

# 한국 인터넷 AS 레벨 위상 분석

오 동 익\*, 이 강 원°

## Analysis of Korea Internet AS-Level Topology

Dong Ik Oh\*, Kang Won Lee°

### 요 약

한국의 인터넷은 매우 빠르게 진화해 등록된 AS 수는 755개에 이르지만 국내에서 AS 수준의 인터넷 위상을 분석한 연구는 거의 전무하다. 본 연구에서는 한국 인터넷 AS 위상 구축과 power-law 특성 분석을 시도 하였다. 보다 정확하고 최신의 AS 수준 위상 구축을 위해 사용 가능한 데이터들을 모두 조사 하였다. 이들에 대한 비판적인 분석 과정을 통해 한국의 실정에 적합하고 또한 획득 가능한 데이터 소스 들을 유도 하였다. UCLA의 “Internet Topology Collection”을 이용한 BGP 데이터 그리고 IRR 과 IXP 데이터로부터 실제 국내 인터넷에 최대의 근접한 위상 데이터를 추출하고 검증 절차를 거쳐 한국 인터넷의 AS 수준 위상을 구축 하였다. 전 세계 AS 수준 위상의 차수 분포가 power-law 분포를 따르는 것으로 알려져 있다. 본 연구에서는 전 세계 AS 수준 위상의 일부분인 한국 AS 수준 위상의 차수 분포가 power-law 분포를 따르는지 조사하였다.

**Key Words** : AS-level topology, Autonomous System, Power-law, BGP, Internet topology, 인터넷 위상

### ABSTRACT

Over the last decade, the Internet in Korea has developed rapidly. The number of ASes in Korea is 755. However, there has been no significant effort to obtain an accurate picture of the Internet AS-level topology. To construct more complete set of Korea AS-level topology data all the available data sources are critically reviewed and compared. Three data sources, which can be used to infer Korea Internet AS-level topology, are selected; BGP data of UCLA “Internet Topology Collection”, IRR data and IXP data. It is known that global Internet AS-level topology follows power-law distribution. In this paper, we also investigate if the Korea Internet AS-level topology follows power-law distribution.

### I. 서 론

인터넷의 급격한 성장으로 inter-networking과 관련된 라우팅, 자원예약, 관리운영과 같은 문제들이 광범위하게 발생하여 왔다. 이러한 문제들을 해결하기 위해 실제 인터넷 상에서 새로운 알고리즘, 관리 정책들을 적용하여 테스트 하는 것이 가장 좋은 방법이겠지만, 실제 인터넷 상에서 테스트는 불가능하

다. 때문에 인터넷 구조와 유사한 모델에서 시뮬레이션을 통해 알고리즘과 정책들을 분석, 평가하는 방법이 사용되어진다. 인터넷과 유사한 특성을 가지는 모델을 만들기 위해서 실제 인터넷의 특성을 알아야 하며, 동일한 특성을 가지는 위상을 생성할 수 있어야 한다. 자연스럽게 우리는 “인터넷은 무엇처럼 보이는가?”, “미래에도 변하지 않는 인터넷의 특성은 무엇인가?”, “시뮬레이션을 위해 인터넷과 닮

※ 이 연구는 서울과학기술대학교 교내 학술연구비 (일부)지원으로 수행되었습니다.

• 주저자 : 서울과학기술대학교 산업정보시스템공학과 통신망 분석 실험실, cyber7511@hanmail.net, 정희원  
 ° 교신저자 : 서울과학기술대학교 산업정보시스템공학과 통신망 분석 실험실, kwlee@seoultech.ac.kr, 정희원  
 논문번호 : KICS2012-07-335, 접수일자 : 2012년 7월 31일, 최종논문접수일자 : 2012년 10월 8일

은 그래프를 어떻게 생성할 수 있는가?”와 같은 질문들을 하게 된다. 지난 몇 년 동안의 연구에도 불구하고 인터넷 위상의 모델링은 아직도 중요한 공개 문제로 남아있다<sup>4)</sup>. 정확한 인터넷 위상을 이해하고 파악 하는 것은 다음의 관점에서 중요한 의미를 갖는다. 첫째, 인터넷의 망 관련 다양한 알고리즘을 분석하고 평가하는데 중요한 역할을 하고 둘째, 위상의 특성을 이용하여 더욱 효율적인 알고리즘을 개발할 수 있게끔 하며 셋째, 네트워크의 사고와 공격에 더욱 신뢰할 수 있는 네트워크 구조를 구축하는데 도움을 줄 수 있다. 마지막으로 향후 급격하게 성장할 인터넷에 적합한 전략을 계획하고 설계할 수 있게 해준다<sup>5)</sup>.

인터넷 위상은 인터넷 상의 host-host, router-router, AS-AS(Autonomous System) 간에 서로 연결되어 있는 구조로 정의할 수 있다(그림1). 이중에서도 AS 수준의 위상에 대한 연구가 주요 연구 대상이다. AS 수준의 인터넷 위상 연구가 중요한 이유는 첫째, AS는 인터넷의 가장 상위에 위치한 구조이며, 다른 수준의 인터넷 위상은 AS 수준 위상에 의존적이다. 둘째, 인터넷의 end-to-end QoS(Quality of Service)는 AS간 트래픽 흐름과 정책에 의존적이다. 셋째, AS간 라우팅(BGP Routing)과 트래픽 흐름은 AS 운영 기관의 경제적, 정책적인 요인에 의해 설계되어지기 때문에 예측하기가 힘들다. 하지만 AS 수준의 위상 분석을 통해 AS의 위상 특성과 AS간의 상호 관계가 규명 된다면 어느 정도 라우팅과 트래픽 흐름을 예측 할 수 있다. 마지막으로 AS 수준의 인터넷 위상 데이터가 상대적으로 얻기 쉽다. 다른 수준의 위상은 때때로 개인적인 정보로 간주되며, 데이터 획득이 어렵다<sup>12)</sup>.

일반적으로 AS 수준의 인터넷 위상분석을 위하여 사용되는 데이터는 BGP(Border Gateway Protocol) 라우팅 테이블 및 업데이트 메시지, Route Servers, Looking Glasses, IRR(Internet Routing Registry) 데이터베이스, Traceroute, IXP(Internet Exchange Point)를 통해 수집된 데이터이다<sup>1,3,6,7,12)</sup>. 이들 데이터는 각각 자신만의 장점을 가지고 있지만, 반대로 한계도 가지고 있다. 예를 들어 BGP 라우팅 테이블과 업데이트 메시지를 통해 얻은 데이터는 정확도가 높지만 hidden 링크와 invisible peer 링크를 포함하지 못하는 단점이 있다<sup>11,12)</sup>. IRR 데이터베이스는 BGP 데이터에서 얻지 못하는 링크를 추가적으로 얻을 수 있지만 AS 보유기관 관리자의 자발적 참여에 의존하여 데이터를

구성하기 때문에 다수의 불완전한 또는 기한이 만료된 데이터들을 포함하고 있을 수 있다<sup>18,10)</sup>. 서로 다른 방법으로 수집한 인터넷 위상 데이터를 이용하여 구축한 그래프는 각각 다른 모양과 특성을 갖는다. 때문에 인터넷 위상 특성 연구에 어떤 위상 데이터를 사용하는가가 연구 결과에 절대적인 영향을 미친다. 이렇게 위상 데이터의 종류에 따라 상이한 결과를 보임에도 불구하고, 한 종류의 위상 데이터만을 이용하거나, 기간이 만료된 데이터, 불완전한 데이터까지 포함한 데이터를 이용하여 AS 수준의 위상 구축을 시도한 여러 연구들이 진행되었다<sup>12,4,5,9)</sup>. 따라서 보다 정확한 인터넷 AS 수준의 위상 구축과 특성 분석을 위해 무엇보다 먼저 선행되어야 할 것은 보다 정확한 그리고 최신의 위상 데이터를 확보하는 것이다.

현재 한국에 할당된 AS번호는 755개(2012년4월)로, 세계에서 12번째로 많은 수의 AS를 보유하고 있다. 하지만 국내에서 AS수준의 인터넷 위상을 분석한 연구는 거의 전무한 상태이다. 지금까지 국내 인터넷 위상에 관한 연구는 공개된 BGP 데이터를 이용하여 한국 인터넷 위상의 power-law의 특성을 분석한 연구[13]가 전부이며, 연구의 결과에 결정적인 영향을 끼치는 인터넷 위상 데이터에 대한 검증, 분석에 대한 연구는 없었다. 본 논문에서는 먼저 인터넷 AS 수준의 위상 구축을 위해 사용 가능한 데이터들을 모두 조사 하였다. 이들에 대한 비판적인 분석 과정을 통해 한국의 실정에 적합하고 또한 획득 가능한 데이터 소스 들을 유도 하였다. UCLA의 “Internet Topology Collection”<sup>14)</sup>을 이용한 BGP 데이터 그리고 IRR 과 IXP 데이터로부터 실제 국내 인터넷에 최대한 근접한 위상 데이터를 추출하고 검증 절차를 거쳐 한국 인터넷의 AS 수준 위상 구축을 시도 하였다.

본 논문 II장에서는 인터넷 위상 데이터 주요 소스 들에 대해서 설명하고 이들에 대한 비판적인 분석 과정을 통해 한국의 실정에 적합하고 또한 획득 가능한 데이터 소스 들을 유도 하였다. III장에서는 II장에서 결정한 데이터 소스 들로부터 한국 인터넷 AS 수준의 위상 구축을 하였다, IV장에서는 국내 인터넷 위상의 power-law 특성과 global 인터넷의 power-law 특성을 비교 분석하였다. 마지막 V장에서 결론과 추후 연구 과제를 논하였다.

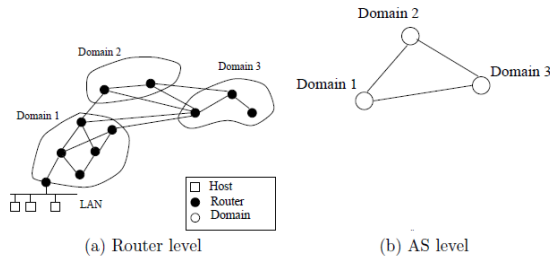


그림 1. AS 수준 인터넷 위상  
Fig. 1. AS-level internet topology

## II. 인터넷 위상 주요 데이터 소스

본 장에서는 인터넷 위상 분석에 일반적으로 사용되는 데이터 소스 들을 소개하고, 각 데이터 소스들의 특성과 장, 단점들을 분석 하였다.

### 2.1. 주요 위상 데이터 소스들

#### 2.1.1. BGP Trace Collectors

BGP(Boarder Gateway Protocol)는 AS의 라우팅을 위한 프로토콜이다. BGP 데이터 덤프(Dump)는 AS 수준의 인터넷 위상 연구에 가장 많이 사용되는 데이터 이다. BGP 라우팅 테이블에는 AS 경로가 저장돼 있다. 예를 들어 다음과 같은 AS 경로 (“701 1299 34211 41856”) 가 있다면, 우리는 위 경로로부터 701-1299, 1299-34211, 3411-41856 이렇게 세 개의 링크를 확인할 수 있다. 다수의 BGP 라우터에서 BPG 라우팅 테이블과 update 메시지를 수집하는 프로젝트가 진행 중 이며, RouteView, RIPE/RIS가 대표적인 예이다<sup>[7]</sup>. 이들 BGP trace collector들로부터 수집된 위상 데이터는 신뢰도가 높지만 많은 수의 invisible peer 링크와 hidden 링크들을 찾아 낼 수 없는 문제를 가진다<sup>[12]</sup>. 이것은 밸리-프리 라우팅 정책(Valley Free Routing Policy)을 따르는 BGP의 행동 방식 때문이다. 예를 들어 (그림2) A의 경우 일반적인 상황에서 5번 노드에서 1번 노드로 데이터를 전송하기 위해 2번 노드를 이용한다. 4번 노드를 통해서 데이터를 전송하는 경우는 2번 노드를 통해 데이터를 전송할 수 없을 때 백업용으로 사용하는 것이다. 이러한 경우 4번 노드를 통한 링크를 발견하는데 오랜 시간이 걸린다. (그림2) B의 경우에 8번과 9번 노드는 앞에서 언급한 밸리-프리 라우팅 정책으로 인해 그들 사이의 peer 링크를 노드 10번으로 링크를 광고하지 않기 때문에 관찰 될 수 없다. 위와 같은 이유 때문에

우리는 전체 AS 수준 인터넷 위상의 일부분만을 파악할 수 있다. Public View만을 사용할 경우 모든 AS들의 단지 4%만 peer링크를 포함한 완전한 연결을 추론 할 수 있고, 나머지 96%인 AS들은 hidden 링크를 제외한 provider-customer링크는 확인할 수 있지만 invisible peer 링크는 확인할 수 없다<sup>[7,8]</sup>.

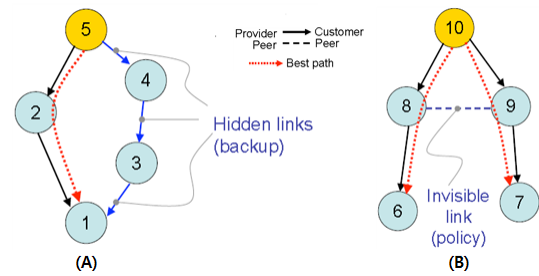


그림 2. AS 수준 링크 관찰  
Fig. 2. Observation of AS-level links

#### 2.1.2. Router Servers

Router Servers는 망에서 발생하는 문제들을 해결하기 위하여 ISP 망 운영자들이 공격적으로 사용할 수 있는 라우터들이다. 사용자들은 telnet을 이용하여 접속이 가능하며, 몇 가지 라우터 명령을 실행할 수 있다. 이러한 Router server들은 RouteView<sup>[19]</sup>나 RIPE/RIS<sup>[20]</sup>와는 peer 관계를 가지고 있지 않기 때문에 추가적인 링크나 노드를 발견할 수 있다.

#### 2.1.3. Looking Glasses

Looking Glasses는 웹상에서 매우 제한된 명령어를 이용하여 라우팅 정보를 제공하는데 사용자는 특정 prefix의 경로를 점검할 수 있다. 비록 Looking Glass가 라우팅 테이블을 제공하지는 않지만 특정 네트워크의 이웃 네트워크를 확인할 수 있다. Looking Glass는 Router Server에 비해 매우 적은 양의 데이터를 제공하지만, Router Server의 수보다 많은 수의 Looking Glasses가 존재한다. 대부분의 Looking Glasses는 RouteView나 RIPE/RIS와 peer 관계를 가지고 있지 않게 때문에 추가적인 링크나 노드를 발견 할 수 있다.

#### 2.1.4. Traceroute

Traceroute는 ICMP probe packet을 이용하여 출발지에서 목적지까지 IP 경로를 보여준다. CAIDA[18]는 traceroute 기반의 인터넷 위상 데이터를 수집하는 Skitter라는 툴을 개발하였다. Skitter

는 약 백 만개의 IPv4 주소를 관리하며, 이들 주소를 목적지로 traceroute를 수행하여 인터넷 위상을 추론한다. Skitter는 전 세계에 25개의 모니터를 운영하고 있다. 각각의 모니터는 probe packet들을 목적지 리스트로 보내어 IP 경로를 모은다. 이렇게 수집된 데이터는 실제 인터넷 경로를 보여주기 때문에 수집된 데이터의 신뢰도가 높다. 그러나 traceroute server는 자신의 위치에서 전 세계 목적지까지의 경로를 수집하기 때문에, 다수의 링크가 빠지는 문제점이 있다. 또한 RouteView에서 제공하는 BGP 테이블을 사용해서 추출된 IP 주소를 AS번호로 매핑 하는데 이는 간단한 문제가 아니며 왜곡이 나타날 수 있다. 특히 multi-home AS들은 AS와 IP주소가 정확하게 매칭 되지 않는 문제를 가진다.

2.1.5. Internet Routing Registry(IRR)

IRR은 AS간의 연결과 라우팅 정책을 저장하기 위한 데이터베이스이다. IRR은 글로벌 인터넷 라우팅의 보안, 일관성, 안전성을 향상시킬 목적으로 만들어졌다. IRR 데이터는 ISP 운영자들의 참여로 만들어지기 때문에 기한이 만료되거나 불완전한 데이터들을 많이 포함하고 있다고 알려져 있다. 하지만 잘 관리된 IRR 데이터는 다른 위상 데이터에서 제공하지 못하는 정보들을 제공해 준다. 때문에 많은 논문에서 IRR 데이터를 BGP 라우팅 테이블에서 추론한 인터넷 토폴로지의 빠진 링크를 보완하는데 사용하고 있다<sup>1,3,6,8)</sup>. 라우팅 정책은 RPSL(Routing Policy Specification Language)를 통해서 표현된다. RPSL은 객체 지향 언어로 13개의 클래스로 정의되어 진다. 라우팅 정책은 aut-num 객체의 import, export, default의 속성으로 표현된다. aut-num 객체는 라우팅 정책을 특정하기 위해, as-sets나 peering-sets와 같은 다른 객체를 참조 할 수 있다. <표1>은 간단한 RPSL의 예이다. aut-num의 값인 AS1111은 이후에 나오는 import와 export의 값들이 AS1111의 라우팅 정책임을 나타낸다. import와 export 속성을 통해서 간단한 peering 관계를 파악할 수 있다. <표1>의 경우 AS1111과 AS2222는 peering 관계가 있는 것을 알 수 있다. 한국의 IRR 데이터는 APNIC(Asia Pacific Network Information Centre)의 FTP 서버에서 획득 할 수 있다.

표 1. RPSL의 예  
Table 1 . Example of RPSL

aut-num:	AS1111
as-name:	ABC
descr:	ABC company
country:	KR
import:	from AS2222 action pref=10 accept ANY
[.....]	
export:	to AS2222 announce AS1111
[.....]	
changed:	hostmaster@nic.or.kr
source:	KRNIC

2.1.6. Internet Exchange Point(IXP)

IXP는 서로 다른 AS간의 트래픽을 원활하게 소통시키기 위한 인터넷 연동 서비스를 말한다. 최근 대부분의 peer 링크들은 IXP를 통해 연결 상태를 유지한다. IXP는 크게 두 가지 종류로 나누어진다. 레이어3 방식의 IXP는 하나의 라우터를 설치하고 여기에 각 AS의 라우터를 연결하는 방식이며, 레이어2 방식의 IXP는 AS의 라우터 사이를 스위칭 허브로 접속하는 방식이다. Packet Clearing House(PCH), Peeringdb, 그리고 Euro-IX와 같은 웹사이트에서 IXP 리스트를 관리하고 있다[21,22,23]. IXP의 참여 AS목록은 참여 AS의 자발적 참여에 의존하여 작성되기 때문에 불완전하거나, 기한이 지난 목록이 있을 수 있다. 그러나 BGP table에서 발견되지 못한 peer-peer 링크의 95%를 찾을 수 있다고 알려져 있다[11]. 때문에 IXP에서 AS간의 연결 상태를 조사하는 것은 인터넷 위상 데이터 수집에 도움을 줄 수 있다.

2.2. 한국 인터넷 위상 데이터 구축을 위한 데이터 소스 선택

전 절에서 일반적으로 사용되는 위상 데이터 소스에 대해서 알아보았으며 위상 데이터에 대한 다음과 같은 사실을 확인할 수 있었다.

첫째, 좀 더 완벽한 AS 수준의 인터넷 위상을 구축하기 위해서는 가능한 한 II-1에 소개한 여러 데이터 소스 들을 통합하여 사용하는 것이 좋다. 본 연구에서는 AS 수준의 한국 인터넷 위상을 구축하기 위해 가능한 많은 위상 데이터 소스들을 통합하여 사용할 것이다. BGP trace collector의 현재 데이터에 라우팅 테이블을 더 많이 추가 하더라도 많은 도움이 되지 않는다. 이것은 라우터가 지역적인 연결에 관한 정보만 제공하기 때문이다.

둘째, BGP 라우팅 테이블의 snapshot은 단지 snapshot 당시 최적 경로 상에 포함된 연결 정보만을 제공한다. 라우팅 경로는 링크 단절, 라우팅 정책 변화, 위상 변화에 따라서 변경될 수 있다. 때문에 BGP 라우팅 테이블과 업데이트 메시지 데이터는 일정 기간 동안 수집 되어야 할 필요가 있다.

셋째, RIPE/RIS나 RouteView와 같은 BGP collector들로부터 수집된 BGP 데이터를 사용하면 multi-homed AS들의 hidden 링크(backup 링크)들과 stub AS 사이 또는 stub AS와 small AS 사이의 peer 링크(invisible 링크)들을 제외한 provider-customer 링크들은 잘 파악 된다. 한편 hidden 링크들은 일반적으로 BGP update 메시지 수집 기간을 연장함으로써 수집할 수 있다. BGP 데이터로 부터 invisible peer 링크를 찾아내는 것은 거의 불가능 한데 이는 AS의 밸리-프리 라우팅 정책(Valley Free Routing Policy) 때문이다. 따라서 stub AS 사이 또는 stub AS와 small ISP 사이의 peer 링크를 찾아내는데 보다 많은 노력을 경주해야 한다.

넷째, IRR 데이터와 IXP 데이터는 라우터의 지역적 연결과 hidden 링크, invisible peer 링크를 발견하는데 중요한 데이터이다. 따라서 본 연구에서는 IRR 데이터와, IXP 데이터를 사용하여 BGP 데이터에서 발견하지 못하는 링크들을 보완하였다. IRR과 IXP 데이터를 사용하기 위해서는 각 데이터의 정확성과 최신성을 확보해야만 한다.

다섯째, Looking Glasses는 local connectivity에 대한 정보를 제공한다. 그러나 한국은 아직 Looking Glasses를 갖고 있지 않기 때문에 본 연구에서 Looking Glasses 데이터는 사용할 수 없었다.

여섯째, 국내 AS는 대부분 작은 사이즈의 AS이거나 stub AS들이기 때문에 대부분의 AS들이 address block을 광고하지 않는다. 따라서 앞서 언급한 것처럼 traceroute는 이런 AS들을 잡아내지 못할 것이다. 또한 traceroute는 IP 주소와 AS 번호와의 매핑 문제를 가진다. 이것은 간단히 해결될 문제가 아니고 또한 많은 왜곡이 발생할 수 있다. 따라서 본 연구에서는 traceroute 데이터를 사용하지 않았다.

결과적으로 본 논문에서는 BGP 데이터, IRR 데이터, IXP 데이터를 사용하여 한국 AS 수준 인터넷 위상 구축을 시도 하였다.

### III. 국내 인터넷 위상 구성

#### 3.1. BGP 데이터

BGP 데이터 획득을 위해 UCLA의 “Internet Topology Collection”을 사용하였다. BGP routing tables과 update 메시지, Route Server, Looking Glasses의 데이터들을 이용하여 전 세계 AS 위상을 구축하는데 이 데이터들은 매일 update 된다. 본 연구에서는 2012년 1월 한 달간의 데이터를 통합한 뒤 이 데이터에서 한국에 할당된 AS 노드와 AS간의 링크만을 추출하여 한국 AS 수준 위상을 구축하였다. 국내 AS 목록은 KISA(Korea Internet Security Agency)<sup>[15]</sup>에서 제공하는 자료를 사용하였다. 최종적으로 UCLA “Internet Topology Collection” 데이터에서 추출된 한국 AS 위상 데이터는 678개의 AS와 1,273개의 링크이다. 링크의 1,125개는 provider-customer관계이며, 145개는 peer-peer관계를, 그리고 남은 3개의 링크는 관계가 미확인 되었다.

#### 3.2. IRR(Internet Routing Registry) 데이터

한국 IRR 데이터는 APNIC의 FTP서버<sup>[17]</sup>에서 제공하고 있는 데이터를 사용하였다. 전 장에서 언급한 것처럼 IRR 데이터는 AS 관리자의 자발적 참여에 의존하기 때문에 만료된 데이터들과 불완전한 데이터들이 다수 포함되어 있다. 때문에 IRR 데이터를 사용하기 전에 데이터를 검증하는 작업을 수행하였다. IRR 데이터의 검증 방법은 Beichuan Zhang et al.[1]이 제안한 IRR 데이터 4단계 검증 방법을 이용하였다. 각 단계마다 IRR 레코드들을 검사하고 각 단계를 통과하지 못한 레코드들은 최종 위상 데이터 구성 시 제외 된다. IRR 데이터 검증 4단계 방법은 다음과 같다.

1단계 Void: IRR에 있는 특정 AS를 위한 레코드가 BGP 라우팅 테이블이나 BGP update에서 한 달 이상 발견되지 않으면 그 AS는 Void로 간주한다.

2단계 Void neighbors: 특정 AS와 관계를 가지는 이웃 AS가 BGP 라우팅 테이블이나 BGP update에서 적어도 한 달 동안 발견되지 않으면 두 AS 모두 Void 이다.

3단계 Incomplete: 특정 AS 레코드의 최종 변경 시간 이후 특정 AS의 모든 이웃 AS들을 BGP 라우팅 테이블이나 BGP update에서 찾아낸다. 만약 BGP 테이블이나 BGP update에는 존재하는 특정

AS의 이웃들이 IRR의 특정 AS 레코드에서 발견되지 않는다면 incomplete로 간주한다.

4단계 Neighbor conflict: 위 세 단계를 모두 통과한 후에, IRR 레코드 A에 AS B가 이웃으로 등록되어 있지만, 레코드 B에는 AS A가 이웃노드로 등록되어 있지 않다면 레코드 A와 B 중 하나 또는 양쪽 모두는 잘못된 것이 확실하다. 이런 경우 neighbor conflict라 하고 두 레코드 모두를 삭제한다.

4단계 검증을 하기 전 국내 IRR 데이터로부터 800개의 AS와 1,595개의 링크를 획득 하였다. 4단계 검증을 모두 마친 결과 모든 데이터가 IRR 데이터 검증 4단계를 통과하지 못했다. 이러한 결과의 이유는 다음과 같다. 한국의 대형 ISP 업체(KT, SK, LGU+)들은 수 백 개의 이웃 노드들을 가지고 있다. IRR 데이터를 일일이 분석한 결과 대형 ISP 업체의 이웃들인 small AS 혹은 stub AS의 IRR 레코드에는 대형 ISP의 AS 번호가 이웃 노드로 등록 되어 있었다. 하지만 대형 ISP 업체의 IRR 레코드에는 그들의 이웃으로 있는 AS들이 등록되어 있지 않았다. 때문에 위 검증의 네 번째 단계에서 모든 AS들이 neighbor conflict가 되어 삭제된 것이다. 이 문제를 수정하기 위하여 대형 ISP의 이웃들을 수작업으로 조사하였는데, small AS 혹은 stub AS들의 IRR 레코드에 이웃으로 대형 ISP가 등록되어 있다면, 두 AS간에 링크가 반드시 존재하는 것을 확인 하였다. 따라서 대형 ISP의 IRR 레코드에 small AS 혹은 stub AS가 이웃으로 등록되어 있지 않더라도 이를 neighbor conflict로 간주 하지 않았다. 이런 방식으로 다시 neighbor conflict 검증을 실행한 결과 215개의 AS와 408개의 링크가 남았다 (<표 2> neighbor conflict 괄호안의 값). 본 연구에서는 수정된 neighbor conflict로 검증한 결과를 사용하였는데 단계별 결과는 <표 2>와 같다.

표 2. IRR 데이터 검증 결과  
Table 2. Validity Analysis of IRR data

	No. of ASes	No. of Links
All Records	800	1,595
- Void	727	1,425
- Void Neighbor	644	1,239
- Incomplete	293	588
- Neighbor Conflict	0(215)	0(408)

최종적으로 남은 데이터를 살펴본 결과, 408개의

링크 중 49개의 링크는 다른 위상 데이터에는 존재하지 않고 오직 IRR에만 존재하는 링크로 나타났다, 49개 링크 모두 provider-customer 관계이며, back-up 링크인 것으로 조사되었다.

3.3. IXP 데이터

KISA 홈페이지는 한국 AS 업체들의 연결 지도 [16]를 제공하고 있다. KISA에서 제공하는 데이터는 추론된 데이터가 아니기 때문에 신뢰할 수 있는 데이터다. AS 연결 지도는 각 AS의 이름과 AS 번호를 알려주며, 지도에서 AS 이름에 마우스를 올리면 다른 AS와의 연결 상태를 보여준다(그림 3).

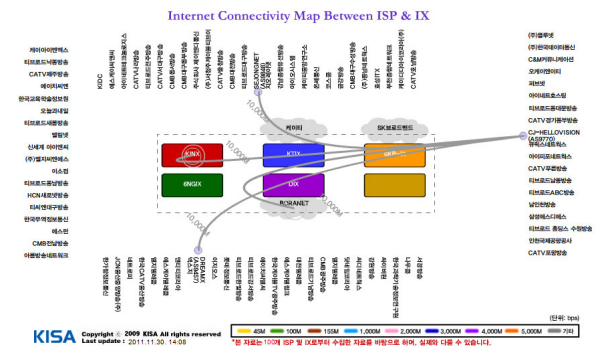


그림 3. KISA 한국 AS 연결 지도  
Fig. 3. Map of Korea AS connection at KISA

AS 연결 지도는 국내 ISP 업체들이 매달 제공하는 자료를 바탕으로 KISA에서 정리한 데이터다. 이는 한국 인터넷의 위상 분석을 하는데 기존의 데이터에서 얻을 수 없는 중요한 정보를 제공한다. 2011년 12월 기준으로 KISA의 한국 AS 연결 지도는 106개의 노드와 261개의 링크를 보여주고 있다. 전체 링크 중 101개의 링크는 오직 KISA 지도에서만 관찰 가능한 데이터였으며, provider-customer 링크가 69개, peer-peer 링크가 32개로 나타났다. Provider-customer 링크는 layer-3 router를 통해 연결된 것이며, peer-peer 링크는 layer-2 switch hub를 통해 연결된 것들이다.

3.4. 최종 한국 AS 수준 위상 데이터

앞서 소개한 3가지의 데이터를 정리하여 최종 한국 AS 그래프를 구축하였다. 각 데이터들 간에 중복되는 데이터를 제외하고 최종적으로 685개의 AS와 1,428개의 링크를 확인하였다. 기존에 발표된 연구[13]의 노드와 링크 수가 각각 659 개와 1,304 개임을 감안 한다면 이는 4%와 9.5% 증가한 수치다.

<표 3>은 각 그래프별 노드와 링크를 비교한 것이다.  $G_B$ 는 BGP 데이터를 토대로 구축한 AS 그래프,  $G_R$ 은 IRR 데이터를 토대로 구축한 그래프,  $G_X$ 는 IXP 데이터를 토대로 구축한 AS 그래프를 나타낸다.

표 3. 그래프 별 노드와 링크 비교  
Table 3. Comparison of  $G_B$ ,  $G_R$  and  $G_X$

	Number of Nodes	Number of Links
Only $G_B$	376	773
Only $G_R$	0	49
Only $G_X$	7	101
$G_B \cap G_R$	215	354
$G_B \cap G_X$	99	155
$G_R \cap G_X$	12	14
$G_B \cap G_R \cap G_X$	12	9
Sum	685	1,428

<표 4>는 각 소스별 노드와 링크의 수를 정리한 것이다. BGP 데이터는 678개의 노드와 1,273개의 링크로 구성되었으며, BGP 데이터와 IRR 데이터를 합치면, 노드 수는 변화가 없고 링크 수만 4.2% 증가하였다. 마지막으로 모든 소스의 데이터를 토대로 구축한 그래프에는 7개의 노드와 155개의 링크가 추가 되었다.

표 4. 데이터 소스별 그래프 비교  
Table 4. AS Graph Built From Various Sources

Sources	Number of Nodes	Number of Links
BGP data	678	1,273
BGP + IRR data	678	1,327(+4.2%)
BGP + IRR + IXP data	685(+1%)	1,428(+12.2%)

KISA에서 발표한 국내 AS의 총 수는 755개이며(2012년 1월 6일) 우리의 최종 결과는 685개로 70개의 AS가 발견되지 않았다. 이와 같은 결과의 원인은 다음의 세 가지 경우로 생각해 볼 수 있다. 첫째, BGP 데이터를 수집하는 UCLA의 “Internet Topology Collector”의 관측점이 한국 AS들의 정확한 연결 정보를 수집하지 못했거나 둘째, AS번호를

할당 받고 있지만 AS를 운영하지 않는 경우, 셋째, 상위 제공자와 포인트-포인트로 연결하고 BGP를 운영하지 않는 경우 이다<sup>12)</sup>. <표 2>의 IRR 데이터 검증 결과를 보면 800개의 AS중 156개가 “void”나 “void conflict” 테스트를 통과 하지 못함을 보여준다. 특정 AS가 한 달 이상 BGP 라우팅 테이블이나 update에 나타나지 않으면 해당 AS 기록들은 제외된다. 따라서 본 연구에서 발견하지 못한 70개의 AS 노드들도 대부분 AS를 할당받고 운영하지 않거나, 상위 제공자와 포인트-포인트로 연결하고 BGP를 운영하지 않는 경우라서 본연구의 위상 데이터에 나타나지 않은 것으로 판단된다. 이와 같은 상황을 몇 개 기관에서 확인할 수 있었다.

위에서 구한 한국 위상 데이터로부터 AS의 차수(j)를 기준으로 국내 AS를 <표 5>와 같이 분류하였다. 국내 AS의 94%는 5이하의 차수를 가지는 stub로 분류되며, 5%의 Small ISP, 1%의 Large ISP로 구성되어 있다. 이중 차수가 50이상인 Large ISP가 6개로 나타났는데 국내 IXP를 포함한 대형 ISP 업체들임을 확인 할 수 있었다.

표 5. 한국 AS 분류  
Table 5. Korea AS Classification

AS type	Criteria	Number
Tier-1	no provider	0
Large ISP	$j \geq 50$	6
Small ISP	$5 \leq j < 50$	36
Stub	$j < 5$	643

#### IV. 한국 AS 수준 인터넷 위상의 power-law 특성 분석

본 장에서는 BGP, IRR, IXP 위상 데이터로부터 구축한 한국 AS수준의 위상 그래프를 이용하여 한국의 AS 수준 인터넷 특성을 분석하였다. 전 세계 AS 수준의 인터넷 특성은 UCLA에서 제공하는 “Internet Topology Collector”의 한 달간(2012년 1월) 자료를 통합한 데이터를 이용하여 분석하였다. 가장 기본적인 특성인 노드 수, 링크 수, 평균 노드 차수를 분석하고, 위상 특성 분석에 가장 빈번하게 사용하는 차수분포를 분석하였다. <표 6>은 한국 AS 수준의 인터넷 위상에 대한 특성을 분석한 결과이다.

그래프의 가장 기본적인 특성은 노드의 수(n)와 링크의 수(m) 그리고 노드가 갖고 있는 평균 링크

의 수인 노드의 평균 차수( $\bar{k} = 2m/n$ ) 이다. 노드의 평균 차수는 위상의 연결을 나타내는 가장 기본적인 특성으로 한국의 평균 노드 차수는 4.17로 전 세계 평균 차수인 7.09보다 낮게 나타났다. BGP 데이터만으로 분석할 경우 평균 노드 차수가 3.75( $= 2*1,273/678$ )로 더 낮은 값을 갖는다. 평균 차수가 크면 평균적으로 노드 간에 잘 연결되어 있음을 나타낸다. 그러나 두 개의 그래프가 같은 평균 차수를 가지더라도 두 그래프의 위상 구조가 서로 다를 수 있기 때문에 위상의 세부적인 특성을 나타내는 데는 한계가 있다.

표 6. 노드 평균 차수 특성  
Table 6. Characteristic of Average Node-degree

		Global AS graph	Korea AS graph
Average degree	Number of Nodes (n)	40,023	685
	Number of edges (m)	141,842	1,428
	Avg. node degree ( $\bar{k}$ )	7.09	4.17

임의의 노드를 선택 했을 때 특정 차수의 확률을 나타내는 노드 차수 분포가 위상 특성연구에 가장 빈번하게 사용되고 있는데 평균 차수보다 그래프의 더 많은 연결 정보를 제공해 준다. Faloutsos et al.[4]은 1997년 말에서 1998년 말까지 RouteView를 이용하여 구한 BGP 데이터를 토대로 AS 수준에서 인터넷 위상이  $y \propto x^\alpha$  형태인 power-law에 의해 묘사될 수 있음을 보였다. 인터넷의 차수 분포가 power-law 분포를 따른다는 사실은 인터넷 위상 연구에 큰 영향을 주었다. Power-law는 높은 차수의 노드 수는 적고 낮은 차수의 노드는 많은 것을 의미한다. 한국 AS의 93.7%는 평균 차수 4.17보다 작은 것으로 조사되었다.

국내, 전 세계 AS 수준의 인터넷 위상의 노드 차수 분포의 power-law 특성을 살펴보았다. 모든 노드를 고려하여 power-law 특성을 살펴보면 상관계수가 국내 0.718, 전 세계 0.885로 power-law를 따른다고 보기에 무리가 있다. (그림 4) 에서처럼 그래프의 오른쪽 하단의 높은 차수의 노드들은 차수의 변화에 따라 빈도가 거의 변하지 않기 때문이다. Faloutsos et al.[4]은 그래프 내에서 높은 차수를 가지면서 출현 빈도가 1인 노드들을 제거한 후 분석을 시도 하였는데 본 연구에서도 같은 방법을 사용 하

였다. 상관계수가 가장 크게 나올 때 까지 노드를 제거 하였는데, 한국 데이터 에서는 빈도수가 1인 높은 차수의 노드 12개(1.8%)를 제거하였고, 전 세계 데이터는 상위 264개(0.6%) 노드를 제거하였다. 상위 차수의 노드를 제거한 뒤 power-law 특성을 살펴보면 한국의 power-law 지수는 2.49, 상관계수 0.931, 전 세계 인터넷 power-law 지수는 1.92, 상관계수 0.979로 양쪽 모두 power-law 특성을 따르는 것을 확인 할 수 있었다(그림 5). 한편 최대 노드 차수는 세계가 3,675 그리고 한국이 384로 나타났다. 한국 인터넷 AS 그래프에서 비록 모든 노드를 포함한 노드 차수 분포는 power-law 분포를 따른다고 볼 수 없지만, 상위 1.8%의 노드를 제외한 나머지 98.2%의 노드의 노드 차수 분포는 power-law 분포를 따르는 것으로 나타났다. 노드 차수 분포를 살펴보면 2보다 작은 차수인 1이 상당 수 존재한다. AS번호의 할당 정책(stub AS의 경우 노드 차수가 최소 2개 이상일 경우 AS 번호 부여)이 정확하게 시행되고 있다면, hidden 링크와 invisible peer 링크이기 때문에 발견되지 않은 경우가 대부분일 것이다. 따라서 모든 hidden 링크와 invisible peer 링크가 포함된 완전한 위상 데이터를 구축할 수 있다면 모든 노드는 2 이상의 차수를 갖게 될 것이다.

### V. 결 론

본 논문에서는 먼저 인터넷 AS 수준의 위상 구축을 위해 사용 가능한 데이터들을 모두 조사 하였다. 이들에 대한 비판적인 분석 과정을 통해 한국의 실정에 적합하고 또한 획득 가능한 데이터 소스 들을 유도 하였다. UCLA의 “Internet Topology Collection”을 이용한 BGP 데이터 그리고 IRR 과 IXP 데이터로부터 실제 국내 인터넷에 최대한 근접한 위상 데이터를 추출하고 검증 절차를 거쳐 한국 인터넷의 AS 수준 위상 구축을 시도 하였다.

순수하게 BGP 데이터만을 이용했을 때보다 7개의 노드(1% 증가)와 155개의 링크(12.2% 증가)를 추가로 확인할 수 있었다. 최종적으로 이번 연구에서 얻은 국내 인터넷의 AS 수준 그래프는 685개의 노드와 1,428개의 링크를 갖고 있다.

기존 연구[13]는 BGP update 메시지만을 사용하여 노드와 링크 수를 각각 659 개와 1304 개로 발표 하였다. 따라서 본 연구 결과가 각각 4%와 9.5% 증가한 수치를 보여준다. UCLA의 IRL 데이터를 이용 했을 때보다 IRR 데이터를 토대로 49개

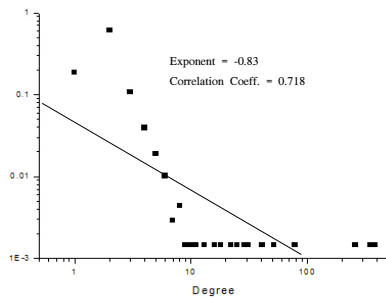


의 hidden 링크인 backup 링크를 찾아냈으며 KISA의 IXP 데이터를 토대로 69개의 provider-customer 링크와 32개의 peer-peer 링크를 더 수집 하였다. 아울러 본 연구에서는 보다 정확하고 객관적으로

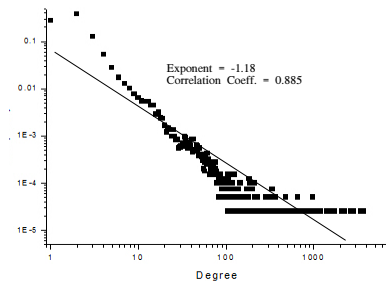
표 7. 노드 차수 분포 특성

Table 7. Characteristic of Node Degree Distribution

		Global AS graph	Korea AS graph
Degree Distribution	Max node degree	3,675	384
	Exponent of P(k)	1.92	2.49

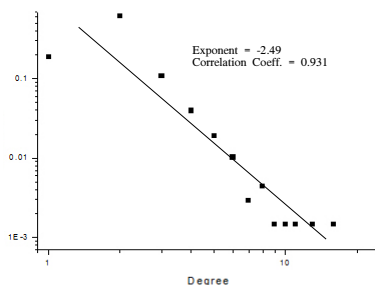


(a) Korea

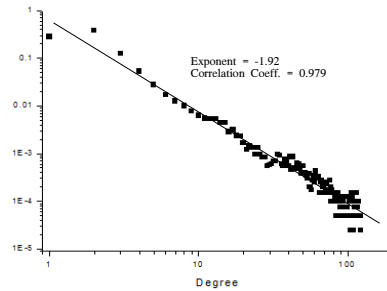


b) Global

그림 4. 차수 빈도 분포(모든 노드 포함)  
Fig. 4 . Node Degree Distribution (all nodes are included)



(a) Korea



(b) Global

그림 5. 차수 빈도 분포(차수가 1인 상위 노드 일부 제거)  
Fig. 5. Node Degree Distribution (eliminating high degree nodes)

타당한 AS 수준의 위상 구축을 위해서 사용하는 인터넷 위상 데이터 소스 들에 대한 검증과 분석을 진행하였기 때문에 연구 결과의 정확도나 최신성도 확보 할 수 있었다.

국내 인터넷 AS 노드의 평균 차수는 4.17로 전 세계 평균 차수인 7.09보다 낮게 나타났다. 노드 차수가 1인 상위 차수 1.8%의 노드를 제외한 나머지 노드들만으로 분석한 국내 노드 차수 분포의 power-law 지수는 2.49이며 상관계수가 0.931로 비교적 power-law 분포를 잘 따른다고 볼 수 있다.

본 연구에서는 인터넷 위상 특성을 분석하기 위해서 그래프의 평균 차수와 노드 차수 분포에 관한 특성을 살펴보았다. 하지만 평균차수와 노드 차수 분포 특성은 인터넷 토폴로지 특성 중 일부뿐이며, 인터넷의 위상을 설명하는데 한계가 있다. 따라서 한국 AS 수준의 인터넷 위상을 심층적으로 이해하기 위해서는 향후 노드 차수 관련 특성 외에 노드 간의 연결성(connectivity) 이나 망의 안전성(robustness) 라우팅 효율성(efficiency) 등을 나타내 줄 수 있는 다른 특성에 관한 연구가 추가적으로 필요하다.

현재 국내에서는 인터넷 위상 연구를 위한 기초 데이터를 수집, 제공하는 어떠한 단체나 기관이 존재하지 않는다. 본 연구에서 가장 많은 정보를 제공하는 BGP 데이터는 UCLA에서 제공하는 데이터 중 한국 AS 관련 부분만을 발췌하여 사용하였다. 그런데 관측 지점들(monitoring points)의 위치가 한국과는 무관하게 설정 되어있기 때문에 데이터의 정확성이나 최신성이 떨어질 수 밖에 없다. III-2절의 <표 2>에서 확인 하였듯이 APNIC에서 제공하는 한국 IRR데이터도 조사 결과 많은 수가 사용할 수 없는 데이터였다. 국내에서 인터넷 위상 관련 연

구가 보다 활발히 진행되려면 우선적으로 보다 정확한 그리고 최신의 위상 데이터를 획득할 수 있는 체계가 정립되어야 할 것이다.

### References

- [1] Beichuan Zhang, Raymond Liu, Daniel Massey, Lixia Zhang, "Collecting the internet AS-level topology," *ACM SIGCOMM, Computer Communication Review*, vol. 35 2005, pp. 53-61
- [2] Hamed Haddadi, Damien Fay, Almerima Jamakovic, Olaf Maennel, Andrew W. Moore, Richard Mortier, Steve Uhlig, "On the importance of local connectivity for internet topology models," *inn 21st International Teletraffic Congress(ITC 21)*, 2009
- [3] Hyunseok Chang, Ramesh Govindan, Sugih Jamin, Scott J. Shencker, Walter Willinger, "Toward capturing representative AS-level Internet Topologies," *Computer Networks* 44(6), 2004, pp. 737-755
- [4] M. Faloutsos, P. Faloutsos and C. Faloutsos, "On power-law relationships of the internet topology," *In Proc. of ACM SIGCOMM, 1999*, pp. 251-256.
- [5] Georgos Siganos, M. Faloutsos, P. Faloutsos and C. Faloutsos, "Power-laws and the internet topology," *IEEE/ACM Trans. On Networking* vol.11, no. 4, 2003, pp. 514-524.
- [6] Priya Mahadevan, Dmitri KRIoukov, Marina Fomenkov, Bradley Huffaaker, Xenofontas Dizmitropoulos, Kc clsffy, Amin Vahda, "The internet AS-level topology: Three data source and one definitive metrics," *ACM SIGCOMM Computer Communication Review*, 2006, pp. 17-26
- [7] Ricardo Oliveira, Dan Pei, Walter Willinger, Beichuan Zhang, Lixia Zhang, "Quantifying the completeness of the observed internet AS-level structure," *Technical Report TR-080026, Computer Science Department UCLA*, 2008.
- [8] Ricardo Oliveira, Dan Pei, Walter Willinger, Beichuan Zhang, Lixia Zhang, "In search of the ground truth: The internet's AS-level connectivity structure," *In Proc. ACM SIGMETRICS'08*, 2008.
- [9] Shi Zhou, Guo-Qiang Zhang, Guo-Qing Zhang, "Chinese internet AS-level topology," *IET Communications*, vol.1 no. 2, 2007, pp. 209-214
- [10] Tian Bu, Don Towsley, "On distinguishing between internet power law topology generators," *In Proc. IEEE INFOCOMM*, 2002, pp. 638-647.
- [11] Yihua He, Georgos Siganos, Michalis Faloutsos, Srikanth Krishnamurthy, "A systematic framework for unearthing the missing links: Measurements and impact," *In proc. USENIX/SIGCOMM NSDI'07*, 2007, pp. 187-200
- [12] Yihua He, Georgos Siganos, Michalis Faloutsos, "Internet topology," *In R. Meyers, editor, Encyclopedia of complexity and System Science*, 2009, pp. 4930-4947
- [13] KooHong Kang, "Power-Laws exponents of the domestic AS-level internet topology based on the valley-free BGP routing policy," *Korean society for Internet Information*, 11-4, 2010, pp.41-49
- [14] UCLA IRL Internet Topology Collection, <http://irl.cs.ucla.edu/topology/>, 2012
- [15] The list of AS number in Korea, <https://ip.kisa.or.kr/main.html>, 2012
- [16] Map of Korea AS connection at KISA, <http://isis.kisa.or.kr/>, 2011
- [17] APNIC FTP server for Korea IRR data, <http://ftp.apnic.net/apnic/dbase/data/>, 2012
- [18] CAIDA website, [www.caida.org/measurement/skitter/](http://www.caida.org/measurement/skitter/), 2012
- [19] RouteView website, <http://www.routeviews.org/>, 2012
- [20] RIPE/RIS website, <http://www.ripe.net/data-tools/stats/ris/routing-information-service>, 2012
- [21] Packet Cleaning website, House <http://www.pch.net>, 2012

[22] Peeringdb website, <http://www.peeringdb.com>, 2012

[23] Euro-IX website, <http://www.euro-ix.net>, 2012

오 동 익 (Dong Ik Oh)



2002년 2월 수원대학교 전자  
재료공학과 졸업

2007년 8월 서울과학기술대학  
교 산업정보시스템 (공학석  
사)

2009년 3월~현재 서울과학기  
술대학교 산업정보시스템

박사과정

<관심분야> 정보통신, 무선통신, 통신시스템 성능평  
가

이 강 원 (Kang Won Lee)



1980년 서울대학교 공과대학  
산업공학(공학사)

1982년서울대학교대학원(공학  
석사)

1985년 Kansas State Univ.  
U.S.A 공학박사(산업공학)

1985년 한국전자통신연구원 TDX

개발단

2012년 현재 서울과학기술대학교 산업정보시스템공  
학과 교수

<관심분야> 정보통신, 품질 및 신뢰성, O.R.