Then, the federation for the independent management of resources and resource information sharing among national research networks have been examined.

***Keywords*: Federation, GMPLS, Path Computation Element, GFO, ICE**

A preliminary version of this paper was presented at APIC-IST 2013, was selected as an outstanding paper.

http://dx.doi.org/10.3837/tiis.2014.02.0021

1. Introduction

Information & telecommunication user environment is changing rapidly and information networks in various sectors of our society have enabled more flexible and diverse big-data transfers. To accommodate dramatically increasing Internet traffic, optical communication technologies including Wavelength Division Multiplexing (WDM) have been developed as well.[1] This kind of technological advancement has made it possible for IP routers to provide tens of terabit bandwidth with a single optical fiber. As the dependence on optical transport network rapidly increases, and since the internet traffic is being congested with communication network components, it's been critical to keep the optical transport network alive.[2] It is very important for a network service provider to provide proper connectivity for end-to-end users through optimal routing between networks.[1][3]

The current networks do not disclose information about the management domain due to scalability, manageability and commercial reasons. Therefore, it is very hard to compute an optimal path to the destination. Also due to poor information sharing, if an error occurs in the intermediate path, it is very difficult to re-search the path and find the best path. Hence, to manage each domain more efficiently, an architecture of top-level path computation that obtains information of separate nodes are highly needed.

This study aimed to investigate the federation to build a united network led by the NRN for managing and sharing the network resources among countries. Considering first the aspects that can be accessed from the perspective of a national research network, ICE(Information Control Element) and GFO(Global Federation Organizer)-based architecture is designed as a top-level path computation element to support traffic engineering and applied to the multi-domain network. Then, the federation for the independent management of resources and resource information sharing among national research networks have been examined.

2. GMPLS-based Path Management

2.1 GMPLS Routing

In existing MPLS(Multi-Protocol Label Switching), LPS(Label Switched Path) layer relation among the same interfaces was provided. However, GMPLS(Generalized Multi-Protocol Label Switching) is able to set LSP layer relation among diverse interfaces as well as among the same interfaces. In addition, the routing protocols such as OSPF(Open Shortest Path First) and IS-IS(Intermediate System to Intermediate System) can be applied. [2][4]

One of the main goals of GMPLS is to set LPS, the optical channel that satisfies the specific link metric at path setting. Therefore, it should be able to transfer the constraint-related link state information as well as the general link state information. In terms of a way to extend routing protocols to support the GMPLS that includes these functions, there are OSPF, IS-IS and BGP(Border Gateway Protocol).

2.2 Path Computation Element (PCE)

In GMPLS, the PCE(Path Computation Element) can calculate topology- based network path or route. It refers to an object to that constrains for the calculation are applied.[5]

In a label-switched network such as MPLS and GMPLS, most traffic engineering solutions are operated in a single routing domain. These solutions stop when they get off the routing area from ingress node to egress node or escape the AS of ingress node. Under these situations, it is impossible to obtain complete routing information from the network.[6][7] The path computations can become complex because service providers are reluctant to disclose routing information for reasons such as scalability constraints or confidentiality concerns.[8]

Even though there are several methods to use PCEs in single domain, single layer, multi domains or in multi layers,[9][10] The PCE can be configurated so that it takes into account all layers that exist in a single domain which can be managed dynamically with a single domain. PCE is used in such way to find more efficient way for using network resources.

The specific reasons are as follows:

① CPU-intensive path computation

Considering economic parameters, the provider edge optimized node may not have sufficient capacity. If there is too much load, significant CPU computing power may be required for path computation. [10]

② Partial visibility

From the node that is responsbile for path computation to the destination node, there could be various cases where network topology is unknown. In this case, a loose path is available. However, optimal path setting is not guaranteed.

③ Absence of TED(Traffic Engineering Database) / Use of IGP(Interior Gateway Protocol) that is hard to apply TE(Traffic Engineering)

TED is a huge information storage where resources of network node are stored and it requires large amount of memory and CPU power. Therefore, it is effective for each PCE to manage individual TED by dispersing TED.

④ Backup path computation to protect bandwidth

PCE can calculate faster reroute backup path for protecting TE LSP.

⑤ Path selection policy

Each PCE has a local policy that can have an effect on path computation and path selection.

2.3 Multi-domain Path Computation

The biggest duty and goal of PCE are to solve problems associated with multi-domain connection setting.[7] For multi-domain path computation, the following methods are usually used:

A. Per-domain path computation

The per-domain path computation starts at the entry point as shown in **Fig.1**. Path computation is performed within the domain and a good-enough path is selected after taking into account the destination. [11] This process is conducted in assumption that the connection between domains is known. When selecting the exit point, the computation uses the same information used in IP routing.

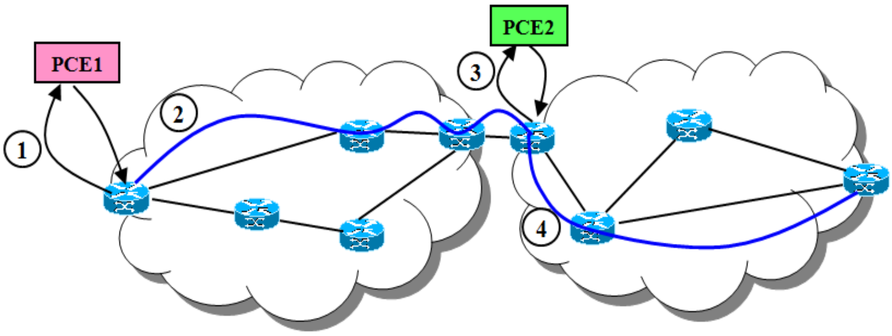

Fig. 1. Per-domain Path Computation

However, PCE1 doesn't know where destination is situated in the per-domain path computation method. To solve this kind of disadvantage of the per-domain path computation, therefore, a crankback-like signaling method can be adopted.[9][12] This method is relatively more complicated and takes more time.

B. Simple cooperating PCEs

The simple cooperating PCEs is a method used for communicating between PCEs and finding an optimized path among adjacent domains.

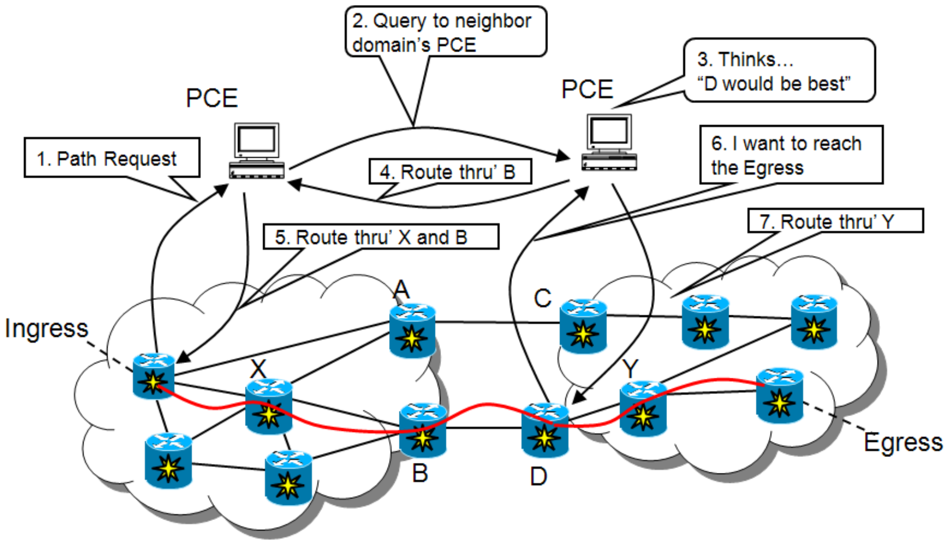

Fig. 2. Example of Simple Cooperating PCEs

First, if the path computation is requested to the PCE by the ingress node of the source, PCE does not select a path with its domain but query is performed on the path selection against the PCE that controls neighbor domain.[13][14] In the PCE of neighbor node, an optimal path in its domain is selected and returned to the neighbor egress.

In the simple cooperating PCEs method, however, if more than two domains are in a series, it becomes very complicated.[9] In fact, it is unable to provide an optimized path. In other words, the optimized path may not be a real optimized path for each domain.

C. Backward recursive path computation (BRPC)

BRPC is a method used for cooperating between the PCEs and it provides an optimal path using crankback signaling in the absence of full visibility of the network.[15] In this method, path computation is performed from destination domain and deliveres a set of configurable path to the neighbor domain. Based on this process, each PCE computes the configurated optimal path from entry point to exit point and forms the path tree based on the destination. However BRPC requires an assumption that destination domain must be known. [16][17]

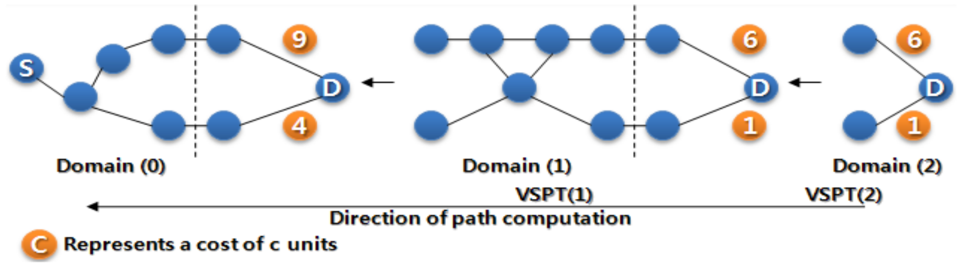

Fig. 3. Backward Path Computation

3. GFO Model

In this section, the architecture of a top-level path computation node, GFO(Global Federation Organizer) and ICE(Information Control Element) related GFO are proposed.

3.1 GFO / ICE Terminology

ICE that collects node information and controls each domain for path computation in the multi-domain environment and the top-level path computation node GFO architecture-based network that can obtain all information among ICEs are shown in Fig.4.

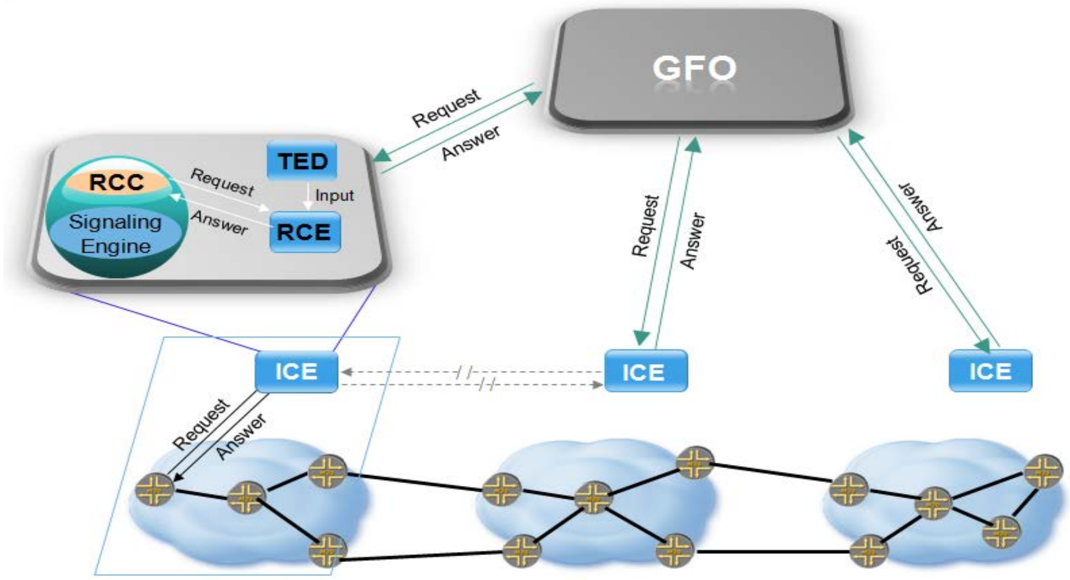

Fig. 4. GFO/ICE Architecture

3.2 GFO Architecture

The elements of GFO architecture can be defined as follows.

① IIR(ICE Information Repository)

IIR stores basic domain information in each ICE and sends the informations reqired at path computation such as ingress/egress node information and network performance elements information to ICE in each domain. Then, IIR requests to make ICE to compute paths in total number of cases within the domain. IIR only stores minimum amount of information ofr path computation.

② IMS(ICE Management System)

IMS registers, deletes and corrects ICE informations in IIR.

③ PCP(Path Calculation Processor)

PCP combines the path result data set in the domain that IIR requests to each ICE and computes an optimal path. Then, it sends the results to ICE that requests path computation.

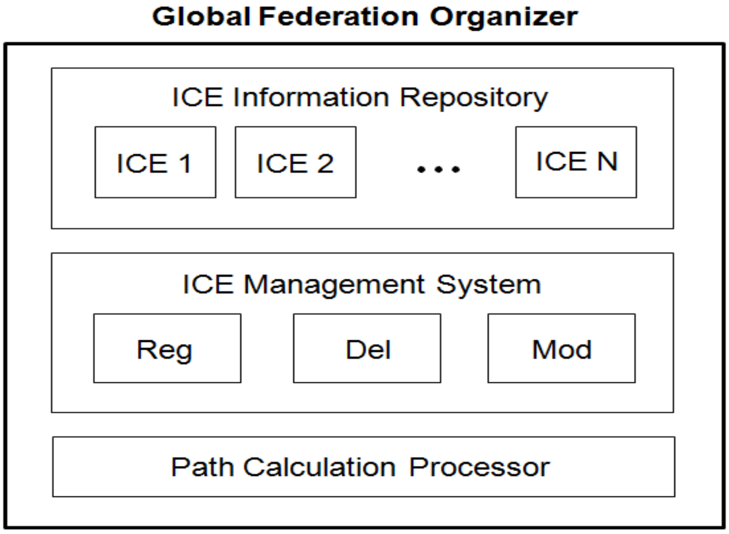

Fig. 5. Architecture of GFO

3.3 ICE Architecture

The elements of ICE architecture can be defined as follows:

① RCC(Route Computation Client)

If route initialization is requested by a user or a certain application, RCC requests the routing information to ICE. In general, this refers to the edge/ingress node of network.

② RCE(Route Computation Element)

After the IIR in GFO requests path computaton to ICE, the RCE in ICE computes the ingress/egress node information and network performance information based on all paths from IIR and sends the results to IIR.

③ TED(Traffic Engineering Database)

TED is created by the domain resources and network topology information. It includes bandwidth, delay and jitter, etc. Using these data, ICE selects an optimal path that satisfies the requested needs.

④ Node/Device Information Database

It stores basic network connection status information in the domain for fast path computation and updates on a regular basis

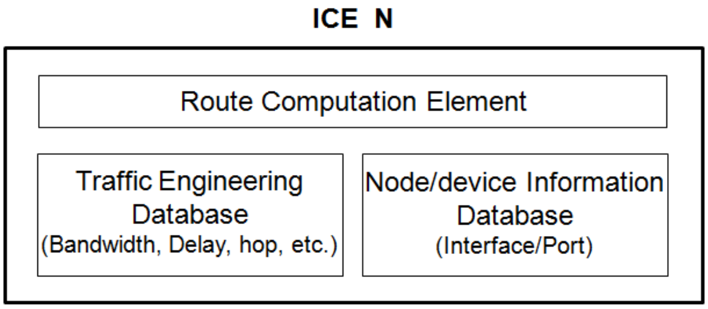

Fig. 6. Architecture of ICE

3.4 GFO-based Multi-domain Path Computation

In order to analyze the GFO/ICE application on multi-domain path selection during data transfers, this section will discuss the chatacteristics of research fields and explain alternative ways of GFO/ICE application on multi-domain different characteristics.

In case of general data transfer, this would not be the main issue caused by Qos because if the packet is transferrd by TCP/IP, the packet that is lost due to the characteristics of transfer protocol will be retransferred. However transfer delay and delay jitter in multi-network services become a critical problems in voice communication, live streaming or realtime communication services. Therefore, QoS must be guaranteed in these services.

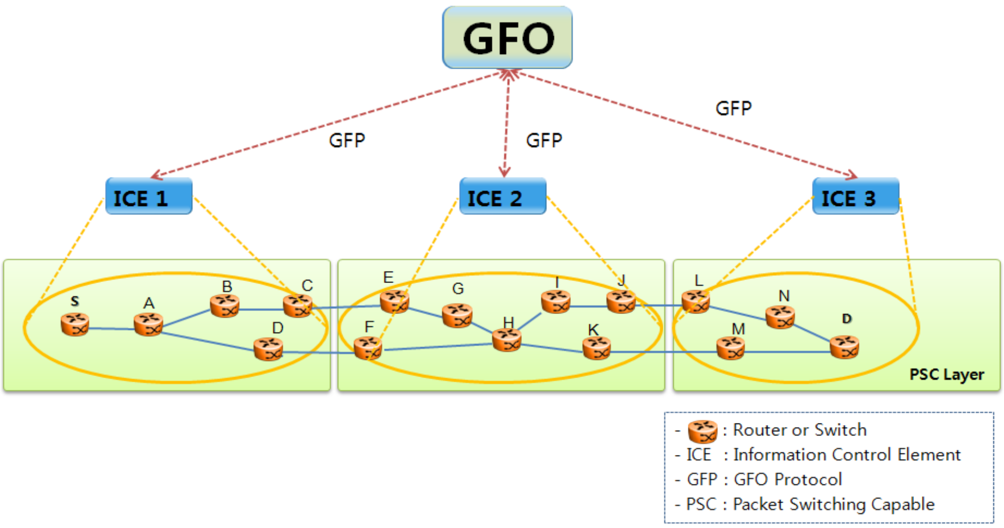

Fig. 7.Example of Multi-domain Path Calculation

Fig. 7 shows GFO/ICE configuration for optimal path computation by the characteristics of data transfer in multi-domain environment.

First, ICE1 makes a request to GFO from the source router ‘S’ to the destination router ‘D’ and then to GFO. Then, GFO sends the information needed for optimal path computation(ex: Destination information, etc.) to ICE 1, 2 and 3. It calculates all possible path results among ingress/egress nodes in each domain and sends them to GFO. Since GFO knows that the destination router ‘D’ belongs to the domain and the path management system on this domain is ICE3, GFO computes based on the path computation results that are provided by ICE3 and calculated by the backward recursive method.

Based on the path result information required for path computation by all ICEs, GFO computes all possible paths. Then, GFO selects a taking into account the cost and delay among the all possible paths depending on selection of parameters by the characteristics of data transfer.

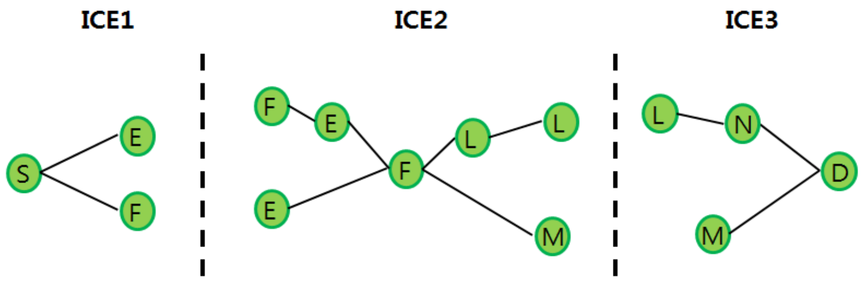

Fig. 8. PSC Layer Information

As shown in **Fig. 8**, the cost and delay in a multi domain can be calculated by combining path set of ICE1, ICE2 and ICE3, and the results are shown in **Fig. 9**.

- GFO calculates the results received from the ICE1, 2, 3
- GFO returns optimal path to ICE1

- (SABCE)	(EGHIJL)	(LND)	→ **(S...E...L...D)**
(5, 7)	(12, 8)	(5, 1)	→ **(22, 16)**
- (SABCE)	(EGHKM)	(MD)	→ **(S...E...M...D)**
(5, 7)	(9, 6)	(7, 2)	→ **(21, 15)**
- (SADF)	(FHIJL)	(LND)	→ **(S...F...L...D)**
(7, 3)	(8, 15)	(5, 1)	→ **(20, 19)**
- (SADF)	(FHKM)	(MD, 4)	→ **(S...F...M...D)**
(7, 3)	(5, 17)	(7, 2)	→ **(19, 22)**

- GFO choose a path with Min(cost or delay or hop) by data characteristics

Fig. 9. Final Path Computation in GFO

Fig. 10. shows the path selected in consideration of cost(throughout) as the transfer parameter among path groups are calculated on the PSC layer.

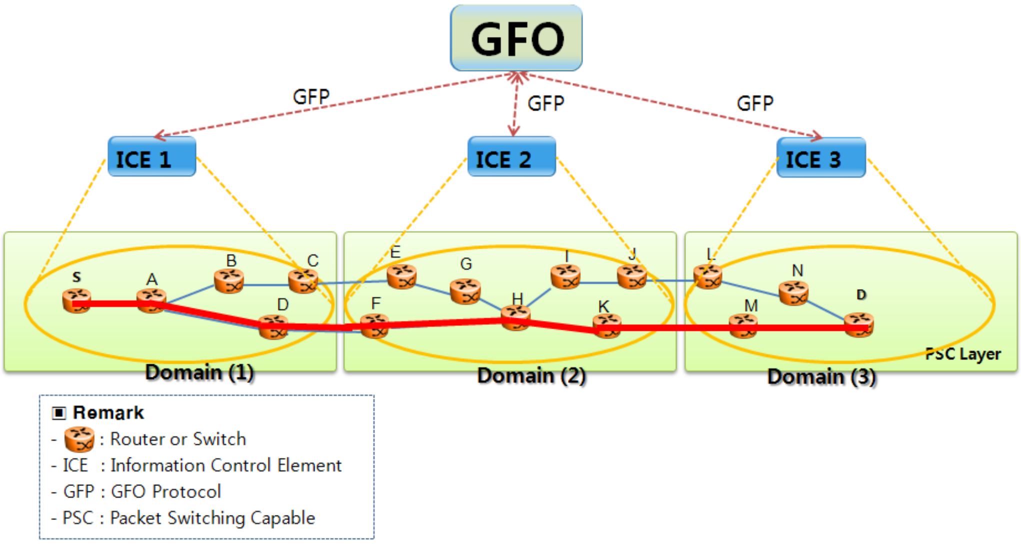

Fig. 10. Path selection by throughput

Fig. 11. shows the results on the PSC layer that is selected by considering delay as the transfer parameter.

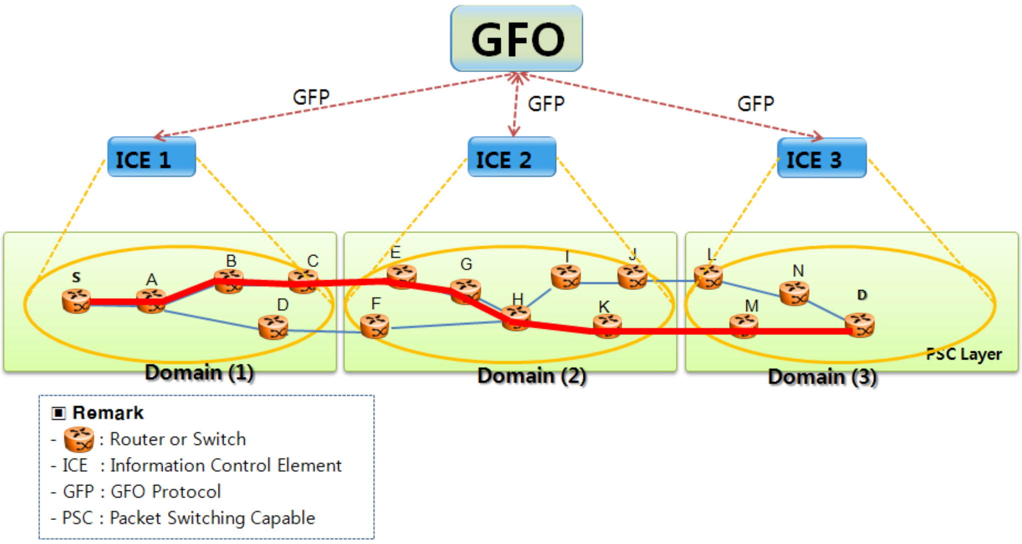

Fig. 11. Path selection by delay

4. Performance Evaluation

To verify an optimal path computation model by the transfer characteristics, a testbed has been configured as shown in **Fig. 12**. Actually, it is a single domain with a title 'KREONET.' However, the multi-domain network environment has been configured by classifying it into logical domain by each zone.

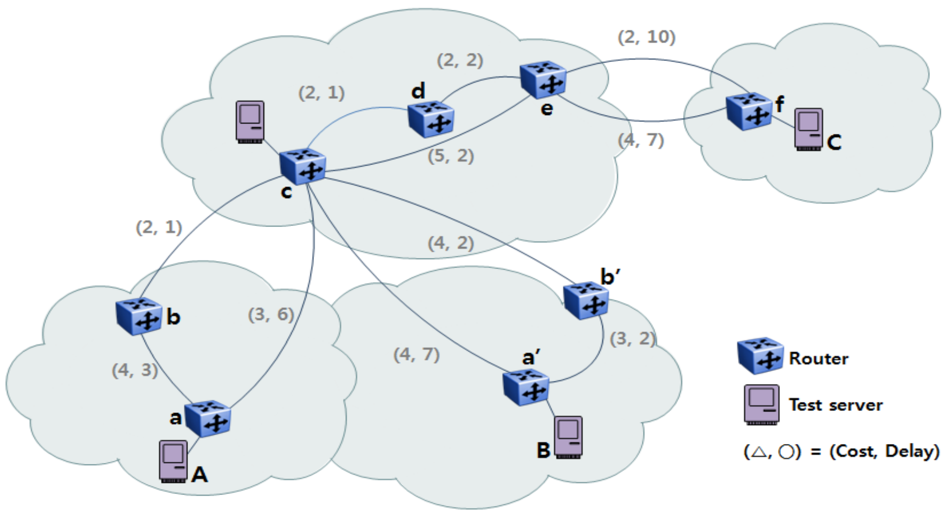

Fig. 12. GFO/ICE testbed

A testbed was implemented based on NRN as shown above and two cases were selected among currently active research fields and tested. In other words, a test was performed on 'transfer of big data' that was limited by transfer speed and bandwidth and 'transfer of high-resolution images(HD or higher)' that is sensitive to delay and jitter.

To set-up a multi-domain environment, as in **Fig.12**, the testbed was configurated so that there are 4 domains linked with 1Gbps and 10Gbps and test server starting points as A and B. The optimal path to the final destination, Test Server C which is linked to Seattle PoP, were set-up with the minimum hop-based method and with the newly-proposed method. Then, the transfer speed and delay(jitter) of two methods were compared and analyzed.

The marks (△,○) in **Fig. 12** shows the cost and delay of each path in relative values. The data transfer where transfer rates and bandwidth are relatively more important than delay such as in 'transfer of large-volume data,' transfer from the source 'A' to the destination 'C' has been tested. The data transfer where transfer delay such as 'realtime transfer of HD (or higher level) media' is more important, transfer from the source 'B' to the destination 'C' has been analyzed in assumption that there would be no changes in network state such as dramatic increase of traffic.

Then, the following results were obtained.

Table. 1. Test result

		Optimal Path	Throughput	Packet loss rate	Jitter (Image, Voice)
TEST1	path1 (min hops)	*A-a-c-e-f-C*	*31.4MB/s*	-	-
	path2 (cost)	*A-a-c-d-e-f-C*	*38.1MB/s*	-	-
TEST2	path3 (min hops)	*B-a'-c-e-f-C*	-	*3.0%*	*13m/s, 12m/s*
	path4 (delay)	*B-a'-c-d-e-f-C*	-	*0.0%*	*3m/s, 3m/s*

The path1 and 3 in table.1 represents optimal paths based on the minimum hop while path2,4 shows optimal paths based on cost and delay. During the data transfer, the transfer speed in the cost-considered path(path2) was improved by about 20% compared to that in the minimum hops-considered path(path1). In terms of the realtime transfer of image data, packet loss decreased from 4,000 to 160 in the delay jitter-considered path(path4), compared to the minimum hops-considered path(path3). In terms of jitter as well, video and audio jitters were about 12*ms* and 3*ms*, showing appropriate network conditions for realtime data transfer.

5. Conclusion

At present, a network can only be configurated by each managemet domain due to scalability, manageability and commercial reasons. This management domain consists of IGP domain and AS(Autonomous System).[10] The information in each domain is kept confidential with security and policy reasons. Even though the trust-based path information are shared through mutual confidentiality policy among ASs, most users are unable to get complete end-to-end path information.[11] [18] Therefore, it is extremely difficult to compute an optimal path to final destination.

Considering these aspects, this study proposes an architecture and protocol of ICE that controls information in each domain and a GFO that handles the ICE. GFO combines the results of path computation performed by ICE in each domain and computes an optimal path. In this method, CPU processing overhead can be reduced by spreading excessive processes.

After designing ICE and GFO-based structure by considering the national research network environment and applying it to the multi-domain network, the following matters have been conducted; efficient use of resources, selection of optimal path among multi-domains and optimization of end-to-end connection. By establishing a logical multi-domain enviornment testbed which includes specific sector for “large-data transfer’ and ‘real-time media transfer’ and also by designing an infractructure that could select optimal path with minimum information transfer between nodes, we created an optimal path computation model that could not only use network resources more efficiently but also offer much faster network service.

Moreover, there have been studies on network linkage and operation for efficient resource management under international collaborative studies between national research networks such as GLORIAD and GEANT. Thus, we can expect more diverse studies on network linkage and operation that are related to the findings of this study.

References

[1] Hyuncheol Kim, "A Study of Survivable Traffic Grooming in Broadband Convergence Networks," *Thesis of Doctoral course*, Sungkyunkwan Univ., 2005. Article (CrossRef Link).

[2] A. Bonerjee, J. Drake, J. P. Lang, and B. Turner, "Generalized Multiprotocol Label Switching: an Overview of Routing and Management Enhancements," *IEEE Commun. Magazine*, vol. 39, no. 1, pp. 144-150, Jan. 2001. Article (CrossRefLink).

[3] A. Shaikh and K. Shin, "Destination-driven Routing for Low-cost Multicast," *IEEE Journal on Selected Areas in Communications*, vol. 15, No 3, Apr. 1997. Article (CrossRefLink).

[4] I. Nishioka, S. Ishida, and Y. Iizawa, "End-to-end path routing with PCEs in multi-domain GMPLS networks," *IPOP2008*, July. 2008. Article (CrossRef Link).

[5] Sophie De Maesschalck, etal., "Intelligent Optical Networking for Multilayer Survivability," *IEEE Communications Magazine*, pp.42-49, Jan. 2002. Article (CrossRefLink).

[6] N. Bouabdallah, G. Pujolle, E. Dotaro, N. Le Sauze and L. Ciavaglia, "Distributed Aggregation in All-Optical Wavelength Routed Networks," in *Proc. Of 2004 IEEE International Conference on Communications*, vol. 3, pp. 1806-1810, June. 2004. Article (CrossRef Link).

[7] Brehon, Yannick, and Daniel Kofman. "Bus-label switched paths, an approach to reduce the cost of multilayer networks," in *Proc. Of 2006. ICC'06. IEEE International Conference on Communications*. Vol. 5. IEEE, 2006. Article (CrossRef Link).

[8] Cha, Meeyoung, et al. "Path Protection Routing with SRLG Constraints to Support IPTV in WDM Mesh Networks," INFOCOM 2006. in *Proc. Of 25th IEEE International Conference on Computer Communications*, pp.1-5, 2006. Article (CrossRef Link).

[9] R. Aggarwal, D. Papademitriou, and S. Yasukawa, "Extensions to RSVP TE for Point to Multipoint TE LSPs," Work in Progress, 2006. Article (CrossRef Link).

[10] Y. Seok, Y. Lee, Y. Choi, and C. Kim, "Explicit Multicast Routing Algorithms for Constrained Traffic Engineering," in *Proc. Of 2002 Proceedings of the 7th ISCC*, pp. 455-461, July. 2002. Article (CrossRef Link).

[11] J.-L. L. Roux et al. "OSPF Protocol Extensions for Path Computation Element (PCE) Discovery," RFC 5088, Jan. 2008. Article (CrossRef Link).

[12] Eiji Oki, Ichiro Inoue, Kohei Shiomoto, "Path Computation Element (PCE)-based Traffic Engineering in MPLS and GMPLS networks," in *Proc. Of Sarnoff Symposium*, 2007 IEEE, Apr. 2007. Article (CrossRef Link).

[13] P. Pan, G. Swallow, and A. Atlas, "Fast Reroute Extensions to RSVP-TE for LSP Tunnels," Internet Draft, draft-ietf-mpls-rsvp-fastreroute-07.txt, 2005. Article (CrossRef Link).

[14] E. Rosen, and Y. Rekhter, "BGP/MPLS IP VPNs," internet draft, draft-ietf-13vpn-rfc2547bis-03.txt, Oct. 2004. Article (CrossRef Link). http://tools.ietf.org/html/rfc4364

[15] K. Shiomoto, E. Oki, D. Shimazaki, and T. Miyamura, "Multilayer Traffic Engineering Experiments in MPLS/GMPLS Networks", IEEE BROADNETS 2007, Sep. 2007. Article (CrossRef Link).

[16] Sebastian Gunreben, Franz Rambach, "Assessment and Performance Evaluation of PCE-based Inter-Layer Traffic Engineering," in *Proc. Of ONDM conference*, 2008. Article (CrossRef Link).

[17] Richard Douville, Jean-Louis Le Roux, Stefano Secci, "A Service Plane over the PCE Architecture for Automatic Multidomain Connection-Oritented Services," IEEE Communications Magazine, Jun 2008. Article (CrossRefLink).

[18] Oliver Yu, "Inter carrier inter domain control plane for global optical networks," in *Proc. Of IEEE International Conference on Communications*, Volume 3, pp.1679–1683, Jun. 2004. Article (CrossRef Link).

Jinhyung Park is an Senior Researcher of KREONET Center at The Korea Institute of Science and Technology Information(KISTI). He received his M.S. and Ph.D. degrees in Communication Engineering from the Kyungpook National University, in 2009 and 2013, respectively. From 2004 to 2007, he has been with Fusionsoft Inc., where he worked as a senior engineer. His current research focuses on Future Internet and Software Defined Network, Large-scale Research Networking.

Byoung Ju Yun received the B.S. degree in electronics engineering from Kyungpook National University, Korea, in 1993 and the M.S. and Ph.D. degrees in electrical engineering and computer science from KAIST (Korea Advanced Institute of Science and Technology), Korea, 1996 and 2002, respectively. From 1996 to May 2003, he has been with Hynix Semiconductor Inc., where he worked as a senior engineer. From June 2003 to February 2005, he has been with the Center for Next Generation Information Technology Kyungpook National University, where he worked as assistant professor. Since March 2005, he has been with the school of Electronics Engineering, where he works as an invited professor. His current research interests include image processing, color consistency, multimedia communication system, HDR color image enhancement, and HCI.