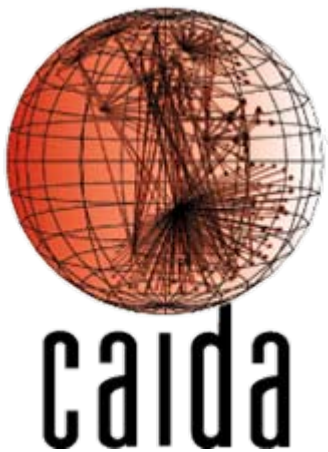
.CL anycast switching experiment

Sebastian Castro
secastro@caida.org
secastro@nic.cl

CAIDA
NIC Chile



8th CAIDA/WIDE workshop – July, 2007

Introduction

- The anycast technology has been widely deployed in various DNS services around the world.
  - With the objective of provide improved reliability.
- Several research has been conducted to get a deeper understanding of its behavior
  - Degrees of importance among different nodes (RIPE), geographic influence (CAIDA), etc.
- Wanted to know how a client is affected when an anycast node is shut down
  - Time of no-service

.CL anycast deployment

- One anycast cloud (a.nic.cl) with three nodes located in Chile
  - santiago, located in Santiago
  - valparaiso, located in Valparaiso, 150 km to the west of Santiago.
  - tucapel, located in Concepcion, 650 km to the south of Santiago.
- Additionally one unicast server (ns.nic.cl)
- 5 additional nameservers located around the world.

Methodology

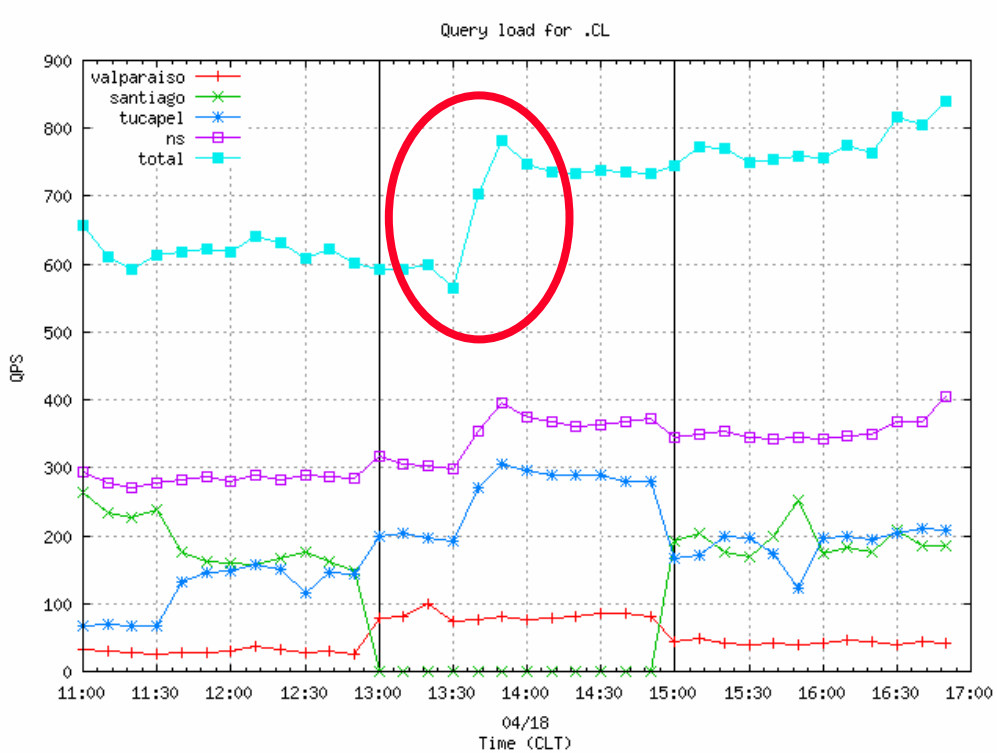
- 6-hour packet traces on every node (including the unicast server)
- Shut down one node two hours after the beginning of the trace (withdrawing the BGP announcement)
- Restore the node two hours after the shutdown (reinjecting the BGP prefix)
- Nodes time-synchronized with NTP

The Experiments

- Three different attempts
- Taking one anycast node on each instance
- First two were used as preparation
  - Didn't work as expected by operator mistake
- The third try took the busiest node down "**santiago**".
  - On April 18th, 2007.
  - From 11:00 to 17:00 CLT
    - 15:00 to 21:00 UTC
- Packet traces included queries and responses, UDP and TCP.

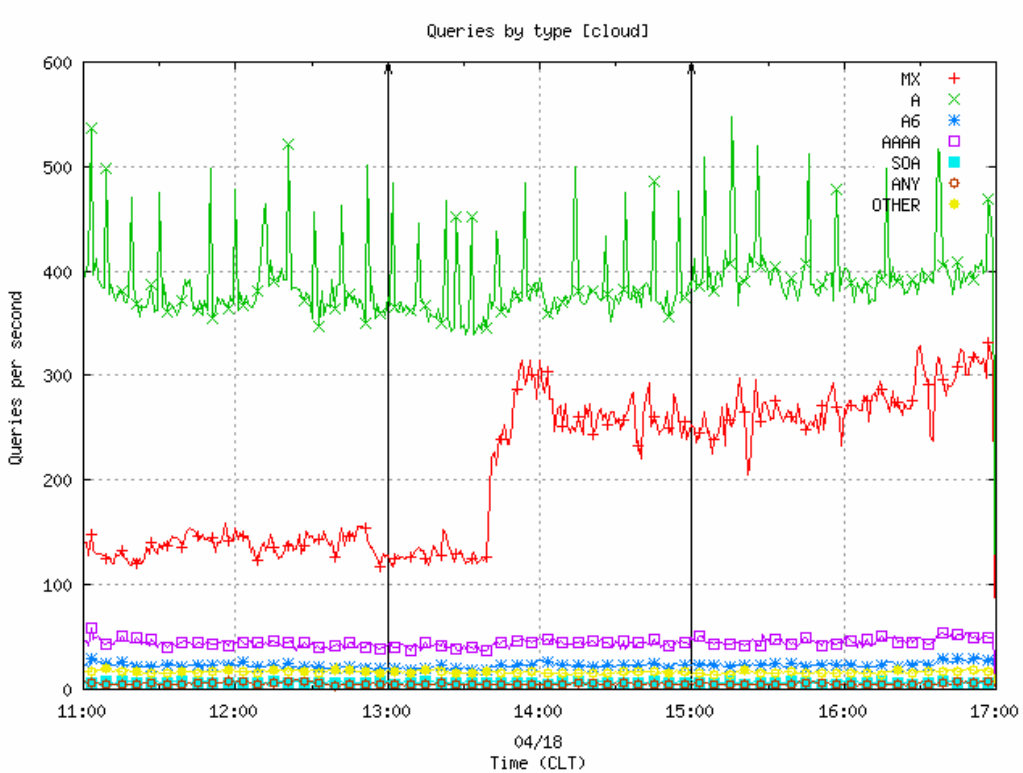
Results

- Query load
  - Within the cloud, **santiago** receives a biggest part of the load.
  - Within the country, the unicast server receives similar load of the anycast cloud.
  - An interesting spike at 13:40 CLT caught our attention

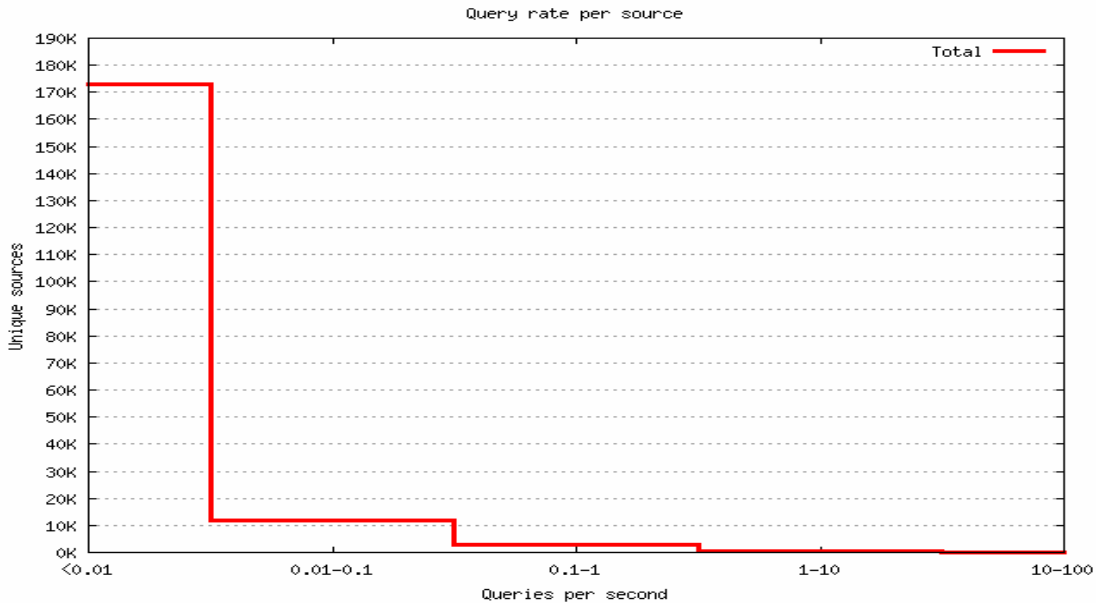


The spike

- The spike was produced by a burst of MX queries.
- Mainly coming from Chile, USA and Brasil.
  - Other countries also contributed.
  - The phenomenon lasted beyond the duration of the experiment.

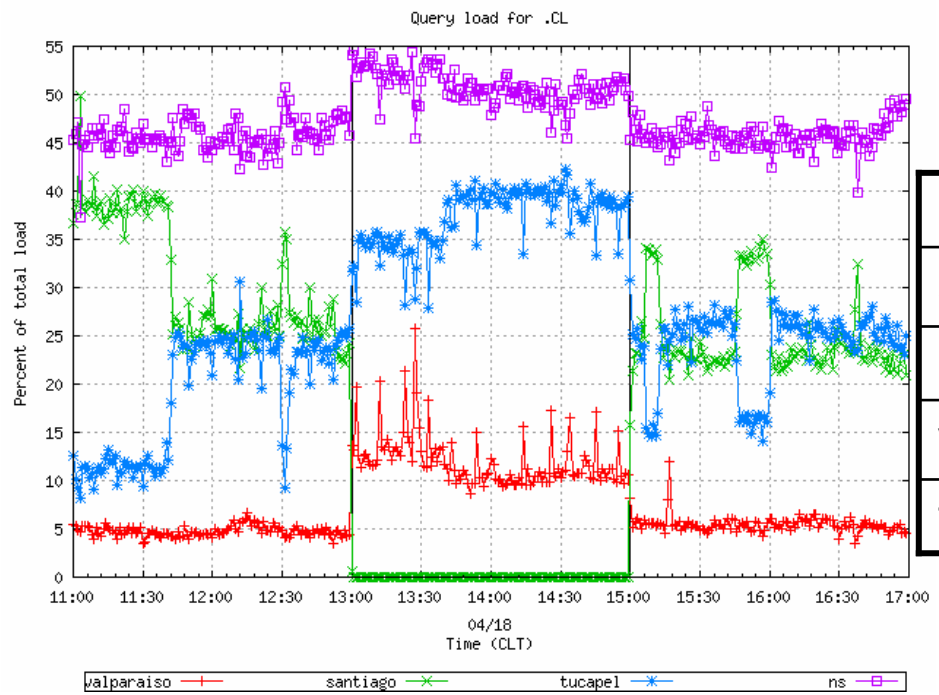


Query load per source address



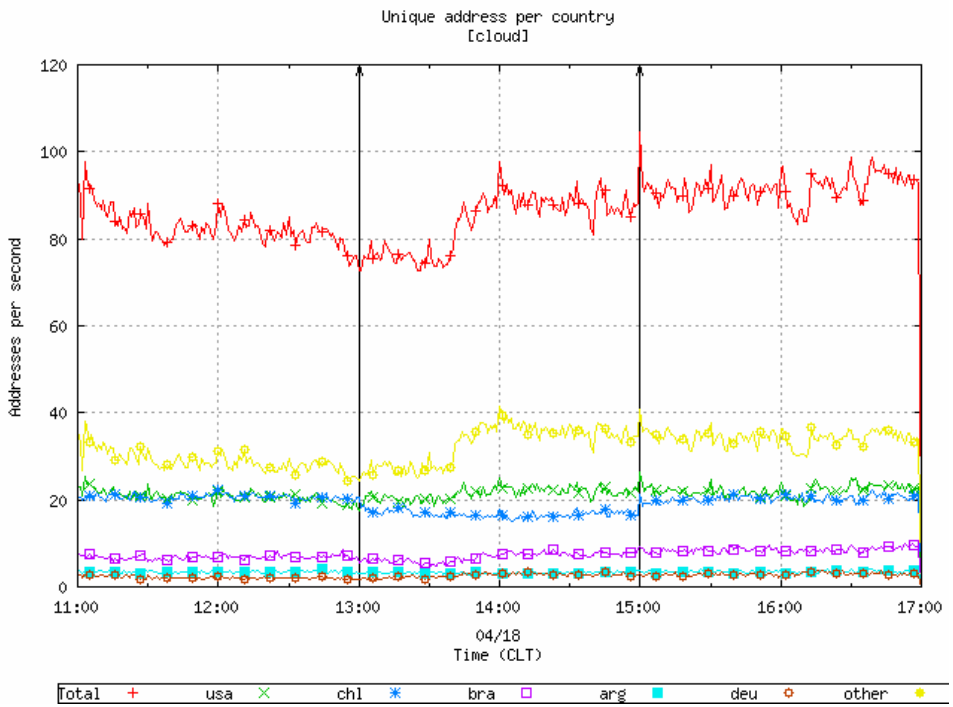
Queries per second	Number of unique sources	Query load [# of queries]	Percentage of total load	Queries per address
< 0.01	172 790	647 944	8.207	3.750
0.01 - 0.1	11 594	726 899	9.207	62.696
0.1 - 1	2 602	1 518 605	19.235	583.630
1 - 10	346	3 130 768	39.656	9 048.462
10 - 100	2	1 870 693	23.695	935 346.500
Total	187 334	7 894 909	100.000	**42.143**

Load redistribution



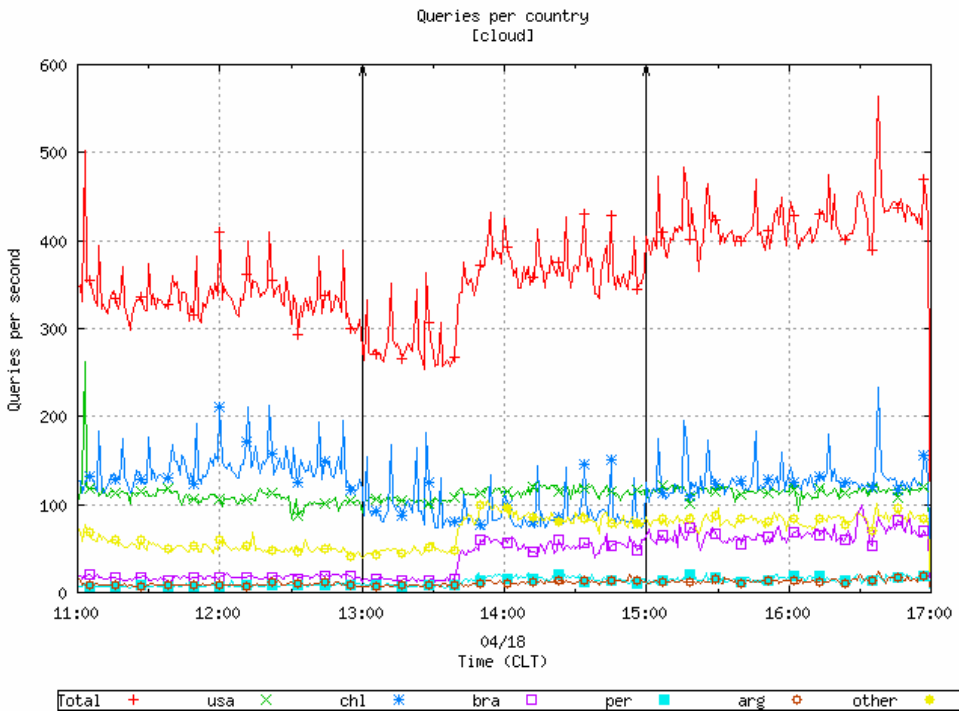
% of load	Before	After	Diff
santiago	25.11	0	-25.11
ns	46.57	52.56	+5.99
valparaiso	4.54	13.38	+8.84
tucapel	23.77	34.03	+10.26

Geographic characterization



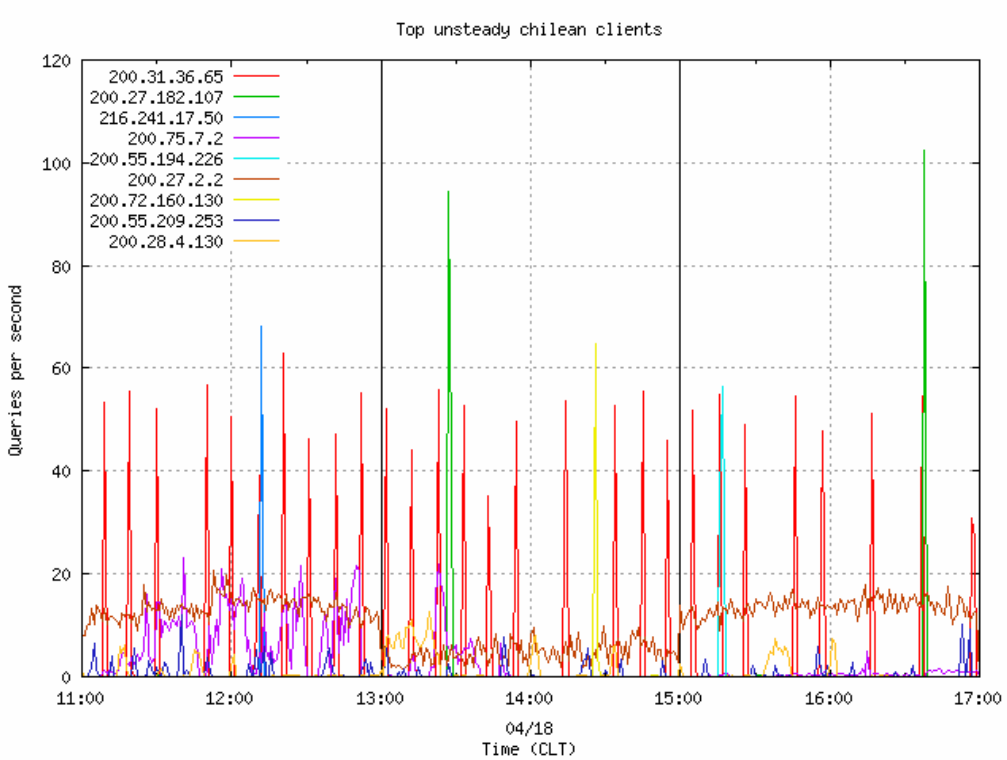
- The number of clients per minute aggregated by country is relatively steady
- The number of clients coming from USA to the anycast cloud is surprisingly high (considering there are other authoritative nameservers for .CL closer to those clients).

Geographic characterization



- In terms of queries per second aggregated by country, we could see the traffic coming from Chile is clearly “unstable” compared to any other country.
- Checking further on that, we found…

Spiky local load



- Nine clients were primarily responsible for the high variation.
- We plotted the clients with the highest standard deviation of the query rate.

Spiky local load

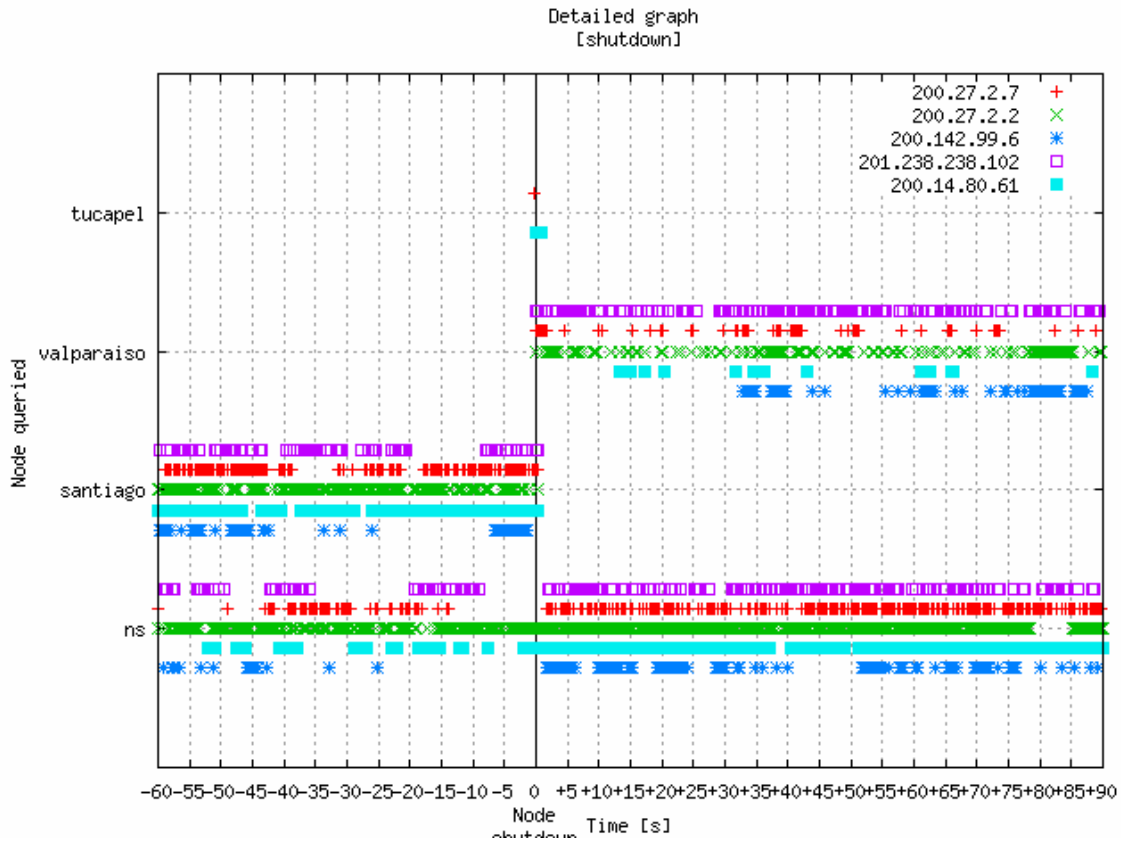
- One client (200.31.36.65), sending the same query in bursts of 60 queries per second, during 20 seconds each 10 minutes.
  - IN, A, EXCH_STGO.viconto.cl
- Four clients querying for the same pair of queries, probably due to a zone operator mistake.
  - IN, A, ns5.chileadmin.cl.imm.cl
  - IN, A, ns6.chileadmin.cl.imm.cl
- One client sending MX queries, iterating over an alphabetic list of names (SPAM activities perhaps?)

Client switching

- 7 894 909 total queries
- 187 334 unique source addresses
- 44 968 unique source addresses switched
- 72 390 total switches

	val -> san	val -> tuc	san -> val	san -> tuc	tuc -> val	tuc -> san
One way switch	245	251	540	8586	42	9752
Two way switch	1720	47	5	129	84	24501
Reverse two way switch	5	84	1 720	24 501	47	129
Total	1 971	382	2 265	33 217	173	34 382
Percentage of queries generated by the clients switching to:	0.952	1.654	4.059	22.269	0.816	2.660

Switching time



- This graph represents the queries sent by the top 5 most prolific sources. The graph is focused on the moment of the shutdown of the node.

Switching time

- Elapsed time: time between the shutdown and the last query seen on that node.
- Switching time: time between the last query seen on the shutdown node and the first query seen in any other member of the cloud from the same client.

IP Address	Elapsed Time [s]	Switching Time [s]
200.27.2.7	0.307	0.071
200.27.2.2	0.341	0.114
200.142.99.6	< 0	32.767
201.238.238.102	0.355	< 0
200.14.80.61	0.355	13.319 [0.039]

Conclusions

- On query load graphs and shutdown graph, there is some evidence of clients leaving the cloud and querying somewhere else. That load shifts from the anycast to the unicast server, probably selected by a lower RTT.
- The convergence time seen on this experiment is surprisingly short, if we think about the BGP updates convergence. The presence of the unicast server seems to be an added factor of stability.