More Control Over Network Resources: An ISP Caching Perspective

D. Tuncer, M. Charalambides, R. Landa and G. Pavlou
Department of Electronic & Electrical Engineering
University College London, UK

CNSM
14-18 October 2013
Motivation

- Content Delivery Networks (CDNs) have been the prevalent method for efficient content delivery across the Internet.

- Internet Service Provider (ISP) network resources severely affected by current content delivery services.
  - CDNs perform management operations with limited information about ISP networks.

- Target: provide more control to the ISP for a win-win situation.
Proposed Approach

• ISP-operated content distribution service of limited capacity.

Request for \( x_1 \)

Cache Manager

Cache

1

2

3

4

5

6

7

8

9

10

Request for \( x_1 \) redirected to node 4

Request for \( x_1 \) served locally

ISP_2

Inter-domain link

CDN

• The ISP can implement its own cache management strategy.
Content Distribution Management (1/3)

• Limited total caching capacity in the network.
  ➢ Only a subset of content items can be stored in the network
  ➢ Information about content items should be available to the ISP

• Limited caching capacity in each caching location.
  ➢ Content items distributed across the different caching locations
  ➢ Content placement can affect resource utilization

• The use of appropriate management approaches can allow a better control of network resources utilization.
Content Distribution Management (2/3)

• Content Placement Problem

Given a set of $M$ caches and a set of $X$ contents, determine

- the number of copies of each content item to store in the network
- the location of each copy

so that network resources can be better utilized.

• Simple and lightweight decentralized strategies to compute new content placement
Content Distribution Management (3/3)

- Decisions taken by cache managers that coordinate among themselves through a *management substrate* [1][2].
  - Logical infrastructure used to facilitate communication between decision-making points

- Reconfiguration decisions taken at a timescale of the order of hours.


Influence of User Demand Characteristics

• The performance of a content placement strategy can be affected by several parameters.
  ➢ Content size, cache size, cache location etc.

• Investigate the influence of user demand characteristics
  ➢ Content popularity
  ➢ Geographic distribution of interests

• Modelling choices and assumptions
  ➢ Uniform cache size and content size
  ➢ Total volume of contents to cache less than total caching capacity
User Demand Model

• In the absence of traces…

• How to characterize the interests for the contents?
  - Global popularity of the contents usually modelled according to a Zipf distribution of parameter $\alpha$.
  - How to model the geographic distribution of interests ($GDI$)?
    - Uniform distribution commonly assumed but not realistic
Modelling the GDI (1/2)

Very popular contents are requested everywhere

Less popular contents are requested from specific locations

**UNLIKELY**

Most Popular Content

Less Popular Content

Geographic heterogeneity

Popularity
Modelling the GDI (2/2)

\[ f_\beta (r) = 1 + (M - 1) \left( \frac{(X - r)}{(X - 1)} \right)^\beta \]

- $M$ the total number of caching locations
- $X$ the total number of contents to cache
- $r$ the global popularity rank of the content
Placement Strategies

• Two categories of placement strategies
  - Local popularity driven strategy
  - Global popularity driven strategy

• Placement constraints
  - There is at least one copy of each content cached in the network.
  - A content \( r \) is copied in a cache only if there is enough caching capacity to accommodate \( r \).
  - A content \( r \) is copied in a cache only if \( r \) is not already cached at this location.
Local Popularity driven Strategy

- Each cache manager maintains a list of locally requested contents ordered by decreasing local popularity.

- Two-phase process
  1. Selection of the content items to cache locally as follows:
     - They have the highest local popularity.
     - They are not already cached somewhere else in the network.
     - There is enough caching space.
  2. Fill the potential remaining caching space to replicate the locally more requested contents not already cached.
Global Popularity driven Strategy

• Each cache manager has global information about:
  - Content global popularity.
  - Geographic distribution of interests for each content.

• Two-phase process
  1. Selection of the content items to cache locally as follows:
     - Items for which there are locally the highest aggregated number of requests.
     - There is enough caching space.
  2. Fill the potential remaining caching space by replicating contents in the $n$ caches with the higher aggregated number of requests.
Performance Evaluation Objective

• Is it possible to identify, for a given user demand profile, a strategy that always outperforms the others?
  
  ➢ The best strategy is the one that results in the minimum maximum utilization (min max-u) in the network.

• Four strategies are considered.
  
  ➢ Local-popularity driven Strategy (LPS)
  ➢ Global-popularity driven strategy with parameter n = 1 (GPS_1)
  ➢ Global-popularity driven strategy with parameter n = 2 (GPS_2)
  ➢ Global-popularity driven strategy with parameter n = 3 (GPS_3)
Evaluation Settings

• We use the Abilene network, 12 nodes/caches, \( \approx 100 \) of contents.

• We consider 20 values of \( \alpha \in [0.1; 2.0] \) and 20 values of \( \beta \in [0.2; 10.0] \), i.e. a total of 400 pairs of \((\alpha, \beta)\).

• For each \((\alpha, \beta)\) pair, we consider 100 realisations, i.e. we randomly select different locations from where each content is requested.

• We work with a total of 40,000 samples which cover a wide range of profiles.
Metrics $H_P$ and $H_G$ (1/4)

• How to retrieve the values of $\alpha$ and $\beta$ in practice?

• We define two metrics to represent the popularity distribution and geographic distribution.
  
  ➢ Metric $H_P$: heterogeneity in terms of popularity
  
  ➢ Metric $H_G$: geographic heterogeneity of the interests
Metrics \( H_P \) and \( H_G \) (2/4)

- Metric \( H_P \): heterogeneity in terms of popularity

\[
H_P = \sqrt{\frac{1}{X} \sum_r \left( \sum_m d_{mr} - \mathbb{E} \left( \sum_m d_{mr} \right) \right)^2} / \sum_r \sum_m d_{mr}
\]

- Metric \( H_G \): geographic heterogeneity of the interests

\[
H_G = \frac{1}{X} \sum_r \bar{h}_g (r)
\]

with

\[
\bar{h}_g (r) = \sqrt{\frac{1}{M} \sum_m \left( d_{mr} - \mathbb{E}_m (d_{mr}) \right)^2} / \sum_m d_{mr}
\]
Metrics $H_p$ and $H_g$ (3/4)
Metrics $H_P$ and $H_G$ (4/4)

- Correlation between $\alpha$ and $H_P$, and between $\beta$ and $H_G$.
  - The value of $H_P$ increases with the heterogeneity of the global popularity distribution.
  - The value of $H_G$ increases with the heterogeneity of the geographic distribution of interests.

- However, the formulae used to determine $H_P$ and $H_G$ are independent of $\alpha$ and $\beta$.
  - Can be used to analyze the popularity characteristics in any content distribution scenario.
  - Can be directly applied to the monitored data.
Performance Comparison (1/3)

• What is the best strategy for each sample?
Performance Comparison (2/3)

• Average deviation from the minimum max-u
Contributions and Future Work

• We propose an approach to empower an ISP with caching capabilities.

• We demonstrate how simple and lightweight content placement strategies can lead to better control of network resources.

• We propose a methodology to evaluate the influence of user demand characteristics.

• Future work
  
  ➢ Investigate the influence of other parameters (cache size, content size etc…).
  
  ➢ Investigate cooperation models between ISP and CDN.