Optical BGP Routing Convergence in Lightpath Failure of Optical Internet

Sangjin Jeong, Chan-HyunYoun, Minho Kang, Kyoung-Seon Min
Hyun Ha Hong, and Hae Geun Kim

Optical Border Gateway Protocol (OBGP) is an extension to BGP for Optical Cross Connects (OXCs) to automatically setup multiple direct optical lightpaths between many different autonomous domains. With OBGP, the routing component of a network may be distributed to the edge of the network while the packet classification and forwarding is done in the core. However, it is necessary to analyze the stable convergence functions of OBGP in case of lightpath failures. In this paper, we first describe the architecture of the OBGP model and analyze the potential problems of OBGP, e.g., virtual BGP router convergence behavior in the presence of lightpath failure. We then propose an OBGP convergence model derived from an inter-AS (Autonomous System) relationship. The evaluation results show that the proposed model can be used for a stable OBGP routing policy and OBGP routing convergence under lightpath failures of the optical Internet.

Manuscript received Sept. 21, 2001; revised Mar. 11, 2002.

Sangjin Jeong (phone: +82 42 866 6186, e-mail: sjjeong@icu.ac.kr), Chan-Hyun Youn (e-mail: chyoun@icu.ac.kr), and Minho Kang (e-mail: mhkang@icu.ac.kr) are with the Information and Communications University, Daejeon, Korea.

Kyoung Seon Min (e-mail: minks@kt.co.kr) is with Korea Telecom, Daejeon, Korea.

Hyun Ha Hong (e-mail: hhhong@etri.re.kr) and Hae Geun Kim (e-mail: hgkim@etri.re.kr) are with ETRI, Daejeon, Korea.

I. INTRODUCTION

Network architecture for high-spe ed IP backbones has recently been studied extensively because of the emergence of next generation Internet. The growth, performance, and survivability requirements of the Internet are mandating that IP networks be cost effective, survivable, and scalable, and also provide control capabilities that facilitate network performance optimization. Some of these requirements are being addressed by the Internet Engineering Task Force (IETF) under their Multiprotocol Label Switching (MPLS) traffic engineering working group [1], [3]. The MPLS traffic engineering control plane is a synthesis of IP traffic engineering enabled by MPLS and the conventional IP network layer control plane. The MPLS traffic engineering control plane includes components such as resource discovery, path selection dissemination, and path management. In order to provide inter-domain connectivity, a label is assigned to every prefix in the IP routing table of a router acting as the MPLS Label Switch Router (LSR), the only exception being routes learned through the Border Gateway Protocol (BGP) [4]. No label is assigned to the BGP next hop to forward packets toward BGP destinations. Where the customer controls the optical routing, using exterior routing protocols may be possible.

Currently, optical networks are primarily used for the interconnection of large network domains such as enterprise networks, Internet Service Providers (ISPs), and GigaPOPs. Most of these networks already use BGP to manage the interconnection of their respective networks. More importantly, these large enterprise customers and ISPs are likely to be the first to use dark fiber and operate Dense Wavelength Division Multiplexing (DWDM) networks. Therefore, routing protocols for inter-

domain networking might also be useful for interconnecting optical networks.

The Optical Border Gateway Protocol (OBGP) is an extension to BGP for manipulating Optical Cross Connects (OXCs) to permit them to be automatically setup and configured as BGP speaking devices to support multiple direct optical lightpaths among many different Autonomous Systems (ASs). With OBGP, the routing component of a network may be distributed to the edge of the network, while the packet classification and forwarding is done in the core. OBGP also allows customers at the edge to control a subset of lightpaths within another network's wavelength cloud, so that customers can manage their own lightpath routing within that cloud. With the large number of adjacencies possible using OBGP, lightpaths themselves may be used as a direct peering and transit mechanism between consenting ISPs. The proposed protocol extensions allow carrier free networks, where the customers at the edge control and route lightpaths directly across an optical wavelength cloud [1], [2].

These architectural approaches for customer empowered networks may require a fundamentally different architecture from the traditional ones. For example, caching and multihoming can provide better reliability than fast restoration and protection on individual optical links for enterprise customers. Interconnection and direct peering also allow the enterprise or small ISP network to bypass the traditional hierarchical carriers and ISPs to establish direct peering with destination ISPs. One possible solution is to treat each OXC as a direct path between a pair of OBGP speakers. However, this significantly increases the session complexity of the OBGP, particularly for multiple parallel lightpaths. The alternative solution is to treat each OXC as an independent virtual BGP router with one input port and one output port. A virtual BGP router can then be set up for each OXC and separate OBGP sessions are initiated with peers of the virtual BGP router. This approach is much more scalable because each virtual BGP router configuration can be easily cloned from other virtual BGP routers.

The key for OBGP to scale to a very large network lies in the stability of inter-AS routing. If AS paths fluctuate frequently— a phenomenon called route flapping [8]—then the virtual BGP routers spend a great deal of time to update their routing tables and to propagate the routing changes. Unstable inter-AS routing can cause unstable end-to-end routing [5]. The analysis of both the topology of the routing system and the instability of the routing system is important to evaluate network performance between end users. Since the customers of each network are connected to domains and provided various services, the instability of domains can severely affect the performance of their customers.

In this paper, we discuss the instability in OBGP and propose a routing convergence model to effectively reduce the instability in the virtual BGP router. After describing the architecture of the OBGP model in section II, we analyze, in section III, the potential problem of a virtual BGP router in case of OXC failure and propose requirements for a virtual BGP router to prevent potential routing problems. In section IV, we describe the OBGP convergence policy model and present an experimental analysis. Finally, in section V, we describe the applicability of our proposed model.

II. ARCHITECTURE OF THE OBGP MODEL

There are two approaches for inter-domain optical networking, BGP/GMPLS [6] and OBGP [3], [7].

The Generalized Multiprotocol Label Switch (GMPLS) architecture extends MPLS signaling protocols to circuit-switched networks. In this fashion, it enables functions that transform the optical transport network into an automatically switched transport network. Current research on GMPLS has concentrated on the setup and management of Label Switched Paths (LSPs) within a single administrative domain (i.e., in the intra-domain context). However, in reality, many end-to-end LSPs span multiple service provider networks, and therefore we require inter-domain signaling, for example, the carriers provide this function as a service to ISPs, global enterprises, and so on [6].

OBGP is intended to allow customers to control the routing of their lightpaths through another entity's optical wavelength cloud, as an overlay to an interior wavelength management protocol. For example, a carrier may have a large managed wavelength cloud, but rather than hiding the routing of the wavelengths from the customer, the customer may be given a limited view of the network topology or a choice of possible routes which are subsets of all possible routes. In addition, the customer may have optical routes transiting two separate carrier networks and may wish to interconnect its routes through these clouds at some mid point. As a consequence, the customer's ideal optical wavelength topology may vary according to the ideal optimized topology of the individual carrier networks. OBGP allows the customer's topology to take precedence over the carrier's preferred topology (Fig. 1). Large single domain wavelength clouds become unmanageable and are too difficult to optimize for traffic engineering purposes as a large single domain. The common solution is to break those single domains into many small domains that can be individually optimized. In each such domain, OBGP could be used as a more modest optimization mechanism [3], [7].

OBGP requires a simple OXC switch as depicted in Fig. 2. OBGP routers with multiple paths in the OXC path are given preference over any path that goes through an electrical for-

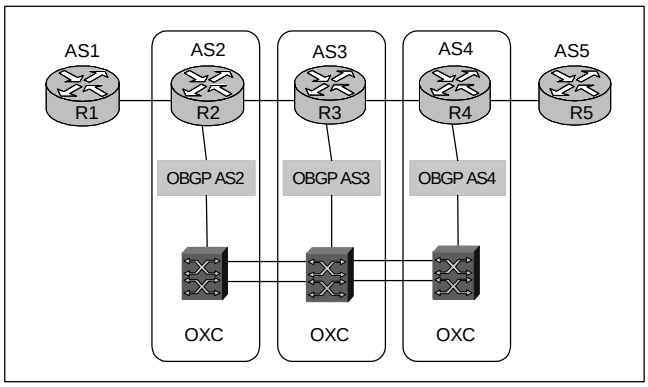

Fig. 1. OBGP configuration in optical Internet.

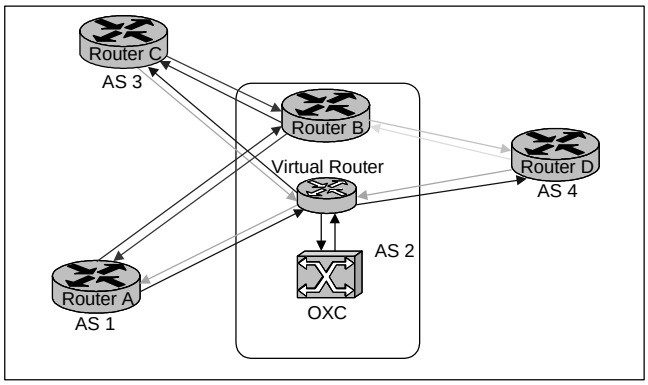

Fig. 2. Virtual router for OBGP.

warding engine using standard BGP techniques for selecting the shortest AS path, local preferences, and such. To multiplex and demultiplex wavelengths, Router B must use optical filters to separate out the individual wavelength. By using a simple optical switch, the individual light path can in effect be treated as an alternative path to router B. There are two ways to configure Router B. One is to treat each OXC as a direct path between a pair of BGP speakers. However, this significantly increases the complexity of any single BGP session, particularly for many parallel lightpaths. Another is to treat each OXC as an independent virtual BGP router with only one input port and one output port. A virtual BGP router can then be set up for each OXC and separate BGP sessions initiated with its peers. This approach is much more scalable because each virtual BGP router configuration can be easily cloned from other virtual BGP routers.

To date there has been little effort spent on addressing the requirement for configuring, setting up and managing wavelengths between domains, and allowing enterprises to manage their own wavelength configuration across a wavelength cloud. The conventional solution to date is for a carrier to operate a wavelength cloud and offer a managed lightpath service to the customers at the edge (Fig. 2).

A number of mechanisms have been proposed for the management and control of such wavelength cloud systems [3]. Most of these systems have been designed on variations of link state interior routing protocols, such as OSPF (Open Shortest Path First), IS-IS (Intermediate System to Intermediate System), and PNNI (Private Network to Network Interface) or complementary extensions of MPLS, such as GMPLS. For complex single domain networks, these protocols allow the optimized configuration and establishment of lightpaths. Because these networks provide a common carrier to many downstream customers, the networks require survivable and fast restorable lightpaths. An attribute of these networks is the capability to initiate and route end-to-end optical channels in near real time and to provide capabilities that enhance network survivability [3].

The main operation of OBGP consists of two phases. The first phase is the lightpath reachability phase. During this phase, sites advertise the availability of the optical lightpath to their sites through BGP. These announcements contain information on the OXC and the available lightpath through the OXC. The information is encoded using multi-protocol BGP extensions and extended community. This first phase allows sites to build up a lightpath Routing Information Base (RIB) that is used to determine if a lightpath is available across a number of OXCs in different sites. The second phase is the lightpath establishment. This phase uses the information received from the lightpath reachability phase and then uses a BGP UPDATE message to communicate the lightpath establishment to the OXC sites on the path. The information is encoded using BGP extensions and extended community.

III. POTENTIAL PROBLEMS OF OBGP WITH A VIRTUAL ROUTER

The basic concept of virtual BGP routers is to treat each individual OXC as a separate BGP router. The virtual router advertises itself independently of Router B in Fig. 2 with its own loopback address and its own set of IP addresses for its interfaces. Contrary to a normal BGP multi-router configuration, the virtual BGP router does not establish any Internal BGP (IBGP) connectivity even though it is within Router B's AS. It acts and behaves as an independent router by carrying its own set of routes, metrics, and such. The use of a virtual router for each OXC allows us to use standard BGP routing with no modifications being necessary to support optical lightpaths. In fact, the virtual BGP router assigns its own private (or public) AS such that AS path metrics are used for basic traffic engineering.

By instantiating a virtual BGP router, at first the owner of the OXC can establish OXCs between neighbors that reduce the load on its electrical forwarding engine. Over time it can recon-

figure the virtual BGP router to interconnect with other neighbors if traffic patterns change. More importantly, the owner of the OXC can establish OXCs between neighbors automatically. More intriguingly, the virtual BGP router can also be easily reassigned into other routers' AS domains. The main purpose of the OBGP OXC is to announce routes, perform route filtering, and classify and provide standard BGP traffic network engineering capabilities to OBGP peers. As there is only one input port and one output port, there is no need to create a forwarding table within the OXC.

The loopback address for the virtual router used for OBGP connectivity is not the same as the data forwarding address of the OXC. As such, under normal circumstances the virtual router is not aware of any failures on the optical cross connect link between Routers A and C (Fig. 2). Therefore, if the interface card on either Router A or C detects a link failure, it immediately terminates the OBGP session with the neighbor – the virtual router. Either router A or C can terminate the session by sending an OBGP NOTIFICATION message to the virtual router. The virtual router then updates its routing information database and sends Network Layer Reachability Information (NLRI) UPDATE messages to the other edge router indicating that the those addresses are unreachable across the OXC. Once the problem has been cleared, the router tries to re-establish the link across the OXC. To re-establish the link, three routers can re-initiate the OBGP sessions between the virtual router and routers A and C. The re-initiation of the BGP sessions starts immediately after the receipt of the NOTIFICAION message, even before the link failure has been cleared.

These processes demand that OBGP updates for topology information be exchanged among domains. OBGP updates contain the reachability information for destination IP address prefixes. Each OBGP UPDATE message can be classified into two types: route announcements and route withdrawal. A route announcement discloses that a router has either learned of a new network attachment or has made a policy decision to prefer another route to a network destination. A route withdrawal is sent when a router makes a new local decision that a network is no longer reachable. Furthermore, each OBGP update message contains an AS_PATH list for storing traversed domains. Each AS_PATH list identifies by route domains traversed [4]. Thus, the OBGP operation maintains network reachability when the virtual router is not aware of any failures on the optical cross connect link. This can give rise to routing instability in the OBGP by generating excessive update messages when there is a failure of the virtual router or OXCs.

1. Instability Propagation

Internet routing instability is an important problem currently facing the Internet engineering community. High levels of network instability can lead to packet loss, increased network latency, and time to convergence. At the extreme, high levels of routing instability have led to the loss of internal connectivity in wide-area networks [8], [18].

Since OBGP is not deployed in the global Internet, it is hard to experimentally analyze the behavior of OBGP. However, since OBGP is an extension to BGP, it is sufficient to evaluate BGP for identifying OBGP behavior [3], [7]. Therefore, before investigating the potential problems of OBGP, we present an inter-domain routing instability analysis of the domestic Internet in 1999 [9]. This analysis is based on the conventional BGP data from January 1999, which was collected in the major Internet eXchange Point (IXP) of the domestic network.

According to the standard, BGP generates routing update messages when the network topology or policy is changed [4]. Frequent generation of BGP UPDATE messages about a particular IP prefix means that the IP prefix is unstable.

As we explained earlier, BGP UPDATE messages are of two types, prefix announcement and prefix withdrawal. Figure 3 shows the cumulative distribution of IP prefix update counts. The vertical axis denotes the fraction of the total number of IP prefixes during the measurement period and the horizontal axis denotes the number of IP prefix updates for each IP prefix. In the figure, about 85% of the total IP prefixes are announced less than 20 times, but about 97% of the total IP prefixes are withdrawn less than 20 times. Thus, we can conclude that the number of prefix announcements per IP prefix is more than the number of prefix withdrawals. As for prefix announcements, 78.6% of the total IP prefixes are less than or equal to 10 announcements during the measurement period. As for prefix withdrawals, 90.3% of the total IP prefixes are less than or equal to 10 withdrawals.

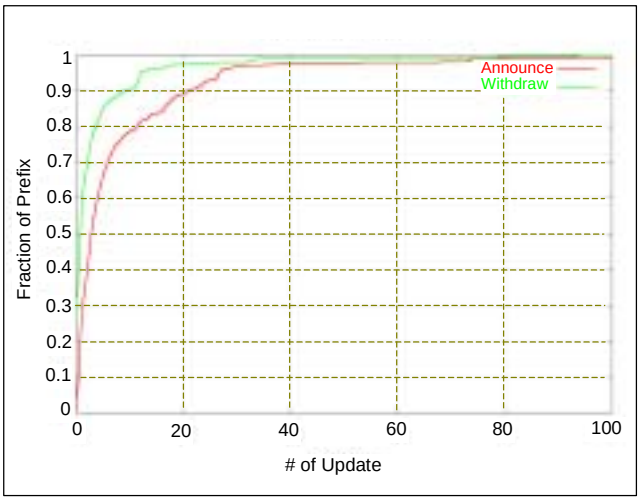

Fig. 3. Cumulative distribution of IP prefix update counts for each prefix.

According to the above results, we can conclude that most routes in most domains are stable. But some IP prefixes generated over 100 BGP update messages during measurement; this contributed 1.1% of the total BGP update messages. This means that some domains are very unstable. Thus, reducing the instability of these domains can decrease the overall instability. From this analysis, we can infer that some ASs showing BGP routing instability in the inter-domain network can be the origin or the propagation of BGP routing instability.

To analyze the propagation behavior of OBGP routing instability, we consider the instability propagation scheme based on the inter-AS relationship suggested by Gao [10]. To reveal the AS relationships in the Internet, Gao classified the inter-AS relationship as follows.

Two ASs that exchange traffic have a customer-to-provider, provider-to-customer, or peer-to-peer relationship.

Customer-to-provider or *provider-to-customer* relationship: the customer typically belongs to a smaller administrative domain that pays a larger administrative domain for access to the rest of the Internet. The provider is an AS that belongs to the larger administrative domain.

Peer-to-peer relationship: the two peers typically belong to administrative domains of comparable size and find it mutually advantageous to exchange traffic between their respective customers.

Since the virtual BGP router inherits the characteristics of BGP, we can claim that the behavior of the virtual BGP router is strongly coupled with the BGP router. Therefore, by analyzing the behavior of BGP, we can infer the behavior of the virtual BGP router, i.e., OBGP.

To analyze the instability distribution, we use Gao's inter-AS relationship inference algorithm to identify the customer-to-provider and peer-to-peer relation among ASs. Using these two algorithms, we extracted the AS relationships from the AS_PATH field in BGP UPDATE messages.

Our analysis is based on BGP data collected in the global Internet for two years from 1998. These data are provided by NLANR [11], [12].

Figure 4 depicts the mapping of an example inter-AS topology to leveled topology. The left and right sides of the figure indicate an example inter-AS topology and leveled topology, respectively. The level of each domain is defined as follows.

a) Level 0 domain indicates the source of BGP routing instability.
b) Level 1 domain represents the neighbor having a peer-to-peer relationship with level 0.
c) Level 2 domain depicts the neighbor having a customer-to-provider relationship with level 0 or level 1 AS.

Figure 5 shows the domain Temporal Topology Variation

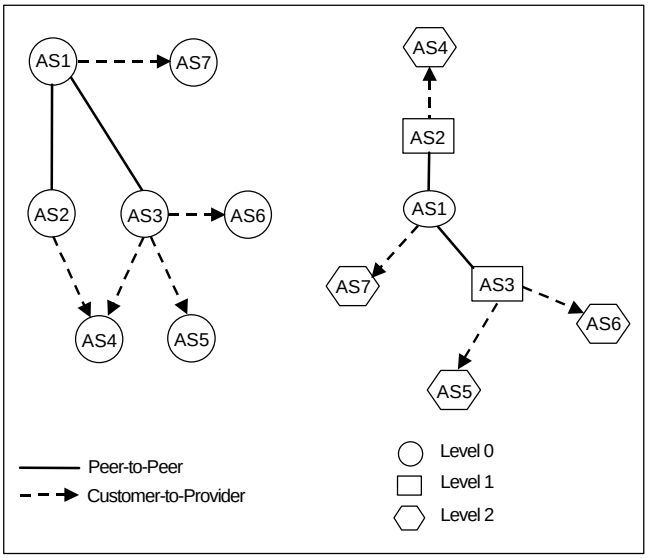

Fig. 4. Level architecture of inter-AS topology.

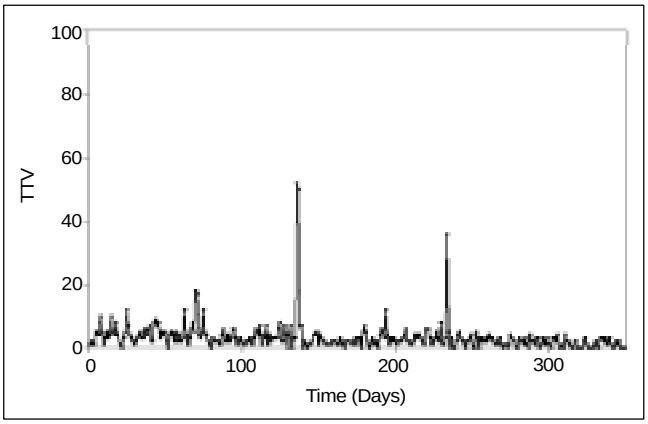

Fig. 5. Temporal topology variation of level 1 AS (AS1).

(TTV) of one of the peer-to-peer ASs with AS1239. TTV denotes the absolute value of the degree difference between consecutive days, namely, a high value of TTV showing a rapid change of topology indicates network instability. AS1239 is identified as highly unstable during 1998 [13]. According to our analysis, other neighbors of AS1239 show similar behavior.

As we can see in the Fig. 5, level 1 ASs show rapid domain degree variation on the 137th day (May 20, 1998), the same as AS1239. From the figure, we assume that the high instability of certain ASs can be generated by the effect of a peer-to-peer relation between AS pairs.

Table 1 summaries the domain degree of several of AS1239's neighbor ASs. These ASs have peer-to-peer relationships with AS1239, that is, peer-to-peer ASs of AS1239 show a rapid domain degree change. This behavior implies that the high BGP routing instability of certain ASs can impact the do-

Table 1. Domain degree of level 1 ASs.

AS Number	Domain degree (TTV)		
	May 19, 1998	May 20, 1998	May 21, 1998
1	205 (3)	153 (52)	201 (48)
701	802 (1)	514 (288)	806 (292)
1673	74 (0)	71 (3)	74 (3)
1740	101 (2)	92 (9)	103 (11)
2548	171 (7)	107 (64)	184 (77)
2914	137 (2)	125 (12)	135 (10)
3561	630 (6)	416 (214)	627 (211)
3847	48 (2)	35 (13)	51 (16)

main degree of its peer-to-peer AS. Furthermore, with the results of Govindan and Reddy [14], we can classify certain domains as backbone nodes if their degree is over 28 and understand that peer-to-peer relationship ASs of AS 1239 are backbone nodes.

Table 2 summarizes the analysis of domain degree distribution of level 1 and level 2 ASs. To analyze the degree change of level 2 domains, we select one of the peer-to-peer domains of AS1239 and then investigate the neighbor ASs of a selected AS. In this paper, we choose AS2548 as the origin of the level 2 AS. We classify BGP peers of AS2548 into peer-to-peer and customer-to-provider ASs.

According to analysis results, 58% (or 92 customers) of the customer domains show a change in TTV and 92% of the backbone domains show a change in TTV. Although most backbone domains experience a change in domain degree, BGP peerings between backbone domains are preserved. In other words, the topology of backbone nodes, which is made up of peer-to-peer relationships, does not change with the occurrence of BGP routing instability.

Many research results on the topological characteristics of inter-AS routing systems and the origins of Internet routing instability have been reported [5], [8], [10], [14], [15]. However, there is no work on the propagation of Internet routing instability among domains or the development of a systematic model which can represent the propagation of BGP routing instability. In this section, we analyze the propagation behavior of BGP routing instability using inter-AS relationships suggested by [15] and propose a BGP routing policy model to preserve stability in the inter-AS network during the occurrence of high level instability. It is important to analyze the propagation of instability, because with a systematic propagation model of instability, it is possible to set up a routing policy that can efficiently decrease the BGP instability and fast converge to a stable network.

Table 2. Domain degree of level 2 ASs.

Domain Class	Number of domains (%)	Number of domains whose TTV is nonzero (%)	Number of domains whose TTV is zero (%)
Backbone (or Level 1 ASs) degree $\geq$ 30	13 (7.6)	12 (92.3)	1 (7.7)
Customer (or Level 2 ASs)	158 (92.4)	92 (58.2)	66 (41.8)

2. Available Policy for Stable Convergence in OBGP

Arnaud et al. [3] discussed several possible schemes for Router A to signal to Router B its preferred connection:

1) Static configuration at setup.
2) Establish at configuration knowledge of the destination router.
3) Use BGP UPDATE information such that Router A can make a dynamic decision.
4) Let Router A control the OXC on Router B and advertise a virtual router that is part of Router A's domain.

Scheme 1) is clearly straight forward and the connection can be established as part of the standard configuration. Scheme 2) requires prior knowledge of the destination router by Router A but requires no further configuration setup. In this case Router A can signal to Router B in its OPEN message its desirability to connect to a specific router.

Router B still retains the decision authority as to where it will cross connect to serve its own needs. However, everything else being equal, Router B can designate the virtual router to cross connect the routers as indicated in the OPEN message from Router A or C. The loopback address or the actual interface of the designated router can be used to indicate the required destination.

However, it would be attractive to have a signaling protocol that allows Router A to indicate a cross connect preference to Router B without any prior knowledge of the other routers attached to Router B. More importantly, it would be ideal if Router A could also change its preference over time. One possible technique is after initial configuration, Router A performs a route flap. If initially Router A is connected to Router C, but instead it would prefer to be connected to router D, it can terminate the existing BGP process with the virtual router. Router A only knows Router C and D by their ASs. Router C and D in

effect could be an abstraction of an entire network. Router A then initiates a new BGP process by sending a BGP OPEN message to the virtual router but indicating in the options field that its preference is to connect to Router D (or any other subsequent routers further down the path).

The virtual router, in turn, then closes down its BGP session with Router C and initiates a BGP session with Router D. Router D also has done a soft reboot of its BGP process at the beginning and since then has been trying to establish a BGP session with the virtual router controlling the OXC. If Router D were also an integral OXC and router like Router B, it could further propagate any special routing requests from Router A using the same technique of route flaps. Router B in its OPEN message with Router D could carry a list of ASs that Router A can have a direct optical cross connection with.

IV. MODEL DESCRIPTION FOR OBGP CONVERGENCE POLICY

As time passes, an AS may change the nature of its relationships with its neighbors. For example, a customer may grow large enough to renegotiate its relationship with a provider, and the AS pair may transit to a peer-to-peer relationship. As part of evolving to a new relationship, the two ASs may need to change their import and export policies. Ideally, these changes would occur simultaneously. However, in practice, each AS configures the routers independently from the others. As a result, the BGP system may go through a transition period where one AS has changed its configuration and the other has not. Since these changes occur on a human time scale, it is important to carefully study the influence of the transition period on system stability.

Our proposed model can be used to identify potential convergence problems and to determine which kind of routing policy should be used to reduce the convergence time it takes to return to a normal state when BGP routing instability occurs.

In this section, we discuss our OBGP convergence policy model based on state automaton for reducing Internet OBGP routing instability and verify our model with BGP data that are collected in the global Internet.

To analyze OBGP routing instability propagation, we use AS relationships represented in the following properties that were proposed by Gao [10].

[Property 1] The valley-free is defined as follows. After traversing a provider-to-customer or peer-to-peer edge, the AS path cannot traverse a customer-to-provider or peer-to-peer edge.

[Property 2] If an AS path in any OBGP routing table entry satisfies the valley-free condition, the AS path of a BGP routing table entry has one of the following patterns:

(a) an uphill path,
(b) a downhill path,
(c) an uphill path followed by a downhill path,
(d) an uphill path followed by a peer-to-peer edge,
(e) a peer-to-peer edge followed by a downhill path,
(f) an uphill path followed by a peer-to-peer edge, which is followed by a downhill path.

A downhill path is a sequence of edges that are provider-to-customer edges and an uphill path is a sequence of edges that are customer-provider edges. For example, in Fig. 6, AS paths (1, 2, 3) and (1, 2, 6, 3) are valley-free while AS path (1, 4, 3) is not valley-free.

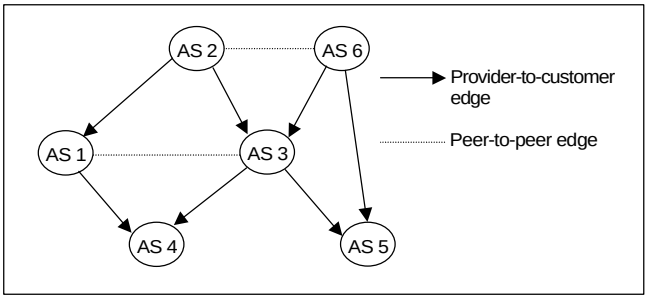

Fig. 6. Example topology showing valley-free.

As proposed by [16], if AS_PATHs listed in an OBGP routing table satisfy the valley-free condition, it is possible to classify an inter-AS topology into peer-to-peer or customer-to-provider relationships.

On the other hand, these relationships translate into rules that determine whether an AS exports its best routes to a neighboring AS, e.g., by normal export rules [9]. In addition, the interaction of locally defined routing policies can have global ramifications for the stability of the OBGP system. Conflicting local policies among a collection of ASs can result in OBGP route oscillation. To avoid local oscillation, we consider a policy model in the theorem provided in the subsequent section.

The definition of the safety of a path is defined as follows [15]. An instance of the path is safe if the protocol Simple Path Vector Protocol (SPVP) can never diverge. SPVP is defined in [17].

1. Policy Management for Stable Convergence

The transition between states happens during the traversing of the inter-AS topology. According to the AS relationship between two ASs, there are at most two possible transitions for each state. The output is generated during the state transition. Output indicates the probability of pruning the customer's logical connection to the provider. Output consists of two values, 0 and p_{tr} , where 0 means that the provider does not prune the

connection of the customer that has a customer-to-provider AS relation. p_{tr} represents the probability that the provider prunes the connection of the customer. Furthermore, while analyzing the state transitions in the automaton, the conceptual model of topology reconfiguration can be used to determine state transition probability p_{tr}.

If we define a state automaton with output $(E, X, \Gamma, f, x_0, Y, g)$, each parameter is formulated as follows. The set of events *E* consists of two events α and β. α means the neighbor AS is a peer-to-peer relationship and β means the neighbor AS is a customer-to-provider relationship. The set of states *X* has three entities, P_s, P, and C, which represents a starting state, a current state that has a peer-to-peer relationship with the previous state, and a current state that has a customer-to-provider relationship with the previous state, respectively.

The set of feasible events Γ is formulated as follows:

$$\Gamma(P_s) = \Gamma(C) = \{\alpha, \beta\}$$
$$\Gamma(P) = \{\alpha\}. \quad (1)$$

Let state transition function f^x be represented as follows:

$$f^x(P_s, \alpha) = f^x(C, \alpha) = P$$
$$f^x(P_s, \beta) = f^x(C, \beta) = f^x(P, \beta) = C. \quad (2)$$

The set of outputs *Y* has two entities. 0 means there is no outage of BGP connection between two states, and p_{tr} means the probability that the provider prunes the connection of the customer.

The output function *g* maps the pair of the set of states *X* and the set of events *E* to the set of outputs *Y*. The formal definition is given as follows:

$$g^x(P_s, \alpha) = g^x(C, \alpha) = 0$$
$$g^x(P_s, \beta) = g^x(C, \beta) = g^x(P, \beta) = p_{tr}. \quad (3)$$

Finally, the initial state x_0 represents the source of OBGP routing instability, namely P_s.

As proposed in [16], the state transition probability p_{tr} can be determined experimentally by statistical analysis of a large number of samples using inter-AS relationships.

Our proposed model can be applied to network topology as shown in Fig. 7. Let AS1 be the source of BGP routing instability. As the instability propagates from AS1 to its neighbors, each AS sets its routing policy according to the AS relationship with its neighbor.

As we defined before, event α means that two ASs have a peer-to-peer relationship and β means that two ASs have a provider-to-customer relationship. In the above example, AS1

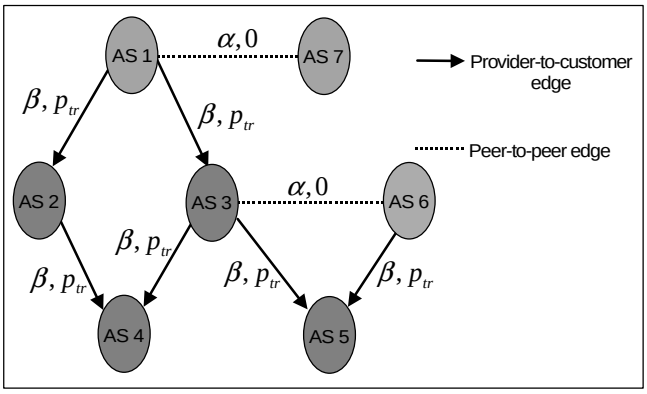

Fig. 7. Example of BGP routing instability propagation.

has one peer-to-peer AS (AS7) and two provider-to-customer ASs (AS2 and AS3). Since AS1 has a peer-to-peer relationship with AS7, AS1 does not change the connection for AS7. But in the case of AS2 and AS3, the edge is provider-to-customer, so AS1 (provider AS) sets its routing policy to cut its customers according to probability p_{tr}. Therefore, by pruning its customers, AS1 can prevent customers from sending their traffic to it, fast recover from high instability, and control the propagation range of the instability. Furthermore, by choosing the optimal p_{tr}, the BGP routing instability propagation range can be adjusted.

To discuss the convergence of the proposed model we assume the following.

[Assumption 1]

1) We assume that inter-domain topology shows two types of OBGP relationships, e.g., peer-to-peer and provider-to-customer.
2) We assume the probability that the occurrence of each OBGP relationship is equally distributed.

The routing policy following the proposed model converges the propagation of OBGP routing instability in the inter-domain network. Moreover, the convergence rate can be restricted by the following theorem.

[Theorem] If the propagation of routing instability follows Assumption 1, the event probability for a fast convergence policy is restricted by the following:

$$P(\textit{converge within Level n domain}) = 1 - \frac{(2 - p_{tr})^n}{2^n},$$

where *n* is the domain level and p_{tr} is given.

Proof : Since there are two types of inter-AS relationships and each relationship can occur in equal probability, there are 2^n routing paths according to the classification of inter-AS

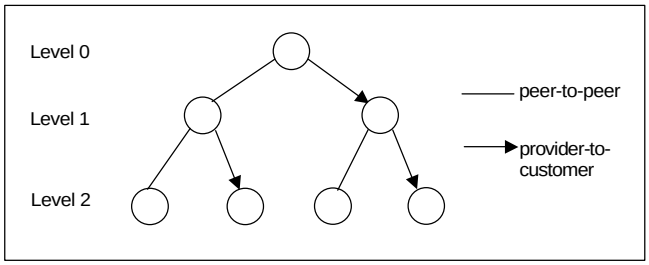

Fig. 8. Two level inter-AS topology.

relationships from level 0 domain to level *n* domain. Figure 8 shows a level 2 example.

In our proposed model, the instability is not absorbed in a peer-to-peer relationship.

From Fig. 8, we can see that there are four kinds of routing paths and the probability of instability propagation is computed as follows:

Peer-to-peer → peer-to-peer: 1*1
Peer-to-peer → provider-to-customer: $1*(1-p_{tr})$
Provider-to-customer → peer-to-peer: $(1-p_{tr})*1$
Provider-to-customer → provider-to-customer: $(1-p_{tr})*(1-p_{tr})$

Let q_{tr} be $1-p_{tr}$, then the average probability of routing instability propagation from the level 0 domain to the level *n* domain is given as

$$P = \frac{\sum q_{tr} \text{ for each path}}{\text{Total number of paths}}. \quad (4)$$

Since the denominator of (4) is 2^n in the level *n* domain and the numerator can be represented as $(1+q_{tr})^n$ according to the binomial distribution, the average probability of instability propagation is equal to $\frac{(1+q_{tr})^n}{2^n}$. Thus, the convergence probability is represented as follows:

$$1-\frac{(1+q_{tr})^n}{2^n} = 1-\frac{(2-p_{tr})^n}{2^n}. \quad (5)$$

Therefore, our proposed model converges as the domain level increases. Furthermore, there is a routing policy that can converge BGP routing instability. (q.e.d)

2. Experimental Analysis and Discussion

Potential problems in lightpath failure recovering techniques were discussed in [3]. Because of the inherent characteristics of a virtual BGP router, it is not easy to implement. However, since the virtual BGP router inherits the characteristics of the BGP router, it is possible to analyze the behavior of BGP [3],

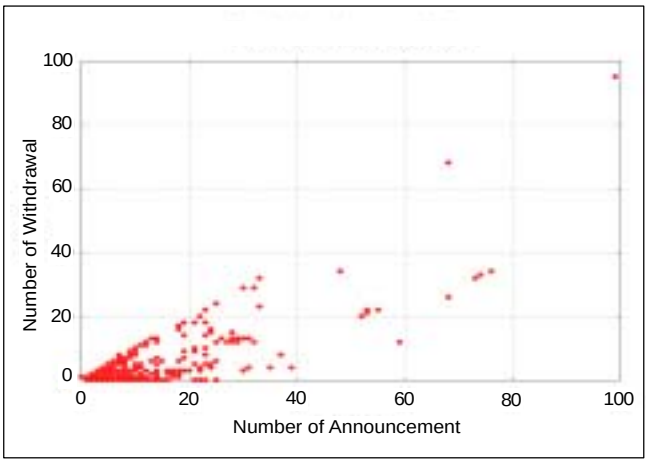

Fig. 9. Correlation graph of prefix announcements and withdrawals.

[7]. Therefore, before analyzing the proposed OBGP instability propagation model, we investigate the correlation of the number of BGP UPDATE messages using BGP data measured at XP in the domestic network in January 1999. Figure 9 depicts the correlation between the number of announcements and withdrawals of BGP UPDATE messages. The figure indicates that the number of prefix announcements is greater than the number of prefix withdrawals. This means that once a prefix is withdrawn, there is a high probability that it will be announced again.

Figure 10 shows the regression result of Fig. 9. The horizontal axis denotes a logarithmic scale. Let x be the number of prefix announcements and y be the number of prefix withdrawals. We use power-laws as a regression function, i.e., $y = ax^b$. Figure 10 shows that the relationship between announcements and withdrawals follows the power-law function [9].

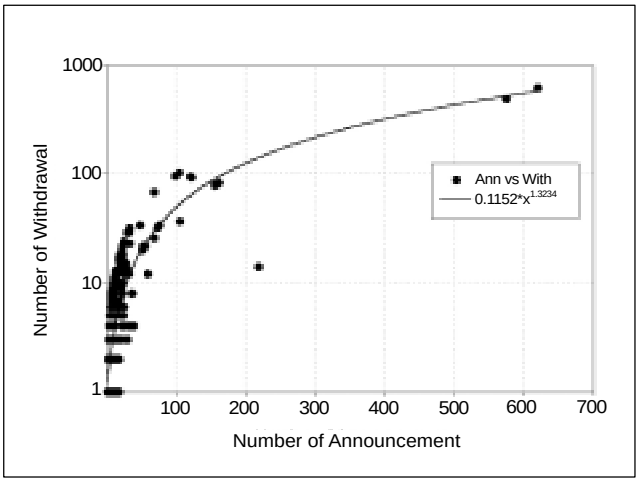

Fig. 10. Regression graph of prefix announcements and withdrawals.

Table 3. Regression parameters.

Parameter	Value	Stand Error	Coefficients of Variation (%)
A	0.1152	0.00587	5.095
B	1.3234	0.00826	0.6244

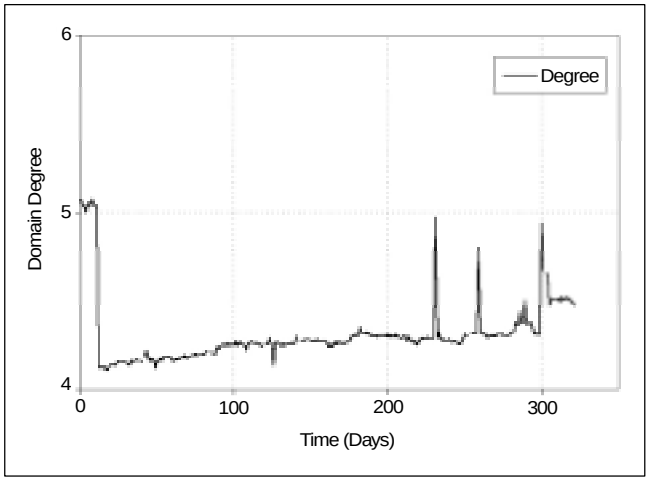

Fig. 11. Day-to-Day distribution of domain degree in 1999.

$$y = 0.1152 \times x^{1.3234}. \quad (6)$$

The Eq.(6) implies the existence of a correlation between prefix announcements and withdrawals in an inter-domain network.

Table 3 shows the regression parameters and errors.

Since the data used in Figs. 9 and 10 consist of all the BGP UPDATE messages during the measurement interval, we can conclude that they show microscopic behavior of BGP routing. Figure 10 shows that the distribution of the number of prefix announcements and withdrawals follows the regression line. Thus, it is possible to infer that the relation between prefix announcements and withdrawals shows linearity in relation to the power-law function.

Since the data used in Figs. 12 and 13 were gathered from the global Internet daily at midnight during two years, they show macroscopic behavior of BGP routing in the global Internet. Figure 12 shows that the variation of domain degree follows the regression line, i.e., it shows linearity in the temporal domain.

Figure 13 shows the convergence rate according to various p_{tr}, the probability of instability propagation to the provider-to-customer relationship with respect to the domain level. The figure reveals that the convergence rate of our proposed model increases according to the levels of the domain. This behavior

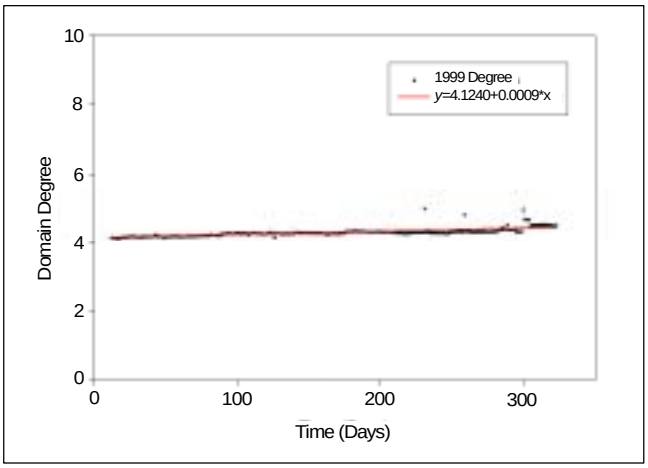

Fig. 12. Regression analysis of day-to-day distribution of domain degree in 1999.

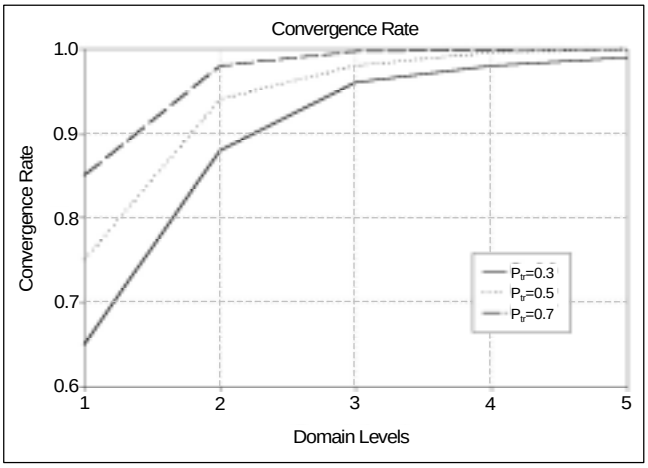

Fig. 13. Comparison of convergence rate versus domain level.

implies that the routing instability decreases by propagating domains that have a routing policy based on the proposed model.

Thus, on the basis of the results in Figs. 10, 12, and 13, we claim that our proposed model can be used for setting up a routing policy which converges to a stable state of OBGP routing.

The analysis scheme based on IP prefix update counts in [9] can identify the existence of BGP routing instability, but it cannot investigate the propagation behavior of the instability. The propagation behavior of the instability can be used for setting up a routing policy in the domain. Without considering the propagation of the instability, high levels of routing instability in a certain domain can cause the disconnection of its neighbor domains. However, our proposed model can be used for setting up a routing policy to manage the propagation of routing instability.

The applicability of the proposed model can be summarized

as follows: when OBGP routing instability occurs, by pruning some customer connections, i.e., the virtual BGP router peering, backbone domains reduce incoming customer traffic. In this way, backbone topology can be preserved even when OBGP routing instability occurs. To preserve reliability in inter-AS network, the routing path between peer-to-peer ASs should be guaranteed. Moreover, it is important to keep redundancy and an available routing path in the inter-AS topology.

V. CONCLUSIONS

Internet routing instability is an important problem in the Internet engineering community. High levels of BGP routing instability can lead to packet loss, increased network latency, and increased time to convergence. At the extreme, high levels of routing instability may lead to the loss of internal connectivity in wide-area networks. In spite of the seriousness of these problems, few studies have investigated the propagation of the instability in inter-AS networks or a routing policy that can preserve the reliability despite the occurrence of high level instability.

OBGP is an extension to BGP for the manipulation of OXCs to permit them to be automatically setup and configured as BGP speaking devices to support multiple direct optical lightpaths between many different autonomous domains. Since the virtual BGP router that connects the BGP router and OXC inherits the characteristics of BGP, the virtual BGP router is highly likely to show the same behavior, i.e., routing instability.

In this paper, we discussed the instability in OBGP and proposed a routing convergence model to reduce the instability in a virtual BGP router. We described the architecture of OBGP and the operation mechanism of virtual BGP routers. We also analyzed the potential problems of OBGP, especially the failure of OXCs, and proposed requirements for a virtual BGP router that would prevent each potential routing problem.

Our instability propagation model and its analysis demonstrate that an effective policy management scheme improves OBGP routing instability in case of lightpath failure. We conclude that our proposed model can be used to set up a routing policy in an autonomous system for the purpose of minimizing the effect and propagation of OBGP routing instability in the optical Internet.

REFERENCES

[1] J. Luciani, B. Rajagopalan, D. Awduche, B. Cain, B. Jamoussi, "IP over Optical Networks – A Framework," draft-ip-optical-framework-00.txt, Sept. 2000.

[2] Jong Hyup Lee, "Design and Configuration of Reconfigurable ATM Networks with Unreliable Links," ETRI J., vol. 21, no. 4, Dec. 1999.

[3] Bill St. Arnaud, Rene Hatem, Wade Hong, John Coulter, Marc Blanchet, Abdul Abdalla, Ian MacDonald, Florent Parent, Tom Tam, Mike Weir, "Optical BGP Networks," CA*Net3 News Archive, 2001.

[4] Y. Rekhter, "A Border Gateway Protocol 4," Request for Comments, 1771, Internet Engineering Task Force, Mar. 1995.

[5] C. Labovitz, A. Ahuja, A. Bose, and F. Jahanian, "Delayed Internet Routing Convergence," Proc. ACM SIGCOMM, Aug. 2000.

[6] Yangguang Xu, Anindya Basu, Yong Xue, "A BGP/GMPLS Solution for Inter-Domain Optical Networking," draft-xu-bgp-gmpls-01.txt, Internet Engineering Task Force, July 2001.

[7] Marc Blanchet, Florent Parent, Viagenie, Bill St-Arnaud, Canarie, "Optical BGP: InterAS Lightpath Provisioning," draft-parent-obgp-01.txt, Internet Engineering Task Force, Mar. 2001.

[8] C. Labovitz, R. Malan, and F. Jahanian, "Internet Routing Instability," Proc. ACM SIGCOMM, Sept. 1997.

[9] Sangjin Jeong, Chan-Hyun Youn, "An Analysis of Routing Instability Based on the Internet Topology," Proc. ICEIC, Aug. 2000.

[10] Lixin Gao, "On Inferring Autonomous System Relationships in the Internet," Proc. IEEE GLOBECOM, Nov. 2000.

[11] http://www.nlanr.net/

[12] http://www.antc.uoregon.edu/route-views/

[13] http://www.merit.edu/

[14] R. Govindan, A. Reddy, "An Analysis of Internet Inter-Domain Topology and Route Stability," Proc. IEEE INFOCOM, 1997, pp. 850-857.

[15] Lixin Gao, Jennifer Rexford, "Stable Internet Routing Without Global Coordination," Proc. ACM SIGMETRICS, June 2000.

[16] Sangjin Jeong, Analysis of BGP Routing Convergence Using Inter-AS Relationship, M.S. Thesis, Information and Communications University, Korea, June 2001.

[17] T.G. Griffin, G. Wilfong, "A Safety Path Vector Protocol," Proc. IEEE INFOCOM, Mar. 2000.

[18] Sangbum Kim, Chan-Hyun Youn, "Delay-Constrained Bottleneck Location Estimator and Application to Scalable Multicasting," ETRI J., vol. 22, no. 4, Dec. 2000.

Sangjin Jeong received his MS (2001) from Information and Communications University, Korea in communications and his BS (1999) from Korea Advanced Institute of Science and Technology (KAIST) in computer science. During 1999, he served at Korea Telecom (KT) Traffic Engineering Research Laboratories as an Invited Researcher and at the University of Southern California, Department of Electrical Engineering as a Visiting Scholar in 2001. Since 2001, he has been enrolled as a PhD candidate at Information and Communications University, Korea. His current interests include grid middleware, routing stability, and network performance measurement.

Chan-Hyun Youn received BS and MS degrees in electronics engineering from Kyungpook National University, Taegu, Korea, in 1981 and 1985, respectively. He also received a PhD in electrical and communications engineering from Tohoku University, Japan, in 1994. He served in the Korean Army as a communications officer, First Lieutenant, from 1981 to 1983. From 1986 to 1997, he was Leader of the high-speed networking team at Korea Telecom (KT) Telecommunications Network Research Laboratories, where he was involved in the research and development of switching systems maintenance systems, high-speed networking, and HAN/B-ISDN network testbed. Especially, he was a Principal Investigator of high-speed networking projects including ATM technical trial between KT and KDD, Japan, Asia-Pacific Information Infrastructure (APII) testbed, KOREN and APAN, respectively. Since 1997, he has been an Associate Professor at Information and Communications University, Daejeon, Korea. He is also Chair of resource management and scheduling WG in grid Forum Korea. Currently, he is interested in grid middleware, high performance routing, multicasting, optical Internet, and network performance measurement. He was a recipient of IEICE PAACS friendship prize, Japan, in 1994.

Minho Kang is Director of the Optical Internet Research Center at the Information and Communications University, Daejeon, Korea. He received the BSEE, MSEE, and PhD degrees from Seoul National University, University of Missouri-Rolla, and the University of Texas at Austin in 1969, 1973, and 1977, respectively. From 1977 to 1978, he was with AT&T Bell Laboratories, Holmdel, NJ. During 1978 and 1989, he was Department Head and Vice President at Electronics and Telecommunications Research Institute. Also during 1985 and 1988, he was the Electrical and Electronics Research Coordinator at the Korean Ministry of Science and Technology. During 1990-1998, we was Executive Vice President at Korea Telecom in charge of R&D, quality assurance, and overseas business development groups. In 1999, he joined the Information and Communications University, where he is a Professor of Engineering. Dr. Kang served as chairman at the Asia Pacific Telecommunity Study Group of Bangkok for 1996-1999 and the General Assembly of Korean Telecommunications Technology Association for 1995-1997. Dr. Kang is a member of National Academy of Engineering in Korea, Vice Chairman of the Optical Society of Korea, and Co-Chairman of the first International Conference on Optical Internet 2002. He was awarded the Order of Merit-DongbaekJang by the Korean Government and Grand Technology Medal by the 21st Century Management Club in 1983 and 1991, respectively, for the contribution to optical communications development.

Kyoung-Seon Min received his BS in electronic engineering from Korea University and MS in computer science from Hanyang University, Korea, in 1980 and 1991, respectively, and his PhD degree in electronic engineering from Aju University, Korea, in 1997. In 1980, he joined ETRI, where he worked on the design of MTIP (Magnetic Tape Interface Processor). In 1984, he moved to Korea Telecom, where he worked on the development project for the TDX-1/1A/1B/10 switching system during 1984 to 1993. From 1994 to 1996, he was in charge of designing and developing ATM switching systems. He is currently working on an optical internet project and NOC (Network Operation Center) for KOREN (Korea Research and Education Network). His research interests include optical internet and GMPLS, optical ethernet, optical switching.

Hyun Ha Hong was born in Yuncheon, Korea, in 1956. He received his BS degree from Kwangwoon University, in 1979, and the MS degree from Yonsei University, Seoul, in 1981, both in electronic engineering. In 1985, he joined Electronics and telecommunications research Institute (ETRI). He is now a Team Leader of the optical packet router team at Optical Communication Department there. His current research interests include optical packet/burst switched networks for Internet traffic, optical switch architectures, and performance analysis.

Hae Geun Kim received his BS degree in electronics engineering from Kyungbook National University, Korea in 1977, and his PhD degree in electrical engineering from the University of South Florida, USA in 1994. Since 1980, he has been with Electronics and Telecommunications Research Institute (ETRI) as a Principal Member of technical staff. Since 1994 he has worked as a Project Leader in research projects such as ATM switch, Optical Cross Connector, Optical Packet and Burst switching. His research interests include Optical Switching, Optical Cross Connector, OCDMA, Subcarrier Multiplexing, Coding, and Modulation / Demodulation.