Investigating anomalous DNS traffic

A proposal for a address reputation system

Sebastian Castro
sebastian@nzrs.net.nz

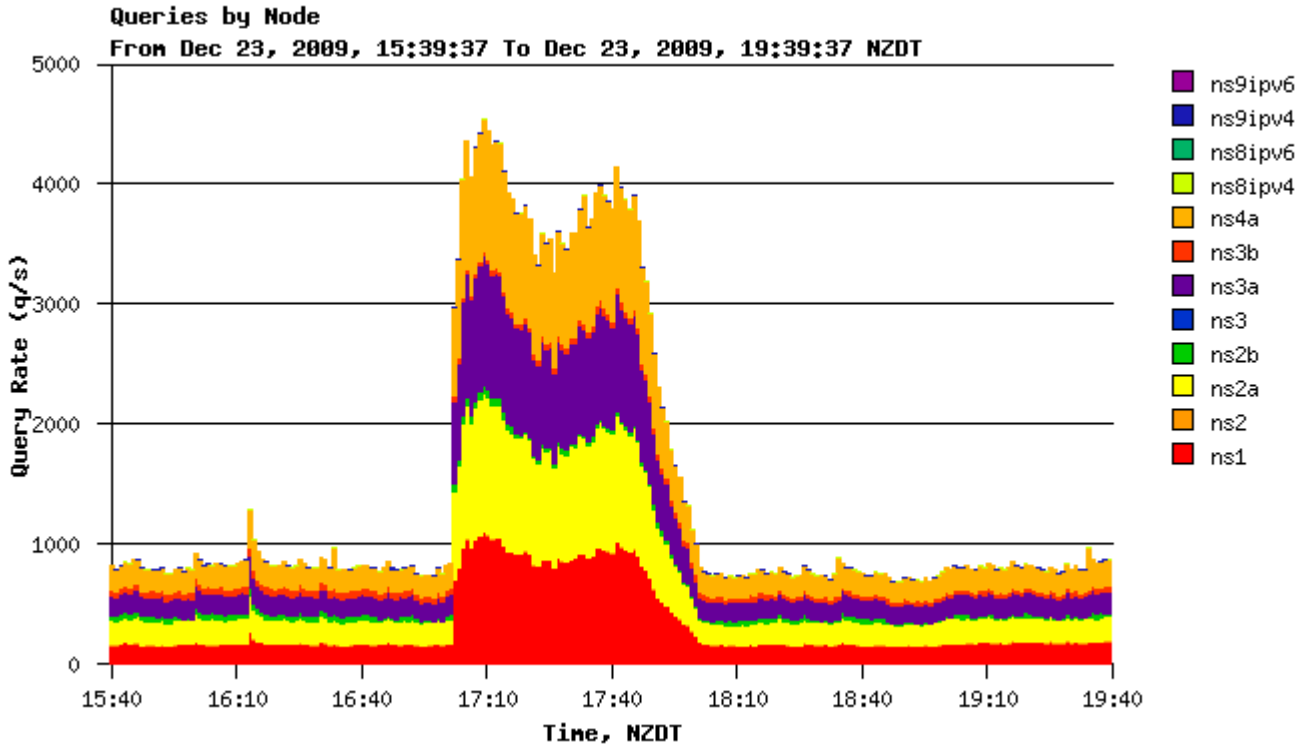
.nz Registry Services

Briefly about NZRS

- The New Zealand Registry Services
  - Handles registrations for the 14 SLD under .nz
  - 74 registrars
  - 7 nameservers
    - 1 anycast clouds provided by Autonomica
    - 2 anycast clouds provided by UltraDNS
    - 2 anycast clouds located in New Zealand
    - 2 unicast servers located in New Zealand

Motivation

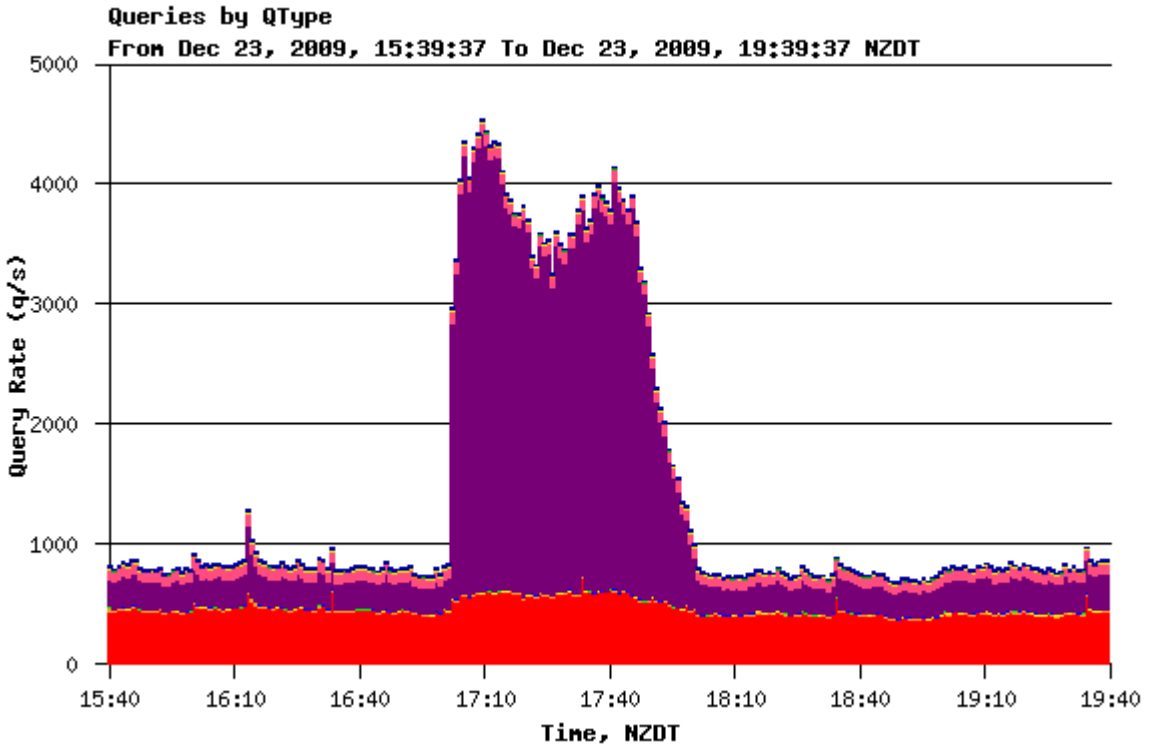
- From time to time we see peaks like this in the local nameservers



- Dec 23, from 17:00 to 18:00 NZDT
- Triggered alerts in the Nagios Monitoring

Motivation

- Where the type of traffic look like this



Breaking the usual distribution

- ➔ 58% A-queries
- ➔ 26% MX-queries
- ➔ 12% AAAA-queries
- ➔ 4% others

Motivation

- The pausible explanation: spammers checking for domain names
  - Usually in dictionary-based scans
- Other TLD operators have seen this
- No further investigation was done.... until now

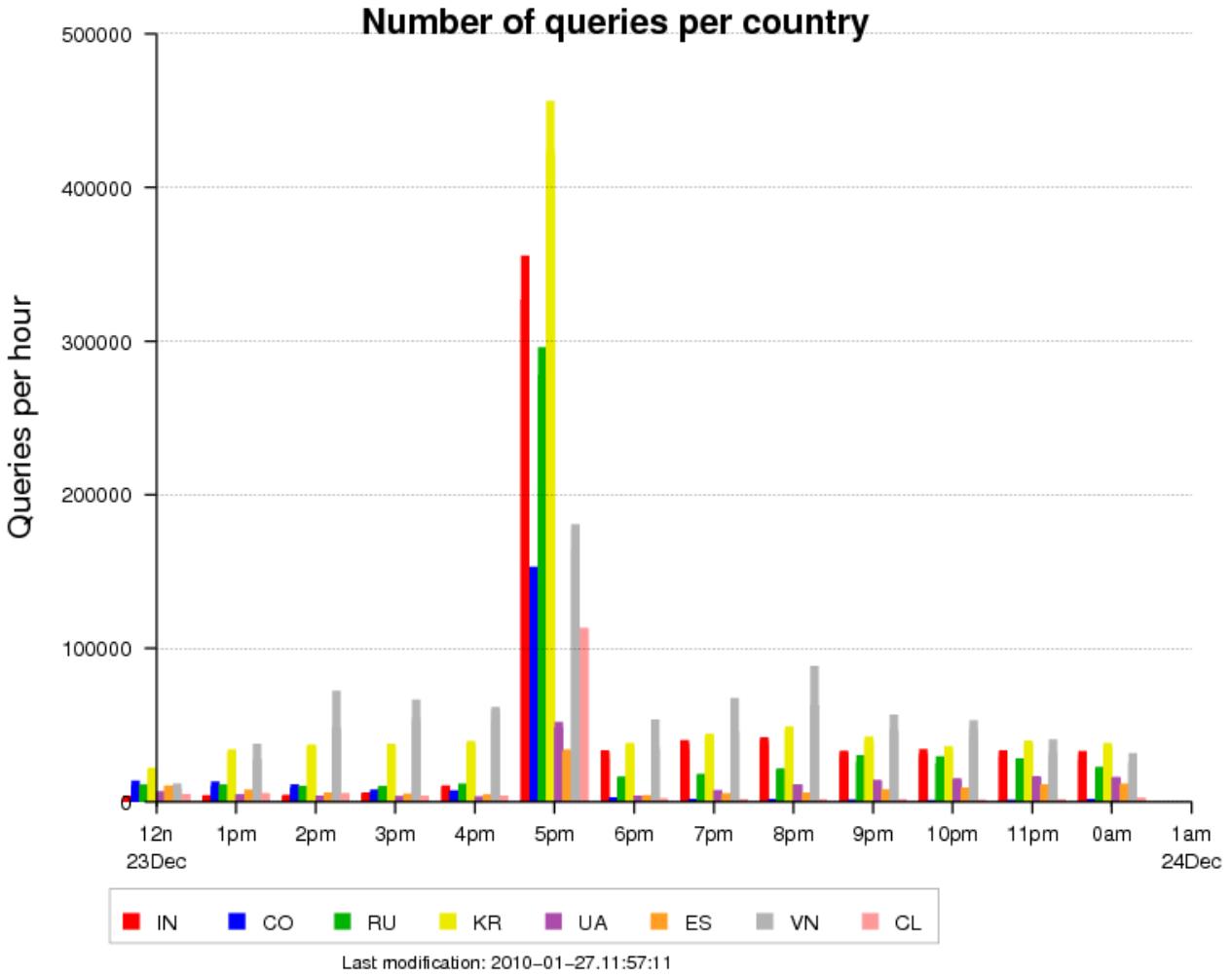
Data used

- Aggregated hourly files with
  - Destination node, source address, query type, number of queries
  - We don't keep the name of the query
    - Not possible to correlate to look for lexicographic sequences
  - We lose the time granularity

Investigation

- We analyzed the sources of traffic
  - By country using a GeoIP database
  - By origin AS using the BGP routing tables
    - We do this regularly since the last months to understand better placement for nameservers
- Selected AS/CC based on “normal” behavior
  - Calculated average and stddev for the month
  - Filtered the sources exceeding $dp > \bar{x} + 3 * \sigma$
  - Data represents the traffic observed in all NZ-based nameservers

Queries per country

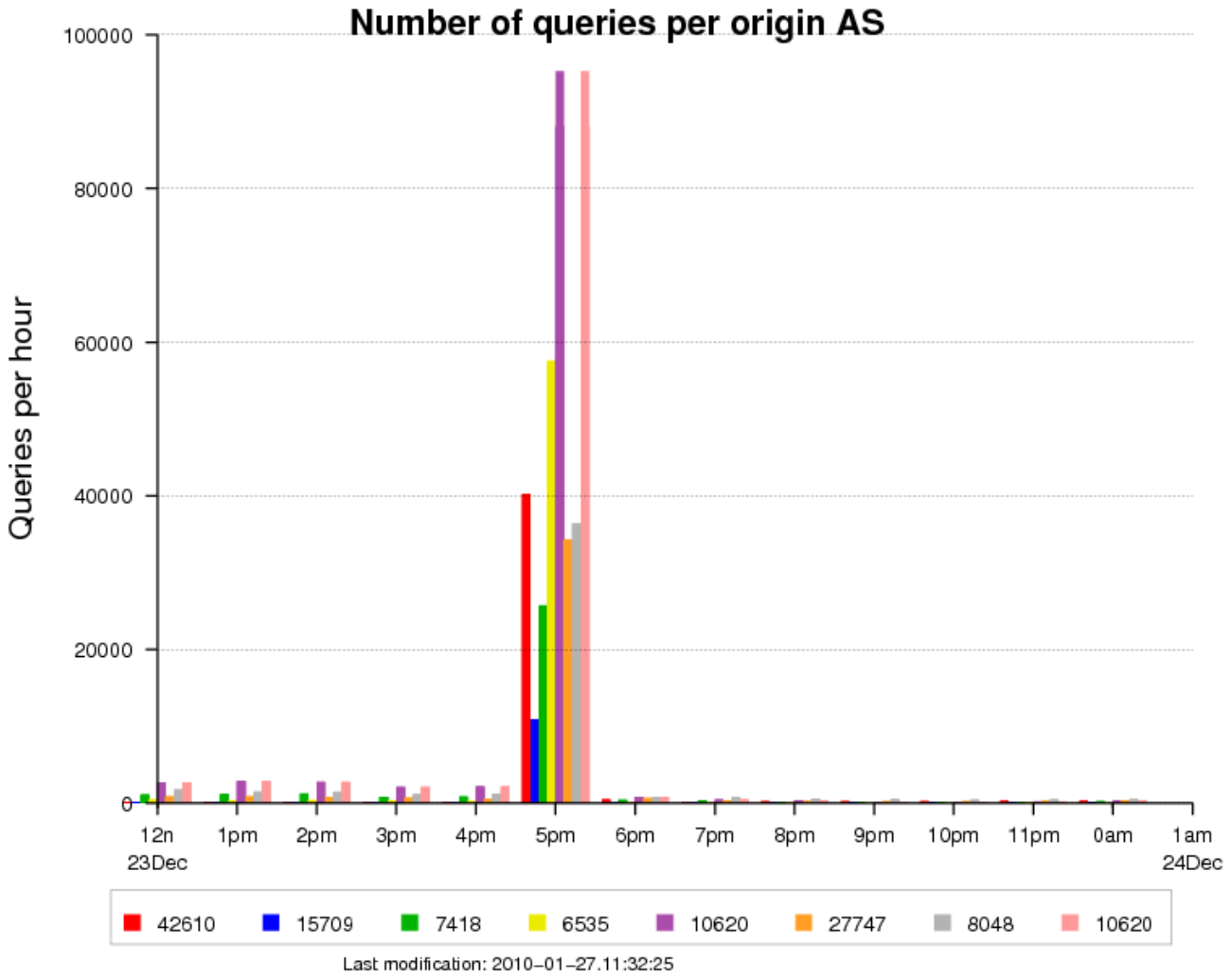


The figure shows the top countries with the higher difference between the mean queries and the selected data point

Two interesting points:
➔Countries from which hardly see any traffic (CO, ES, CL, UA)
➔Countries which are normal clients dramatically increasing their traffic (KR, IN, RU, VN)

Shows some coordination: 5pm NZDT is 1am CLST or 5am CET

Queries per origin AS



Different view using the origin AS

Countries for each AS
42610: Russia
15709: Germany
7418: Chile
6535: Chile
10620: Colombia
27747: Argentina
8048: Venezuela
10620: Colombia

Usual suspects?

- Can we discover is there are common sources behind this?
- Analyzed two extra events to correlate sources
  - Dec 17, 21:52 – 23:00 NZDT
  - Dec 25, 0:00 – 0:30 NZDT
- Anomalous sources
  - If query count during the event > avg+3*stddev, the source is considered “suspicious”
  - 10,000 sources matched the criteria in the three events
  - 171 sources were present in two events.

Making some “useful”

- We can analyze particular events
  - And create some knowledge on the source for long term analysis
  - But this approach is limited if acting alone
  - Others see this kind of event and even do some analysis by themselves
- Why don't we create a reputation system?

IP-based reputation systems

- Mainly used by mail servers
- One of the validation steps when a incoming SMTP connection is received
  - Check the source address against one or more DNS BlackLists
    - To verify if others have seen that source generate SPAM or massive e-mailing
  - Creates an address reputation system
    - According to some authors, helps to reduce SPAM by 80%

Reputation for DNS

- The DNS is a different service
  - Different transport (UDP can be spoofed)
  - Different way of working
    - Repeated queries can be considered annoying but not harmful
    - The origin of the queries are expected to be cache resolvers
- At CAIDA we tested correlations between misbehaved sources and spam
  - Found very little correlation, not relevant
  - We can't use the DNSBL to assign reputation to DNS clients

Proposal

- Establish a cooperative environment to share this kind of findings
  - Perhaps use DNS-OARC as a platform?
- Build a reputation score based on the events reported
  - X-type of queries over the chart
  - Constant hammering
  - Others
- At this point, not proposing a policy
  - Like block, throttle, delay a bad source

Open Questions

- How to build the reputation
- Are there privacy issues involved?
  - Original or anonymized address
- Structure language to report events
- Willingness of operators to cooperate