The Problem

- Potentially Lots of Internet Traffic
- ISP DNS Cache Good
- Satellite Internet ~550ms Latency
- Local DNS Cache Limited or Nonexistent
- Increased Page Load Times on High-Latency Networks
- Potentially Lots of Internet Traffic

The Problem
A Solution - Enter Cloud Computing Technologies

- Apache Kafka
- Spark Streaming
- Machine Learning
- Push-Based Local DNS Cache
Algorithms/Approaches

- Similarity Analysis - Starting Points
  - Streaming K-Means
  - Locality Sensitive Hashes

- Metrics to Examine
  - Query Source
  - Query Domain
  - Query Time

- Similarity Analyses - Starting Points

- Streaming K-Means
Related Work

- Accelerating Last-Mile Web Performance with Popularity-Based Prefetching (Sundaresan, Magharei, Feamster, Teixeira)
- Behavior-based tracking: Exploiting characteristic patterns in DNS traffic (Herrmann, Banse, Federerath)
- Streams and Funnels (Raghav, AuYoung, Smith)
- Many Others - Focused Primarily on Spam Detection, Security, Tuning

Streams and Funnels (Raghav, AuYoung, Smith)
Data Sets

- Toy Dataset: personal BIND servers(s) pushing logs to Kafka

- Enterprise Servers
  - Anonymized Datasets: 24 hours of DNS Requests from Medium-Sized
    - No anonymity
    - unlikely to yield useful results
  - Traffic Simulator
  - Web Crawler

- Possibly Others
  - Web Crawler
  - Traffic Simulator

- Anonymized Datasets: personal BIND server(s) pushing logs to Kafka
Evaluation

● Testing Accuracy
  ○ Streaming Hit Rate

● Testing Performance
  ○ Collect metrics on throughput, latency
  ○ Test in different environments
  ▪ Azure Cluster
  ▪ AWS Cluster
  ▪ Local desktop
  ▪ Collector metrics on throughput, latency

Comparison to Caches?
  ○ Could do in-place analysis of data to analyze hit rates assuming LRU sizes of different sizes
  ▪ # of Queries Answered Since This Domain Seen Last

Testing Accuracy
  ○ On-Line Calculation of Predictive Accuracy
  ▪ Streaming Hit Rate

Testing Accuracy
  ○ Different Starting Points
  ▪ Local desktop
  ▪ AWS Cluster
  ▪ Azure Cluster


Citations

Latency - Why Is It a Big Deal for Satellite Internet?


Citations


