Performance evaluation study for QoS-aware triple play services over entry-level xDSL connections

Chryssa A. Papagianni, Nikolaos D. Tselikas *, Evangelos A. Kosmatos, Stauros Papapanagiotou, Iakovos S. Venieris

School of Electrical and Computer Engineering, National Technical University of Athens, 9 Heroon Polytechniou Street, Zografos, Athens, Greece

A R T I C L E   I N F O

Article history:
Received 27 July 2007
Received in revised form 4 March 2008
Accepted 24 March 2008

Keywords:
Triple play
H.264
Packet scheduling
PQ
WFQ
WFQ-LL
Packetization schemes
Performance evaluation

A B S T R A C T

Nowadays IT and telecom worlds and specifically all network, service, and content providers show special interest in multi-service delivery. The concurrent provision of video or TV content, voice, and fast Internet via the same network and user equipment, known as “triple play service”, constitutes one of the major concerns in modern telecommunications due to the latest developments in the areas of video and audio encoding which strain the capacity of content delivery systems. The combination of the above technologies has raised significant quality of service related issues. Problems such as full exploitation of available bandwidth, providing adequate quality of service to subscribers and meeting the requirements of all three supported services (video, voice, and data), must be addressed. Based on this framework, the paper examines the performance of the most common packet scheduling techniques (PQ, WFQ, WFQ-LL) used in IP triple play architecture, considering also two different packetization schemes applied to single-layer video. The performance of these scheduling algorithms is assessed, in order to propose the most appropriate solution, supporting the triple play bundle. The evaluation of schedulers is based on simulations designed taking into account existing triple play networks, while trace files are used for the simulation of video flows.

© 2008 Elsevier Ltd. All rights reserved.

* Corresponding author. Tel.: +30 210 772 2318; fax: +30 210 772 1092.
E-mail address: ntsel@telecom.ntua.gr (N.D. Tselikas).

1084-8045/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jnca.2008.03.006

Please cite this article as: Papagianni CA, et al. Performance.... J Network Comput Appl (2008),
doi:10.1016/j.jnca.2008.03.006
1. Introduction

Over the last decade numerous innovative services have emerged, utilizing Internet as the delivery mechanism, whereas traditional services such as telephony and television are migrating onto IP, enabling it to become the wide area network communication protocol of choice. This happens because, following the continuous long lasting success of Internet and broadband access worldwide, governments, telecommunication companies, network operators, content, and service providers are driven by a new challenge. The new challenge is arising by the broadband evolution and the need to optimize cost, which leads to the migration of the existing—and often separated—networks’ services (e.g. fixed telephone, television, video, and Internet access) into one competing multiservice, the so-called triple play service (Altgeld and Zeeman, 2005). Via the triple play service, the user can have a relish for all the aforementioned services; IPTV (video on demand or commercial—grade TV), voice over IP (VoIP) telephony and high-speed Internet access, all at the same time.

There are multiple approaches to deliver triple play service to the end user (Xiao et al., 2007). The usage of fibers is one alternative, either by assimilating an active last-mile architecture—the so-called active Ethernet—in which each user is provided with a dedicated fiber connection to a switch at a neighborhood aggregation point, or by inheriting a passive last-mile architecture based on passive optical networks. The other alternative is to exploit the existing access network, based on copper loops, that is widely available to end users in most developed countries. Telecommunication companies worldwide invested a lot of money on digital subscriber line (DSL).

Through the investments network operators have managed to cut down on cost per Mbyte, in order to provide high rate xDSL services. However, in many countries a 4 Mbps xDSL connection can still be characterized as the “average connection speed”. Thus, the utilization of the MPEG-2 (Motion Picture Engineer Group 2) video deployed globally for video compression applications, which is “bandwidth voracious”, still remains a serious problem. Nevertheless video compression technology is advancing with the introduction of techniques like MPEG-4 AVC and VC-1 that are far more efficient concerning bandwidth requirements. The H.264 video codec standard, also known as MPEG-4 advanced video coding (AVC), is now the accepted standard for communications, broadcast, and streaming applications. MPEG-4 AVC cuts the bandwidth requirements for digital video delivery in half and along with its extension—known as scalable video coding (SVC) allows the parallel provision of video, VoIP, and broadband internet access at the same time even in low rate DSL connections.

In Xiao et al. (2007), quality of service (QoS) guarantee and traffic management is acknowledged as a technical challenge for the successful deployment of triple play services. The goal of traffic management is to support QoS requirements for the diverse set of triple play services, including policing, scheduling, flow control, traffic differentiation, admission control, etc. For that purpose commonly used packet scheduling techniques (PQ, WFQ, WFQ-LL) are evaluated for providing triple-play services over an IP-QoS aware architecture, considering also two different packetization (PKZ) schemes applied to single-layer video. Scheduling ensures the timely and error free delivery of the triple play service bundle as well as efficient utilization of bandwidth.

Section 2 features the network architecture, specifically designed for triple play service provisioning. The architecture has been simulated using the Opnet simulation tool. In Section 3 the relative rising concerns of the triple play service specifics are presented, such as video and audio codecs for triple play service provision, service separation and packet forwarding issues, QoS alternatives and differential treatment scheduling. In Sections 4 and 5 the simulation study and the evaluation results are cited, respectively, regarding various scheduling strategies applied to two different PKZ schemes. Conclusions and future steps are summarized in Section 6.

---

1 The average speed of residential broadband connections (downlink) in the UK at the end of 2007 was 4 Mbps [http://www.prompt-communications.com/impromptu_weekly/30_Nov_07.html](http://www.prompt-communications.com/impromptu_weekly/30_Nov_07.html).
2. Architecture overview

In the telecommunications industry triple play encompasses the provisioning of three services: high-speed Internet, television (unicast/multicast video), and telephony through the use of a single broadband connection. Key factors in the triple play service delivery are the technological advances in relevant equipment. VoIP Gateways and soft switches enable circuit-to-packet migration for voice services as well as VoIP service switching, respectively. Digital head-ends deliver IP video while broadband loop carrier (BLC) systems allow the convergence of voice, data, and video in a single access network infrastructure.

In order to evaluate triple play service over DSL for various queuing algorithms, part of a triple play network architecture adopted by major industry players such as Cisco (Cisco Systems, 2005; TR-101, 2006) was simulated (Fig. 1).

Fig. 1. Triple Play Architecture over xDSL.
The video delivery enabling segments of the architecture are structured hierarchically. The super head-end is typically residing in the core network, where live feeds for the broadcast video service are located and real time and offline encoding of raw material takes place for multicast video and for VoD services. Usually it is unique on a nation level for IPTV deployments. The video switching office (possibly located at the Central Office) interconnects the distribution-aggregation network though the use of aggregation routers (ARs) that collect subscriber traffic from DSL access multiplexers (DSLAMs). In addition local/popular content may be cached for on demand use at the VSO. The boundary of the distribution network is the core edge router (ER). ER is a convergence device at the edge of the network, which provides not only transport for multimedia flows between the IP/MPLS core network and the distribution network components but also routes data and voice traffic. According to the depicted topology, ER is responsible for the appropriate DiffServ tagging of the flows (voice, data, or multimedia) that are distributed to the users. It is assumed that ER and AR have multicast routing capability and can process IGMP requests. If either of them is configured as a rendezvous point where downstream subscribers obtain multicast traffic, their performance becomes a significant part of the network join/leave time equation. The broadband services router (BSR) is responsible for network access control functionality (user authentication, access control, etc.).

The home access gateway (HAG) is located at the edge of the access domain and one of its functionalities is to mark traffic coming into the domain with the appropriate QoS tag. The DSL Layer 1 connection from the HAG is terminated at the DSLAM. The DSLAM switches the traffic to the Aggregation Router. HAG is responsible for identifying each flow and properly forwarding it to the appropriate home device (e.g. video flow to the IP set-top box, voice flow to the Foreign Exchange Station and then to the home telephone wire infrastructure, data flow to a personal computer).

3. Triple play architecture concerns

3.1. Video and audio codecs for triple play service

One of the challenges that VoIP carriers deal with early in their network planning is to choose the most appropriate voice coding standard in order to provide good voice quality and adequate network efficiency. From uncompressed G.711 at 64 kbps to G.726 at 16 kbps, G.729 at 8 kbps and the highly compressed G.723.1 at 5.3 kbps, the VoIP service providers can choose the level of voice compression that will be applied to their customers. In this particular study the G.711 codec is employed. G.711 is the international standard for encoding telephone audio on a 64 kbps channel (ITU-T, 1989). It is a pulse code modulation (PCM) scheme, operating at 8 kHz sample rate, in order to encode frequencies between 0 and 4 kHz, with 8 bits per sample, used to convert an audio signal from its native analog format into the digital domain. Two encoding laws are defined in the standard, both designed as logarithmic algorithms and these are commonly referred to as the A-law and the μ-law. According to both variants, lower signal values are encoded using more bits than higher signal values which require fewer bits. G.711 μ-law compresses frames of 14-bit linear PCM samples into frames of 8-bit logarithmic PCM code words. G.711 A-law compresses 13-bit linear PCM samples into 8-bit logarithmic PCM code words and is the international scheme for encoding telephone audio on standard PCM channel (Bellamy, 2000). Initially G.711 was preferred from VoIP carriers due to the fact that it provides PKZ flexibility, payload efficiency; speech quality and media gateways cost reduction. However, as it shows poor network efficiency it is surpassed by lower bit rate coding algorithms (Percy, 2007).

The choice of the appropriate encoding schemes is vital for the bandwidth save in triple play architecture. The video codec H.264 standard, which is being adopted by all major video service operators, is utilized in the performance evaluation of triple play service over xDSL access network. (ITU-T Rec. H.264, 2003). It was jointly developed by ITU Video Coding Experts Group (VCEG) and ISO Moving Picture Experts Group (MPEG). H.264 is used in fixed and wireless network environments.
It is divided into two distinct layers, the video coding layer (VCL) and the network abstraction layer (NAL). Roughly, the VCL generates an efficient representation of the video data as a sequence of bits. On the other hand, NAL deals with the appropriate packaging of the coded data, based on the underlying network characteristics and provides header information in a manner appropriate for conveyance by particular transport layers (such as Real-Time Transport Protocol). The NAL encoder encapsulates the slice output of the VCL encoder into NAL units, which are suitable for transmission over packet networks.

H.264 has proven to be more resilient to error prone networks through the use of flexible macroblock ordering, slice interleaving and data partitioning (Richardson, 2003). Moreover, in order to prevent error propagation through inter prediction, H.264 provides mechanisms such as redundant slice and spare macroblocks that carry duplicate information to be used when there is an error in transmission. H.264 is relatively simple in its implementation. In addition it attains enhanced compression performance therefore is a “network-friendly” standard. It is capable of providing good video quality at substantially lower bit rates than other standards. Compared to MPEG-2 video, it cuts down transmission bit rate by half, while the coding gain over H.263 and H.263+ is in the range of 24% up to 47% (Kamaci and Altunbasak, 2003; Schäfer et al., 2003).

In addition to high compression gain, encoder technology supports PKZ schemes in order to relax the high fluctuations of the traffic produced at the output, contributing in smoothing the burstiness of self-similar sources. According to the RTP/UDP/IP PKZ guidelines for H.264 AVC, there are three different basic payload structures. These are single NAL unit packet, aggregation packet and fragmentation unit (Wenger et al., 2005). The single NAL unit packet contains only a single NAL unit in the payload. The NAL header type field will be equal to the original NAL unit type. The aggregation packet is used to aggregate multiple NAL units into a single RTP payload and the fragmentation unit is used to fragment a single NAL unit over multiple RTP packets.

On top of H.264 AVC standard, MPEG and VCEG agreed to jointly finalize the SVC project as an amendment of their H.264/MPEG-4 AVC standard (Schwarz et al., 2007; ITU-T Rec. H.264). The scalable extension of H.264/MPEG-4 AVC generates an H.264/MPEG-4 AVC compliant base layer and one or several enhancement layers. Therefore, it implements a layered coding scheme with switchable inter-layer prediction mechanisms, supporting quality granularity by using progressive refinement layers and extending the usage of the NAL unit concept of H.264/MPEG-4 AVC.

3.2. Quality of service

QoS refers to the nature of the different types of traffic delivery provided, as described by parameters such as achieved bandwidth, packet delay, and packet loss rates. The simulated QoS architecture is based on the IETF differentiated services architecture described in RFC 2475, which assumes that all nodes where congestion may occur are capable of implementing QoS functionality at the IP layer. Therefore, QoS handling is defined on a per-hop behavior based on IP differentiated services code point. In the simulated architecture, the DiffServ domain extends to the aggregation/distribution network including access gateways, IP–DSLAMs and ARs. ER acts as a domain boundary capable of implementing the basic DiffServ functionality. That is to classify packets by means of a DiffServ code point (DSCP) and implement corresponding per-hop behavior. Additionally on the upstream direction, the HAG is the administrative boundary of the domain. Currently, QoS architecture is usually extended to the data link layer by mapping DiffServ PHBs to Layer 2, at the edges of the Layer 2 multiservice network (e.g. ATM Class of Service, Ethernet 802.1p value) (DSL Forum TR-059, 2003).

3.3. Service separation and packet forwarding

In a converged triple play architecture, where each service has its own QoS delivery requirements, the isolation of each service is crucial for the transport network to ensure fairness and operating efficiency (Cisco Systems, 2006c). Isolation can be translated as packet forwarding/routing per service type and service separation.
In most DSL transport architectures (Ethernet/ATM), where the aggregation/access networks include nodes that are not capable of supporting IP-layer QoS, packet forwarding with Class of Service at Layer 2 is implemented. However, in many cases especially for video services Layer 3 forwarding is also utilized among the set-top boxes and the video infrastructure modules. As far as service separation is concerned a DiffServ QoS architecture that determines per-hop behavior based on packet priorities is considered as a flexible, scalable distributed design that lacks however the capability to provide intra service administrative separation. On the other hand, through service topology separation each service uses separate physical or virtual interfaces and forwarding instances allowing for advanced service administration or bandwidth control (Cisco Systems, 2006c). Many technologies exist at both Layers 2 and 3 to achieve topology-based separation, including VLANs, MPLS traffic engineering, etc.

3.4. Differential scheduling

Within the DiffServ domain, nodes following the same policy for QoS must be capable of implementing the specified per-hop behavior accordingly. Typically this is achieved by means of complementary mechanisms such as queue management and scheduling. Queue management algorithms handle the length of packet queues by dropping packets when necessary or appropriate. Scheduling algorithms determine the priority among packets ready for transmission residing in different queues and are used primarily to manage the allocation of bandwidth among flows. In a packet-switched network, typically, packets are queued and scheduled for transmission on a first-in-first-out (FIFO) basis that applies the same priority to all packets, independently of their performance objectives. Several other scheduling mechanisms have been proposed in order to determine which queue will be given the opportunity to transmit, such as the fixed priority queuing (PQ), fair queuing, weighted fair queuing (WFQ), and weighted fair queuing with low latency (WFQ-LL) queue, worst-case fair weighted fair queuing, round robin, weighted round robin, earliest deadline first algorithm, etc. In the particular paper we present a comparison among Cisco’s premier queuing algorithms such as PQ, WFQ, and WFQ-LL queue for varying video buffers and different PKZ techniques, over a triple play DiffServ architecture (Cisco Systems, 2006b).

4. Triple play performance evaluation

4.1. Triple play service

For the simulation of triple play over xDSL, a specific bundle of services is provided uniformly to the simulated users including different types of traffics. The simulated traffic flows include bulk data applications, such as FTP, e-mail, and web browsing involving large data transfers and real-time applications, such as VoIP and unicast video on demand. Concerning the simulation of VoIP, G.711 encoded data streams were simulated, as it provides the best voice quality possible, with an average transmission rate of 64 kbps, packet size of 32 transmission, rate of 250 frames per seconds and no voice activity detection (Percy, 2007). Simulation of multimedia traffic was performed using H.264 encoded trace files. Trace files provide a parsing/decoding of the bitstream of each clip, into the bitstream fields as listed in the H.264 standard. Specifically trace files from “Starship Troopers” film were used as input in the Opnet simulation engine (Video trace research group; Opnet technologies). Trace file’s statistics are shown in Table 1.

Because in the MPEG traces each frame is a slice, the size of each slice is usually larger than the maximum transmission unit (MTU) of typical networks, where MTU is the maximum packet size a particular network can accept without imposing any fragmentation. Thus, larger frames are fragmented in order to conform to the MTU requirement when these are received by the router.

Two different PKZ schemes have been used in the simulations. The default, defined as complete frame (CF), includes packets of variable length as shown in Table 1. In the second scheme called constant PKZ, every video frame of the H.264 flow is segmented to packets of constant length.

Please cite this article as: Papagianni CA, et al. Performance.... J Network Comput Appl (2008), doi:10.1016/j.jnca.2008.03.006
(Chen, 2002), i.e. 552 bytes, which is a common maximum transmission unit (MTU) size in real networks (Thompson et al., 1997). Following the PKZ at the transport layer, the transmission of packets belonging to the same frame is uniformly distributed in the first half of the frame duration (each frame’s duration is 40 ms since the frame rate is 25 frames per second). Spreading packet transmission assists to avoid sudden congestion in the router buffer, if the distribution duration is too long, delay is augmented. On the other hand, short distribution duration imposes heavy traffic. Since there is a trade-off between delay and congestion the selection of the packet size should result in a compromised solution. PKZ methods have significant influence on the simulation results, as they affect the flow of traffic and its statistical attributes.

4.2. Simulation architecture

The simulation architecture was built based on the architecture described in Section 2 using Opnet simulation engine (Opnet technologies). Since the purpose of the simulation was to denote the performance of various packet-scheduling techniques using different PKZ schemes for H.264 over various traffic loads, only essential components of the triple play infrastructure were employed. In order to simulate the triple play service, a video server was utilized for the delivery of video flows as well as a VoIP and a data traffic source. In the distribution network reside four ERs and an AR. In the access network an IP DSL access multiplexer is utilized. Home network equipment consists of a typical client-side network termination device—an xDSL modem and an Ethernet switch. In order to support triple play, users are serviced over a 4096/512 kbps xDSL connection with a contention ratio of 2:1.

4.3. Simulation scenarios

A set of simulation scenarios were executed on the aforementioned triple play architecture with varying traffic load conditions. The motivation for conducting the simulation study was to investigate the way in which different traffic loads, queuing algorithms and PKZ schemes affect the most important QoS parameters (i.e. packet loss, delay, and jitter) for each service. We assume that congestion is more likely to happen across the link (Cisco Systems, 2006a) which interconnects the AR to the DSLAM. In the mean time the DSL link connecting each home network to the DSLAM contributes no losses according to the simulated triple service profile used. Specifically, the AR-DSLAM link utilization ranged from 55% to 95%. As the main objective is to determine which scheduling algorithm and PKZ technique meets effectively the requirements of triple play service under heavy traffic conditions, only the results of the worst-case simulation scenario (95% link utilization) are presented.

In the simulated triple play service deployment, VoIP retains the highest EF (expedited forwarding—RFC 3246) priority; AF42 (assured forwarding—RFC 3246) offers medium packet drop precedence for video on demand service and is second on the priority list and bulk data transfer is default PHB (best-effort) traffic (Cisco Systems, 2006a).

The packet queuing and scheduling techniques that were explored are PQ, WFQ, and WFQ-LL for four different levels of video buffer size on the AR (50, 80, 110, and 220 packets). The buffer size for the EF class is 10 packets while for the BE is 30 packets. The problem of optimal PKZ for H.264 was addressed by examining the two aforementioned PKZ schemes (CF and PKZ).

<table>
<thead>
<tr>
<th>Trace File Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frame per second</td>
</tr>
<tr>
<td>Minimum frame size (bytes)</td>
</tr>
<tr>
<td>Maximum frame size (bytes)</td>
</tr>
<tr>
<td>Average frame size (bytes)</td>
</tr>
<tr>
<td>Average transmission rate (Mbps)</td>
</tr>
<tr>
<td>Burstiness (Max TR/Min TR)</td>
</tr>
</tbody>
</table>
5. Simulation results

The simulation results for the video service illustrate the fact that there is a trade-off between delay and packet loss, for each PKZ scheme, scheduling scheme and buffer size. As the buffer size increases, the packets’ presence in the queue is extended, decreasing the number of discarded packets.

End-to-end latency and jitter are very important for a VoIP service. A typical end-to-end jitter requirement for a carrier-class VoIP service is 60 ms (Cisco Systems, 2006a). The ITU G.114 specification recommends less than 150 ms one-way end-to-end delay for high-quality real-time traffic such as voice. Moreover concerning packet loss, even a 1% loss can significantly degrade the...
user experience with the ITU-T G.711 voice coder (Intel Corporation, 2003). From the corresponding figures we notice that the aforementioned targets are met for the various scheduling/PKZ schemes tested (see Fig. 2).

Regarding PQ, it is derived from Fig. 3 that, especially in constant PKZ (PQ_PKZ) scenario, best effort flows are excluded from using the available bandwidth, as the data service has the lowest priority in the hierarchy of scheduling. The problem becomes more intense for high values of video buffer sizes (110,220).

As far as the weight fair queue scheme is concerned, the assigned weights grant 90.5% of the available bandwidth to the video flows, 7% to VoIP service and 2.5% to TCP traffic. We assume that voice traffic flow will encounter no losses, as the weight assigned to the voice traffic flows retains sufficient bandwidth for the overall voice service requirements. According to the results for constant PKZ, when small video buffer size is used, there is only a slight deterioration of the video service, due to a small increase in packet loss ratio, while there is a major improvement of the best effort data services in comparison to PQ. Therefore, WFQ does not exclude best effort traffic when the network is highly utilized.

In WFQ scenarios, for both PKZ schemes, the delay of the VoIP service is slightly worse than in the PQ scheme, while as the buffer size increases the packet loss percentage drops. Fine tuning of the WFQ scheme is possible, provided that we have a priori knowledge or we can statistically predict the actual bandwidth requirements for each service. However, the need for prior knowledge of bandwidth requirements for the provided services makes WFQ impractical as it is very difficult to adjust a proper weight for each service.

A solution can be found with the use of the WFQ-LL scheduling. Adequate bandwidth is allocated permanently to the voice service, providing high QoS for the subscribers through constant low packet latency. It also guarantees that the data traffic flows are never excluded from the bandwidth usage. The inevitable cost is a slight augmentation in the packet loss for the video and voice services (Fig. 2). This can be easily compensated with an increase of the video buffer size as well as with modern video error resilience schemes applied in the set-top box.

Regarding the different PKZ schemes, the constant PKZ scheme achieves far better performance in regard to average delay and jitter for both VoD and VoIP services (Fig. 2). The packet loss results show that it avoids high packet losses of video service in small video buffer sizes, while values seem to converge in medium and high values. Considering data applications a considerable packet loss occurs only with PQ and constant PKZ.

![Fig. 3. Packet Loss for Data Flows.](image-url)
The comparison of all results leads to the selection of the combination of constant PKZ scheme and WFQ-LL as the most efficient solution. The WFQ-LL-PKZ scheme presents far better performance on the basis of maintaining low average delay and inter-arrival jitter values and low packet losses even at small buffer sizes for media streams. In addition concerning best effort traffic handling, it exhibited average packet loss rate close to zero.

6. Conclusion—future work

The paper presents the performance evaluation of the three most common scheduling schemes (PQ, WFQ, WFQ-LL) in the framework of a triple play architecture. The design of the simulation topology was based on commercial triple play architectures, while trace files were used in order to simulate the troublesome video applications. In addition a special packetization scheme was used called constant packetization and its efficiency was demonstrated. Assuming that congestion is more likely to happen across the link that interconnects the aggregation router to the DSLAM, the behavior of the system was examined using different link utilization and configuration of the routers (buffer sizes and scheduling algorithms) while maintaining a traffic mix of the three services. During the simulation and evaluation phase we kept in mind the inviolable law of a triple play network, which is to meet the needs of all three supported services (video, voice, and data). The comparison of all results led to the selection of constant packetization scheme for MPEG-4 and WFQ-LL as the most efficient solution for triple play delivery.

Our future work will utilize the presented DiffServ architecture for additional comparison with custom-made scheduling strategies. In addition the application of per-hop QoS techniques is investigated, which may lead to enhanced QoS granularity of the DiffServ architecture and smoother delivery of voice and video flows in the framework of a triple play service delivery architecture.

References


DSL Forum TR-059. DSL evolution—architecture requirements for support of QoS-enabled IP services; 2003.


Intel Corporation. Overcoming Barriers to High-Quality Voice over IP Deployments, 2003


