Project Number: 248495
Project acronym: OptiBand
Project title: Optimization of Bandwidth for IPTV Video Streaming

Deliverable reference number: D8.3
Deliverable title: Proof of Concept Report

Due date of deliverable: M34
Actual submission date: 31/10/2012 (M34)

Start date of project: 1 January 2010 Duration: 34 months

Organisation name of lead contractor for this deliverable: CORRIGENT SYSTEMS LTD - CSL
Name of the lead author for this deliverable: Yossi Barsheshet, Yuval Harduf

Project co-funded by the European Commission within the Seventh Framework Programme (2007-2013)		
Dissemination Level PU		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n 248495

Contributors

Participant #	Participant short name	Name of the Contributor	E-mail
1	CSL	Yossi Barsheshet	yossibs@corrigent.com
2	CSL	Yuval Harduf	yuvalh@corrigent.com
3	FTW	Froehlich Peter	froehlich@FTW.at
4	INTEROUD	Miguel Ángel Francisco	miguel.francisco@interoud.com

Table of Contents

1. Table of figures

2. Executive Summary

The main goal of OptiBand project is to develop an efficient solution for IPTV operators to provide more HD video streams to users which have a limited access bandwidth, specifically ADSL.

Following 27 months of operating as a consortium, studying, researching, investigating, developing, simulating and integrating OptiBand could start a live test as planned in order to provide the proof for the OptiBand technology.

During the last months OptiBand conducted a live test at TILab, led by TiS with the assistance of the consortium members, with a close cooperation of FTW in order to assess the technology QoE performance compared with the defined objective of the project.

From QoE point of view, the most critical objective is the reduction of 33% bandwidth while delivering three HD over 15 mbps ADSL line while maintaining a reasonable QoE. For this reason OptiBand introduces QoE assessment model and which was applied during the live test, under real use cases scenarios.

In this report we dive into the details and the deep explanation of the results provided in the QoE final report (i.e., D2.5 or D8.2), and provide the summary and the conclusions of the OptiBand performance.

The bottom line of our conclusions is that the technology works. We have placed the technology infrastructure for ADSL operators to provide IPTV services with multi stream of HD quality.

3. Introduction

This document provides the summary, conclusions and recommended actions of the live test. In fact this document can be referred to as the final report of the OptiBand research and development results.

In general, in order to summarize and conclude the proof of concept, the document refers to D2.5 and to D8.2. Then it provides OptiBand proof of concept conclusions followed by some recommendations.

Chapter 5, Live Test Conclusions, provides a brief review of the live test and of the QoE conclusions.

Chapter 6, Conclusions based technical and QoE achievements, provides a brief review of the live test and of the QoE conclusions.

Chapter 7, Recommendations for further actions, provides recommendations for possible next steps.

4. Live Test Conclusions

In order to answer the QoE research questions (see D2.5 1.1) a live test was defined, planned (see D8.1) and executed at TILab based on the technology that was developed in OptiBand D8.2 (see D8.2 Live Test Report).

This chapter provides the summary of the QoE conclusions from the live test which were reported at D2.5 (Final QoE research recommendations report).

It is recommended to read D2.5 and D8.2 for a more comprehensive understanding of the live test results.

4.1 Live Test Execution

Four use cases were tested and evaluated during the live test under real live conditions. The use cases were designed in order to evaluate the technology performance in terms of bandwidth reduction rate and QoE, as well as for enabling absolute and relative QoE assessment. Use cases differ from each other by the ADSL line rate, the number of TV sets at the household and the zapping pattern. Hence different level of complexity use cases were evaluated, including aggressive ones.

For more details on the use cases and assessment methodology please refer to D8.1.

The live test was executed in two phases, each with a different size of users' panel who were individually interviewed and provided their quality rating score as well as acceptance rating.

During the preparation for the live test we faced technical issues which have the potential to negatively affect the QoE assessment results.

- The main issue we faced is a phenomenon of artefacts that appeared on played video stream, mainly during zapping scenarios. This occurred due to an implementation bug we tried to fix, but due to schedule constraints of the live test we had to proceed to execute the live test without being able of fixing the bug (though most of the issue was fixed).
- Another issue is the incompliant video formats between the original HD SOTA sources we initially tested and reported at D2.4 and with the equipment which TILab is equipped with, hence a transcoding operation was required and performed. This has the potential of reducing the quality of the original video (HD SOTA).

4.2 QoE conclusions following the live test

For convenience, this section provides the QoE live test conclusions as they were reported in D2.5 section 2.5 (Conclusions).

In the following, the above presented results are summarized and interpreted, in order to provide answers to the formulated research questions.

4.2.1 Absolute quality in the baseline condition

For the interpretation of the performance and experience of the OptiBand solution, it is of course important to first review the results for use case 0, which represented 'optimal' system conditions with only 1 TV set and available 8.5 Mbps, i.e. the video at HD SOTA. Here, the average experienced quality of 4.0 MOS was good. However, in comparison with the scores for the state-of-the-art sequences within the previous testing iteration, MOS scores were about 0.2 lower (4.0 vs. 4.2 for action, 3.7 vs. 3.9 for soccer). This means that only with slight relative impairments of 0.3 MOS, the absolute MOS threshold would be reached.

In principle, as an explanation for these issues, the reference videos employed for the test could have been of bad quality and thus be responsible for these lower subjective ratings. In fact, as the same video sources had received good scores in previous QoE studies [2], this cannot serve as a valid explanation in this case. However, we do not know whether the transcoding of the videos conducted at TIS premises have resulted in potential damages. Anyway, the same equipment and transcoding method were in use for use-case 0 and for the rest of the use-cases, hence resulting with same video quality.

4.2.2 Relative quality impairments in challenging conditions

One of the main OptiBand objectives was to reduce the necessary bandwidth of an HD stream by more than 33% while preserving acceptable perceived quality. To this end, the comparison of use case 3 with use case 0 show that a reduction by 1.5 Mbps did not materialize in noteworthy degradations of the mean opinion score (compare section 2.4 at D2.5). In the more challenging use cases 1 and 2, where the available bandwidth was decreased by more than 33% per TV set, average quality values tended to be lower by about 0.3 MOS. However, we also found that with slow zapping patterns, the decline of perceived quality is almost invisible (compare the very small effect size in Figure 18 at D2.5).

It is the place to mention that due to PDD implementation resources considerations it was decided to use strict GOP size requirement of 2 seconds, without keeping the option for configuration. This has the potential to affect the QoE during zapping scenarios.

Thus, we can conclude that with optimized tuning and zapping strategies (such as decreasing the GOP time and supporting this in the implementation), small relative impairments can be achieved with the OptiBand solutions.

4.2.3 Absolute quality in challenging conditions

While the focus of the OptiBand QoE investigations is on assessing relative quality degradations, it is of course also important to look at the absolute achieved quality. In order to interpret these results, we had defined a threshold of 3.7 MOS and applied throughout the project. The result of the previous testing iteration, as reported in D2.4, was that all 1080i sequences exceeded this threshold, with values of 3.8–4.0 MOS, even with reductions of 55%. Based on these experiences, our expectation was that with the selected bandwidth levels the subjective service perception should not have been below 3.9 MOS, thus comfortably surpassing the 3.7 MOS threshold.

The results of the live test for use case 3 show that a reduction by 1.5 Mbps using the OptiBand packet dropping solution resulted in a satisfactory absolute level of subjective quality (~4.0 MOS). The results for the more demanding use cases 1 and 2 are more differentiated, and they seem to be especially dependent on the underlying zapping scenario. We can conclude that absolute quality is satisfactory in slow zapping situations, as the quality threshold of 3.7 MOS is slightly surpassed in the respective test conditions (with an average MOS score of 3.7, compare Figure 18 in section 2.4 of D2.5). Also, disconnections and reconnections of other set top boxes do not seem to have a strongly disturbing impact. This picture is supported by the acceptance testing results, where the lowest acceptable threshold of 0.80 was slightly surpassed in slow zapping conditions (0.83). We expect that a higher absolute perceptual quality of baseline conditions of use case 0 would also have resulted in better quality within use cases 1 and 2.

An open point for the OptiBand solution is still how to deal with fast zapping conditions. Here, relatively low rating values of 3.1 MOS and an acceptance ratio of 0.63 have been obtained. The selected bandwidth levels for H.264 streaming services should have resulted in excellent streaming performance, however this was not the case in the presented test. Possibly, this may be related to problems with the co-ordination of involved system components (PDD, network, and encoding technology), or more likely, due to implementation constrains in terms of resources and time for system stabilization. As can be seen in the analysis of the video stream characteristics (section 1.5.3 in D2.5), the drop of STB 1, the too long channel switching intervals, the high variance on many streams, and the high variance on the documentary streams may have caused interferences and impairments resulting in limited quality.

4.2.4 Content classes

Regarding the content classes, we found that the documentary sequences received the best MOS and acceptability ratings, although they usually had lower average payloads. Action and soccer contents were very similar to each other. This result cannot be explained in a straight-forward manner. On the one hand we could assume that documentary contents might have been more appealing to users, which could have resulted in higher ratings. However, the strength of the effect is still surprising, especially when considering the fact two previous studies had not found significant differences between content classes within the same original video materials.

Another possibly decisive factor may have been the applied encoder: TVN provided the encoder for the documentary materials, and OPT provided the encoder for the action and soccer sequences. Thus, in principle, the encoders may have had different efficiency levels with regard to achieving high QoE levels.

However, as mentioned before, there are several missing information layers that make an interpretation difficult. As an important example, we do not know about the effects of the transcoding that had to be executed during live test preparation for the integrated prototype, as described in chapter 3 of D8.2 (Remarks).

5. Conclusions based technical and QoE achievements

In this chapter we will compare the results with the project objectives, and then we will dive into the gaps between the expected results and the results we got from the live test.

Conclusions regarding the proof of concept will be provided and explained with regard to the project execution and to the execution of the live test.

5.1 Project Objectives

The main concept behind the OptiBand project is to enable the delivery of multiple HD streams to a household over a single ADSL. This should enable the delivery of HD IPTV application by ADSL operators and increase the competition in the IPTV market.

In order to fulfil this concept four objectives were defined at the project definition phase (see Annex I – "Description of Work").

Please note that all of the objectives were tested and achieved during the simulations OptiBand conducted during the first two years of the project. At the following chapter we will provide the explanation for the live test PoC achievement against the objectives.

5.1.1 Objective #1: Providing more HD video over existing access technology and IPTV networks

Specifically this objective defines 3 HD streams for a single 15Mbps ADSL line (normally 3 HD streams require 24Mbps). Each video stream should suffer only minimal degradation of QoE (SSIM greater than 0.7 for 70% of the time, or a better – subjective – QoE measurement selected by WP2 and described in D2.3).

The measuring criteria for this objective:

The PoC (Proof of Concept) should demonstrate the operation of 3 HD streams over 3 TV screens fed by a single 15Mbps ADSL line. QoE degradation should be measured by SSIM metric, using measurement equipment (or a better – subjective – QoE measurement selected by WP2 and described in D2.3).

The criteria were already achieved in previous testing we have conducted in OptiBand and as reported at D3.4.2 and D4.5.2.

5.1.2 Objective #2: Providing Scalable Personalized Quality of Experience

Specifically this objective should be achieved through the data dropping algorithm that will allow adjusting, in real time, the level of data dropping of each video stream according to the needs of the subscriber at any given time (e.g. adjustment to the number of TV screens that are switched on and to the type of video content (HD or SD) displayed on any of the screens).

The measuring criteria for this objective:

The PoC will demonstrate that different levels of data dropping should be applied to different households, according to their individual preferences. The Live Test Plan (D8.1) includes specific test cases to measure this objective (see also use cases 0 to 3 that are mentioned earlier in this document as well).

Examples of such test cases are:

1. Two TV sets are showing full bandwidth HD video. A third TV set is switched on. The OptiBand system responds by reducing bandwidth on each stream and enabling the 3 HD streams to the 3 TV sets
2. One of the 3 TV sets is switched off. The OptiBand system responds by restoring full HD bandwidth to the 2 remaining sets

3. The video quality is monitored by SSIM test equipment. In addition, the viewers of the TV sets are interviewed and asked if they have perceived any degradation of the video quality during the above 2 tests.

The criteria were already achieved, tested and reported as part of WP3 and can be reviewed at D3.3.2 and D3.4.2.

5.1.3 Objective #3: Providing content aware data drop algorithm

Specifically the content aware data drop algorithm should be able to reduce the data rate of a HD stream by 33% on the average. The measurement is “on the average” since the data dropping may vary during the time because of changing in preferences and video streams. Therefore only an average measure has any significance.

The bandwidth reduction should be achieved with minimal degradation of QoE (SSIM greater than 0.7 for 70% of the time, or a better – subjective – QoE measurement selected by WP2 and described in D2.3).

The measuring criteria for this objective:

The PoC should measure the bandwidth of a HD video stream before and after the data dropping done by D3.4.2. This measurement will show a reduction of 33% in the bandwidth.

SSIM for the reduced bandwidth will be greater than 0.7 for 70% of the time (or a more accurate – subjective – QoE measurement selected by WP2 and described in D2.3 may be used).

This objective was tested by simulation of the content aware data drop algorithm in M12 and later on it should be demonstrated by the prototypes (D3.4.2 and D4.5.2) at the POC (WP8).

5.1.4 Objective #4: Achieving all the project objectives while maintaining the existing IPTV ecosystem intact

Specifically all the other project objectives should be achieved without any need to upgrade the existing IPTV equipment such as home equipment (STB) or last mile equipment (DSLAM).

The measuring criteria for this objective:

The PoC should demonstrate the achievement of the other project objectives without the need to upgrade any existing DSLAM or STB equipment.

This objective was tested by simulation of the content aware data drop algorithm in M12 and later on it should be demonstrated by prototypes (D3.4.2 and D4.5.2) at the POC (WP8).

5.2 PoC compliance with the Objectives

5.2.1 Objective #1

In use case 1 we tested the scenario of 3 TV sets each presenting HD stream connected to a single home gateway, connected to the IPTV network using a single 15 mbps ADSL line.

The subjective model, which was defined by WP2 and described at D2.3 (Initial QoE research recommendations report), defined a MOS rating threshold of 3.7 and an acceptability rating threshold of 0.8.

5.2.1.1 Use Case 1 Complying

For use case 1 we have several dimensions to look at:

The documentary genre achieved satisfying results and it is consistently above the thresholds, while results are below the thresholds for soccer and action genres except for slow zapping where action genre achieved good result.

On fast zapping we are much below the defined thresholds, while the situation improves for slow zapping and more than that for disconnecting scenarios.

As for the full average result it looks like we fail to surpass the thresholds when considering also the fast zapping, while for slow zapping we surpass all thresholds, including for use case 2 which is the second challenging one.

5.2.1.2 Objective #1 Compliance Conclusion

Although not all parameters surpassed the defined threshold we can claim that we have proved that the technology complies with the objective 1. The technology absolutely could show, at least in the documentary genre case, that we can deliver 3 HD over a single 15 mbps ADSL line, with reasonable QoE degradation.

After analysing the causes of the low ratings of part of the tested cases, we can say that by adding some more R&D resources we can get over the issues and provide fully acceptable ratings. However such an effort is out of the scope of the project.

- **Zapping issue:**

Following the difference between the zapping scenarios we can point out that the main problem is within the implementation aspects rather than in the technology.

As mentioned in D8.2, during the live test preparation we found out that from time to time we face artefacts at the video stream while zapping on another STB. We can tell that this problem is less visible at the disconnecting scenario compared with zapping since the disconnecting scenario frees a full stream bandwidth from the line while in zapping it only changes the stream and doesn't free bandwidth.

We have tried to fix this issue (which was indicated as a bug probably at the PDD) but without any success due to time limit of the live test.

The difference at the results between the zapping scenarios strengthens our assumption regarding that root cause which leads to such low results. It is trivial to see that disconnecting scenario results with better score than slow zapping, and that slow zapping results with better score than the fast zapping.

Moreover, during the implementation phase it was agreed to use a 2 seconds GOP (group Of Pictures) size due to implementation consideration. Under degradation scenarios, the errors propagate longer and have a bigger impact than if the GOP size was smaller.

It should also be commented that the most typical and interesting scenario is the slow zapping scenario, though fast zapping is also in use.

- **Streams aspect:**

As mentioned above and in D2.5, for the live test we have used the same video sequences we have used for the early QoE assessments and which were reported at D2.4 (Intermediate QoE research recommendations report). As can be learned from D2.4, the MOS and acceptability ratings were dramatically higher than in the live test, including for lower rates.

For example, SOTA Action genre got ~4.3 MOS rating while Soccer genre at the same rate got ~4 MOS rating. The comparable live test results are based on use case 0 where no bandwidth limitation is available, the stream plays SOTA level and no packet drop is being applied by the PDD. The results of use case 0 are:

MOS – 3.9 for Action, 3.7 for Soccer, 4.3 for documentary

As concluded by D2.5, the live test results introduce a degradation of scoring inherently compared with the original video sequences. This can partially explain the lower results we got at the live test.

Another stream related note is the transcoding we have generated at the live test preparation phase due to TILab equipment video coding comparability issue. All the three genre streams were transcoded, hence a degradation may be caused by this operation.

While trying to explain the difference between the documentary genre and the soccer+action genres we arrived to the encoder equipment. TVN encoder was responsible for the documentary streams while the

OPTEC encoder was responsible for the soccer+action streams. We cannot tell what the impact of this difference was, but we do can tell that it is something with huge potential for influencing the results.

5.2.2 Objective #2

In use cases 0, 1, 2 and 3 we have proved that every household is applied with different level of data dropping according to its preferences.

As can be learned from Figure 1 and Figure 2, based on different available bandwidth ADSL line characteristic, the algorithm applies different level of data dropping to each household.

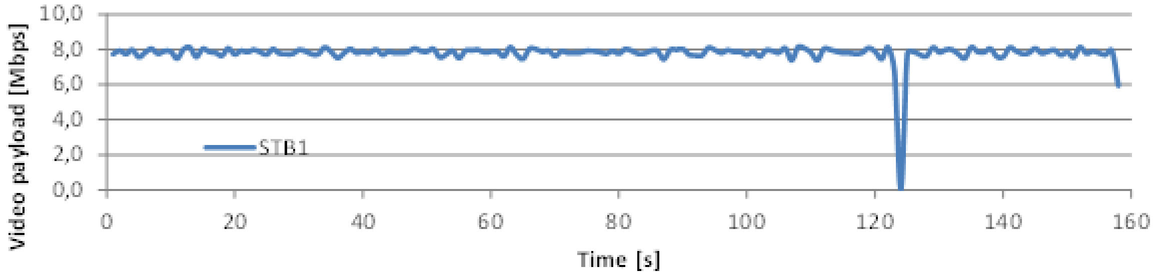

Figure 1: use case 0, action genre, with available 15 mbps

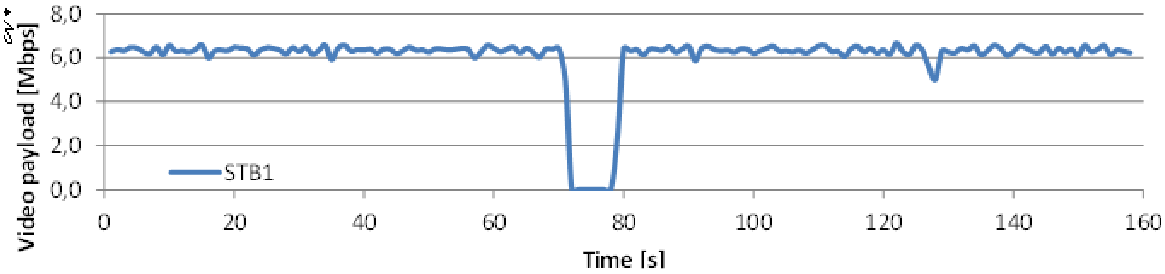

Figure 2: use case 3, action genre, with available 10 mbps

Another example is the disconnecting scenarios of use cases 1 and 2. At Figure 3 below we can see that when only STB1 is set on then no data dropping is applied (second 0 - 40), but when all the STB are set on then STB1 data dropping increases and stream bandwidth decreases to 4.95 mbps. At seconds 115 – 155 STB3 is set on and only STB2 remain off, hence STB3 data dropping decreases and stream bandwidth increases to ~6 mbps. However, using alternative weights preferences it could result that STB3 would increase to 8 mbps and STB1 reduce to ~6 mbps.

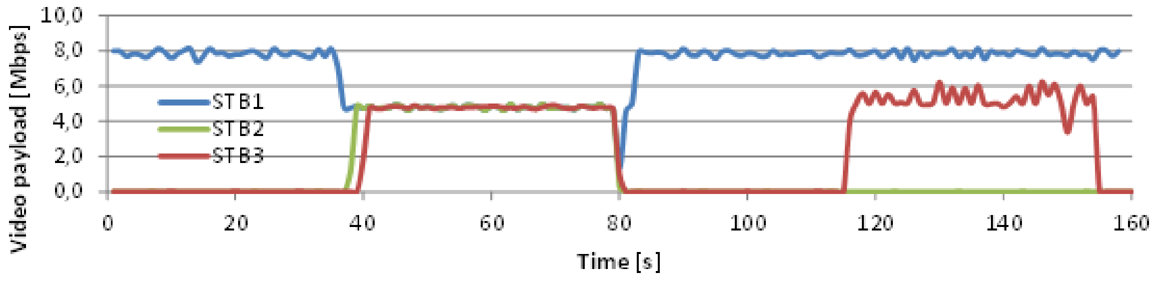

Figure 3: use case 1, disconnecting scenario

Furthermore, as demonstrated and reported at D3.4.2, the algorithms enables fine tuning parameters which are related to subscriber preferences per genre and per STB (e.g. preferring better Soccer quality at the living room than at the kids room). Such preferences configuration capabilities were developed as part of the prototypes and were supported in the middleware, but we didn't find a reason to use live test resources for demonstrating it.

5.2.3 Objective #3

Compliance to objective 3 was demonstrated at the simulations and during the prototype testing and was reported accordingly. Here we focus on the compliance as was proven by the live test.

5.2.3.1 Use Case 1 Complying

At use case 1 the live test had demonstrated three HD streams delivered over a single 15 mbps ADSL line. Although the live test results suffered from some low ratings, it was proven that when a household is watching three HD streams, an acceptable level of QoE can be achieved (see 5.2.1 Objective #1 above).

Regarding the bandwidth reduction rate calculations, we provide two perspectives:

- As can be study from Figure 4 (in this case, the documentary test which surpassed the thresholds) and from **Error! Reference source not found.** (all genres as was proven in earlier testing at OptiBand), the stream plays with 4.95 mbps while achieving reasonable scoring. Assuming the SOTA is 8 mbps, we can say that bandwidth was reduced at (8-4.95)/8 = 38%, much more than the required 33%.
- Playing three HD SOTA streams requires at least 3 * 8mbps = 24mbps, while in use case 1, as per Figure 4 below, OptiBand played three HD streams with 15mbps. This is a reduction of (24-15)/24 = 37.5%, much more than the required 33%.

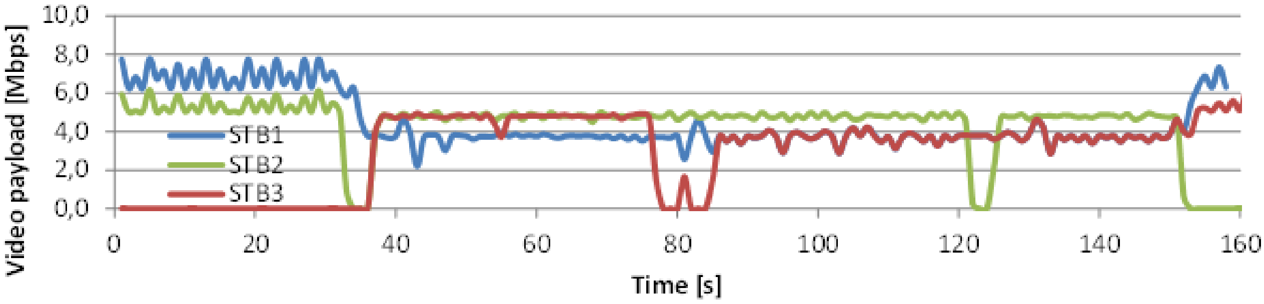

Figure 4: use case 1 with slow zapping, while documentary genre is active

5.2.3.2 Use Case 2 Complying

At use case 2 the live test had demonstrated two HD streams delivered over a single 10 mbps ADSL line. Although the live test results suffered from some low ratings, it was proven that when a household is watching two HD streams, an acceptable level of QoE can be achieved (similarly to the explanation of Objective #1).

Regarding the bandwidth reduction rate calculations, we provide two perspectives:

- As can be study from Figure 5 (in this case, the documentary test which surpassed the thresholds) and from **Error! Reference source not found.** (all genres as was proven in earlier testing at OptiBand), the stream plays with 4.95 mbps while achieving reasonable scoring. Assuming the SOTA is 8 mbps, we can say that bandwidth was reduced at (8-4.95)/8 = 38%, much more than the required 33%.
- Playing two HD SOTA streams requires at least 2 * 8mbps = 16mbps, while in use case 1, as per Figure 5 below, OptiBand played two HD streams with 10mbps. This is a reduction of (16-10)/16 = 37.5%, much more than the required 33%.

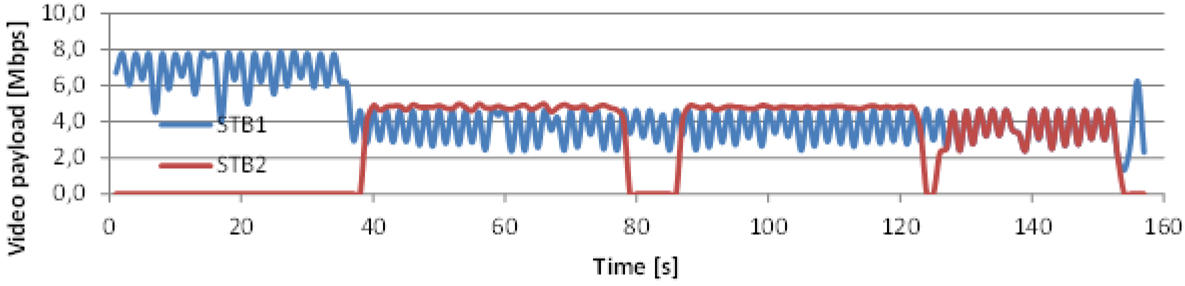

Figure 5: use case 2 with slow zapping, while documentary genre is active

5.2.4 Objective #4

At the entire live test, the only modified equipment are the middleware, the encoders and the PDD. In order to deploy the OptiBand technology it is only required to install the PDD and to upgrade the encoders. Upgrading the middleware is just an optional and it is not mandatory by the OptiBand, unless the operator wants some more OptiBand related added values such as genre priority configurations and etc. The mass deployed equipment such as the DSLAM equipment and the set-top-boxes do not require any change.

By leaving the mass deployed equipment as is, OptiBand enables the project objectives and functionality while maintaining the IPTV ecosystem intact.

Although the live test didn't use the encryption system during the testing, it was verified at TILab that encryption operation does not affect the functionality or the video quality. By saying that we can tell confidently that none of the encryption components all over the network should be replaced or upgraded.

6. Recommendations for further actions

Following the live test results analysis some recommendations for further actions can be suggested for those who would like to leverage the technology developed under OptiBand project.

Although we faced some difficulties with the live test results, we can tell that all the four OptiBand objectives were proven all over the technical parts of the project.

The technology works and can be commercialized within a reasonable time. However, although commercializing is out of the scope of this document we will mention that from marketing point of view, there are some doubts that were raised regarding the worthiness of the IPTV investments, which lost momentum.

6.1 Technical Recommendations

- The PDD implementation should migrate to a larger FPGA with more memory. This is required for enabling better shaping functionality and in order to enable better logic functionality with small GOPs. It was found that not all the DSLAM provides good shaping capabilities, and some of them assume the upper equipment (e.g. PDD) implements the shaping.

- R&D resources should be invested in order to have a better solution during zapping scenarios, where the FPGA should switch between streams. This was found as a non-trivial scenario.

- As for the live test itself, it is highly recommended to implement such a test using a much larger sample audience.

7. References

[1] Fröhlich, P., Egger, S:, Schatz, R., Mühlegger, M., Masuch, K., and Gardlo, B. QoE in 10 seconds: Are short video clip lengths sufficient for quality of experience assessment In Quality of Multimedia Experience (QoMEX), 2012 Fourth International Workshopon, Australia, July 2012.

[2] Gardlo, B., Hoßfeld. T., Schatz, R. and Ries M. Microworkers vs. Facebook: The impact of crowdsourcing platform choice on experimental results. In Quality of Multimedia Experience (QoMEX), 2012 Fourth International Workshop on, Australia, July 2012.