TCP/IP Networking
Domain Name System

10강

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

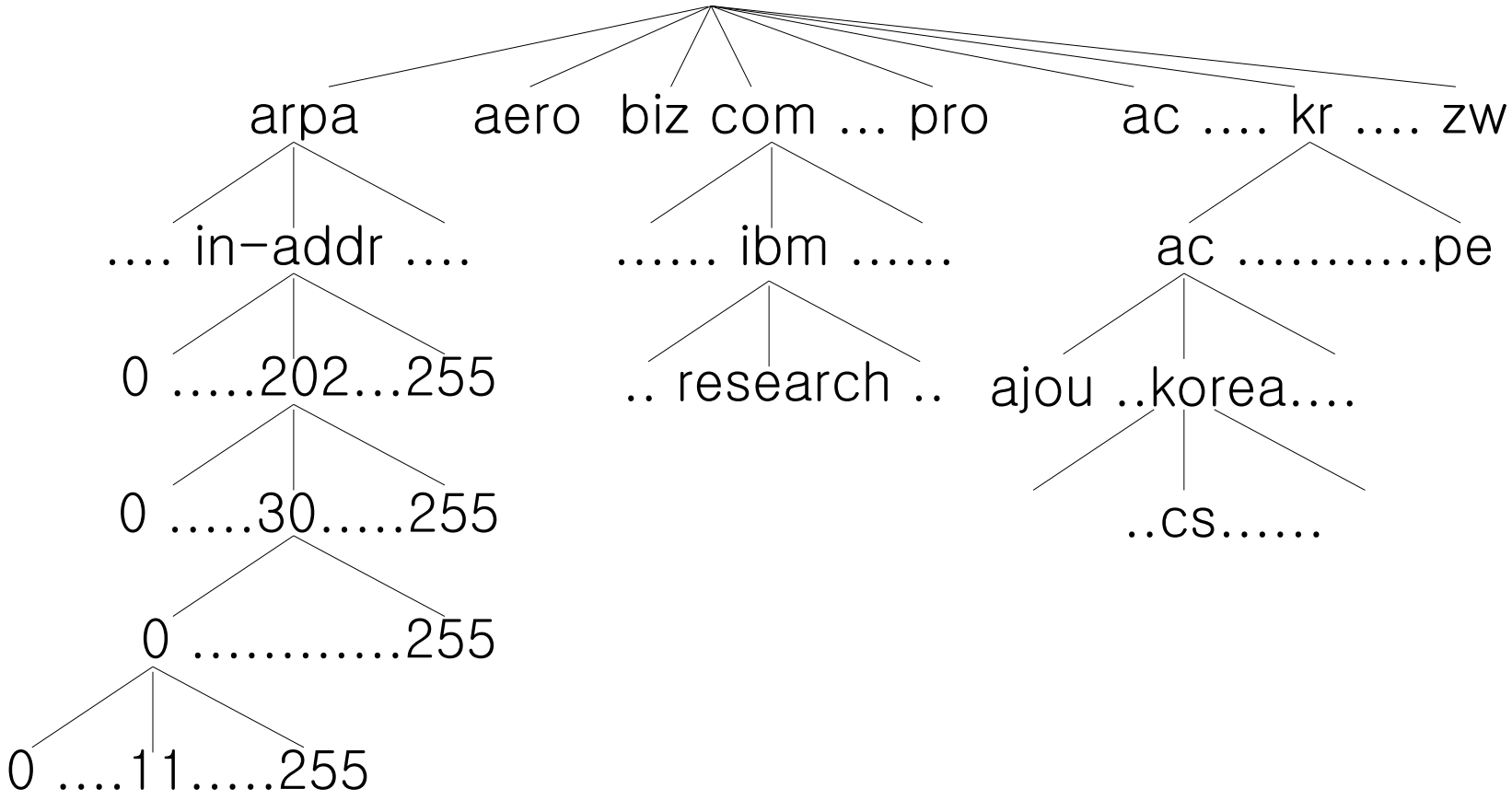
Why domain name?

- 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

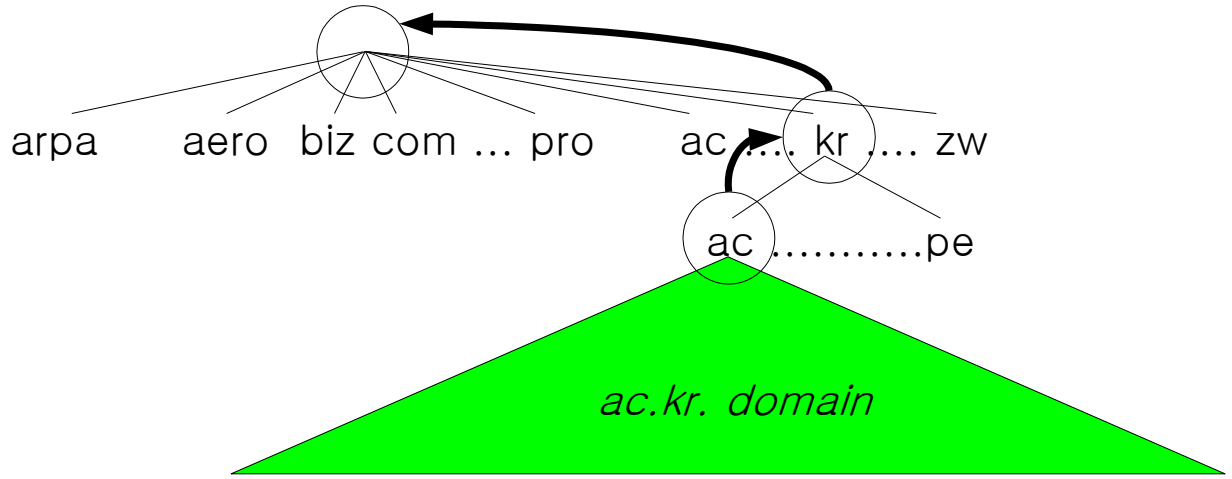


What is a domain name?

- Each node in the domain name tree has a *label*, which is up to 63 bytes
  - The root node has a null label “”
- Punycode for international labels
  - “우리나라” → “xn--910bs4k7yb12w”
- 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

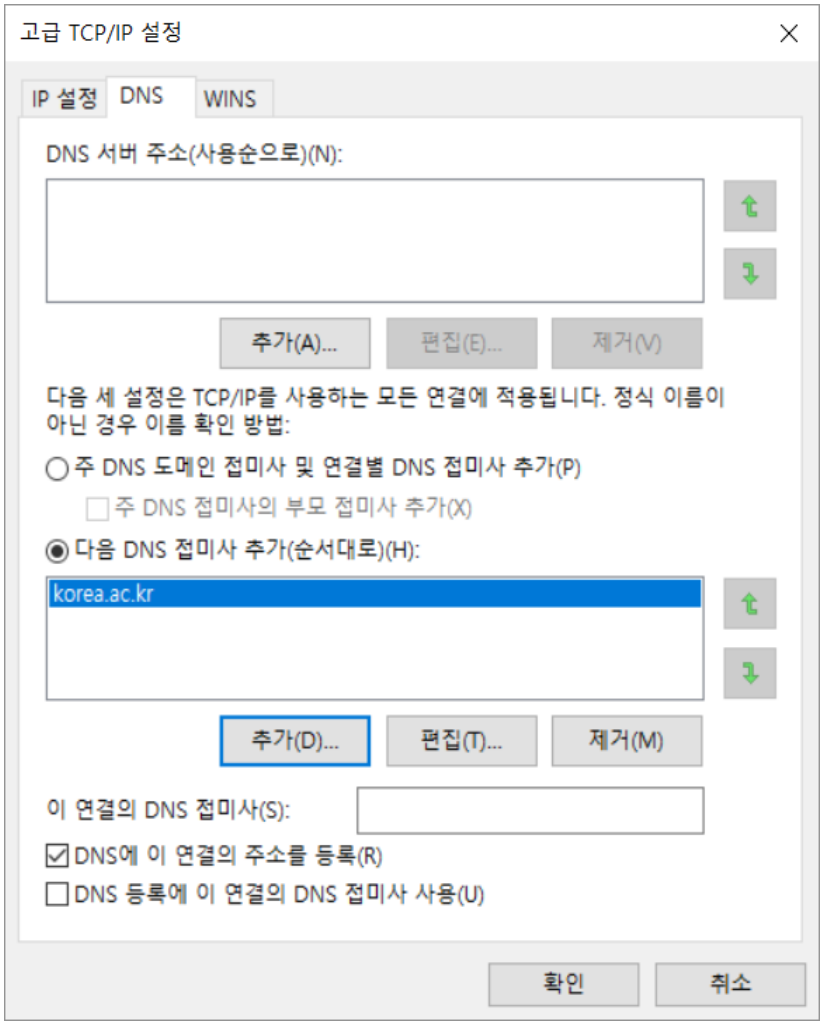
FQDN

<hyogon>53% more /etc/resolv.conf

domain ajou.ac.kr

nameserver 202.30.0.11

nameserver 168.126.63.1



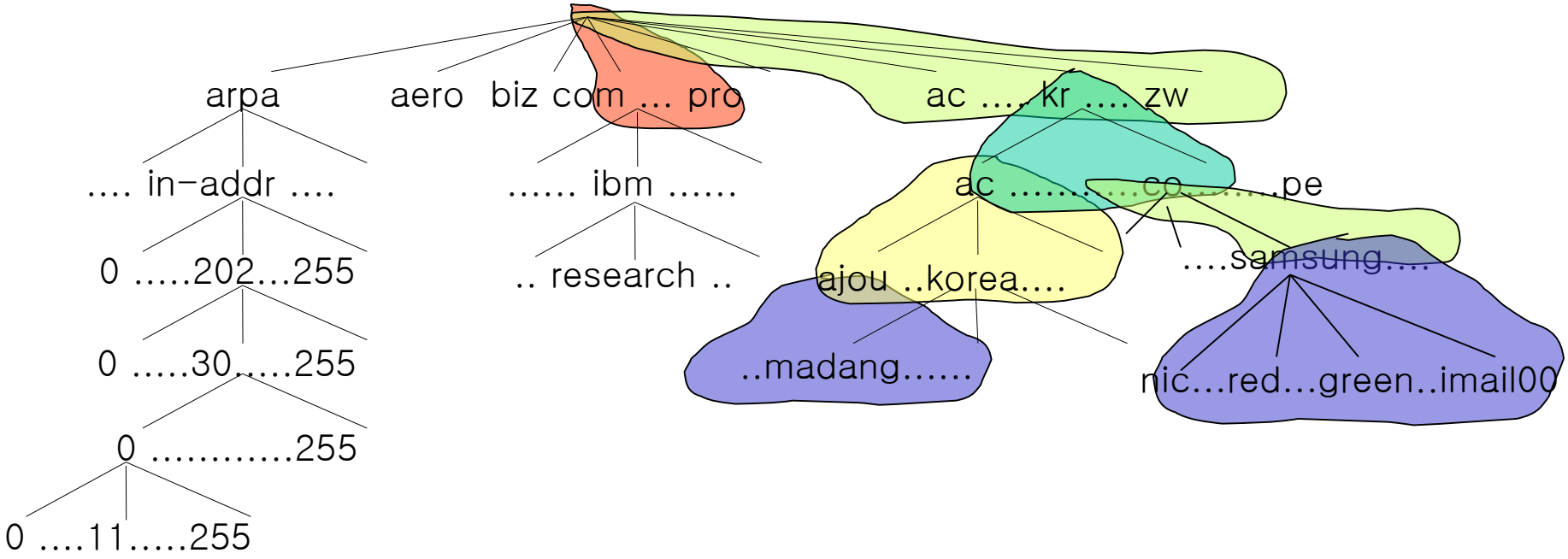
What's in a domain name?

- 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

Secondary authoritative servers

```
C:₩Users₩사용자>nslookup
기본 서버:  kns.kornet.net
Address:  168.126.63.1

> set querytype=ns
> korea.ac.kr
서버:    kns.kornet.net
Address:  168.126.63.1

권한 없는 응답:
korea.ac.kr     nameserver = kucc.korea.ac.kr
korea.ac.kr     nameserver = kuccgx.korea.ac.kr

kucc.korea.ac.kr        internet address = 163.152.11.6
kuccgx.korea.ac.kr      internet address = 163.152.1.1
```

```
> mit.edu
서버:    kns.kornet.net
Address:  168.126.63.1

권한 없는 응답:
mit.edu nameserver = asia2.akam.net
mit.edu nameserver = usw2.akam.net
mit.edu nameserver = asia1.akam.net
mit.edu nameserver = use5.akam.net
mit.edu nameserver = eur5.akam.net
mit.edu nameserver = use2.akam.net
mit.edu nameserver = ns1-37.akam.net
mit.edu nameserver = ns1-173.akam.net

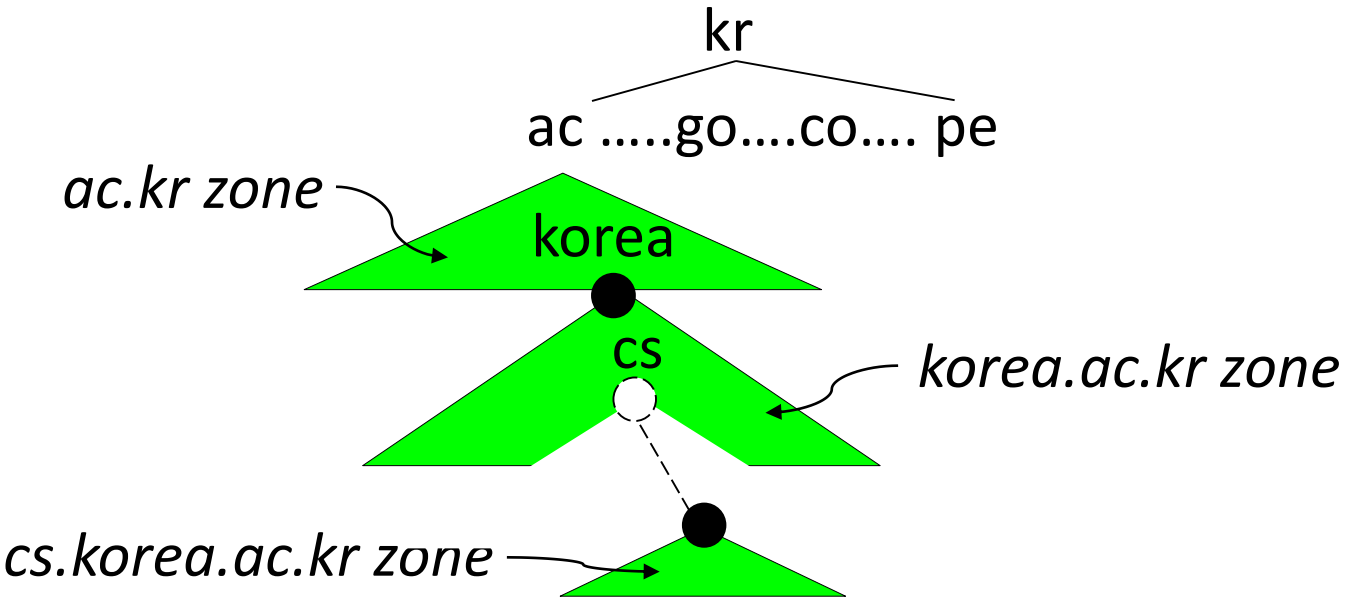
eur5.akam.net   internet address = 23.74.25.64
use2.akam.net   internet address = 96.7.49.64
use5.akam.net   internet address = 2.16.40.64
usw2.akam.net   internet address = 184.26.161.64
asia1.akam.net  internet address = 95.100.175.64
asia2.akam.net  internet address = 95.101.36.64
ns1-37.akam.net internet address = 193.108.91.37
ns1-173.akam.net        internet address = 193.108.91.173
```





Delegation

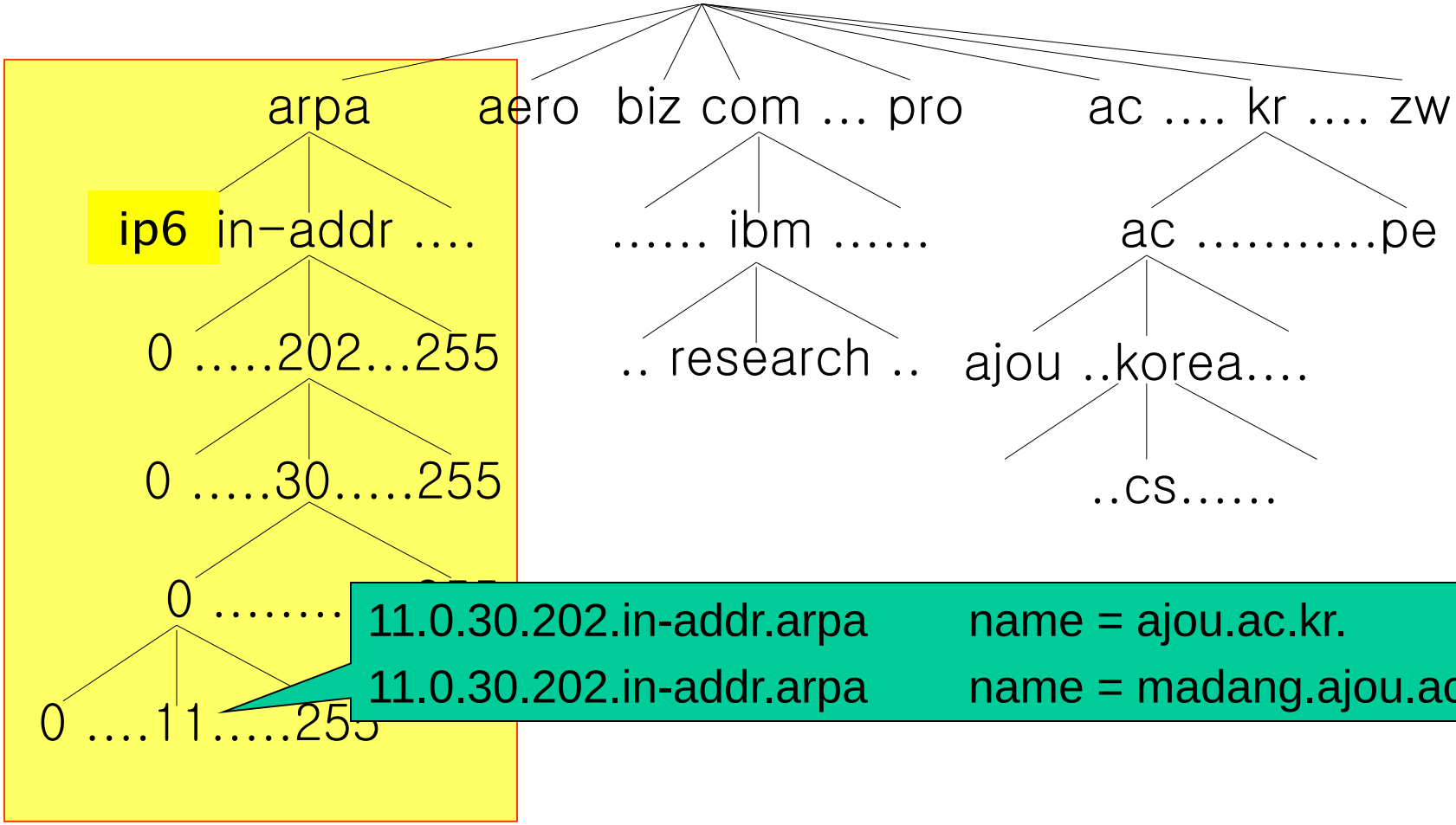
- A zone can be split into multiple sub-zones
- Zones can be separately managed



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
  - More than 1,200 now (e.g. “삼성”)
- ccTLD (country-code Top Level Domains)
  - .kp : North Korea – uses its “own (juche?)” DNS
  - .tv : Tuvalu
  - .ch: Swiss (Confœderatio Helvetica)

Domain Name Space



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

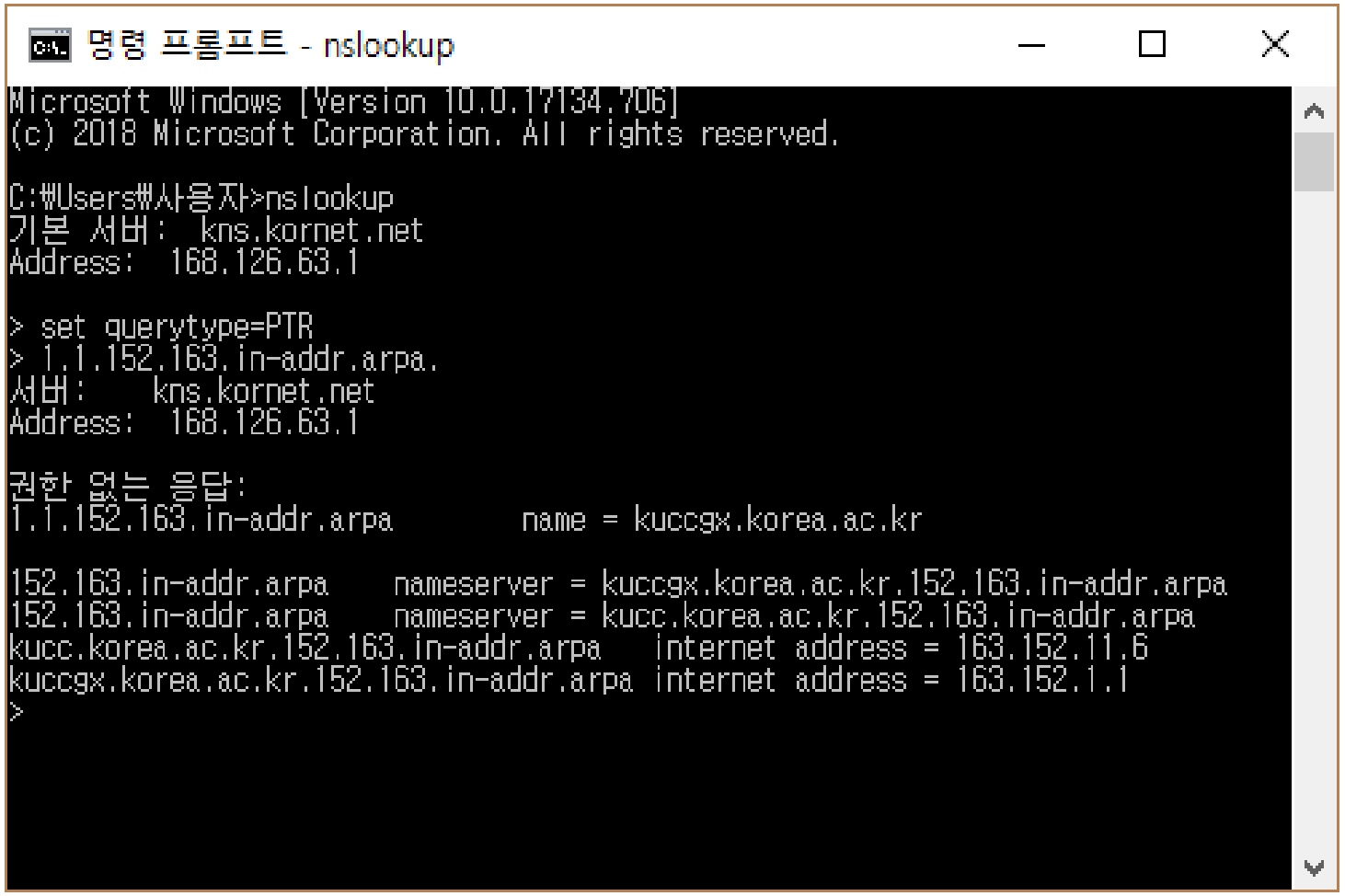
ARPA domain

- 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
  - 163.in-addr.arpa → APNIC server
  - 152.163.in-addr.arpa → KU server

ARPA zones

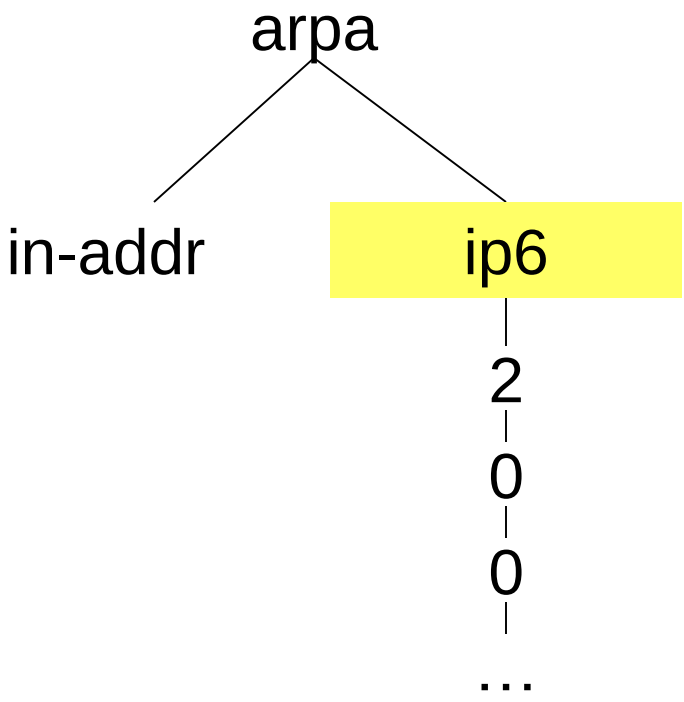
- 11.0.30.202.in-addr.arpa and widen.korea.ac.kr are in totally different zones!
- Registration of ARPA domain names is done separately
  - A record registration does not mean PTR record registration will be done automatically
  - Frequently done for free - in the future?

PTR query


명령 프롬프트 - nslookup
Microsoft Windows [Version 10.0.17134.706]
(c) 2018 Microsoft Corporation. All rights reserved.
C:₩Users₩사용자>nslookup
기본 서버: kns.kornet.net
Address: 168.126.63.1
> set querytype=PTR
> 1.1.152.163.in-addr.arpa.
서버: kns.kornet.net
Address: 168.126.63.1
권한 없는 응답:
1.1.152.163.in-addr.arpa name = kuccgx.korea.ac.kr
152.163.in-addr.arpa nameserver = kuccgx.korea.ac.kr.152.163.in-addr.arpa
152.163.in-addr.arpa nameserver = kucc.korea.ac.kr.152.163.in-addr.arpa
kucc.korea.ac.kr.152.163.in-addr.arpa internet address = 163.152.11.6
kuccgx.korea.ac.kr.152.163.in-addr.arpa internet address = 163.152.1.1
>

ARPA domain

- Can be used for IPv6
- And others such as ENUM : E.164 identifier to * mapping [RFC 2916]
  - Email address
  - URL
  - SIP address
  - Fax number
  - Telephone number
  - Etc.



General TLD (gTLD) before 2012

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 (~2000)
  - Com, net, org, gov, int, mil, edu
- 7 new gTLDs ratified by ICANN (~2003)
  - General use: biz, info
  - Personal use: name
  - Profit restricted domain: pro
  - Non-profit restricted domains: museum, aero, coop
- 2012, floodgate is open → now > 1000

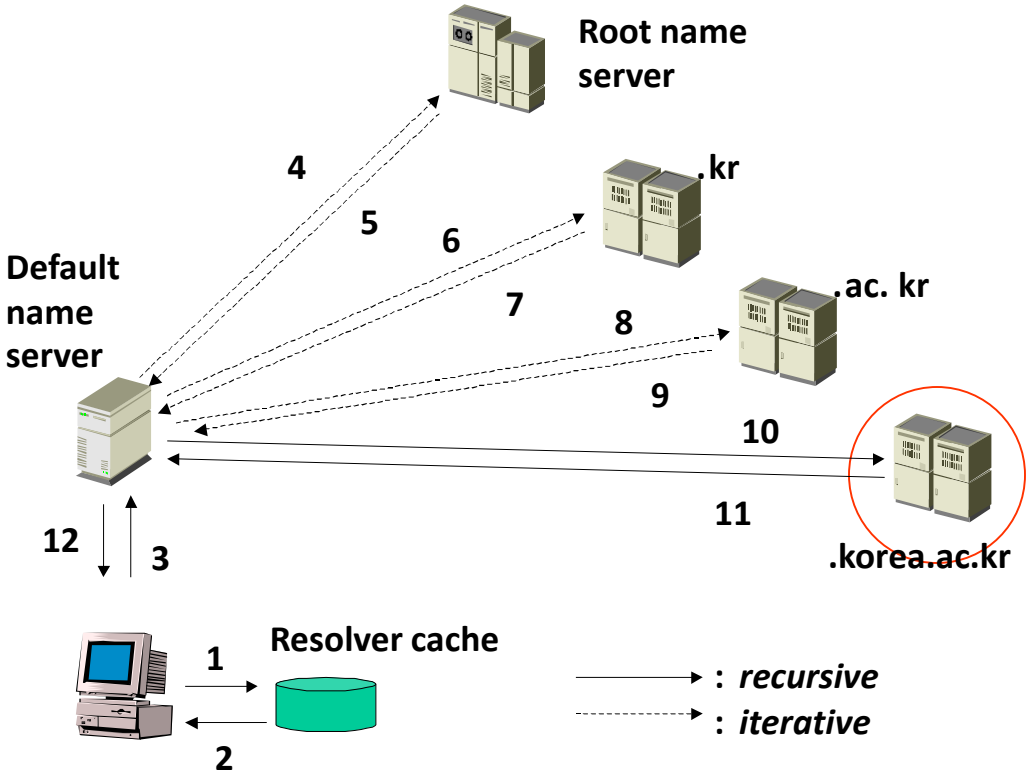
ccTLD

- 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	

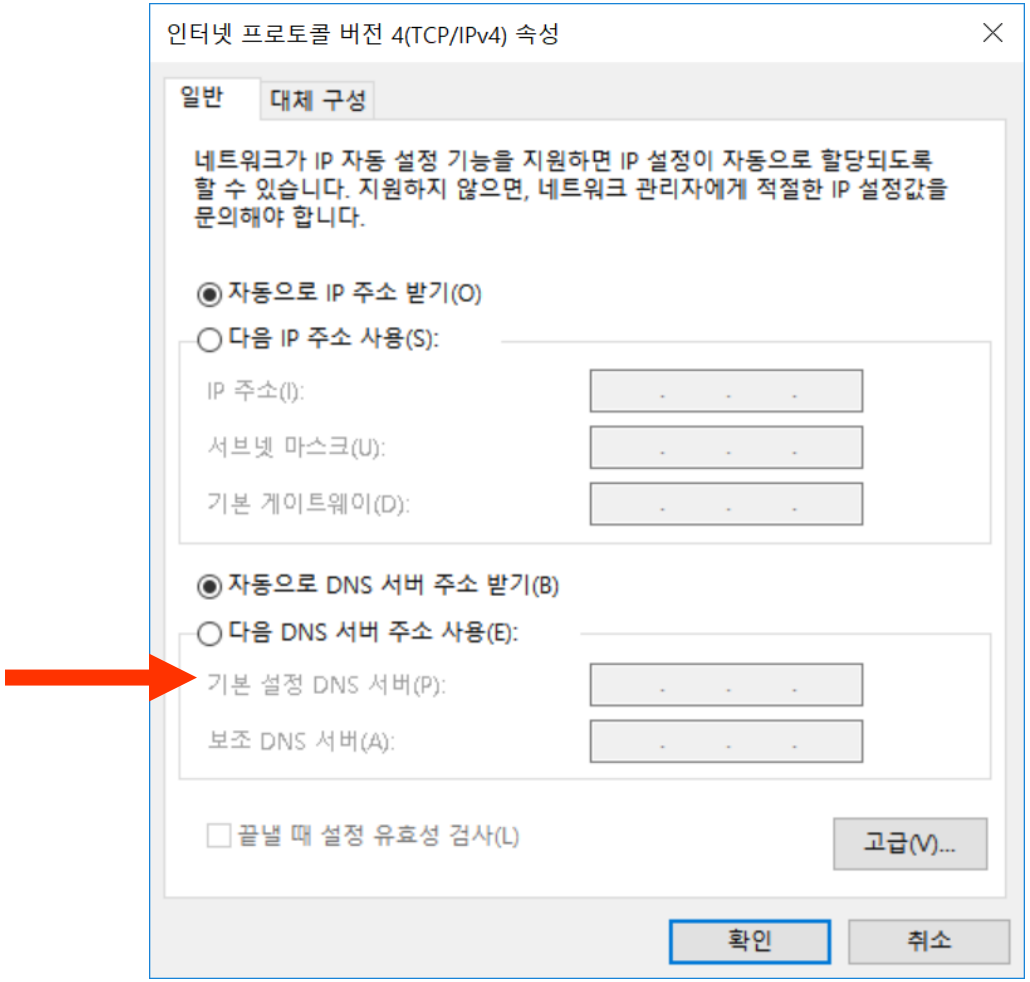
What happens when we click on a hyperlink?

http://widen.korea.ac.kr/talks/2019.html



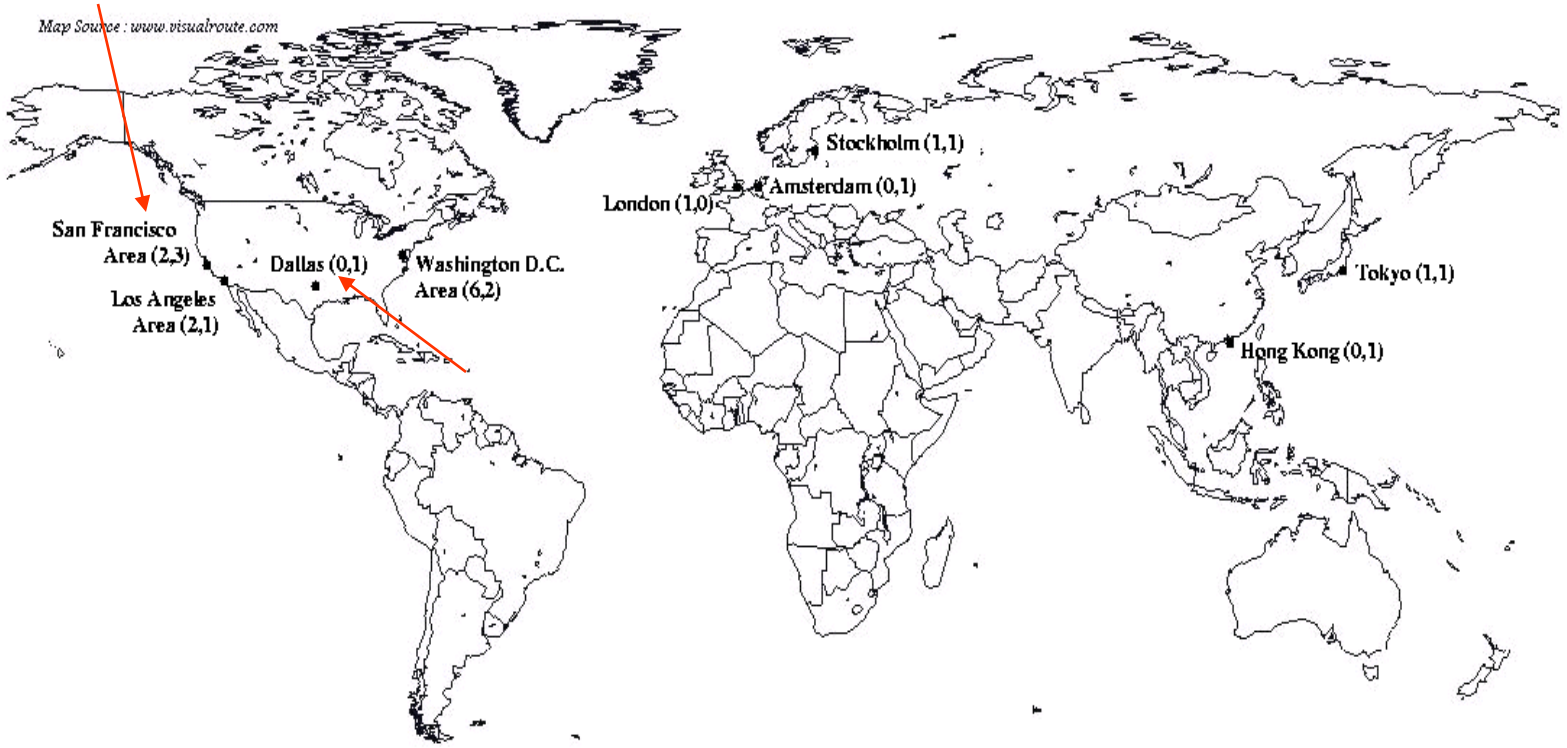
“Local” name server

- Is the *default* name server



Servers in the higher hierarchy

- 13 root servers and 13 gTLD servers

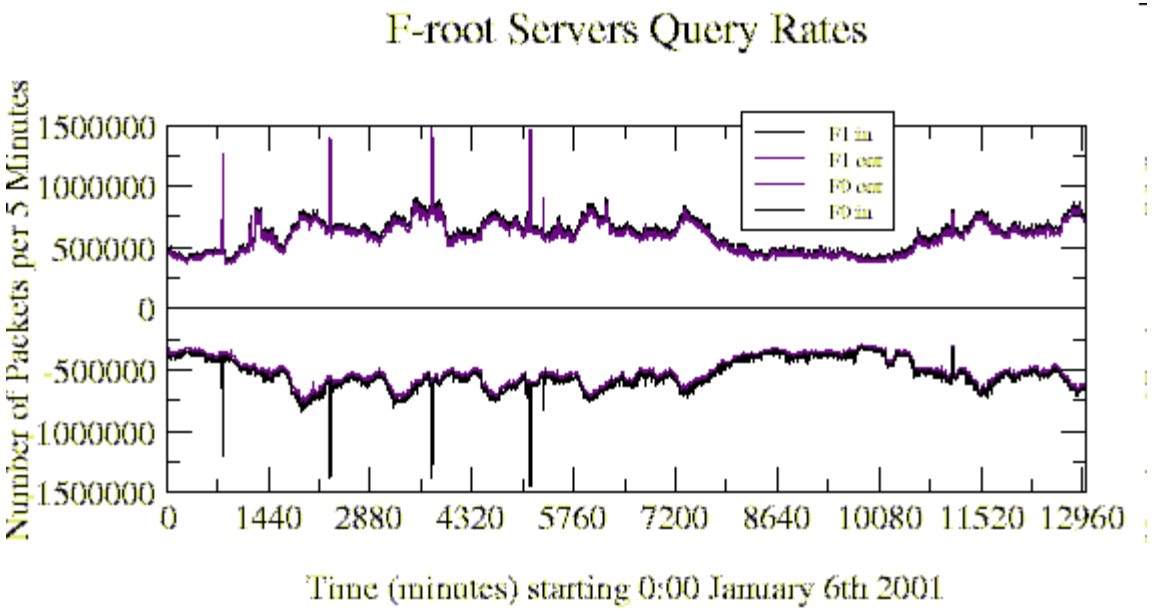


Root name servers

- 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	
...				

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

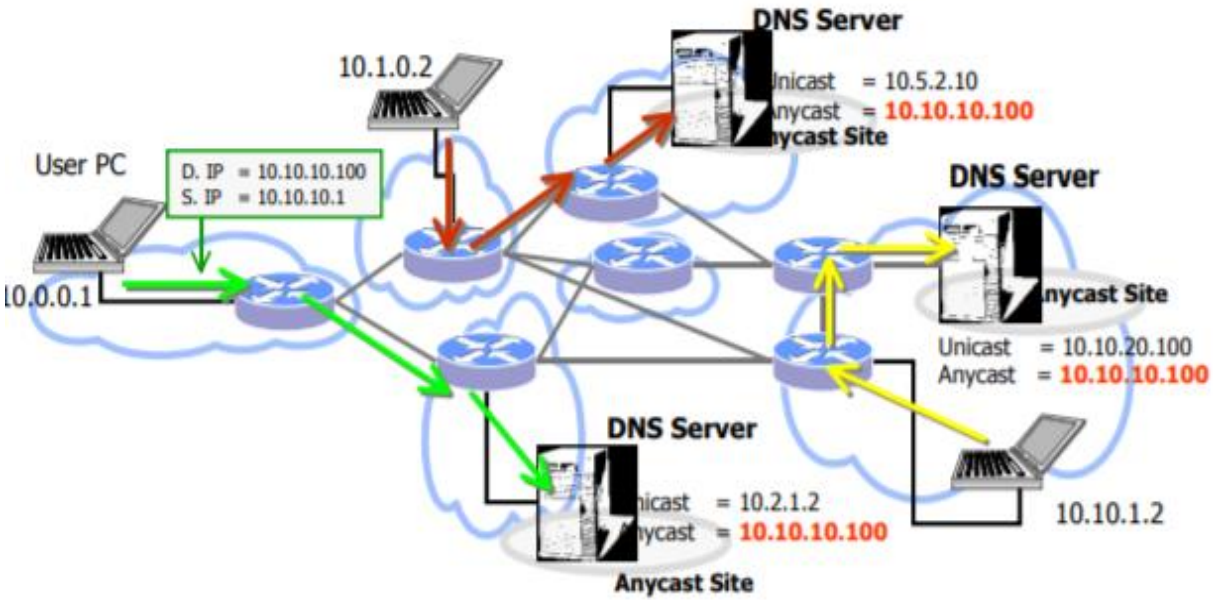
- 13: [A-M].gtld-servers.net
- Under the administration of VeriSign
- As of June 2002
  - .com 77.70%
  - .net 13.50%
  - .org 8.79%

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

Replication

- Numerous replicated root/gTLD servers
  - Korea: replicates F/J/M root servers
- Big default servers (e.g. 168.126.63.1)
- Anycast!



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

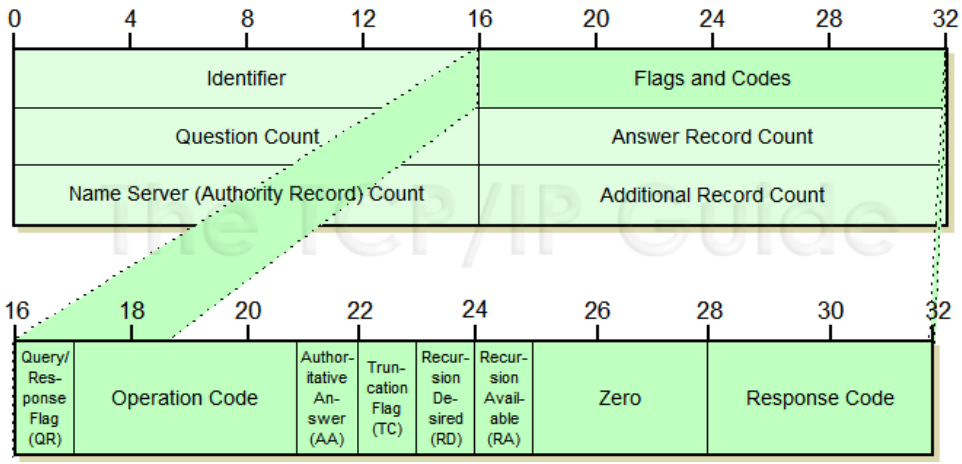
DNS packet format

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

Truncated!

	Source	Destination	Protoc	Length	Info
1	192.168.0.11	163.152.1.1	DNS	79	Standard query 0x0006 TXT cnn.com.korea.ac.kr
2	163.152.1.1	192.168.0.11	DNS	127	Standard query response 0x0006 No such name TXT cnn.com.korea.ac.kr SOA kuccgx.korea.ac.kr
3	192.168.0.11	163.152.1.1	DNS	67	Standard query 0x0007 TXT cnn.com
4	163.152.1.1	192.168.0.11	DNS	480	Standard query response 0x0007 TXT cnn.com TXT TXT TXT TXT TXT TXT TXT TXT TXT TXT
10	192.168.0.11	163.152.1.1	DNS	79	Standard query 0x0007 TXT cnn.com
13	163.152.1.1	192.168.0.11	DNS	793	Standard query response 0x0007 TXT cnn.com TXT TXT TXT TXT TXT TXT TXT TXT TXT TXT TXT TX…

```
> Internet Protocol Version 4, Src: 163.152.1.1, Dst: 192.168.0.11
> User Datagram Protocol, Src Port: 53, Dst Port: 58602
v Domain Name System (response)
    [Request In: 3]
    [Time: 0.020690000 seconds]
    Transaction ID: 0x0007
  v Flags: 0x8380 Standard query response, No error
      1... .... .... .... = Response: Message is a response
      .000 0... .... .... = Opcode: Standard query (0)
      .... .0.. .... .... = Authoritative: Server is not an authority for domain
      .... ..1. .... .... = Truncated: Message is truncated
      .... ...1 .... .... = Recursion desired: Do query recursively
      .... .... 1... .... = Recursion available: Server can do recursive queries
      .... .... .0.. .... = Z: reserved (0)
      .... .... ..0. .... = Answer authenticated: Answer/authority portion was not authenticated by the server
      .... .... ...0 .... = Non-authenticated data: Unacceptable
      .... .... .... 0000 = Reply code: No error (0)
    Questions: 1
    Answer RRs: 10
    Authority RRs: 0
    Additional RRs: 0
  > Queries
  v Answers
```

OP codes

- 0: standard query
- 1: no longer used
- 2: server state report request
- 3: not used
- 4: notify
  - Primary authoritative name server notifies secondaries of changed zone file
- 5: update → see dynamic DNS

DNS packet format

- Representation of domain names in the packet



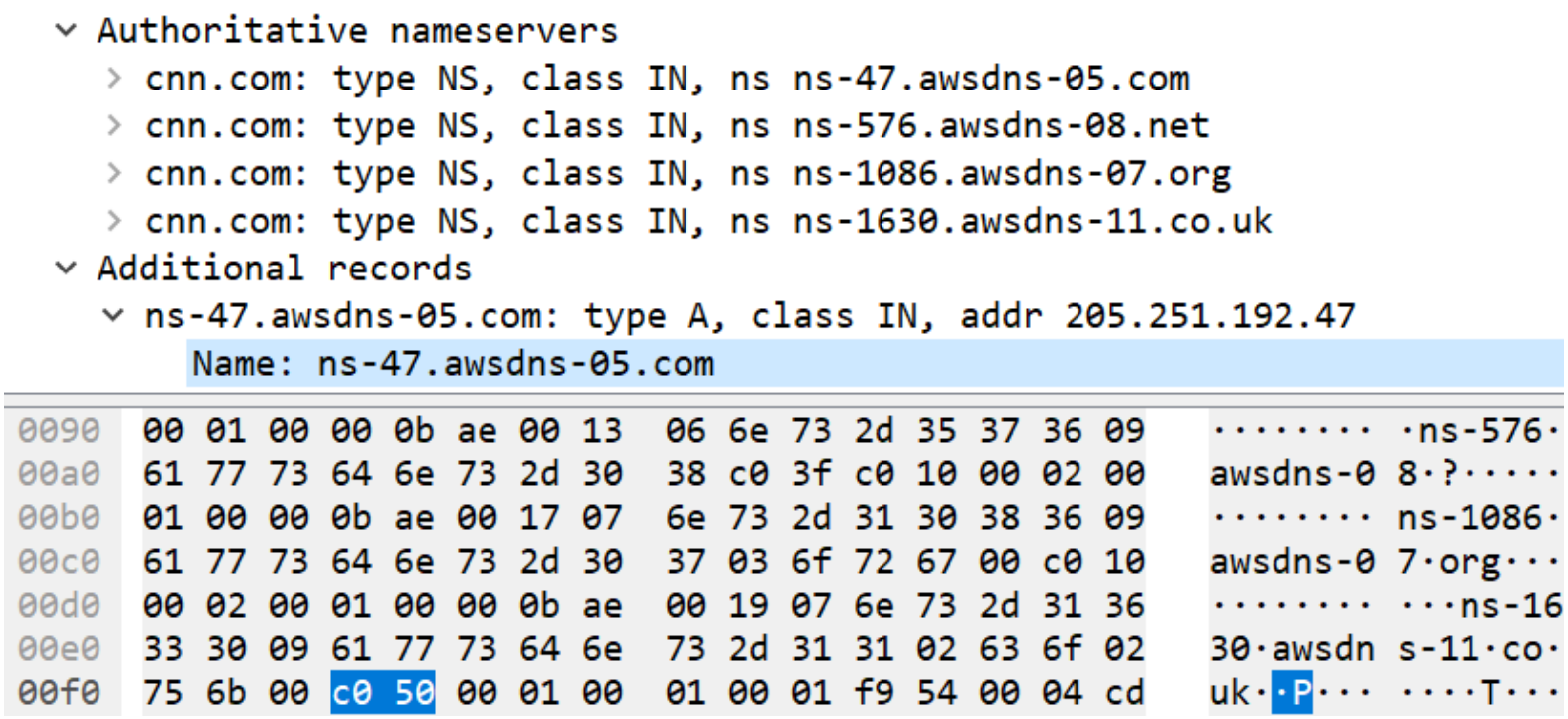
- Only 1 “long” representation
  - Repetitions are coded as 2B pointers

FQDN – long format

```
Domain Name System (query)
    Transaction ID: 0x0006
  > Flags: 0x0100 Standard query
    Questions: 1
    Answer RRs: 0
    Authority RRs: 0
    Additional RRs: 0
  Queries
    www.cnn.com: type A, class IN
        Name: www.cnn.com
        [Name Length: 11]
        [Label Count: 3]
        Type: A (Host Address) (1)
        Class: IN (0x0001)
```

```
0000  00 07 89 82 0d aa 72 f0  ec 4b d8 70 08 00 45 00   ·····r· ·K·p··E·
0010  00 39 6a 4f 00 00 80 11  3b ba ac 1e 01 0d a8 7e   ·9jO···· ;·····~
0020  3f 01 c4 00 00 35 00 25  44 2c 00 06 01 00 00 01   ?····5·% D,·····
0030  00 00 00 00 00 00 03 77  77 77 03 63 6e 6e 03 63   ·····w ww·cnn·c
0040  6f 6d 00 00 01 00 01                              om·····
```

FQDN – pointer



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 :
  - Back-off after each cyle

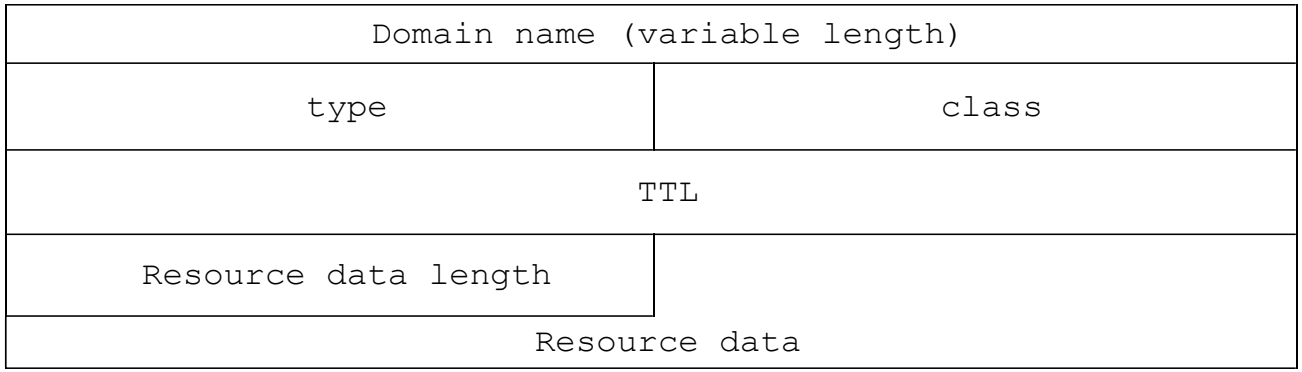
$$T_{base} = \max(4, 2 \times E[R])$$

Questions (in queries/replies)

Query name (variable length)	
Query type	Query class

```
> Internet Protocol Version 4, Src: 172.30.1.4, Dst: 168.126.63.1
> User Datagram Protocol, Src Port: 56743, Dst Port: 53
v Domain Name System (query)
    Transaction ID: 0xffa5
  > Flags: 0x0100 Standard query
    Questions: 1
    Answer RRs: 0
    Authority RRs: 0
    Additional RRs: 0
  v Queries
    v chat-pa.clients6.google.com: type A, class IN
        Name: chat-pa.clients6.google.com
        [Name Length: 27]
        [Label Count: 4]
        Type: A (Host Address) (1)
        Class: IN (0x0001)
```

Answers (in replies)



Answers (in replies)

```
▾ Queries
  ▸ cdn.onenote.net: type A, class IN
▾ Answers
  ▾ cdn.onenote.net: type CNAME, class IN, cname cdn.onenote.net.edgekey.net
      Name: cdn.onenote.net
      Type: CNAME (Canonical NAME for an alias) (5)
      Class: IN (0x0001)
      Time to live: 2993
      Data length: 26
      CNAME: cdn.onenote.net.edgekey.net
  ▾ cdn.onenote.net.edgekey.net: type CNAME, class IN, cname e1553.dspg.akamaiedge.net
      Name: cdn.onenote.net.edgekey.net
      Type: CNAME (Canonical NAME for an alias) (5)
      Class: IN (0x0001)
      Time to live: 580
      Data length: 24
      CNAME: e1553.dspg.akamaiedge.net
  ▾ e1553.dspg.akamaiedge.net: type A, class IN, addr 104.75.33.244
      Name: e1553.dspg.akamaiedge.net
      Type: A (Host Address) (1)
      Class: IN (0x0001)
      Time to live: 3
      Data length: 4
      Address: 104.75.33.244
```

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

타입	의미	번호
A	IPv4 주소	1
AAAA	IPv6 주소	28
NS	DNS 서버	2
MX	메일 서버	15
CNAME	본명 (canonical name)	5
PTR	IP 주소에 대응되는 도메인 이름	12
SOA	도메인 관리 정보	6
SRV	NS, MX 이외의 서버 정보	33
TXT	임의의 텍스트 정보	16

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



  - “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

```
> _ldap._tcp.openldap.org.
서버:    kuccgx.korea.ac.kr
Address:  163.152.1.1

권한 없는 응답:
_ldap._tcp.openldap.org SRV service location:
          priority       = 0
          weight         = 0
          port           = 389
          svr hostname   = www.openldap.org
```

RR types

- SOA = Start Of Authority

```
> set querytype=SOA
> cnn.com
서버:    kns.kornet.net
Address:  168.126.63.1

권한 없는 응답:
cnn.com
        primary name server = ns-47.awsdns-05.com
        responsible mail addr = awsdns-hostmaster.amazon.com
        serial  = 1
        refresh = 7200 (2 hours)
        retry   = 900 (15 mins)
        expire  = 1209600 (14 days)
        default TTL = 86400 (1 day)
```

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



Negative caching

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

Resolver cache

- Ipconfig/displaydns, flushdns

```
C:₩Users₩USER>ipconfig/displaydns

Windows IP 구성

    www.korea.ac.kr
    ----------------------------------------
    데이터 이름 . . . . . : www.korea.ac.kr
    데이터 유형 . . . . . : 1
    TTL(Time To Live) . . : 2254
    데이터 길이 . . . . . : 4
    섹션 . . . . . : 응답
    (호스트) 레코드 . . . : 163.152.6.10

    데이터 이름 . . . . . : kucc.korea.ac.kr
    데이터 유형 . . . . . : 1
    TTL(Time To Live) . . : 2254
    데이터 길이 . . . . . : 4
    섹션 . . . . . : 추가
    (호스트) 레코드 . . . : 163.152.11.6

    데이터 이름 . . . . . : kuccgx.korea.ac.kr
    데이터 유형 . . . . . : 1
    TTL(Time To Live) . . : 2254
    데이터 길이 . . . . . : 4
    섹션 . . . . . : 추가
    (호스트) 레코드 . . . : 163.152.1.1
```

Round-robin DNS

```
> set querytype=A
> cnn.com
서버:    kuccgx.korea.ac.kr
Address:  163.152.1.1

권한 없는 응답:
이름:    cnn.com
Addresses:  151.101.65.67
          151.101.129.67
          151.101.1.67
          151.101.193.67

> cnn.com
서버:    kuccgx.korea.ac.kr
Address:  163.152.1.1

권한 없는 응답:
이름:    cnn.com
Addresses:  151.101.1.67
          151.101.129.67
          151.101.65.67
          151.101.193.67
```

Dynamic DNS

- DHCP and DNS working together
- Allows remote and dynamic update
  - Used to be manual

