

Challenges and Requirements in Preserving DNS Digital Data

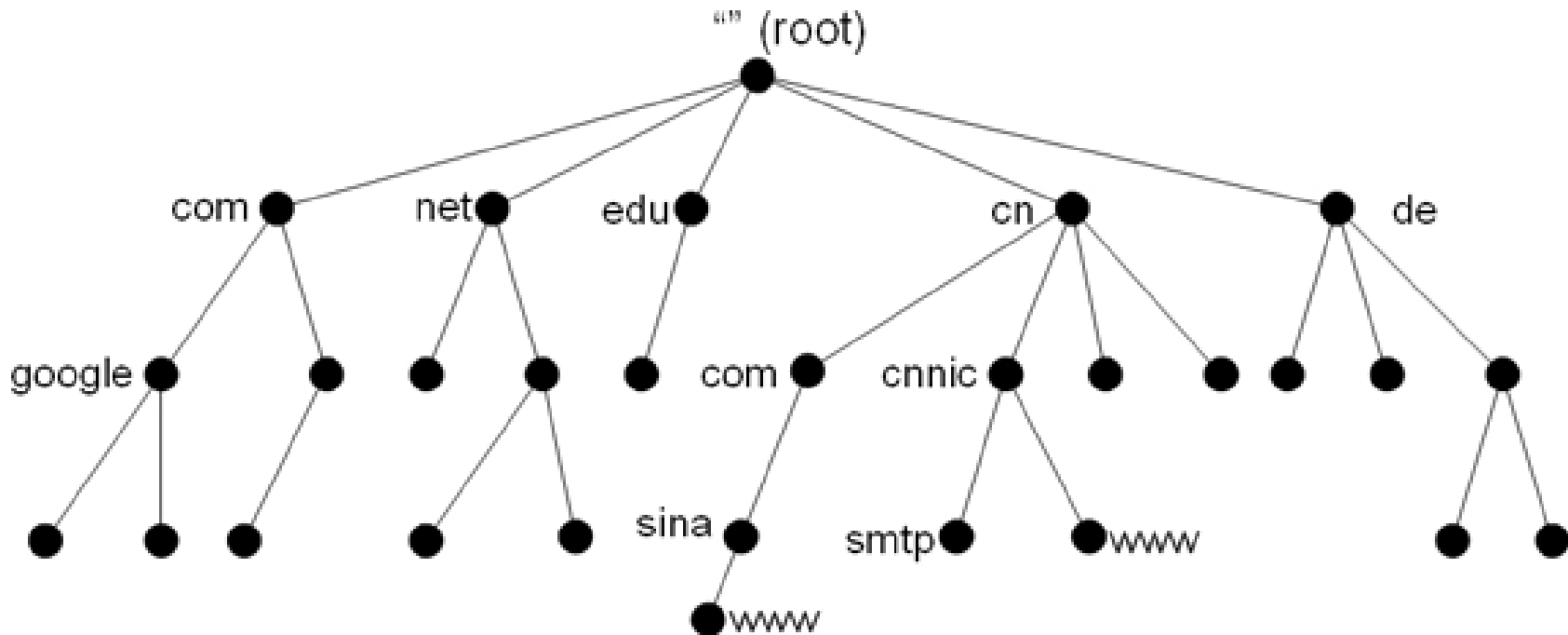
CNNIC
2007-10

Summary

- DNS Theory
- DNS Log
- DNS Information Analysis
- Challenges on Preserving DNS Digital Data

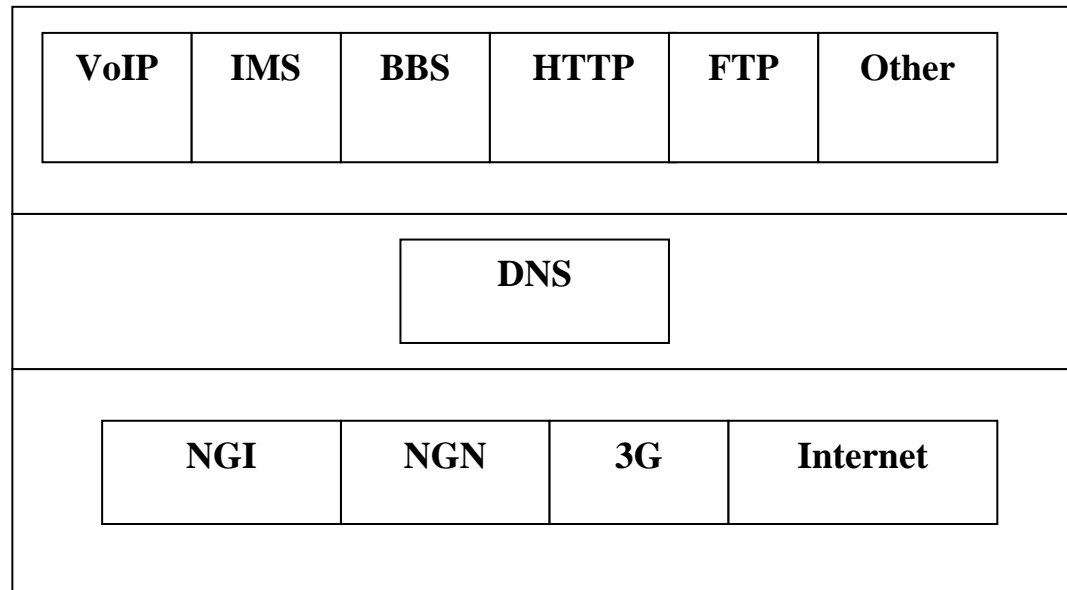
DNS

- Name -> IP Address
- Global Distributed
- Multiple Levels

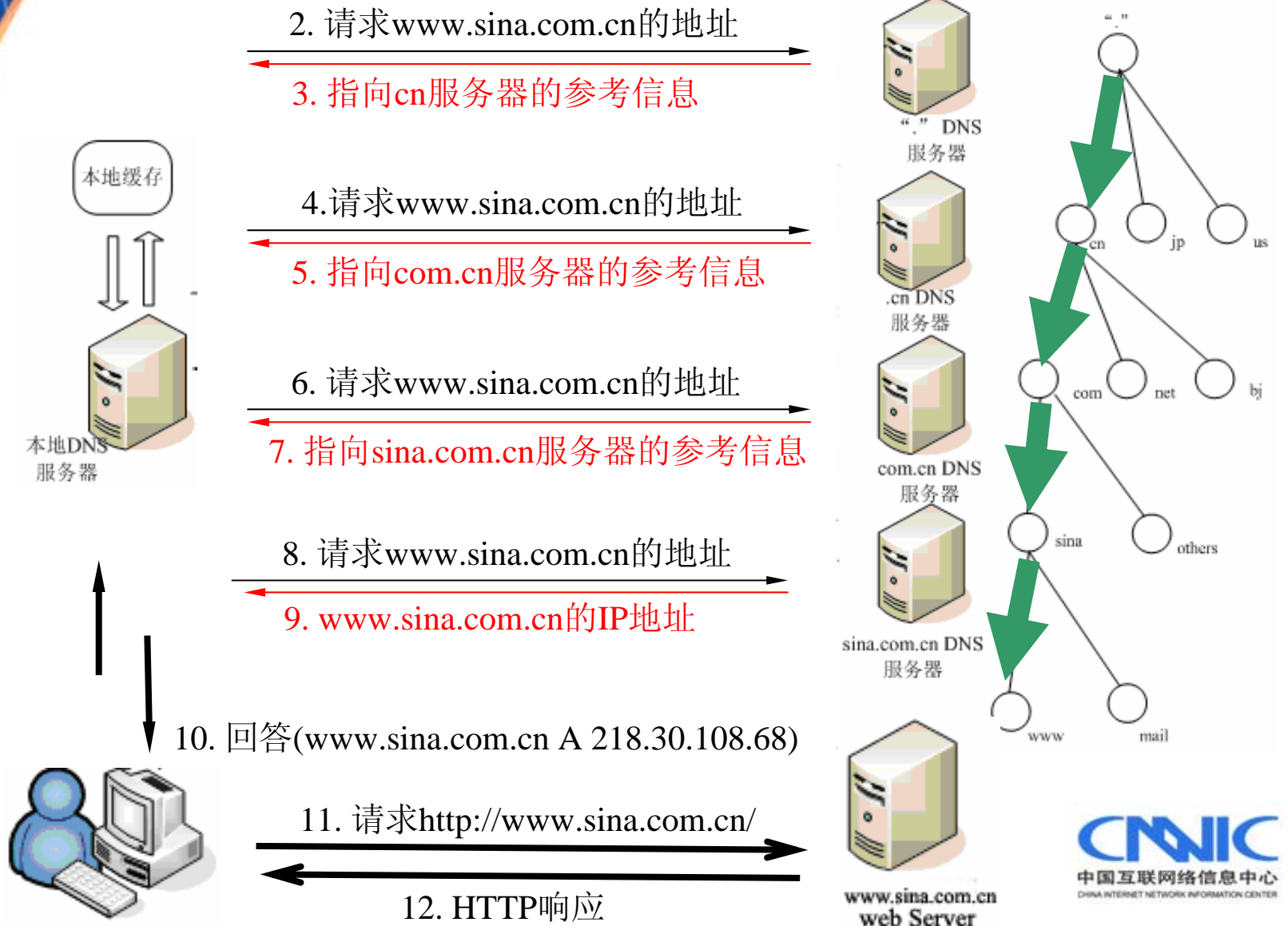


DNS

- Vital for all network applications



DNS Resolution Mechanism



DNS LOG

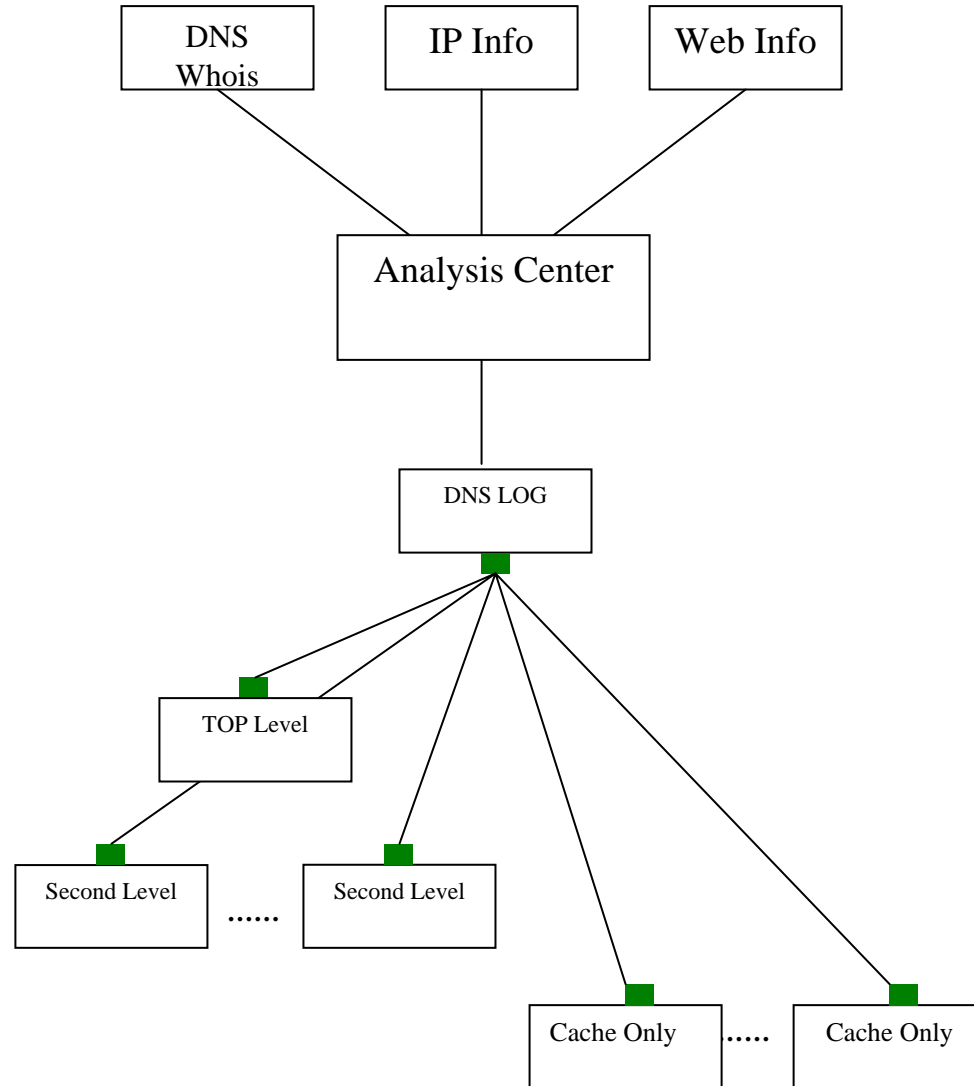
■ Authority DNS Server

- 08-Oct-2007 16:00:01.856 queries: info: client 202.188.0.132#59583: query: cpst.com.cn IN MX –
- 08-Oct-2007 16:00:01.866 queries: info: client 202.106.0.104#32771: query: MXJ.dxstats.gov.cn IN A –
- 08-Oct-2007 16:00:01.866 queries: info: client 202.96.69.41#33023: query: www.x150.cn IN A -

■ Cache-Only DNS Server

- Oct 11 14:58:11.999 general: debug 60: socket 0x68bd60 193.0.14.129#53: packet received correctly
- Oct 11 14:58:12.005 general: debug 60: socket 0x68bd60 203.119.26.1#53: packet received correctly
- Oct 11 14:58:12.010 general: debug 60: socket 0x68bd60 202.106.184.166#53: packet received correctly
- Oct 11 14:58:12.036 general: debug 60: socket 0x68bd60 61.172.201.254#53: packet received correctly

DNS Information Analysis

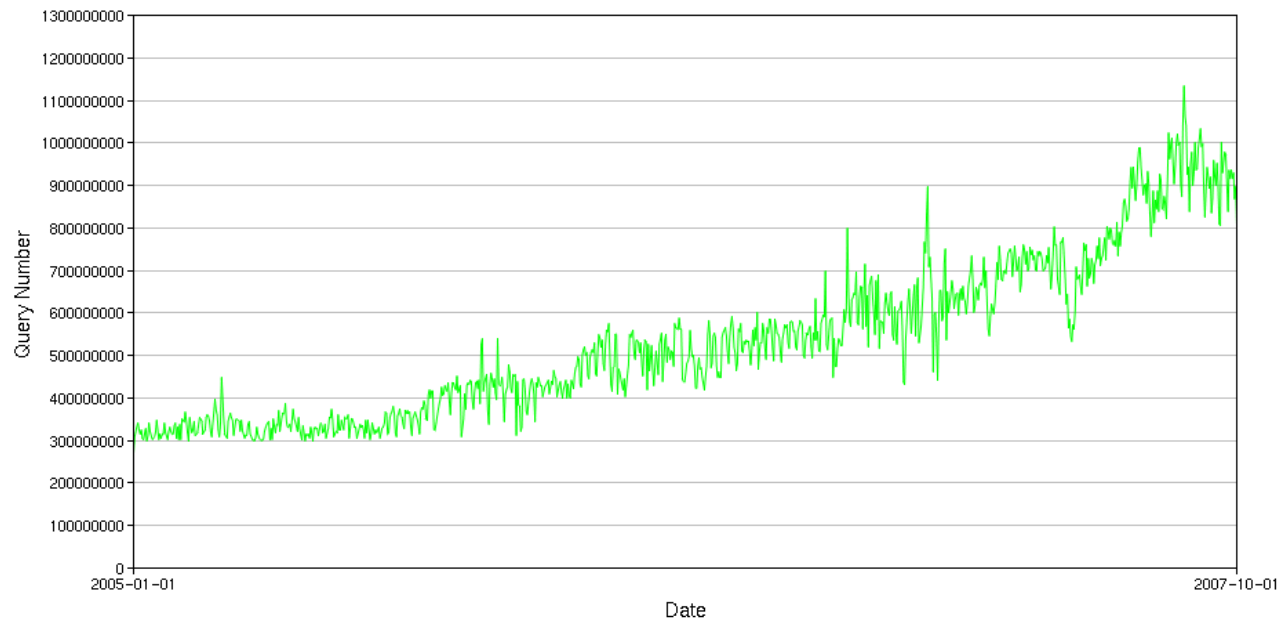


DNS Information Analysis

- Mass Data Analysis
 - When
 - Where
 - Who
 - What
- Mass Data Preservation
 - DNS LOG
 - Whois Info
 - IP Info
 - Website Info

Challenges 1

- Increasing Mass Data



Total Query Number is 535709758492

- 120G plain text
- 900 million query record

Challenge 2

- IP Address Info
 - IP -> geographical position
 - IP -> Internet User
 - IP allocation and withdraw

Challenge3

- Website livability

	1998	1999	2000	2001	2002
1998	100%	52%	28%	15%	8%
1999		100%	50%	27%	14%
2000			100%	47%	26%
2001				100%	46%
2002					100%

- Website Content

- Webpage's life-span:44 days – 2years
- webpage's content changes

Challenge 4

- Preserving Media
 - Tape
 - Disk
 - Array
- Preserving format
 - Plain text
 - XML
 - Database



- Q&A