Vertical Caching: Web Caching for Challenged Networks

Jay Chen (jchen@cs.nyu.edu)  Lakshmi Subramanian (lakshmi@cs.nyu.edu)

Problem: A single cache miss during browsing may cause a stall on the order of minutes or hours!

Do we need both pages?

Solution Idea: The same or similar content is aliased by different URLs.
We propose a caching methodology called “vertical caching” that extends existing caching mechanisms based on URLs to aggregates of cached pages across topics.

What is a Topic?
A topic is a set of pages that represent some particular interest of users. This definition is broad to encourage different methods of defining types of topics. We currently have three types of topics: domain, search, and content topics. Each page is associated with a “cost”, we define a novel cost function based on the time taken to download or access the page.

1. Domain Topics
2. Search Topics
3. Content Topics

How are Topics Useful?
Topics are aggregates that are used when deciding when to evict pages from the cache. e.g. pages from LRU topics are evicted first, and within each topic the LRU page is evicted first. We use our cost function to evict least used/useful pages first. Similarly, in the case of prefetching, topics are aggregates that are used to decide which pages to prefetch and how to find them and the cost is used to prioritize between topics.

Preliminary Evaluation
Using log data from an intermittently connected secondary school classroom in Kenya, we show that we can extract useful topics that persist over time. Our log contains over 100,000 requests and over 1,400 search queries gathered over a period of four months.

Do Topics Exist?

Do Topics Persist?