GoalBit Plus

Performance measures:

Testing Plan & Results

Executive update summary

This document presents the testing plan and testing results that were executed by GoalBit Solutions in order to measure the performance of GoalBit Plus technology. The tests are developed in order to emulate real scenarios. The metrics collected inform about scalability, bandwidth save, and quality of experience.

Content Index

1. Introduction 2
2. Test deployment environment 2
3. Test cases 4
4. Test results analysis 5
5. Test results summary 11

1. Introduction

In order to quantify the benefits of our GoalBit Plus video optimization technology, during 2013, GoalBit-Solutions developed a set of test cases in order to analyze in real-time, more than three thousand video streams from viewers and emulated viewers. Emulated viewers are created in order to generate background traffic and in order to stress the delivery service.

GoalBit Plus records detailed information on each and every viewer's experience during every viewing session. We use this important information in order to present metrics about performance, scalability, bandwidth save, and quality of experience provided by our products. We compare our results with the "Conviva Viewer Experience Report"[1], from where our tests and this document were inspired. Conviva proved the direct correlation between quality, viewer engagement and profitability; showing the importance of the results provided in this document and applied to the GoalBit Plus platform.

As Conviva said, there are three broad categories that have large impact into the video quality of experience:

- Buffering or interruptions to smooth playback.
- Play failures, long video start times and abandonment before video starts.
- Low bit-rate, poor resolution and picture quality.

In order to focus the tests into our objectives, we will limit our study to live streaming (no video on demand) and we will not consider multi-bit-rate, resolution and picture quality.

Beside quality of experience metrics, we will also study the performance and scalability of our technology and specially the bandwidth save in the delivery due to the Peer-to-Peer technology.

In summary, we will present our metrics in the following groups:

1. Live streaming monitor
2. Bandwidth Transfer
3. Quality of Experience metrics

A detail of these metrics is presented in Section 4.

2. Test deployment environment

We keep the test-bed environment as simple as possible. The components and roles are the following:

- **GoalBit Media Server:** a single server with a minimal installation of the GoalBit Plus technology, i.e. the GoalBit Media Server. It does not include an administration interface, but it could be configured by command line. The administration interface is included in the GoalBit Manager product, and it is not a part of this environment.
- **Load Testing Tool:** a single server that will emulate end users connections. It should emulate 100 simultaneous users using P2P technology approximately (this varies with the technical specifications of the server). It also emulates Internet network conditions like packet lost and bandwidth uploads limits per user. It is installed for testing proposes and it will not be part of a production environment.
- **Android Set-Top-Boxes:** several set-top-boxes (up to 50) in order to measure the performance with production hardware in this work-loaded test environment. It is also expected to watch the video output of few of these set-top-boxes in order to verify the measures reported.

Figure 1 shows the topology of the test deployment environment.

[1] The Conviva Viewer Experience Report. February 2013. http://www.conviva.com/

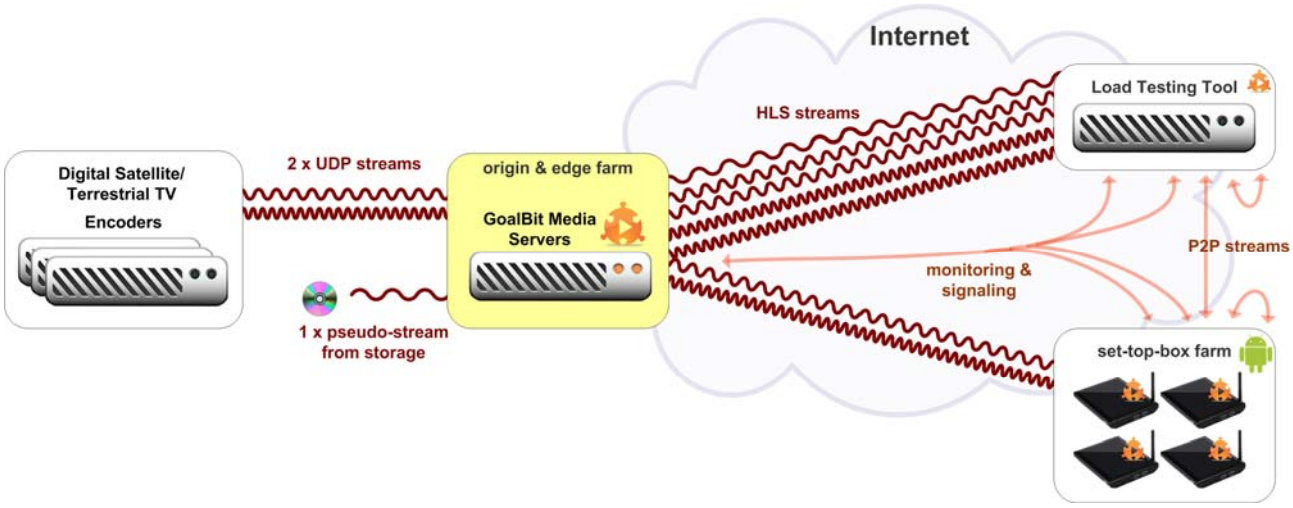


Figure 1. Test deployment environment. The GoalBit Media Servers is used as a traditional streaming server, for: video source acquisition, video processing, and content delivery. The GoalBit Load Testing Tools emulate several clients connected to the GoalBit Media Server, using P2P technology. A set of set-top-boxes also are connected to the GoalBit Media Server downloading the content and sharing it throw the P2P technology.

The hardware specifications of each component are:

- **GoalBit Media Server:**
  - Location: Vietnam
  - Platform: Ubuntu 12.04.3 LTS 64 bits
  - IP address: 118.69.57.139
  - DNS: goalbitplus1.quehuongtv.net
  - CPU: 8 Cores, Intel(R) Xeon(R) CPU E5620 @ 2.40GHz
  - RAM: 24068 MB ~ 24 GB
  - HDD: 250 GB
- **Load Testing Tool:**
  - Location: Vietnam
  - Platform: Ubuntu 12.04.3 LTS 64 bits
  - IP address: 118.69.57.140
  - DNS: goalbitplustest.quehuongtv.net
  - CPU: 8 Cores, Intel(R) Xeon(R) CPU E5620 @ 2.40GHz
  - RAM: 24068 MB ~ 24 GB
  - HDD: 250 GB
- **Android Set-Top-Boxes:**
  - Location: Vietnam
  - Platform: Android 2.3.4
  - CPU: ARM Cortex – A9 process (AM8726) @ 1.00GHz. MALI – 400 GPU, Open GL ES1.1 & 2.0 API Support (3D Games Support)
  - RAM: 512 MB DDRII 800
  - Audio/Video Outputs: 1 x HDMI 1.3, 1 x Optical SPDIF, Y/PbPr component, CVBS composite
  - Resolutions: HD 1080p/1080i, HD 720P, SD 576i/480i/576p/480p

3. Test cases

The objective of these tests is to quantify the start-time and re-buffering metrics of GoalBit Plus technology. All these tests will be conducted from Android set-top-boxes and from the load tool provided by GoalBit Solutions.

In order to simplify the tests, we will have the following assumptions:

- All the clients have enough download bandwidth (it means a download bandwidth larger than the bit-rate of the stream).
- In each test case, we will work with only one stream channel. The number of channels is not a variable in performance, quality or other metric of interest.
- The load generated by the Load Testing Tool is called “background traffic”, and the load generated by the set-top-boxes is called “user traffic”.
- Most test cases use a moderate volume of background traffic. We assume that it will be 100 simultaneous users approximately. But, the number of simultaneous users will depend on the server capacity and it will be defined exactly later in order to not stress the Load Testing Tool.
- Most test cases use a full volume of user traffic. It means that it is possible to connect as much as possible set-top-boxes into the test. For reading proposes, we assume that it could be 50 set-top-boxes. It is not possible to use more than 50 set-top-boxes in each test because it will impact in the measures that we are looking for; for example does not have sense to measure the impact of a very dynamic background traffic if it is negligible respect to the user traffic.

In the following, we present a summary of the test case called **“average viewing time - constant simultaneous users”**:

- **Objective:** To test the system in a steady state, with an average time of user sessions.
- **Test duration:** 30 minutes.
- **Background traffic:** Moderate volume of traffic. Approximately 100 simultaneous users, with a viewing time of 25 minutes in average.
- **User traffic:** Full volume of traffic. As much as possible simultaneous set-top-boxes (up to 50), with a viewing time of 25 minutes in average.
- **Network profile:** Four network profiles are used. Each user is in only one profile during the test period. There is no constrain in the download bandwidth; it matches with the common scenario of real end-users, because the download is very high respect to the current use. Because technical reasons (firewalls, proxies), in a real situation not all the end-users could open some port in order to accept incoming connections (knows at NAT traversal). In this test we configure that 37% of end-users accept incoming connections, a value close to our previous knowledge in real situations.

Network Profiles	Number of users:	Upload Bandwidth:	Other network conditions:
1	30%	1800kbit/s	duplicate & re-ordering (duplicate 0.4% gap 5 delay 10ms)
2	20%	1200kbit/s	loss & corrupt (loss 0.1% corrupt 0.1%)
3	30%	800kbit/s	duplicate & loss (duplicate 0.2% loss 0.1%)
4	20%	500kbit/s	re-ordering & corrupt (delay 10ms reorder 10% 50% corrupt 0.2%)

FENETEC S.A.
RUT 215311440011
Luis Cavia 2724, piso 8.
Montevideo 11300. Uruguay.
+(1)754-3009094 +(598) 2707-9369

4. Test results analysis

In this Section we will provide a summary of main results, metrics collected, etc. after the execution of test cases. A technical version of this section could be provided in HTML format in order to be able to access to detailed data (navigating in graphics and plots).

For the test case we will present the information and metrics:

1. Live streaming monitor
    - 1.1. Simultaneous Users
        - 1.1.1. Simultaneous Users: average
        - 1.1.2. Simultaneous Users: by time
    - 1.2. Viewing Time
        - 1.2.1. Viewing Time: average
        - 1.2.2. Viewing Time: by time
        - 1.2.3. Viewing Time: per buffering percent
        - 1.2.4. Viewing Time: per start time

2. Bandwidth Transfer
    - 2.1. Bandwidth Transfer: average

3. Quality of Experience metrics
    - 3.1. Start time
        - 3.1.1. Start time: average
        - 3.1.2. Start time: by time
    - 3.2. Views with Buffering
        - 3.2.1. Views with Buffering: average
        - 3.2.2. Views with Buffering: by time
    - 3.3. Viewers that abandon before play begins
        - 3.3.1. Viewers that abandon before play begins: average
        - 3.3.2. Viewers that abandon before play begins: by time

The test started at `2013-12-11 04:49:00`, and finished at `2013-12-11 05:20:00`.

In the period, the total number of users is `98` (all of them as auto-generated as background traffic), where `94` users start their session inside the period, and where `63` users start and stop their session inside the period.

1. Live streaming monitor

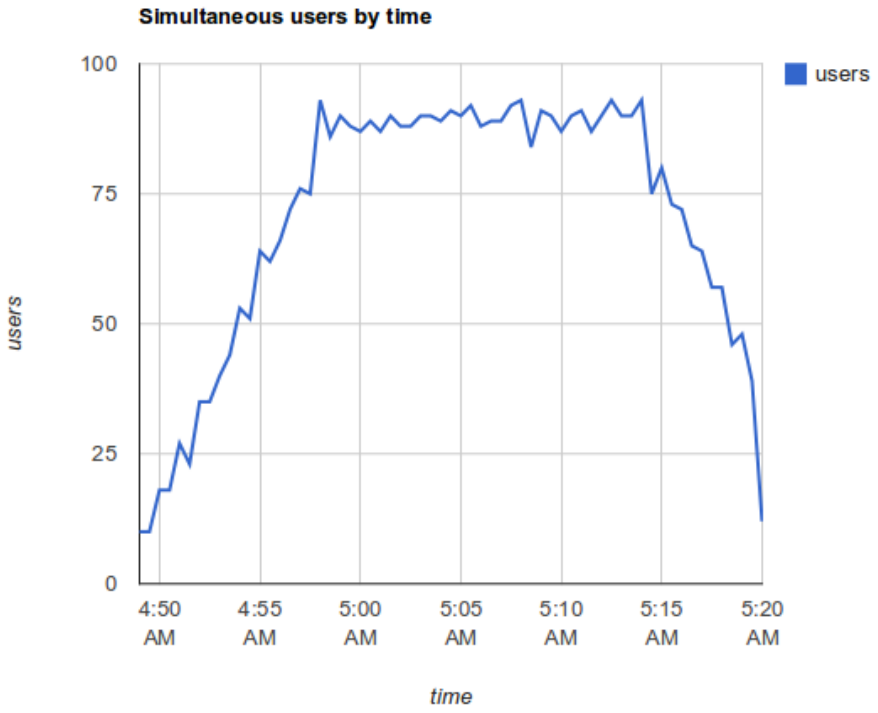
1.1. Simultaneous Users

1.1.1. Simultaneous Users: average

The following Table summarizes the distribution of simultaneous users in time. The maximum number of simultaneous users was 93, and in average it was 70.2.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
10.0	55.0	86.0	70.2	90.0	93.0

1.1.2. Simultaneous Users: by time



From 4:50AM until 4:58AM the test was bootstrapping (and end-users were connecting to the streaming). From 5:14AM until 5:20AM the end-users were disconnecting from the test.

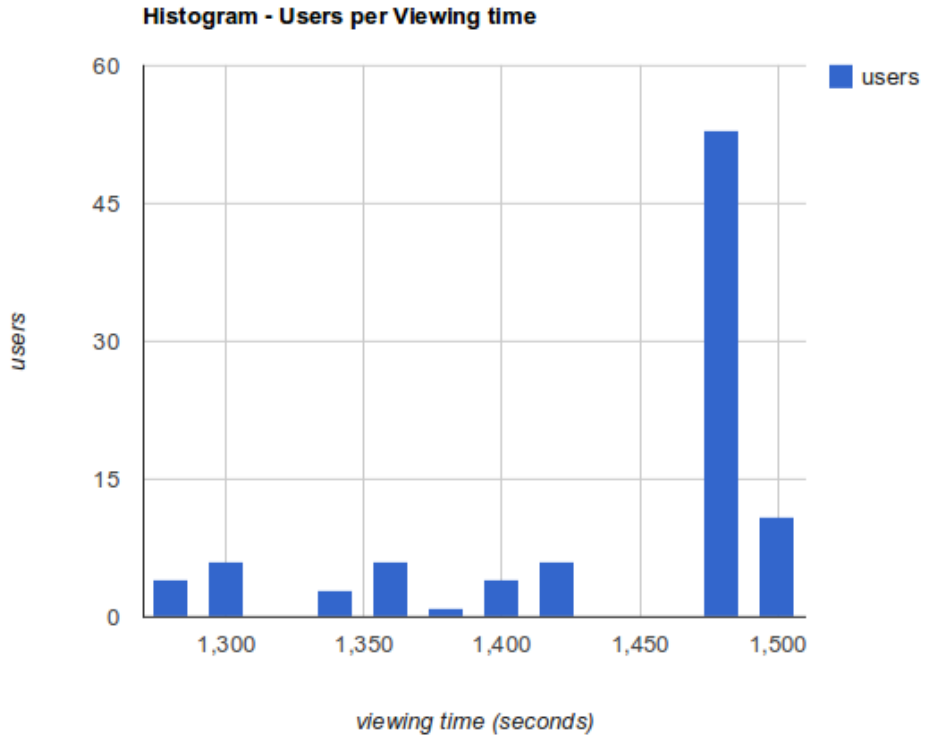
1.2. Viewing Time

1.2.1. Viewing Time: average

The following Table summarizes the distribution of viewing time per user. As expected, end-users were connected for 1500 seconds (25 minutes). Last connected used could not reach the 25 minutes because the test finished before.

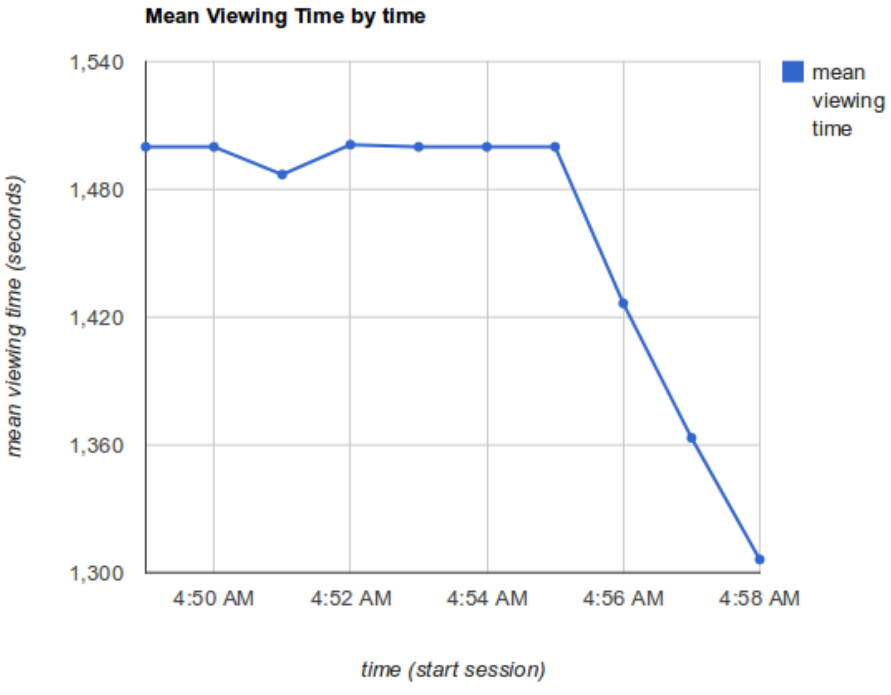
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1290	1420	1500	1460	1500	1500

Same information is showed in the following histogram.



1.2.2. Viewing Time: by time

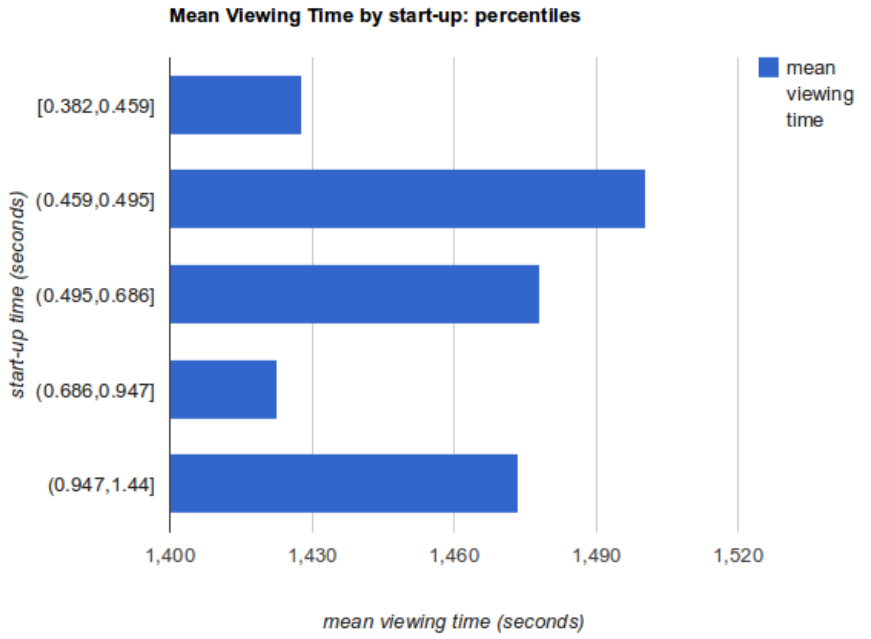
The following Figure shows that only the last connected end-users are being in the system for less than 25 minutes.



1.2.3. Viewing Time: per buffering percent

In the test, no client has re-buffering. Therefore, this Section does not apply to this test.

1.2.4. Viewing Time: per start time



In an overload system is expected that end-users with largest start-up time will be connected for shortest periods (lower viewing time). In our test, there is not correlation between start-up time and mean viewing time. It is expected because the behavior of end-users (automatic generated as background traffic) does not change with quality.

2. Bandwidth Transfer

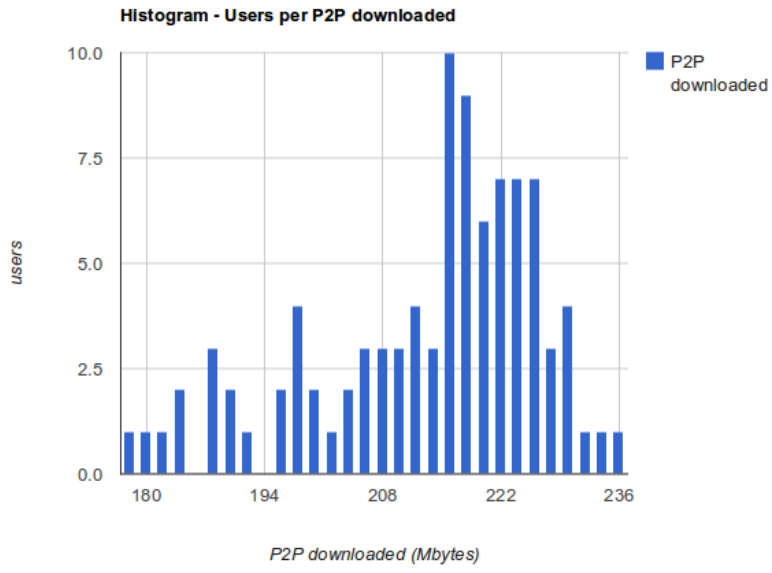
2.1. Bandwidth Transfer: average

The total data transferred is 24476 MBytes, where 4335 Mbytes are transferred form the CDN and the rest from the P2P protocol. This implies a bandwidth saves of 82.29% (20141 MBytes). The following Table summarizes the distribution of bandwidth save between users in percentage.

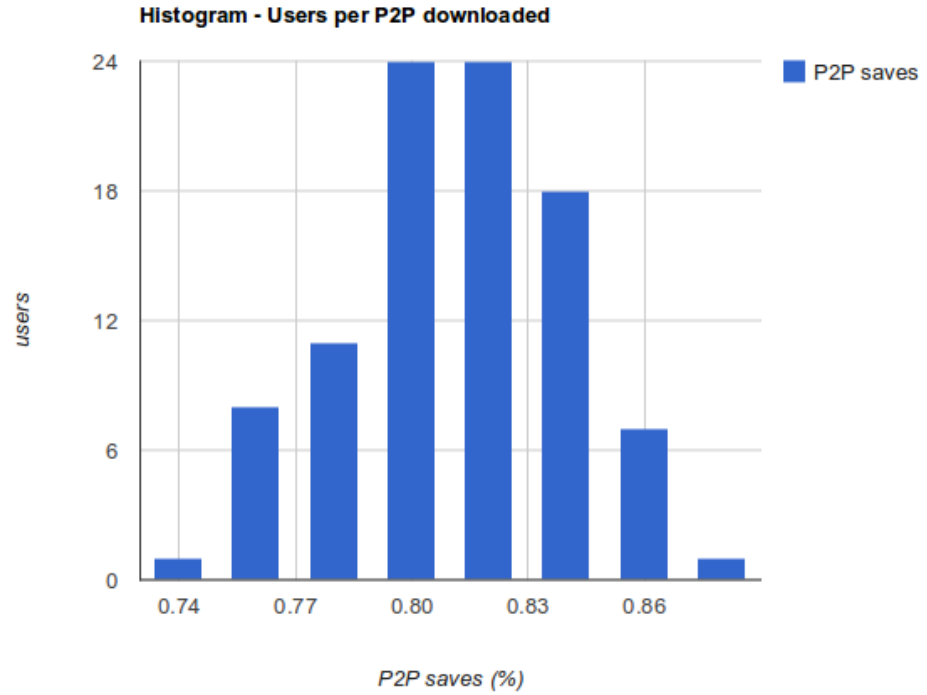
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
74.3	80.9	82.3	82.3	84.3	88.4

The same information is showed in the following histograms.

The histogram about Mbytes transferred by P2P protocol does not show clearly when an end-user shares few Mbytes because he was in the system for a short period or because he could not share enough due a protocol problem. The histogram about saves is much more illustrative about this point.



The following histogram shows the number of users that save the same percentage of bandwidth.
The end-user that shares less, shares 74%; and the end-user that shares most, shares 88%.



3. Quality of Experience metrics

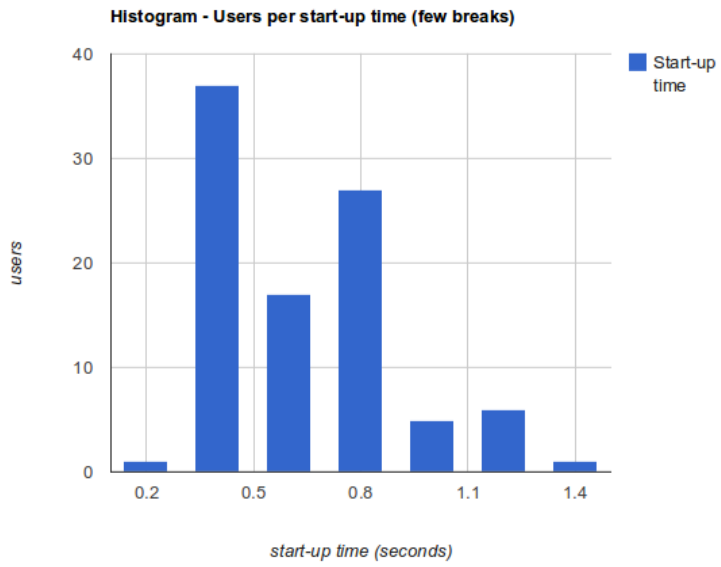
3.1. Start time

3.1.1. Start time: average

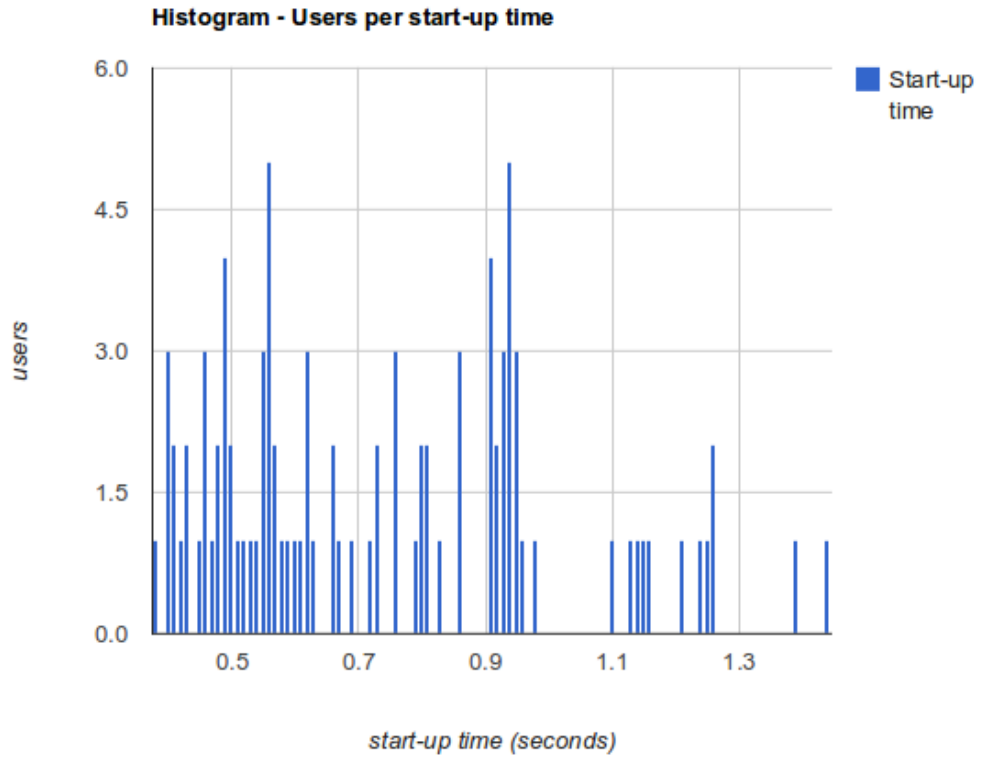
The following table summarizes the distribution of start time per user. In average, the start time was 0.745 seconds. The maximum start time was one second and a half (1.440 seconds), and the minimum start time was less than a half of second (0.382 seconds)

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.382	0.529	0.686	0.745	0.937	1.440

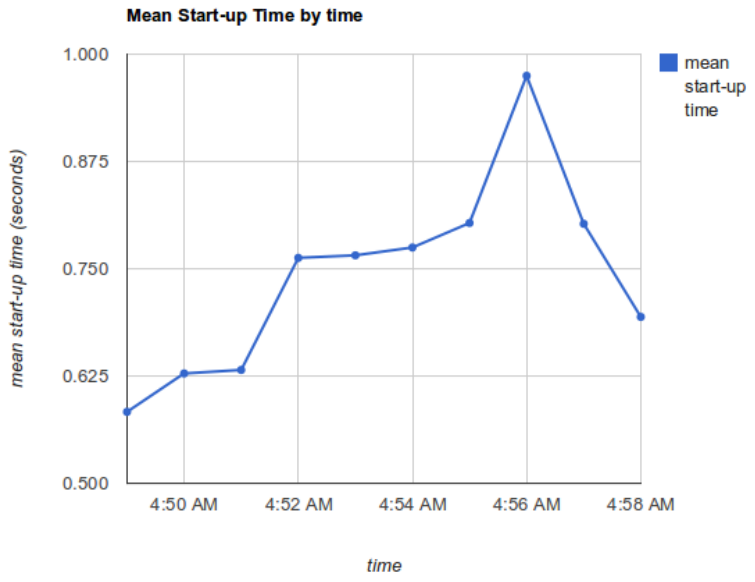
Same information is showed in the following histogram.



Same histogram with more breaks (granularity).



3.1.2. Start time: by time



The Figure shows how the start time changes with the time and load of the system. As not expected, the start time was increasing with time. We think that it happens because the "Load Testing Tool" increases its CPU load with time and it impact in the computation time of new peers. We does not consider that this happen because a degradation in the service with the load.

3.2. Views with Buffering

3.2.1. Views with Buffering: average

In the test, no client has re-buffering. Therefore, this section does not apply to this test.

3.2.2. Views with Buffering: by time

In the test, no client has re-buffering. Therefore, this section does not apply to this test.

3.3. Viewers that abandon before play begins

3.3.1. Viewers that abandon before play begins: average

The percentage of clients that abandon before play begins is 0% (the total clients that abandon are 0).

3.3.2. Viewers that abandon before play begins: by time

This section does not apply to the test case because abandonment does not happen.

5. Test results summary

We will compare our results with the performance information provided by the "Conviva Viewer Experience Report"[2]; this report uses more than 22.6 billion video streams from viewers in more than 190 countries watching content on more than 150 of the most recognizable video sites in the world during 2012. In summary, for live streaming we know:

- Views with Buffering: 48% in average.
- Conviva has established a critical threshold at which viewers react the most negatively: when buffering exceeds 2% of the total length of the viewing session. 12.3% of buffering in live streaming is above the 2% threshold.
- The time spent buffering in 90 minutes of viewing, a staggering 10.8 minutes lost to buffering. The best site in each content type had: less than 32.4 seconds of buffering for 90 min of content.
- A start time of 2 seconds or less does not have a large effect. A start time that exceeds 3 seconds, is deadly. Viewers who abandon a stream do so most of the time between 2 and 3 seconds of waiting. For the views abandoned, approximately 20% of them are before the 2 seconds, and 40% of them abandon between 2 and 3 seconds.
- Viewers that abandon before play begins: 18.3%. And 4.1% of live streams fail to start.
- 42% of streams are delivered without buffering impact and with high quality picture (HQ).
- Experience varies by device and platform, in Android (including live and VoD content):
  - 49.1% of views with buffering,
  - 17.5% of views that never successfully started,
  - 56.7 seconds in average spent in buffering per 10 minutes of content viewed,
  - 2.9 seconds start-up time in average
- The impact of buffering on viewing: 18.5 min per viewer impacted by low quality, and 121.8 min per viewer not impacted.

In our test case "average viewing time - constant simultaneous users":

- The mean start-time was 0.745 seconds (below of the recommend 2 seconds Conviva's threshold). In a real situation (not a simulated scenario) we expect an increase of this value for unconsidered factors as server overload, large network delay, etc.
- The clients did not experiment re-buffering. It shows a clear benefit of GoalBit Plus video optimization technology.
- No viewer abandons before play begins (i.e. not live stream session fails to start).
- In addition to the previous quality of experience measures, the system saves more than 82% of bandwidth at the server side (20141 MBytes over 24476 MBytes in total).

The results of this test show the benefits of new P2P protocol included in GoalBit Plus.

[2] The Conviva Viewer Experience Report. February 2013. http://www.conviva.com/