Caracterização do Comportamento dos Clientes de um Sistema de Vídeo ao Vivo Durante um Evento de Larga Escala na Internet

Alex Borges Vieira
alex.borges@ufjf.edu.br

Breno Almeida; Gustavo Carnivali; Wanger Almeida; Jussara M. Almeida; Italo Cunha; Alex Vieira
Introduction

Large scale Internet services are a class of applications which impose strong requirements on their management infrastructure and on their execution environment.

Large scale internet services are services deployed in Data Center environments and running on a large number of machines.
Introduction

Large scale internet services are a class of applications which impose strong requirements on their management infrastructure and on their execution environment.

Large scale internet services are services deployed in Data Center environments and running on a large number of machines.

Some Issues...

- Scalability
- Cost
- Monitoring
Large Scale Events...

Generates Ultra Mega Huge Hyper Peta
Amount of Data
“Since its early stages, about then years ago, Internet streaming has been pointed as a killer application. Indeed, the large number of current commercial applications indicates a trading point, where users are replacing traditional media ---such as radio, TV, and satellite--- to the current Internet streaming scenario. In part, users prefer Internet due to its low cost, high flexibility and capillarity.”
A few years back - 2002
A few years back - 2002 Streaming!
A few years back - 2002

Video On Demand (VOD): A Killer App or “Too Little, Too Late”?

Video on Demand Is Ready, but the Market Is Not.
The New York Times
Back to the past - 2002

In 2000, Netflix was offered for acquisition to Blockbuster for $50 million, however Blockbuster declined the offer. Later in 2002, Netflix opened up an initial public offering (IPO), selling 5.5 million shares of common stock at $15.00 per share. That same year, the company sold an additional 825,000 shares of common stock at the same price.

During its first few years, Netflix incurred substantial losses. Netflix did not post its first profit until the 2003 fiscal year. The company earned $6.5 million profit on revenues of $272 million. By 2005, 35,000 different film titles were available, and Netflix shipped one million DVDs out every day.
In 2000, Netflix was offered for acquisition to Blockbuster for $50 million, however Blockbuster declined the offer. Later in 2002, Netflix opened up an initial public offering (IPO), selling 5.5 million shares of common stock at $15.00 per share. That same year, the company sold an additional 825,000 shares of common stock at the same price.

During its first few years, Netflix incurred substantial losses. Netflix did not post its first profit until the 2003 fiscal year. The company earned $6.5 million profit on revenues of $272 million. By 2005, 35,000 different film titles were available, and Netflix shipped one million DVDs out every day.

Currently, Netflix’s market value stands at $32.9 billion.
In 2002

Lots of service

Lots of research on Vod architecture

but...

No one knows how users behave during live streaming large scale events!
Back to the past... 2002

In fact...
“Here you come again”...

The *HTTP* age (201X)
The *HTTP* age (201X)
The **HTTP** age (201X)
Key problem in the “new” age

Large scale events are no longer exception.

User behavior changed during the last decade.

Does anyone still doubt it?

When did you last bought a newspaper?
Do you still use your cellphone to phone?
Key problem in the “new” age

Large scale events are no longer exception.

User behavior changed during the last decade.

*Does anyone still doubt it?*

When did you last bought a newspaper?
Do you still use your cellphone to phone?

*In sum: past works may be really outdated.*
Let’s work again

FIFA WORLD CUP
Brasil
Our dataset
Number are impressive
We focus on …

Argentina x Switzerland (Tuesday)
- the most streaming match by the content provider in Br.

- Alemanha x Algeria (Monday)
  - *awesome match*
The most “streamed”

ESPN Live Stream Crashes During USA-Germany World Cup Match

first half: 1.4 million
peak 1.7 million*

* There was concurrent streams
Germany's destruction of Brazil became the most discussed sports match in Twitter's history as 35.6M tweets were posted during the game. Most of them were laughing at Brazil.
- The Guardian.
Germany's destruction of Brazil became the most discussed sports match in Twitter's history as 35.6M tweets were posted during the game. Most of them were laughing at Brazil. - The Guardian.
A brief looking
A brief looking
A brief looking
Ok... let’s talk about cliente behavior modeling
A very simple model I
A very simple model II
Number of sections
Number of sections

- Users present very few sections.
Session duration
Session duration

- Session tends to be large
Session duration

- Session tends to be large
Off times

- User idle period is short.
Off times

- User idle period is short.
- Bias? NOP...
What’s going on? Compared to past few years...

Number of sessions is lower:

- Before: 40% of clients - a single section;
- Now: about 80% of clients - a single section.

On time is longer:

- Before: 90% of sections <= 10 min;
- Now - mean session time ~ 40 min.

Off time is lower

- Before: 35% of clients - Off time > 20 min;
- Now: 6% of - Off time > 20 min;
In sum...

Client behavior is useful to develop better applications;

To evaluate server performance;

But, there are so many variables... so many information...

Quite impossible to precisely reproduce server load dynamics.
In sum...

Client behavior is useful to develop better applications;

To evaluate server performance;

But, there are so many variables... so many information...

Quite impossible to precisely reproduce server load dynamics.

We are able to reconstruct workload during steady state only!
Caracterização do Comportamento dos Clientes de um Sistema de Vídeo ao Vivo Durante um Evento de Larga Escala na Internet

Alex Borges Vieira
alex.borges@ufjf.edu.br

Breno Almeida; Gustavo Carnivali; Wanger Almeida; Jussara M. Almeida; Italo Cunha; Alex Vieira
Do we have more time?
Synthetic Workload Generator

(a) Ips únicos
Synthetic Workload Generator

(b) Volume (GB)
Data prediction
Client behavior is not the only thing we should be aware.

Newcomers rate?
Model time dependencies?