Improving Content Delivery Using Provider-aided Distance Information

Authors: Ingmar Poese, Benjamin Frank, Bernhard Ager, Georgios Smaragdakis, Anja Feldman

Presenter: John Rula

Overview

- Show more than half of total traffic can be delivered from multiple servers at diverse network locations
  - Propose PaDIS as a content location recommendation system within an ISP
  - Quantify Content Delivery performance improvement of PaDIS

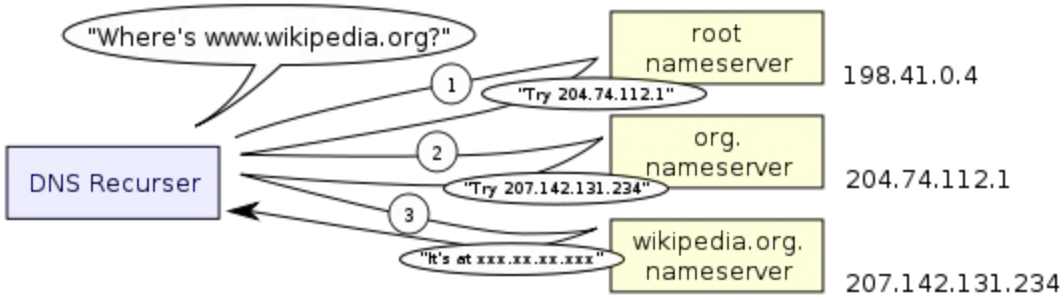
Content Delivery Networks

- Also known as CDNs deliver content through a worldwide network of ‘edge servers’
- Host content from content providers and serve it to clients from the “best” possible server (closest or lowest latency).
- Most popular use **DNS Redirection** to find closest server.

What is DNS

- Doman Name System
  - A distributed database of mappings from domain names to IP addresses
  - There are different levels of DNS Servers
    - Authoritative name servers
      - Cache servers
      - Subdomains can be run by different Name Servers
  - A Query is called by a client and is recursively resolved

DNS Query

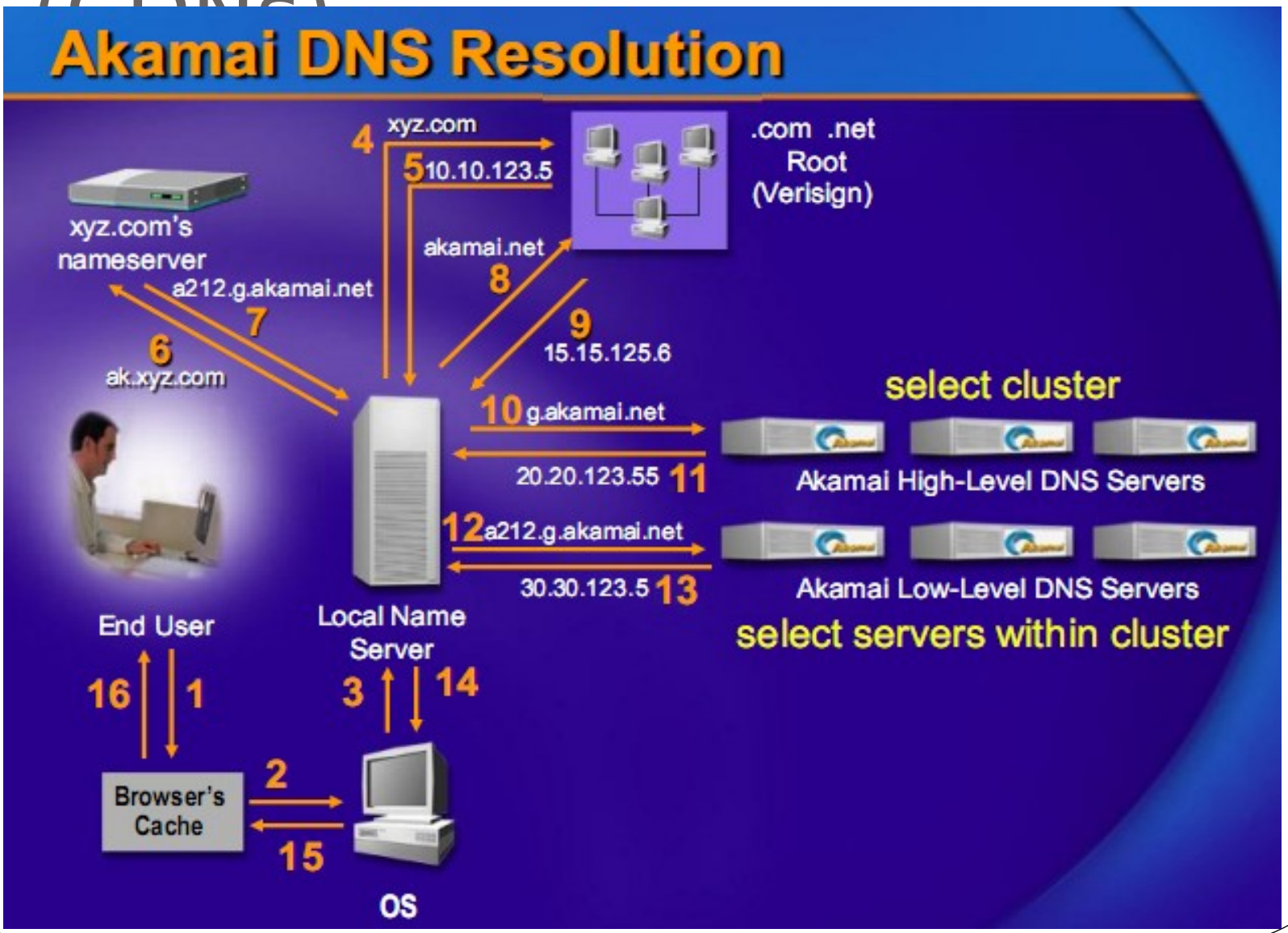


- A Recursive DNS Query
- Query is first sent to root nameserver
- Then is recursively sent down the hierarchy of name servers until it is resolved

CDNs and DNS

- CDNs utilize this recursive structure to first encode information into a DNS query
  - A CDN DNS query is resolved to the edge server which has been determined by the CDN to be the “best” server (for the initially requesting DNS nameserver)
    - Typically the ISP’s DNS Server
  - DNS queries are cached across the network, and as such, the TTL (Time-To-Live) value is used (abused) to make sure the queries are not cached long.

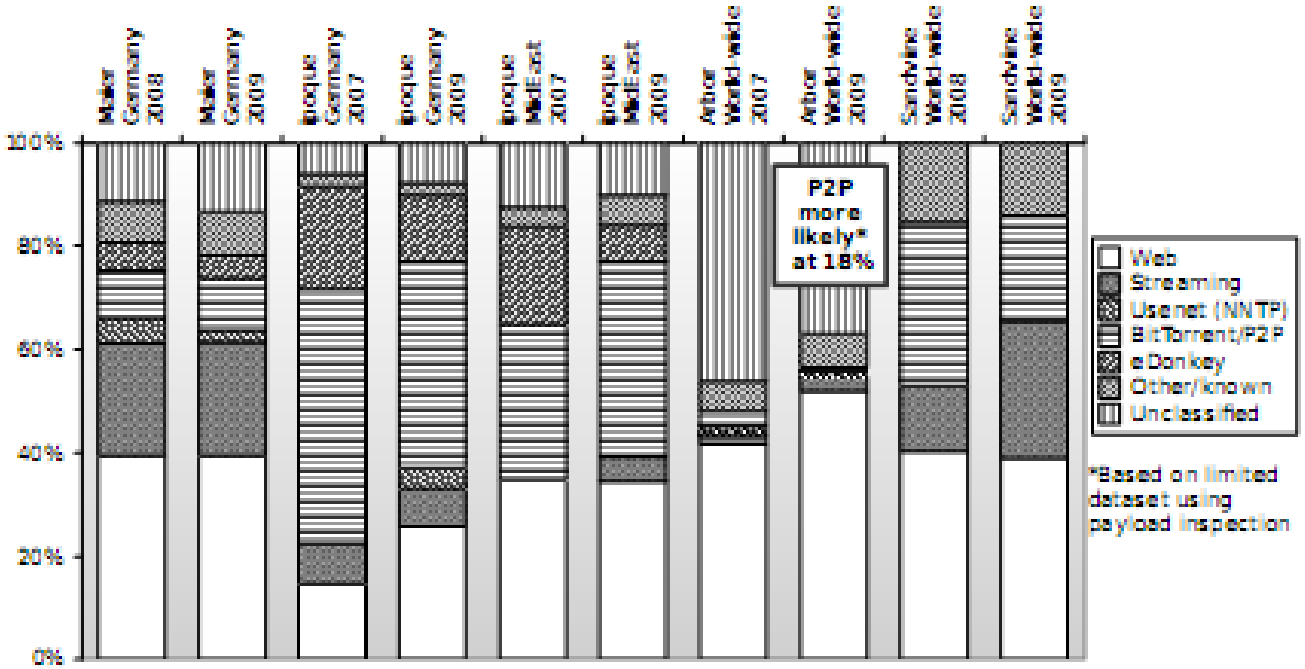
Content Delivery Networks (CDNs)



Dataset

Table 1: Summaries of anonymized traces.

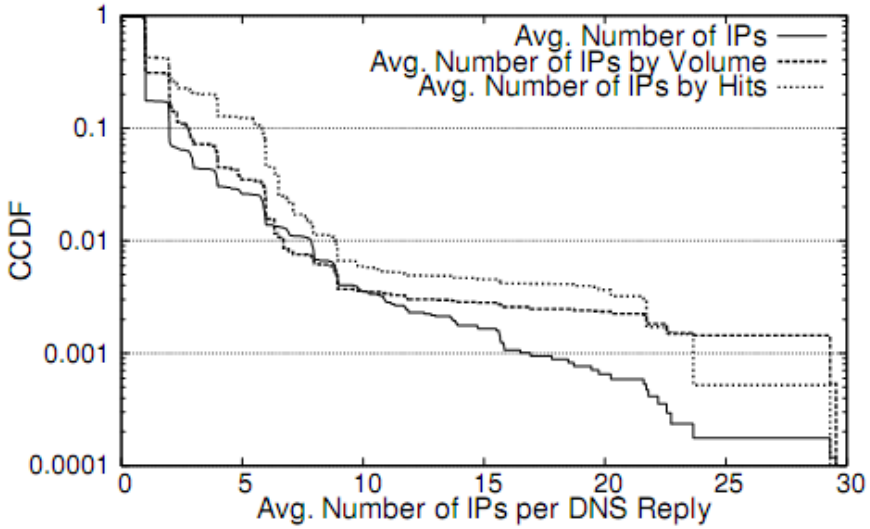
Name	Type	Start date	Dur	Size	Application Volume
MAR10	packet	Thu 04 Mar'10 2am	24 h	>5 TB	> 3 TB HTTP, > 5 GB DNS
HTTP-14d	log file	Wed 09 Sep'09 3am	14 d	> 200 GB	corresponds to > 40 TB HTTP
DNS-5d	packet	Wed 24 Feb'10 4pm	5 d	>25 GB	> 25 GB DNS

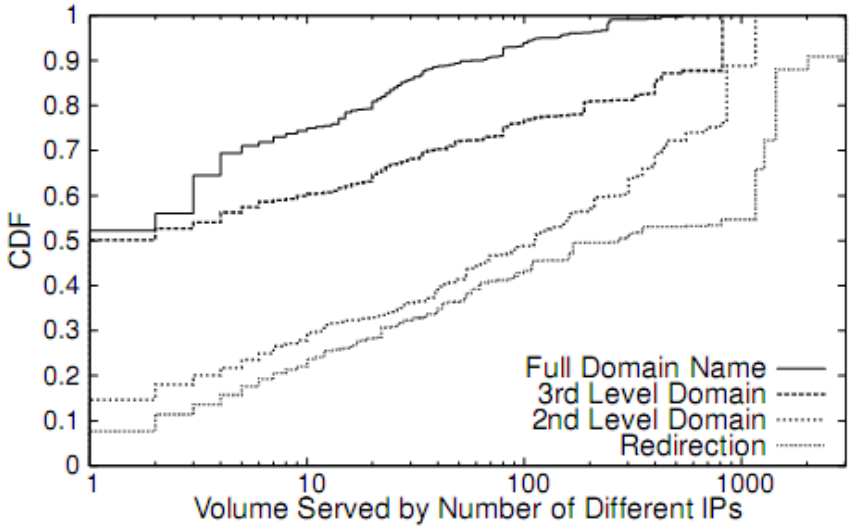


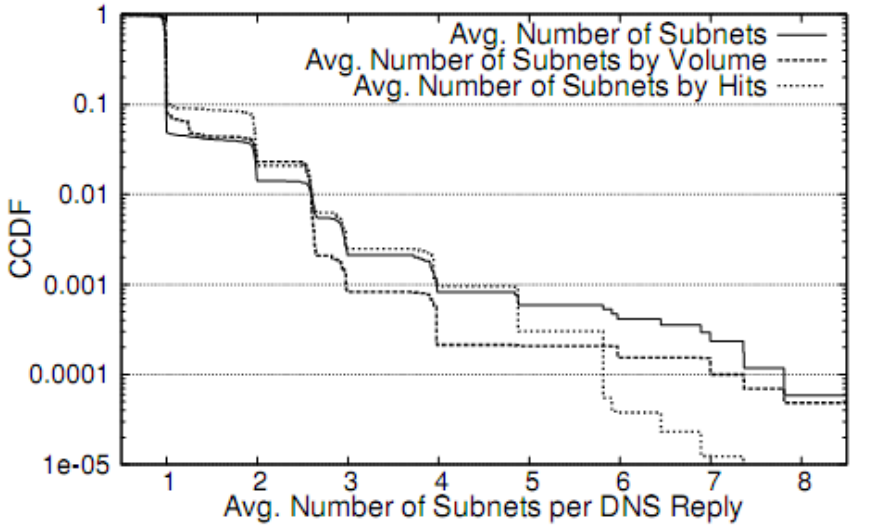
DNS Analysis

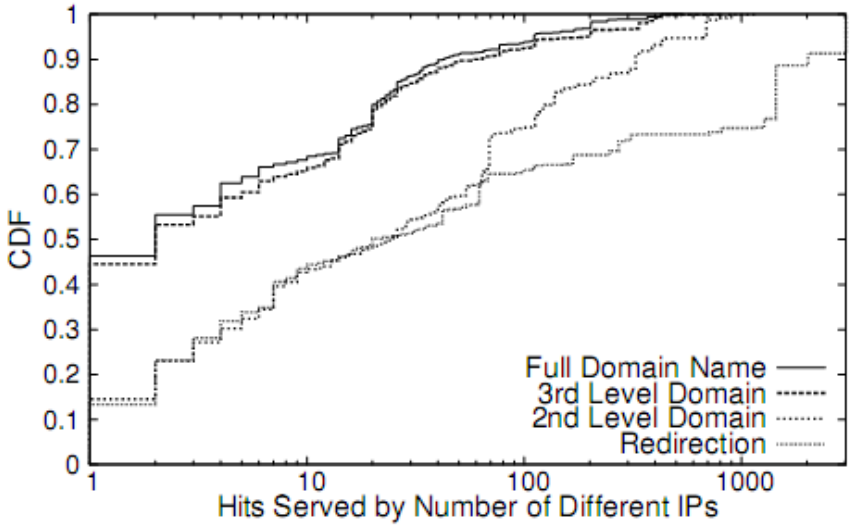
- MultiQuery - Returns multiple IP addresses witihin a single DNS response
- CrossQuery – Can return different IP addresses for repeared queries and this perform DNS redirection
- Internal DNS Resolver is ISP’s, External is something like OpenDNS or GoogleDNS.
  - Found that 95% of clients use ISP’s DNS Resolver as default
    - Use of external DNS decreases performance of CDN delivery

DNS Analysis cont...









Impact of Traffic Localization

- Labeling hostnames
  - Using external DNS amounts to 12% of traffic within same AS
  - Using ISP DNS amount to 25% of traffic in AS
  - Potential is if any IP found within same AS is used all the time

	ISP DNS		OpenDNS		GoogleDNS	
Metric	observed	potential	observed	potential	observed	potential
IPs	12.3 %	24.2 %	5.8 %	16.0 %	6.0 %	9.7 %
requests	14.9 %	33.2 %	4.7 %	18.8 %	4.8 %	6.4 %
volume	23.4 %	50.0 %	12.0 %	27.7 %	12.3 %	13.4 %

What is PaDIS?

- Provider-aided Distance Information System
- Can be used by client, CDN, and ISP
- Acts as a middle layer between an ISP and a Content Delivery Network.
- Allows ISP to optimize CDN traffic for their benefit (cost).

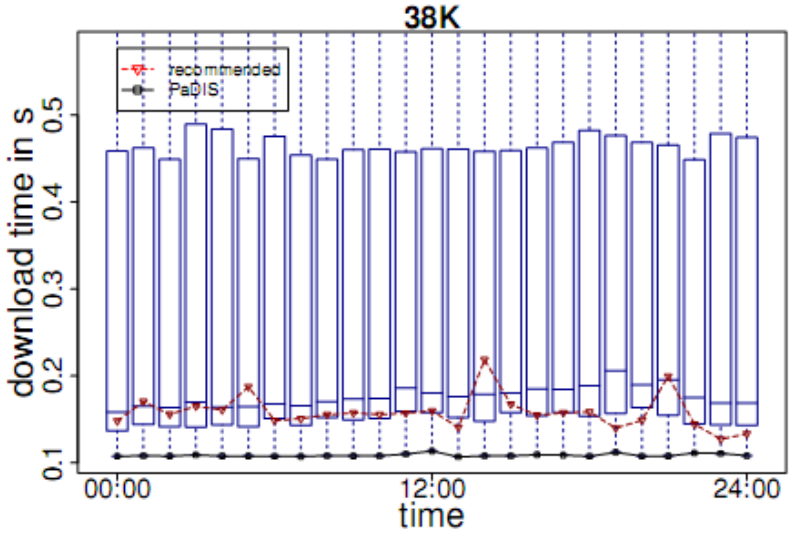
How does PaDIS work?

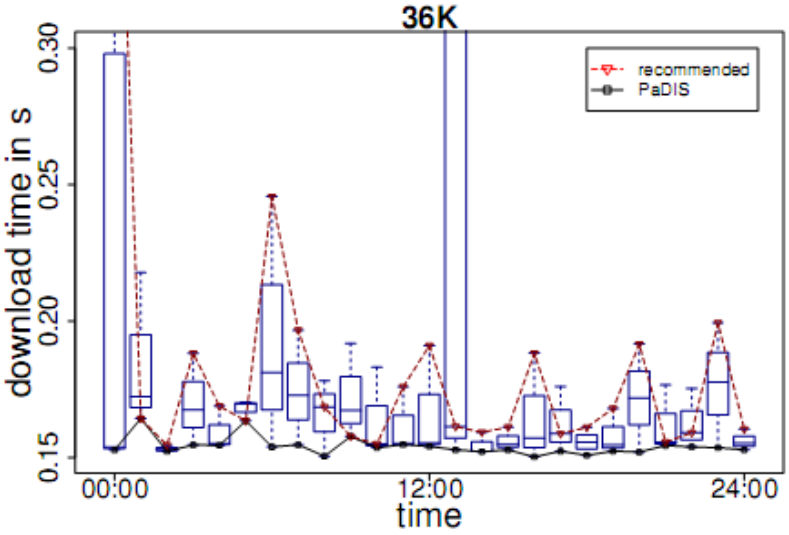
- Consists of a Data Management System and a Query Processing System
- Data Management System
  - Information Retrival
    - Network Map Generator
- Query Processing System
  - Request Translator
  - Query Manager
  - Path Ranker
  - Frequent hitter Detector

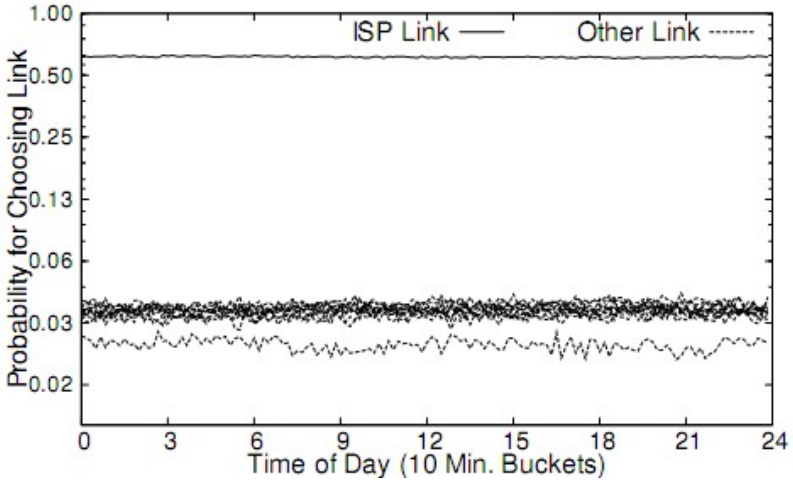
PaDIS Use Cases

- Clients
  - Can have all DNS replies sent to PaDIS server for re-ranking and returning IP address of ISP's preference
- CDNs/CDPs
  - Collaborate with ISP through PaDIS by contacting them before returning server choices to the DNS resolver
- ISP
  - Enhance DNS Resolver to contact its PaDIS server and reorder the IP addresses if needed before returning any answer to the client

PaDIS Evaluation







Problems with PaDIS

- Potential for abuse by “bad” ISPs.
  - Unable to correctly implement the low TTL times used by CDNs.

Questions.....?