CNAME in the wild

Kazunori Fujiwara, JPRS
fujiwara@jprs.co.jp
Akira Sato, University of Tsukuba
akira@cc.tsukuba.ac.jp

CNAME Response Wire Format (RFC 1035)

- CNAME Resource Record
  - (Owner)NAME: <domain-name>
  - TYPE: 16bit (5)
  - CLASS: 16bit (1)
  - TTL: 32bit
  - RDLENGTH: 16bit
  - RDATA: <domain-name>
- <domain-name>
  - <domain-name> is a domain name represented as a series of labels, and terminated by a label with zero length
  - <character-string> is a single length octet followed by that number of characters
- Message compression
  - Message compression targets are (owner)NAME and NS/MX/CNAME/SOA RDATA

Example of long (7) CNAME chains

- dig @8.8.8.8 www.brother.in A

;; flags: qr rd ra; QUERY: 1, ANSWER: 8, AUTHORITY: 0, ADDITIONAL: 1

;; QUESTION SECTION:

;www.brother.in. IN A;; ANSWER SECTION:

www.brother.in. 600 IN CNAME mc-12265895-control-tm.trafficmanager.net.

mc-12265895-control-tm.trafficmanager.net. 30 IN CNAME mc-12265895-6e1d-4319-8534-7755-cdn-endpoint.azureedge.net.

mc-12265895-6e1d-4319-8534-7755-cdn-endpoint.azureedge.net. 1800 IN CNAME mc-12265895-6e1d-4319-8534-7755-cdn-endpoint.afd.azureedge.net.

mc-12265895-6e1d-4319-8534-7755-cdn-endpoint.afd.azureedge.net. 60 IN CNAME reserved-g01.afd.azureedge.net.

reserved-g01.afd.azureedge.net. 60 IN CNAME star-t-g.trafficmanager.net.

star-t-g.trafficmanager.net. 60 IN CNAME shed.dual-low.s-part-0018.t-0009.t-msedge.net.

shed.dual-low.s-part-0018.t-0009.t-msedge.net. 40 IN CNAME s-part-0018.t-0009.t-msedge.net.

s-part-0018.t-0009.t-msedge.net. 40 IN A 13.107.246.46

Tcpdump output of www.brother.in A response

```
000:  4500 018a 4ec7 0000 7211 26e2 0808 0808  E...N...r.&.....
010:  c0a8 0102 0035 6764 0176 a7ca 0545 8180  .....5gd.v...E..
020:  0001 0008 0000 0001 0377 7777 0762 726f  .........www.bro
030:  7468 6572 0269 6e00 0001 0001 c00c 0005  ther.in.........
040:  0001 0000 0258 002b 166d 632d 3132 3236  .....X.+.mc-1226
050:  3538 3935 2d63 6f6e 7472 6f6c 2d74 6d0e  5895-control-tm.
060:  7472 6166 6669 636d 616e 6167 6572 036e  trafficmanager.n
070:  6574 00c0 2c00 0500 0100 0000 1e00 392c  et..,.........9,
080:  6d63 2d31 3232 3635 3839 352d 3665 3164  mc-12265895-6e1d
090:  2d34 3331 392d 3835 3334 2d37 3735 352d  -4319-8534-7755-
0a0:  6364 6e2d 656e 6470 6f69 6e74 0961 7a75  cdn-endpoint.azu
0b0:  7265 6564 6765 c052 c063 0005 0001 0000  reedge.R.c......
0c0:  0708 0033 2c6d 632d 3132 3236 3538 3935  ...3,mc-12265895
0d0:  2d36 6531 642d 3433 3139 2d38 3533 342d  -6e1d-4319-8534-
0e0:  3737 3535 2d63 646e 2d65 6e64 706f 696e  7755-cdn-endpoin
0f0:  7403 6166 64c0 90c0 a800 0500 0100 0000  t.afd...........
100:  3c00 0f0c 7265 7365 7276 6564 2d67 3031  <...reserved-g01
110:  c0d5 c0e7 0005 0001 0000 003c 000b 0873  ...........<...s
120:  7461 722d 742d 67c0 43c1 0200 0500 0100  tar-t-g.C.......
130:  0000 3c00 2c04 7368 6564 0864 7561 6c2d  ..<.,.shed.dual-
140:  6c6f 770b 732d 7061 7274 2d30 3031 3806  low.s-part-0018.
150:  742d 3030 3039 0874 2d6d 7365 6467 65c0  t-0009.t-msedge.
160:  52c1 1900 0500 0100 0000 2800 02c1 27c1  R.........(...'.
170:  2700 0100 0100 0000 2800 040d 6bf6 2e00  '.......(...k...
180:  0029 0200 0000 0000 0000
```

IP Header　　underline shows DNS compression

UDP Header　　DNS Header

DNS Header　　QuerySection

QuerySection　　CNAME1(owner/compressed, type)

CNAME1(class 0001, TTL 32bit, rdlength 002b, RDATA)

CNAME1(RDATA)

CNAME1(RDATA)

CNAME1(RDATA)　CNAME2(owner, ..)

CNAME2 RDATA

CNAME2 RDATA

CNAME2 RDATA

CNAME2 RDATA　　CNAME3(owner,type,class,ttl, ..)

CNAME3 RDATA

CNAME3 RDATA

CNAME3 RDATA

CNAME3 RDATA　　CNAME4(owner,type,class,ttl,rdlen

CNAME4 RDATA

CNAME5(owner,tyle,class,ttl,rdlen)

CNAME5 RDATA　　CNAME6(owner,type,class,ttl,rdlen

CNAME6 RDATA

CNAME6 RDATA

CNAME6 RDATA

CNAME7(owner,type,class,ttl,rdlen)　CNAME7 RDATA

A(owner,type,class,ttl,rdlen)　A RDATA

It is too complicated, and parsing this response is hard for me

Who parses the CNAME chain?

- If the network application written in C, programmers usually use gethostbyname*() or getaddrinfo() to get the target IP adddresses
- Where are gethostbyname*() or getaddrinfo() ?
  - gethostbyname*() and getaddrinfo() are located in libc, are called from application programs, and operate with the privileges of the application programs.
  - That is, application programs need to parse responses containing complex CNAME chains.
- Programs that use their own asynchronous resolvers, such as browsers, need to parse the CNAME chains by the resolvers

Execution time of getaddrinfo() with/wo CNAME

- Debian machine
  - on Intel N100 (12th-gen E core)
  - kernel: 6.1.0-17-amd64
- Run unbound on localhost (127.0.0.1)
- first getaddrinfo(),then
- measure 10,000 gethostbyname2() and getaddrinfo() execution times
  - Within TTL time
- query names
  - www.google.com (No CNAME)
    - TTL 300
  - www.facebook.com (1 CNAME)
    - TTL 60
  - www.asahi.com (2 CNAME chains)
    - TTL 20
  - www.brother.in (7 CNAME chains)
    - TTL 60

	num of CNAMEs	Average exec time (microsec)	
		getaddrinfo	gethostbyname2
www.google.com	0	57.52	40.36
www.facebook.com	1	58.89	42.11
www.asahi.com	2	60.05	45.57
www.brother.in	7	63.27	47.33

As the number of CNAMEs increases, the execution time of gethostbyname2 and getaddrinfo increase.
7 CNAME chains add 5 or 7 microseconds processing time.

CNAME in the Wild: top domain list analysis

- To understand number of CNAME chains in the wild,
- I analyzed CNAME chains by top domain list
  - Generated popular hostnames from top domain lists
    - July 1,2024
    - Merged three top domain name lists (tranco, majestic, umbrella)
    - The name with "www." in front and the bare domain name
  - sent all host names "A" queries to local/public resolvers
  - analyzed responses
- Result
  - 2,413,235 hostnames had final A resource records

Number of CNAME chains: from top list

nCNAME	nQname	Ratio
9	2	0%
8	5	0%
7	35	0%
6	5,645	0%
5	6,417	0%
4	10,807	0%
3	25,450	1%
2	108,917	5%
1	456,922	19%
0	1,799,034	75%

- The longest CNAME chains seen at top domain lists is 9
- ＞ 3 CNAME chains
  - 22,912 (1%) hostnames have ＞3 CNAME chains
- 2 hostnames have 9 CNAME chains
  - ba.dn.nexoncdn.co.kr
  - hit2tw.dn.nexoncdn.co.kr
- 5 hostnames have 8 CNAME chains
  - dtstockphotos.azureedge.net
  - lightsail.aws.amazon.com
  - dc.services.visualstudio.com
  - clientconfig.passport.net
  - prod-files.acompli.net
- 35 hostnames have 7 CNAME chains
  - www.brother.in (9 CNAMEs in April)
  - …

Who offer long CNAME chains: from top domain list

ASN	Number of CNAME chains									
	0	1	2	3	4	5	6	7	8	9
8075	41,795	17,516	8,596	7,593	6,323	5,369	5,417	9	4	0
16509	164,582	76,774	16,140	3,226	1,253	365	28	3	1	0
8068	273	1,237	1,074	649	537	251	47	4	0	0
15133	752	1,247	1,151	545	607	46	16	0	1	2
20940	9,201	4,631	17,751	3,090	341	41	7	0	0	0

- This table shows the number of CNAME chains for each AS to which the A record corresponding to the query name
- AS8075 offers 17122 hostnames that use >3 CNAME chains
- AS16509 offers 1650 hostnames that use >3 CNAME chains

CNAME in the wild: real users’ data

- To understand real users' data
  - We analyzed resolvers' data at the University of Tsukuba
    - analyzed qtype A responses
    - 2024/6/1-14 data
    - The resolvers receive
      around 3000 queries/sec
      from around 10,000 query source IP addresses, daily
- We got
  - 2,481,681 hostnames have final A resource records
    - ignored all error/NODATA responses

Number of CNAME chains: from the university's data

Number of CNAME chains	Number of QNAMEs	ratio of number of QNAMEs	ratio of number of queries
10	1	0.00%	0.00%
9	8	0.00%	0.05%
8	53	0.00%	0.26%
7	482	0.02%	0.81%
6	2041	0.08%	1.21%
5	4976	0.20%	1.79%
4	23253	0.94%	8.31%
3	84157	3.39%	8.48%
2	118421	4.77%	15.53%
1	1502488	60.54%	22.90%
0	745801	30.05%	40.65%

- The longest CNAME chains seen at university's data is 10
- ＞3 CNAME chains
  - 30,814 (1.2%) hostnames have ＞3 CNAME chains
  - They generate 12.4% of valid A queries

Who offer long CNAME chains: from the university's data

ASN	Num of QNAMEs	Ratio of Num of Qnames	Ratio of Num of Queries	Num of Qnames >3 CNAME chains	Ratio of Qnames >3 CNAME chains	Ratio of Queries >3 CNAME chains
15169	1,160,009	46.74%	17.52%	0		
8075	321,366	12.95%	16.96%	17,428	0.70%	4.57%
16509	203,197	8.19%	10.64%	645	0.03%	0.06%
20940	49,076	1.98%	10.38%	476	0.02%	4.88%
16625	16,961	0.68%	4.76%	204	0.01%	0.70%
714	5,450	0.22%	4.58%	1	0.00%	0.03%

- This table shows the number/ratio of query names, the ratio of query volumes, and the number of query names with ＞3 CNAME chains
  for each AS to which the A record corresponding to the query name belongs.
- AS15169 does not use >3 CNAME chains
- AS20940 offers 476 hostnames that use >3 CNAME chains, 4.88% of valid A queries
- AS8075 offers 17428 hostnames that use >3 CNAME chains, 4.57% of valid A queries

Number of CNAME chains: from the university's data, AS20940

Number of CNAME chains	number of QNAMEs	ratio of number of QNAMEs	ratio of number of queries
7	2	0.00%	0.00%
6	12	0.02%	0.45%
5	75	0.15%	4.84%
4	387	0.79%	41.67%
3	2,133	4.35%	28.74%
2	36,428	74.23%	10.29%
1	6,054	12.34%	10.78%
0	3,985	8.12%	3.23%

- Some popular services that use Akamai CDN use longer CNAME chains
  - *.spotify*.com
  - *.bing.com
  - *.office.com
  - *.amazon.com *.aws.com

Number of CNAME chains: from the university's data, AS8075

Number of CNAME chains	Number of QNAMEs	ratio of number of QNAMEs	ratio of number of queries
10	1	0.00%	0.00%
9	8	0.00%	0.05%
8	53	0.00%	0.26%
7	482	0.02%	0.81%
6	2041	0.08%	1.21%
5	4976	0.20%	1.79%
4	23253	0.94%	8.31%
3	84157	3.39%	8.48%
2	118421	4.77%	15.53%
1	1502488	60.54%	22.90%
0	745801	30.05%	40.65%

- 95% of hostnames, 79% of queries have <3 (or no) CNAME chains
- Some services that use AS8075 use longer CNAME chains
  - {LABEL}.sharepoint.com
    - uses 5 CNAME chains
    - check: icannorg.sharepoint.com
  - Some of www.brother.{ccTLD}
    - use AS8075 and results 7 CNAME chains

Conclusion

- Long CNAME chains not only complicate name resolution for full-service resolvers, but also increase the execution time of stub-resolver/application software.
- A quick survey of the number of CNAME chains revealed that around 9-10 chains were in use.
- We also found that some well-known services use CNAME chains of 5 or more.
- In order to improve the efficiency of name resolution, we would like to consider setting an upper limit on the number of CNAME chains in the future
- Things change. For example, www.brother.in had 9 CNAME chains in April to 7 CNAME chains in July.