TCP/IP Networking
Domain Name System

Hyogon Kim

Korea University

Introduction

- DNS is the most frequently used application level protocol
- But unlike other application level protocols, it forms the Internet infrastructure
- Born in 1984, standardized in 1987 [RFC 1035]
  - Pre-DNS era : hosts.txt file maintained at SRI

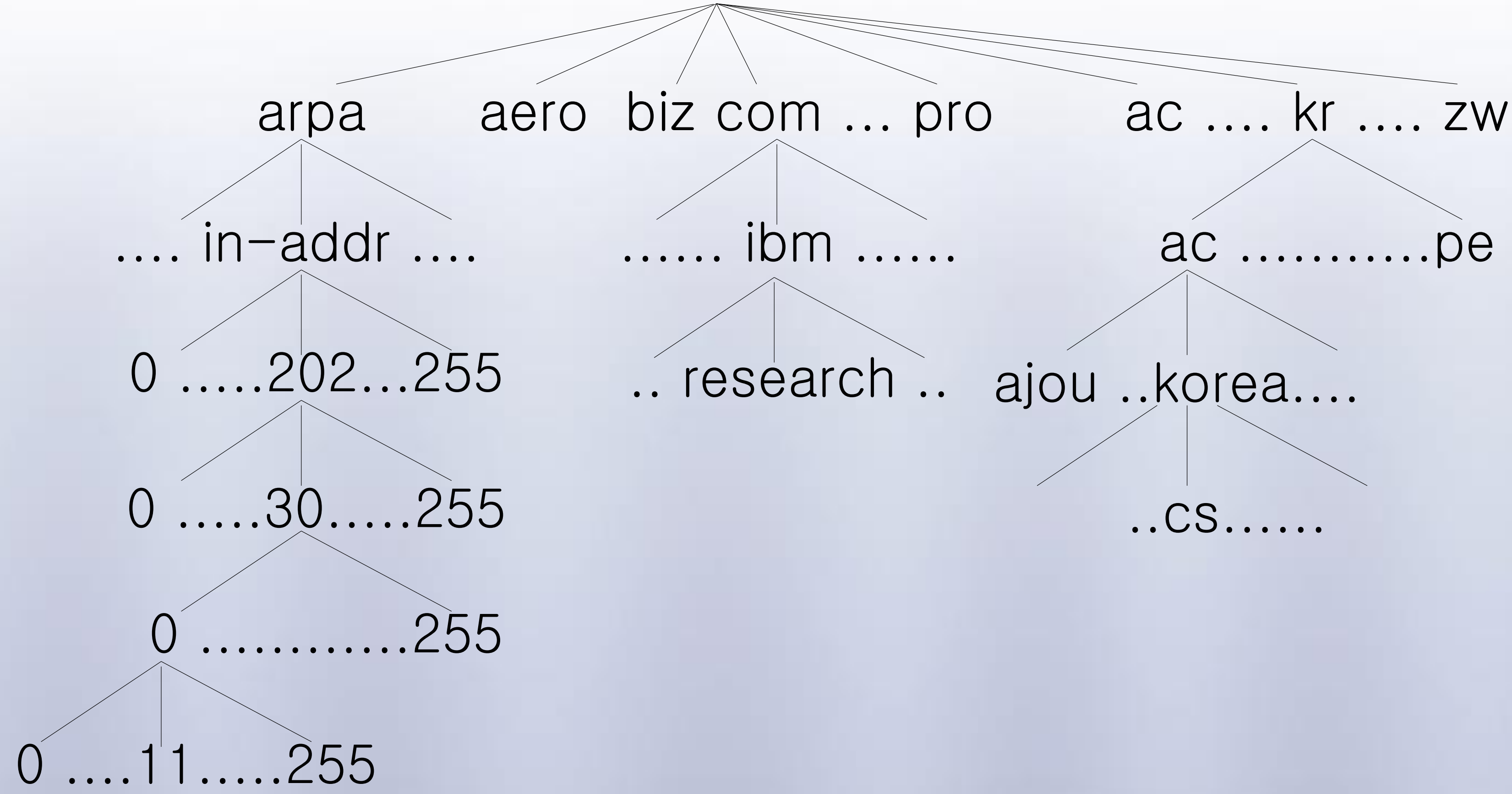
Why domain name?

- The IP Internet only recognizes *IP address*, a 32-bit number, for delivery
  - Likewise, telephone networks only recognize *telephone numbers* for call setup
- Unfortunately, humans are not good at memorizing numbers, so let’s have a “nickname” for an IP address
- For *translation*, let’s have DNS

What is a “domain”?

- A domain is a *naming* domain
  - In Korea University, only .korea.ac.kr allowed
- *Domain name space* is the global, logical, and hierarchical (tree-shaped) naming structure
  - A domain is a subtree of the domain name space
  - The domain name is the “name” of the domain
- Physical manifestation of the domain name space is a *distributed database*

Domain Name Space



arpa aero biz com … pro ac …. kr …. zw

…. in-addr …. …… ibm …… ac ……….pe

0 …..202…255 .. research .. ajou ..korea….

0 …..30…..255 ..cs……

0 ………..255

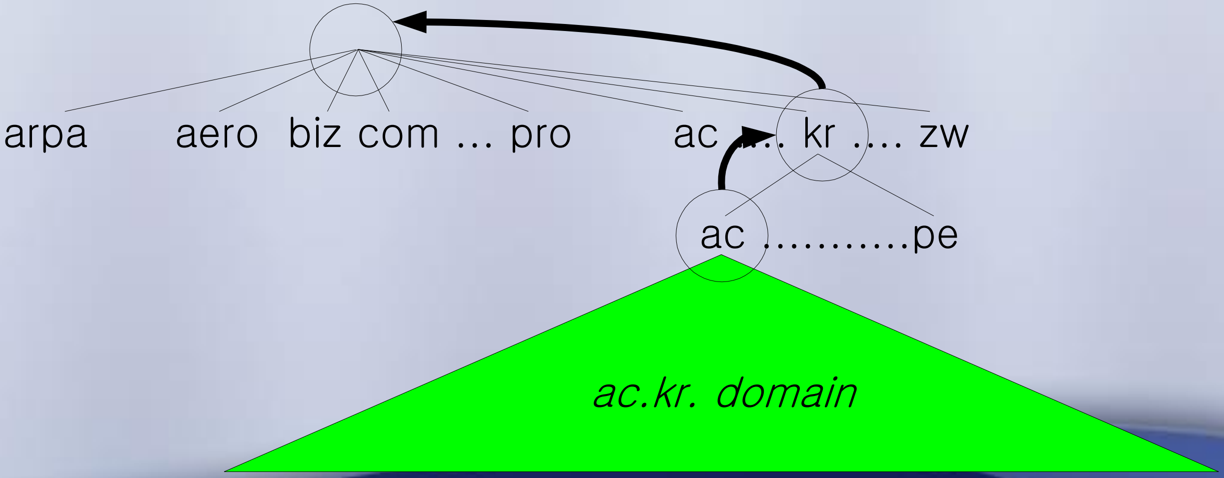
0 ….11…..255

What is a domain name?

- Each node in the domain name tree has a *label*, which is up to 63 bytes
  - The root node has a null label “”
- Domain name is case-insensitive
  - SaMsUnG.Co.kR
  - Samsung.CO.kr
  - SAMSUNG.CO.KR

Domain name

- A domain name of a node is the concatenation of the labels, read from the node through the root node
  - Labels are delimited by "."
  - E.g. "nic.samsung.co.kr."

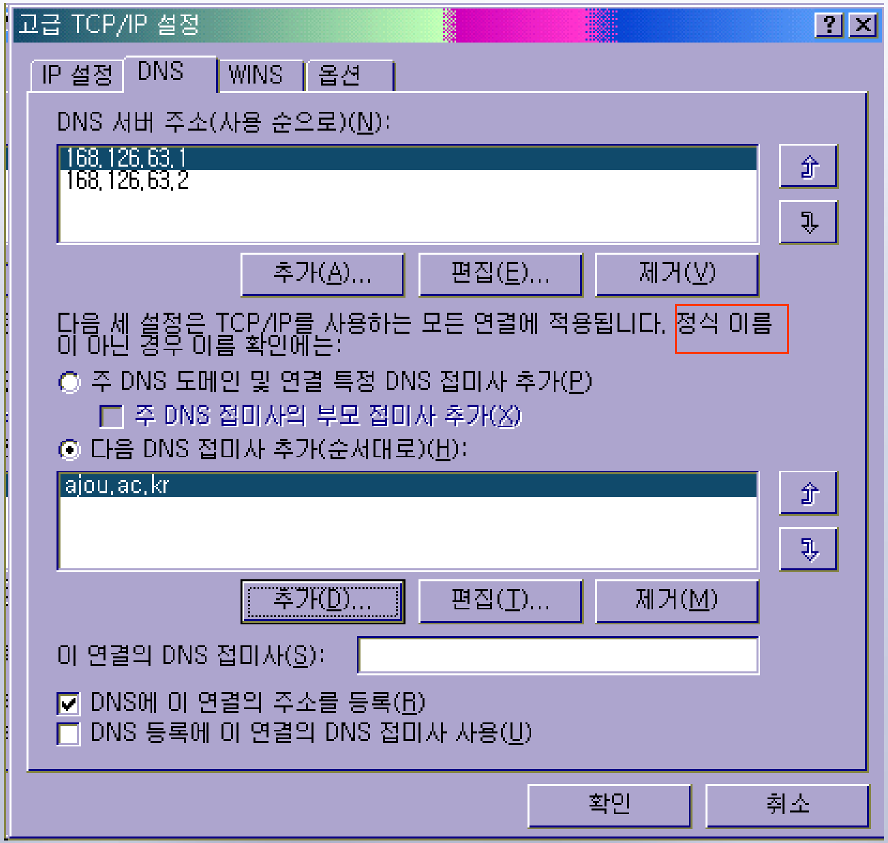


FQDN

- Fully qualified domain name contains the root label
  - imail00 (X), imail00.samsung.co.kr (X)
  - imail00.samsung.co.kr. (O)
- DNS protocol uses only FQDN
- The resolver must complete if the given domain name is not FQDN
  - /etc/resolv.conf

FQDN

```
<hyogon>53% more /etc/resolv.conf
domain ajou.ac.kr
nameserver 202.30.0.11
nameserver 168.126.63.1
```



고급 TCP/IP 설정

DNS 서버 주소(사용 순으로)(N):

168.126.63.1
168.126.63.2

다음 세 설정은 TCP/IP를 사용하는 모든 연결에 적용됩니다. 정식 이름이 아닌 경우 이름 확인에는:

○ 주 DNS 도메인 및 연결 특정 DNS 접미사 추가(P)

□ 주 DNS 접미사의 부모 접미사 추가(X)

● 다음 DNS 접미사 추가(순서대로)(H):

ajou.ac.kr

이 연결의 DNS 접미사(S):

☑ DNS에 이 연결의 주소를 등록(R)

□ DNS 등록에 이 연결의 DNS 접미사 사용(U)

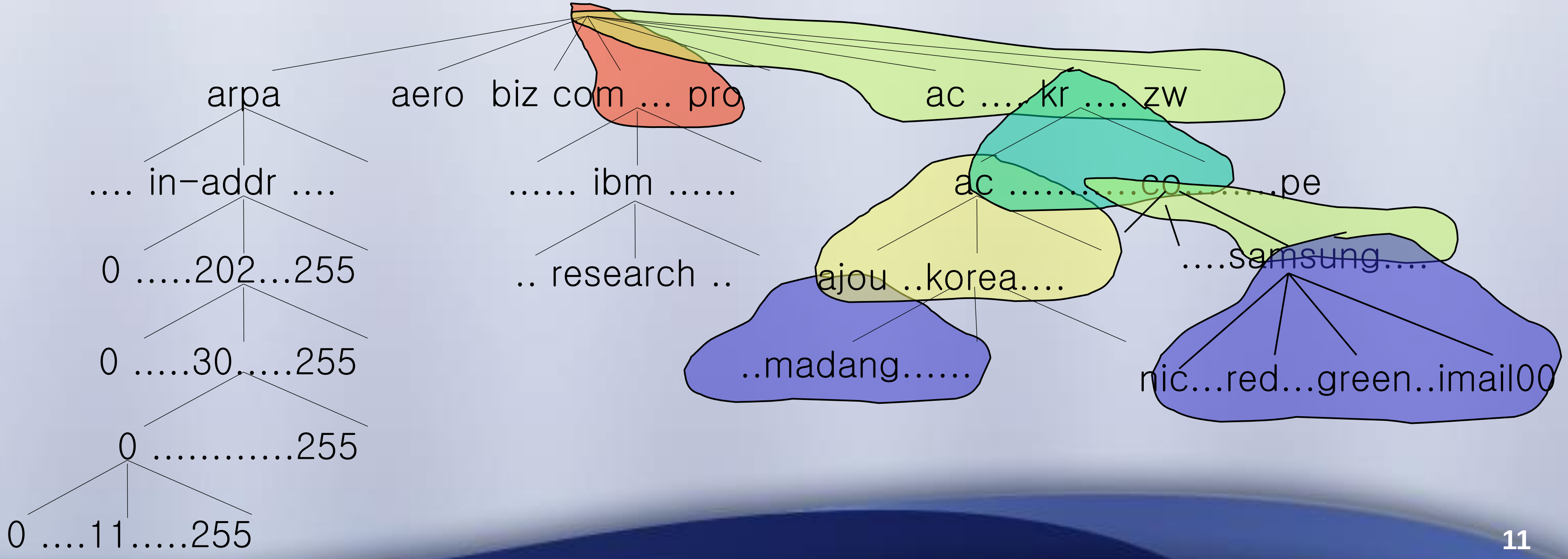
What’s in a domain name?

- A domain name (and the denoted node thereby) can have a set of associated “resource records (RRs)”: e.g., samsung.co.kr has

A	203.254.192.15
NS	nic.samsung.co.kr red.samsung.co.kr green.samsung.co.kr
MX	imail00.samsung.co.kr
SOA	Postmaster: root@nic.samsung.co.kr, etc.

DNS is a distributed database

- A zone is a subset of the domain name space that is physically managed in the same database
  - Each zone has an authoritative name server



Authoritative name servers

- For reliability, multiple authoritative name servers are placed in different locations

```
skku.ac.kr
Name Server:  ajou.ac.kr
Address:  202.30.0.11

Trying DNS
skku.ac.kr      preference = 20, mail exchanger = yurim.skku.ac.kr
skku.ac.kr      nameserver = yurim.skku.ac.kr
skku.ac.kr      nameserver = ns.kreonet.re.kr
skku.ac.kr      nameserver = ns.kaist.ac.kr
yurim.skku.ac.kr        internet address = 203.252.57.2
ns.kreonet.re.kr        internet address = 134.75.30.1
ns.kaist.ac.kr  internet address = 143.248.1.177
```

Authoritative name servers

- When the authoritative name servers are not physically dispersed, the affected domain loses logical connectivity
  - Physical connectivity still exists
  - Access by IP address works fine
- Microsoft incident, Jan. 4, 2002
  - Microsoft authoritative servers are on the same subnet, and the router to the subnet fails
  - Global access to .msnbc.com & .microsoft.com blocked for 2 days

Domain name registration

- Means writing RRs for the domain name in the authoritative name server(s)
- Only the authoritative name server(s) for the registered domain need to update
  - Other authoritative name servers are not affected

Zone

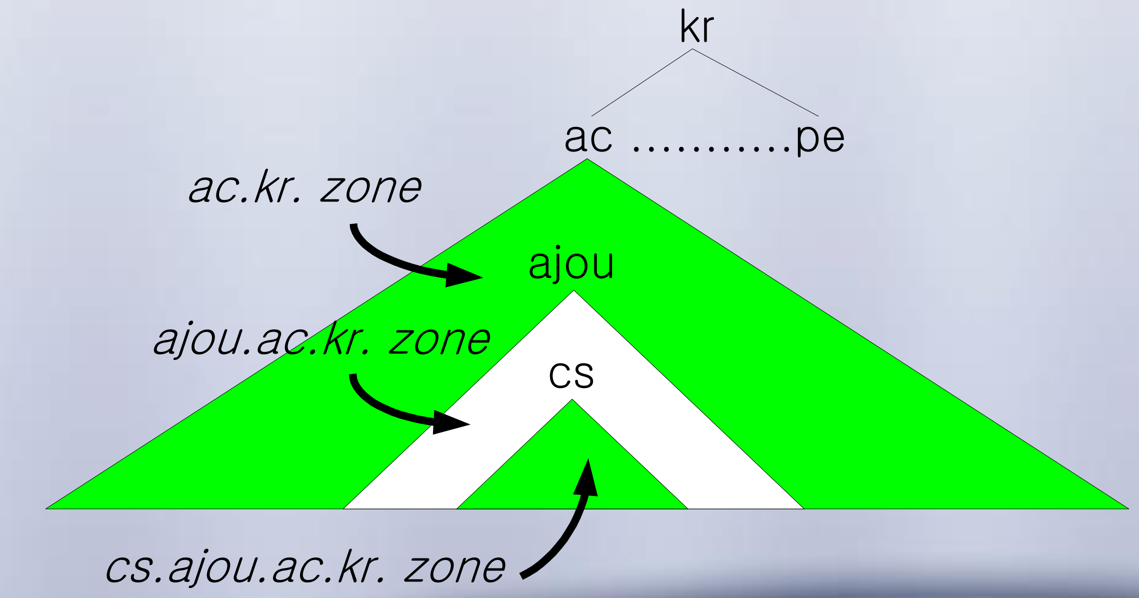
- Authoritative name servers maintain the zone file
- Primary and secondary (both are authoritative)
  - Primary has the zone file in hard disk
  - Secondary gets it from primary upon boot-up – zone transfer

Primary integrity

- Primary authoritative server must maintain integrity!
  - It is the single source of the master copy
- NSI incident, July 18, 1997
  - .com & .net master copy corrupted
  - Distributed to secondaries (secondary root servers)
  - Internet access to .com & .net globally blocked for 4 hours

Delegation

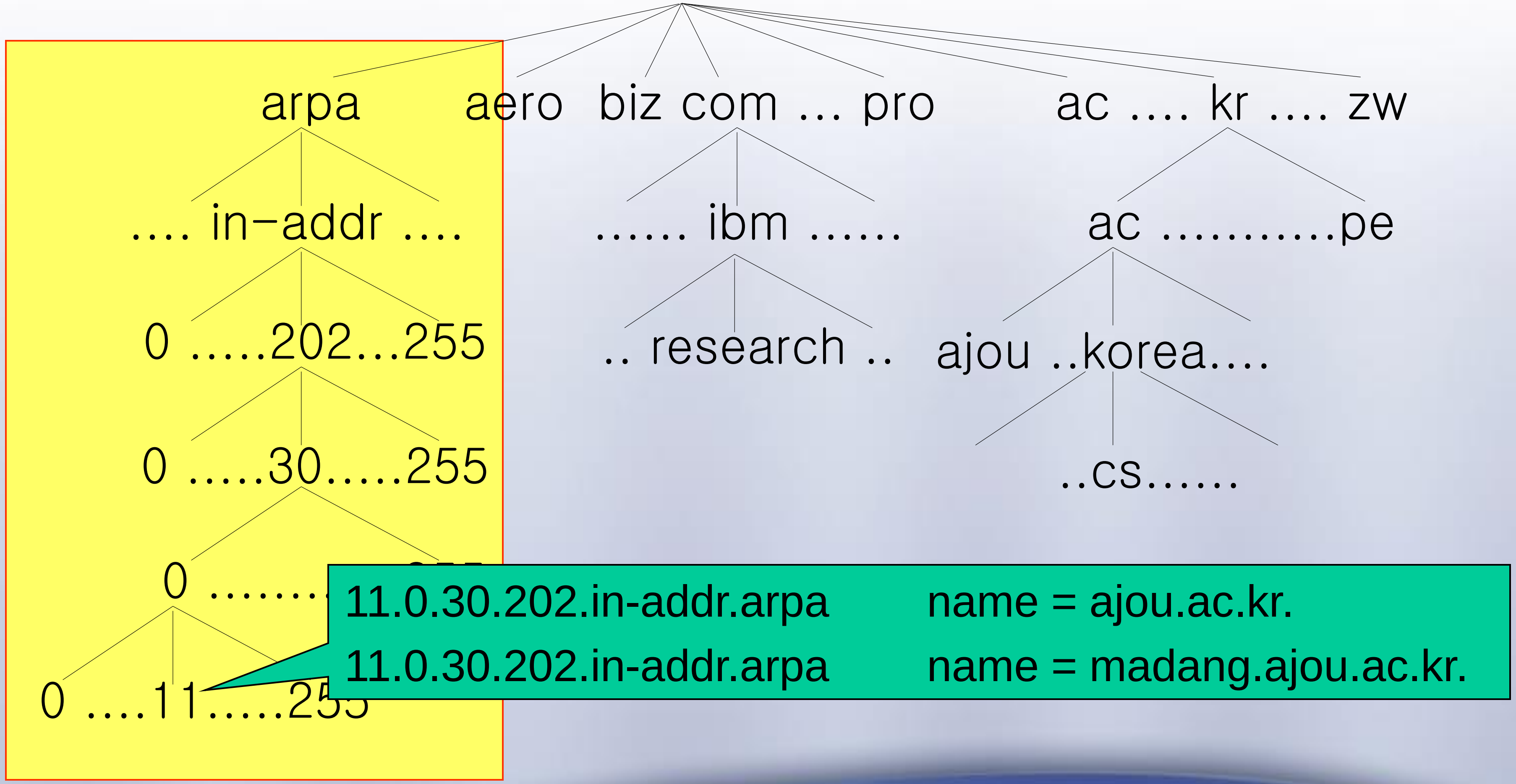
- A zone can be split into multiple sub-zones
- Zones can be separately managed



Top-level domains

- ARPA (address and routing protocol area)
  - Not Advanced Research Project Agency
  - Used for number-to-name mapping
- gTLD (general Top Level Domains)
  - .com, .net, .org maintained by VeriSign
- ccTLD (country-code Top Level Domains)
  - .kp : North Korea – uses its “own (juche?)” DNS
  - .to : tonga – frequently used by “Warez” sites

Domain Name Space



arpa aero biz com … pro ac …. kr …. zw

…. in-addr …. …… ibm …… ac ……….pe

0 …..202…255 .. research .. ajou ..korea….

0 …..30…..255 ..cs……

0 …….

0 ….11…..255

11.0.30.202.in-addr.arpa name = ajou.ac.kr.

11.0.30.202.in-addr.arpa name = madang.ajou.ac.kr.

ARPA domain

- Currently, used for IP address ➔ domain name translation
- If this domain does not exist, reverse lookup (“PTR query) will have to search all A record for the given IP address
- Look like IP addresses but they are domain names

ARPA domain

- IP address w.x.y.z is corresponds to the domain name z.y.x.w.in-addr.arpa
- Why not w.x.y.z.in-addr.arpa? Delegation!
  - in-addr.arpa → root servers
  - 202.in-addr.arpa → APNIC server
  - 30.202.in-addr.arpa → KRNIC server
  - 0.30.202.in-addr.arpa → Ajou server

ARPA zones

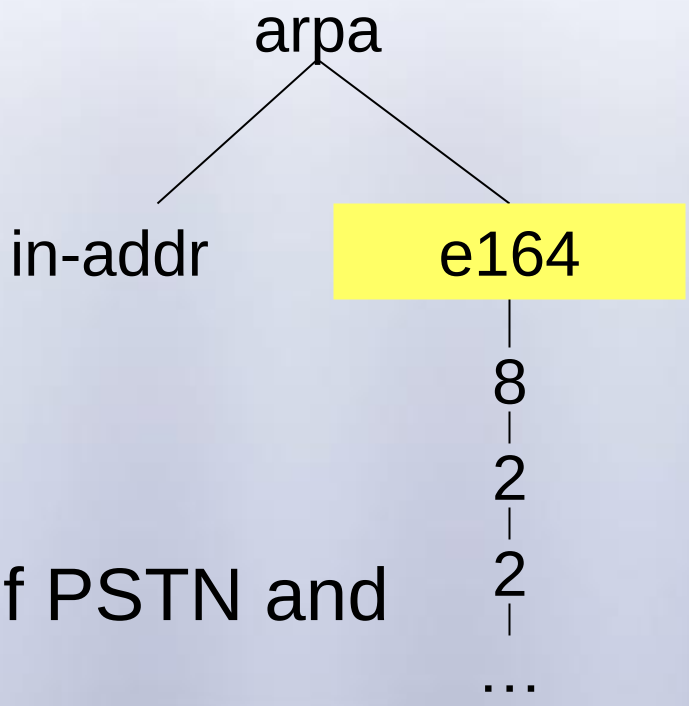
- 11.0.30.202.in-addr.arpa and madang.ajou.ac.kr are in totally different zones!
- Registration of arpa domain names is done separately
  - A record registration does not mean PTR record registration will be done automatically
  - Frequently done for free - in the future?

PTR query

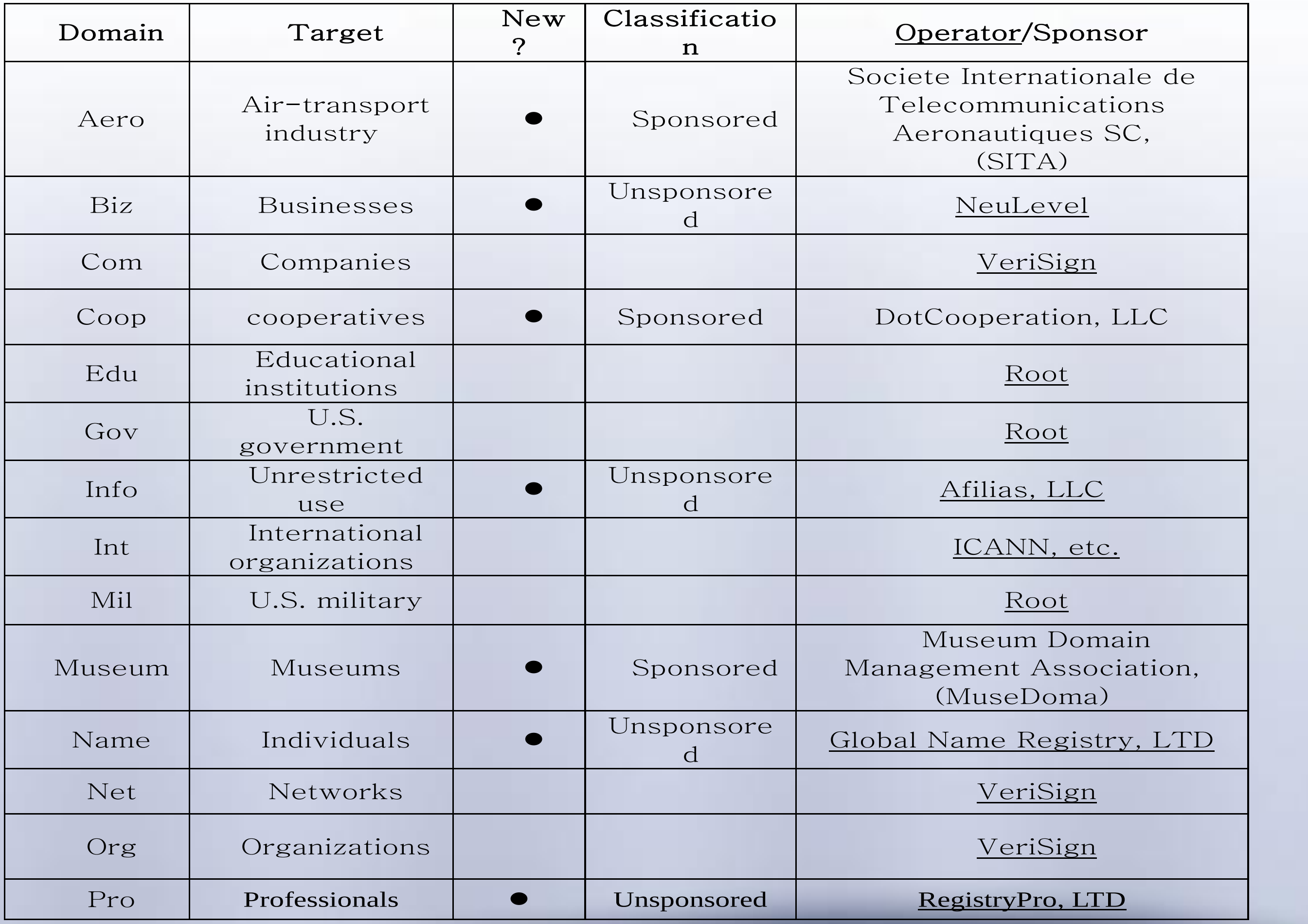
```
> set querytype=PTR
> 11.0.30.202.in-addr.arpa.
Server:         202.30.0.11
Address:         202.30.0.11#53

11.0.30.202.in-addr.arpa        name = ajou.ac.kr.
11.0.30.202.in-addr.arpa        name = madang.ajou.ac.kr.
```

ARPA domain

- Being augmented with ENUM : E.164 identifier to * mapping [RFC 2916]
  - Email address
  - URL
  - SIP address
  - Fax number
  - Telephone number
  - Etc.
- Basically, convergence of PSTN and Internet



General TLD (gTLD)



Domain	Target	New?	Classification	Operator/Sponsor
Aero	Air-transport industry	●	Sponsored	Societe Internationale de Telecommunications Aeronautiques SC, (SITA)
Biz	Businesses	●	Unsponsored	NeuLevel
Com	Companies			VeriSign
Coop	cooperatives	●	Sponsored	DotCooperation, LLC
Edu	Educational institutions			Root
Gov	U.S. government			Root
Info	Unrestricted use	●	Unsponsored	Afilias, LLC
Int	International organizations			ICANN, etc.
Mil	U.S. military			Root
Museum	Museums	●	Sponsored	Museum Domain Management Association, (MuseDoma)
Name	Individuals	●	Unsponsored	Global Name Registry, LTD
Net	Networks			VeriSign
Org	Organizations			VeriSign
Pro	Professionals	●	Unsponsored	RegistryPro, LTD

gTLD

- Incumbent 7 gTLDs
    - Com, net, org, gov, int, mil, edu
- 7 new gTLDs ratified by ICANN
    - General use: biz, info
    - Personal use: name
    - Profit restricted domain: pro
    - Non-profit restricted domains: museum, aero, coop

ccTLD

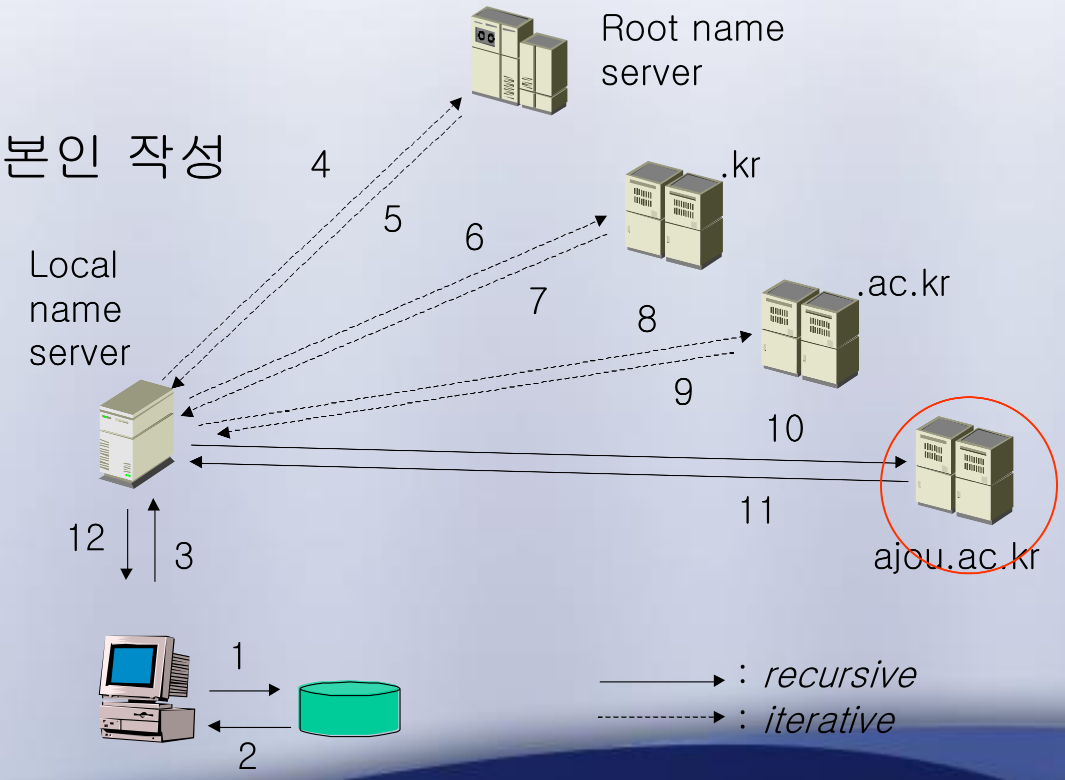
- 244
- Given to every country even if it does not have any Internet infrastructure

Domain	Country	Whois information
Ac	Ascension Island	
Ch	Swiss	
Kp	North Korea	None
Kr	South Korea	Sponsor: KRNIC, Administrative contact: K. Chon (KAIST), Technical contact: C. Park (KRNIC)
Ro	Romania	
To	Tonga	Sponsor: Government of Tonga, Administrative and technical contact: E. Gullischen (Gov. of Tonga)
Zw	Zimbabwe	

What happens when we click on a hyperlink?

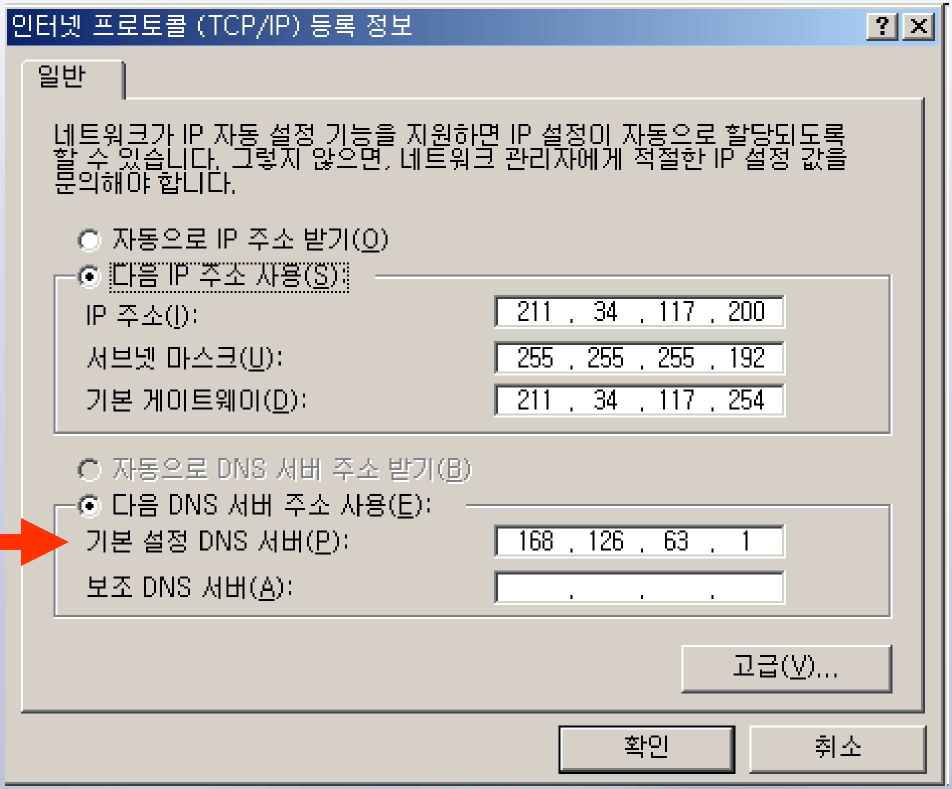
- http://ilab.ajou.ac.kr/talks/200301.html

그림 출처: 본인 작성



“Local” name server

- Is the *default* name server



인터넷 프로토콜 (TCP/IP) 등록 정보

네트워크가 IP 자동 설정 기능을 지원하면 IP 설정이 자동으로 할당되도록 할 수 있습니다. 그렇지 않으면, 네트워크 관리자에게 적절한 IP 설정 값을 문의해야 합니다.

○ 자동으로 IP 주소 받기(O)

● 다음 IP 주소 사용(S):

IP 주소(I): 211 . 34 . 117 . 200

서브넷 마스크(U): 255 . 255 . 255 . 192

기본 게이트웨이(D): 211 . 34 . 117 . 254

○ 자동으로 DNS 서버 주소 받기(B)

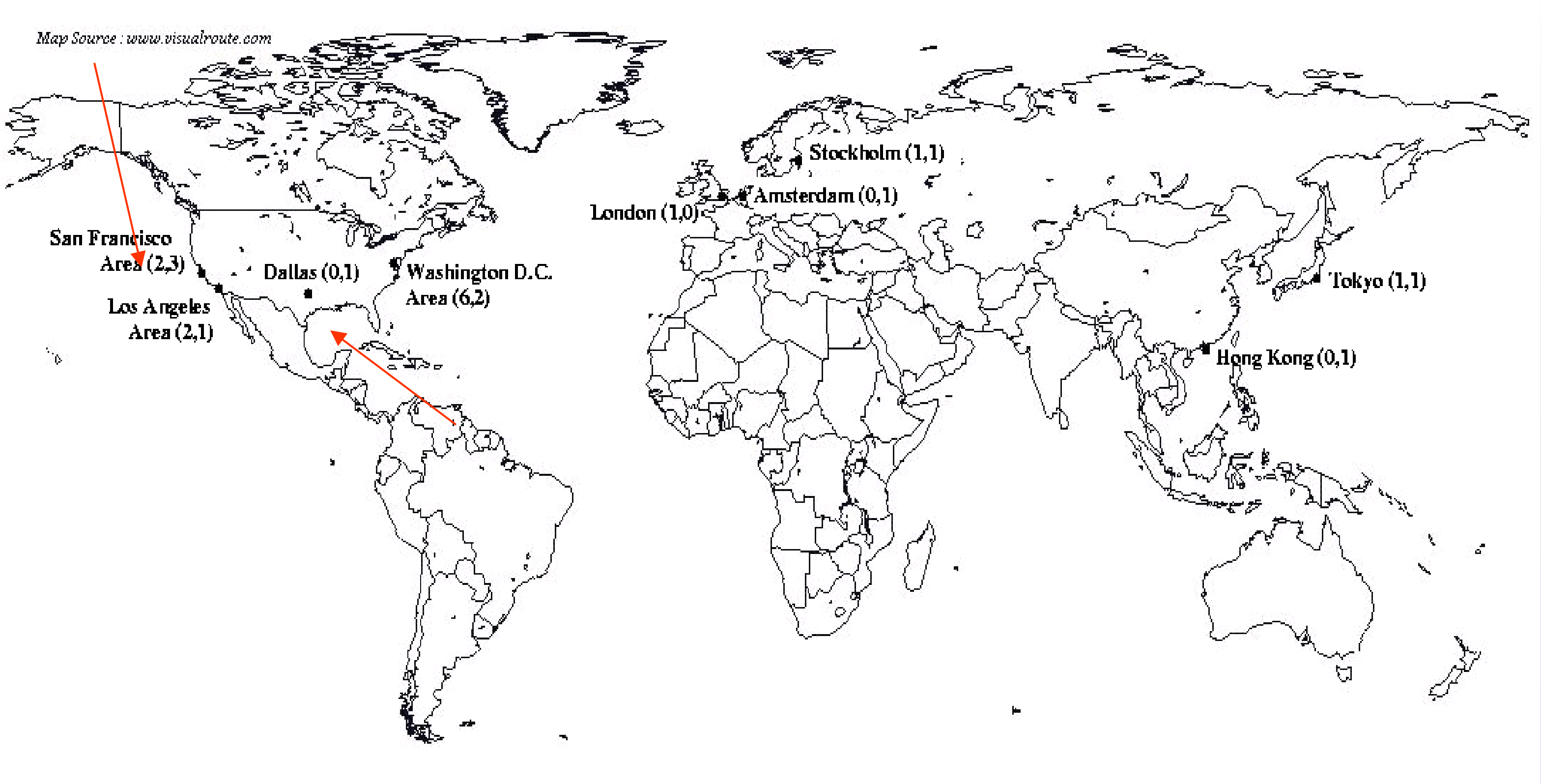
● 다음 DNS 서버 주소 사용(E):

기본 설정 DNS 서버(P): 168 . 126 . 63 . 1

보조 DNS 서버(A): . . .

Servers in the higher hierarchy

- 13 root servers and 13 gTLD servers

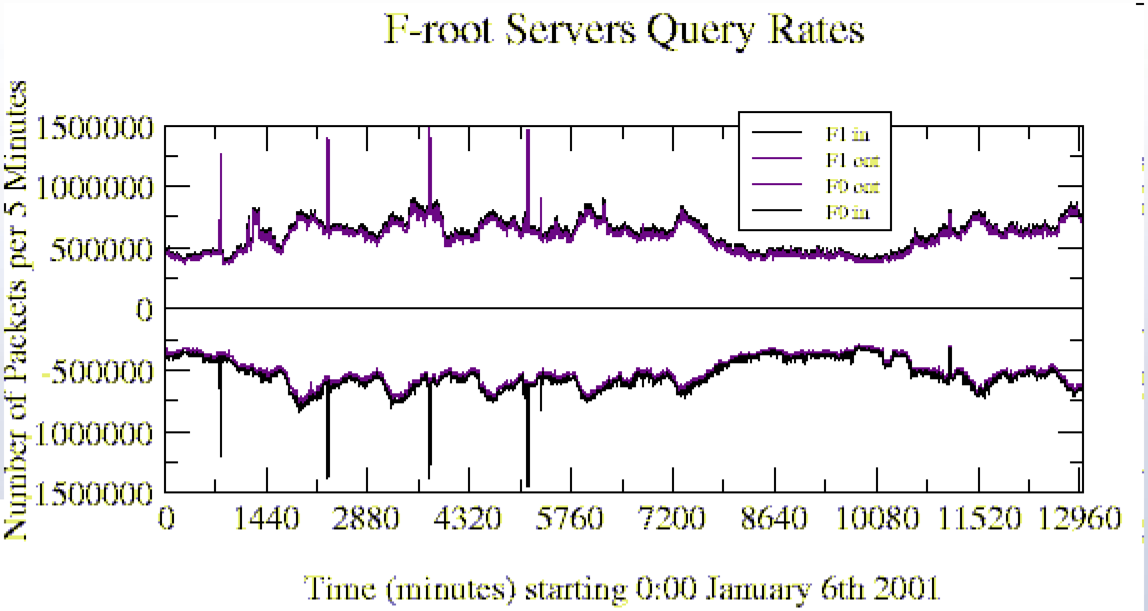


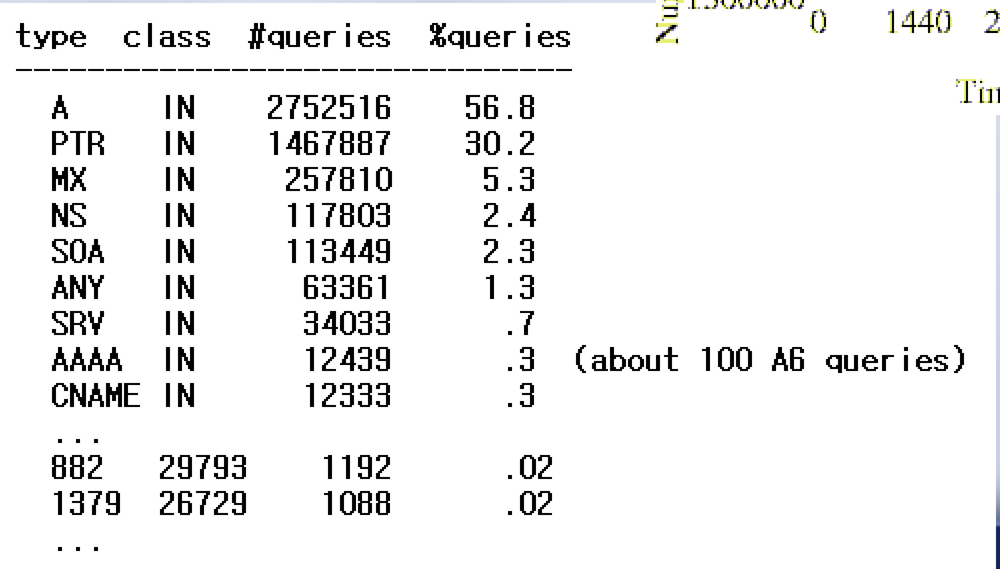
출처: CAIDA
"http://www.caida.org/publications/presentations/ietf0112/dns.damage.html"

Root name servers

- 13: [A-M].root-servers.net
  - A (Virginia, US), D (Maryland, US), H (Maryland, US), I (Stockholm, SE) are most popular
  - C (Virginia, US), G (Virginia, US), J (Virginia, US), K (London, UK), L (California, US), M (Tokyo, JP) are not
- Have NS records for TLDs
  + .edu, .gov, .mil data

Root name servers





```
type  class  #queries  %queries
-------------------------------
  A     IN    2752516    56.8
  PTR   IN    1467887    30.2
  MX    IN     257810     5.3
  NS    IN     117803     2.4
  SOA   IN     113449     2.3
  ANY   IN      63361     1.3
  SRV   IN      34033      .7
  AAAA  IN      12439      .3  (about 100 A6 queries)
  CNAME IN      12333      .3
  ...
  882   29793    1192      .02
  1379  26729    1088      .02
  ...
```

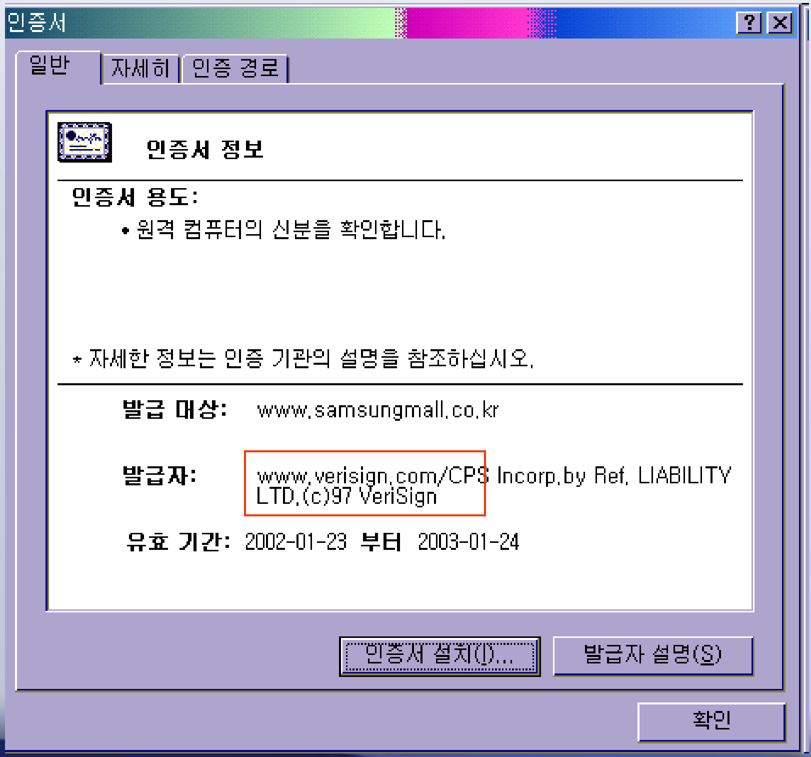
출처: CAIDA "http://www.caida.org/publications/presentations/ietf0112/dns.damage.html"

Root name servers

- 5K queries/s (F), 12K queries/s (A)
- 20% of all queries bogus TLD
  - E.g., .local, .localhost, .msft, .domain, etc.
- 14% are bogus A queries
  - E.g., asking for IP address of an IP address

gTLD servers

- 13: [A-M].gtld-servers.net
- Under the administration of VeriSign
- As of June 2002
  - .com 77.70%
  - .net 13.50%
  - .org 8.79%



인증서 정보

인증서 용도:

- 원격 컴퓨터의 신분을 확인합니다.

* 자세한 정보는 인증 기관의 설명을 참조하십시오.

발급 대상: www.samsungmall.co.kr

발급자: www.verisign.com/CPS Incorp.by Ref. LIABILITY LTD.(c)97 VeriSign

유효 기간: 2002-01-23 **부터** 2003-01-24

DNS protocol

- Runs on UDP, port 53
  - Except for zone transfer and TC=1
  - Packet size is limited to 512B
- Very simple transaction-style
  - Send 1 packet, receive 1 packet

DNS packet format

- Identification matches queries with answers
  - Server and client can be the same, answer can arrive out of order

Identification	Flags
# of questions	# of answer RRs
# of authority RRs	# of additional RRs
Questions (variable #)	
Answers (variable # of RRs)	
Authority (variable # of RRs)	
Additional information (variable # of RRs)	

DNS packet format

- Flags

QR	opcode	AA	TC	RD	RA	0	rcode

  - QR: 1=query, 0=response
  - Opcode=0 [1: inverse → deprecated]
  - Authoritative Answer
  - TrunCated
  - Recursion Desired, Recursion Available
  - Rcode=0

DNS packet format

- Representation of domain names in the packet

6	m	a	d	a	n	g	4	a	j	o	u	2	a	c	2	k	r	0

- Only 1 “long” representation
  - Repetitions are coded as 2B pointers

Queries

- Recursive query asks for answers
  - 70-80% of authoritative servers answers to recursive queries
- Iterative query gets referrals
  - Root servers do not allow recursive queries

Retransmissions

- DNS uses UDP
  - Packet loss must be dealt with by DNS protocol itself
- DNS does not say much about …
  - Client must try other servers before retransmitting the query
  - Retransmissions must be spaced between 2 to 5 seconds

Retransmissions

- BSD client policy
  - Do not send the same query to more than 3 servers
  - Exponentially back-off retransmission timeout after each cycle (3 servers lookup)
  - Servers looked up determines timeout
  - Stop at 4 cycles
    - Max = 4 * 3 = 12

Retransmissions

- BIND (server)
  - Trace RTTs for up to 16 higher level servers
  - Sort the expected RTTs in increasing order
  - Maximum 3 queries per server
    - Max = 16 * 3 cycles = 48
  - Before back-off : $T_{base} = \max(4, 2 \times E[R])$
  - Back-off after each cyle

Questions (in queries/replies)

Query name (variable length)	
Query type	Query class

Domain name (variable length)	
type	class
TTL	
Resource data length	
Resource data	

Answers (in replies)

- Authoritative positive
  - Normal answer from authoritative server
- Positive
  - Answer from non-authoritative server
- Referral
  - Next server to ask (“authority”)
- Negative
  - Even authoritative name server cannot find the answer

Resource records (RR)

- Records corresponding to a domain name
- A domain name can have multiple resource records
    - Not just IP address!

RR types

RR type	용도	Zone file 내 표현 예
A	호스트 도메인 이름 → IP 주소	www.yahoo.co.kr. A 211.32.119.151
NS	Zone의 도메인 이름 → Authoritative name server의 도메인 이름	yahoo.co.kr. NS ns0.yahoo.co.kr.
MX	도메인 이름 → mail server의 도메인 이름	yahoo.co.kr MX 0 mx1.mail.yahoo.com
PTR	IP 주소 → 도메인 이름	151.119.32.211.in-addr.arpa. PTR rc.yahoo.co.kr
CNAME	도메인 이름 (별명) → 정식 이름	www.yahoo.co.kr CNAME rc.yahoo.co.kr.
SRV	서비스,프로토콜, 도메인 → 그 서비스를 제공하는 호스트의 이름과 포트, 우선 순위등의 정보	_http._tcp.example.com. SRV 10 5 80 www.yahoo.com.
HINFO	호스트 도메인 이름 → 호스트 타입 및 OS	
SOA	도메인 (zone) 이름 → Primary authoritative server의 책임자등 정보	

RR types

- NS = Name Server
  - Each zone must have a NS RR with the same name with the zone
  - After delegation, mother zone has NS RR to the child + A record of the child NS (“glue record”)

RR types

- MX = Mail eXchange
  - Small priority values have precedence
- Mail server always tries MX query before sending an email
  - wykim@hp.com → wykim@smtp.hp.com
  - If fails, use A record

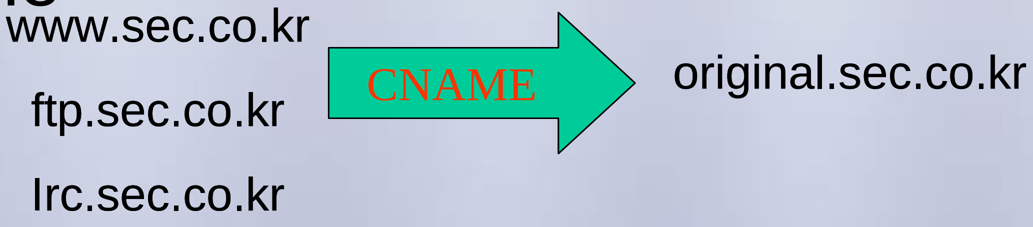
```
> set querytype=MX
> hp.com
Non-authoritative answer:
hp.com  mail exchanger = 50 atlsmtp.hp.com.
hp.com  mail exchanger = 50 palsmtp.hp.com.
hp.com  mail exchanger = 10 smtp.hp.com.
hp.com  mail exchanger = 30 smtpx.hp.com.
```

RR types

- CNAME = Canonical NAME
  - To remember a domain name easily (25% of popular domain names)

```
> www.yahoo.co.kr
Non-authoritative answer:
www.yahoo.co.kr canonical name = rc.yahoo.co.kr.
```

  - To run multiple servers on the same machine



www.sec.co.kr

ftp.sec.co.kr

Irc.sec.co.kr

CNAME

original.sec.co.kr

  - “Alias chain” (even length 4)

RR types

- PTR = PoinTeR

```
$ ftp ftp.tislabs.com
Connected to portal.gw.tislabs.com.
520- This FTP server requires the ability to perform reverse DNS lookups
on all addresses connecting to it.   We cannot perform this on the current
connection.
421 Service not available, remote server has closed connection
ftp>
```

RR types

- SRV = SeRVice
  - Weight for load sharing between equal priority servers

```
_service._protocol.domain priority weight port hostname

_http._tcp.example.com. SRV 10 5 80 www.yahoo.com.
```

RR types

- SOA = Start Of Authority

```
> set querytype=SOA
> ajou.ac.kr
Non-authoritative answer:
ajou.ac.kr
            origin = madang.ajou.ac.kr.
            mail addr = root.madang.ajou.ac.kr.
            serial = 258
            refresh = 10800
            retry = 3600
            expire = 604800
            minimum = 86400
```

  - Origin: primary authoritative name server
  - Hostmaster: root.madang.ajou.ac.kr → root@madang.ajou.ac.kr

RR types

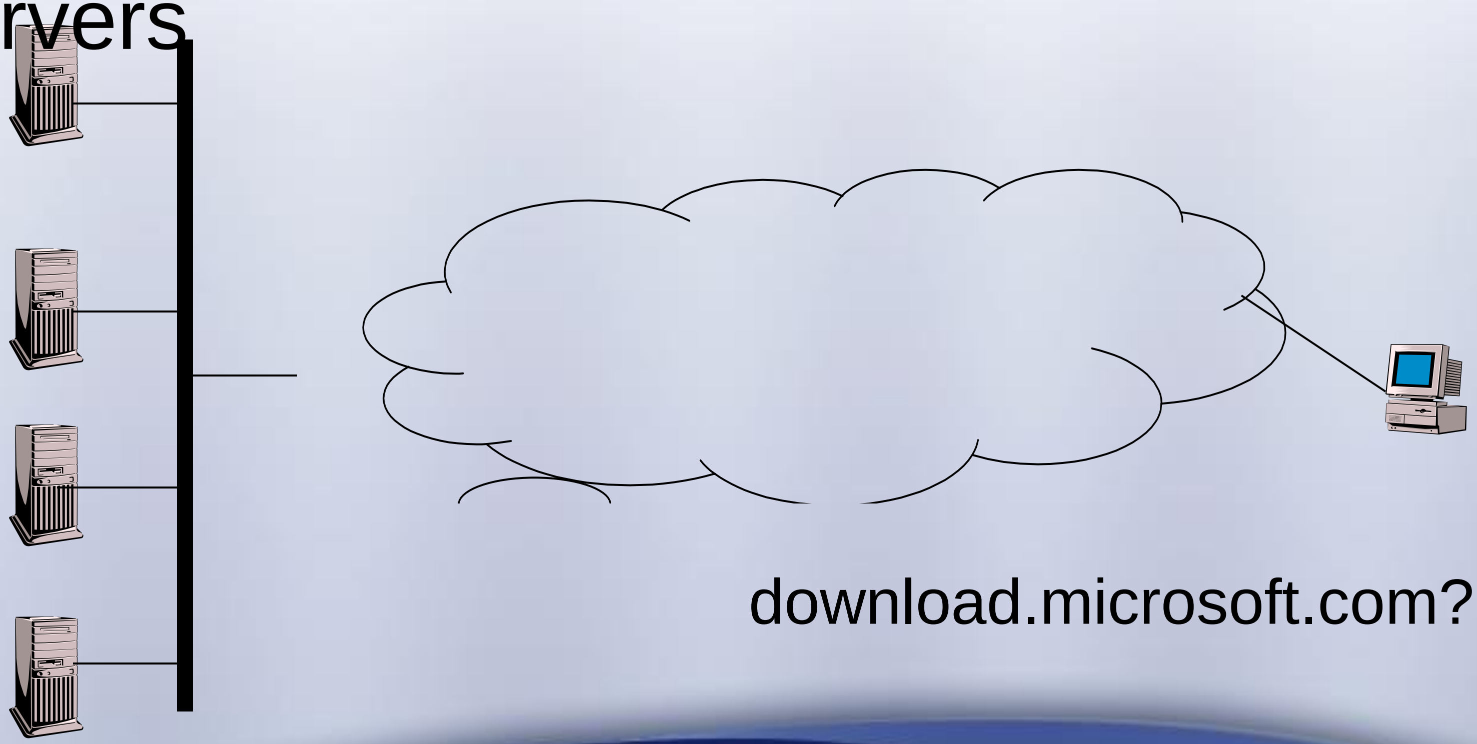
- SOA = Start Of Authority
  - Serial number: zone file version at primary
  - Secondary checks primary for zone transfer every refresh
  - If failed, check after retry until expire
  - Minimum: TTL of RRs in the zone
    - 24 hrs most popular, followed by 1 hr and 1-2 hrs
    - During zone update: < 10min

Caching

- RRs obtained from the authoritative server is kept for a prescribed duration (“TTL”)
  - Reduces the load on the DNS infrastructure
- Microsoft incident, Jan. 4, 2002
  - Microsoft authoritative servers are on the same subnet, router to the subnet fails, load on an observed root server explodes 750-fold for .msnbc.com and .microsoft.com
  - Global access to .msnbc.com / .microsoft.com blocked for 2 days

Caching

- After 2-hour TTL expires, everyone begins to knock on the root name servers



download.microsoft.com?

Negative caching

- BIND 8 and 9, Windows 2000
  - About 90% of servers implement it as of 2001
- When authoritative name server gives <u>negative</u> answer, resolver caches the negative answer for a preset duration
  - Typically 10 minutes
  - Otherwise, top level servers will be harassed by retransmissions

Round-robin DNS

```
Server:  ns.hananet.net
Address:  210.94.0.7


Name:    cnn.com
Addresses:  64.236.16.20, 64.236.16.52, 64.236.16.84,
64.236.16.116, 64.236.24.4, 64.236.24.12, 64.236.24.20,
64.236.24.28
```

Attack on root name servers

- Oct. 22, 2002 but "test" run already recorded on Oct. 7
- DDoS using ICMP (smurf?)
- Last only an hour – stopped before TLD NS RR TTLs expire
  - No visible impact *this time*
- In the wake, VeriSign moves one of its 2 root servers to another location, to a different part of its network
  - These guys obviously didn't learn from the Microsoft incident

Conclusion

- DNS started as a distributed database and a companion protocol mainly to implement name-to-address mapping
- DNS has evolved into a critical infrastructure for modern Internet
  - Indispensable for N2A, A2N, email, VoIP, etc.
  - Imagine a world without Phonebook nor 114!
- Caching plays a vital role to maintain the performance
- U.S. has a vested interest in managing TLDs to not lose the control of the Internet