SCTE·ISBE CABLE-TEC EXPO 2018 ATLANTA, GA OCTOBER 22-25

2018 Fall Technical Forum SCTE·ISBE • NCTA • CABLELABS

Real-Time Analytics for IP Video Multicast

A Technical Paper prepared for SCTE•ISBE by

Dr. Claudio Righetti
Chief Scientist
Telecom Argentina
Agüero 2392, Buenos Aires, Argentina
Phone: +5411 5530 4468
crighetti@teco.com.ar

Emilia Gibellini
Data Scientist
Telecom Argentina
egibellini@teco.com.ar

Florencia De Arca
Data Scientist
Telecom Argentina
fdearca@teco.com.ar

Mariela Fiorenzo
Data Scientist
Telecom Argentina
mafiorenzo@teco.com.ar

Gabriel Carro
Senior VP R&D
Telecom Argentina
gcarro@teco.com.ar

Table of Contents

Title **Page Number**

List of Figures

Title **Page Number**

List of Tables

Title **Page Number**

Abstract

In order to understand the impact of multicast implementation, it is necessary to collect data on key indicators such as the number of concurrent streams, the average bitrate, and the average bandwidth, among others. We use these indicators to estimate the gain, in terms of bandwidth, at a service group level. The aim of this paper is to analyze the way in which the gain varies according to the service group size and its location, and to obtain –through the usage of statistical modeling– a model that describes and quantifies this relationship. In addition, the gain is estimated under a wide variety of scenarios, to know how many channels should be set to multicast, and if there is any gain in having a real-time analytics system that updates what channels should be delivered using Multicast.

Contents

1. Introduction

It has been more than thirty years since the IP (Internet protocol) Multicast standardization work started [RFC] [1]. Much research has been conducted into the benefits of IP Multicast versus Unicast for Live video in access networks with xDSL (Digital Subscriber Line), FTTH (Fiber To The Home), DOCSIS (Data Over Cable Service Interface Specification) and Wireless technology. In particular, cable operators have been using technology for years to distribute digital video over IP backbone networks to multiple head-ends and hubs to feed broadcast QAMs (Quadrature Amplitude Modulation).

CableLabs-IP Multicast Working Group- published a document ("IP Multicast Adaptive Bit Rate Architecture Technical Report" [2]) describing how to put together two network concepts: Multicast and Adaptive Bitrate delivery, in what is called M-ABR (Multicast Adaptive Bitrate).

This approach enables IP video subscribers in the same node to consume a common linear video stream over the access network, thus reducing access network bandwidth requirements over Unicast delivery (where a separate stream is delivered to each subscriber). The adaptive video streaming is a type of technology responsible for delivering video through the Internet in an efficient way. This is done by selecting the image quality according to the resources of each user. Adaptive Bitrate streaming technologies are almost exclusively based on HTTP (Hypertext Transfer Protocol).

However, in the world of cable operators there are still some questions with regard to the benefit of implementing M-ABR in their networks. How convenient is that Multicast migrate to IP Video Service? If service areas tend to be reduced, does that situation justify the implementation of this technology? What policy is used to define what channels are Multicast and what are Unicast? Should it be reached with a static policy or a dynamic policy in real time? If this assignment is adaptive, must the analysis of the demand be done in real time? Must we apply machine-learning technologies? Do client behaviors change significantly from one service area to another? Through an updated analysis of the behavior of video subscribers and the incorporation of machine-learning (ML) technologies, our work is aimed at finding the answers to the above-mentioned questions.

There are several works related to the video subscribers' behavior in HFC (Hybrid Fiber Coaxial) networks. In most cases they have been made by the vendors with samples of some operators – Cable Labs in 2009 and 2012 as well [3].– Our analysis includes the behavior of Legacy STB (Set Top Box), Hybrid STB (Video QAM and Control IP) and the behavior of our OTT (Over The Top) subscribers –the latter are part of our service called "FLOW".– In this paper, we also include the behavior in major events, such as the 2014 and 2018 FIFA World Cup.

2. Motivation and Backgrounds

Telecom Argentina (former Cablevisión Argentina) has already moved from legacy Digital TV (DTV) to Hybrid (DTV+IP) and OTT system and now, we are finally starting to deploy Full IP Video delivery. Our biggest challenge in migration to full IP video is to deliver fully managed linear TV services to any device. The primary motivation for this migration to be based on IP Multicast is the expected improvement in efficiency over Unicast.

IP Multicast WG defines *Best Practices as the techniques that the working group has identified as generally being the preferred design approach in a specific area.* In this work, we seek to see how we can apply these best practices in light of the analysis of our clients behavior.

2.1. Definitions

The IP Multicast CableLabs Working Group suggests multicast live linear TV as the best practice and identified three main approaches to determine what content should be delivered using Multicast:

- *Viewership Driven Multicast*: any stream with more than one consumer will be multicast regardless of bit rate.
- *Policy Driven Multicast*: *n* configured channels are available for request via multicast (typically, these are the *n* most popular channels for a given time period and location)
- *Hybrids*: There are hybrids between the two previous models, the two possible ones that the working group would like to highlight are:
  - *Viewership Driven with Maximum Number of Multicast Channels*: the set of multicast channels at any given time is driven by *real-time requests* for content. However, like Policy-Driven multicast, there is a maximum number of channels allowed to be multicast.
  - *Viewership Driven with Limited Bit Rates*: This hybrid model adds to the pure Viewership Driven model a policy component that limits the number of bit rates which are available for multicast. Typically, in this model, bit rates are limited to HD-only or HD- and SD-only.

2.2. Background

Maximizing efficiency was the motivation for the development of IP Multicast. This efficiency is directly related to our video subscribers' behavior. This means that we must determine what the most popular channels are, and those will be the ones delivered using Multicast. The Pareto principle –or the 80-20 rule– is often referred to when describing video popularity and the concentration of user interest towards a few popular programs [2] [3].

Many authors have adjusted this popularity following a Zipf distribution [4], and based on that, they have determined the gain of using Multicast in the most popular channels. The distribution is as follows:

$$P_i = \frac{1}{\sum_{i=0}^{N} i^{(1-\propto)}}$$

Where α is the skew factor and i is the rank.

Multicast gain is a measure of the efficiency of multicast delivery compared to unicast. The multicast gain achieved depends on a variety of factors, especially, the number of viewers per service group and the popularity of the programming.

With α = -1, there are just a few very popular channels at a particular time and the potential for high Multicast gain >> 8[5]. Multicast gain of 8 indicates that the Unicast approach requires 8X the numbers of streams.

If α = 0.5, we will have more popular channels at a particular time and potential for low Multicast gain >> 3. For example, in [6] it was reported a gain of 5 under certain SG size conditions, popularity, etc.

Through this example, we want to illustrate in a simple way how the skew factor influences the gain; having a *long tail* and a *tall head* in the distribution. The tall head during prime time –observed in [6], for instance–corresponds to 60% of viewers watching the top 10 channels.

Zipf-Mandelbrot is the most appropriate model to replicate video popularity distributions –as presented in [7] and subsequent work [8].–

3. Systems Overview and Data Description

Telecom Argentina S.A. provides Live TV (or linear TV) and VoD (Video On Demand) services over two systems: Flow and DTV. There are about 500 Live channels and over 50,000 videos available. The users of both systems pay a monthly subscription fee to use Live TV and VoD services and they have to pay extra fees for some VoD contents. There are many differences between the systems; by way of example, Flow has functions as Catch up TV, Restart TV and NDVR (Network Digital Video Recorder), while Legacy platform has a TV guide where users can choose a Live channel or search for a specific VoD content by browsing into a couple of folders.

3.1. Data Description

In order to analyze the TV user behavior, we collected a large amount of logs from the two platforms from July 2017 to July 2018 and then selected particular days and weeks to conduct our study. There are about 3 million subscribers, taking into account STB –Legacy and Hybrid–and OTT devices, and the average number of daily records is about 55 million, so the sample that has been chosen is representative of general TV system users.

The logs contain many fields and those differ according to the type of system. Table 1 shows the format of Legacy system logs and Table 2 shows the format of Flow logs.

Table 1-DTV (Legacy) log sample

Fields	Examples
Date	06/26/2018
Hour	00:00.0
IP Address	10.132.34.53
Flag	w
Set Top ID	0004c96740
Service ID	788
Channel Number	4612
Time	61
Idle	61
Data	*Telediario 10 minutos*
Region	SANTA_FE_8

Table 2-Flow log sample

Fields	Examples
Account ID	3101671
Customer ID	788840
Device ID	3632563
Type	PHONE
OS Type	ANDROID
OS Version	7
Brand	SAMSUNG
Model	SM-G610M
Firmware	1.10.1-173531
Channel ID	277
Channel Name	DISNEY XD
Program ID	MV00000000153771
Program Title	*Un gran dinosaurio*
Quality	SD

Tunein	25/05/2018 09:28
Tuneout	25/05/2018 09:30
Duration	17

4. TV User Behavior Analysis

In this section, we explain some of the analysis we carried out related to TV user behavior from our Flow and DTV systems. As we are planning the migration to a Full IP Video platform, we focus our attention on Multicast gain at CDN (content delivery network) and SG (service group) levels. In order to estimate the impact on the CDN and SG sizing, we studied some parameters, described as follows:

- *Concurrence*. Number or percentage of STB or OTT devices using a service at the same time (day, hour, minute, etc.).
- *Access frequency*. Number of tunings of each channel or videos during a certain time window.
- *Type of requests*. It refers to Live TV or VoD.
- *Bitrate*. Streaming bitrate of Live channels or VoD videos (in bps). It depends on the quality of the contents, the quality of the channel and the type of device that reproduces the content.
- *Popularity*. Probability of tuning a certain channel or video. It is calculated as the number of tunings to this channel or video divided by the total tunings.
- *Busy hour*. Hour slot with the greatest concurrence in all day, which generally happens from 9 p.m. to 10 p.m. or from 10 p.m. to 11 p.m. It differs from the US's prime time, because in Argentina people tend to have dinner after 8 p.m.

For Legacy STB, we analyzed access frequency and concurrence by hour, and for Flow STB, only concurrence by 10 minutes. The reason why we studied different indicators for each system is the structure and complexity of each log. Calculating the concurrence for Flow is quite simple through an elaborated algorithm but it is not possible to reproduce the same algorithm for Legacy system. Therefore, for Legacy STB we calculated access frequency that is the most similar indicator to the concurrence in the lowest time possible that is an hour.

4.1. Comparison on Live TV and VoD User Behavior

To understand user behavior and traffic of both systems, we show in Figure 1 the concurrence of OTT devices on Flow ([a]) and the proportion of them in Live TV and VoD ([b]). In Figure 2,we represent the Live TV and VoD tunings, expressed as the percentage of total STB on the Legacy system([a]) and the proportion of STB that were streaming Live TV and VoD from 8 p.m. to 11 p.m. ([b]).

We conducted a weekly analysis to study concurrence in both cases and we found clear differences between the systems, mainly due to the granularity of data and the types of devices analyzed in each case. For the week represented in Figures 1 and 2, we observed that Legacy STB follows the

typical pattern of access frequency all the week while OTT devices present an irregular pattern, but they all have in common a peak in the middle of the week, which appears very pronounced in the Flow system, due to a soccer match.

Then, we carried out a daily analysis to study the proportion of Live TV and VoD tunings –we show the soccer match day and the previous day. – In both cases, the proportion of Live TV tuning is higher than VoD. For Flow, VoD tunings are about 20% and in the Legacy system they represent less than 1%. This result is in line with the configuration of the platform, which is able to support up to 12K simultaneous VoD tunings, which is 1% of all active Legacy STB.

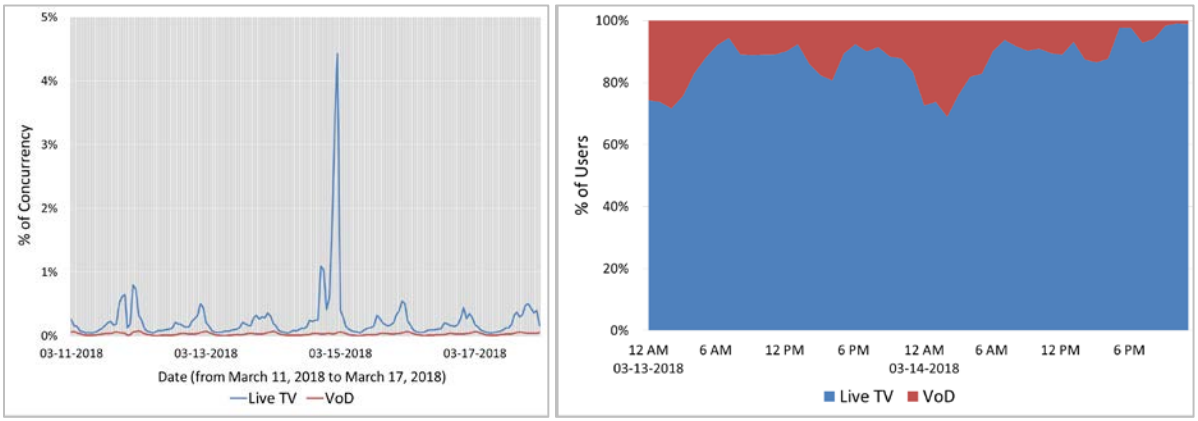

Figure 1-[a] Concurrence of OTT devices from Flow, colored by type of request. [b] Proportion of Live TV and VoD tunings.

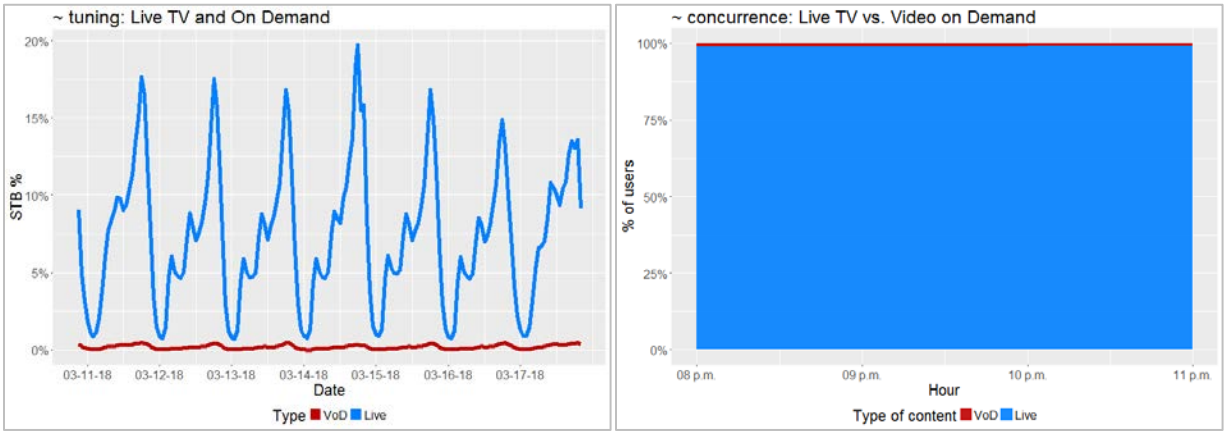

Figure 2-[a] Distribution of requests from STB of the Legacy system. [b] Proportion of Live TV and VoD tunings on March 14, 2018 from 8 p.m. to EOD.

One of the questions we have to answer is if it is necessary a multicast configuration for some channels and if so, how many channels are needed to be configured as multicast. We founded through the analysis that live contents are what users tend to view the most on the STB. Figures 1 and 2 show that while VoD represents around 20% of the views in OTT, in STB the same

percentage is around 1%. In [9] it was observed that more than 80% of viewers were found to be watching live TV between 7 p.m. and 9 p.m.

As Multicast is the best practice for Live TV, we conclude that it is necessary to implement it.

In the next sections, we make a deeper analysis of Live TV users behavior from both systems. We focus our attention on a day with a major event and on a regular day. The results of this analysis answer most of the above-mentioned questions.

4.2. Regular Weekday

In order to understand if the users behavior changes depending on the day of the week or when some particular event happens, we performed a study during regular weekdays –'regular' means a day with no soccer match or any other major event. We selected May 24, 2018 as an example.

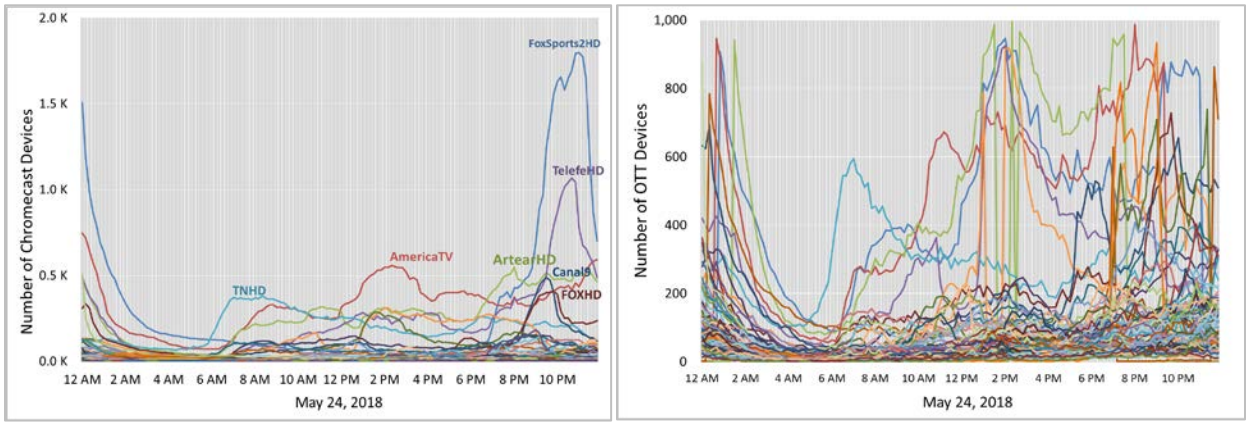

Figure 3-[a] Concurrence of Chromecast devices. [b] Concurrence of other OTT devices, on Flow, every 10 minutes. May 24, 2018.

We can see in Figure 3 that the subscribers who accessed the live contents via Chromecast tended to choose sports channels –such as *Fox Sports*– or general interest channels –such as *Telefe, America TV, Artear*, among others. On the other hand, when we look at the rest of the OTT devices, we observe a substantial difference in users behavior, as the views are distributed among many channels.

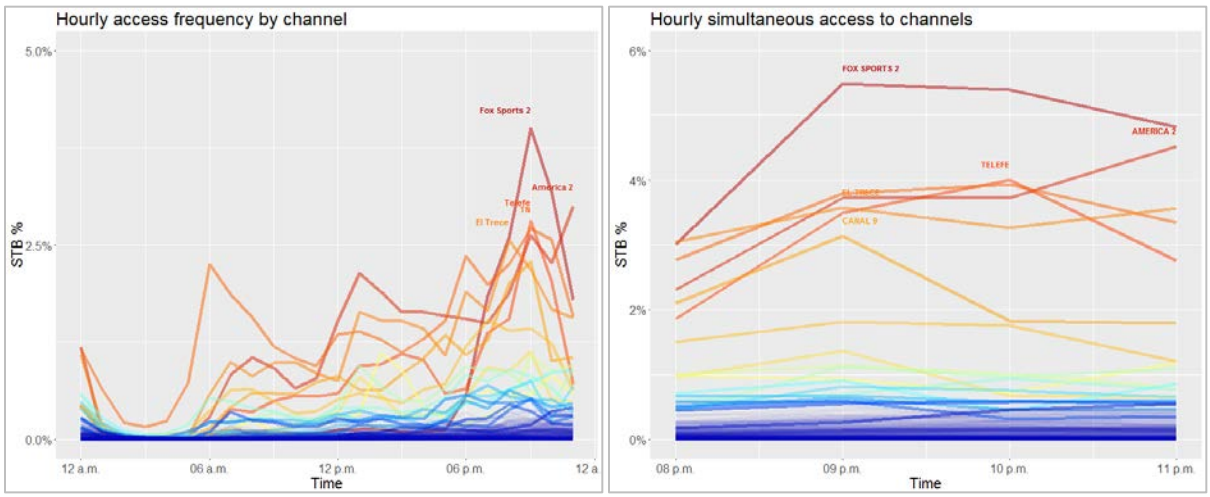

Figure 4 - [a] Hourly access frequency to Live TV in Legacy system, on May 24, 2018. [b] Legacy STB playing Live TV channels simultaneously, on May24, 2018 from 8 p.m. to EOD.

The conclusions we get from the observation of tunings on the Legacy platform are similar to the ones driven from the Chromecast case. It is clear from Figure 4 that the majority of the requests go towards the sports channel, *Fox Sports*, and the general interest and news channels *America 2, Telefe*, *El Trece* and *TN*. It is important to mention that on that night, *Fox Sports* was transmitting a Spanish soccer match from 8 p.m. to 11 p.m., which is not a major event in Argentina, but still gets many viewers.

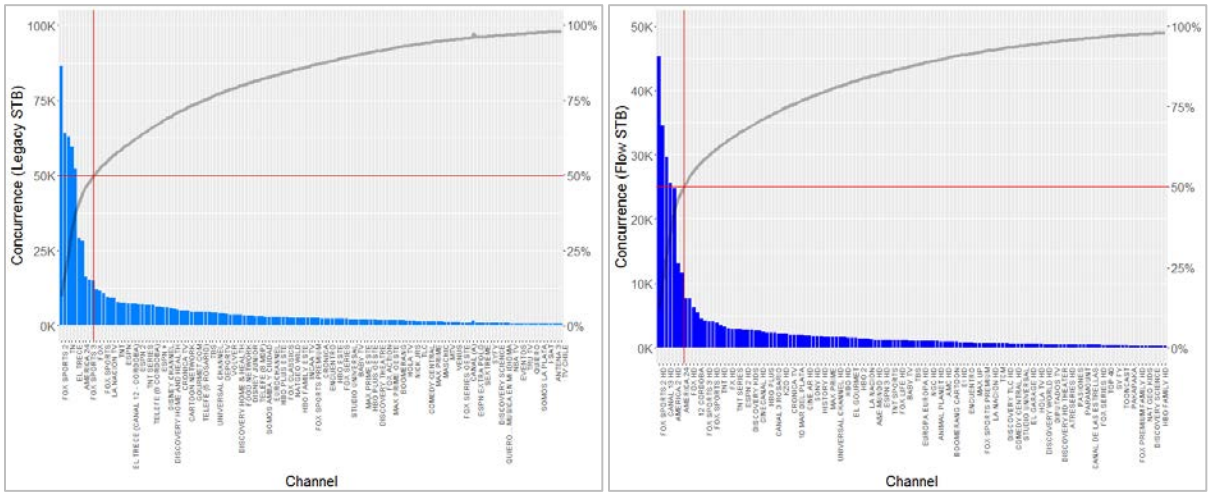

Figure 5-[a] Concurrent tunings from Legacy STB. [b] Concurrent tunings from Flow STB. May 24, 2018between 10 p.m. and 11 p.m. (busy hour).

Figure 5 shows that there is not much difference among the percqentage of views that the top channels get on a regular day. One channel concentrates 10% of the views on Legacy and 12% on Flow. It is followed by a set of four channels that get between 9% and 6% of the views. On the

legacy system, we observed that 10 channels get 50% of all simultaneous views, while on Flow this happens with eight channels.

We compared the distribution of the simultaneous views to several Zipf and Zipf-Mandelbrot theoretical distributions. The Zipf-Mandelbrot is:

$$p_i = \frac{C}{(i+b)^a}$$

Where p_i is the probability that a certain STB would tune in the *i-th* most popular channel,$a > 1$ and $b \geq 0$. The parameter C is a normalizing constant that depends on a and b:

$$C = b^{1-\frac{1}{a}} \cdot (a-1)^{\frac{1}{a}}$$

We found that the one that approximates the most to the data observed is a Mandelbrot-Zipf with parameters a=1.11 and b=0.56, which is shown in Figure 6.

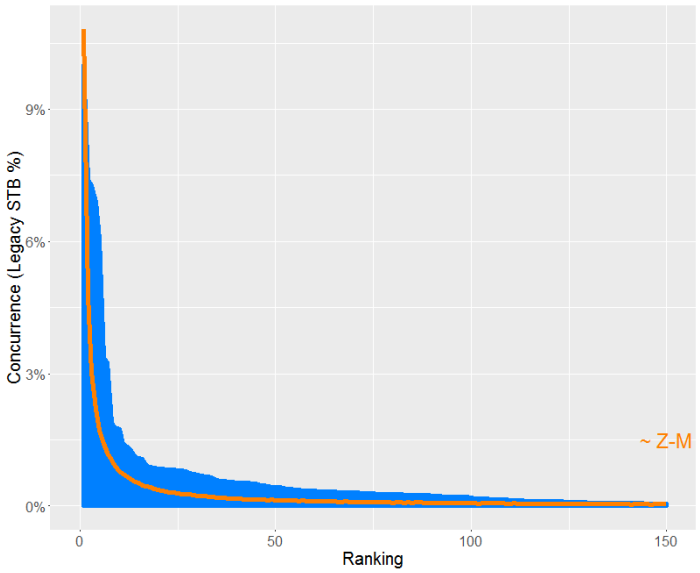

Figure 6 - Comparison of observed distribution to Zipf-Mandelbrot.

This agrees with the results obtained in related works, and shows that under normal circumstances a few channels –10 or less– concentrate a high percentage of the total views. We proceed to show how this distribution is affected when a major event occurs.

4.3. Major Events

In Argentina, soccer matches really drive TV usage and can introduce several variations to the channel ranking. In order to investigate the impact of these sports events, we analyzed two world cups and other important sports events. We picked two dates: March 4, 2018 (a typical soccer Sunday) and March 14, 2018 (Argentinian Super Cup). This event faces the winning teams from

previous tournaments and, in the latest edition, it faced *Boca Juniors* and *River Plate*, the two soccer teams with the most fans.

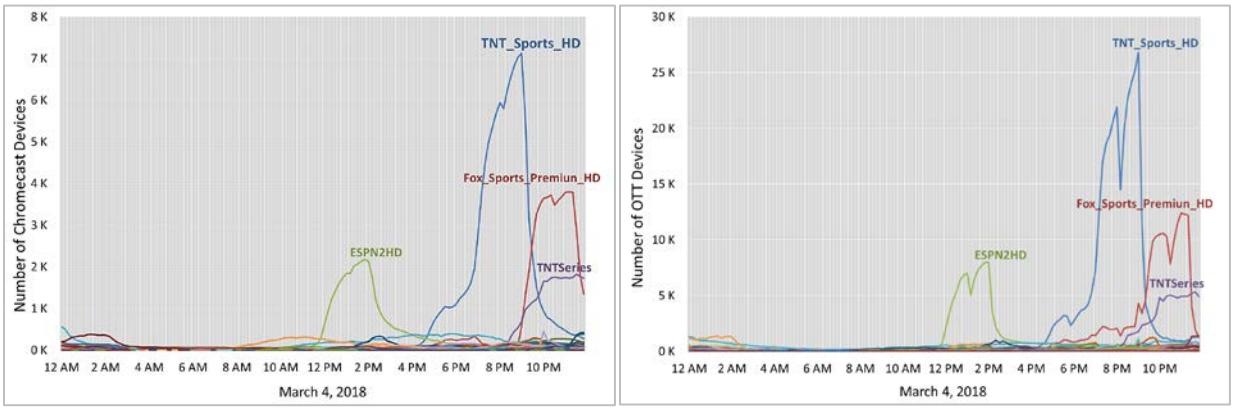

Figure 7 - [a] Concurrence of Chromecast devices. [b] Concurrence of other OTT devices of Flow, every 10 minutes. March 4, 2018.

Figure 7 shows how Flow Live TV consumption occurred via Chromecast and other OTT devices (phones, tablets and computers) on a typical Sunday with matches. The most visited channels were *TNT Sports* and *Fox Sports Premium*, around the prime time. The other two channels that stand out are *ESPN* and *TNT Series*. Therefore, three out of the four signals that accumulate the most views are sports channels.

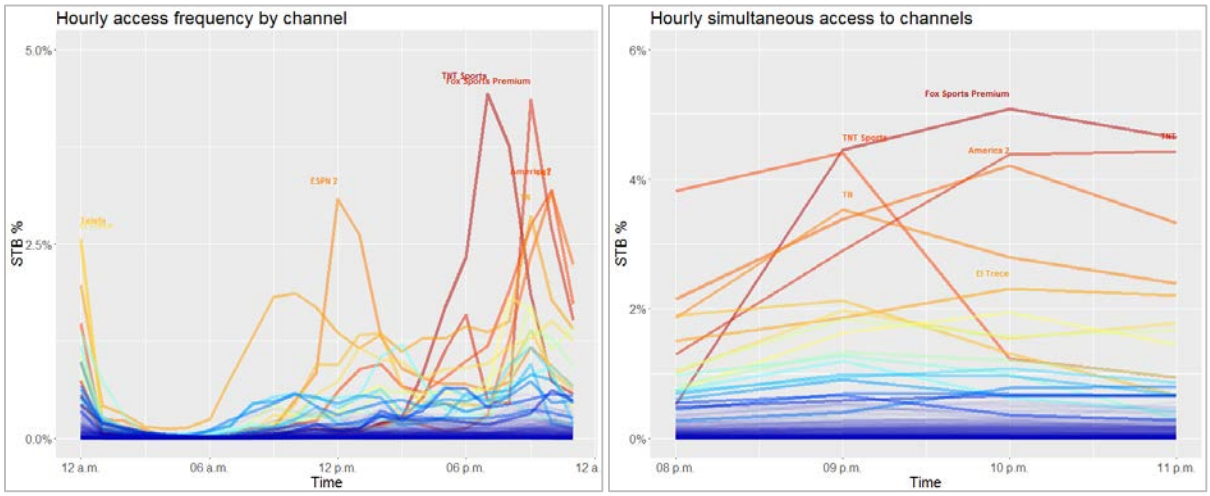

Figure 8 - [a] Hourly access frequency to Live TV in Legacy system, on March 4, 2018. [b] Legacy STB playing Live TV channels simultaneously, on March 4, 2018 from 8 p.m. to EOD (end of the day).

With regard to the Legacy system, Figure 8 leads us to reach similar conclusions: the channels that were tuned the most around the prime time are *TNT Sports* and *Fox Sports Premium*, and during the afternoon, *ESPN*. It is easy to see in the second chart that during the night there are some other series (*TNT*), news (*TN*) and general interest (*America 2* and *El Trece*) channels that concentrate views.

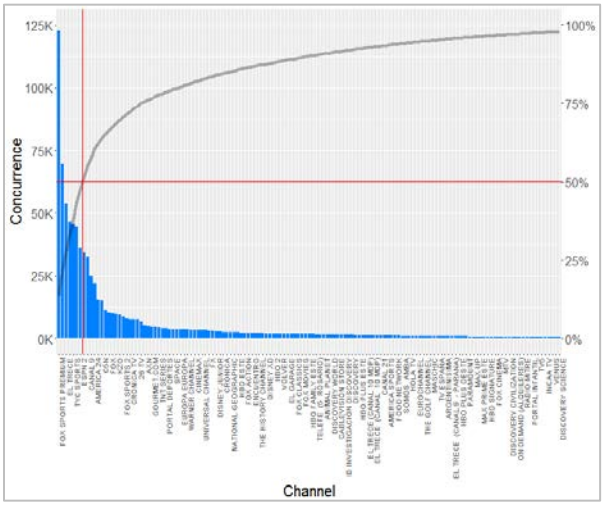

Figure 9 - Concurrent tunings from Legacy STB on March 14, 2018 between 9 p.m. and 10 p.m. (match hour).

On March 14, the match started at 9.10 p.m., so we analyzed the concurrence around that time. According to the Legacy system data, showed in Figure 9, the most visited channel during that time slot is *Fox Sports Premium*. It consolidates 125K simultaneous views, which represents around 15% of the STB. Meanwhile, the second most viewed channel is *El Trece* –general interest, – with an 8% concurrence.

We would like to highlight the difference with respect to the regular day ranking. In this case, the first channel doubles the second channel's concurrence. Besides, seven channels concentrate 50% of the views. It should be taken into account that on a regular day, around 10 channels get 50%.

To sum up, we conclude that major events not only introduce a variation in the channels that appear on the top of the ranking, but also modify the distribution of the tunings during a certain time interval.

4.4. Variation of Rankings in Time

To explore the variations of the rankings through time, we calculated each channel's popularity in one-hour intervals, and established a daily ranking based on the maximum popularity achieved by each channel in one hour. Then, to compare the rankings from different dates, we used the Spearman's rank correlation coefficient [10]. Figure 10 [a] shows the correlations obtained after comparing the top 10 channels on July 1, 2017 versus the top 10 calculated for the following 180 days. This data corresponds to the Legacy system.

It is easy to notice that during the first 40 days the correlation is high, indicating that the list of the top 10 generally consists of the same channels. After that, a series of soccer matches and TV series appear on screen. Correlation is lower when we compare July 1, 2017 to soccer days or when a popular series is first aired.

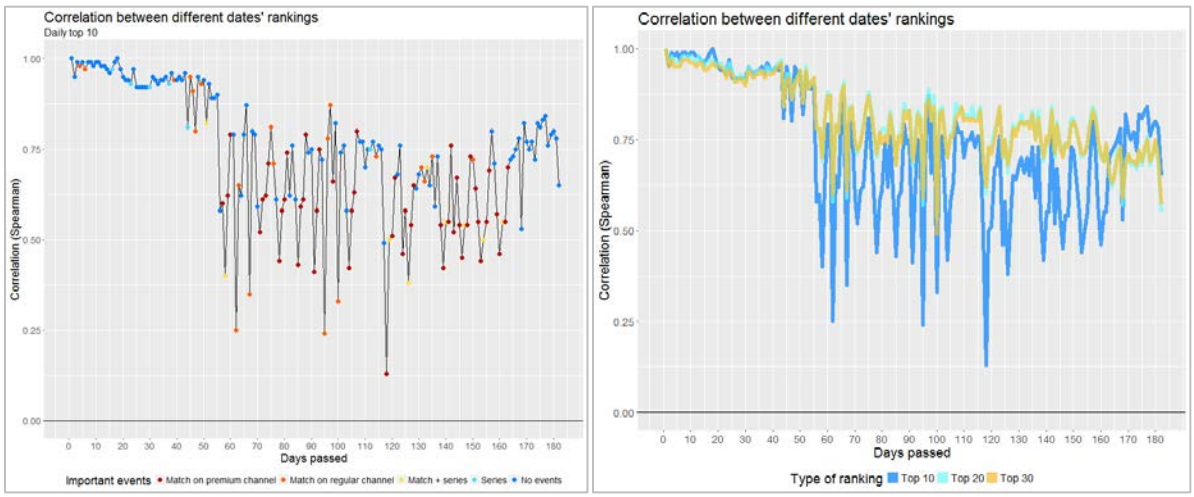

Figure 10 - [a] Correlation between the top 10 channels on July 1, 2017 and the top 10 on the 180 following days.[b] Correlations between the top 10, top 20 and top 30 channels, on July 1, 2017 versus the same rankings on the following 180 days.

In Figure 10 [b], we compared the correlation series when we use the top 10, top 20 and top 30 channels to calculate it. The top 10 series is the one with the highest variation, meaning it is more sensible to changes in the ranking. Nevertheless, the daily top 10 is similar throughout the days, even when an important series is transmitted. What introduces more changes is the transmission of a major sports event.

From the user behavior analysis, we conclude that it would be enough to set by multicast a limited set of signals, consisting of the most popular general interest, news, movies and sports channels. According to the Legacy data, there are 11 general interest and news, 6 movies and 8 sports channels that regularly appear among the top 10.

After that, we continued to analyze the gain, in terms of capacity, that multicast implementation would bring. This analysis aims at looking for an optimum number of channels that should be delivered using multicast.

5. Multicast Gain

We evaluated the multicast gain at CDN and at service group levels as a percentage of the capacity needed under a 100% Unicast scheme, which we define as follows:

$$Capacity\ 100\%\ Unicast = \sum_{\substack{All \\ channels}} Concurrence \cdot Avg\ bitrate$$

When working with data from the Legacy system, we counted on the access frequency to approximate the concurrence, and we assumed that the average (Avg.) bitrate is 4 Mbps.

We analyzed three scenarios: in the first one, called *Top 10*, the 10 most popular channels are delivered to Multicast, and the rest remain Unicast. For the second and third, named *Top 20* and *Top 30*, we set the 20 and 30 most popular channels to Multicast and the rest of the grid in Unicast.

The capacity that we would need at CDN level is calculated as follows:

$$Capacity\ Top\ \text{"X"}\ Scenario = \sum_{\substack{Top\ "X" \\ channels}} Avg\ bitrate + \sum_{\substack{Other \\ channels}} Concurrence \cdot Avg\ bitrate$$

Finally, we defined the multicast gain as:

$$Multicast\ gain\ for\ Top\ \text{"X"}\ Scenario = \frac{Capacity\ 100\%\ Unicast - Capacity\ Top\ \text{"X"}\ Scenario}{Capacity\ 100\%\ Unicast}$$

To calculate the gain at CDN level, we used the total concurrence –or total access frequency– for each channel. On the other hand, to calculate the gain at service group level we used the concurrence observed within the service group.

In addition, to estimate the maximum possible multicast gain, we proposed a theoretical scenario in which all the channels are transmitted via multicast. This is useful to determine whether the gain in other scenarios is close to the maximum or not.

$$Capacity\ 100\%\ Multicast = \sum_{\substack{All \\ channels}} Avg\ bitrate$$

So, the maximum gain is estimated as:

$$Maximum\ Multicast\ gain = \frac{Capacity\ 100\%\ Unicast - Capacity\ 100\%\ Multicast}{Capacity\ 100\%\ Unicast}$$

We would like to highlight that all the results in this section are based on the Legacy data. As seen in the previous section, the VoD views represent less than 1% of all views, so the results may differ in case the distribution of views between VoD and Live is other than 1%-99%.

5.1. Analysis at CDN Level

In order to address the question of how many signals should be delivered using Multicast, we analyze how multicast gain varies according to the scenarios and time in this section. Figure 10 shows the variation of the gain at CDN level under the three scenarios, during one week, from May 20 to May 28, 2018. We observed, as expected, that gain is greater at peak times, for all scenarios.

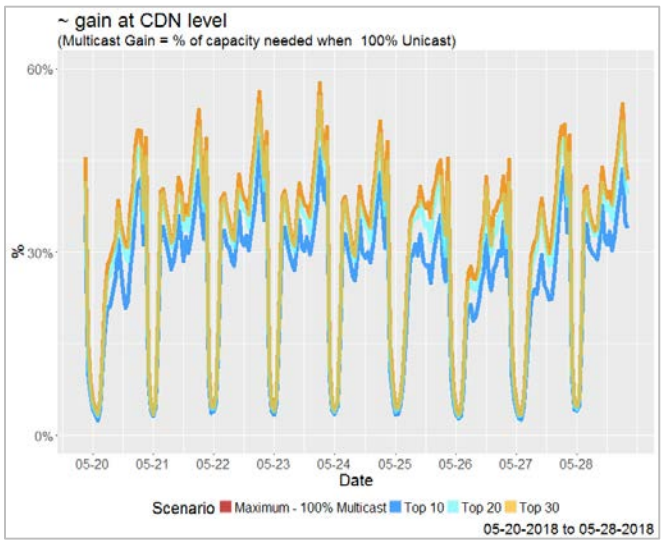

Figure 11 - Multicast gain, as a percentage of the capacity needed with 100% unicast scheme.

In Figure 11, it should be noted that the Top 30 line overlaps with the maximum. It seems that there is not much difference when there are 20, 30 or even if all channels are delivered using multicast. The difference between the Top 10 scenario and the maximum is, on average, 6% with a standard deviation of 2.8%.

It is intuitive that when a few channels are delivered using multicast, there is gain, but then if we add more channels, after a certain point the gain does not suffer a drastic increment. It is our task to find out where that cutoff is. In order to do that, we plotted the gain by the number of channels sent via multicast, as if each hour slot was a new sample. It is clear from Figure 12 that such cutoff should be between 10 and 30.

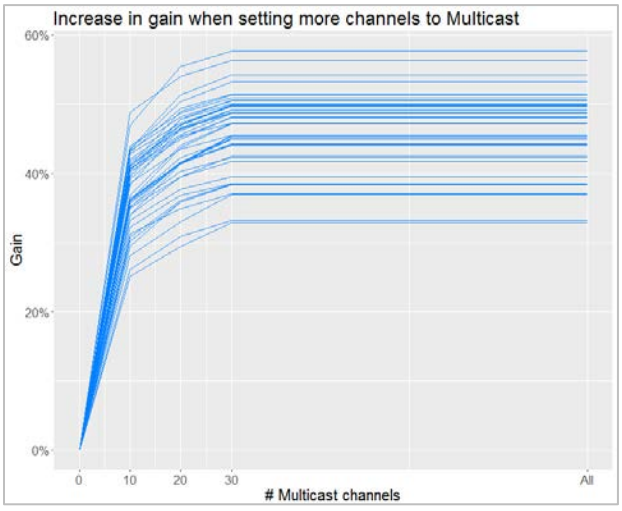

Figure 12 - Multicast gain versus number of channels that are set to multicast. Based on data from May 20 to May 28, 2018 gain calculated for hour slots from 8 p.m. to midnight.

If 25 channels were delivered using multicast, as we proposed at the end of section 4, the multicast gain at CDN level would be near its maximum.

5.2. Analysis at Service Group Level

We know that at service group level the gain is subject to the service group size. On the other hand, service areas tend to be smaller in time. Therefore, it is a frequent question whether there is multicast gain at this level. In this section, we tried to find the answer.

Figure 13 shows the distribution of Telecom Argentina S.A. service groups' size, in terms of households passed (HHP), by region ([a]), and the relationship between the SG size and the maximum possible gain under a 100% multicast scheme ([b]). The STB count per service group is higher in Buenos Aires than in other regions, because this is a highly populated area. It is clear from the scatter plot in Figure 13 [b] that the relationship between the gain and the service group size has a logarithmic shape.

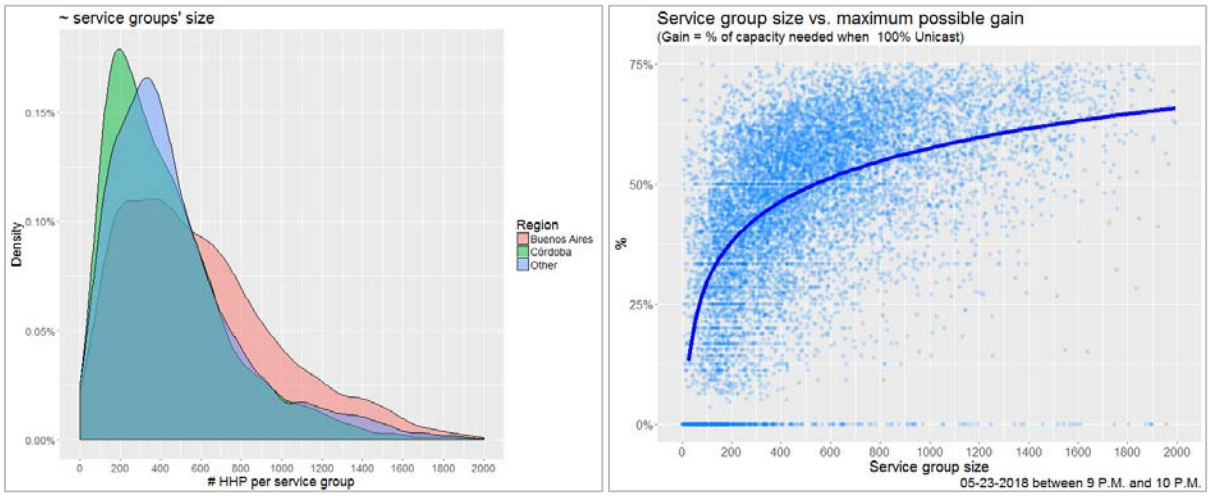

Figure 13 - [a] Distribution of service group's size (HHP) by region.[b] Maximum multicast gain versus service group size.

In Figure 13, we plotted the maximum multicast gain at service group level, and in [a] we colored each line according to the service group size. It should be notice that plot [a] is not an area plot, it just contains so many series that it looks like one. The warmest colors, that are the biggest service groups' series, indicate higher maximum possible gain. Chart [b] summarizes chart [a] by showing the minimum, maximum and average gain at service group level.

Figure 14 - [a] Maximum multicast gain at service group level, colored by service group size. [b] Mean multicast gain at service group level.

The gain may also be determined by the region where the service group is located. We know that channels' popularity tends to vary according to the region. By way of example, Figure 14 compares the popularity in Buenos Aires versus Córdoba ([a]) and the popularity in Buenos Aires versus La Plata ([b]). When a point is near the diagonal line, this means that the channel is as popular in the other region as it is in Buenos Aires. Otherwise, when a point gets far from the diagonal and near the edges, this means that the channel is very popular in one place but very unpopular in the other.

Local versions of the news and general interest channels tend to be popular within the regions where they are from. In the case of May 24, 2018 data, there are some news channels that are popular in Córdoba and not so much in Buenos Aires, and the sports channels are popular at both locations. Then, in La Plata there are more coincidences –most of the higher popularity points are near the diagonal.

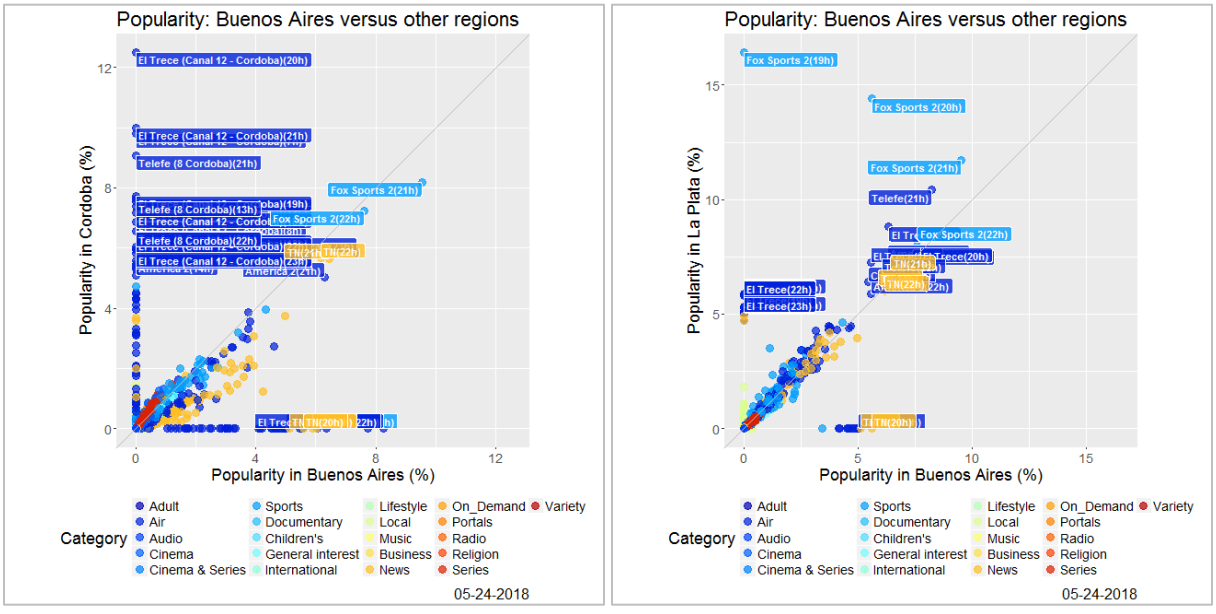

Figure 15 - Popularity in Buenos Aires region versus other regions, on May 24, 2018.[a] Buenos Aires versus Córdoba. [b] Buenos Aires versus La Plata.

To get an idea of multicast gain under the different scenarios, Figure 15 shows the average during the week from May 20 to May 28, 2018. The *Top 30* series overlaps with the *Maximum - 100% multicast* series, indicating that when there are 30 channels delivered using Multicast it may not make a difference to continue to add more signals.

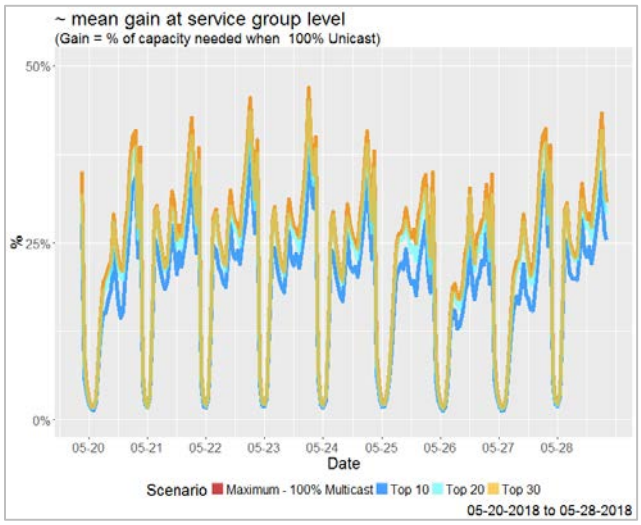

Figure 16 - Average multicast gain at service group level for different scenarios.

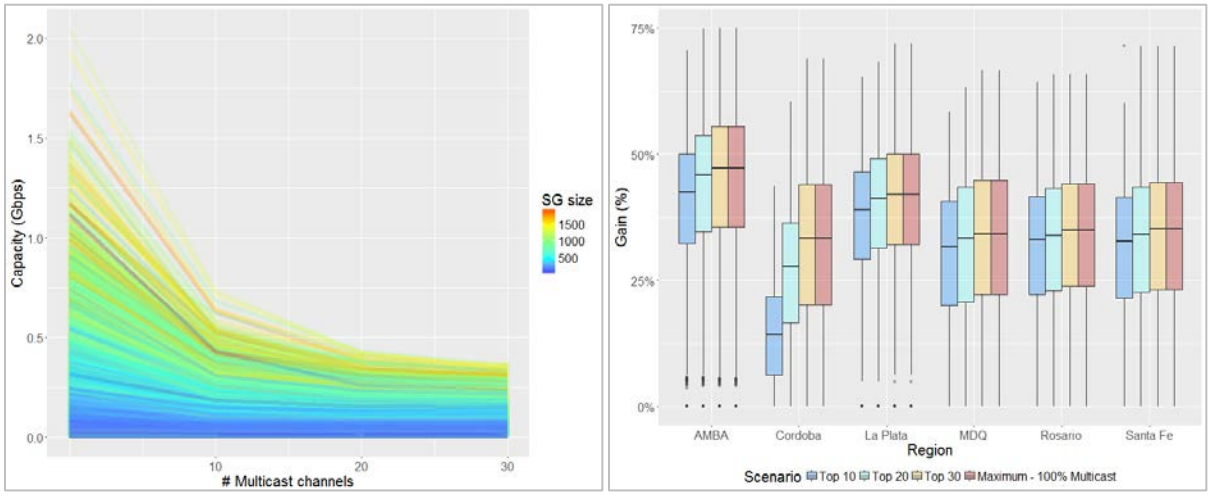

Figure 17-[a] Capacity needed at service group level versus multicast channels count, by SG size. [b] Multicast gain distribution by region and scenario on May 23, 2018 from 9 p.m. to 10 p.m.

From Figure 17 we conclude that when there are more HHP in the SG, there is a need for more capacity to support their activity, but also the decrease after setting channels by multicast is drastic.

In addition, the region variable seems to have a greater impact on gain when the region is Córdoba and there are fewer channels delivered using multicast.

To summarize the relationship between gain, region and area size, we estimated a statistical model [11] that has the general form:

$$Capacity = \exp(\alpha + SG_i + \beta \cdot X_i + \gamma \cdot \log HHP_i + \delta \cdot X_i \cdot \log HHP_i + \varepsilon_i)$$

Where HHP_i represents the HHP count in the *i-th* service group, it should be noted that we used the logarithm to denote the relationship described in Figure 12 [b]. The variable $T10_i$ should be replaced by 1 for the Top 10 scenario –and 0 in other cases–, $T20_i$ should be replaced by 1 for the *Top 20* scenario and $T30_i$ should be replaced by 1 for the *Top 30*case.The term$SG_i$is a random effect, which varies from one service group to the other, and the random errorε_i. This linear mixed effects model is subject to the assumptions of Gaussian distribution of the error and the SG_i random effect, in other words:

$$\varepsilon_i \sim N(0, \Sigma)$$

$$SG_i \sim N(0, \sigma_{SG}^2)$$

What is actually fruitful to get from the model is the parameter estimation, which is shown in Table 3. According to this information, for a SG of size 500 located in Buenos Aires, under 100% unicast scenario, the capacity needed would be of 210 Mbps. For the *Top 10* case, we would need 164 Mbps (46 Mbps less) to sustain all of the STB's activity. With the Top 20 scheme, the saving is on average 83 Mbps, which means that if we add ten more channels, we can save an extra 37 Mbps. Following the same logic, when we add ten more (totaling 30 channels) we reduce the need, on average, by an extra 28 Mbps. It should be noted how the marginal difference between one scheme and other declines when the total count of channels delivered using Multicast increases.

Table 3- Fixed parameter estimation for the mixed-effects model

	Buenos Aires	Córdoba	Other
$\boldsymbol{\alpha}$	1.66	0.50	1.96
$\boldsymbol{\beta}$	0.02	0.02	0.02
$\boldsymbol{\gamma}$	0.59	0.74	0.50
$\boldsymbol{\delta}$	-0.01	-0.01	-0.01

For the case of a SG with the same size (500 HHP) but located in Córdoba, when passing from a 100% unicast scheme to the *Top 10*, the capacity that we would need would be reduced by 29 Mbps (approximately a 60% of what we observed for the Buenos Aires area). The *Top 20* plan would sum another 24 Mbps to the saving –less than what we estimated for Buenos Aires. The *Top 30* scenario would add another 20 Mbps.

In the Córdoba region, the percentage of gain when adding more multicast channels does not decrease as fast as in Buenos Aires. It is a subtle difference, but the explanation for it resides in the fact that Córdoba is a big region yet a very different one in terms of habits. The most viewed

channels may differ to the ones that are popular elsewhere but may still appear in the top 10, 20 or 30.

For the other regions, the interpretation is quite similar to Buenos Aires, except that on average, the impact of multicast implementation would be slightly reduced. Going from all-unicast to the *Top 10* scenario, would reduce the demand on average about 31 Mbps. Adding ten more channels, with the *Top 20* scenario, would add another 25 Mbps to the saving. Then, with the *Top 30* case, this would increase by another 20 Mbps. More details about this example can be found in Table 4.

Table 4–Estimation of the capacity (Mbps) for a 500 HHP service group, by multicast channel count and region.

Multicast channels\Location	Buenos Aires	Córdoba	Other
0	210	163	159
10	164(-22%)	134(-18%)	128(-19%)
20	127(-17%)	110(-15%)	103(-16%)
30	99(-13%)	90(-12%)	83(-13%)

Table 5 and Table 6 show capacity estimations for service groups of size 128 and 64, respectively. Throughout the examples, it is clear that when the SG is smaller, the gain –as a percentage– decreases. In addition, Córdoba is increasingly different from the rest of the regions. It should be noted that the examples in Table 4, Table 5 and Table 6 correspond to the blue lines in Figure 17 [a].

Table 5 - Estimation of the capacity (Mbps) for a 128 HHP service group, by multicast channel count and region.

Multicast channels\Location	Buenos Aires	Córdoba	Other
0	94	60	80
10	81 (-14%)	54 (-10%)	70 (-13%)
20	70 (-12%)	49 (-9%)	61 (-11%)
30	60 (-10%)	44 (-8%)	54 (-10%)

Table 6 - Estimation of the capacity (Mbps) for a 64 HHP service group, by multicast channel count and region.

Multicast channels\Location	Buenos Aires	Córdoba	Other
0	62	36	57
10	56(-9%)	34(-5%)	52(-9%)
20	51(-8%)	32(-5%)	47(-8%)
30	46(-8%)	30(-5%)	43(-7%)

The percentage of gain in the 128 HHP service group in Córdoba is practically the same as in the 64 HHP in Buenos Aires. Therefore, we conclude that not only the service group size but also its location determines the multicast gain. In order to boost the gain to its maximum, the regional channels –especially the ones that are popular in Córdoba– should be considered.

6. Real-Time Analytics

According to Gartner's definition: *"**Real-time analytics** is the discipline that applies logic and mathematics to data to provide insights for making better decisions quickly. For some use cases, real time simply means the analytics is completed within a few seconds or minutes after the arrival of new data. **On-demand real-time analytics** waits for users or systems to request a query and then delivers the analytic results. **Continuous real-time analytics** is more proactive and alerts users or triggers responses as events happen".*

We observed certain consistency in the rankings through time, we identified the sports channels and we know they should be delivered using multicast as the best practice. In addition, we believe it is convenient to consider the channels that are popular in other regions –especially Córdoba– and that may not appear on top of the general ranking. Therefore, it seems to be a better strategy to set a policy driven multicast approach and use the real time analytics to monitor its functioning.

In this section, we propose the idea of continuous real-time analytics in order to create alerts related to linear TV channels that are not delivered using multicast but in some cases –due to a major event– they would need to be. In order to do so, we looked for an unsupervised machine-learning algorithm to detect changes in user behavior, so that it allows us to make a decision about how to adjust the multicast plan.

6.1. K-means Clustering

Cluster analysis is a concept that encompasses a variety of machine learning techniques that aim to group a set of units or objects so that the ones within the same cluster have similar characteristics and the clusters are as different as possible from one another. This technique is unsupervised, which means that data is not previously labeled; it is used to find hidden underlying structure in the data.

One of the most well-known clustering algorithm is *k-means*. It is very useful to execute exploratory analysis on a large number (millions) of cases, when we want to classify them but the classes are unknown a priori. It has been applied in the past on pattern recognition, image analysis, data compression, among others.

A good cluster analysis has two main characteristics:

- *Efficient:* uses as few clusters as possible.
- *Effective:* captures all statistically and commercially important clusters.

The k-means method seeks to minimize the distances between the observations in the same cluster, and maximize the distances to observations in other clusters [12][13].

The algorithm consists on the following steps:

1. Place K points into the space determined by the variables measured on the units that we want to cluster. These points represent initial group centroids.
2. Assign each unit to the group that has the closest centroid.
3. After assigning all units, recalculate the positions of the K centroids.
4. Repeat Steps 2 and 3 until the centroids no longer move.

There is a variety of definitions of the distance that can be used to execute a k-means analysis. In this paper, we use the Euclidean distance.

6.2. K-means Clustering applied to the selection of multicast channels

In this application case, we will look for two groups: one that contains the channels with higher access frequency –that should be delivered using multicast, – and the other with the rest of the channels –that should remain delivered using unicast.– This means that the parameter has to be$K$=2.

If we wanted to split the complete channel list into unicast, multicast and variable multicast, we would set K=3. Provided that we found that the top channels are most of the time the same, we do not want to have a variable section, and we only want to monitor that the top channels are delivered using multicast in all cases, so the results showed in this section were obtained for K=2.

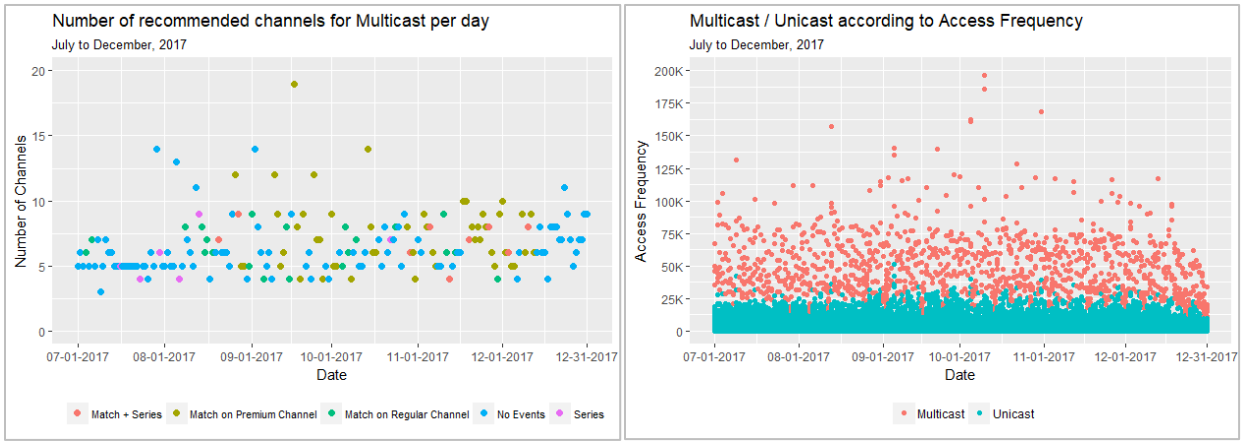

Figure 18- [a] Size of the cluster that groups the high access frequency channels - multicast cluster- by date, colored by type of event. [b] Access frequency versus date, channels colored by cluster. Data from July 1, 2017 to December 31, 2017.

Figure 18 shows part of the results obtained after running k-means on six months of data. On the left, figure [a] shows that the sample size of the multicast cluster varies mainly between 4 and 9 channels. Major events do not influence the number of channels that should be set to multicast; this situation is expected since the algorithm only takes into account the access frequency. On the right, we show a daily detail. As it was previously stated, a few channels capture most of the views. To understand what channels they are, we expose a sample with Flow data and another with Legacy data, in Figure 18.

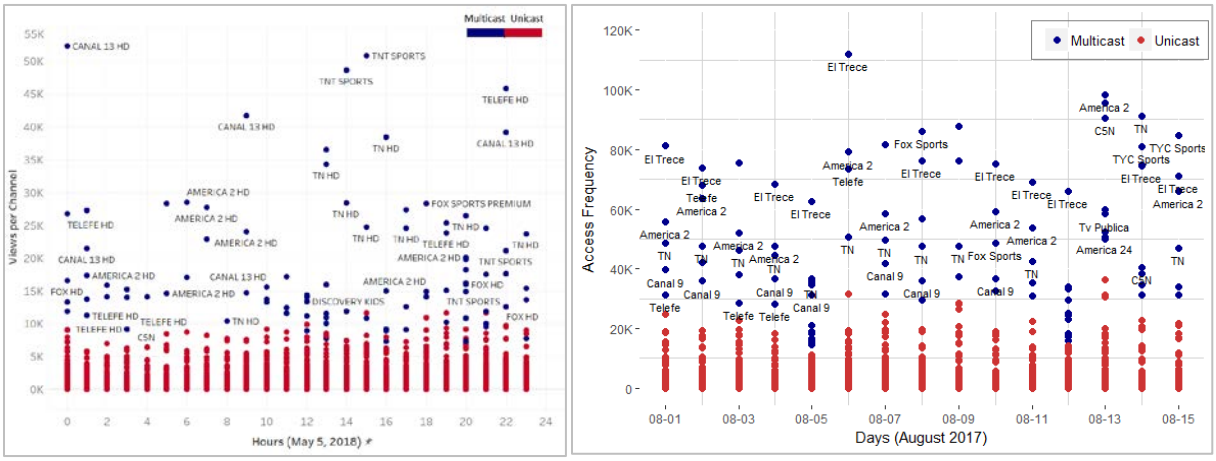

Figure 19-[a] K-means clustering applied to the views per channel by hour for OTT devices. [b] K-means clustering applied to the access frequency per channel by day for the Legacy system. Algorithm used to classify the signals between multicast and unicast. Blue dots represent multicast channels and red dots unicast.

Note in Figure 19 that the names of the channels grouped as multicast tend to repeat execution after execution. In [a], common labels are *Telefe, Canal 13, America 2, TN, TNT Sports* and *Fox Sports.* In [b], which is from a different date and the k-means was executed based on the daily ranking, we see some repeated channels: *Telefe, El Trece* (which is the same as *Canal 13*), *America 2, TN, Canal 9* and *Fox Sports*. This leads us to think that channels that concentrate the most views have little variations in time, and a continuous monitoring system would be good enough to keep the system on track.

After taking into consideration these observations, we suggest the following algorithm:


```
Every 10 minutes repeat:
     Calculate {Cluster_Unicast; Cluster_Multicast}
     If Cluster_Multicast not in Multicast_fixed then:
          PrintCluster_Multicast
```

The clusters should be based on the concurrence variable, calculated using the k-means method and K=2.We would like to clarify that this is a theoretical design and the implementation of the alerts represents a new challenge.

Conclusion

Through the analysis of user behavior, we have found that the proportion of Live TV versus VoD tunings, as well as the most viewed channels vary according to time slot, region and the device used by the subscribers. The percentage of VoD tunings is around 20% on the Flow platform, and below 1% on the Legacy system.

The distribution of concurrent views on regular days follows a Zipf-Mandelbrot distribution, as observed in related works. Major events modify the ranking, and tend to alter this distribution. When one of these events takes place, it increases the difference between the most tuned channels and the rest.

In a more general approach, we have used the Spearman's coefficient to study the relationship of rankings from different dates, for a six-month period. We have found that the correlation is, most of the time, high (above 50%), except when a sports event or a series is aired.

The following conclusions about multicast gain are based on the subscribers' behavior while using the STB. The variation of the device may introduce alterations.

After estimating the multicast gain at CDN level, we observed that when there are around 10 channels delivered using multicast, the gain is around 50% during the busy hour. If more channels are delivered using multicast, the marginal gain tends to decrease. We found that with 25 channels delivered using multicast, the gain approximately reaches its maximum.

We studied the relationship between multicast gain at SG level and its size. We found that, given the variation of channels' popularity among regions, the gain is not only conditional to the STB count but it also depends on its location. For a SG located in Córdoba, the marginal gain is lower than a similar-sized SG in Buenos Aires. Meanwhile, in other regions, the tendency is the same as in Buenos Aires but on a lower scale, the absolute capacity and gain are in all cases smaller due to smaller SG. We concluded that regional channels should be taken into consideration to boost the gain.

In order to explore the channel count that would be delivered using multicast if it was an automatic and unsupervised process, we executed the k-means algorithm on ranking data. After analyzing six months of data, we found that this technique grouped between 4 and 9 channels as the most popular.

Given the high correlation between rankings from different days, the sports channels being already identified and considered for multicast as the best practice, and the fact that it would be necessary a complex process –hence non-scalable in real time– to capture the regional specificities, it seems inconvenient to apply real time analytics for a viewership driven multicast approach.

Nevertheless, real time analytics provide an efficient alternative for monitoring a policy driven multicast approach, since there could be a special event not considered so far (not a sporting event, or a popular TV series) which could drastically shift the ranking for a few hours and then return to usual.

We proposed a continuous process, which consists of a k-means clustering algorithm to be executed every 10 minutes. The program looks for the channels that get the most views and checks whether they are included in the multicast channel list. In case there is one or there are more channels that are being accessed aggressively, and do not belong to the multicast list, it sends an alert and reports the list of the most popular channels.

Abbreviations

ABR	adaptive bitrate
avg	average
bps	bits per second
CDN	content delivery network
DOCSIS	data over cable service interface specification
DSL	digital subscriber line
FTTH	fiber to the home
HD	high definition
HFC	hybrid fiber coaxial
HHP	household passed
HTTP	hypertext transfer protocol
IP	Internet protocol
IPTV	Internet protocol television
ML	machine learning
NDVR	network digital video recorder
OTT	over the top
QAM	quadrature amplitude modulation
SD	standard definition
SG	service group
STB	set top box
VOD	video on demand

Bibliography & References

[1] S. Deering, "Host Extensions for IP Multicasting," Network Working Groupb-RFC 1112 , Stanford University, August 1989.

[2] Cable Television Laboratories, "IP Multicast Adaptive Bit Rate Architecture Technical Report," OC-TR-IP-MULTI-ARCH-C01-161026, October 26, 2016.

[3] Ron Reuss, "IP Unicast v. Multicast Modeling Overview," CableLabs, Liousville, Colorado, September 2012.

[4] Kunwadee Sripanidkulchai, Bruce Maggs, and Hui Zhang, "An analysis of live streaming workloads on the Internet," in *Proceedings of the 4th ACM SIGCOMM conference on Internet measurement (IMC '04)*, ACM, New York, NY, USA, 2004.

[5] Guillermo Wilkinson, "IP Video Topics," in *Seminario Internacional de Telecomunicaciones - SIT 2016*, Buenos Aires, March 2016.

[6] J. Horrobin and G. Shah, "Pioneering IPTV in Cable Networks," in *SCTE Cable-Tec Expo*, October 2013.

[7] Ulm and P. Maurer, "IP Video Guide - Avoiding Pot Holes on the Cable IPTV Highway," in *SCTE Cable-Tec Exp*, October 2009.

[8] Weisenborn, Hildebrand J, "Video Popularity Metrics and Bubble Cache Eviction Algorithm Analysis," PhD thesis, University of Essex., http://repository.essex.ac.uk/22350/, 2018.

[9] Amit Eshet, John Ulm, Uzi Cohen, Carol Ansley, "Multicast As A Mandatory Stepping Stone For An IP Video Service To The Big Screen," in *NCTA/SCTE Technical Sessions*, spring 2014.

[10] S. Sprent and N. C. Smeeton, Applied Nonparametric Statistical Methods.--3rd edition., Chapman & Hall/CRC, 2001.

[11] G. M. Fitzmaurice, N. M. Laird and J. H. Ware, Applied Longitudinal Analysis, Hoboken, Hudson, U.S.: J. Wiley & Sons, 2004.

[12] J. A. Hartigan, Clustering Algorithms, [New Haven, Estados Unidos]: John Wiley & Sons, 1975.

[13] D. Peña, Análisis de datos multivariantes, España: McGraw-Hill Interamericana de España S.L., 2002.

[14] H. Yu, D. Zheng, B. Zhao, and W. Zheng, "Understanding User Behavior in Large-Scale Video-on-Demand Systems," in *In Proceedings of EuroSys2006*, Leuven, Belgium, 2006.

[15] M. Cha, H. Kwak, P. Rodriguez, Y. Ahn, and S. Moon, "Analyzing the Video Popularity Characteristics of Large-Scale User Generated Content Systems," *IEEE Transactions on Networking*, vol. 17, p. 1357–1370, October 2009.

