Big Data Initiative: Effective Caching in Online Video Platforms

Rongrong Bao Atabak Hafeez Tom Wiesing
Jinbo Zhang

September 29, 2015

1 Background

2 Motivation

Caching is very important when it comes to streaming videos online. According to an article [1](1), in 2007, 50% traffic came from several thousand sites and by 2009, 50% traffic came from 150 sites. Furthermore, by 2013, the 50% of all internet traffic came form 35 or more sites. From this we can see that the traffic of websites that are popular is being aggravated. A lot of this increase in traffic to a smaller number of websites is due to online video streaming.

Before 2012, video streaming was done using using Real Time Messaging Protocol (RTMP)-based servers. This requires a streaming server and a near-continuous connection between the server and player. Requiring a streaming server can increase implementation cost, while RTMP-based packets can be blocked by firewalls. In 2012, this was replaced by an (Hypertext Transfer Protocol) HTTP-based servers known as MPEG DASH (Dynamic Adaptive Streaming over HTTP). By using this technology, the servers were able to use HTTP-Caching. This latter capability should both decrease total bandwidth costs associated with delivering the video, since more data can be served from web-based caches rather than the origin server, and improve quality of service, since cached data is generally closer to the viewer and more easily retrievable.

3 Big Data Initiative: Research Questions

1. Which techniques have been implemented in YouTube, in order to reduce the bandwidth consumption?
   – keywords: Distributed Caching, Standalone Caching,

2. Are there any optimizations have been used for the popular videos and non-popular videos? if yes, which one?
   – keywords: Content Delivery Network,

3. Which buffering techniques are effective from YouTube side, buffering all the time, or buffering just a little bit ahead of current position?
   – keywords: Dynamic Adaptive Streaming over HTTP,

4. Can big data analysis help YouTube offer better service? if yes, in which aspects?
   – keywords: Machine Learning, Unsupervised Learning

4 Formalities

4.1 Time Table

4.2 Roles

5 References

[1] Tim Suglin. Online video jumps on the big data bandwagon. `http://www.streamingmedia.com/Articles/Editorial/Featured-Articles/Online-Video-Jumps-on-the-Big-Data-Bandwagon-91353.aspx`, August 2013. [Online; posted August/September 2013].