

Original Research Paper

Implementation of IP Video Streaming Software to Identify Availability and Audience

^{1,2,3}Vitor Chaves de Oliveira, ⁴Sérgio Bimbi Junior,
^{3,5}Andreiwid Sheffer Corrêa, ⁶Inácio Henrique Yano,
²Mauricio Becker, ¹Paulo Batista Lopes and ¹Gunnar Bedicks Junior

¹Mackenzie Presbyterian University – UPM, Brazil

²Pontifical Catholic University of Campinas – PUC, Brazil

³Federal Institute of Education, Science and Technology of São Paulo – IFSP, Brazil

⁴Institute for Technological Research/University of São Paulo – IPT/USP, Brazil

⁵Institute of Mathematics and Statistics/University of São Paulo – IME/USP, Brazil

⁶Embrapa Informática Agropecuária, Brazil

Article history

Received: 20-03-2018

Revised: 23-05-2018

Accepted: 07-06-2018

Corresponding Author:

Vitor Chaves de Oliveira

UPM, PUC, IFSP,

Brazil

Email:

vitor_chaves_oliveira@yahoo.com.br

Abstract: The perpetual rise of the video on demand is currently one of the leading challenges the telecommunications industry faces. It per passes the eternal comparison with a service that continuously set's the bar at a highly elevated consumer quality. And the user, advertiser and all stakeholders involved not only are used to it, but demand equal and/or similar value, i.e., Broadcast Television. Such dichotomy has made this relatively new medium create a long list of technologies to make this as viable as possible. However, the solutions only work to a certain extent and critical problems remain not yet addressed. One in particular is delivery assurance in Internet Protocol networks, which affects every stakeholder on these New Media outlets. Keeping this issue in mind, this work developed a range of experiment scenarios through a software-based apparatus in order to convey a technical assessment of key variables in this ecosystem. Afterwards, from these tests an analysis was conducted which brought about a series of discoveries in regards to the technology performance in terms of availability and audience. Ultimately, it culminated in one the central contributions of this research, that is, how to mathematically interpret this type of data indicating its statistical relations. Such methods unveiled impacts and feasibility of: Privacy protocols, mobile and landline connection, latency, delay, loading time, interactions volumes, content history, channel characteristics and user attributes. In other words, this work developed a tool to measure audience and availability in IP delivery. And, it also forged a method to interpret and model the measurements into statistical patterns that can provide predictability. Additionally, it is stated that the significance of this research was confirmed through the Law of Large Numbers, which showed that the data has statistical validity to interpret and envision behavior. That is, the work presents data with a reliability of 97% with a margin of error of 3.03% which confirms this, to the best of our knowledge, as the most comprehensive accurate study of this nature in comparison to the state of the art in the literature. It is also necessary to ascertain that the tool is restricted to measuring quantities related to each content displayed for the largest platform of online video and the most desktop utilized Web Browser, i.e., YouTube and Google Chrome.

Keywords: Audience, Connection Availability, IP Stream, Data Analysis, Software Assessment

Introduction

Obtaining sustained advance through quality-like Broadcast Television (TV) is the outmost wish for the Internet Media companies such as Netflix, Daily Motion, Facebook, YouTube, Amazon Prime Video, HBO GO, Instagram and so on (Statistic Brain, 2016; de Oliveira *et al.*, 2016; Sharma and FitzGerald, 2014; Oliveira *et al.*, 2015a; 2015b; TECHNAVIO, 2014; Trend Micro, 2015; Frost, 2014) and, an industry so new, with a little more than half a decade, is customarily rivaled with what they brought to the scenario we live in today, motion picture as seen over the tube. Reminding that audiovisual content was only known and dominated by TV until their arrival. So now, people can watch “TV” and even “broadcast” from and to anywhere at any given time as they choose to. However good and compelling as these promises sound, making it work is not as simple as the new industry would like. One of the major difficulties is to rely on a fragmented environment where each provider is in charge of monitoring and consequently evaluating itself. Therefore, this research aims to mitigating this moral hazard starting the accountability with a software assessment to capture relevant information from IP Streams. Then, a data analysis was carried out identifying mathematical trends and relationships. It is also necessary to mention that this paper conducted the widest study, to the best of our knowledge correlated (Krishnan and Sitaraman, 2013), in the IP delivery field. It provides a tool to monitor 16 key metrics and segments it into 6 different scenarios, while providing the coding methods and a mathematical modelling to interpret results (Buyya *et al.*, 2008; Oliveira *et al.*, 2013a; 2013b; Chai, 2014; Bedicks Jr., 2008; Cloonan and Allen, 2011; Edwards, 2013; Ellacott, 2014; Greenfield, 2012; Kaduoka, 2016; Katel, 2014; Statista, 2016; McMillan, 2013)

Firstly, it was required to expand the functionalities of an existing software that monitors YouTube videos accessed via Google Chrome Browser for desktop in order to capture all the intended variables. It is worth mentioning that the authors of this article also developed the earlier version of this tool. After the development phase, the next one was to define scenarios for the experiments. Then, with these setups, numerous tests were performed to assure statistical validity. Subsequently, the data was analyzed to unveil which characteristics and to which degree they affect each other as well as in what manner they relate to the IP video delivery. And, it is noted that such exploration was focused on the influences on content availability and audience (de Oliveira *et al.*, 2016; Statista, 2015a; 2015b; YouTube, 2017).

Quality in Video Transmission

The quality of video in packet networks, i.e., IP networks, has been a constant concern for a satisfactory content flow. This section seeks to discuss these concepts.

In computer networks, that is, Internet protocol networks, Quality of Service is a set of technical conditions that permeate a qualitative operation for a given application that is based on an IP network to function (Oliveira *et al.*, 2013a; Baldini *et al.*, 2014; Tanenbaum and Wetherall, 2010). These requirements translate into specific QoS parameters in computer networks, such as throughput, latency, jitter and losses (Oliveira *et al.*, 2013b; Khanafer *et al.*, 2014). The first parameter, known as throughput, refers to the amount of data transmitted per time interval, measured in bits per second (bps) or bytes per second (Bps). Latency, or delay, indicates the time used for packet transmission from the point of origin (or node) to reach the destination point (or node). While jitter corresponds to the time of variation of the delay, that is, the time difference spent by data packets to travel from the transmitter to the receiver. Meanwhile the losses correspond to the detriment of data that occurs in the transmission that detract from bit-segment information. Losses can be represented by the Bit Error Rate (BER), Packet Error Rate (PER), or Frame Error Rate (FER), perceive that all these rates are generally measured in percentage (%) (Oliveira, 2012; Nunes, 2011).

Availability

All these parameters are related to the concept of availability. Essentially, it precedes it, so that for any of these parameters to be measured, the connection must first be available (Doeven, 2013). Therefore, it is established that the concept of availability is classified as the first concern with quality, even for basic service standards. In this way, it becomes prominent to define and quantify this. And, a mathematical calculation of probabilities is how this statistic is estimated and only through actual experimental measures that can be provided (N-TRON, 2014; Oliveira *et al.*, 2015a).

Audience

To recognize the concept of audience it is indispensable to recall that it is a ramification of Quality of Experience. So, when we check the definition of Quality of Experience as dictated by the International Telecommunication Union (ITU), the following is observed:

“The overall acceptability of an application or service, as perceived subjectively by the end-user.

NOTE 1 – Quality of Experience includes the complete end-to-end system effects (client, terminal, network, services infrastructure, etc.).

NOTE 2 – Overall acceptability may be influenced by user expectations and context.” Completely extracted from Definition of Quality of Experience (QoE) (ITU, 2003).

The aforementioned parameters of QoS, throughput, latency, jitter, losses and availability, transform the effects of sensitivity and perception of a video transmission, affecting what is assessed as Quality of Experience that the end user intuitively and receives. However, it is emphasized that QoE, although closely linked to QoS, is not exclusively understood by Quality of Service, since it is something subjective, that is, there are variations from one end user to another (ITU, 2003).

Moreover, it is worth noting that in the IP Stream context the issue of availability is consistently evidenced, since it demonstrates that the quality of the audiovisual content signal is representative in audience engagement. And if the availability is deteriorated, the loss of views accompanies this trend consistently. This is evidenced with data and mathematically correlated (Krishnan and Sitaraman, 2013) and, for copyright reasons and for using corporate information, this information and graphics cannot be reproduced in this article.

The concepts listed so far in this study are closely related to the definition of audience for audiovisual content, described below:

*"Meaning of Audience
s.f. Meeting of individuals (listeners or spectators) who by assumption, by research, by estimation or by statistical proof, attend certain television or radio programs. "
Extracted integrally from Meaning of Audience (DICIO, 2016).*

With this mind, to effectively measure the audience of content transmitted via TV Broadcast, a device called People's meter or audience meter is used. This is installed in the home of certain viewers connected to the television, it continuously registers and sends information to a central station on which channel is being watched, at what time and for how long. These users are chosen to have such devices installed in their homes so that all of them represent a portrait, a sampling, of the whole society or the segment of society that one wishes to measure. This entire equipment scheme is called Television Rating Points (TRP), which aggregates this entire measurement system. Thus, the amount of success of an audiovisual content is statistically estimated. The importance of measuring the audience is given because it is the engine of the commercial part of what programs a broadcaster should invest and the impact or reach of advertising campaigns. Therefore, a relevant commercial interest for the

purposes of pricing transmission periods in this and other segments such as IP Streaming exist (KANTAR IBOPE MEDIA, 2018; GFK, 2018).

Materials and Methods

The method used in this research is detailed in this section are the construction of control mechanisms and data analysis for video on demand in the IP ecosystem. In addition, it is necessary to point out that this research is supported and based on the named DSR, or 'Design Science Research' and 'The Sciences of the artificial' that deal with the construction of artifacts to solve problems of the real world, thus representing software, algorithms, methods, analyzes, information systems and entities form a scientific method (Dresch *et al.*, 2015; Simon, 1996).

Software for joint Capturing of Audience and Availability of Video Stream

Requirements Analysis

A solution was designed considering the aforementioned importance of obtaining information at the user's end on videos delivered via IP from the largest provider of online media on the Internet, i.e., YouTube (Statista, 2015a; 2016). In addition, it has been delimited that the application is compatible with one of the most used browsers currently, Google Chrome (Statista, 2015b). Considering the requirements, in order to capture the audience of an audiovisual content, a project was decided upon in which the program should be composed of two parts, described below. Besides, it is emphasized that this is a newer and more complete version of the previous software developed by these same authors (de Oliveira *et al.*, 2016) with increased functionalities and features aimed to conceive a more comprehensive solution. The justifications for these implementations are also listed as follows.

The client part, or Front-end, consists in an application to be installed in the browser through an extension format that could be installed on any device with the Google Chrome browser enabled for extensions. It collects and sends certain data to a centralizer. These data are statistics (measurements and information) of videos accessed, that is, viewed by users. Note that this is done from data in public site codes when a page with a YouTube video is accessed, at which time the application becomes active. Such data selected to be captured and sent automatically to the centralizer are: - Total time with the page open; This variable chosen to evaluate how much the user is willing to stay on a page to view content is both associated with audience and availability.- Time to load video; This parameter is associated to measure the initial availability for the content to begin to be seen.- Watched video time; This

metric demonstrates the strong relationship with the audience as it denotes the time the content was seen.- Time with video in Pause; This time is intimately related to the issue of content availability and the audience, as it demonstrates how much a user is willing to wait for content to be viewed. In addition, it allows more in-depth analysis and opens up question so that other possibilities could have occurred as the user paused the video so that it loads, how much to separate by pure forgetfulness when it accesses other content or by pausing too much over the time to load the desired content.- Total Video Time; This metric is needed to compare how long that video was watched in percentage.- Number of Interactions in the Timeline; This number gives us indications referring to availability, because if a user had to make many interactions with the content is something relevant to be observed. As well as tangencies the question of the audience for the case of the person want to watch several times a same video or the same portion.- Post-Video time; This variable combined with the one of the total time collaborates so that the time relative to that video separates, as well as delimits the question to the matter of the availability because this time can be pre-loading of a next video if the person is with some type of automatic reproduction activated. - URL, Universal Resource Locator, i.e., Internet address of the video; Knowing what content is or is not available and whether or not a satisfactory audience was reached are requirements for the business chain involved in delivery via IP.- Visualization, number of views and, consequently, the number of the specific visualization; This metric is necessary to be observed, since it collaborates in the strategy of the Content Delivery Network (CDN) (TECHNAVIO, 2014; Rouse, 2011) and also to preload a demanded content, thus meeting expectations of the ecosystem treated. - Channel, i.e., the producer and/or owner of the content; Knowing which channel proves itself interesting, since it lets you know that even if a content may not yet have many views on this platform, it may have high demand because it is from another provider (e.g.: Consolidated TV and Radio channels and programs, Digital influencer pages on Facebook, etc.) and this must be observed. - Subscribers, that is, the number of registered users that the Channel has; This number values the number of people who will be contacted if the channel launches a new video, so if the channel has many subscribers it is usually because its audience is high, this affects the strategy of the providers for availability.

In addition, other relevant data is captured through a brief questionnaire to be completed by the user when the software is installed. These are:- City; Knowing the physical location of the user helps to

evaluate the connection, since QoS is subjective this can contribute to the strategy issue of CDN (e.g.: A user that accesses from a metropolis versus a user that accesses a remote region). - Age; Since the shift from traditional TV to the Internet TV is generational, that is, the younger the customer the higher the demand is for Internet content, the more availability that person expects. - Gender. Since access to YouTube videos is 68% male and 32% female (Statistic Brain, 2016; YouTube, 2017) it is necessary to know which part of the tester's base we need to increase. That is, if we need more users of which gender.

Additionally, all captured data has a time stamp so that they can be analyzed separately in time, as for example in the debut of a content.

The part of the centralizer, or server, that receives, stores, manipulates and displays all of this data, is called the Back-end. There is a software that acts as a server on the Internet, which will provide a service to receive, save (record) and query the captured data, segmenting them into database tables.

During the software requirements analysis, it became necessary to produce a modeling of the desired solution. The objective was to design the relationship between coding, database, processes involved and communication based on the premises of the UML - Unified Modeling Language. There are several definitions for this method, for the purposes of this project the design was considered in diagram design as proposed in its generic premises (Arthemias, 2017; OMG, 2017; Alhir, 2003).

Coding and Development

To develop such software that is capable of interacting with the API, Application Programming Interface, or YouTube Application Programming Interface, the JavaScript language has been selected. And to store the data we used a MySQL (Oracle Corporation, 2015) database and hence the SQL language to manipulate and treat the data. The following are the main parts of the coding that has been implemented for the Front-end which is the extension installed in the consumer device's audiovisual content browser.

In Fig. 1, we have programming a code that verifies changes on the page, i.e., in the URL and registers, saving the data locally whenever a change to another page is executed. That is, it captures when the user ceases to access certain content and saves in the cookies of the Browser, at the client's machine all the information is monitored and captured.

As for Fig. 2, there is programming a code that verifies the filling of the user's personal data, at the time of first use of the Extension. That is, soon after installing the extension in the Browser, once you access a page with YouTube videos, the following code asks via

questionnaire to fill in the City, Age and Gender of that user. This questionnaire is filled in once, on the first use, then this data is automatically stored in the browser cookies. Through this code it is also possible to observe that this information is configured to be mandatory, being programmed so that the user does not skip or do not stop filling this information.

The main coding, called the Extension programming core, is shown in Fig. 3. Through it, it is possible to observe from the comment "COLLECTS VIDEO INFORMATION" the existing interaction with the API of the YouTube site, being this the codification that presented the greatest challenge to be discovered and implemented, especially the time information contained

in the script that changes over time, that is, throughout the reproduction of a video. In it, we collected video name, watched time, total time, number of views and number of subscribers in the channel. Previously to this, we have this information defined as variables of this larger function, which has the purpose of writing the data in the Back-End. So that immediately after the data collection, we have the function that forwards all this data to a web service maintained by the site configured for this in the domain of own property, the 'tetconsultoria.com.br'. If the communication fails to send this data, the user is issued a message about this problem stating that the web service was not communicated and it was unavailable to send the data.

```
//VERIFIES IF THE USER UPDATED OR WILL LEAVE THE BROWSER
var allowPrompt = true;
$(document).click(function(event) {allowPrompt = true;});
$(document).ready(function() {
    window.onbeforeunload = WarnUser;});
function WarnUser() {
    if(allowPrompt) {
        //SENDS DATA TO SERVER
        SetData("");return ".";
    }else {SetData("");
        allowPrompt = true;
    }
    if (localStorage.clickcount) {
        localStorage.clickcount = Number(localStorage.clickcount) + 1;
    } else {localStorage.clickcount = 1;}}
```

Fig. 1: Coding that Verifies if the user updated or will leave the browser

```
//CHECKS IF THE USER ALREADY FILLED THE REQUESTED DATA
function SetDataUser(){
    //Verifies if the City field is filled
    if (localStorage.getItem('UserCity') == null) {var UserCity = prompt
    ("YT Stats,\n\nInformation Required:\n\nInform your City:", "New York");
    if (UserCity != null) {localStorage.UserCity = UserCity;
    localStorage.setItem('UserCity', UserCity);
    }else{alert('City Field Must be Filled!');}}
    //Verifies if the Age field is filled
    if (localStorage.getItem('UserAge') == null) {
        var UserAge = prompt("Inform your Age:");if (UserAge != null) {
            localStorage.UserAge = UserAge;
            localStorage.setItem('UserAge', UserAge);
        } else {alert('Age Field Must be Filled!');}}
    //Verifies if the Gender field is filled
    if (localStorage.getItem('UserGender') == null) {
        var UserGender = prompt("Inform your Gender:", "M or F");
        if (UserGender != null) {localStorage.UserGender = UserGender;
        localStorage.setItem('UserGender', UserGender);
        } else {alert('Gender Field Must be Filled!');}}
```

Fig. 2: Coding to check if the user filled the requested data

```
/*COLLECTS VIDEO INFORMATION */ $(document).ready(function() {
VideoName = $('#watch-title').text(); var element = $('#ytp-progress-bar');
if (element.length > 0) {$(element[0].attributes).each(function () {
if (this.nodeName == 'aria-valuenow') {WatchedTime = this.nodeValue;}
else if (this.nodeName == 'aria-valuemax') {TotalTime = this.nodeValue;}});}
ViewsNumber = $('#watch-view-count').text();ChannelName=$('#.yt-user-info').text();
SubscribersNumber = $('#yt-subscriber-count').text();}); WatchedArrayTime += ','
+ TimeCounter + '-' + WatchedTime + '-' + new Date().getDate() + '/'
+ new Date().getMonth() + '/' + new Date().getFullYear()+''+ new Date().getHours()
+ ':' + new Date().getMinutes() + ':' + new Date().getSeconds();
console.log("URL - ", window.location.href); console.log("-----");
closeWindow = new Date().getTime();OpenPageTime =(closeWindow - openWindow)/1000;
LoadVideoTime=(LoadVideoTime-openWindow)/1000; if (TotalTime != "0"){
//SENDS DATA TO WEBSERVICE
$.ajax({type: "POST",url: 'https://www.tetconsultoria.com.br/WS.asmx/SetStats4',
data: "{ VideoName:'" + VideoName.trim() + "',"+
"WatchedTime:'" + TimeCounter + "',"+ "TotalTime:'" + TotalTime + "',"+
"URLVideo:'" + urlPlay + "',"+ "ViewsNumber:'" + ViewsNumber + "',"+
"ChannelName:'" + ChannelName.trim() + "',"+
"SubscribersNumber:'" + SubscribersNumber + "',"+
"UserCity:'" + localStorage.getItem('UserCity') + "',"+
"UserAge:'" + localStorage.getItem('UserAge') + "',"+
"UserGender:'" + localStorage.getItem('UserGender') + "'," +
"OpenPageTime:'" + Math.round(OpenPageTime) + "'," +
"PausedTime:'" + PausedTime + "'," + "PostVideoTime:'" + PostVideoTime + "'," +
"WatchedArrayTime:'" + WatchedArrayTime + "'," +
"LoadVideoTime:'" + Math.round(LoadVideoTime) + "" + " }",
contentType: "application/json; charset=utf-8",
dataType: "json",success: function (msg) {},error: function (e) {
alert('Message WS03 - WebService unreachable')}});
WatchedTime = 0; PreviousWatchedTime = -1; WatchedArrayTime = "";
urlPlay = '';TimeCounter = 0;PausedTime = 0;TotalTime = 0;
PostVideoTime = 0;openWindow = new Date().getTime();}}
```

Fig. 3: Coding that collects video information and sends to web service

For the Extension and this software to function properly, a continuous check of which page the user is, to which changed, whether or not he has changed, etc. in order not to lose any content in performed. This, because the information captured by the previously described codes are captured and sent to the centralizer are made to each page with video accessed automatically. That is, follows a sequence of monitoring, capturing and sending it to the Web Service. Thus, to achieve this goal of checking changes on the page, both to send the information already collected and to initiate new collection, the methods shown in Figure 4 below were implemented. In this code, you have to verify that URL has been changed by user request, by automatic means of the YouTube site itself, when the continuous and automatic played of videos in sequence is configured on this site, since it is by default on the site. And an additional one has to check over time every 100 milliseconds if the page has changed. This encoding

triggers the sending of all information as soon as a change occurs and it is only at this moment of change that they are sent.

The back-end programming part is illustrated in the image of Figure 5. In it, there is the encoding that implements the web services server that accepts, receives and inserts into the SQL database the return, i. e., the information collected from all videos accessed from all users who have the client installed on your browser. The programming demonstrates ASMX, an 'Active Method for File Server' which is the Web Service of the .NET language, method implemented in this part of the solution that implements the page and provides the referred service. Next, appears the request that accesses the MySQL database and allocates all the data. Thus, in this insertion the tabulation, parameterization in columns in a table created in the database for the information collected from the clients.

```
(function() { /*VERIFIES CURRENT URL*/ var strLocation=window.location.href;
var strHash=window.location.hash;var strPrevLocation="";var strPrevHash="";
//FREQUENCY TO CHECK THE URL
var intIntervalTime=100;
var fnCleanHash=function(strHash){return(strHash.substring( 1,strHash.length));}
/*VERIFIES USER DATA*/SetDataUser();
//YOUTUBE DOES PAGE TRASITION IN TWO WAYS: INTERNAL AND EXTERNAL. FOR INTERNAL LINKS
//USES AJAX AND THE CAPTURE OF THIS INFORMATION IS DONE USING THE CODE BELOW
//METHOD TO CHECK CHANGES AT THE WINDOW LOCATION
var fnCheckLocation=function(){/*VERIFIES IF THE LOCATION CHANGED
if (strLocation != window.location.href){
//SENDS DATA WHEN CAPTURES PAGE CHANGE VIA AJAX
SetData("INSIDE");// STORES NEW AND PREVIOUS LOCATIONS
strPrevLocation=strLocation;strPrevHash=strHash;strLocation=window.location.href;
strHash=window.location.hash;
// THE LOCATION CHANGED, TRIGGERS CHANGES AN EVENT IN THE LOCAL OBJECT
// INFORMING CURRENT AND PAST LOCATIONS VALUES
$( window.location ).trigger("change",{currentHref: strLocation,
currentHash: fnCleanHash( strHash ),previousHref: strPrevLocation,
previousHash: fnCleanHash( strPrevHash )});}}
// SETS AN INTERVAL TO CHECK LOCATIONS CHANGES
setInterval( fnCheckLocation, intIntervalTime );
})( jQuery );
```

Fig. 4: Coding to verify URL every 100 ms

```
using System;using System.Collections.Generic;using System.Linq;using System.Web;
using System.Web.Services;using System.Collections;using System.Web.Services.Protocols;
using System.Web.Script.Services;using MySql.Data.MySqlClient;
/*WEBSERVICE DEFINITION*/ namespace WService{[WebService(Namespace=
"https://www.tetconsultoria.com.br/ws.asmx/setstats1",Name = "WSExtension")]
[WebServiceBinding(ConformsTo=WsiProfiles.BasicProfile1_1)] [System.ComponentModel.ToolboxItem(false)]
/*METHOD FOR WEBSERVICE*/[System.Web.Script.Services.ScriptService]
public class WS:System.Web.Services.WebService{[WebMethod] public string SetStats(string VideoName,
string WatchedTime,string TotalTime, string URLVideo, string LoadVideoTime,string PausedTime,
string InteractionsNumber,string ViewsNumber, string ChannelName, string SubscribersNumber,
string UserCity,string UserAge,string UserGender, string PostVideoTime,string OpenPageTime)
{string msg="";//DATABASE ACCESS*/using(MySqlConnection conexao = new MySqlConnection
("Server=MYSQLE5015.Smarterasp.net;Database=db_9f5aee_system;Uid=9f5aee_system;Pwd=banco123;"))
{MySqlTransaction transacao; try {conexao.Open(); transacao=conexao.BeginTransaction();
MySqlCommand command=new MySqlCommand();command.Connection=conexao;command.Transaction=transacao;
/*DATABASE INSERT*/command.CommandText="INSERT INTO`db_9f5aee_system`.`youtubestats`(`ysVideoName`,`
`ysLoadVideoTime`,`ysPausedTime`,`ysInteractionsNumber`,`ysWatchedTime`,`ysTotalTime`,`ysURLVideo`,`
`ysViewsNumber`,`ysChannelName`,`ysSubscribersNumber`,`ysUserCity`,`ysUserAge`,`ysUserGender`,`ysData`,`
`ysPostVideoTime`,`ysOpenPageTime`)+"VALUES(@VideoName,@LoadVideoTime,@PausedTime,@InteractionsNumber,@
@WatchedTime,@TotalTime,@URLVideo,@ViewsNumber,@ChannelName,@PostVideoTime,@OpenPageTime,@
@SubscribersNumber,@UserCity,@UserAge,@UserGender,now());";command.Prepare();/*PARAMETERS*/
command.Parameters.AddWithValue("@VideoName", VideoName);
command.Parameters.AddWithValue("@LoadVideoTime", LoadVideoTime);
command.Parameters.AddWithValue("@PausedTime", PausedTime);
command.Parameters.AddWithValue("@InteractionsNumber", InteractionsNumber);
command.Parameters.AddWithValue("@WatchedTime", WatchedTime);
command.Parameters.AddWithValue("@TotalTime", TotalTime);
command.Parameters.AddWithValue("@URLVideo", URLVideo);
command.Parameters.AddWithValue("@ViewsNumber", ViewsNumber);
command.Parameters.AddWithValue("@ChannelName", ChannelName);
command.Parameters.AddWithValue("@SubscribersNumber", SubscribersNumber);
command.Parameters.AddWithValue("@UserCity", UserCity);
command.Parameters.AddWithValue("@UserAge", UserAge);
command.Parameters.AddWithValue("@UserGender", UserGender);
command.Parameters.AddWithValue("@PostVideoTime", PostVideoTime);
command.Parameters.AddWithValue("@OpenPageTime", OpenPageTime);
command.ExecuteNonQuery();transacao.Commit(); transacao.Dispose();
}catch (Exception ex){msg="ERROR ON SAVING DATA, " + msg;}
finally{conexao.Close();} msg = "DATA SUCCESSFULLY SAVED."; return msg;}}
```

Fig. 5: Main back-end programming

Results

Functioning – Software Operation

For the application developed, the user first needs to use the Google Chrome browser to access the Chrome Web Store - Extensions or Chrome Extension Store on the Internet. Once in it, by searching for ‘YouTube Data Stats’, the name of this software, the icon and application described in Fig. 6 will appear as an option to be installed in the user’s browser, available by accessing the address: ‘chrome.google.com/webstore/detail/youtube-data-stats/nmnaacknlpckjppccigekcnogaeoai?’.

Figures 7 and 8 show, respectively, how the user is informed whether the extension is active or inactive on the codes and APIs of the page it is accessing. This occurs when a YouTube video is being played. So in Fig. 7, with the color red extension icon means that the application is

acting, its codes are interacting with those on the page. While in Fig. 8, with the black and white icon we it inactive, which occurs when the user is not accessing a YouTube video.

Captured data from all tests performed by all users is sent to the Web Service, Back-end, accessible through the interface described by the following figure. It is worth mentioning that this data is available on the Internet and can be accessed by anyone with access to the system. In Fig. 9, the following page is displayed after a system-authorized login. In it, the logged in user has the option to click on “Pesquisar” to search and display everything that is present in the configured database to store the information of the collections. The ‘Pesquisar’ performs the selection in the database to display the search just below the word ‘LISTA’, because in this page it is configured to display the selection.

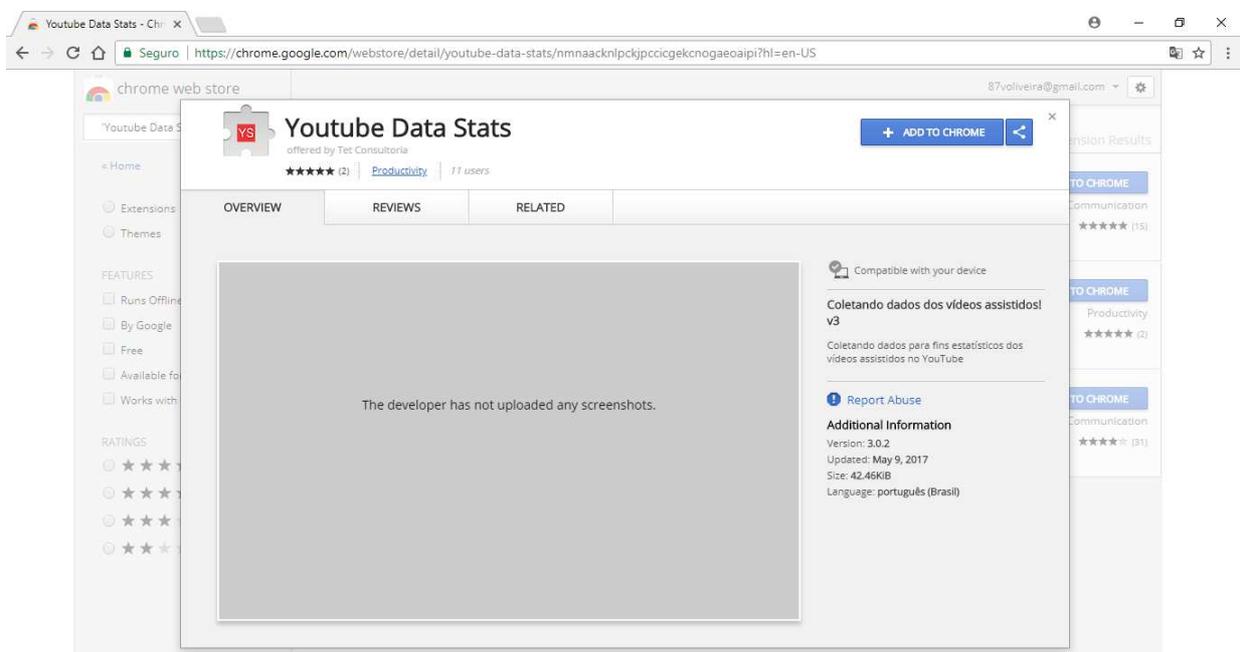


Fig. 6: Page to install the extension from the ‘Chrome Web Store – Extensions’



Fig. 7: Extension icon indicating active application



Fig. 8: Extension icon indicating inactive application

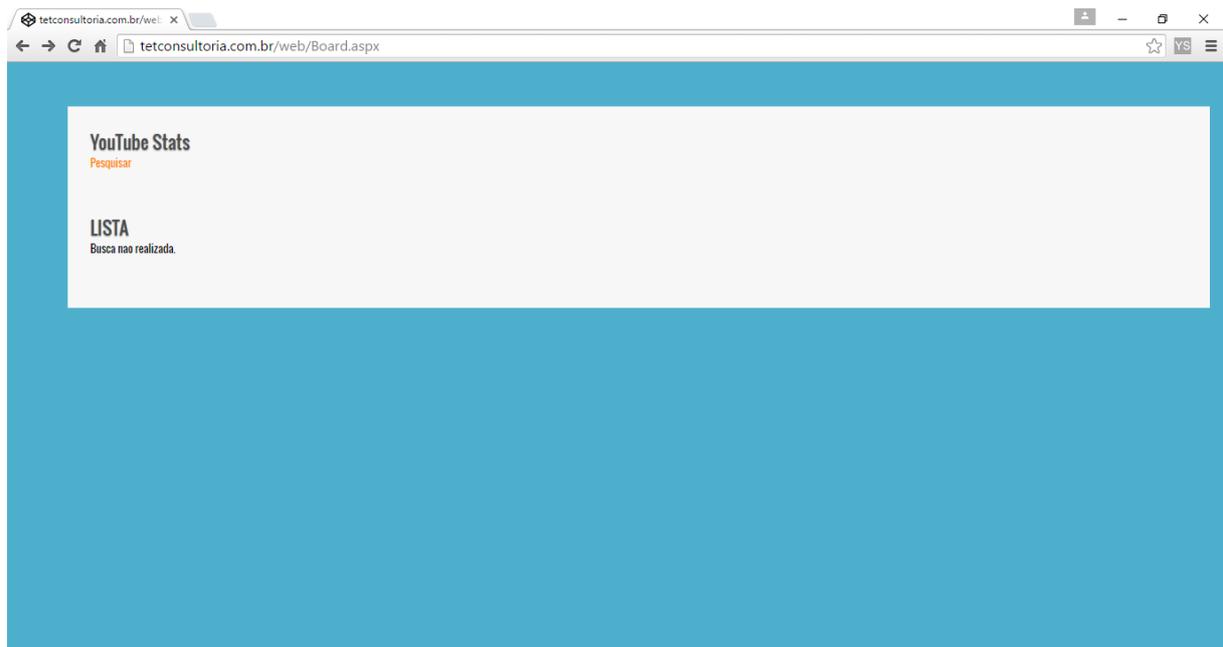


Fig. 9: Input and Search screen for the data captured at the Software back-end

Nome	Tempo Total com a Página aberta	Tempo para carregar o vídeo	Tempo de vídeo assistido	Tempo com vídeo em Pausa	Tempo Total do Vídeo	Qtd de Interações na Timeline	Tempo pós Vídeo	Uri	Visualizacao	Canal	Inscritos	Cidade	Idade	Sexo	Data	Timeline
Florida Georgia Line - Stay 3	3	1	2	0	233	2	0	Link	50.876.957 visualizações	FloGeorgiaLineEVO	1199.879	São Paulo	28	M	2/2/2017	11:20:54 AM
Florida Georgia Line - Anythin...	1993	1	227	1	217	11	1765	Link	1.614.237 visualizações	FloGeorgiaLineEVO	1199.879	São Paulo	28	M	2/2/2017	11:20:51 AM

Fig. 10: Search result screen displaying data captured in the Software Back-end

In Fig. 10, the following page is displayed after clicking on 'Pesquisar', in it we see the result of the search in the database, displaying all the captured data in the form of a table that follows the sequence of the information most recent to the oldest ones sent to the centralizer. From left to right we have, in Portuguese, the columns with Name information, Total Time with the

Page Opened, Time to load Video, Watched Time, Paused Video Time, Total Video Time, Amount of Timeline Interactions, Post-Video time, URL, View, Channel, Subscribers, City, Age, Gender and Date.

Regarding the captured data, it is important to emphasize for the purpose of the audience, the most sensitive data and generators of the evaluation are the

time values collected. That is, the Total Time of the video and Watched Time. This, because if, for example if dividing one by the other already we have an interesting data, that would inform percentage of viewing of a certain video. And with the other data we could relate this in views by date/time, by public, by channel and certainly by specific content. In the examples detailed in Fig. 10, we would have 92 s divided by 325 s and 4 s divided by 325s (sec). With this, we would reach 28.3% and 1.2% of the audience for that viewer of that content at that time. In addition, quality concerns may also be observed, even if partially, by these data. Since in dealing with a subjective question, if the content is of acceptable quality the user watches for longer, if it is not acceptable for him, he will not watch it.

Analysis of Captured Data to Assess Availability

To explore deeper into the issue of availability, that is, how to know if something is available as demanded, i.e., latency, the first data we have is the time to load the Video. Thus, the following metric figures demonstrate how the data with 1281 samples were divided, divided into the metric figures described below.

In Fig. 11, referring to the Time to Load Video - All, we have the arithmetic mean line representing the value of 3.59 s and the initial latency values. That is, the time for content to start showing in the 1281 samples. Of these, 6.32% are equal to or greater than 10 sec, 0.54% are 9 s, 0.54% of 8 s, 0.70% of 7 s, 1.09% of 6 s, 0.54% in 5 s, 1.24% in 4s, 1.79% in 3 s, 9.85 % in 2 s, 48.10% in 1s and 29.29% with 0s. With this, it is noted that 87.24% are in up to 2 sec. Additionally, to the average of 3.59 s, dispersion calculations were performed, obtaining the value of 24.83 s for standard deviation and 616.46 s² for variance. Observing the concentration of values, the weighted, mean and standard deviation calculations were performed, but the values obtained did not show significant changes. Thus, an analysis excluding the samples greater than or equal to 10 seconds, or 6.32% of the data, has an average of 1.09 s, a standard deviation of 1.39 s and a variance of 1.96 s². It is observed an expressive variation in the time to load the videos, presenting different times of bufferization, with tendency of stochastic and increasing variation for data organized over time.

In Fig. 12 we have the graph of the Time to Load Broadband Video, that is, it shows the display of the initial latency for broadband network connection and accesses via Notebook. In this, it is possible to observe the distribution of the data, as well as the comparison with the arithmetic average of 5.01 sec. In addition, dispersion calculations were performed, obtaining a standard deviation of 32.85 sec and variance of 1078.81

s². It is also observed that the tests presented low variation, uniform bufferization, high stability with tendency of slight growth. Despite values higher than the previous graph with the general values, the convergence for stability was expressive.

In Fig. 13 we can see the Time to Load the Video in Connection of Mobile Networks with Static Equipment. This is related to tests with Mobile Networks through a Smartphone Mobile device routing your mobile internet connection in Wi-Fi to the Notebook with both static equipment within an environment built in urban area. And, note that the arithmetic mean was 4.68 sec which was lower than the initial latency for broadband access. However, it presented high variation, besides time for initial buffering with a decreasing tendency of relevant form and with low stability. Given that the computations to observe the distribution of values showed a standard deviation of 13.19s and a variance of 174.10s².

For Fig. 14, the graph titled Time to Load Video in connection of Mobile Networks with Moving Equipment is displayed. In this, as in the previous one, the samples of experiments with Mobile Networks are related through a Smartphone Mobile device routing your mobile internet connection in Wi-Fi to the Notebook. However, in this case, both the Cellular and the Notebook were moving within a motor vehicle traveling through a metropolitan region. Thus, an arithmetic mean of 1.32 sec was pointed out, one of the smallest of the general averages observed in these analyzes. And, following such an arrangement, the mathematically obtained dispersion values showed a standard deviation of 3.49s and a variance of 12.21s². It is worth extracting the interpretation that there was a low variation of the initial loading time of the content, a uniform initial buffering and a high level of stability.

With Fig. 15, or graph of the named No Automatic Playlist, no user logged on to Google, with IP masking and location via Virtual Private Network (VPN). For these tests it is necessary to point out that they were conducted with a broadband connection. The main objective was to demonstrate the impact when the user wants to be anonymous, using robust privacy mechanisms such as those provided by a Virtual Private Network (VPN). This case studied presented the highest arithmetic mean with a value of 7.31 sec and the dispersion calculations followed this increase, being the standard deviation of 57.03s and the variance of 3252.40s². While, for results organized over time, the growing trend is expressive. In addition, it is characterized by the high variation of the initial loading time of the video, by the bufferization of the initial latency increasing and by a low stability.

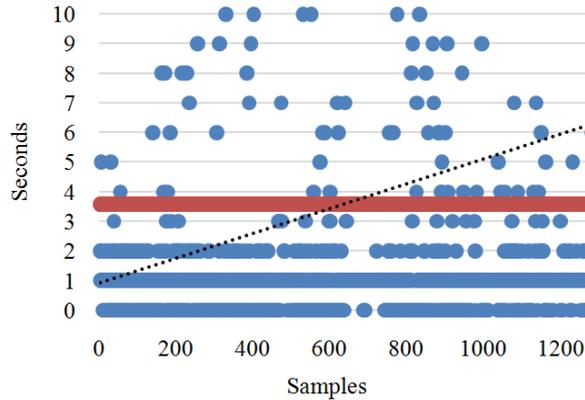


Fig. 11: Time to load video – all; Chart Information Sheet: Time to Load Video - All (Blue); Average (Red); Linear Trend (Time to Load Video - All) (Black); Standard deviation: 24.83s | Variance: 616.46s² | Average: 3.59s | 06.32% ≥10s | 00.54% 9s | 00.54% 8s | 00.70% 7s | 01.09% 6s | 00.54% 5s | 01.24% 4s | 01.79% 3s | 09.85% 2s | 48.10% 1s | 29.29% 0s | 87.24% up to 2s; No significant changes with weighted mean and standard deviation calculations; Excluding the samples ≥10 (06.32%); Standard deviation: 1.39s Variance: 1.96s² Average: 1.09s; Characteristics: Expressive Variation in Time to Load Videos; Different buffering times; Trend of stochastic and increasing variation (for data organized over time); Note: Refers to the initial latency

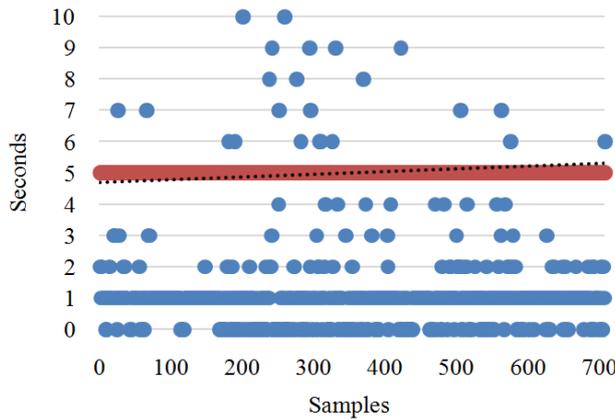


Fig. 12: Time to Load Video via Broadband; Chart Information Sheet: Time to Load Video via Broadband (Blue); Average (Red); Linear Trend (Time to Load Video via Broadband) (Black); Standard deviation: 32.85s | Variance: 1078.81s² | Average: 5.01s; Characteristics: Low Time Variation to Load Videos; Uniform Initial Buffering; High Stability

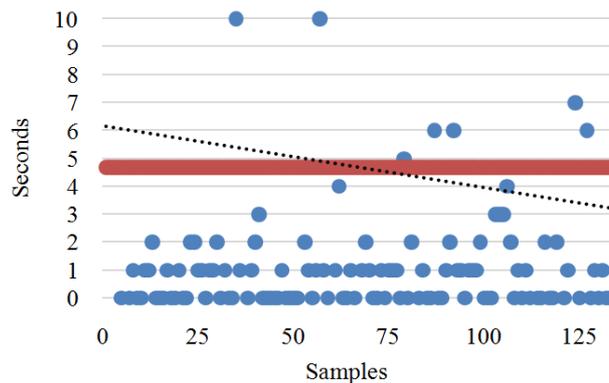


Fig. 13: Time to load video in connection of mobile networks with static equipment; Chart Information Sheet: Time to Load Video in connection of Mobile Networks with Static Equipment (Blue); Average (Red); Linear Trend (Time to Load Video in connection of Mobile Networks with Static Equipment) (Black); Standard deviation: 13.19s | Variance: 174.10s² | Average: 4.68s; Characteristics: High Variation in Time to Load Videos; Initial Buffering Decreasing; Low Stability

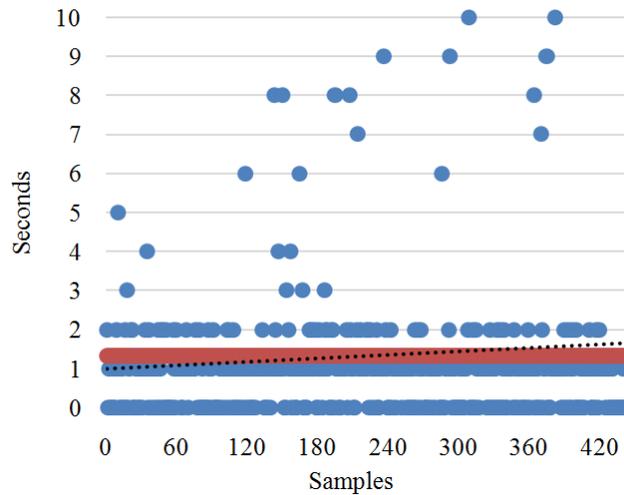


Fig. 14: Time to Load Video in connection of Mobile Networks with Moving Equipment; Chart Information Sheet: Time to Load Video in connection of Mobile Networks with Moving Equipment (Blue); Average (Red); Linear Trend (Time to Load Video in connection of Mobile Networks with Moving Equipment) (Black); Standard deviation: 3.49s | Variance: 12.21s² | Average: 1.32s; Characteristics: Low Time Variation to Load Videos; Uniform Initial Buffering; High Stability

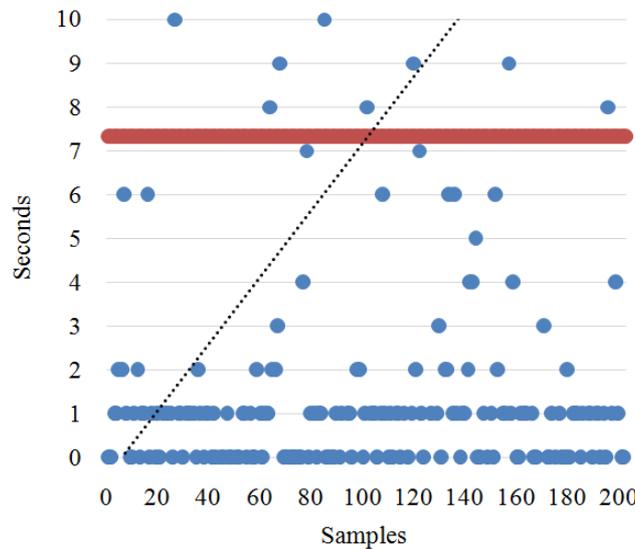


Fig. 15: Time to load Video with Anonymous Browsing - No Automatic Playlist, No User Registered in Google, With IP and locality masking via VPN; Chart Information Sheet: Time to load Video with Anonymous Browsing - No Automatic Playlist, No User Registered in Google, With IP and locality masking via VPN (Blue); Average (Red); Linear Trend (Time to load Video with Anonymous Browsing - No Automatic Playlist, No User Registered in Google, With IP and locality masking via VPN) (Black); Standard deviation: 57.03s | Variance: 3252.40s² | Average: 7.31s; Characteristics: High Variation in Time to Load Videos; Increasing Initial Buffering; Low Stability

Figure 16 shows the graph titled With Automatic Playlist, No User Registered in Google, With IP Masking and Location via Virtual Private Network (VPN). In these tests, we observed the evolution in relation to the study carried out in Fig. 15. This, because Fig. 16 shows the user's discovery, that is, the partial loss of privacy over time, that is, a decrease in the effects of the masking provided by the VPN. Therefore, the mean of these trials was 2.85 sec and the

distribution disposition calculated in 7.81s of standard deviation and variance of 60.96s². Also translating into a low variation of the initial loading time of the audiovisual, uniformity in the bufferization referring to the initial latency and high stability. The overall linear trend, although significantly uniform and stable, shows a slight growth trend over time. Thus, a better result was observed for the availability question, with reduction in the loading period.

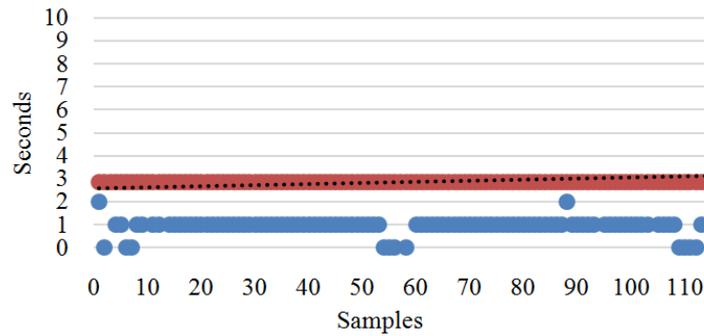
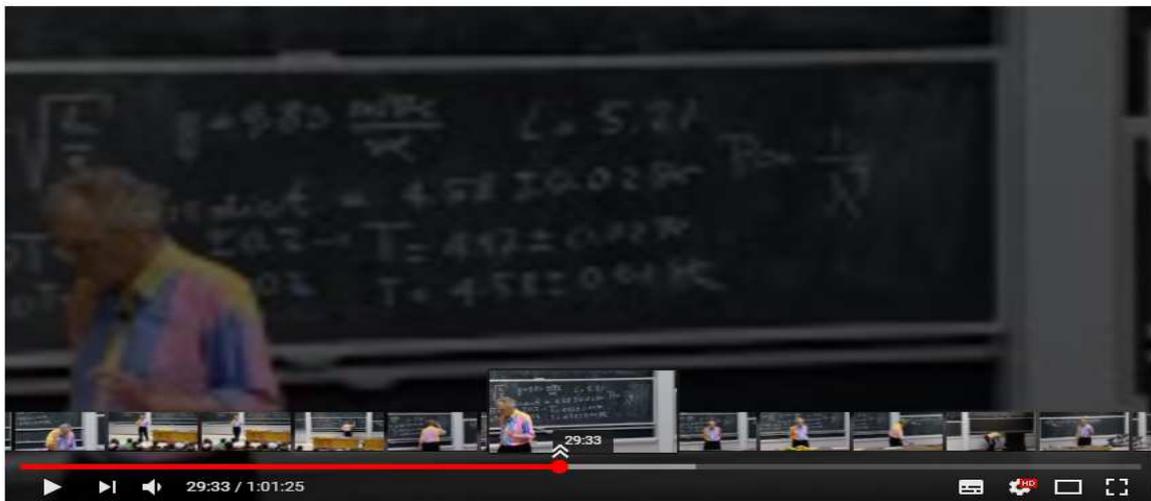


Fig. 16: Time to load Video with Anonymous Browsing - With Automatic Playlist, No User Registered with Google, With IP masking and locality via VPN; Chart Information Sheet: Time to load Video with Anonymous Browsing - With Automatic Playlist, No User Registered with Google, With IP masking and locality via VPN (Blue); Average (Red); Linear Trend (Time to load Video with Anonymous Browsing - With Automatic Playlist, No User Registered with Google, With IP masking and locality via VPN) (Black); Standard deviation: 7.81s | Variance: 60.96s² | Average: 2.85s; Characteristics: Low Time Variation to Load Videos; Uniform Initial Buffering; High Stability



For the Love of Physics (Walter Lewin's Last Lecture)

Fig. 17: Amount of video timeline interactions on YouTube through loaded frames; Source: Reproduced and Adapted from For the Love of Physics (Walter Lewin's Last Lecture) (Lewin, 2014)

Continuing on the issue of availability, the relationship between time with paused video, value that encompasses the sum breaks and user crashes as well as system's ones, with the number of interactions in the timeline. In Fig. 18, Time with Paused Video X Amount of Interactions in the Timeline, relationships are evidenced. In this analysis, the tests evidenced a dependence that the greater the number of interactions with the content, the greater the time with the content paused. That is, more unavailable audiovisual, more crashes and consequent impoverishment of the experience provided to the user. The number of interactions indicates each time the video has been interacted with, whether per user or per system. And it is noted that such a value is increased considerably in the times in which the frames are loaded in the

episodes in which the user and/or the system clicks and/or moves on the time bar of the video. In addition, the value increases proportionately in relation to the amount of frames loaded by the system, such situation is exemplified by Fig. 17.

The analysis of the time data with paused video shows that 53.32% of the tests had 0 sec, 21.08% were in 1s, 0.86% in 2s, 1.56% with 3s, 1.95% for 4s, 1.80% in 5s, 1.87% for 6s, 0.94% for 7s, 1.09% for 8s, 1.17% for 9s and 14.36% for values equal to or greater than 10 sec. A cluster is surveyed with 74.40% of the tests conducted being up to 1 sec. While the arithmetic mean was 78.02s and the dispersion calculations of the data obtained a standard deviation of 949.62s and variance of 901,778.15s². The weighted mean and the weighted standard deviation were also calculated without

significant changes. In this way, the same previously observed cut was performed in order to extract more information from the captured data. That is, in order to observe the behavior of the data concentration, excluding samples greater than or equal to 10s, we obtain an average of 0.95s, standard deviation of 1.92s and variance of 3.69s².

With the study of the number of interactions, it is observed that 2.97% have 1 interaction, 4.29% are in 2 interactions, 29.12% with 3 interactions, 13.19% for 4 interactions, 26.23% in 5 interactions, 6.01% with 6 interactions, 2.11% to 7 interactions, 1.87% in 8 interactions, 1.64% with 9 interactions and 12.57% with 10 or more interactions. It is also observed that 75.80% of samples have up to 5 interactions. And with the calculations of the arrangement and distribution of these data we have an average of 18.08 interactions, standard deviation of 86.56 interactions and the variance of

7,492.75 interactions². Weighted computations were also performed as weighted mean and weighted standard deviation, which did not present any relevant changes. Thus, in order to obtain information about the concentration of data, the values were withdrawn with 10 or more interactions and the arithmetic mean was 4.16 interactions, the standard deviation in 1.55 interactions and the variance with 2.42 interactions².

In addition, a measure of dependence between these two types of data is observed with a positive correlation and a coefficient of 0.06. And, although low, this value presents with a tendency to increase the greater the number of experiments considered in the calculation. In addition, it is observed that when considering a margin of 5% in paused time values or interactions, for more or less, the coefficient tends to present an index referring to a perfect correlation denoting a measure of high dependence.

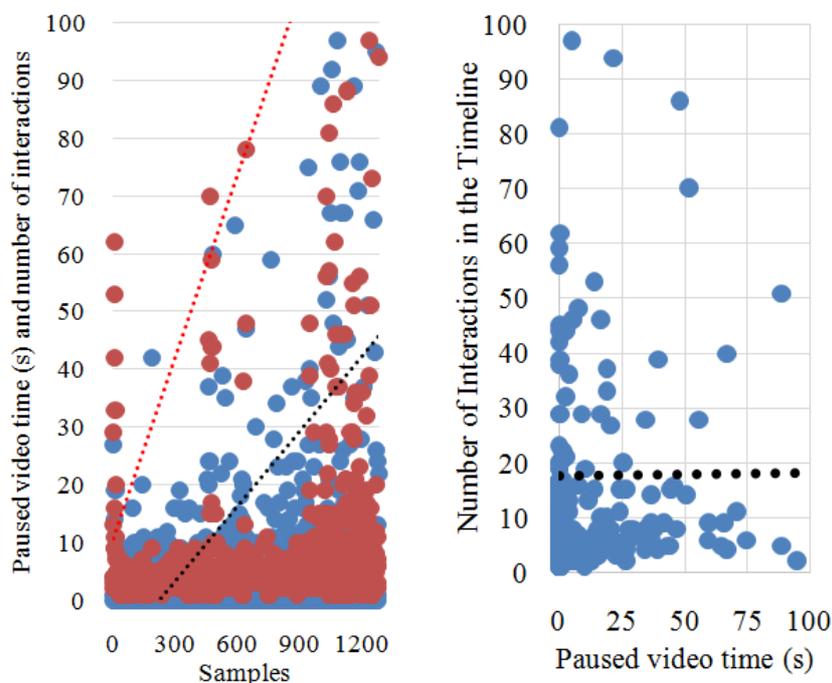


Fig. 18: Paused Video Time X Number of Interactions in the Timeline; Chart Information Sheet - Left: Paused Video Time (Blue); Number of Interactions in the Timeline (Red); Linear Trend (Paused Video Time) (Red); Linear Trend (Number of Interactions in the Timeline) (Black); Chart Information Sheet - Right: Number of Interactions in the Timeline (Blue); Linear Trend (Number of Interactions in the Timeline) (Black); Information Sheet for Pause Video Time: Standard deviation: 949.62s | Variance: 901,778.15s² | Average: 78.02s | 14.36% ≥10s | 01.17% 9s | 01.09% 8s | 00.94% 7s | 01.87% 6s | 01.80% 5s | 01.95% 4s | 01.56% 3s | 00.86% 2s | 21.08% 1s | 53.32% 0s | 74.40% up to 1s |; No significant changes with weighted mean and standard deviation calculations; Excluding the samples ≥10 (14.36%); Standard deviation: 1.92s | Variance: 3.69s² | Average: 0.95s; Information Sheet for Quantity of Interactions (i) in the Timeline: Standard deviation: 86.56s | Variance: 7,492.75s² | Average: 18.08s | 12.57% ≥10i | 01.64% 9i | 01.87% 8i | 02.11% 7i | 06.01% 6i | 26.23% 5i | 13.19% 4i | 29.12% 3i | 04.29% 2i | 2.97% 1i | 75.80% to 5i |; No significant changes with weighted mean and standard deviation calculations; Excluding the samples ≥10 (12.57%); Standard deviation: 1.55i | Variance: 2.42i² | Average: 4.16i; Characteristics: Relationship: The More Interactions The greater the Pause Time. Logo, More Unavailable content, More Crashes and Impoverishment of User Experience; Positive Correlation between Pause and Interactions with coefficient of 0.06 which, with a margin of + -5%, tends to the perfect correlation; Note 1: Paused Video Time encompasses both user and system pauses; Note 2: Amount of Interactions indicates every time the Video has interacted with the Video whether this user and/or system

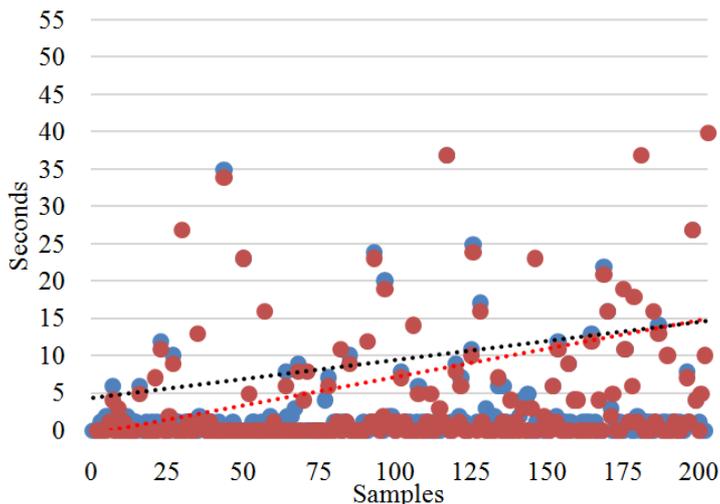


Fig. 19: Time to Load Video X Post-Video Time (VPN w/o Auto Rep); Chart Information Sheet: Time to Load Video (Blue); Post-Video Time (Red); Linear Trend (Time to Load Video) (Red); Linear Trend (Post-Video Time) (Black); Characteristics: Times do not have a positive effect on one another; Relevant relationship; Positive Correlation between Initial Latency Time and Post-Video Time with coefficient of 0.96, close to perfect correlation; As the next videos are chosen randomly, then Video Post Time is not used to Load the next

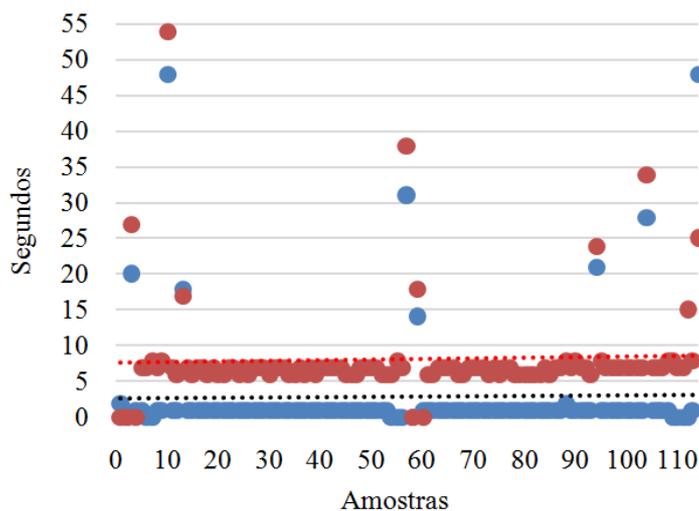


Fig. 20: Time to Load Video X Post-Video Time (VPN with Auto Rep); Chart Information Sheet: Time to Load Video (Blue); Post-Video Time (Red); Linear Trend (Time to Load Video) (Black); Linear Trend (Post-Video Time) (Red); Characteristics: Positive Correlation between Initial Latency Time and Post-Video Time with coefficient of 0.91, close to perfect correlation; Times with significant relationship; Post-Video time reduces loading time; Stability and uniformity of the timings; Note: Dots outside of the trend show crashes. In these it is possible to observe locking to initially load the contents and locking to go to the next video

In the analyzes of Fig. 19 and 20, the following is the comparison between the time to load the video and its relation to the time after video. In them, it is noticeable the sketch that the initial latency time varies with and without automatic reproduction. Since the time to load the next content is lower when the auto play option is selected because it allows the video to be preloaded. Recalling that it is the same tests that the conditions were

elucidated when the metrics of Fig. 15 and 16, respectively, were observed.

When analyzing Fig. 19, referring to the Time to Load Video X Post-Video Time (VPN w/oAut. Rep.), there is an expressive coupling of these two variables. However, the time after the video does not influence the initial latency period of the audiovisual and vice versa. This is because no relationship has been found that can

be considered as being beneficial, but detrimental. Since such a relationship shows that even with high times after the videos the tendency of increase is replicated for the loading time in the beginning of the videos. It is evidenced that the adoption of the total privacy provided through the Private Virtual Network in a schematic without the use of Auto Reproduction, avoided the possible provisioning for a pre-loading and alignment of the transmission. Thus, the value of post-video time, low or high, did not affect positively the growth latency trend during the test period. It is noteworthy that such behavior is due to the randomness with which the videos were chosen, so the time after the video was not used to load the next content. Besides, calculations were made regarding the measure of dependence between the datasets of each time and a high and positive correlation coefficient was obtained, which raises a near perfect correlation with a value of 0.96.

In Fig. 20 of the Video Load Time X Post-Video Time (VPN with Auto Rep.), there is the relevance of the previously cited partial loss of privacy or user discovery by the system. This is evidenced by the expressive relationship between the two variables denoting time, which denotes that both the activation and consequent follow-up of AutoPlay along with Video Post Time act to reduce the Time to Load the Initial Video. Such a decrease contributes to a high stability and uniformity of observed loading periods. Likewise, it is more clearly noted that the detriment of privacy contributes to the transmission being at better levels of availability. In addition, the correlation obtained is close to perfect with a value of 0.91, the difference is that in this case it worked beneficially collaborating to attenuate the loading time.

Remaining in the issue of availability, respectively, in Fig. 21 and 22 with the graphs the Number of Channel Subscribers, the Number of Views against the video's initial latency, i.e., the time to load the page. With this, it observes the influence that these two indicators can have on availability.

For Fig. 21, there is graphical formatting of the Number of Views X Time to Load the Video. Being that the number of times the content was attended presents beneficial effect acting to reduce the loading time. And in this way, it observes that the more visualizations an audiovisual has, the shorter is the trend the period of time it will take to start the reproduction. It is seen that videos with a Number of Views starting at 200 million have the time concentrated in 0.9 sec. Additionally, it is observed that videos with a number equal to or greater than this number of views tend to be within 0 to 2 sec of initial loading latency. It is recalled that 29.29% of the tests had 0s, 48.10% were in 1s and 09.85% with 2 sec.

While in Fig. 22 it demonstrates the Number of Subscribers X Time to Load the Video. In this context, the number of enrollees presented significant influence from 2.5 million with values for initial latency time with

a concentration in 1.0 sec. In a significant way, to establish the tendency of the data sets to be proportionally reversed in some way. That is, in the sense that when the number of users enrolled in a given channel increases, the initial load time tends to be lower, being in the range of up to 2 sec. It is important to remember that up to 2s are 87.24% of the tests.

In addition to the above, it is noted that the influence of both the amounts of views and subscribers is relative, being strongly linked to a status of 'on the rise' or flagged as 'hot'. A prominent content and/or channel is in a growing of the above amounts in a recent short period of time, usually within 1 week. The more current the 'hot', the greater the influence of these quantities, that is, the effect is greater when a recent increase is identified, which classifies the audiovisual and/or channel as 'on the rise'.

Still on the issue of availability the following comparisons are interesting to fully understand it. In Fig. 23, the Total Time with the Page Open and the Total Video Time. In this although there hasn't yet been shown the Watched Time valuation, there is already a glimpse of the audience, perceiving that the less time on the page, the lower the chances of satisfactory audience, in terms of time viewed. In this analysis, a dropout rate with a high level, but lower than the rate reported by the provider, of 90% of evasion up to 10 sec is verified, with 30.68% of users remaining up to 10s. It can be seen that the Total Time of the videos follows the tendency informed by the provider in question with 59.09% of videos lasting up to 300 sec, or 5 min. In the figure of metrics analyzed, it is necessary to point out that the coupling in the center of the graph in the values above the sample of number 600 and below that of value 800. That is, the almost equality of times occurs in for the trials with reproduction set automatically, when the videos were viewed completely. It is also noticed that the concentration of Total Time with the Page Open is in 210s, or 3.5 min, which is the average of these times. And considering a margin of 1 min, for more or less at this time, there are 49.09% of the videos that were shown, excluding those that showed evasion of up to 10 seconds, for videos of up to 299 sec, including (YouTube, 2017).

Figure 24 refers to the Total Time with the Page Open minus the Watched Video Time compared to the Post-Video Time. This provides insight into how long the content required by the user took, i.e., was more likely to be unavailable for delivery. This is because since Post-Video Time is low in comparison, as can be seen below, this total time with the open page subtracted from the Watched Time informs how long the user waited for the content without receiving it, as well as the variation of this from sample to sample. Furthermore, it is observed that the Total Time with the Page Open minus the value of the Watched Video Time corresponds to the Post-Video Time. It presents a positive correlation

of 0.16. Such a value, however low, indicates a behavior. Therefore, with a margin of 5%, for more or less of one set or another, it offers a tendency of perfect coefficient in regards to a measure of dependence. This correspondence collaborates with the discovery of Post-Video Time, which is a relevant factor of availability for cases such as automatic reproduction and provision of

suggested videos on the side, selection that as we observe if followed brings benefits for availability. Also, it is realized that such a mathematical operation provides the data after the video and vice versa. In this way, it supplies the data even without collecting it and serves to check for errors as well as to check the consistency of the captured values.

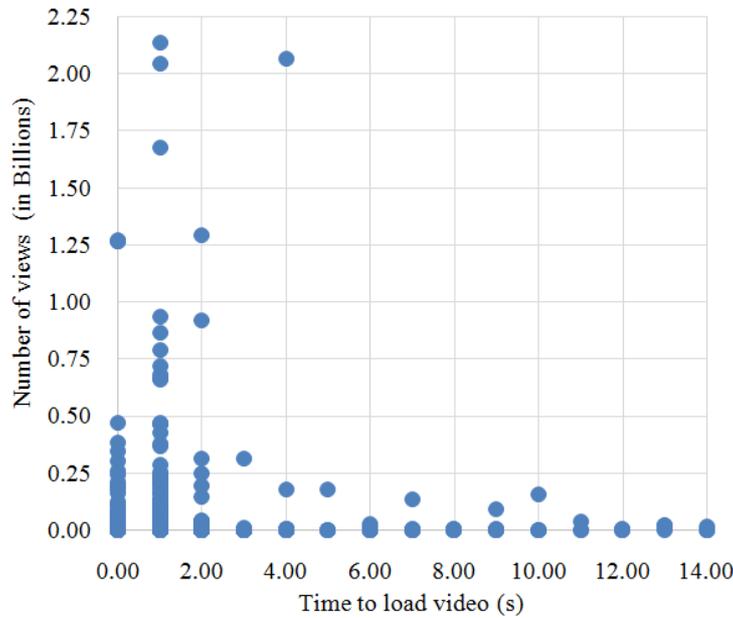


Fig. 21: Number of views X time to load video

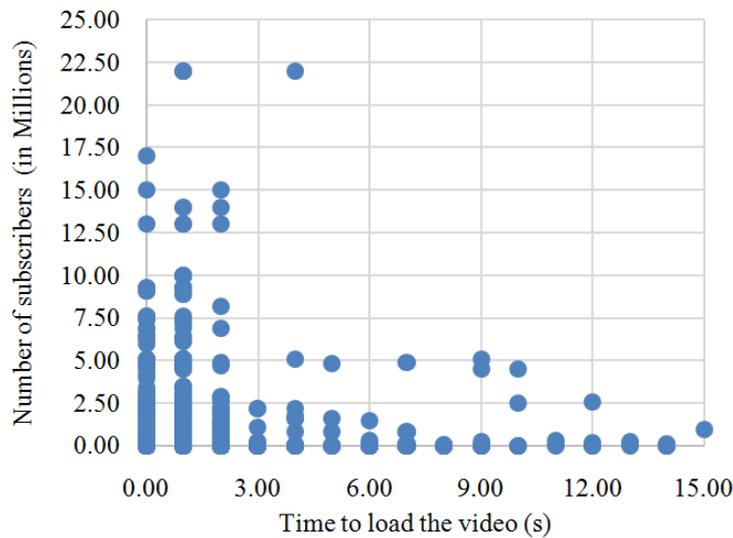


Fig. 22: Number of subscribers X time to load the video; Information Sheet of Metric Figures 21 and 22: 09.85% 2s | 48.10% 1s | 29.29% 0s | 87.24% up to 2 sec.; Characteristics: Influence is relative: If the content is ‘hot’ with a recent increase in views/subscribers, the trend is for content to be available; Both present positive and relatively inversely proportional influence on initial latency availability; Views: for videos with 0.2 billion (and above) views the Time to Load the Video is concentrated in 0.9s (approx.); Subscribers: for 2.5 million subscribers (and above), load times are concentrated in 1.0s (approx.); Number of Views shows greater on availability influence if considered the average loading times; Number of subscribers denotes greater influence on availability, considering its absolute values

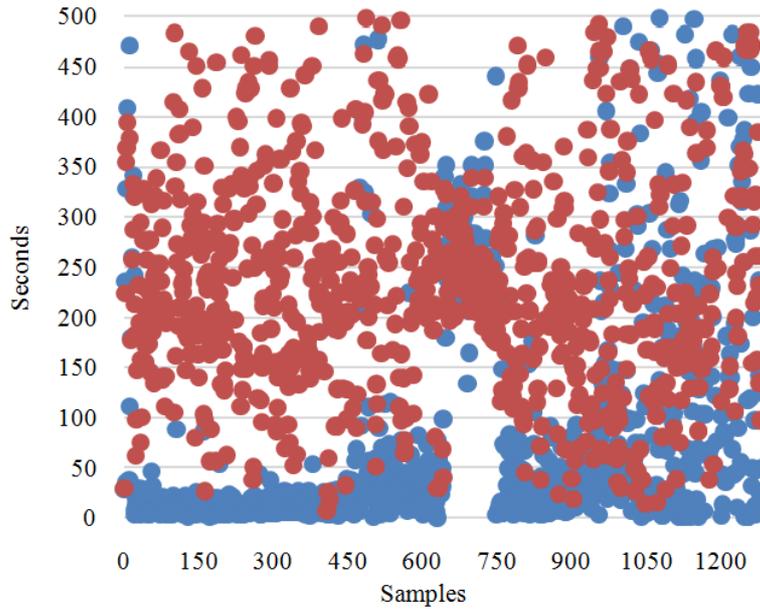


Fig. 23: Total Time with Page Open X Total Video Time; Chart Information Sheet: Total Time with Page Open (Blue); Total Video Time (Red); Average Total Time with Open Page without values that have evaded up to 10s, for videos of up to 299s, inclusive: 210s; Characteristics: Total Time Concentration with Page Open between 150s and 270s, inclusive. 210s with 60s margin; Abandonment Rate: 30.68% leave the page within 10 seconds. Provider: 90%; Total time of videos viewed follows the trend reported by the provider: 59.09% have up to 300s; The junction occurs in a significant way during the tests with auto reproduction, at which time the videos were fully watched; Note: The audience is glimpsed

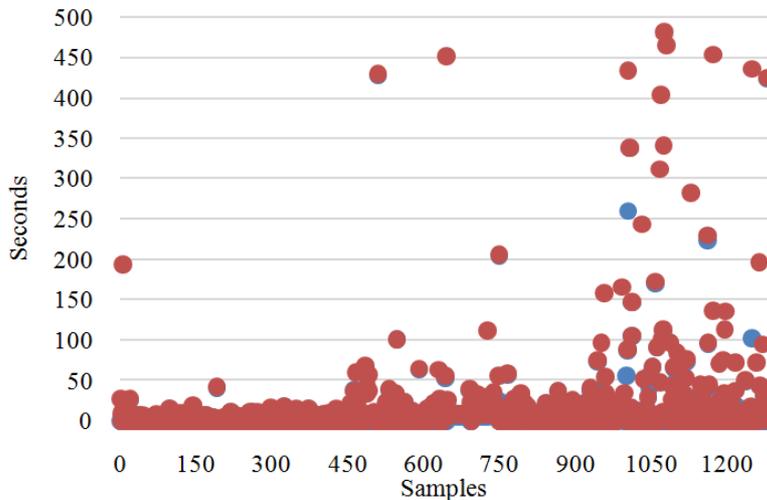


Fig. 24: Total Time with Open Page Minus Watched Video Time X Post-Video Time; Chart Information Sheet: Total Time with Page Open Minus Watched Video Time (Red); Post-Video Time (Blue); Characteristics: Total Time with Page Open Minus Watched Video Time is practically equal to Post-Video Time; Positive Correlation between 'Total Time with Page Open Minus Watched Video Time' and 'Post-Video Time' with coefficient of 0.16 which, with a margin of + -5%, tends to the perfect correlation; This correspondence collaborates with the discovery of Post-Video Time, which is a relevant factor for availability for cases such as automatic played selected and followed with video delivery efficiently; The subtraction provides the possibility of obtaining the same data without collecting it, or if it has an error, as well as checking consistency of captured values

Continuing with this question of evaluating how much IP Stream content is available, it is compared in

Fig. 25, Total Time with Page Open, Watched Video Time and Paused Video Time. In this graph we show

the relation between the times to conceive an understanding about the availability. Once the period with paused video is the added time of all user and system interruptions and blinks after transmission starts. In the chart it is advised how these times vary and how long the user waits the content, that is, the duration of the unavailability.

In completion, it is seen that the content was latently variable to the point of being perceived by the user. This is because the paused time includes intermittent locking, episodes where the page was loading, whether caused by the system and/or by the user. The trends of pause times in ascending order from lowest to highest are shown: VPN AutoPlay, data located in the center of the graph from test 600 and before 800; Broadband, located on the far left from 0 to less than 450; VPN Without AutoPlay, is to the left of the

center between more than 450 to minus 600; Mobile Networks Static Equipment, to the right of the center being from the sample number 800 and less than 940; Mobile Networks Moving Equipment to the far right of 940 onwards. In addition, the measures of dependence between the open page and time periods showed a positive correlation coefficient with a value of 0.51. The proportional gap of this value has a tendency to equate to the paused time, i.e. the difference of the period with the page open with the watched time tends to result in pause time, near perfect positive correlation of 0.88.

Analysis of Captured Data to Assess Audience

In order to delve deeper into the audience issue, it is first necessary to make primary analyzes of checking the total percentage of all videos with their full size.

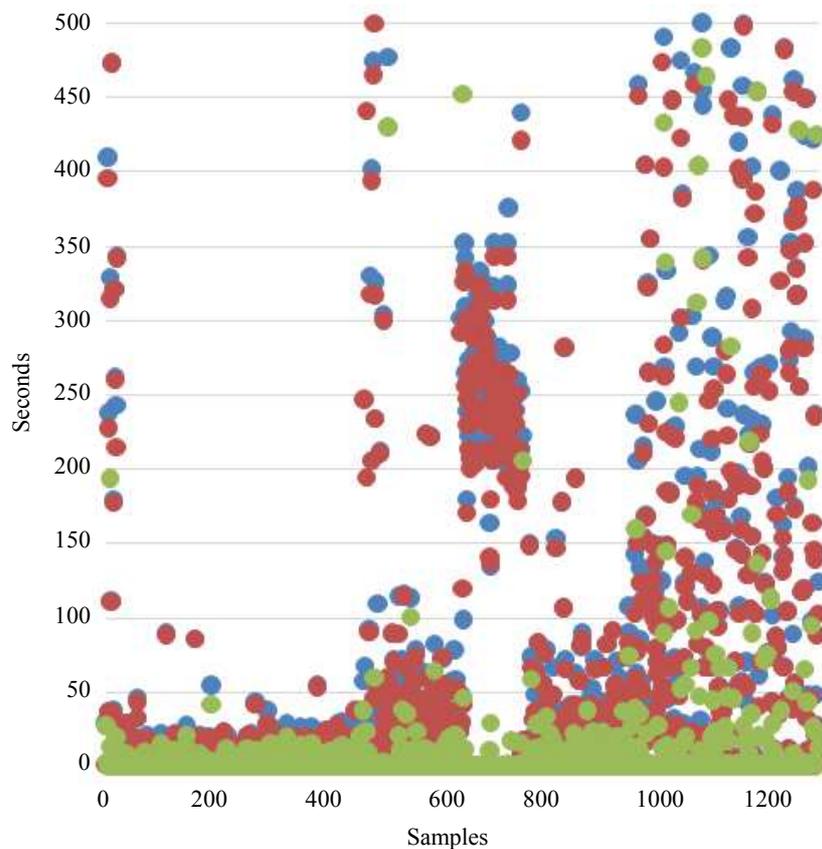


Fig. 25: Total Time with Page Open X Watched Video Time X Paused Video Time; Chart Information Sheet: Total Time with Page Open (Blue); Watched Video Time (Red); Paused Video Time (Green); Positive Correlation Between Total Time with Page Open and Watched Video Time with coefficient of 0.51; Time with the Open Page less Watched Video Time and Paused Video Time have near perfect positive correlation equal 0.88. The values are expressively close, there is a high tendency for the content to remain in pause; Characteristics: Latency of duration significantly variable to the point of being perceived by the user; Pause time, in green, is the summation of intermittent system and user locking; Pause time in ascending order, from lowest to highest break: VPN Autoplay (center of graph) <| Good (Best); Broadband (left) <| Reasonable Availability; VPN No AutoPlay (left center) <| Weak Availability; Mobile Networks Static Equipment (right of center) <| Bad Availability; Mobile Networks Moving Equipment (right) <| Poor Availability (Worst)

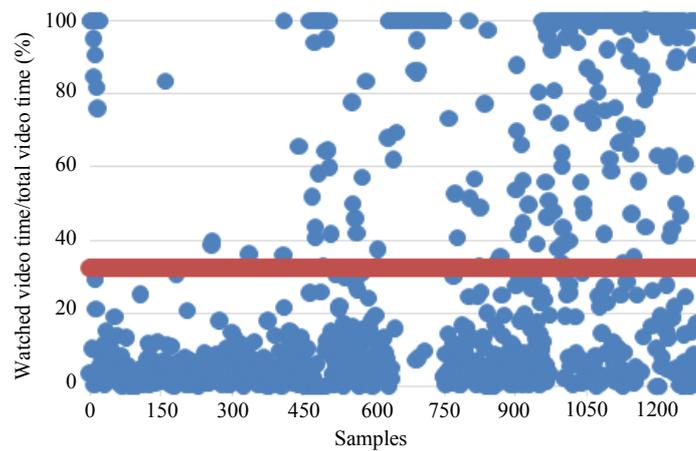


Fig. 26: General Audience; Chart Information Sheet: Audience (Blue); Average (Red); Characteristics: The retention of the user was shown to be greater than that reported by the provider, however with a certain relationship, since the levels of abandonment were relevant; The average general audience of the provider can be estimated at up to 13.33% per weighted average and that obtained was 31.54%; For 35.25% there is evasion in up to 10 sec; With 40.11% viewing is up to 4% of the total duration of the video; And 57.41% assist up to 10% of content; While 27.92% attend at least 50% of the total duration of the audiovisual accessed; Note: It was verified that this hearing was not altered by the automatic reproduction tests, since the other VPN tests neutralized this possible elevation

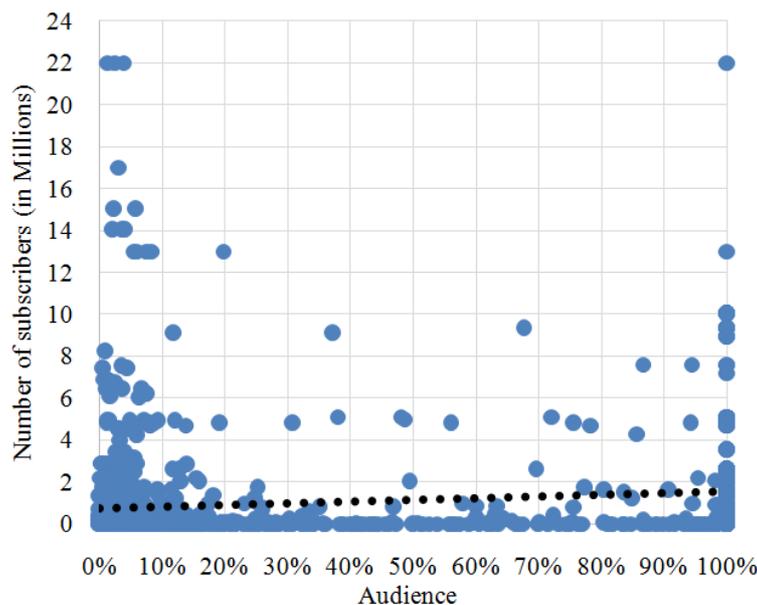


Fig. 27: Number of Subscribers X Audience; Chart Information Sheet: Number of Subscribers (Blue); Linear Trend (Number of Subscribers) (Black)

With this, we have that by the samples, the content had 31.54% audience as shown below by Fig. 26 which demonstrates the sum all watched video times divided by the total time of the video and gives us the individualized audience of each video percentage. It is necessary to remember that 90% of videos viewed on YouTube, do not have 10 seconds of retention and the average size of videos is 4 to 5 min, we can combine these various sources in 270 sec. Thus, the audience of 90% of the

videos would be up to 0.037 or 3.70%. As well as, we can estimate the average overall audience of the provider by up to 13.33%, considering the remaining 10% with 100% audience. Thus, when comparisons and analyzes are performed, it is concluded that user retention has been shown to be greater than that reported by the provider, although with a certain relationship. This, because 35.25% have evasion, in relation to the total time effectively watched, in up to 10 sec. And 40.11% attend to up to 4%

of content. Already 57.41% presented with up to 10% of watched time, while 27.92% viewed 50% or more the video displayed (Statistic Brain, 2016; YouTube, 2017).

Continuing on the question of audience it is seen in Figure 27 the effects of the number of subscribers that the channel, which published the video, has. In this schematic it is evident that this number of subscribers provides influence in the sense that the bigger this number is, the greater the audience. That is, the trend presented is of relatively proportional growth. Since the calculation of the dependence measure between the observed data sets was denoted by a positive correlation coefficient of 0.13, a beneficial but low dependence is thus shown. These results follow an identical justification presented in the issue of the availability regarding the effects of the value of subscribers. That is, the effect observed by the relationship is clearly expressed when the content is classified as 'hot', with the aforementioned recent increase. And, when looking at the graph, as well as the details of the samples, it is possible to see that the contents on the left, even with large numbers of subscribers, did not reach a significant audience. The results of the study show that these contents are old in general, published months and/or years ago and were not classified as 'on the rise'. The results on the right, in particular from 60% of the audience and 1.5 million subscribers, are typified as recent, published up to days and/or weeks ago and/or also because they present larger amounts of videos 'hot', whether new or not.

For Fig. 28, one can see the influence of views amount in an audiovisual over the audience. Such variable, similar to the number of subscribers, the number of views has favorable effects to increase the audience. Since in the dependency calculations there is a positive correlation coefficient with a value of 0.07 which, although classified as weak indicates that the effect exists. However, as in the case of the subscribers number, the influence is moderately relative, depending not only on high values, but whether the content is in a recent upward view, or 'on the rise'. The graph in Fig. 28 shows a trend that the higher the number of views, the larger the audience. However, this effect proves to be expressive from the tests with 85% of audience and 0.3 billion views, it is also relevant to see in this section the increased presence of content classified as 'on the rise'.

Standing at the point of the factors that can be reached over the audience, Fig. 29 examines the Total Time with the Page Open. In this analysis, it is clear that the more time invested in content open, the slightly higher the chance of a high audience rate. Such a characteristic is robustly expressed as can be seen in the trend line present in the said graph. However, it is emphasized that the increase in the audience is not proportional and/or direct, since this value of time when high can also indicate a weak transmission quality, which affects the audience negatively.

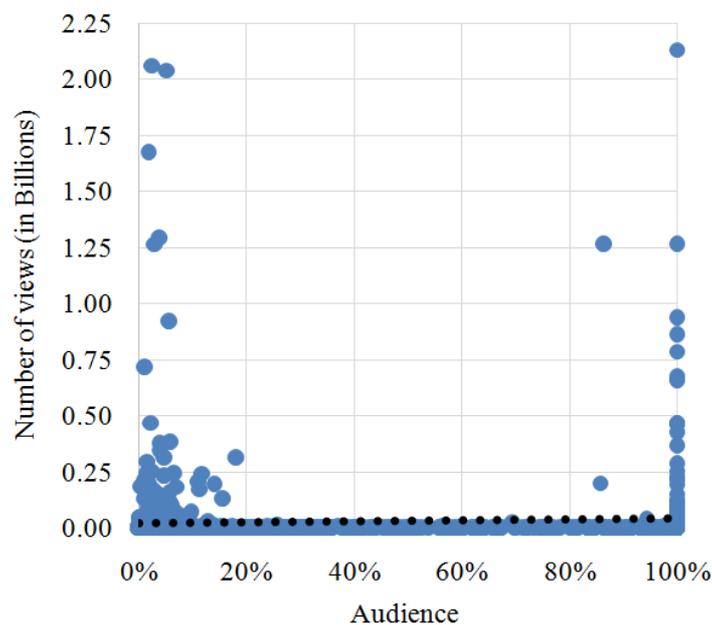


Fig. 28: Number of Views X Audience; Chart Information Sheet: Number of Views(Blue); Linear Trend (Number of Views) (Black); Fact Sheet of Metric Fig. 27 and 28: Characteristics: Influence is relative: If content is 'hot' with a recent increase in viewers/subscribers, the trend is for content to have high levels of audience; Both have a positive and relatively directly proportional influence on the audience; Subscribers: Significant effects from 60% of audience and 1.5 million subscribers (approx.), Positive correlation coefficient of 0.13; Views: from 85% audience and 0.3 billion views (approx.) Trend is expressive, positive correlation coefficient of 0.07; Number of enrollees (subscribers) indicates greater influence on availability, considering absolute values; Number of Views shows greater influence if considered the average loading times

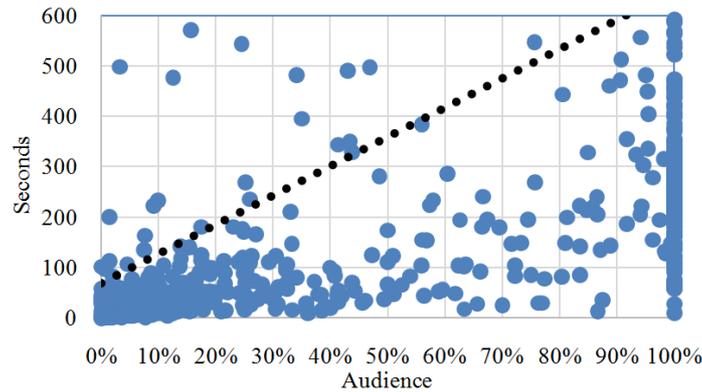


Fig. 29: Total time with Page Open X Audience; Chart Information Sheet: Total time with Page Open (Blue); Linear Trend (Total time with Page Open) (Black); Characteristics: The greater the Total Time with the Page Open, the higher the Audience's elevation tendency, but without an expressive relationship; Low relationship, because a high time can also signal unsatisfactory quality, a factor that reduces the audience; Trend of growth, however weak

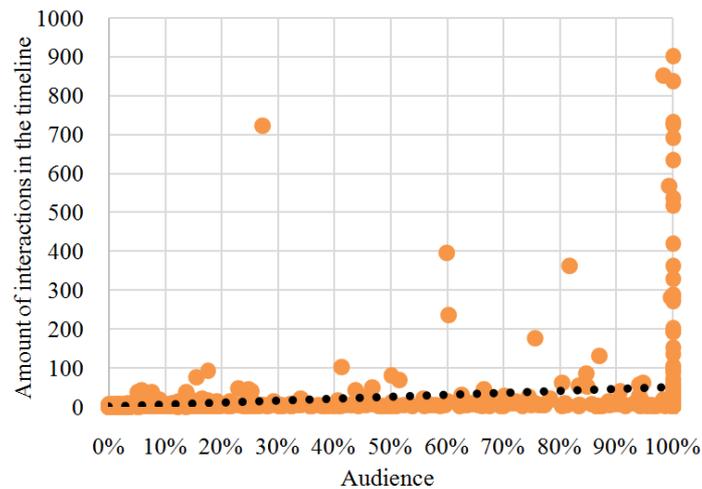


Fig. 30: Amount of Interactions in the Timeline X Audience; Chart Information Sheet: Amount of Interactions in the Timeline (Orange); Linear Trend (Amount of Interactions in the Timeline) (Black); Characteristics: The higher the Amount of Interactions in the Timeline, the less increase occurs in the audience elevation tendency; High number of interactions is indicative of deterioration of quality; Trend of growth, however weak

To analyze the influence of the interactions on the audience question, Fig. 30 was compiled to see the Amount of Interactions in the Timeline versus the Audience. In this sense, it should be noted that if the amount of interactions increases, there is a weak trend of audience growth, since this is indicative of impoverishment of the user experience.

And, in Fig. 31, as to the time to load the content, this did not show any effect that might be relevant. With this, one can see that unavailability causes decay in the audience. However, it is understood that the results show that the user can wait to watch the content he wants, but if this content takes too long to load the intention to watch it completely will be demoted.

Statistical Verification of Samples Quantities

From the analysis received by the data it is necessary to know to what extent such conclusions can be classified

as reliable to represent the whole. The literature gives us that the Mathematical Corollary demonstrating the amount of samples required for an acceptable level of assertiveness is expressed by Equation 1. The margin of error functions as a probability density function, so it is necessary to calculate it for a sample (NSS, 2017; Pegg and Weisstein, 2017; Smith, 2013):

$$\begin{aligned}
 & \text{Number of Samples Required} \\
 & = \left((Z - \text{score})^2 \right) * \text{Standard Deviation} \\
 & * (1 - \text{Standard Deviation}) / (\text{Error Margin})^2
 \end{aligned} \tag{1}$$

The variable described as 'Z-score' denotes the reliability, 'Standard Deviation' for very large or unknown populations should be set to 0.5 and the 'Error Margin' is expressed as a percentage. It should be noted

that such a calculation is valid for a Gaussian distribution. This usually occurs for large populations of data, this concept relates to the Law of Large Numbers. Therefore, it is confirmed that this has validity for an environment of the data order as the one studied. That is, with about 2 quadrillion views a year and 1.3 billion users. Figure 32 shows how a Gaussian distribution, also called a normal distribution, is given. In the said image

of Fig. 32, it is analyzed that 68% of the population are in up to 1 standard deviation of the mean, for more and for less. And 95% within up to 2 standard deviations from the mean and within 3 standard deviations it concentrates 99.7% of the population. Note that the Z-scores are given from the literature, previously calculated tables (NSS, 2017; Pegg and Weisstein, 2017; Smith, 2013; Kernler, 2014).

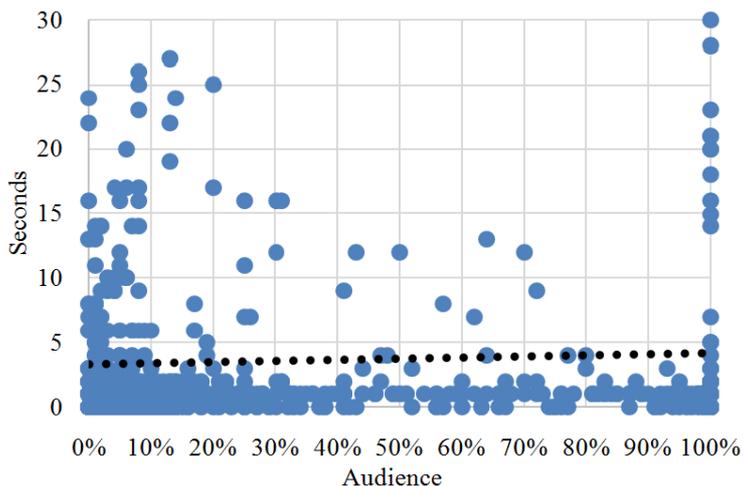


Fig. 31: Time to Load the video X Audience; Chart Information Sheet: Time to Load the video (Blue); Linear Trend (Time to Load the video) (Black); Characteristics: The longer the time to Load the video, the less increase occurs in the Audience elevation trend; A user wants to watch a content he waits for loading and a user who is more sensitive and needs a low waiting period, i.e., satisfactory availability, also have a tendency to watch the video altogether. Both of these characteristics are counterbalanced; Tendency of stability, equality in effects, because the two factors are joint and compensate

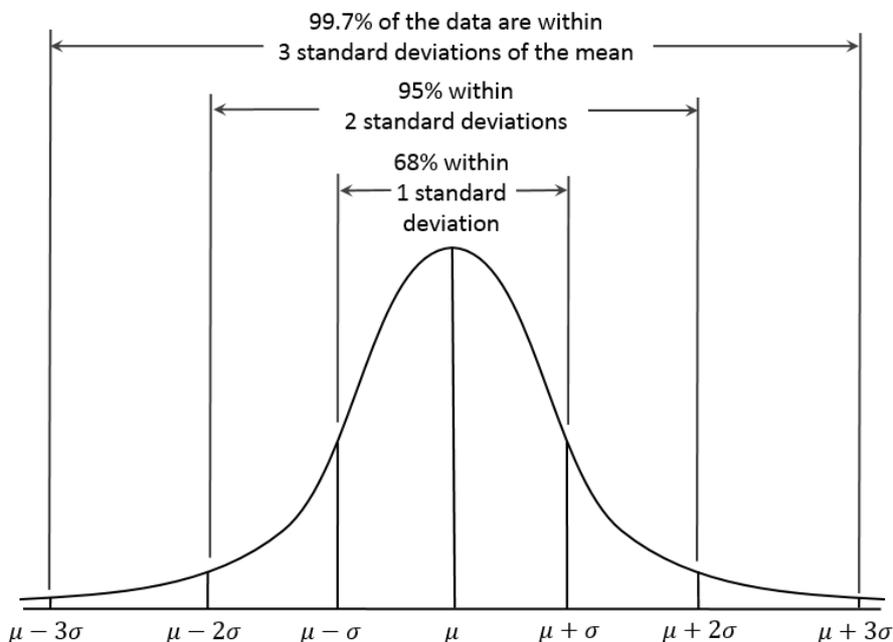


Fig. 32: Gaussian distribution; Source: Reproduced from A visual representation of the Empirical (68-95-99.7) Rule based on the normal distribution (Kernler, 2014)

For a reliability level of 99% with a 1% margin of error, 16,590 samples are required as seen in Equation 2. While for a reliability level of 95% with 5% of margin of error, it is necessary that 385 samples observed by Equation 3 would be necessary. And, for a 90% reliability with a 10% margin of error, 68 samples are necessary, as demonstrated by the calculations of Equation 4. It is also worthy pointing out that the values of sample size need to be completed until the next integer, for there is no way to have fractioned samples for this study:

$$\left(\left((2.576)^2\right) * 0.5 * (1 - 0.5)\right) / \left(\left(0.01\right)^2\right) = 16,589.44 \quad (2)$$

$$\left(\left((1.96)^2\right) * 0.5 * (1 - 0.5)\right) / \left(\left(0.05\right)^2\right) = 384.16 \quad (3)$$

$$\left(\left((1.645)^2\right) * 0.5 * (1 - 0.5)\right) / \left(\left(0.10\right)^2\right) = 67.65 \quad (4)$$

In view of the foregoing, it is necessary to point out that such calculations apply to large quantities, from the millions to more, which are unknown and difficult to measure accurately and tend to infinity. In addition, it is expected that such quantities will follow a normal distribution which, should solidify in accordance with increased sampling (NSS, 2017; Pegg and Weisstein, 2017; Smith, 2013).

Continuing in this matter, it is necessary to quantify in which level the 1281 samples fit. Therefore the resolution of Equation 5 is necessary to obtain these validation answers on this amount of experiments. When observing the equations previously solved and the number 1281 is estimated that the reliability of 97% is adequate, thus, it is verified that the Z-score of this level corresponds to 2.17009. Thus, we validate that the data captured in the experiments have a reliability of 97% with a margin of error of plus or minus 3.03% (NSS, 2017; Pegg and Weisstein, 2017; Smith, 2013):

$$\left(\left(2.17009\right)^2\right) * (0.5) * (1 - 0.5) / \left(\left(x\right)^2\right) = 1281$$

$$x = 0.0303161 \quad (5)$$

$$x = 3.03\%$$

Conclusion

To summarize, the software developed in this research proved to be adequate capturing key variables to identify IP delivery behavior in a wide-range of scenarios. And the tests provided useful data to model network performance with statistical significance. It is also important to point out that the closest study in this field, although deep in the sense of data volume, provides only a limited collection of a limited number of 4 metrics: Failures, Startup Delay, Average Bitrate and

Re-buffer time without scenario segmentation and without providing the methods (Krishnan and Sitaraman, 2013). And, as for the tool and analysis present in this study, measures these 16 variables: Name information, Total Time with the Page Opened, Time to load Video, Watched Time, Paused Video Time, Total Video Time, Amount of Timeline Interactions, Post-Video time, URL, View, Channel, Subscribers, City, Age, Gender and Date with 6 scenarios segmentations. In addition, its perceived that the presented data and corresponding analysis relate confirming the previous study general sense and its discoveries (Krishnan and Sitaraman, 2013). However, this research revealed a series of relationships and trends in IP delivery, allowing it to have a high degree of performance expectedness.

The undeniable and growing importance of internet protocol-based networks for broadcasting audiovisual content requires more and more attention. And, since its birth mechanisms to assign controls have been an incessant and necessary quest. In this study it was proposed to create mechanisms that bring assertiveness to this ecosystem that has assumed the role of Traditional Television Broadcasting through the Internet. Thus, the methodologies proposed to assign control mechanisms to the IP network environment have been shown to be implementable. And, the objectives were achieved through mathematical comparisons and computationally implementable algorithms. And, its products, have been designed in order to capture information in a robust manner.

For the methodology, a codification was programmed to measure the audience associated with the availability of the videos from the largest provider of this type of content currently. In order to observe at the end user's device to reach the goal, following the analogy that is used in the traditional TV broadcast environment. With the differentiation provided by the different types of times properly categorized for this purpose to accompany the interaction with the user. To achieve this goal, the computationally implemented algorithm consisted of two parts, the clients and the server that aggregates all the data of all the tests of all the consumers that have the extension installed.

The commercial and technical motivation to verify the audience is the same as the traditional TV, since content on demand IP have taken this role daily. The results obtained demonstrate that this is a tool that meets this functionality, allowing to bring to the IP world an assertiveness present only on TV so far, according to the literature. Therefore, it is extracted that the Software for Capturing Video Stream Audience Associated with Availability monitors the network endpoint, analogously to the TV meter, allowing to bring this assertiveness to IP Video Delivery. It has been demonstrated to be an efficient, effective and multifaceted tool capturing relevant

times to evaluate how Quality of Service affects the Quality of Experience. The data enables video-to-video analysis, user to device, device to device centralizing everything in the cloud and with low cost in throughput and processing in both the client and the server.

With the Captured Data Analysis to measure Availability and Audience, Comparisons, Calculations and Plottings allowed to draw several relevant conclusions. Like the fact that the relationship between audience and availability is intrinsic so that one feeds the other. It is also noted that the audience is negatively affected by unavailability. The number of subscribers and the number of views positively affect the availability and audience from certain bands. Also, the outages are so frequent that they raise the question about the general low audience, casting doubt on whether it would occur because of the significant amount of negative interference inserted by the IP ecosystem. That is, whether it is due to display or production, since it has been proven that there is a negative effect on packet transmission. In addition, the use of Mobile Networks presents low initial latency, however they present deterioration along the videos, with several pauses for loading and/or buffering. What makes the Quality of Experience becomes Variable, with low reliability.

Remaining in this analysis, it is concluded that the connection masking evidences the lack of CDN Infrastructure even using Broadband. This proves that, even with good connection, if the user-discovery algorithms are prevented from functioning, the tendency is for deterioration. That is, IP-quality video delivery depends directly on these algorithms present in the audiovisual delivery network architecture. In this way, it is established that if privacy is desired in the transmission via IP Stream the quality of the service and consequently of the experience is significantly impaired, even if the user connection is at satisfactory levels. Additionally, it is clarified that the tests performed have statistical validity, denoting relevant indications, besides proving the efficiency and effectiveness of the tool and the analysis.

As seen, the relationship between availability and audience is inherent, the requested parameters and analyzes chosen to be cataloged are satisfactory. Still in this question, it is necessary to emphasize that it is always affected by the place, apparatus and network connection. In addition, it is emphasized that the transmissions that most affect video delivery are demonstrated when we use mobile networks and when we mask the connection in order to obtain privacy. After all, these two situations do not yet have an adaptive solution, since they prevent the algorithms that feed the CDNs with user information from working satisfactorily, since we continually change the route in both cases, forcing a reorganization of the content delivery network.

Therefore, it is concluded that the path to the IP networks is long and the challenge remains great despite the considerable and constant evolutions, because this technology puts a bottleneck in the physical environment, setting a high price for infrastructure construction. Therefore, it is proposed that an alliance for delivery of Hybrid Solutions has better possibilities of delivering audiovisual quality, that is, Broadcast and Broadband must coexist synergistically.

Acknowledgement

The authors would like to thank the support of their respective institutions.

Author's Contributions

Vitor Chaves de Oliveira: Research about state of art in CDN's applications; Research about Software Development and Network functioning; programming with APIs; Experiments execution and refinement; Data-analysis and conclusions about software implementation and Network behavior. Conducted all the studies, experiments, definitions and identification about IP Stream measurement.

Sérgio Bimbi Junior: Research about Software Development and Network functioning; Java Script programming; Experiments execution and refinement; Statistical Analysis; Studies and experiments about privacy protocols; Distribution Analysis.

Andreiwid Sheffer Corrêa: Software Development, UML Modeling, Cloud Implementation, IP Stream Network Modelling, YouTube delivery strategy discovery, Tables and query's definitions (SQL, .NET), Systems administration and refinements.

Inácio Henrique Yano: Infrastructure implementation, Cloud identification relationship, charts plotting and analysis, Research on IP Proprietary Protocols, Overall software adjustments.

Mauricio Becker: Research and mathematical conclusions on algorithms behavior, definition of experiments, IT requirements designs and implementation, Php coding, collections statistics model identification.

Paulo Batista Lopes: Research and development of IPTV problem, Software strategy definition based on People's meter, Mathematical Modelling and interpretation of Results, Overall software adjustments. Quality of Service and Quality of Experience relationship regarding internet stochastic design. Interpretations refinement.

Gunnar Bedicks Junior: Research and development of Digital TV Broadcasting Systems, Software strategy definition based on Television Rating Points, Mathematical Results Modelling and interpretation, Overall software adjustments, experiments discussions and classification on methodology. Interpretations refinement.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and no ethical issues involved.

References

- Alhir, S.S., 2003. Learning UML. 1st Edn., O'Reilly Media, Beijing, ISBN-10: 0596003447, pp: 234.
- Arthemina, 2017. Unified Modeling Language (UML) and software modeling resources, news and tools. Arthemina.
- Baldini, G., S. Karanasios, D. Allen and F. Vergari, 2014. Survey of wireless communication technologies for public safety. *IEEE Commun. Surveys Tutorials*, 16: 619-641.
DOI: 10.1109/SURV.2013.082713.00034
- Bedicks Jr., G., 2008. Tuner-demodulator for the Brazilian digital TV system (in portuguese). PhD Thesis, University of São Paulo.
- Buyya, R., M. Pathan and A. Vakali, 2008. Content Delivery Networks. 1st Edn., Springer, Berlin, ISBN-13: 978-3-540-77886-8, pp: 417.
- Chai, B., 2014. 'True detective' finale crashes HBO go. Wall Street J.
- Cloonan, T. and J. Allen, 2011. Competitive analysis of adaptive video streaming implementations - white paper. ARRIS.
- de Oliveira, V.C., P.B. Lopes and G. Bedicks Jr, 2016. Capturing video stream audience over IP Networks. *Set Int. J. Broadcast Eng.*, 2: 9-12.
DOI: 10.18580/setijbe.2016.2
- DICIO, 2016. Audiência - significado de audiência. Dicio: Dicionário Online de Português, definições e significados de mais de 400 mil palavras. Todas as palavras de A a Z. (in portuguese).
- Doeven, J., 2013. Trends in broadcasting: An overview of developments. ITU - International Telecommunication Union, Telecommunication Development Bureau.
- Dresch, A., D.P. Lacerda and J.A.V. Antunes Jr *et al.*, 2015. Design Science Research: A Method for Science and Technology Advancement. 1st Edn., Springer, ISBN-10: 3319073737, pp: 161.
- Edwards, J., 2013. TV is dying and here are the stats that prove it. Business Insider Inc.
- Ellacott, J., 2014. Video streaming driving impressive growth in video content delivery market: TechNavio trends. TECHNAVIO. Infiniti Research Limited.
- Frost, V., 2014. The TV is dead, long live television. The Guardian.
- GFK, 2018. Smart insights: Media and entertainment. GFK Inc.
- Greenfield, R., 2012. NBC claims super bowl streaming victory, but viewers cry 'fail'. The Wire, The Atlantic Monthly Group.
- ITU, 2003. The birth of broadband - media fact sheet. International Telecommunication Union (ITU).
- Kaduoka, 2016. Streaming and distribution IP: Challenges. *Set Expo Proceedings*, 2: 180-183.
- KANTAR IBOPE MEDIA, 2018. Aferição consumo multimeios.
- Katel, P., 2014. Future of TV Can traditional television thrive in the Internet Age? *CQ Researcher*, 24: 314-336.
- Kernler, 2014. A visual representation of the empirical (68-95-99.7) rule based on the normal distribution. Wikipedia.
- Khanafer, M., M. Guennoun and H. Mouftah, 2014. A survey of beacon-enabled IEEE 802.15.4 MAC protocols in wireless sensor networks. *IEEE Commun. Surveys Tutorials*, 16: 856-876.
DOI: 10.1109/SURV.2013.112613.00094
- Krishnan, S.S. and R.K. Sitaraman, 2013. Video stream quality impacts viewer behavior: Inferring causality using quasi-experimental designs. *IEEE/ACM Trans. Network.*, 21: 2001-2014.
DOI: 10.1109/TNET.2013.2281542
- Lewin, 2014. For the love of physics (walter lewin's last lecture). For the Allure of Physics. YouTube.
- McMillan, G., 2013. Viewers are flocking to streaming video content-and so are advertisers. WIRED.
- N-TRON, 2014. White paper: Network availability. N-TRON Corporation: The Industrial Network Company.
- NSS, 2017. Sample size calculator. National Statistical Service - Australian Bureau of Statistics.
- Nunes, R., 2011. Web portal for continuous Internet access quality of service measurement. Masters dissertation, Technical Superior Institute, Technical University of Lisboa.
- OMG, 2017. Unified modeling language. Object Management Group, Inc.
- Oliveira, V.C., G. Bedicks Jr. and C. Akamine, 2015a. Methodology to asses IP connection availability: A prerequisite for feasible video stream through CDNs. *J. Comput. Sci.*, 11: 426-437.
DOI: 10.3844/jcssp.2015.426.437
- Oliveira, V.C., G. Bedicks Jr. and C. Akamine, 2015b. IP video delivery cost: An analysis on channel availability for live and non-live stream events. *Proceedings of the 64th IEEE Broadcast Symposium, (BS' 15)*, IEEE, Orlando, FL, USA.
- Oliveira, V., A. Mota, L. Mota, 2013a. Impacts of application usage and local hardware on the throughput of computer networks with desktop virtualization. *Am. J. Applied Sci.*, 10: 117-122.
DOI: 10.3844/ajassp.2013.117.122

- Oliveira, V., I. Yano, A. Mota and L. Mota, 2013b. Feasibility of desktop virtualization per software services and local hardware based on the network throughput. *J. Comput. Sci.*, 9: 827-837.
DOI: 10.3844/jcssp.2013.827.837
- Oliveira, V., 2012. Impacts of the variation of applications and of local hardware on the flow rate of computer networks with desktop virtualization (in Portuguese). Masters dissertation, Pontifical Catholic University of Campinas.
- Oracle Corporation, 2015. MySQL-The world's most popular open source database. ORACLE.
- Pegg and Weisstein, 2017. "Margin of Error." From MathWorld--a wolfram web resource. Wolfram Research, Inc.
- Rouse, M., 2011. Definition: Content Delivery Network (CDN). TechTarget, Search AWS.
- Sharma, A. and D. FitzGerald, 2014. ABC's oscars streaming outage shows web limitations for TV networks - traffic overload during academy awards tested live-streaming capabilities. *Wall Street J.*
- Simon, H.A., 1996. *The Sciences of the Artificial*. 3rd Edn., MIT Press, Cambridge, ISBN-10: 978-0262691914, pp: 231.
- Smith, 2013. Determining sample size: How to ensure you get the correct sample size. Qualtrics.
- Statista, 2015a. Statista Inc. Statistics and facts about YouTube.
- Statista, 2015b. Statista Inc. Global market share held by the leading internet browsers from January 2012 to December 2015.
- Statista, 2016. Statista Inc. YouTube - Statista Dossier.
- Statistic Brain, 2016. Statistic Brain Research Institute. YouTube Company Statistics.
- Tanenbaum and Wetherall, 2010. *Computer Networks*. 5th Edn., Prentice Hall, New Jersey, ISBN-10: 0132126958, pp: 960.
- TECHNAVIO, 2014. Global video CDN market 2014-2018. TECHNAVIO, Infiniti Research Limited.
- Trend Micro, 2015. BYOD - Consumerization of IT and Mobility - Trend Micro UK. Trend Micro (EMEA) Limited.
- YouTube, 2017. YouTube Press Statistics. YouTube.