Mapping the Expansion of Google's Serving Infrastructure

by Matt Calder, Ethan Katz-Bassett

University of Southern California

as presented by Lennart Weller

Overview

- Background
- Goal of the paper
- Methodology
- Validation
- Mapping Google's Expansion

Background

CDNs and Serving Infrastructure

- CDNs are now used by every major platform on the web
- Serving infrastructure is ideally situated close to clients
- All to provide excellent performance for clients

DNS-based redirection

- Using DNS records to tell the client the closest front-end
- Supoptimal if resolver is a caching NS
- Extendend DNS as solution for the issue

Goal

- Understand large serving infrastructures like Googles
- Measuring the evolution of the serving infrastructure
- Enumerate front-ends and geolocate the servers
    - Using information provided via EDNS

Methodology

Enumerating the Front-Ends

- Enumerate all addresses
- No distinction between infrastructure types
- Modified EDNS requests are used to receive front-ends
  - Client-subnet extension
- Client prefixes from 10 million /24-Subnets

Methodology

Client-centric Front-End Geolocation

- Geolocation databases deliver poor performance
- Improve using CCG (client-centric geolocation) and CBG (constrait-based geolocation)
- Obtain common prefixes for EDNS via Ono (BitTorrent)
- Load Balancing is detected and excluded

Methodology

Clustering Front-Ends

- City-wide serving infrastructures are clustered into data centers
- Tests are done from a known vantage point using RTTs
- Clustering is done with Manhattan distance using OPTICS

Validation

Coverage of Front-End Enumeration

- Validate EDNS results using 200,000 rDNS resolvers
- Adds another 15-20% more Front-Ends
- Resolvers from all previously selected subnets
- Daily HEAD-Requests to check liveness of Front-Ends

Validation

Accuracy of Client-Centric Geolocation

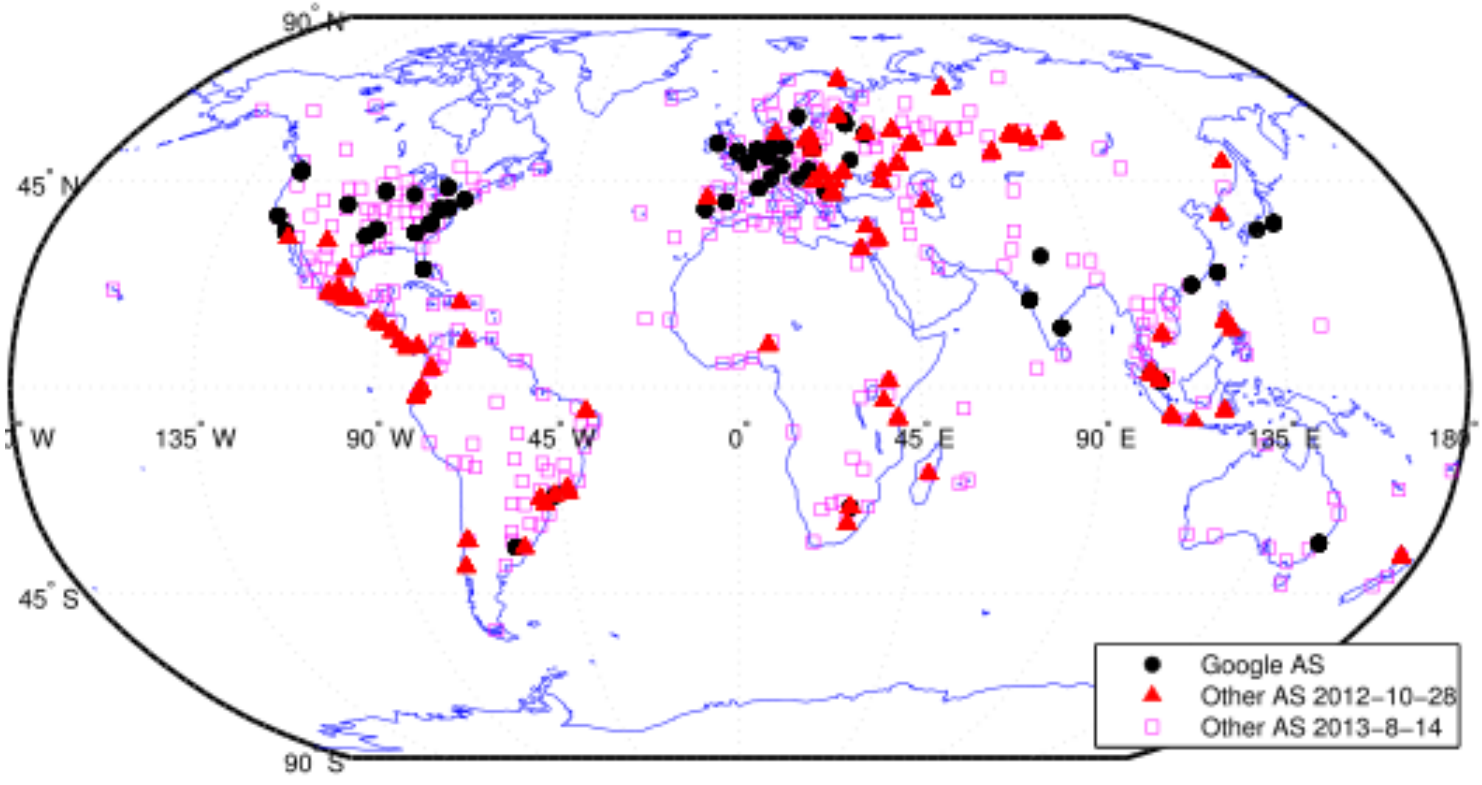
- Front-Ends with Airport codes are verified
- CCG and CBG results are compared with databases
- Filtering results by using multiple methods reduced distance
- Median distance of 305km and errors less than 1000km of 86%

Validation

Accuracy of Front-End Clustering

- Previous results are compared with known clustering points
- Rand Index is used to measure accuracy
- 97% accurate clusters over all clusters
- Some misses with airport codes while clustering

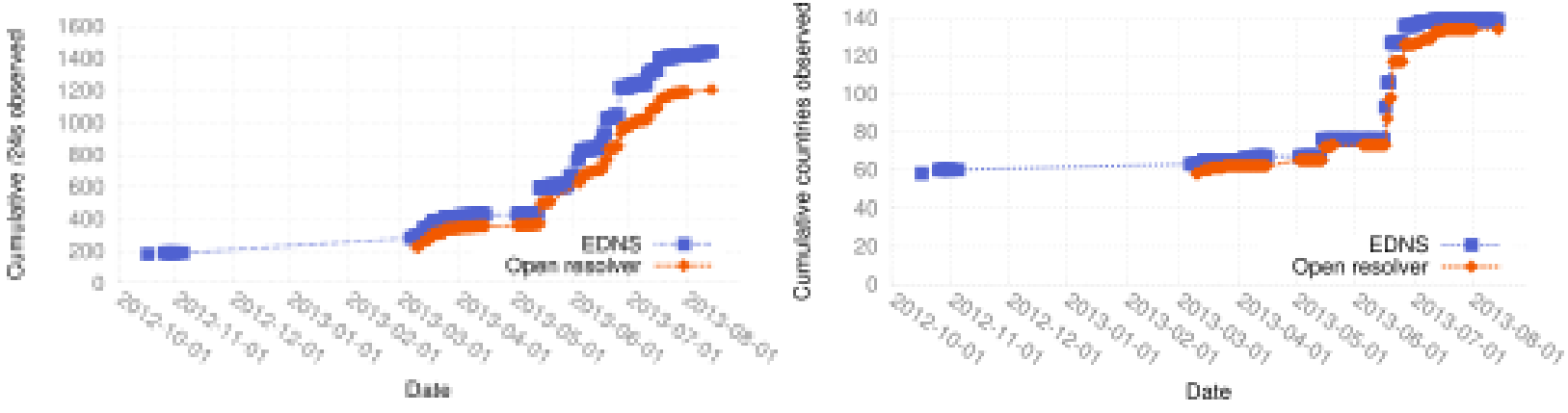
Mapping Google's Expansion



The expansion of the public Google Infrastructure

Mapping Google's Expansion

Growth over time



Number of ASes & overall presence in countries

- A large increase in server infrastructure was observed
- Sevenfold increase in reachable addresses and ASes
- New countries were served from inside their borders
    - Mostly ASes other than Google

Mapping Google's Expansion

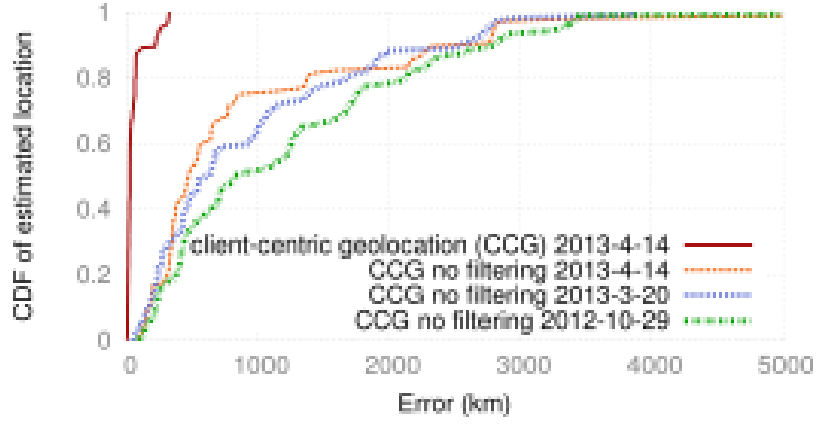
Characterizing the Expansion

- Google started using external infrastructure
- They are re-using Servers installed as YouTube video caches
- Local ISPs serving Google content to customers

Mapping Google's Expansion

Impact on Geolocation Accuracy

- Thanks to the new front-ends overall geolocation is better
- Using the filters a median error of only 22km is measured



Overall the distance to to clients was reduced over time

Questions & Answers