Root Zone Augmentation Impact Analysis

(The Other Root Scaling Study)

Geoff Sisson
Duane Wessels
DNS-OARC

RIPE 59, Lisbon
October 8, 2009

DNS-OARC

Interesting Times Ahead for the DNS Root

- IPv6 Glue
- DNSSEC
- New TLDs
- IDNs

Also...

- Continued anycast deployment
- Continued increase in query rates

This Study of Root Zone Changes

- ICANN hired OARC to simulate changes to the root zone and explore how they affect:
  1. The size of the root zone
  2. Server response latency
  3. Server start and reload times
  4. Bandwidth requirements for AXFR and IXFR
  5. Changes in response size, with an eye toward EDNS0, fragmentation, and TCP

Hardware

- DNS-OARC Testbed[*]
- 16 HP Proliant DL140 G3 servers
  - 4-cores of 3GHz Xeons
  - Most with 16 GB RAM, one with 32 GB
- Pair of 1000Base-T switches

[*]Thank you National Science Foundation, grant OCI-0427144, CAIDA, and ISC

Software

- Testing authoritative nameservers
  - BIND 9.6.0-P1
  - NSD 3.2.1
- Mostly on CentOS 5.3, a little on FreeBSD-7.1
- dnsperf, tcpreplay, NIST Net, and various custom tools.

Zone File Configurations

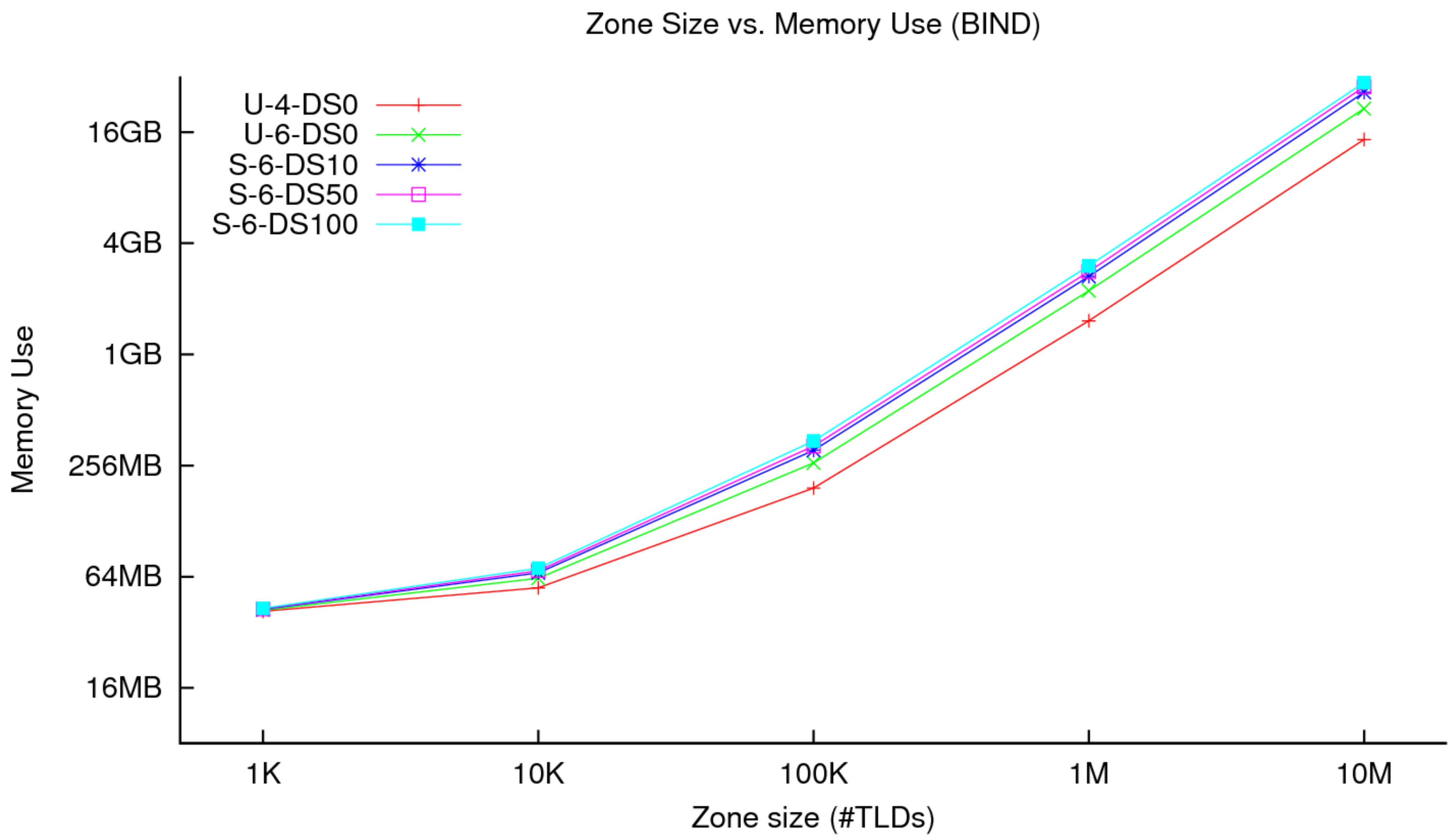
- Five types of zone content

U-4-DS0	Unsigned, mostly Ipv4 glue
U-6-DS0	Unsigned, Ipv4 and Ipv6 glue
S-6-DS10	Signed, Ipv6 glue, 10% DS records
S-6-DS50	Signed, Ipv6 glue, 50% DS records
S-6-DS100	Signed, Ipv6 glue, 100% DS records

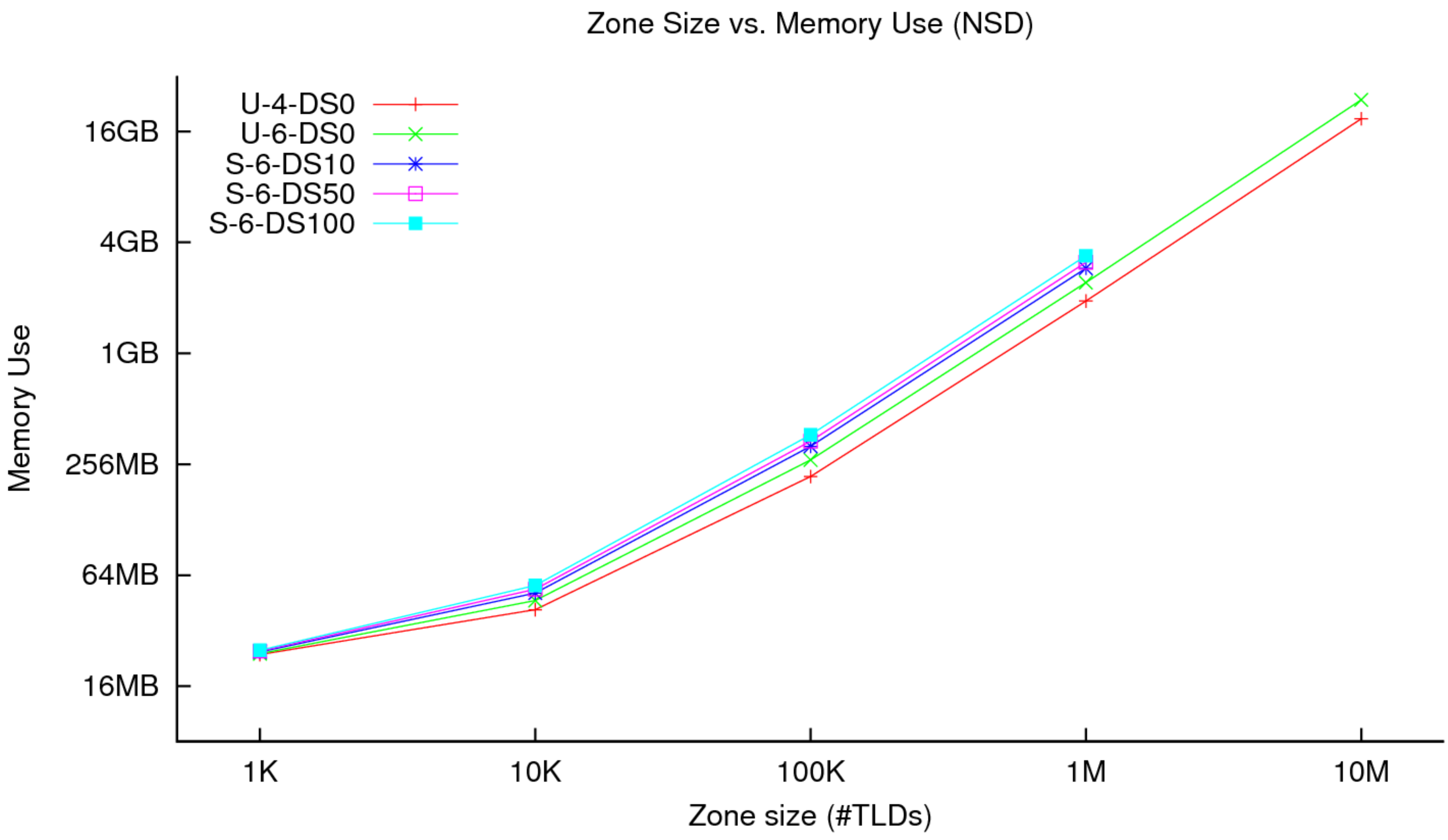
- Five zone sizes (number of TLDs)
  - 1K, 10K, 100K, 1M, 10M

Task 1: Memory Usage

- *How do root zone changes affect zone size and memory usage?*
- Process memory usage measured with *pmap*.
  - Includes memory used by the code segment and shared libraries.



Zone Size vs. Memory Use (BIND)
U-4-DS0
U-6-DS0
S-6-DS10
S-6-DS50
S-6-DS100
16GB
4GB
1GB
256MB
64MB
16MB
Memory Use
1K
10K
100K
1M
10M
Zone size (#TLDs)



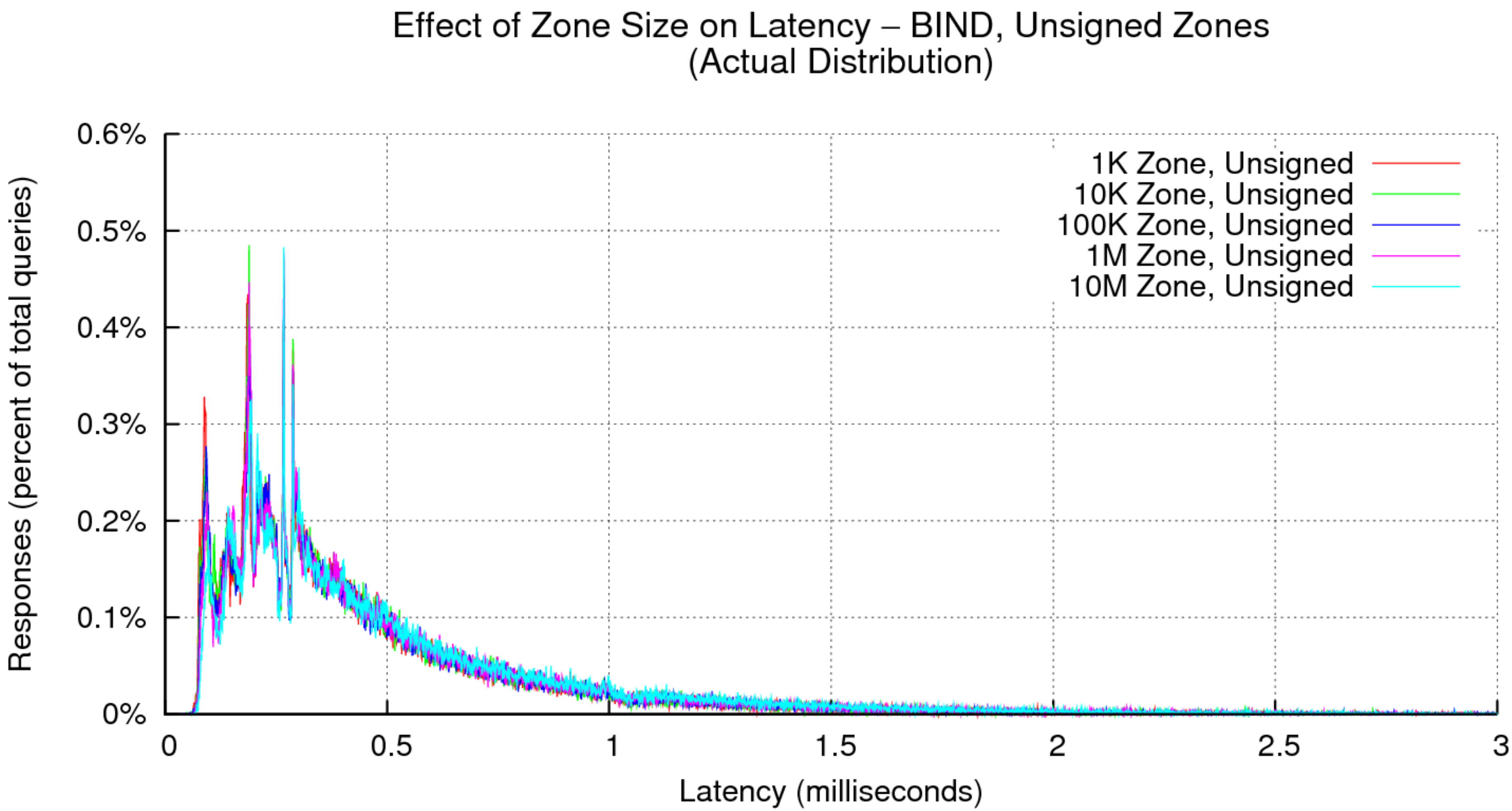
Zone Size vs. Memory Use (NSD)
U-4-DS0
U-6-DS0
S-6-DS10
S-6-DS50
S-6-DS100
Memory Use
16GB
4GB
1GB
256MB
64MB
16MB
1K
10K
100K
1M
10M
Zone size (#TLDs)

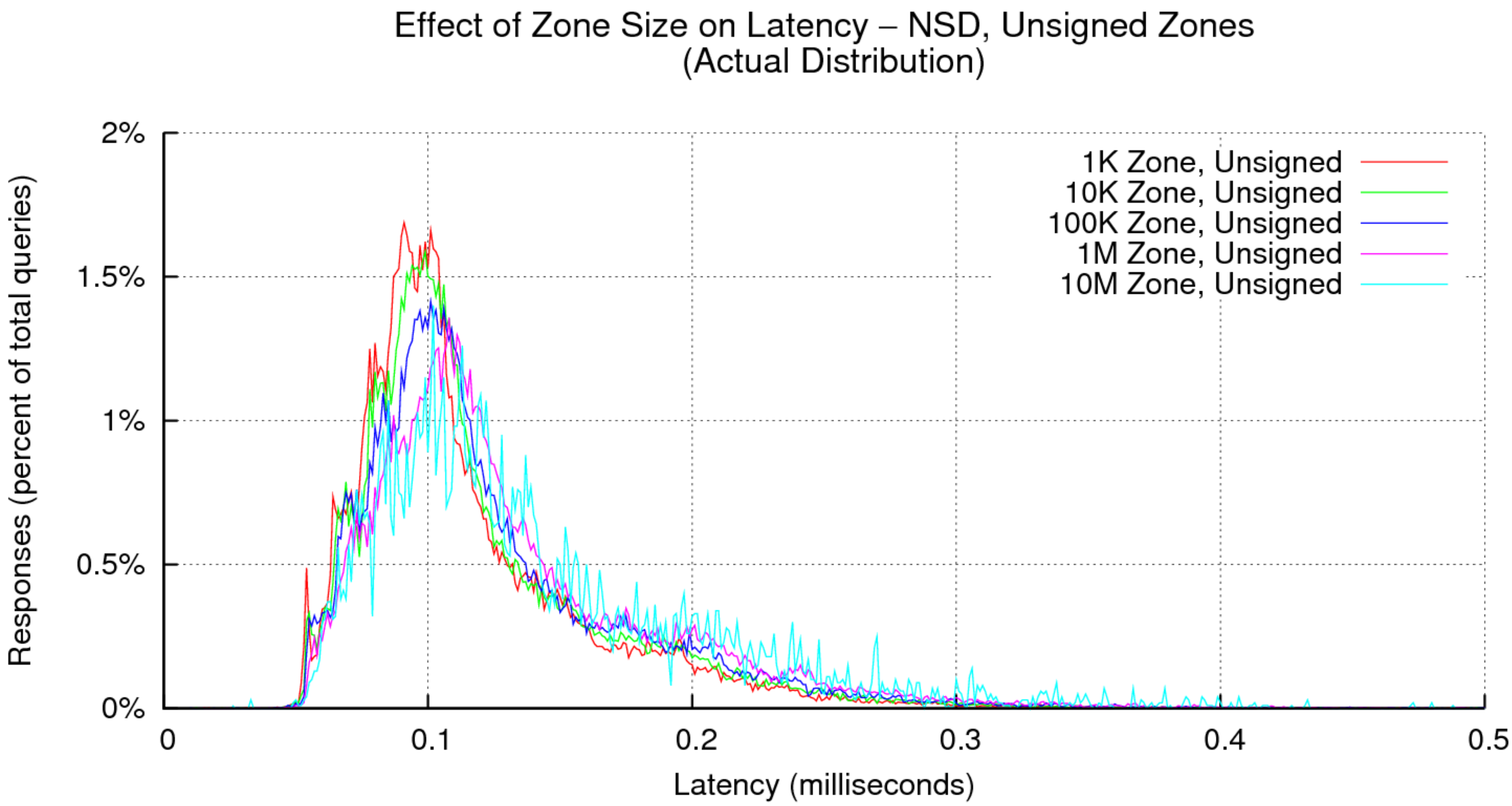
Task 1 Conclusions

- Process memory usage is proportional to zone size.
- A “S-6-DS100” zone uses about twice the memory as “U-4-DS0.”
- NSD needs more than 32 GB RAM to load a 10M TLD signed zone.

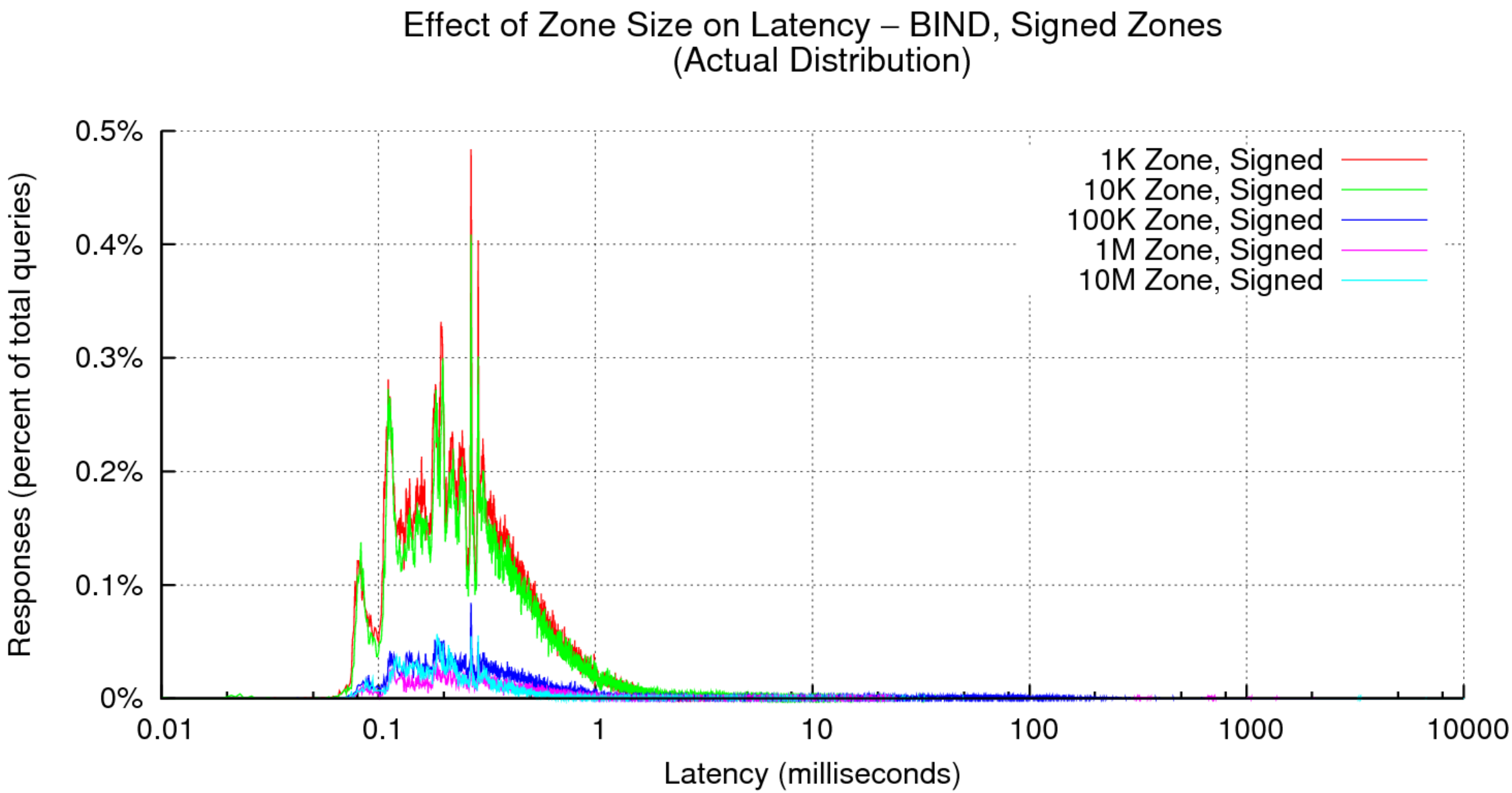
Task 2: Response Latency

- *How does latency of an “L-root analog” vary as a function of zone size?*
- Built pcap files of DNS queries with characteristics based on DITL-200903 data.
- Replayed with *tcpreplay*
- Constant query rate of 5000/sec

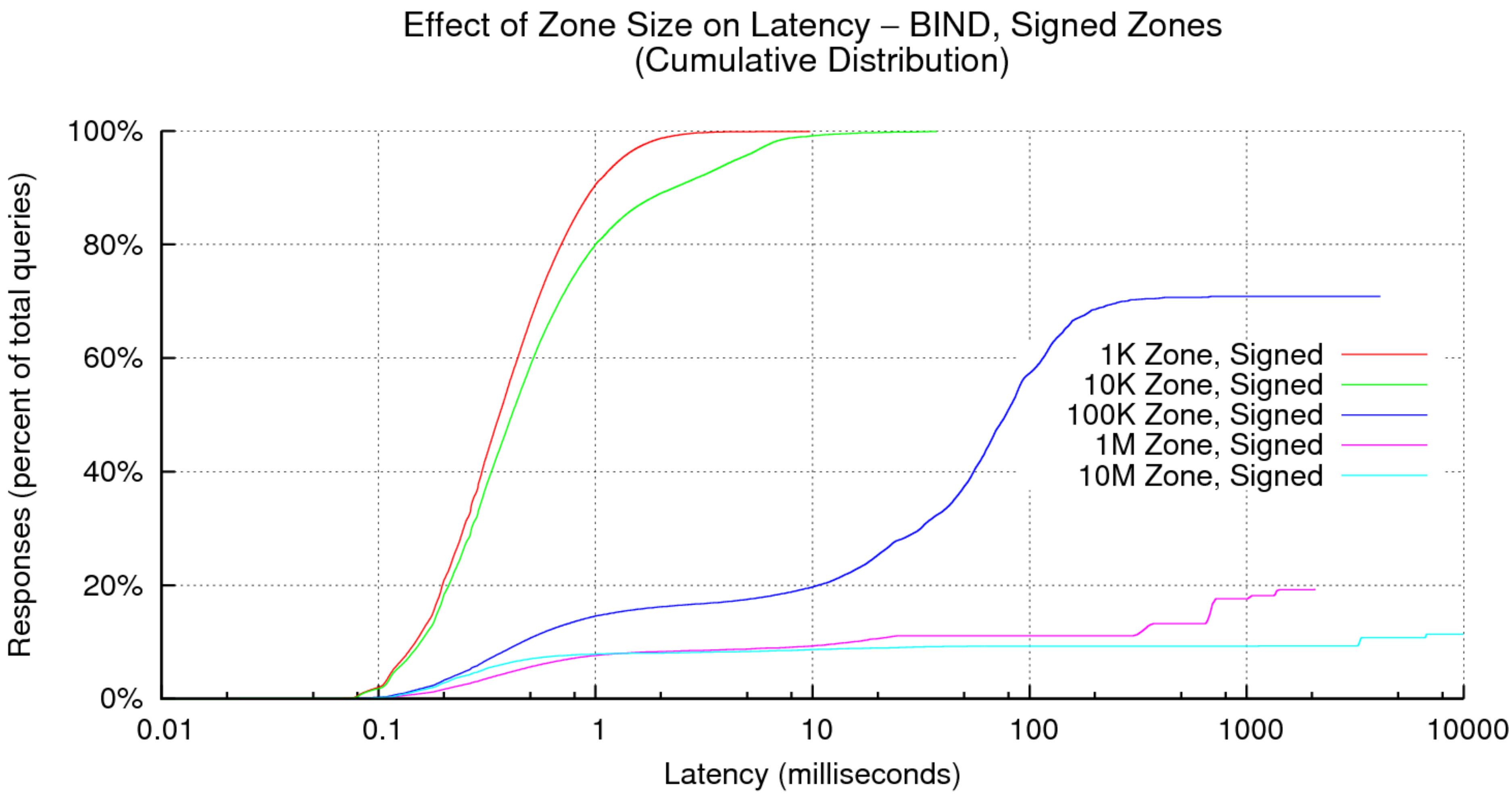

Effect of Zone Size on Latency – BIND, Unsigned Zones
(Actual Distribution)
Responses (percent of total queries)
0%
0.1%
0.2%
0.3%
0.4%
0.5%
0.6%
0
0.5
1
1.5
2
2.5
3
Latency (milliseconds)
1K Zone, Unsigned
10K Zone, Unsigned
100K Zone, Unsigned
1M Zone, Unsigned
10M Zone, Unsigned



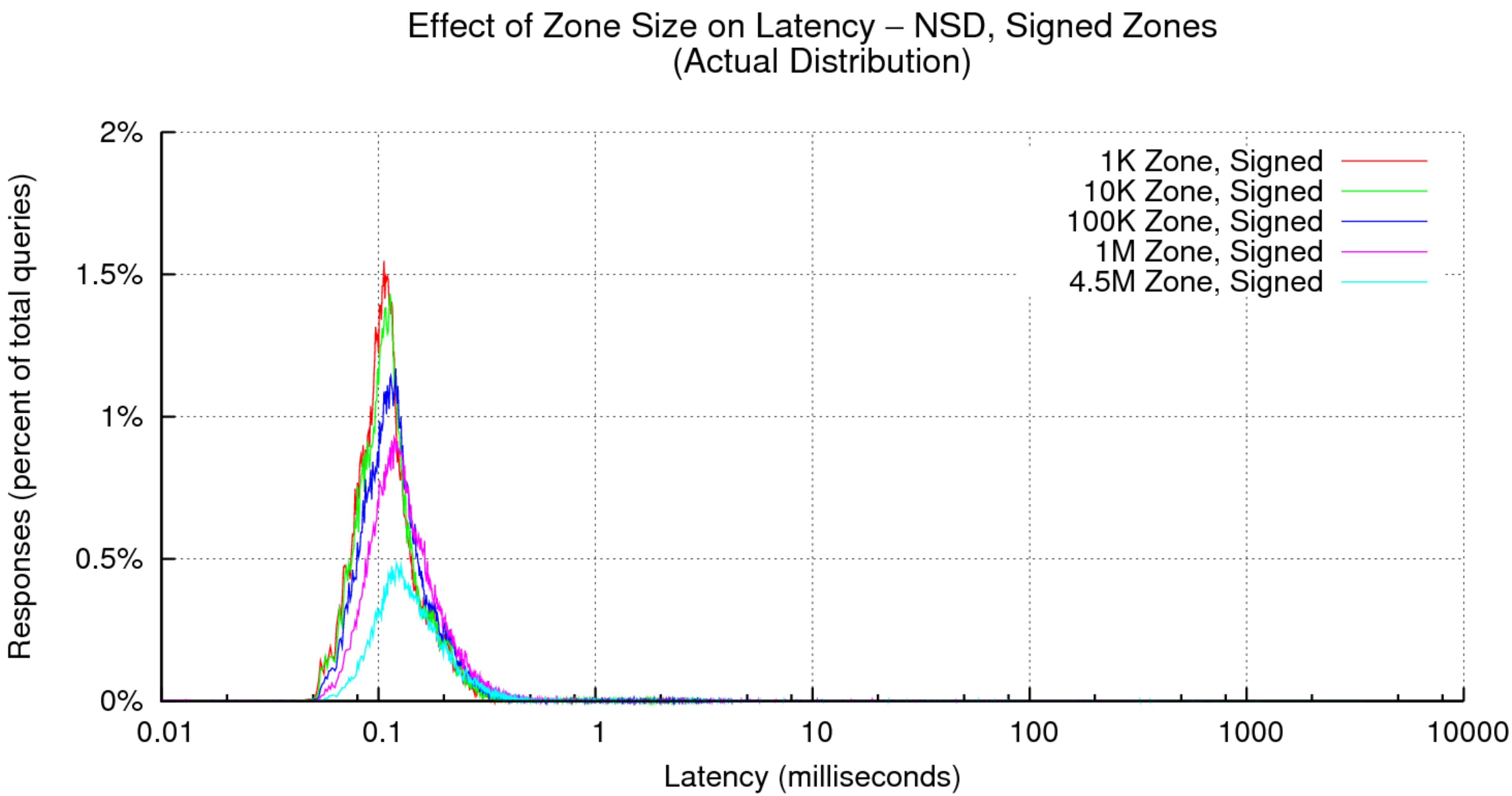
Effect of Zone Size on Latency – NSD, Unsigned Zones
(Actual Distribution)
Responses (percent of total queries)
2%
1.5%
1%
0.5%
0%
0
0.1
0.2
0.3
0.4
0.5
Latency (milliseconds)
1K Zone, Unsigned
10K Zone, Unsigned
100K Zone, Unsigned
1M Zone, Unsigned
10M Zone, Unsigned



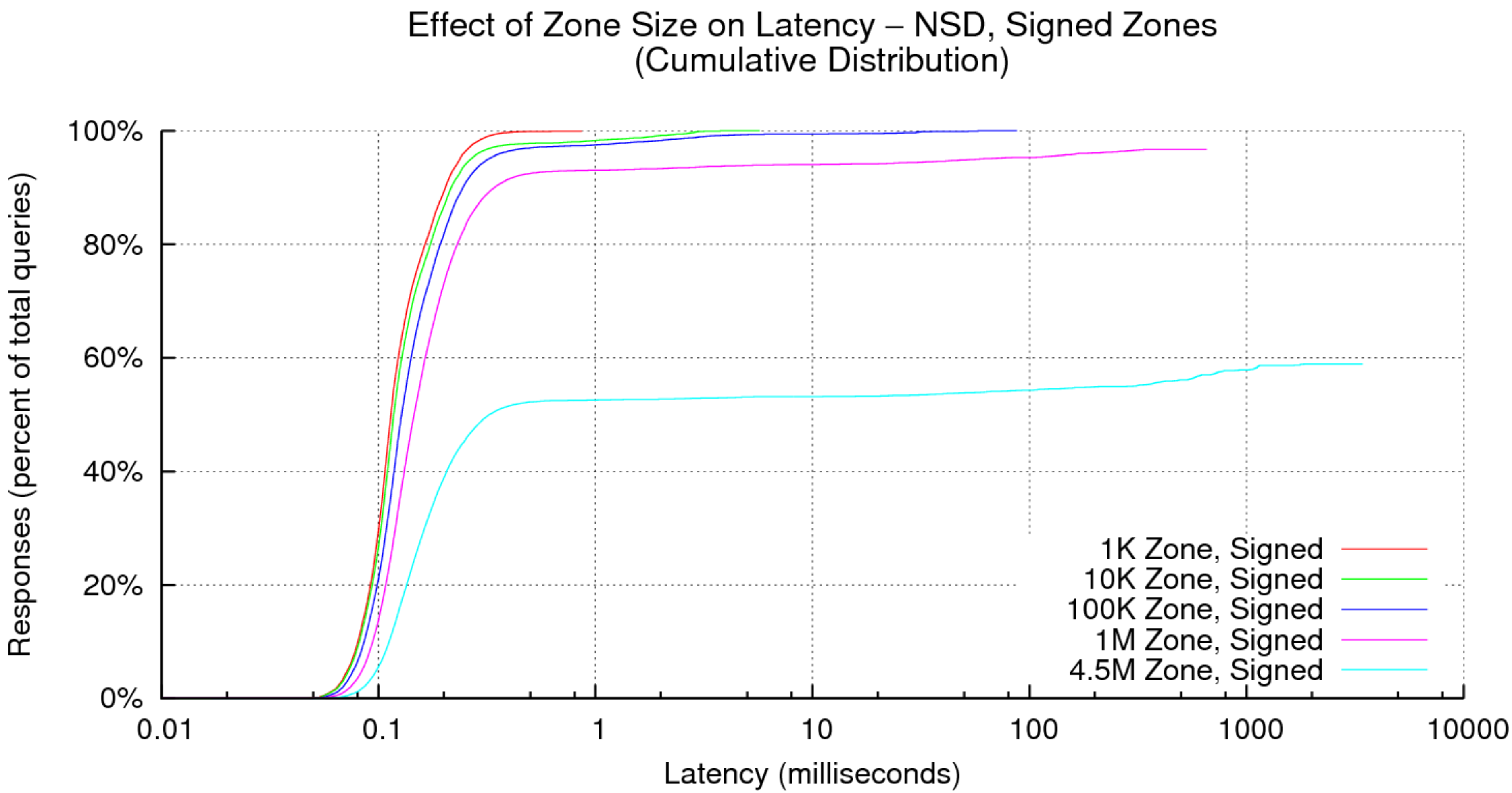
Effect of Zone Size on Latency – BIND, Signed Zones
(Actual Distribution)
Responses (percent of total queries)
0.5%
0.4%
0.3%
0.2%
0.1%
0%
0.01
0.1
1
10
100
1000
10000
Latency (milliseconds)
1K Zone, Signed
10K Zone, Signed
100K Zone, Signed
1M Zone, Signed
10M Zone, Signed



Effect of Zone Size on Latency – BIND, Signed Zones
(Cumulative Distribution)
Responses (percent of total queries)
100%
80%
60%
40%
20%
0%
0.01
0.1
1
10
100
1000
10000
Latency (milliseconds)
1K Zone, Signed
10K Zone, Signed
100K Zone, Signed
1M Zone, Signed
10M Zone, Signed



Effect of Zone Size on Latency – NSD, Signed Zones
(Actual Distribution)
Responses (percent of total queries)
2%
1.5%
1%
0.5%
0%
0.01
0.1
1
10
100
1000
10000
Latency (milliseconds)
1K Zone, Signed
10K Zone, Signed
100K Zone, Signed
1M Zone, Signed
4.5M Zone, Signed



Effect of Zone Size on Latency – NSD, Signed Zones
(Cumulative Distribution)
100%
80%
60%
40%
20%
0%
Responses (percent of total queries)
0.01
0.1
1
10
100
1000
10000
Latency (milliseconds)
1K Zone, Signed
10K Zone, Signed
100K Zone, Signed
1M Zone, Signed
4.5M Zone, Signed

Task 2 Conclusions

- BIND performance is stable for all sizes of unsigned zones.
- BIND performance degrades with larger signed zones.
  - ISC has already identified the code related to this problem and is working on a solution.
- NSD performance is stable for all sizes of unsigned zones.
- NSD shows some degradation in a 4.5M signed zone.

BIND Performance Issue

- Ony with NSEC. No issue with NSEC3
- Only with a zone like the root which is likely to have a large number of glue owner names that get sorted between non-glue.
- Only for a larger (ie 100K TLD) root zone.
- Plenty of time until this fix will really be necessary in production.

Problematic Zone Data

```
...
COM.                            172800 IN NS      M.GTLD-SERVERS.NET.
COM.                             86400 IN NSEC    COMBATSON. NS RRSIG
NSEC
COM.                             86400 IN RRSIG   NSEC 5 1  ...
NS2.00MAPDATEANYTHING7.COM.     172800 IN A       204.115.66.58
NS2.00MAPDATEANYTHING7.COM.     172800 IN AAAA    2001:838:8d:3::9a
NS2.00VOTESC3FLYBELTIF.COM.     172800 IN A       65.53.226.151
NS2.00VOTESC3FLYBELTIF.COM.     172800 IN AAAA    2001:838:8b:5::ed
A.NS.01ITANIITSROME.COM.        172800 IN A       216.36.92.178
A.NS.01ITANIITSROME.COM.        172800 IN AAAA    2001:470:b1:5::28
NS1.01ONFIELDREALTHELIE9.COM.   172800 IN A       81.126.47.187
....
(~10000 more glue records)
...
COMBATSON.                      172800 IN NS      B.COMBATSON.
COMBATSON.                      172800 IN NS      D.OURCOSTSGONEDID.COM.
```

Task 3: Start and Reload Times

- *How does nameserver startup and reload time vary with zone size and characteristics?*
- Start or restart nameserver.
- Record time taken to serve a record at the end of the zone file.
- NSD times include zone compilation.

BIND Start Times (seconds)

Zone Type	1K	10K	100K	1M	10M
U-4-DS0	<1	<1	8	87	950
U-6-DS0	<1	<1	11	113	1153
S-6-DS10	<1	<1	14	157	1581
S-6-DS50	<1	<1	16	170	1723
S-6-DS100	<1	2	17	190	1911

BIND Reload Times (seconds)

Zone Type	1K	10K	100K	1M	10M
U-4-DS0	<1	<1	8	90	1012
U-6-DS0	<1	<1	11	122	1240
S-6-DS10	<1	2	16	168	N/A
S-6-DS50	<1	2	18	203	N/A
S-6-DS100	<1	2	18	200	N/A

NSD Start Times (seconds)

Zone Type	1K	10K	100K	1M	10M
U-4-DS0	<1	2	13	147	1601
U-6-DS0	<1	2	15	173	1763
S-6-DS10	<1	2	18	197	N/A
S-6-DS50	<1	3	19	210	N/A
S-6-DS100	<1	3	21	227	N/A

NSD Reload Times (seconds)

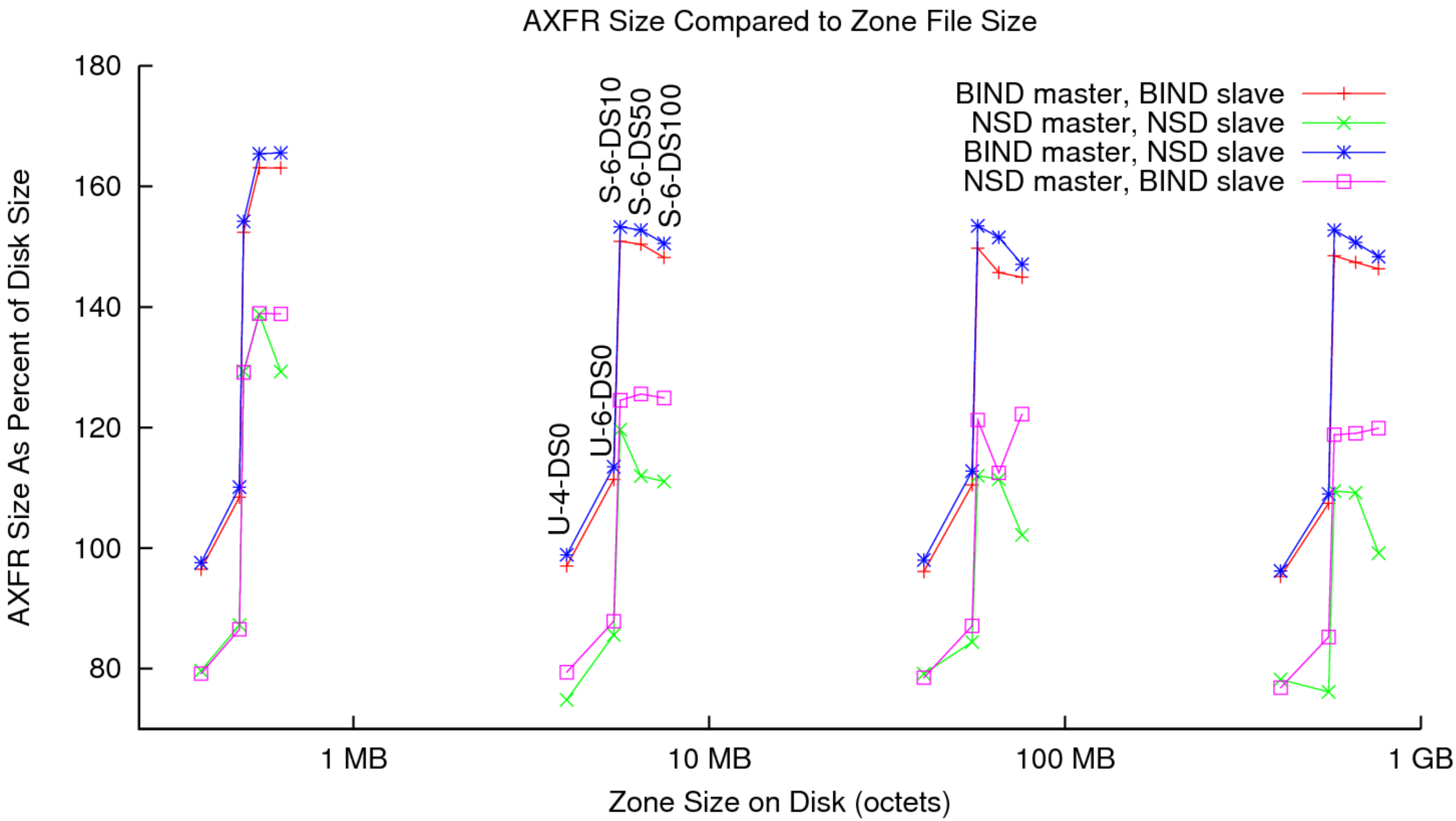
Zone Type	1K	10K	100K	1M	10M
U-4-DS0	<1	2	14	147	1603
U-6-DS0	<1	2	16	175	1778
S-6-DS10	<1	2	18	203	N/A
S-6-DS50	<1	2	21	211	N/A
S-6-DS100	<1	3	22	231	N/A

Task 3 Conclusions

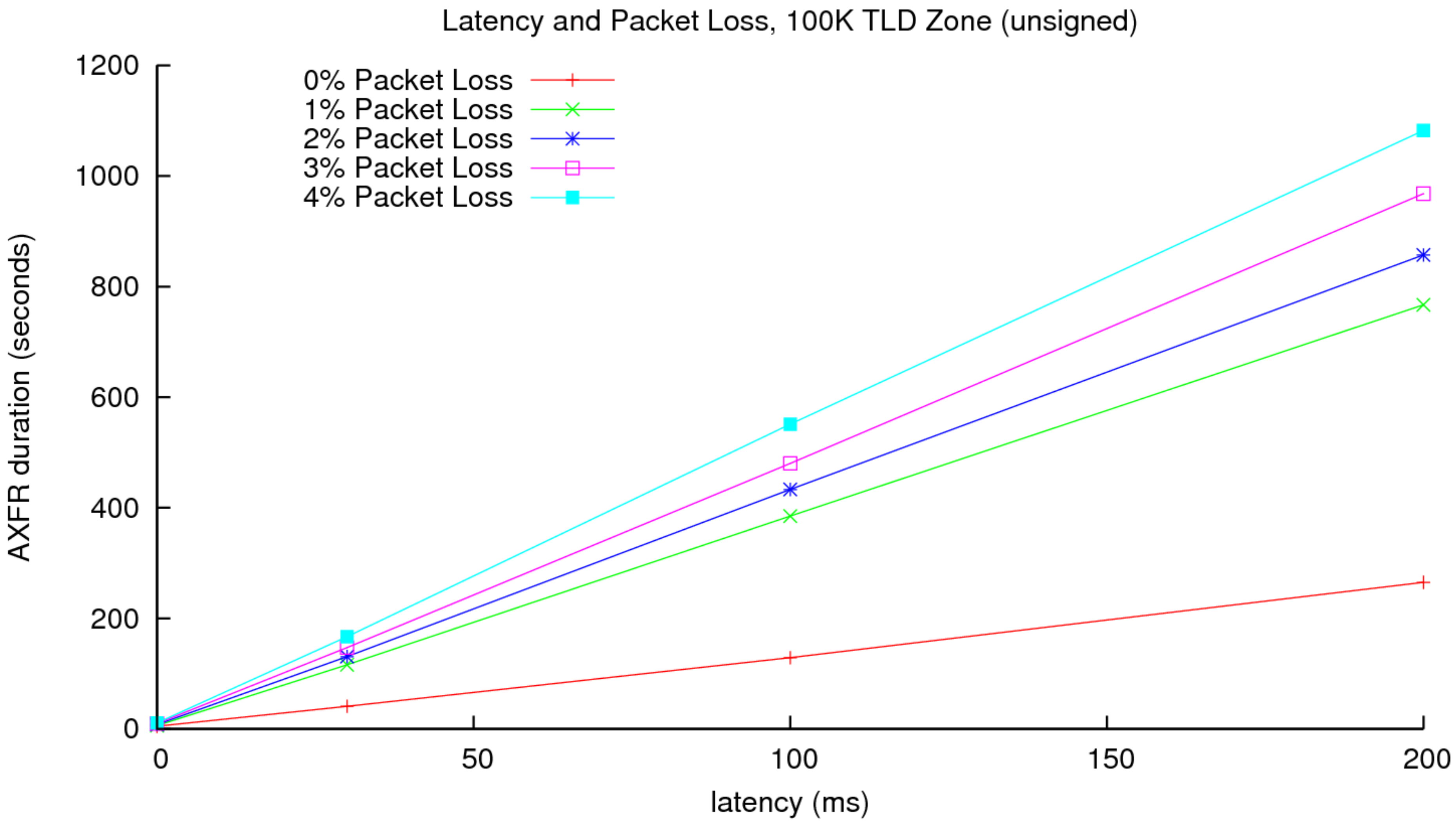
- Start and Reload times are proportional to zone size.
- BIND requires more than 32 GB RAM to reload 10M TLD signed zones.

Task 4: Bandwidth & Transfer Times

- *What are the remote node bandwidth requirements for an increased number of TLDs using AXFR and IXFR?*
- Bandwidth and duration measurements taken from pcap files captured during simulations.



AXFR Size Compared to Zone File Size
AXFR Size As Percent of Disk Size
Zone Size on Disk (octets)
180
160
140
120
100
80
1 MB
10 MB
100 MB
1 GB
BIND master, BIND slave
NSD master, NSD slave
BIND master, NSD slave
NSD master, BIND slave
S-6-DS10
S-6-DS50
S-6-DS100
U-6-DS0
U-4-DS0

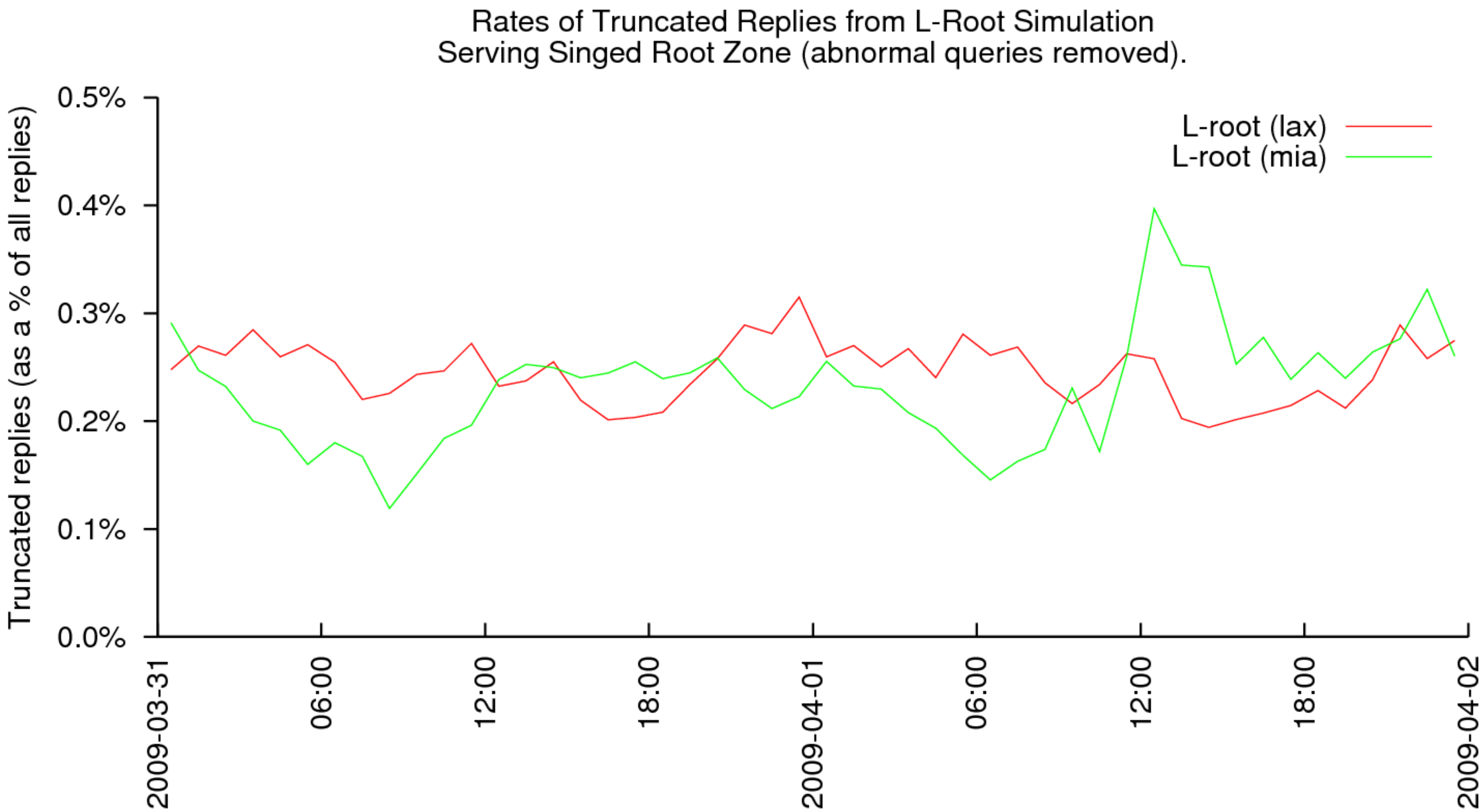

Latency and Packet Loss, 100K TLD Zone (unsigned)
0% Packet Loss
1% Packet Loss
2% Packet Loss
3% Packet Loss
4% Packet Loss
AXFR duration (seconds)
latency (ms)
0
200
400
600
800
1000
1200
0
50
100
150
200

Task 4 Conclusions

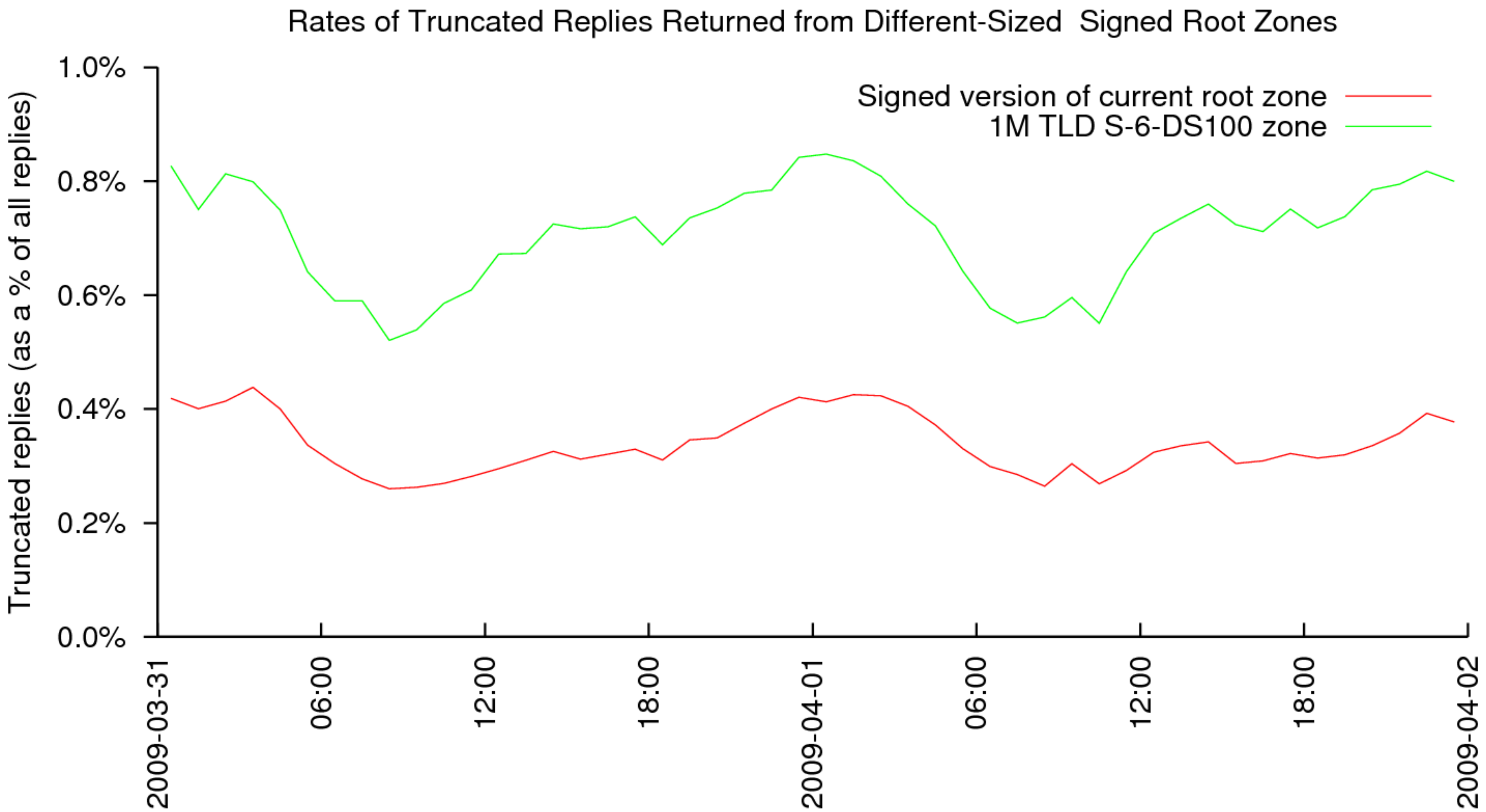
- An NSD master uses 20-30% less bandwidth for AXFR due to name compression.
- For unsigned zones, IXFR incurs an overhead of 20-50% compared to AXFR.
  - For example, updating 10% of zone contents with IXFR uses 14% as much bandwidth as a full AXFR.
- For signed zones, the IXFR overhead is closer to 100%.
- 1% packet loss increases AXFR times by a factor of 3. 4% by a factor of 4.

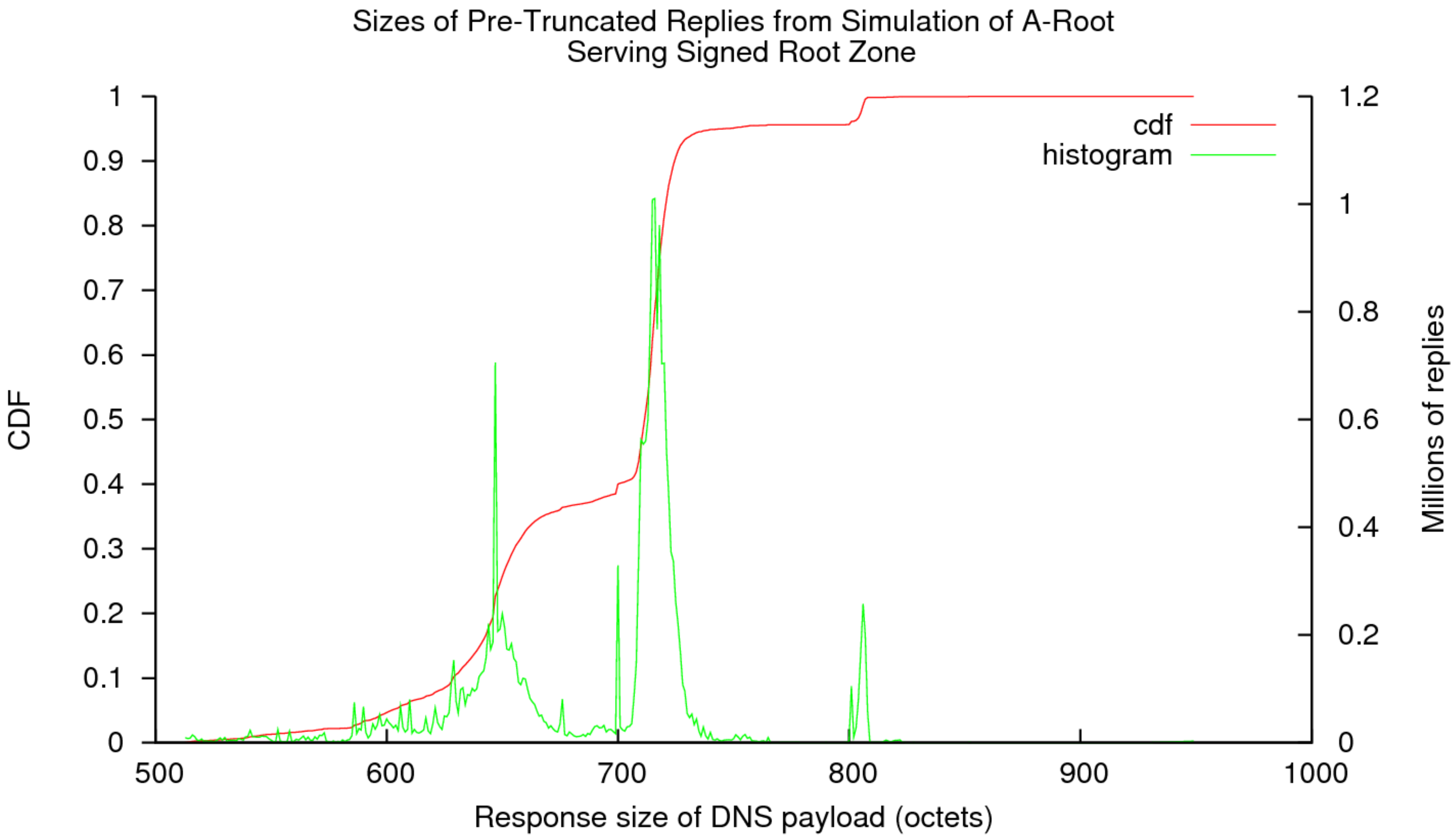
Task 5: TCP Usage

- *To what extent will DNSSEC and IPv6 glue increase TCP usage?*
- Replayed DITL client traces against larger root zones.



Rates of Truncated Replies from L-Root Simulation
Serving Singed Root Zone (abnormal queries removed).
Truncated replies (as a % of all replies)
0.5%
0.4%
0.3%
0.2%
0.1%
0.0%
L-root (lax)
L-root (mia)
2009-03-31
06:00
12:00
18:00
2009-04-01
06:00
12:00
18:00
2009-04-02


Rates of Truncated Replies Returned from Different-Sized Signed Root Zones
Truncated replies (as a % of all replies)
1.0%
0.8%
0.6%
0.4%
0.2%
0.0%
Signed version of current root zone
1M TLD S-6-DS100 zone
2009-03-31
06:00
12:00
18:00
2009-04-01
06:00
12:00
18:00
2009-04-02



EDNS 512 queries replayed with larger EDNS size

Task 5 Conclusions

- Root servers can expect about an order of magnitude increase in queries over TCP when the root is signed.
  - Study predicts A root will go from 5/sec to 50/sec.
- Increasing number of TLDs also appears to increase TCP traffic
  - Due to generally longer names in NSEC records?
- UDP Responses that might be truncated (i.e., EDNS size 512 with DO bit set) would be smaller than 825 bytes if not truncated.

Please read the full report

https://www.dns-oarc.net/files/rzaia/rzaia_report.pdf

Contact Us

Geoff Sisson <geoff@geoff.co.uk>

Duane Wessels <wessels@dns-oarc.net>

DNS-OARC