

Performance Issues in Video Streaming Environments*

Giuliano Casale, Paolo Cremonesi, Giuseppe Serazzi, Stefano Zanero
Dipartimento di Elettronica ed Informazione
Politecnico di Milano
via Ponzio 34/5, I-20133 Milano, Italy
{casale, cremonesi, serazzi, zanero}@elet.polimi.it

Abstract

The main motivation for applying capacity planning techniques to the design of Internet-based video services is that very often these systems incur significant delays that decrease the offered Quality of Service below an acceptable level. However, the very high fluctuations of the Internet traffic make the identification of the correct capacity of the server and network components a very hard task to be accomplished.

In this paper we analyze the metrics used to evaluate the performance perceived by users of video streaming applications and we introduce and validate some new metrics. These variables may be used to parameterize the models required by capacity planning studies. A characterization of users' behavior for live and stored video streams is presented. The knowledge of the behavior of these two streams is important to identify the capacity to be allocated in order to satisfy the QoS requirements. Several issues concerning the security of video streaming are also discussed.

1. Introduction

The latest technical advances in high-speed networks, data compression and signal processing techniques make it possible the delivery of video/audio streaming through Internet to a very high number of concurrent users. However, the highly dynamic environment typical of Internet, either in terms of number or type of processing requests competing for the resources, make the achievement of the performance levels needed by streaming applications a very difficult task to be accomplished.

Usually a degradation in the performance, e.g. a very high response time, of a typical web-based application does

not affect the results of the application itself, but rather influences the user's perception of the QoS. This is not the case of video streaming workload. Indeed, when severe degradations in throughput and response time are present, not only the QoS decreases, but also the visual representation of the streaming could degrade below an acceptable level.

The *bursty* behavior of the Internet and web traffic is a main source of server congestion and lack of network bandwidth and thus of *performance degradation*. Unfortunately, skews in data access that create very high load spikes are typical of *live* and *stored* streaming transmission. For example (as we will show in the following) during events like live shows or sport matches, spikes of requests up to twenty times larger than the average values are observed. In such a condition, servers and Internet components get overwhelmed and very long delays are experienced. Also the traffic generated by the streaming of stored media, e.g., Video-on-Demand, exhibits a bursty behavior showing some periodicity (e.g., daily, weekly) and a duration of the bursts longer than in live streaming.

The evaluation of the quality of video streaming on the Internet involves the analysis of a number of specific factors. Number of delayed and lost packets, rebuffering during playback, bandwidth delivered and bandwidth received are among the variables that should be measured on top of the usual performance indices like end-to-end response time, throughput and resource utilization. Furthermore, the streaming content itself is more complex than the content of a regular web application: a larger amount of data is transferred, the data is much more sensitive to latency, multiple data types are considered (audio and video), different default delivery protocols are involved (UDP vs. TCP) and multiple software technologies are used (Real Media, Windows Media, QuickTime).

An important concern of a streaming application designer is not only the system performance but rather the quality of service perceived by its users. Usual metrics are focused on the static aspect of the video quality (e.g., ar-

* This work was supported by the Italian Ministry of Education, Universities and Research (MIUR) in the framework of the FIRB-Perf project.

tifacts introduced by the compression algorithm) and do not consider the dynamic aspects of the streaming (e.g., video fluidity, audio-video synchronization) that influence the perceived quality. Typically they compare the reduced quality of a coded video-stream with a data loss compression algorithm with respect to the original non-compressed video-stream and analyze the artifacts introduced in every frame. Such metrics do not consider the degradation of performance due to network congestion (e.g. insufficient bandwidth, packet delay, packet loss).

The most accurate metrics used to evaluate the quality of media objects received by end users are the *subjective metrics*. Subjective metrics standardized in ITU-R Recommendation BT.500 [14] have been used for over twenty years to evaluate video quality in television services. More recently, the ITU-T developed Recommendation P.910 [15] to standardize metrics for multimedia quality assessment. Although they are the most realistic evaluations, subjective metrics are difficult to obtain, since they require the organization of multiple tests on several subjects.

Alternative to subject metrics are the *objective metrics*. The main purpose of objective quality assessment is to provide a set of quantitative metrics that can predict the perceived quality from the user's point of view. The goal is to develop a metric without requiring access to the original media.

In the last two decades, a lot of objective metrics have been proposed to assess image and video quality. The easiest way to give a quality value is to compute some simple statistics on the numerical errors between the distorted video and a reference video.

Widely used statistics are Mean Squared Error (MSE) and Peak Signal to Noise Ratio (PSNR). However, MSE and PSNR do not correlate well with subjective quality measures because human perception of distortions and artifacts is unaccounted for [8]. A major emphasis in recent research has been given to a deeper analysis of the Human Visual System (HVS) features [23]. Although HVS is too complex to fully understand, the incorporation of even a simplified model into objective measures reportedly leads to a better correlation with the response of the human observers. Another important factor for the development of any video quality metric is the flexibility for practical implementations. Some of the metrics consider only some special types of distortions or special video coding methods [34, 17]. The implementation of a practical video quality assessment metric is difficult because of the computational complexity.

Traditional objective measurement systems are focused on the collection of data concerning the quality of digitally compressed video systems, because video-stream coding may introduce visually noticeable artifacts [36].

In this paper we address several issues associated with the problem of the performance prediction and evaluation

of video streaming applications and we analyze their impact on the perceived quality of service [10, 11, 21].

The paper is organized as follows: In Section 2 the metrics used to evaluate video streaming are described. New metrics able to describe the dynamic aspect of a video streaming are introduced. We correlate subjective metrics with measured data in order to validate the users' perceived quality with respect to the objective metrics.

In Section 3 we present a workload characterization study based on the traces collected on a popular Italian video streaming provider. Logs of either live and stored streaming transmissions have been collected. The qualitative behavior of the traffic and the correlation between live and stored streaming behavior is also shown.

The security issues typical of video streaming environments are discussed in Section 4. The problems due to the use of UDP and the performance degradation introduced by TCP are considered. Digital rights management techniques for streaming transmission are also analyzed.

Section 5 concludes the paper with a summary of our work.

2. Performance Indices

With the shift in technology from local video-streams (e.g., DVD) to network-transmitted video-streams (e.g., Video-on-Demand), a change in the types of objective metrics is required. Traditional objective metrics, while still necessary, are not sufficient to measure the quality of video streams transmitted over the Internet. In particular, errors or losses during the transmission introduce distortions whose effects must be considered together with the influence (e.g., delays) of the network system [37, 38].

In sections 2.1 and 2.2 a number of new objective metrics are proposed to assess video quality. The metrics are appealing because they can be collected without the need of the original video and because they take into account the effects of network transmission on the quality evaluation technique. Indeed, packet losses and packet delays have a strong influence on the reduction in quality of a video-stream. Together with the objective metrics, two new subjective metrics are proposed that can be easily derived from the objective metrics.

In section 2.3 the two derived subjective metrics are compared against the real subjective metrics obtained from a panel of users.

2.1. Metrics

Objective measurement methods implement algorithms that measure video quality usually based on the comparison of a source and a processed sequence. The algorithms, referred to as models, may incorporate characteristics of the

human visual system in an attempt to systematically measure the perceptible degradation occurring in the video.

Objective quality metrics can be classified according to the amount of information required from both the original and the received videos. Depending on the side information required, three generic classes of objective metrics can be described:

Full reference metrics (FR) The quality evaluation system has access to both the original video and the received video. Full-reference metrics perform a frame-by-frame comparison between a reference video and the video received by the end user. They require the entire reference video to be available, usually in uncompressed form. This is quite an important restriction on the applicability of such metrics. Typical metrics within this class are the PSNR (Peak Signal-to-Noise Ratio) and the MSE (Mean Square Error). The measurement environment of this type of metric is outlined in Figure 1.

No-reference metrics (NR) The evaluation system has no access to any information regarding the original video. This makes it possible to measure video quality of any video, anywhere in any existing compression and transmission system. This type of metrics is the most promising in the context of video broadcast scenario, since the original images or video are in practice not accessible to quality evaluation system. The measurement environment of this type of metric is outlined in Figure 2).

Reduced reference metrics (RR) Reduced-reference metrics lie between the two type of metrics previously described. The evaluation system has access to a limited number of side information regarding the original video, i.e. features or descriptors extracted from the original media. In Figure 3 the measurement environment of this type of metric is outlined.

2.2. Experiments

In order to measure the end-to-end performance of video-streaming applications, a client-server distributed measurement tool has been developed (Figure 4). The tool is named VPET (Video Performance Evaluation Tool).

VPET has been used to describe the quality of video-stream services as a function of the network and video-stream parameters (e.g., available bandwidth, coding format, communication protocol, video server, frame rate, percentage of packet loss).

A VPET client is composed of a number of Visual Basic agents for the collection of objective and subjective measurements. Each client is an enhanced and instrumented Mi-

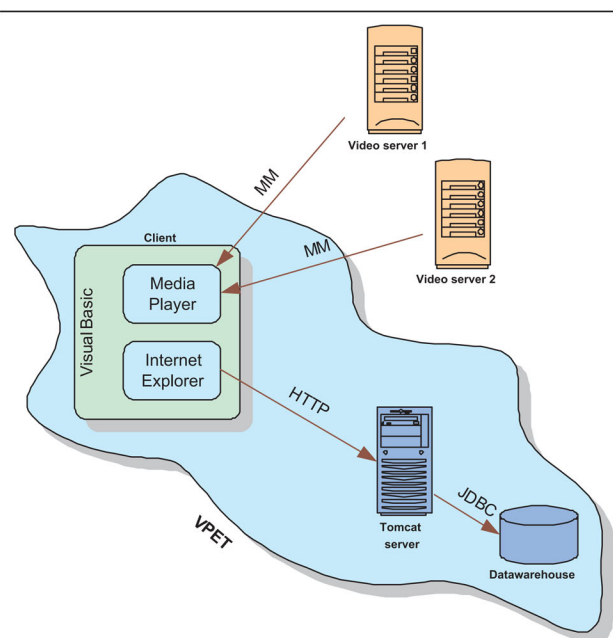


Figure 4. VPET architecture

crosoft Windows Media Player. The players can be used to visualize any Microsoft Media Streaming (MMS) video. During the visualization, the player collects a number of objective and subjective metrics. Each client is connected to Internet through a different *connection type* (e.g., ISDN, xDSL, cable, backbone), from different *geographical locations* and through different *providers*.

All the data collected by the players are stored in a centralized database and analyzed in order to extract statistics.

The metrics collected by the clients can be classified into two categories: instantaneous and global. *Instantaneous subjective metrics* are events signaled in real-time by the users during the visualization of a video:

- audio–video de-synchronization
- video is still
- audio is still.

Global subjective metrics are collected by VPET clients at the end of the video. Users are asked to rate a number of aspects about the video:

- α_{start} : start-up time
- α_{res} : video-clip resolution
- α_{jitter} : video-stream fluidity
- α_{audio} : audio quality
- α_{overall} : overall quality.

Ratings range from 0 (totally unacceptable) to 1 (complete satisfaction).

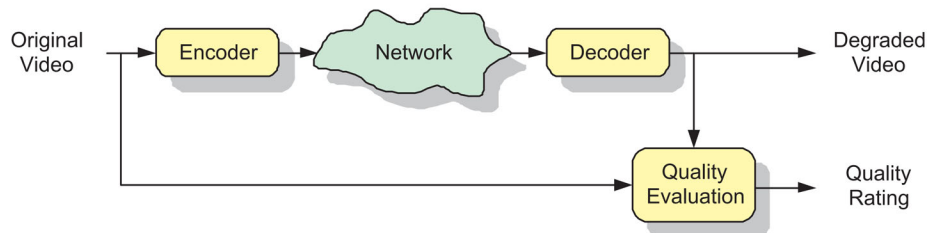


Figure 1. Full reference video quality measurement environment

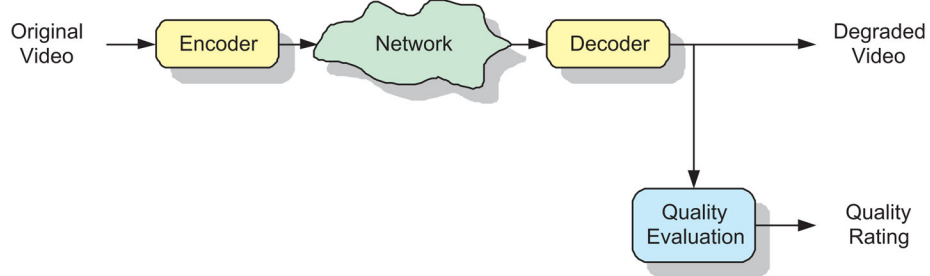


Figure 2. No-reference video quality measurement environment

Instantaneous objective metrics are collected automatically by the VPET client by sampling periodically, e.g., every second, the player engine. Consider a set of n instantaneous measurements labelled from 1 to n . Each sample is collected at time, t_i with $i \in 1 \dots n$:

- $f(t_i)$: the number of frames per second (fps) received by the client. Television and movies display video at 30 frames/sec, which is the rate at which humans discern full motion
- $p_{\text{tot}}(t_i)$: total number of packets transmitted by the video server since the beginning of the video up to time t_i . This information can be obtained by the player by looking at the sequence number of each packet
- $p_{\text{rec}}(t_i)$: number of packets received correctly since the beginning of the video up to time t_i . The difference between the packets transmitted and the packets received is caused by late and lost packets. Late packets are received by the client's play buffer but too late to be processed: not only the client cannot process the packets, but also the packets steal bandwidth to other packets. Lost packets never reach the client. Both late and lost packets have a very negative effects on audio and video, including pixelation, jitters, frozen video, audio popping or audio static
- $b(t_i)$: available bandwidth, measured in bits per second (bps), at time t_i .

- τ_i : movie time at time t_i . Each frame is labelled with a time stamp τ . The player's goal is to preserve the correspondence between movie time τ and real wall-clock time t . In times of plentiful resources, the player can meet the goal. When bandwidth is scarce, the player may reduce frame quality and/or drop frames to keep up to the pace of the movie.

Global objective metrics are measured by the VPET player at the begin or at the end of the video.

- b_{max} : encoded bit rate
- f_{max} : encoded frame rate
- d : integer specifying the width of the video stream, in pixels.
- t_{start} : startup time. Video players offers client-side buffering in order to take into account for network fluctuations. Buffering results in a delay at start-up time, while the buffer fills. The startup time is the time elapsed from the instant in which the user presses the play button until the video begins
- T : the video duration.

Additional metrics are used as reference metrics:

- f_{ref} : reference frame rate
- d_{ref} : reference video width.

The values of the reference metrics depends on content and type of the video clip (e.g., news, movies, sports). Video

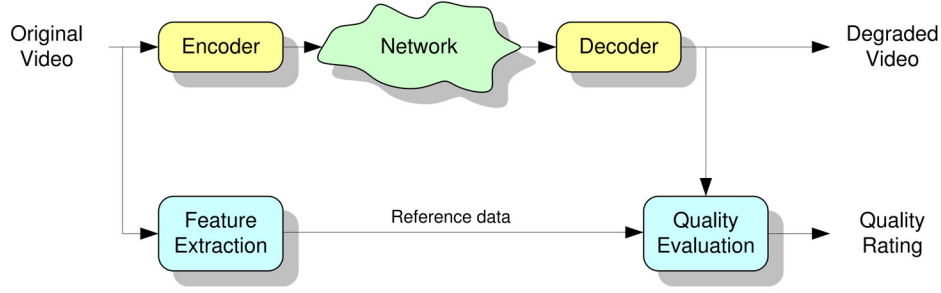


Figure 3. Reduced reference video quality measurement environment

	d_{ref} (pixels)	f_{ref} (fps)
Video conference	352	15
Static video	480	25
Dynamic video	720	25

Table 1. Reference video widths d_{ref} and frame rates f_{ref} for three categories of videos

content can vary from static to highly dynamic [1]. Example of static videos are video conferences, news clips or interviews, where there are few changes between successive frames. On the other hand, dynamic videos are fast changing video sequences, such as action movies or sports scenes. According to the dynamism of the content, videos can be classified in

- video conferences (e.g, videophone)
- static videos (e.g., news, drama and comedy movies)
- dynamic videos (e.g., action movies, sports)

The reference metrics for the above classes are shown in Table 1.

We are now interested in deriving two subjective metrics, i.e., α_{jitter} and α_{start} , from the VPET objective measurements. These two metrics belong to the subjective categories, but their values are obtained directly from the objective VPET measurements.

The first derived metric is $\hat{\alpha}_{\text{jitter}}$ which takes into account the jitter, i.e., quick fluctuations in some aspects of the video stream resulting in a video frame rate that is not stable. Jitter results in a video stream with frames freezing or accelerating with respect to the real wall-clock time. The example of Figure 5 shows a video with a small jitter. The video, after a transient in which the video time τ is slower with respect to the real time t , becomes stable with video time running at the same speed as real time. On the contrary, the example of Figure 6 shows a highly jittering video. The pro-

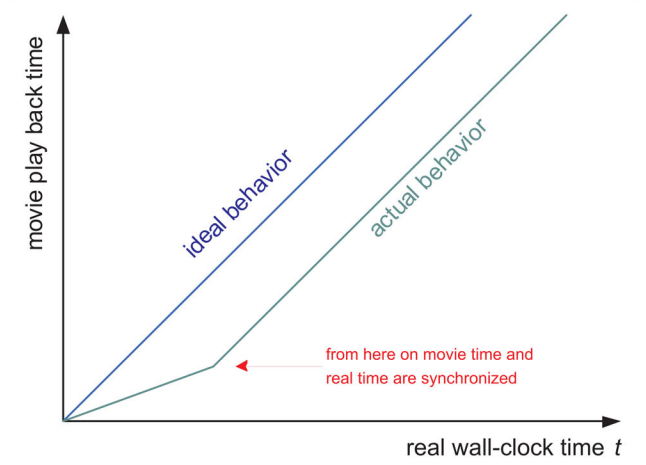


Figure 5. Video with a small jitter

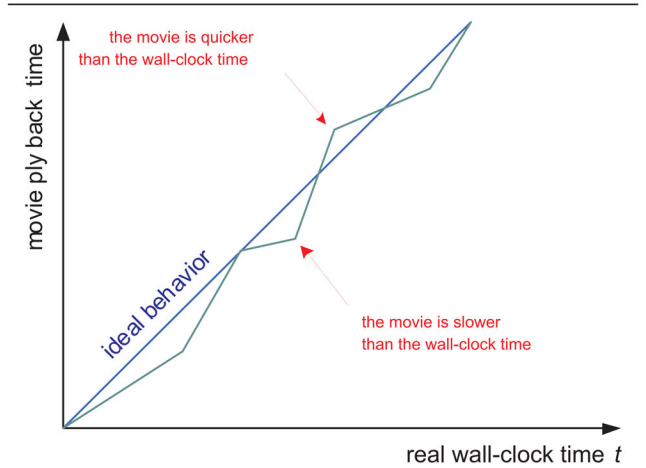


Figure 6. Video with a large jitter

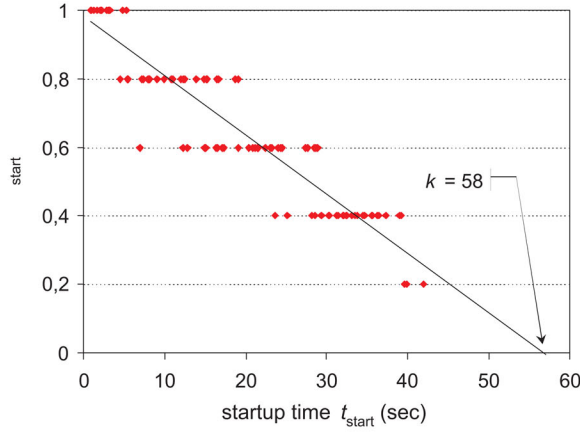


Figure 7. Start-up time validation: the linear regression indicates a value of $k = 58$ seconds to be used in equation (2)

posed estimate for the jitter quality metric is $\hat{\alpha}_{\text{jitter}}$

$$r(t_i) = \frac{\tau(t_i) - \tau(t_{i-1})}{t_i - t_{i-1}} \quad i = 2 \dots n$$

$$\hat{\alpha}_{\text{jitter}} = \sqrt{\frac{\sum_{i=2}^n r(t_i)^2}{n-1}} \quad (1)$$

where $r(t_i)$ is a measure of the video and real temporal displacement between two consecutive frames.

The second derived metric subjective is

$$\hat{\alpha}_{\text{start}} = \begin{cases} 1 - \frac{t_{\text{start}}}{k} & t_{\text{start}} < k \\ 0 & t_{\text{start}} \geq k \end{cases} \quad (2)$$

where k is a time parameter that needs to be estimated by real measurements (see the next section).

2.3. Validation

We selected 27 male participants for the evaluation experiment, between 19 and 62 years old. It was essential to select homogenous group of users. Indeed, users with different background have different expectations: we restricted our sample to users who declared to use the Internet for at least 1 hours per day.

Participants were asked to visualize a set of video clips and rate the performance for each video. Participants gave feedback on the video clips performance through:

- Interaction with the video player during the visualization of the video clip. An interface was developed to

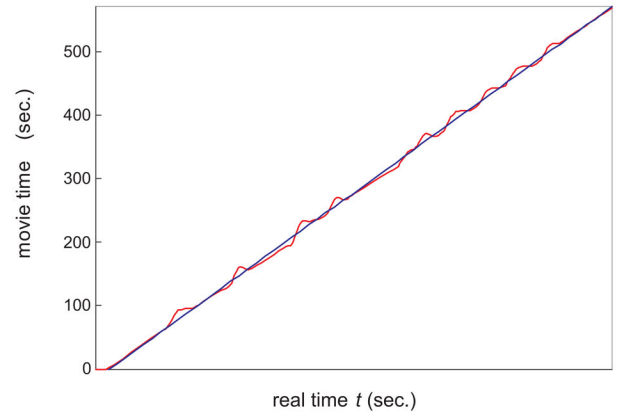


Figure 8. Jitter validation: equation (1) suggests a value of the coefficient $\alpha_{\text{jitter}} = 0.4$

register ratings during and at the end of the clip. The interface contained grey buttons labelled "Video is still", "Audio is still", "Audio and video are not synchronized". Participants were asked to click one of the buttons whenever one of the events happened.

- Compilation of a final evaluation form. A total number of 132 evaluations have been collected: 106 measurements using the university network, 8 measurements using an ADSL connection (640 Kbit/s) and 18 measurements using a dial-up connection (56 Kbit/s).

Figure 7 shows the startup subjective indices α_{start} measured with VPET (dots). The linear fitting confirms the model of (2) and suggests a value of $k = 58$ seconds to be used in the $\hat{\alpha}_{\text{start}}$ model.

Figure 8 shows, as an example, the instantaneous indices τ_i (vertical axes) and t_i (horizontal axes) for one video clip. The jitter model of (1) suggests a value of $\hat{\alpha}_{\text{jitter}} = 0.4$ consistent with the subjective value of α_{jitter} metric provided by the user.

3. Workload Characterization

Capacity planning techniques should be applied in order to identify the adequate capacity of server and network bandwidth to be allocated to maintain the expected response time of a streaming application at an acceptable level. Several solutions can be adopted. For instance, in order to avoid site congestion and lack of bandwidth, servers may be dynamically added or removed depending on the intensity of workload peaks. However, dimensioning hardware components and the number of media servers can be done properly only if the workload to be processed is well understood. Since the resource consumption of most audio/video

streams is usually constant, load variabilities in media services are often determined only by fluctuations in the input workload. Therefore, the capacity planning of streaming media services requires a comprehensive and accurate workload characterization activity.

Despite this, only a limited number of papers in the literature have tried to characterize streaming media workloads (see e.g. [4, 24, 28, 31, 33]), and most of the research efforts have focused on services offering stored streams to paying customers, rather than services that broadcast live events.

In this section we present an analysis of a media service offering *both* stored and live streams. This is innovative with respect to previous work since we consider the mutual interactions between these two kinds of workloads. Furthermore, since several companies, e.g., on-demand television providers, offer both types of streams, we believe that our analysis may be representative of the actual workload of a large number of Video-on-Demand (VoD) systems.

3.1. Log Characteristics

We analyzed a set of logs from a popular Italian on-demand television provider that offers VoD services using a dedicated WAN. Subscribers receive a set-top box (STB) and through a graphical interface select the desired contents. These include either popular live events (mostly soccer matches) or stored videos. Data was collected from two groups of servers that handle all stream requests and monitor the activity of several thousands of STBs in the metropolitan areas of six major Italian cities, including Milan, Rome and Turin.

We were able to obtain logs for 4 days where popular soccer matches were scheduled and for a typical day where only stored streams were available. The workload under exam consisted of two request classes: *stream* requests and *STB* requests.

Stream requests notify the VoD service that a user is requesting a video stream and can be classified in *unicast* and *multicast* requests. Unicast requests represent users asking for a stored stream. Multicast requests are instead generated by users registering to a multicast stream to watch a live event. It is thus reasonable for unicast requests to be quite regular during working days, with a general increase of the number of requests during the weekends. Within a single day, it is also reasonable to expect an increase of the load during evening hours. The behavior of multicast requests should instead reflect the characteristics of the broadcasted event.

Considering STB requests, these are generated by the STB at power on or power off to register or deregister from the VoD service. Hence, *registration* and *deregistration* requests may be considered as an indicator of the user activity cycle. Nevertheless, the inferred number of active streams

Type	#	Σ	$\mu/30\text{min}$	$\sigma/30\text{min}$	σ/μ
Uni Sun	48	35654	742.80	430.67	0.58
Uni Mon	48	25868	538.92	358.75	0.67
Multi Sun	48	3418	71.21	192.78	2.71
Uni w-ends	192	146511	763.07	440.02	0.58
Multi w-ends	96	10524	109.62	302.97	2.77

Table 2. Stream requests statistics: number of requests (#), cumulative sum (Σ), mean (μ), standard deviation (σ) and coefficient of variation σ/μ

Type	#	Σ	$\mu/1\text{min}$	$\sigma/1\text{min}$	σ/μ
STB Reg	240	2993	12.47	12.31	0.99
STB Dereg	240	2782	11.59	18.26	1.58
Users Rome	240	93398	389.1583	18.26	1.58
Users Milan	240	52555	218.9792	12.31	0.99
Users Turin	240	38369	159.8708	18.26	1.58

Table 3. STB requests statistics: number of requests (#), cumulative sum (Σ), mean (μ), standard deviation (σ) and coefficient of variation σ/μ

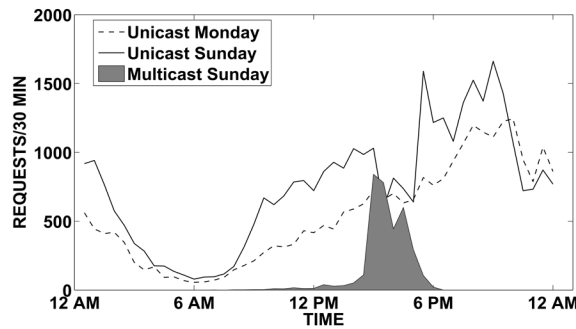
should be considered an estimate of the actual value, since a registered STB may still be inactive without requesting any download from the VoD service.

The characteristics of our logs are as follows. Most traces had a coarse sampling interval, reporting cumulative statistics every 30 minutes. However, for selected peak periods (i.e., those corresponding to soccer matches) we were able to use a sampling frequency of 1 minute. In both cases, we did not apply any filter to avoid further reducing the size of the sample. For some classes of requests, namely for registrations and deregistrations, we were also able to discriminate the geographic area of origin of the requests. Tables 2 and 3 show the statistics for the requests considered in the following section.

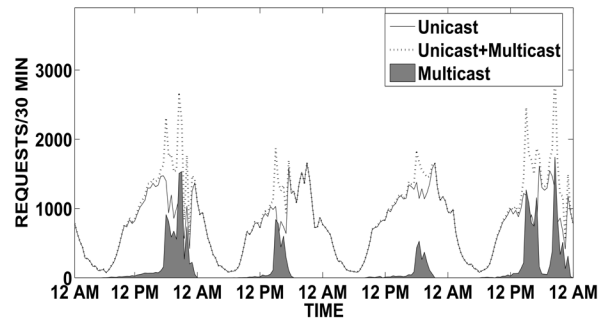
3.2. Media Workload Analysis

The behavior of stream and STB requests for the considered VoD system is shown in Figures 9(a)-(b). Figure 9(a) shows the behavior of unicast and multicast requests for a Sunday where both stored and live streams were available. The traffic shape can also be compared with the unicast traffic of the following Monday, where no live multicast streams were scheduled.

As expected, for both days the arrival process of unicast requests is similar during night hours, where the loads con-

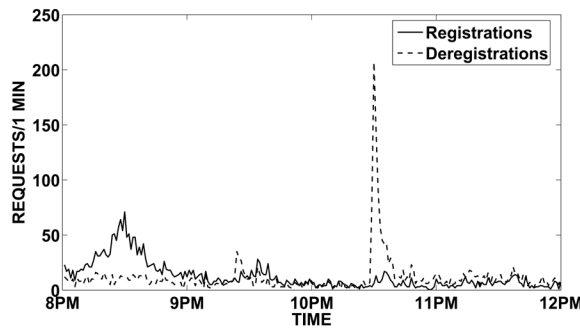


(a) Number of unicast and multicast requests in a working day and a weekend day (sampled every 30 minutes)

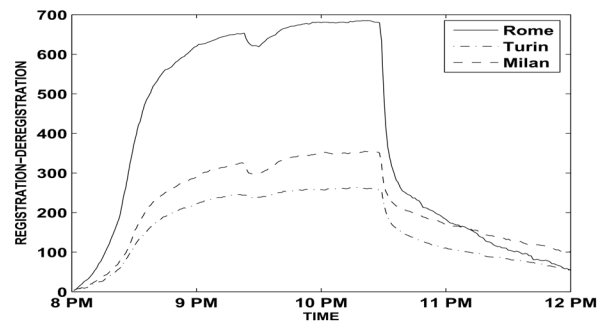


(b) Number of unicast and multicast requests on two consecutive weekends (sampled every 30 minutes)

Figure 9. Traffic behavior for a large on-demand television service offering stored and live streams



(a) STB requests during and after a live event (samples every minute)



(b) Number of active users during and after a live event (samples every minute)

Figure 10. Service registrations and deregistrations during and after a live event

stantly decrease up to a minimum of few hundreds requests around 6 AM. Despite the different slope of the Sunday and Monday graphs, the decrease with respect to the maximum in the period 12AM-6AM is very similar, with a 91% decrease on Sunday and 87% decrease on Monday.

A first significant difference is seen early in the morning hours (6AM-9.30AM), where Sunday workload shows a 740% relative growth, compared to the much smaller 387% growth on Monday. This may suggest that the large difference of unicast requests between Sunday and Monday is due to individuals who are unable to watch TV programs at morning during week days and with a regular sleep cycle during the weekend (e.g. adult workers or children). Therefore, such rapid growth must be taken into account for proper dimensioning of unicast servers. Note that the

gap between Sunday and Monday traffic is partially compensated in the following hours (10AM to 2.30PM), where the Monday unicast requests grow of 129%, compared to a 47% growth on Sunday.

The behavior of unicast requests in the period between 3PM and 5PM on Sunday is instead affected by the broadcasting of the soccer matches. In fact, as shown by the graph of multicast requests, the fluctuations of unicast requests seem to correspond to a similar increase or decrease in the number of multicast requests. In particular, if we linearly interpolate the missing unicast requests using the same trend of the period 10 AM-2.30 PM, we can conclude that approximately the 66% of the over 3000 multicast requests between 3PM and 5PM should be seen as unicast viewers that moved to the multicast stream of the live event.

Sunday traffic shape after the conclusion of the live event (5PM-12AM) has a similar behavior to that of Monday, except for a burst around 6PM. By interpolating the Sunday unicast traffic with the slope 10AM-9PM we would have expected approximately 1250 unicast requests at 5PM, instead of the observed 1600. This indicates that most of the multicast population switched to unicast streams after the conclusion of the live event. Indeed, this proves that a strong correlation between stored and live streams workload exists, and sudden peaks or falls in the arrivals cannot be fully understood without comparing the two workloads.

Evidence of the burstiness of media traffic is shown in Figure 9(b) where unicast and multicast traffics are plotted together with their sum over two consecutive Saturdays and Sundays. Among these, two had multiple live events at different times of the day. Once again we see on all four samples that unicast traffic decreases when the live event starts, and shows a sudden burst at the end of the multicast stream. Therefore, we conclude that the observation on the data of Figure 9(a) are consistent with the behavior shown by the system on the days of Figure 9(b).

Finally, Figures 10(a)-(b) give an idea of the behavior of the STB traffic during the period of a live event. Figure 10(a) illustrates the spikes of this type of workload, which is strongly correlated to the number of users that power on the STB to watch the soccer match. Figure 10 (b) illustrates the number of active users for the same interval of time considered in Figure 10 (a). Each curve represents a different geographic area of origin of the requests. The decrease of active users in the central part of the event around 9.30PM is due to the interval between the two halves of the soccer match. As observed above, the impact of live streams on the overall workload is not limited to the duration of the event, but also influences the number of active users until late night. Indeed several customers remain registered to watch unicast streams after the end of the soccer match. This further enforces our conjecture about the mutual correlation between stored and live streaming services.

4. Security Problems

The solution of security problems is nowadays recognized as a key enabler to the adoption of any new technology. There are many unresolved issues in securing digital media transmissions over public transmission networks [20].

A first range of problems is caused by the fact that nearly all video streaming protocols require a UDP port to be opened through any firewall on the transmission path. In addition, supporting transmission of a UDP stream through a device which performs Network Address Translation (NAT) is difficult, since UDP is a stateless protocol. While support for stateless protocols is being implemented with various workarounds into most modern firewalls, a number of

streaming protocols and servers can switch to a TCP fallback transmission method if UDP is not allowed. Some gateways for automatic translation have also been proposed [16]. However, as it is well known, TCP transmission is not nearly as efficient as UDP transmission. Thus, this fallback mechanism heavily impacts on end-to-end performance.

A second problem deals with the protection of the content of the media being streamed. The development of so-called “digital rights management” (DRM) techniques is usually considered of foremost interest for commercial content distribution. We can identify at least two broad goals for an effective DRM scheme:

1. conditional access to the content, meaning that only authorized parties (e.g. those who have subscribed to a service and paid for it) can access the streaming media
2. copy protection after the content has been delivered, meaning that the stream cannot be saved, copied or re-distributed except under well-specified restrictions

While granting conditional access to the media through an appropriate use of cryptographic or scrambling techniques is not difficult by itself [39], even in multicast environments [3], satisfying the second constraint is more difficult. As always, in security, we need to define an attack model in order to evaluate and weigh the effectiveness of the control mechanisms proposed. We can roughly distinguish three types of threats:

1. the casual aggressor, usually a user who has paid for access but tries to violate the rules he has agreed to, or who tries to make more or less legal copies of the content for a few friends; casual aggressors have limited resources, limited technical knowledge, and are usually not highly motivated;
2. the so-called “garage pirates”, creating copies for a small circle of friends or paying customers; more skilled than the casual aggressors, their resources are still limited;
3. large distributors of illegal materials, who have resources and a wide market for reselling illegal copies of copyrighted contents; skills and resources make it difficult to design adequate protection schemes.

As of today, most “copy prevention” schemes are easily bypassed, and work only against the first class of threats. The only viable solutions against the other classes is to enforce digital watermarking schemes that allow to trace back copied works to their original source, such as watermarking [5]. Watermarking a media means incorporating copyright and access control informations into it in such a way that they cannot be removed without substantially deteriorating the quality of the content itself. This technology can also be used, in compliant devices and programs, to enforce access

restrictions [6], or in general to trace back the illegal leakage of information to the original source. However, a number of attacks demonstrated the fragility of such schemes [25]. It is still to be seen whether digital watermarking will be accepted as an element of proof by courts or not. However, in one recent case (Palladium Music, Inc. vs. EatSleepMusic, Inc., 398 F.3d 1193, 10th Cir., 2005) the court found digital watermarks as admissible evidence for tracking copyrights.

Consequently, most standards for video and audio streaming and encoding today incorporate capabilities for supporting DRM technologies. For instance, the MPEG standards for video encoding and streaming specify a subsystem (IPMP, Intellectual Property Management & Protection) which allows integration of watermarking, encryption and authorization control into the media [18]. Watermarking can also be applied on streaming media by the use of proxies, without interfering with an existing video server infrastructure [13].

Wireless networks normally show an erratic behavior, which makes it difficult to design efficient streaming protocols. In addition, if encryption is added, transmission errors can be amplified to the point that streaming is made impossible. In [35] a secure wireless streaming scheme is proposed. The authors use scalable video coding and packetization to overcome the wide variability in the bandwidth offered by the physical transport layer, and a progressive encryption scheme with resynchronization on the key frames of an MPEG4-encoded stream, in order to avoid that a single lost packet can cause a chain effect on the ability to decrypt the rest of the stream.

In [9] the diffusion of DRM technologies is studied, and requirements and effectiveness of these schemes are discussed on the basis of an interview with various content providers.

We can try to identify the key performance metrics for DRM technologies:

robustness: the resistance of the proposed scheme to tampering, removal and circumvention: this property has to be evaluated end-to-end, because complex security systems may fail to provide the required properties even if their components and algorithms, by themselves, are secure and fit to their role

renewability: the possibility of seamlessly substituting a broken control technology without extensive redeployments in the infrastructure; a good example is the use of smart-cards in satellite receivers, which allows the substitution of keys and management algorithms and protocols without replacing the whole receiver

cost: any DRM technology must have a cost which is inferior to the reduction of losses the content provider risks without deployment of the technology.

While the first two metrics are purely technological, the key performance indicator is the third: to be economically convenient, a DRM scheme must demonstrate a ROI (Return On Investment), which may be difficult to prove, if we count all the direct and indirect costs of such technologies: research, development and deployment, but also the perceived reduction in user freedom which result from access control technologies of any kind.

In fact, DRM technologies adoption has been slowed down by a widespread discontent by users, who feel that their fair use rights are threatened by copy and use restrictions imposed by DRM technologies. Hardware and software vendors have also been reluctant to adopt these measures, fearing that their customers would be very disappointed. In fact, as discussed in [26], DRM technologies seem more oriented to restricting the rights of users (i.e. fair use rights) than protecting the real, legal rights of copyright holders [22]. In fact, in our broad categorization of threats, current DRM technologies work very well against common people, while piracy organizations are more or less unharmed by them. This fact becomes even more disturbing if we take into account the copyright legislation, and in particular the heavily criticized Digital Millennium Copyright Act (DMCA) [7, 12]. The DMCA, amongst other provisions, grants a special status to any “technology” that is being used to protect copyrighted work. It is illegal under the DMCA to discuss or develop techniques that can circumvent such technologies. This basically means that research on vulnerabilities of such technologies is de facto illegal [32]. A good example of how such laws can adversely affect research is the Felten case [30], in which researchers were threatened of prosecution for a work they were presenting at a scientific workshop. Another good example is the querelle between the Motion Picture Association of America and various individuals, concerning the distribution of DeCSS, a decryption algorithm for the CSS (Content Scrambling System), the encryption scheme used on DVDs [27, 29]. The issue here is evidently what Lessig calls “code as code” [19]: computer code (the one in DRM technologies) is supported as being legally binding, well beyond the real legal boundaries of copyright owned by the content producer.

As a final note, the real economic mechanisms of piracy are still under dispute [2]. As more and more evidence grows in support of the fact that online distribution of music through peer-to-peer networks does not really harm commercial distribution, all the research on Digital Rights Management techniques could become less important. In the meanwhile, these issues must be considered as one of the key enablers for the widespread adoption of video streaming techniques in commercial applications.

5. Conclusions

In this paper we have analyzed several issues of video streaming transmissions either live or stored. New methods able to obtain subjective indices of the QoS perceived by users from objective data measures have been introduced and validated. We have shown a correlation between unicast and multicast streams that should be accounted in the design and the capacity planning of a VoD service, as well as security issues that may impact on design decisions.

Acknowledgments

The authors wish to thank Filippo Scognamiglio for providing the data concerning the video streaming traffic.

References

- [1] R. Apteker, J. Fisher, V. S. Kisimov, and H. Neishlos. Distributed multimedia: user perception and dynamic QoS. In A. A. Rodriguez, M. Chen, and J. Maitan, editors, *Proceedings of IST/SPIE Symposium on Electronic Imaging: Science and Technology, Workshop on High-Speed Networking and Multimedia Computing*, pages 226–234, San Jose, February 6–10 1994.
- [2] S. Bhattacharjee, R. D. Gopal, K. Lertwachara, and J. R. Marsden. Economic of online music. In *ICEC '03: Proceedings of the 5th International Conference on Electronic Commerce*, pages 300–309, New York, NY, USA, 2003. ACM Press.
- [3] H. Chu, L. Qiao, K. Nahrstedt, H. Wang, and R. Jain. A secure multicast protocol with copyright protection. *SIGCOMM Comput. Commun. Rev.*, 32(2):42–60, 2002.
- [4] C. P. Costa, I. S. Cunha, A. Borges, C. V. Ramos, M. M. Rocha, J. M. Almeida, and B. Ribeiro-Neto. Analyzing client interactivity in streaming media. In *Proceedings of the 13th international conference on World Wide Web*, pages 534–543. ACM Press, 2004.
- [5] I. J. Cox and M. L. Miller. A review of watermarking and the importance of perceptual modeling. In *Proc. of Electronic Imaging 97*. NEC Research Institute, February 1997.
- [6] J. Dittmann, P. Wohlmacher, and R. Ackermann. *Communications and Multimedia Security Issues of The New Century*, chapter Conditional and User Specific Access to Services and Resources using Annotation Watermarks, pages 137–142. Kluwer Academic Publishers, 2001.
- [7] Digital Millennium Copyright Act. Published by The Berkman Center for Internet & Society, Harvard Law School, 2001.
- [8] A. Eskicioglu and P. S. Fisher. Image quality measures and their performance. In *IEEE Transactions on Communications*, volume 43, pages 2959–2965, Dec 1995.
- [9] M. Fetscherin and M. Schmid. Comparing the usage of digital rights management systems in the music, film, and print industry. In *ICEC '03: Proceedings of the 5th International Conference on Electronic Commerce*, pages 316–325, New York, NY, USA, 2003. ACM Press.
- [10] G. Ghinea and J. P. Thomas. QoS impact on user perception and understanding of multimedia video clips. In *MULTIMEDIA '98: Proceedings of the 6th ACM International Conference on Multimedia*, pages 49–54, New York, NY, USA, 1998. ACM Press.
- [11] G. Ghinea and J. P. Thomas. *Adaptive Evolutionary Information Systems*, chapter Adaptable QoS Management Communication Architectures For User Perception, pages 192–210. Idea Group Publishing, Hershey, PA, USA, 2003.
- [12] A. Grosso. Why the Digital Millennium Copyright Act is a failure of reason. *Commun. ACM*, 45(2):19–23, 2002.
- [13] F. Huang, H. M. Hosseini, H. C. Chua, and Y. L. Guan. Watermarking of streaming video for finger-printing applications. In *IEEE International Symposium on Circuits and Systems, ISCAS 2002*, volume 2, pages 452–455, May 2002.
- [14] ITU-R recommendation BT.500-11: “Methodology for the subjective assessment of the quality of television pictures”. The International Telecommunication Union, Geneva, Switzerland, 2002.
- [15] ITU-T recommendation P.910: “Subjective video quality assessment methods for multimedia applications”. The International Telecommunication Union, Geneva, Switzerland, 2002.
- [16] M. Johanson. A RTP to HTTP video gateway. In *WWW '01: Proceedings of the 10th International Conference on World Wide Web*, pages 499–503, New York, NY, USA, 2001. ACM Press.
- [17] S. A. Karunasekera and N. G. Kingsbury. A distortion measure for blocking artifacts in images based on human visual sensitivity. *IEEE Transactions on Image Processing*, 4(6):713–724, June 1995.
- [18] J. Lacy, N. Rump, and P. K. (eds.). MPEG-4 intellectual property management & protection (IPMP) overview & applications document. Working document of the ISO/IEC JTC1/SC29/WG11 workgroup, December 1998.
- [19] L. Lessig. *Code and Other Laws of Cyberspace*. Basic Books, Inc., New York, NY, USA, 2000.
- [20] E. T. Lin, G. W. Cook, E. J. Delp, and P. Salama. An overview of security issues in streaming video. In *ITCC '01: Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC '01)*, pages 345–348, Washington, D.C., USA, Apr 2001. IEEE Computer Society.
- [21] X. Lu, R. Morando, and M. E. Zarki. Understanding video quality and its use in encoding control. In *12th International Packet Video Workshop*, April 2002.
- [22] D. K. Mulligan, J. Han, and A. J. Burstein. How DRM-based content delivery systems disrupt expectations of “personal use”. In *DRM '03: Proceedings of the 2003 ACM workshop on Digital rights management*, pages 77–89, New York, NY, USA, 2003. ACM Press.
- [23] N. Nill. A visual model weighted cosine transform for image compression and quality assessment. *IEEE Transactions on Communications*, 33:551–557, Jun 1985.

- [24] J. Padhye and J. Kurose. An empirical study of client interactions with a continuous-media courseware server. Technical Report UM-CS-1997-056, University of Massachusetts, 1997.
- [25] F. A. P. Petitcolas, R. J. Anderson, and M. G. Kuhn. Attacks on copyright marking systems. In *Proceedings of the Second International Workshop on Information Hiding*, pages 218–238, London, UK, 1998. Springer-Verlag.
- [26] P. Samuelson. DRM and, or, vs. the law. *Commun. ACM*, 46(4):41–45, 2003.
- [27] P. Samuelson. Trade secrets vs. free speech. *Commun. ACM*, 46(6):19–23, 2003.
- [28] P. J. Shenoy, P. Goyal, and H. M. Vin. Issues in multimedia server design. *ACM Computing Surveys*, 27(4):636–639, 1995.
- [29] B. Simons. From the president: to DVD or not to DVD. *Commun. ACM*, 43(5):31–32, 2000.
- [30] B. Simons. Viewpoint: the ACM declaration in Felten v. RIAA. *Commun. ACM*, 44(10):23–26, 2001.
- [31] W. Tang, Y. Fu, L. Cherkasova, and A. Vahdat. Long-term streaming media server workload analysis and modeling. Technical Report HPL-2003-23, HP Labs, 2003.
- [32] The Association for Computing Machinery. Computer professionals concerned DMCA stifles academic freedom and speech. Press Release, 2001.
- [33] E. Veloso, V. Almeida, W. Meira, A. Bestavros, and S. Jin. A hierarchical characterization of a live streaming media workload. In *Proceedings of the ACM SIGCOMM Internet Measurement Workshop*, Nov 2002.
- [34] A. A. Webster, C. T. Jones, and M. H. Pinson. An objective video quality assessment system based on human perception. In A. A. W. et al., editor, *Proc. of SPIE Human Vision, Visual Processing, and Digital Display IV*, pages 15–26, Feb 1993.
- [35] S. J. Wee and J. G. Apostolopoulos. Secure scalable video streaming for wireless networks. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 4, pages 2049–2052, Salt Lake City, Utah, May 2001.
- [36] S. Winkler. Visual fidelity and perceived quality: Towards comprehensive metrics. In *Proc. SPIE Human. Vision and Electronic Imaging Conference*, volume 4299, pages 114–125, San Jose, CA, January 2001.
- [37] S. Winkler and R. Campos. Video quality evaluation for Internet streaming applications. In *Proc. SPIE Human. Vision and Electronic Imaging Conference*, volume 5007, pages 104–115, Santa Clara, CA, January 2003.
- [38] S. Winkler and F. Dufaux. Video quality evaluation for mobile applications. In *Proc. SPIE/IS&T Visual Communication and Image Processing*, volume 5150, pages 593–603, Lugano, Switzerland, July 2003.
- [39] W. Zeng and S. Lei. Efficient frequency domain video scrambling for content access control. In *MULTIMEDIA '99: Proceedings of the seventh ACM International Conference on Multimedia (Part 1)*, pages 285–294, New York, NY, USA, 1999. ACM Press.