Khameleon: a new prefetching architecture for highly interactive exploration of massive image datasets

Tracy Wei  
Cornell University

Haneen Mohammed  
Columbia University

Eugene Wu  
Columbia University

Background

- Interactive apps are increasingly deployed in the cloud
- Cloud provides tremendous savings, but hurts interactivity
- Current approach: prefetch what app thinks user will do

Problems

- Although prefetching can reduce response times, its efficiency not only depends on the prediction model, but also the system resources.

Proposed Solution

Insight: Many applications are approximate tolerant. It is better to show a partial result than wait a long time for a perfect result.

Scheduler

The scheduler uses the estimated probabilities of possible requests at different points in the future to allocate fractions of network bandwidth to requests.

Each widget in (A) translates mouse interactions within a region of the screen into modified SQL queries that are executed by the server database. As the cursor in (A) moves along the dotted arrow, (B, C, D) shows how the distributions are updated for the possible 7 widgets in (A) at t=(0, 50, 150) ms respectively.

Research Prototype

Increasingly, High-Resolution Image datasets are generated in many domains. Exploring these collections interactively in a cloud-based environment is not a trivial task. Khameleon enables such applications. In this demo, response data is progressively encoded and modeled as a sequence of data blocks (1) as the user navigates view 1, the predictor continuously sends to the server distribution of user's future requests (2) the server actively schedules (3) and sends partial results to the client.