Multi-Layer quality assessment framework for P2PTV applications

René Serral-Gracià, Anny Martinez Manzanilla, Eva Marin-Tordera, Marcelo Yannuzzi, Xavier Masip-Bruin
Advanced Network Architectures Lab, Technical University of Catalunya (UPC), Spain
{rserral, annym, eva, yannuzzi, xmasip}@ac.upc.edu

***Abstract*—The success of Peer-to-Peer (P2P) overlay networks has led to its broad adoption on different scenarios, thereby increasing the set of features present on such protocols, e.g., audio and video-conferencing, streaming of multimedia content using Peer-to-Peer TeleVision (P2PTV), and so on. The wide acceptance of P2P multimedia streaming by the end-users has captured the attention of content providers, making it a suitable candidate to be adopted as technology for the offered multimedia services. Nevertheless, to deliver a proper service, content providers, need to quantify and assess the quality perceived by the users. In the past, this assessment was embedded into Service Level Agreements under the client/server paradigm, but with the introduction of P2P, these solutions became unsuitable. To overcome this limitation, we present a Monitoring and Management Framework to assess the Quality of Service and the Quality of Experience of multimedia traffic in any P2PTV streaming application. The proposed approach is designed to potentially assist management systems to actively assess the performance of existing streaming deployments. To demonstrate the usefulness of the system, we, as an use-case, deploy it in a real scenario where we analyze the performance of a particular streaming application in a P2PTV environment.**

I. Introduction

Network performance metrics have been used in Quality of Service (QoS) platforms to assess and guarantee the compliance of the Service Level Agreements (SLA), generally to ensure business peering agreements among Service Providers, but without any specific consideration to the underlying traffic, where the network treats equally bulk data transfers, such as web page downloads, and sensible traffic, such as real-time multimedia streaming. Then, with the increasing presence of multimedia traffic on the Internet, the need of quality assessment of multimedia traffic becomes apparent. However, most of the related works in this area only objectively analyze some aspects of the traffic, neither considering specific application restrictions nor the end-users subjective perception of the service.

In parallel to the increase of multimedia content on the Internet, the adoption of Peer to Peer (P2P) systems has grown exponentially due to the broad deployment of file sharing applications. This technology has raised the distribution of information on the Internet to a new level, by enabling efficient and autonomous cooperation among end users through an overlay network over the Internet. Participants in these systems are viewed as logical and functional equals, who transmit data among each other in the overlay regardless of the underlying network topology; in contrast with pure Client-Server (CS) protocols, where participants either consume or serve the content.

The growth of P2P overlays has motivated the development of new protocols and applications. In this paper, we focus on Peer to Peer TeleVision (P2PTV), a P2P technology used for television —and multimedia— broadcasting to end-users who subscribed to a P2P overlay. P2PTV has been adopted by a number of applications (e.g., Coolstreaming, PPlive, PPStream, SopCast, etc.) that aim at providing multimedia content to the users. The commercialization of such services, analogously to classic CS multimedia platforms, stimulates the desire of both end-users and content providers to receive feedback about the quality of the service delivered.

In this regard, the contribution of this work is the proposal of a Monitoring and Management Framework (MMF) that can be used in a P2PTV network to provide a full-fledged on-line quality assessment infrastructure. In particular, the proposed framework sets the basis to a complete management suite offering capabilities such as, debugging of P2PTV protocols, live reporting of perceived video quality, or assisting with billing and accounting processes based on quality assessment.

Then, as a proof of concept, we deploy MMF in the PlanetLab network with two different goals, first to analyze the correctness of MMF as a quality assessment platform, and second to evaluate the suitability of SopCast as a P2PTV broadcasting solution. Our results show that MMF can effectively assess the user perceived quality, and that SopCast has a fairly good performance in PlanetLab, with acceptable video quality in general.

The rest of this paper is structured as follows, Section II reviews the related work, then, in Section III, we present the management framework to perform P2PTV application assessment. Later, we describe the tests and the testbed used to validate our proposal. In Section V, we detail the experimental deployment of SopCast in the PlanetLab network, and finally, in Section VI we conclude and outline our future work.

II. Related Work

Since the explosive growth of multimedia content on the Internet, content providers are concerned about the quality of the offered service. Initially, this concern was addressed by enforcing Service Level Agreements (SLA) between network

This work was partially funded by Spanish Ministry of Science and Innovation under contract TEC2009-07041, and the Catalan Government under contract 2009 SGR1508.

operators and content providers and by setting specific policies over multimedia traffic. As a consequence, several solutions appeared in order to monitor the offered Quality of Service (QoS) by the network, as well as, to manage these SLAs. For instance, in [1] the authors present a passive traffic analysis system to assess the network performance.

During the quick expansion of multimedia content streaming, it was clear that solely measuring the network performance was not sufficient to assess the user satisfaction, given that the subjective user perception might considerably differ from the measured network performance. Therefore, the focus shifted from the resource intensive network performance monitoring, to traffic analysis at the edges of the network, in order to assess the user perceived quality by means of Quality of Experience (QoE) techniques. In the particular case of video quality assessment, there are several solutions to assess the user perceived QoE of IPTV services, e.g., in [2], the authors propose a framework for parametric assessment of the QoE of IPTV services from a content provider perspective. All these proposals provide an objective measurement of end-user perceived QoE which, in general, it is quantified in the same scale as the Mean Opinion Score (MOS) [3], that is, a value between 1 (for bad quality) to 5 (for perfect quality) of the received content.

Regarding the design of a full-fledged multi-layer MMF in a P2PTV overlay network, to the best of our knowledge currently there are no other efforts in this direction. However, with similar objectives, De Vera et. al in [4] propose a full video network monitoring suite, which is aimed at QoE estimation using PSQA [5]. Our work differs from theirs in the sense that we focus on P2PTV rather than CS applications.

III. Monitoring and Management Framework

In this section we present the Monitoring and Management Framework (MMF), a full-fledged quality assessment framework of P2PTV applications that perform multi-layer quality assessment in an efficient monitoring infrastructure. To this end, MMF is designed to operate using two different modes, Integrated Monitoring (IM) and Decoupled Monitoring (DM). IM is used when MMF is embedded into the system and can interoperate with it, while DM is devised as an external helper application, running in a separate process, to support content providers and developers in the evaluation of already existing P2PTV solutions.

Depending on the chosen monitoring approach, different information of the system will be available. In particular, when using IM the monitoring system has access to first-hand information about the internal status of the P2PTV application, e.g., video stream and P2P network status, and therefore, it can tweak the application behavior to improve the overall end-user experience. On the contrary, with DM the system must use estimates and reverse engineering techniques to infer some of that information, without the possibility of directly affecting the application behavior.

To illustrate the general structure of MMF, in Fig. 1 we observe that MMF has two different layers, the Collection (CL) and the Management Layer (ML), CL is in charge of collecting and analyzing the traffic in order to compute the different multi-layer performance metrics, while the ML is responsible for the management of the overlay node depending on the information gathered by the CL.

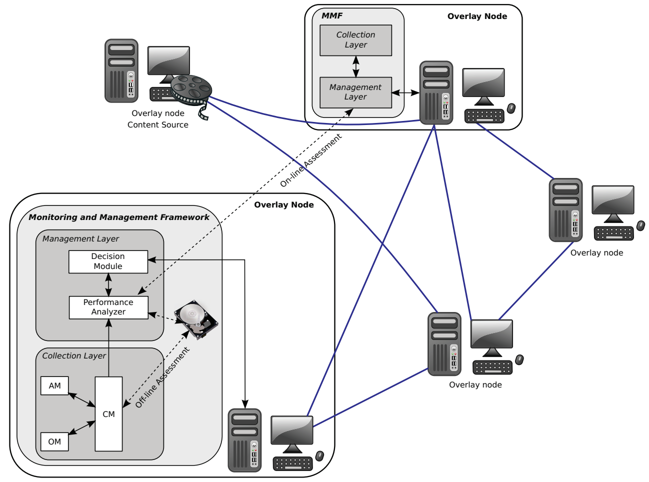


Fig. 1. Sample P2PTV network, with the different MMF components

Complementing the modes of operation, and depending on the requirements of the P2PTV platform we differentiate two assessment mechanisms *(i)* on-line assessment and *(ii)* off-line assessment—both marked in dashed lines in the figure. In on-line assessment, MMF performs all the monitoring and management tasks in real-time distributed across all the nodes the overlay, while the off-line assessment separates the monitoring and management into two different stages, first the monitoring, which continues to be distributed and performed on-line, and second the performance evaluation at the end of the P2PTV session, which is centralized and executed off-line.

A. Layer description

MMF is composed by two different layers with two clearly separated roles, on the one hand there is the monitoring tasks that are performed by the Collection Layer and on the other hand the Management Layer, responsible of assisting the P2PTV management tasks.

1) Collection Layer: The Collection Layer (CL) abstracts all the tasks related with the collection of multi-layer information within the node, e.g., it collects per chunk statistics in the P2P layer, while in the application layer it gathers specific per frame data.

The CL is composed by three different entities:

- *Overlay Monitor (OM):* the OM collects statistics about the received information at the P2P layer. Specifically, it computes the timestamps and the loss ratios of the received chunks of video and audio packets, list of active peers serving each chunk, and required network resources for the streaming.
- *Application Monitor (AM):* analogously to the OM, the AM collects information from the application perspective, e.g., buffer under/overrun status, used video codec, type

of the dropped frames (I, P, or B-frames), presentation timestamp of the frames.

- *Collection Manager (CM):* the CM gathers the information from the AM and the OM and sends it to the Management Layer (when performing on-line assessment), or to the storage subsystem (while using off-line assessment).

The details on how the information is gathered and processed depend on the used assessment mechanism and will be described later in this section.

2) Management Layer: The Management Layer (ML) is responsible of computing the quality metrics by merging the information obtained by the different CLs. This procedure is performed in four different steps, first step is to gather the information collected by the CM in the neighboring peers, the second is to compute the different overlay and application metrics, the third is to detect the periods where the quality drops below predefined boundaries, and finally to take the necessary actions in order to minimize such unsatisfaction periods.

To implement the above steps the ML is composed by two different entities.

- The *Performance Analyzer:* is in charge of performing the first and second steps, thus gathering the neighbor information received from the CM and computing the requested multi-layer metrics, then it reports the obtained results to the decision module.
- The *Decision Module:* decides whether the user is experiencing periods of unsatisfaction (step three above) in case that the quality is below the preset limits, and taking the necessary actions if needed (step four).

In more detail, in the Performance Analyzer each peer node can compute the multi-layer metrics by computing latency at overlay layer—considering information exchanged among the peers—and application layer—considering the video contents once received at the destination—using timestamps, computing the video degradation, e.g., by detecting errored packets or missing frames at reception. Analogously to CL, the particular approach used in order to compute these metrics depends on the used assessment mechanism which will be detailed in Section III-B.

Regarding the Decision Maker, it will determine the tolerable performance metric boundaries at the different layers for the given service as configured by the P2PTV overlay administrators. Then, the Decision Module, stores the computed metrics into a database, and in the case that the preset quality boundaries are not met, an alarm is raised and the system decides the actions to be taken. This database can be queried by different external entities, e.g., by a Network Management System—in the service provider premises—and different actions can be triggered, e.g., to update the resource reservation, or the Traffic Engineering policies to maximize the end-user satisfaction. It is not our goal to detail the internal workings of the action triggering system in order to assist in solving potential quality issues, thus in the rest of this paper we focus on the Performance Analyzer and on the detection of unsatisfaction periods, leaving the action triggering as part of our future work.

B. Quality Assessment mechanisms

Depending on the mechanism used to deploy MMF the capabilities and methods to collect the information will vary. Currently MMF considers three different alternatives: IM with on-line performance assessment, DM with on-line performance assessment, and DM with off-line performance assessment.

1) Integrated Monitoring: When using IM the only supported performance assessment method is on-line with real-time computation of the quality metrics. To perform such operations we propose the following methodology. The video generator embeds within the P2PTV protocol the necessary information to compute the metrics, such as the emission timestamp for each frame (and chunk), and a per frame (and chunk) sequence numbers. Simultaneously, each peer computes basic metrics, e.g., latency, is computed at overlay and at application layers by timestamping the packets and frames once received. In the case of video degradation it is computed by counting lost packets or frames not reaching their destination, the amount of losses per time unit can be computed by using the sequence numbers. Finally, another relevant metric refers to jitter, which is obtained through the differences in the latency of consecutive packets or frames, at overlay and application layers respectively. From this point computing the perceived QoE can be done by using already existing alternatives, e.g., using the method described in [4].

2) Decoupled Monitoring: Opposed to IM, when using DM the system cannot be integrated into the P2PTV protocol, hence MMF must perform the assessment through an external procedure, with the consequent increase in the system's complexity. However, the advantage of this approach is that it is not necessary to perform any modification to the used P2PTV application.

With DM, MMF defines two different performance assessment methods, on-line, which, analogously to the case using IM, monitors the quality metrics simultaneously to the video session, and off-line, which collects detailed information of the video stream on each peer, to process them once the video session is finished.

MMF needs to share some information among the peers in order to compute the necessary metrics. In particular we propose to transmit for each chunk and each frame the following information:

- Identifier: The identifier uniquely distinguishes the different chunks and frames in the stream, they are generated by applying a hash function to specific fields within the packet or frame payload, the field selection is dependent on the P2PTV protocol, in Section IV-B we detail the particular details of the identifier generation for the use case of SopCast.
- Generation and reception timestamps: Each peer in the overlay timestamps all the chunks and the frames within

the video stream in order to compute the different time related metrics.

- Sequence number: Analogously to IM, the sequence number will be used to compute the loss ratios incurred in the video transmission.

For the frames MMF also stores the frame type, which will be useful in order to compute the QoE. It is important to notice that depending on the used P2PTV protocol some of the above information may not be available to MMF. Particularly, most of the time having first-hand application information is not feasible, in such a case, MMF will restrict the analysis to the overlay metrics, leaving out of the study the more detailed application layer analysis.

Independently of using on-line or off-line monitoring, MMF uses the same methodology in order to collect metric information. First both OM and AM in the content generation peer will collect the transmitted information, this is performed by gathering information in the network interface, then identifying the different packets and frames through deep-packet inspection, generating the identifiers, timestamping, and generating sequence numbers. Once the a predefined amount of packets have been forwarded—chunk—, the CM aggregates this information and generates a source *chunk descriptor* $\mathcal{C}_{\mathcal{S}_i}$ that holds all this information. Simultaneously, the rest of peers, perform the same operations with all the received video traffic from the other peers in the overlay generating a destination *chunk descriptor* $\mathcal{C}_{\mathcal{D}_i}$.

a) On-line monitoring: When using on-line monitoring the $\mathcal{C}_{\mathcal{S}_i}$ is sent by any peer when sending a particular chunk to another peer, this way, each peer can autonomously compute the requested metrics. The metrics are computed by matching all the chunks and frame identifiers from $\mathcal{C}_{\mathcal{S}_i}$ against $\mathcal{C}_{\mathcal{D}_i}$. If a particular identifier can be found in $\mathcal{C}_{\mathcal{D}_i}$ then the requested metrics are computed. However, if the identifier is not found then the chunk or frame is considered lost, and then, by using the sequence number it is possible to detect consecutive periods with frame drops in the network.

b) Off-line monitoring: Opposed to on-line monitoring, when performing off-line monitoring each peer stores locally either $\mathcal{C}_{\mathcal{S}_i}$ or $\mathcal{C}_{\mathcal{D}_i}$, and once the video session is finished, all the information is gathered and post-processed off-line, using the same methodology as previously in on-line monitoring.

IV. Testing environment

In this section we discuss the used P2PTV application and the testbed used to evaluate the proposed MMF. To this end, we deployed SopCast in the PlanetLab network, where we streamed a video session among 137 peers of the SopCast P2P overlay.

A. MMF Deployment

Given that we do not have access to the source code of SopCast or to its specification, we deploy our MMF in the Decoupled Monitoring mode of operation, with off-line assessment.

We set up the SopCast overlay with 136 PlanetLab nodes and a server streaming a 300 seconds long video. During the streaming period, we collected the traffic at the content generator and also at the 136 peers. In parallel, we also dumped the received video to each node's local disk, in order to have a reference of *"perfect knowledge"* against which we can compare the performance of MMF.

B. Reverse engineering SopCast

The DM off-line quality assessment was performed by identifying SopCast video packets in all the traces, and off-line running the Performance Analyzer process to compute the performance metrics. The procedure required some reverse engineering of the SopCast protocol, given that it is a closed source application. After deeply studying the generated traffic we realized that SopCast recodes the received frames in each peer, this fact complicates substantially the frames and chunks identification and later the generation of the required identifiers by MMF. However, after a closer inspection, we observed that the protocol keeps an internal sequence number that is sent within the header and it identifies the same video frame in all the nodes in the overlay, which we use as identifier for the frame.

By matching the different descriptors we were able to compute the overlay quality metrics, while the user-perceived QoE was assessed by analyzing the video traces delivered by the multimedia player, in the analysis we considered chunks of 10Kbytes (about 200ms of video). It is important to notice that with access to SopCast's protocol specification, or in the case of using IM, we could have avoided such reverse engineering and accomplished the metric computation more easily.

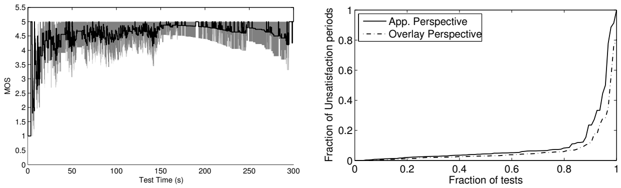
V. Use case: SopCast Performance Analysis

As a proof of concept, this section deploys MMF in the PlanetLAB slice introduced before to perform a throughout study of SopCast as a P2PTV media streaming solution. The experiments are focused on the perceived video quality.

With these experiments we demonstrate as a use case, how MMF can be seamlessly used independently of the underlying protocol to assess its performance. Nevertheless, the results shown in this section cannot be generalized to SopCast but to our particular use in the PlanetLab testbed.

A. Video perceived quality - User satisfaction

In order to evaluate both MMF and SopCast within the PlanetLab network, we first measure the overall MOS as perceived by the end-users over the whole test duration. To this end, in Fig. 2(a) we plot the average MOS evolution and its confidence interval extracted from all the tests spanning the whole duration of the video session. We obtained these quality values from the application perspective by computing the PSNR of the received video packets and by using [6] we extracted the equivalent MOS for each video frame. Since such analysis would not be possible in a real deployment we use these values in the whole section as a reference of "perfect



(a) Average MOS (b) Unsatisfaction periods

Fig. 2. Overall average MOS and Unsatisfaction periods from overlay and application perspectives

knowledge" about the user perceived video quality, and use it in order to compare the obtained values by our MMF.

As it can be noted in the figure during the warm-up phase of the P2PTV protocol, i.e., the first seconds of the session, the perceived quality is noticeably lower than the one obtained once the protocol is fully operative, where the quality is consistently above 4 in average, which indicates good user perception of the service regarding video quality.

To highlight the differences in the perceived quality, in Fig. 2(b) we plot the Cumulative Density Function (CDF) of the total fraction of unsatisfied periods (Y-Axis) per test (X-Axis), for convenience we normalize the tests to 1. The figure underlines that around 80% of the tests have unsatisfaction periods below 10%, while the other 20% imply that either the P2PTV protocol or the network used in the deployment have performance issues, given the larger periods of unsatisfaction for the users during the experiment.

These results show that, in PlanetLab, once the node converges to proper video quality it can cope with the necessary user demands not incurring in any more noticeable service disruptions. Nevertheless, as we notice in the following section, the overlay continues to observe packet drops, which are successfully recovered by the video codecs, masking its effects to the final users.

B. P2P Overlay chunk degradation

In a real deployment, it is not always possible to access application information. Besides, the user-perceived values, sometimes it can be useful for the overlay management to infer the overlay performance directly from the raw metrics such as chunk degradation. We consider that a chunk is degraded (or lost) when the received amount of data is higher than a threshold δ, where δ represents the sensitivity in the estimation of the delivered quality. In particular, the loss threshold δ has a two-fold utility, *i)* for assessing the compliance with different SLA, e.g. setting QoS management strategies, and *ii)* set up δ to align it with the resilience in error concealment of the video codec.

In Fig. 3 we show the CDF of the probability (Y-Axis) of the relative number of corrupted chunks per test (X-Axis) for the different δ. As we notice, the results are aligned with the intuition that bigger δ implies lower chunk disruption. In the effort to model the behavior of the chunk degradation in a SopCast overlay, we also plot in the figure the different normal distribution fits for each δ. As it can be noted, for low values of δ the chunk degradation tightly follows a normal distribution.

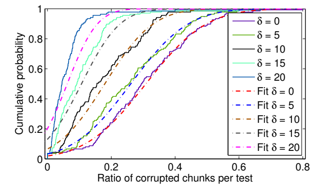


Fig. 3. Ratio of corrupted chunks per δ

In particular, we observe that in 95% of the cases SopCast users will experience less than 60% of corrupted chunks for $\delta = 0$, while in the best case, i.e., $\delta = 20$, the experienced degradation for 95% of the users will be below 21%.

VI. Conclusion

In this paper, we presented MMF, a framework to perform P2PTV quality assessment by using a formalization of multi-layer performance metrics. The deployment of MMF has many potential applications, e.g., to integrate it in the P2PTV protocols for on-line quality assessment, to perform billing and accounting based on the media quality, to analyze P2PTV service deployment feasibility, to study end-user satisfaction, or to profile P2PTV protocol requirements. In this work, and as a proof of concept, we used the MMF and the multi-layer metrics to analyze the performance of SopCast in the PlanetLab research network. We have found that SopCast provides a scalable protocol, which with relatively low amount of resources can provide a good framework for live P2PTV broadcasting.

After the analysis performed in this work, we left a number of open issues which require further development. In particular, we plan to integrate MMF into an existing P2PTV application, e.g., PeerCast, in order to compare the efficiency and accuracy of both IM and DM approaches and evaluate the advantages and shortcomings of each alternative.

VII. Acknowledgment

The authors would like to thank both Yue Lu and Fernando Kuipers from Delft Technical University for the access to the SopCast traces, along with many technical discussions about the performance assessment with the P2PTV system.

References

[1] René Serral-Gracià, Marcelo Yannuzzi, Yann Labit, Philippe Owezarski, and Xavier Masip-Bruin. An efficient and lightweight method for service level agreement assessment. *Computer Networks*, (54):3144–3158, Dec 2010.

[2] Kazuhisa Yamagishi and Takanori Hayashi. Parametric packet-layer model for monitoring video quality of iptv services. In *IEEE International Conference on Communications. ICC '08.*, pages 110–114, May 2008.

[3] ITU-T Recommendation G.107. The E-model, a computational model for use in transmission planning, 03/2005.

[4] Daniel De Vera, Pablo Rodríguez-Bocca, and Gerardo Rubino. Automatic quality of experience measuring on video delivering networks. *ACM SIGMETRICS Performance Evaluation Review*, 36(2):79–82, 2008.

[5] Samir Mohamed and Gerardo Rubino. A study of real-time packet video quality using random neural networks. *Circuits and Systems for Video Technology, IEEE Transactions on*, 12(12):1071–1083, 2002.

[6] Jirka Klaue, Berthold Rathke, and Adam Wolisz. Evalvid - a framework for video transmission and quality evaluation. In *In Proceedings of the 13th International Conference on Modelling Techniques and Tools for Computer Performance Evaluation*, pages 255–272, 2003.