Testing DNS Performance limits the Final Chapter

Research by ISC for CAIDA

Funded by NSF

David Boggs, lead investigator
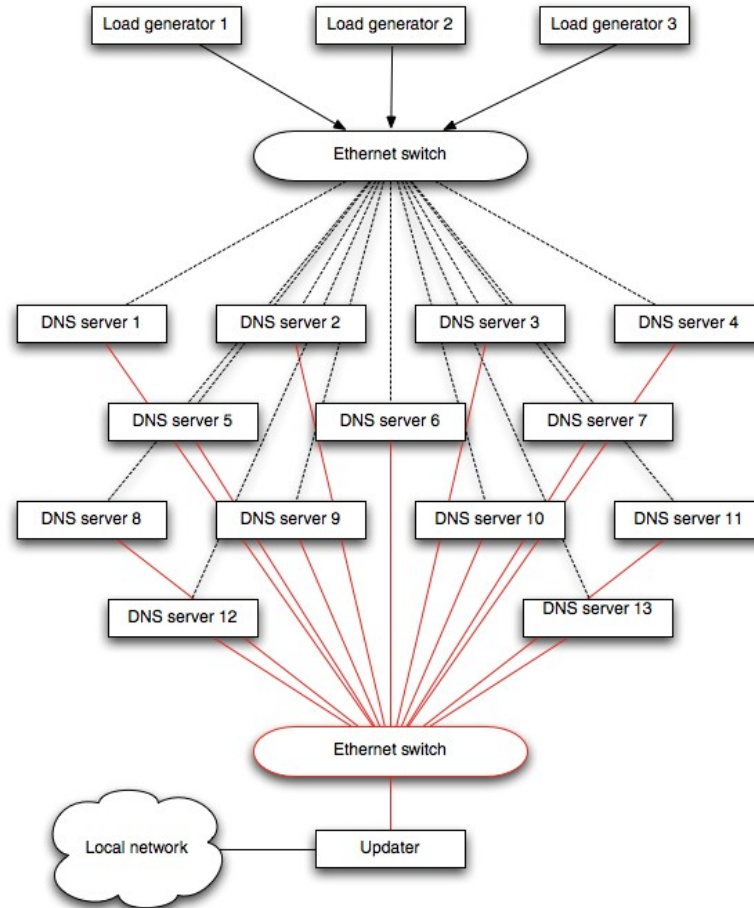
Brian Reid, writer and reporter

DNSPERF Project overview

- Build testbed big enough to test .COM, .NET TLD service

- Test its maximum capacity (query rate at server overload point)

- Reconfigure to use DDNS for updates, IXFR for distribution
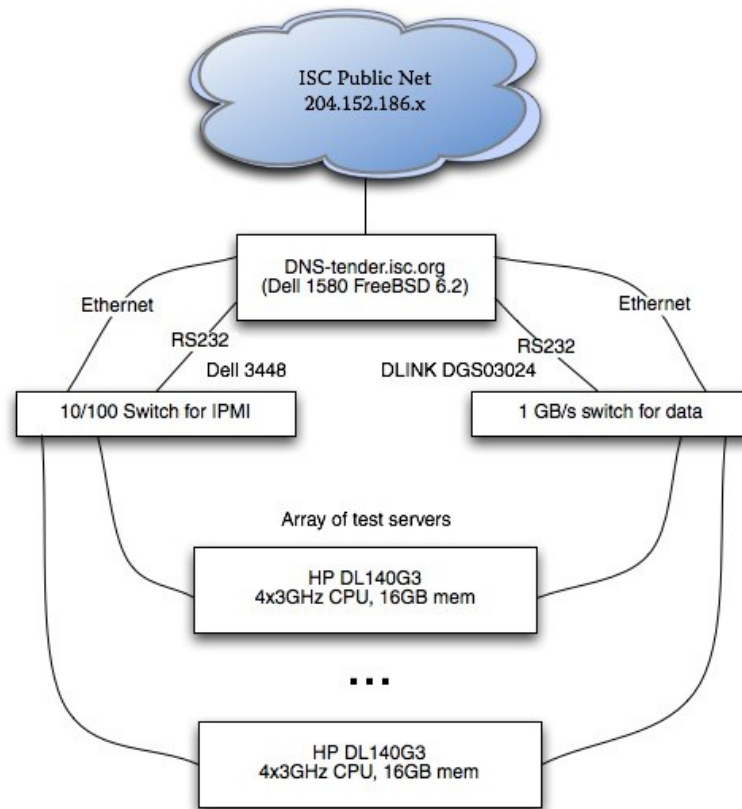
- Test under load, find maximum

Physical testbed

- 13 affordable COTS servers
- 1 Stealth master for IXFR sourcing
- Non-blocking GB ethernet
- Load generator
- Update generator
- Monitoring

Logical diagram

Physical diagram

ISC Public Net
204.152.186.x

DNS-tender.isc.org
(Dell 1580 FreeBSD 6.2)

Ethernet

Ethernet

RS232

RS232

Dell 3448

DLINK DGS03024

10/100 Switch for IPMI

1 GB/s switch for data

Array of test servers

HP DL140G3
4x3GHz CPU, 16GB mem

. . .

HP DL140G3
4x3GHz CPU, 16GB mem

What hardware?

- Affordable under limited budget ($100K available to buy 16 servers)
- Candidates: Sun X4200, HP DL140G3, Iron Systems I-class, M-class (Intel Xeon and AMD28x)
- Must run open-source OS
- Choose by memory performance

Hardware test results

	L1 Memtest MB/sec	LMbench Bandwidth MB/sec	L1 calibrator (NS for miss)	LMbench latency (NS)	STREAM Copy (MB/sec)	STREAM Add (MB/sec)	STREAM Triad (MB/sec)
HP DL140/G3	49058	2984	3.07	72	2586	2884	2890
Sun x4200 AMD254	22886	2316	3.48	83	1724	1896	1893
Sun x4200 AMD285	21251	2368	3.73	83	1816	1994	1958
Iron Systems M (AMD)	19717	-	4.08	-	-	-	-
Iron Systems I (Intel)	19607	2047	6.82	109	1309	1329	1524
HP Celestica	16331	1303	5.07	155	1122	1254	1138

Hardware decision

- HP DL140/G3

- Surprised that Intel processors outperformed AMD for these tests

- Able to afford 16GB RAM in each (8 pairs of matched 1GB parts)

What software?

- BIND 9.4
- OS: Test these, pick the fastest

  Linux (Gentoo, Fedora), FreeBSD (6, 7), Solaris 10, NetBSD 4, OpenBSD 4.1, Windows 2003 Server, Windows XP Pro64

What test?

- Loaded server with .PT zone
- Used queries from 48-hour F-Root capture, sent with queryperf
- Ramped query rate until server limit reached
- Ran test at server limit for 1 hour (1.13 millioin queries)

OS Performance queries/sec

Linux-Gentoo	Kernel 2.6.20.7	92327		Solaris-10	SunOS 5.11 snv-64a	41306
Linux-Fedora	Kernel 2.6.20.7	86732		NetBSD	4.0-beta2	36331
FreeBSD	7-current 200708	83089		OpenBSD	4.1-current 200705	35237
FreeBSD	6-stable 200708	54076		Windows 2003 Server	SP2 5.2.3790	22548
Solaris-10	SunOS 5.10 120012-14	53539		Windows XP Pro	SP2 5.2.3790	19888
FreeBSD	6.2-release	50611		Windows 2000 Pro	SP4 5.0.2195	18957

Test data stream

- 48-hour capture from F-Root

- `414931073 requests (38.8% failed)`

- `Avg rate (req/sec) = 2401.2`
  `95%ile burst = 3011.0`
  `    Max burst = 3921.9`

Test data stream

ISC baseline DNS test data set (15-17 November 2006)

Average rate — Minimum burst rate — Max burst rate
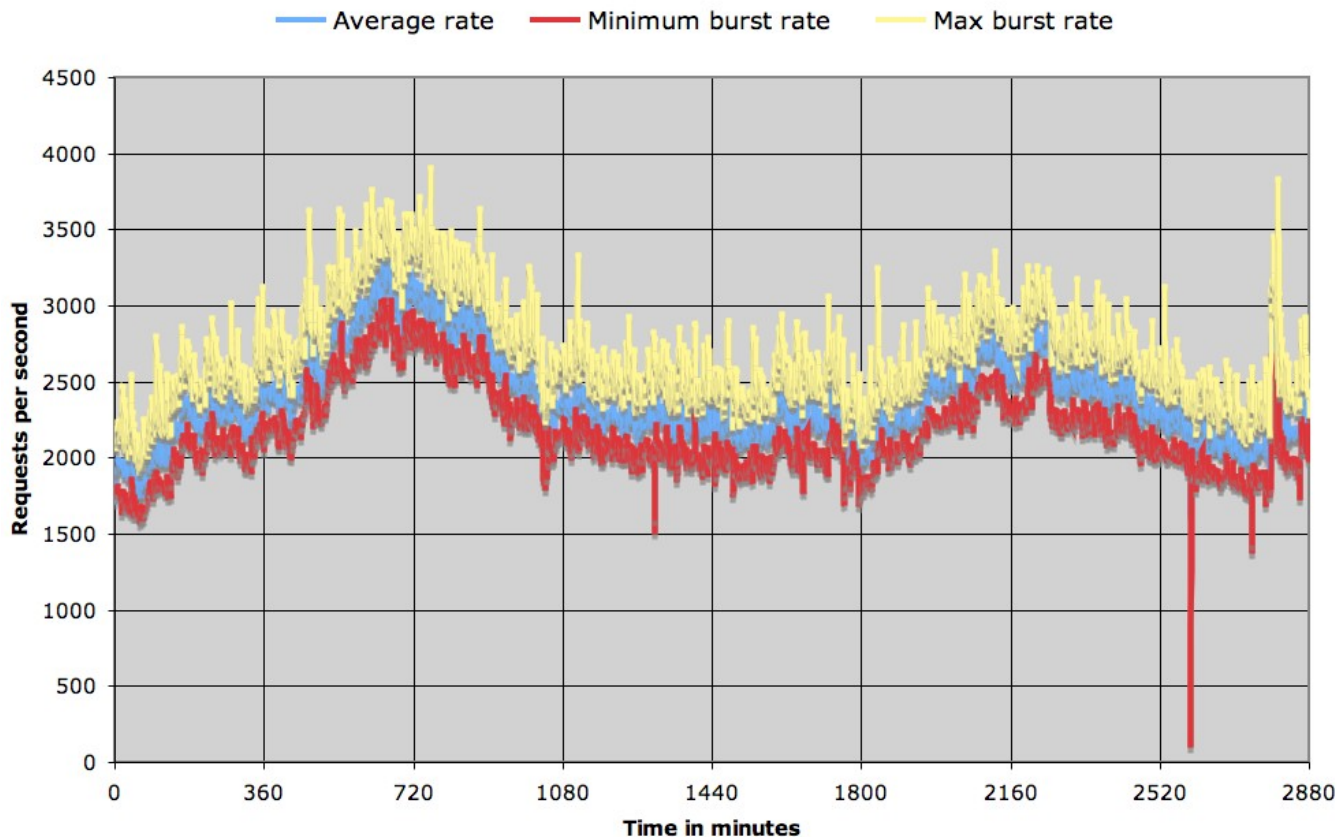
Testing with .COM

- Used COM zone from 5 Oct 2007
  - Split into 2 parts
  - Each part about 87,000,000 entries
  - One server for each part
- Raw zone slice file sizes 3GB each

New numbers

- We presented some numbers in November 2007.

- Today's numbers are different.

- Two sources of differences:
  - COM zone split into slices to fit memory
  - BIND `stderr` now piped to /dev/null (profiling showed lots of time wasted in console error messages)

Testing with COM in 2 slices

OS	Queries/sec average across slices
Linux-Fedora	78,701
FreeBSD-7-RC1	75,177
Linux-Gentoo	70,223
Solaris-10 (Edition of 8/07)	63,954
Solaris-10 (Edition of 5/07)	61,027
FreeBSD 6.2-STABLE	55,866
FreeBSD 6.2-RELEASE	53,127
NetBSD-4-current	32,129

Measuring update performance

- After measuring server performance on fixed zone, we measure rate at which we can update that zone.
- Updates use RFC2136 protocols and standard BIND tools
- Update stream was synthetic

DDNS updates to COM

- We sent 10,000 DDNS updates to BIND 9.1.1-P1 at maximum speed.
- Zone being updated was 50% slice of COM, one server per slice.
- Table shows rate at which updates can be committed to stable storage.
- We believe the FreeBSD 6.2 anomaly to be a compiler or config bug that omitted fsync calls.

COM update rate limits

OS	Slice 1	Slice 2
FreeBSD-7RC2	93.46	82.85
FreeBSD-6.2-Stable	89.52	77.34
FreeBSD-6.2-Release	471.70	446.23
NetBSD-4-Current	80.32	81.77
Solaris-10 5/07 ed	87.80	83.54
Solaris-10 8/07 ed	63.21	64.10
Linux-Gentoo	42.07	42.76
Linux-Fedora	22.29	15.07

Final test:
query during update

- Measured performance of COM zone during DDNS update and IXFR.
- Made slight modification to BIND 9.4.2 for this test (disabled periodic write to stable storage, since this is not master copy of zone)
- Measured for signed and unsigned zones.

Notes on the test and results

- Recall that DDNS updates to COM zone max out around 90/sec
- So the test sequence 1,2,4,8,16,32,64,128 updates/sec is really 1,2,4,8,16,32,64,90
- This test performed only with FreeBSD 7.0-RELEASE

Query reply rate of server under update via IXFR

Updates / second	Unsigned zone	Signed zone
0	69,400	61,200
4	67,600	60,100
8	68,100	58,500
16	67,400	58,000
32	60,900	54,900
64	60,408	51,200
128*	60,361	50,800

*DDNS updates of master maxed at 90/second

Conclusions

- Open source software and commodity hardware are up to the task of serving large real zones
- Real-world COM update rates average only 2-4/second, peaked at 10/second during 2006 and 2007.
- Signed zone service is only 10-15% slower than unsigned
- Slicing a zone for multiple servers works fine.
- Constant zone update does not significantly degrade server's query-response performance.

For more information

- ISC Tech Note TN-2008-1
  ftp://ftp.isc.org/isc/dns_perf/ISC-TN-2008-1.pdf

- Contact info@isc.org to inquire about research access to this testbed (it is available to other researchers)