QoE-Driven Network Management For Real-Time Over-the-top Multimedia Services

Janne Seppänen and Martin Varela

VTT Technical Research Centre of Finland PL 1100 Oulu 90571, Finland Email: janne.seppanen@vtt.fi, martin.varela@vtt.fi

Abstract—This paper introduces a network access point (AP) control solution in the context of over-the-top (OTT) multimedia services. The solution is designed to provide network-level management mechanisms for packet traffic while using Quality of Experience (QoE) as a performance indicator. The results showed that with customer subscription scheme, traffic differentiation and QoE-driven management it is possible to both improve the perceived quality of multimedia traffic and increase the average revenue per user.

I. INTRODUCTION

During the last ten years the amount of multimedia on the Internet has increased significantly. While customers of Internet service providers used to be satisfied with a fraction of the bandwidth they purchased, nowadays consumer traffic, especially multimedia, tends to consume a significant portion of Internet service providers' (ISP) network capacity; According to estimations of [1], mere video streaming will account for over 55 percent of all Internet traffic in 2016. Investing for more bandwidth poses large capital and operational expenditures, therefore network operators must come up with new ways to both deal with bandwidth-demanding multimedia applications and keep the average revenue per user high.

The concept of Quality of Experience has come to supersede QoS when we are interested in the perceived quality, rather than in the network parameters. Generally speaking, QoE is a broad term, which can be applied to numerous scenarios. A fairly comprehensive definition is given in [2]: "Quality of Experience (QoE) is the degree of delight or annoyance of the user of an application or service. It results from the fulfillment of his or her expectations with respect to the utility and / or enjoyment of the application or service in the light of the user's personality and current state". It can be noted from this definition that QoE is not a purely technical term, but it spans over several fields of science.

The author of [3] further describes QoE from two aspects: A person (who in this context has a subscription to ISP) is both a customer and a user of the service. There is a certain way this person perceives the service as a customer, of what the author calls "quality of customer experience" (QoCE). This experience depends on how the person is treated from the customer perspective. If the person calls a customer service line and needs to hold for twenty minutes, the person feels mistreated, not as a user but as a customer. However, if the

service the person received is poor, then this deteriorates the quality of user experience (QoUE).

We should note that in the context of computer networks, the user is not using the network itself, but the application which operates on the network infrastructure. Parameters commonly linked to QoS, such as bandwidth, delay, packet loss etc., do not sufficiently describe the quality perceived by the user, as the quality greatly depends on what happens in the application itself - not only in the network. For example, a packet loss of 3 percent can be almost unnoticeable when browsing the web, but when streaming an RTP video under the same condition, the quality of the multimedia may be intolerable. The same way, an email can be a minute late without aggravating the user, but half-a-second delay in a VoIP call can render the quality unacceptable.

From these observations, we can conclude that QoE is a logical link between the user and the application, and QoS between the application and the network. But since overthe-top services, by definition, are not controlled by network operators, the application traversing their network is beyond their reach; The ISPs lack not only control over the users' equipment, but also control over the servers on the Internet. Then again, QoE is conceptually above the application layer. Remembering that the user is also a customer, the problem becomes the question "what can be done in the network to improve the QoE of a user to meet the expectations he has as a customer?"

This paper tackles this question with an autonomous network access point management solution. The following subsection gives an overview of the state-of-the-art in this research topic. Section II introduces the methods we are using for identifying traffic flows, estimating QoE and performing management tasks. Section III describes the testbed used for testing the designed solution, and section IV presents and discusses the results acquired from the tests. Finally, conclusions and future prospects are presented in section V.

A. Related works

While QoE-driven network AP management for OTT services is quite a new concept, there are already many papers published for IPTV (e.g. [4]–[6]). Nevertheless, there are though a few papers describing various management methods for OTT services; A scenario similar to that in this paper is



Fig. 1. The architecture of the access point.

presented in [7], where TCP-based video streaming is managed with a client-side software and a home gateway, whereas we consider RTP-based video. In [8], the authors demonstrate the efficiency of QoE-driven vertical handover, video codec changeover and bit rate reduction as management methods. A pure QoE-driven handover management solution is presented in [9], while admission control is used to maintain QoE in 802.11 wireless networks in the study conducted in [10].

The field of QoS-based network management is well established, and most of these methods are applicable also in this work. We are especially interested in admission control, scheduling, bandwidth management and congestion control in access points. A good example of a work on an AP management is presented in [11], where, like in this work, the traffic is controlled not only by the traffic type, but also by the user subscription type. The solution consists of a centralized hotspot controller, which balances the load of several IEEE 802.11e wireless access points. A hybrid solution is presented in [12], where two QoS mechanisms are implemented in two parts of the network: fair congestion control and DiffServ in the core of the network, and IntServ-governed admission control at the edge of the network.

II. OUR APPROACH

As the foundation of the management system, we use three aspects of control: customer, user and application. First, we divide customers into two subscription groups: premium users and normal users. Normal users are customers who have purchased a best-effort Internet connection. Premium users, on the other hand, are provided with a high-end connection type, which gives better service quality (in addition to, say, a higher bandwidth). The idea of this kind of differentiation is to make the more expensive subscription type more appealing to "heavy users" and to users who need better reliability than best-effort. Premium customers must get better service quality, but the solution should not deteriorate the experience of normal customers.

Regardless of the customer differentiation scheme, all users ought to enjoy as good quality as possible. Moreover, not all applications can be considered equal in terms of QoE. Users assess the QoE through the application, hence applications traversing the network need to be identified and treated according to their real-time restrictions.

Our proposed solution consists of three components, which can be seen in Fig. 1. First, all incoming traffic from the ISP's network enters the traffic classification and traffic management modules. The classification module identifies the application behind each relevant traffic flow and hands the information to the management module. Just before entering the last link, the traffic is mirrored to the QoE assessment module, which, in real-time, estimates the perceived quality of each relevant flow. By knowing the application of the flow, the client subscription type (which in this setup is identified by the IP address) and the estimated perceived quality, the traffic management module can make appropriate adjustments to increase the perceptual quality (and hence QoE) as needed.

The algorithms of the three modules are described further in the following subsections.

A. Traffic classification

Since there is no simple, unambiguous way of knowing the underlying application of a traffic flow beforehand, we need an algorithm able to classify flows in real-time based on network-level information. In this work, a two-phased traffic classification tool is utilized, which is able to classify network traffic based on the statistical features of flows. This solution is a machine-learning algorithm, which can be trained to identify applications by using a training dataset (i.e. a collection of traffic traces). A detailed description of the tool can be found in [13].

The algorithm operates, as the name indicates, in two phases: in the first phase, the payload sizes of first four packets are observed. In the second phase, the following statistics are collected during the first 1000 packets: average payload size, average downlink inter-arrival time, number of push packets to downlink and the number of data packets to downlink. The collected statistics are then compared to the ones gathered in the training stage in order to find an application where these statistics would match the best.

Once the application has been identified, the flow is assigned with the corresponding tag for the management system. In this work, we used three application classes: bulk, streaming and interactive. The first class contains plain file bulk downloads and it is the lowest class, the second class is reserved for one-way streaming multimedia applications, and the last class contains interactive applications, such as online multiplayer games, VoIP calls and videoconferencing. This class division exists for both normal users and premium users in the following manner (in a descending order in priority):

- 1) Premium interactive media
- 2) Normal interactive media
- 3) Premium streaming media
- 4) Normal streaming media
- 5) Premium bulk traffic
- 6) Normal bulk traffic / Default class

The lowest class is the default class where all traffic is initially directed in. A flow is promoted to a higher class only when a degradation in perceived quality is detected.

B. Quality Assessment

Since QoE is a concept, we need a metric for expressing it quantitatively. Typically, QoE is expressed as a mean opinion score (MOS), which indicates the average opinion score of a group of users. For assessing video quality, a range of 1-5 is often used, where the numbers represent a verbal expression of the perceived quality [14]: In absolute category rating (ACR), 5 stands for "excellent", while 1 stands for "bad" quality. In degradation category rating (DCR), 5 means that impairments in the video signal are "imperceptible", and 1 means that the impairments are "very annoying". In this work we use a DCR scale.

For mapping QoS parameters to a single MOS, we use an algorithm called pseudo-subjective quality assessment (PSQA). PSQA is usually implemented with a feed-forward random neural network (RNN) [15]. The RNN is trained to estimate, or "map", the perceived quality based on selected parameters. The training data was acquired from a subjective video campaign, where participants watched and assessed short RTP video clips which were subjected to varying network conditions. 98 training and 20 validation samples were used in order to teach the neural network the relation between a degraded RTP video and MOS. The parameters used were video resolution, amount of movement (in scale of 0, 0.5 and 1) packet loss burst size and packet loss rate. The resulting RNN model has very good performance, with a correlation coefficient of 0.91 with the validation data.

C. Traffic Management

By default, all traffic flows enter the same FIFO-queue in the traffic management system (i.e. the sixth class). Since the system is QoE-driven, only flows which suffer from low MOS are handled by the system. If the estimated MOS of a flow decreases below 3.0, the management system promotes the flow to a higher class, depending on the underlying application. The flow stays in the class until the estimated MOS has remained above 4.0 for a specific amount of time (in this work, the time was arbitrarily set to 99 seconds).

The management system consists of 4 parts: inter-class scheduling, intra-class scheduling, bandwidth management and admission control. Inter-class scheduling operates between traffic classes. When dequeuing, traffic queues with the highest priority are emptied first, or until they reach their bandwidth limit. Meanwhile, intra-class scheduling is implemented with stochastic fair queuing, which provides almost complete fairness between traffic flows residing in the same class. This algorithm is described in detail a few online sources (e.g. [16], [17]). Bandwidth management is implemented using hierarchal token bucket system [18].

The admission control system operates on a flow-basis rather than on classes. When a new flow is about to enter the AP, the management system must decide whether there's enough capacity to fully support this flow. One should notice though that some flows, like bulk transfers, can be accepted at any time, since these flows are throttled down by the management system during congestion. Therefore, the decision depends on other inelastic flows in the AP. Again, a perceptual quality estimate is used to indicate the situation of the flows. If the estimated MOS of most of the inelastic flows in the



Fig. 2. Testbed



Fig. 3. Flowchart of the packet traffic within the controller.

AP is below 4.0, it can be concluded that the AP is full, and allowing new inelastic flows to enter would cause even more quality degradations.

III. TESTBED

The solution was tested in laboratory conditions using a set-up illustrated in Fig. 2. In this set-up, the server plays the role of third-party OTT RTP video servers. Clients are end-user devices which connect to the server and initiate traffic flows (not unlike an ISP customer would watch for example a Youtube video). The access point consists of three components: router, switch and controller PC.

The actual software used in this work is located in the controller, which is illustrated in Fig. 3. First, all traffic is mirrored to the two-phased traffic classification tool via a libpcap interface. The classifier identifies the flows as interactive, streaming, bulk or unknown. This information is then passed to the control script, which in turn commands the traffic control suite. The QoS measurement is performed in the egress point by Qosmet, a light-weight passive QoS measurement tool [19]. In this particular scenario, the single-point measurement feature of Qosmet is utilized. QosmetService is the measurement agent of the system which monitors selected flows and returns the relevant QoS information to the control script. The control script receives the measurement data and controls the measurement (PSQA is also integrated in here). Together these components allow us to estimate and monitor the perceived quality of desired flows in real-time, based on QoS parameters.

The current Linux traffic control suite, iproute2 [20], was used in the implementation of scheduling and bandwidth management. Bulk transfers were generated with D-ITG [21], and network degradations were emulated with netem [22].



(a) MOS score of the RTP video in both runs



(b) Throughputs of the RTP video and the bulk transfer in both runs Fig. 4. Congestion test results

The test scenarios and results are presented in the following section.

IV. TEST RESULTS

Three tests were performed to validate our approach. In the first test, the prioritization between traffic classes is tested. The second test observes the operation of the client differentiation scheme, and the final test validates the admission control algorithm. Additionally, the performance of the system is measured, mainly the time it takes for the system to react to a quality degradation. The same high-definition RTP video sequence [23] featuring a variable bit rate was used as a reference for all test cases.

A. Congestion Test

In this test, the AP is subjected to a heavy bulk transfer, while an RTP video stream is also traversing through the AP. Ideally, this should not cause any issues since there is enough capacity to support the video, and the leftover is used by the adaptive TCP-based bulk transfer. However, the additional traffic causes the RTP stream to suffer from increased delay and jitter, and a queue overflow in the AP causes packet loss. Rather than trying to allocate a static amount of resources for the stream, the flow is managed according to the estimated perceived quality.

The test was repeated 15 times, each test ending up with results similar to Fig. 4. The first dashed vertical line indicates the start of the RTP video, while the second one marks the beginning of the bulk transfer. The TCP-based bulk transfer was generated with D-ITG with a constant data inter-arrival time of 1 second and uniformly distributed packet sizes between 500 and 1000 bytes. In a 2 Mbps link, this amount

of data can alone create a congestion, even without the RTP video.

During run #1, the traffic management system was disabled in order to see how the AP performs in a normal case. We can see from Fig. 4a that once the bulk transfer starts, the quality of the RTP stream begins to drop heavily. However, when the traffic management is on in run #2, a drop in the estimated quality is detected, and the flow is promoted to a higher class. It should be noted from Fig. 4b that even though the quality of the RTP video increased significantly, throughputs remained roughly the same. This is due to the fact that the RTP stream is very sensitive to even small packet loss percentages, which is reflected in the perceived quality. A bulk download, on the other hand, takes only slightly longer to complete in run #2. A further analysis of all the test runs showed that, on average, the MOS of the RTP stream was increased by approximately 0.9 points.

The test demonstrates very well the benefit of a QoEdriven management system. While the bulk download adapts to the varying bandwidth of the RTP stream even without the management system, both streams lose packets due to the fact that TCP congestion control does not throttle down until some packets are lost. While this is not so severe for the bulk transfer, the quality of the video stream decreases significantly. We could think that the RTP stream can be instantly promoted to a higher traffic class, which would result in the same outcome. For a simple example, this is true to some extent. However, when several different applications enter the network with different types of constraints, we must identify not only the required share of resources per application but also how much we can compromise the resources without a loss in perceived quality. This results in a greater utilization of the link, or in other words, more applications with the same amount of resources. By using QoE as a trigger for management, excess resources will be distributed only to applications which truly need them to keep the users satisfied.

B. Subscriber Priority Test

The previous test demonstrated the efficiency of the management system between different traffic classes, but now we observe the results in the context of customer differentiation. In this test, two RTP streams are started simultaneously. The first flow is initiated by a premium user, and the other by a normal user. A typical result of this test is depicted in Fig. 5. Again, the management system is off in run #1. At 10 seconds (the vertical dashed line), both RTP video streams are initiated. After the flow negotiation phase, the flows are detected by Qosmet. The AP bandwidth is set to 2Mbps — enough to support one of these videos at the peak bit rate. During approximately 25-45 seconds, the video bit rate increases, and both of the flows start to suffer from congestion. This is followed by another congestion event at 60-90 seconds, but in run #2, the premium flow has been promoted to a higher class (premium streaming) than the other flow, which is promoted to normal streaming class. The test was executed 15 times, and the average estimated MOS of the premium video increased



by 1.4 points, while there were no significant changes in the video stream of the normal user.

In addition to the improved quality for the premium users, we should also notice that there is hardly any difference between the two runs when considering the perceived quality of the normal RTP stream. Due to congestion, the perceived quality of the normal stream is already intolerable. Therefore, an additional bandwidth distributed to the premium user does not make things any worse for the normal user. This is one aspect which taking QoE into account provides but QoS does not; we are able to detect when bandwidth goes into waste by observing the perceived quality. If the packet loss of an RTP stream is greater than 5 percent, then we might as well re-distribute the bandwidth elsewhere, because the quality is already too poor (the user would most likely disconnect quite soon if this were the case, but the actual flow may still keep running for some time if the connection is not terminated gracefully).

C. Performance Test

In the third test, we measured the performance of the system. We are interested to know how long it takes for the system to detect and react to a quality degradation. There are four factors contributing to the total reaction time: classifier delay, Qosmet measurement delay, traffic management update and processing and communication delays. Netem was used to emulate a 3% packet loss in the default class, which would



force the traffic management to promote the flow to a higher class.

The test was repeated 40 times, and the same RTP sequence was used as in the previous tests. The flow was promoted to a higher class in 2.8 seconds on average, and the total reaction time was 4.0 seconds on average. The reaction time can be adjusted to be quite small, but it is a compromise in accuracy and performance. If both the measurement intervals of the classifier and Qosmet are set small, more computational power is required, but more importantly, a small measurement interval yields inaccurate results. The smaller the interval, the more packet loss bursts and single packet losses affect the reported packet loss rate. Moreover, due to jitter, Qosmet needs to wait for a short amount of time for late packets before considering them lost. If this interval is too small, the reported loss percentage may be larger than it really is. All in all, the measurement intervals should be large enough for accurate results, but small enough for a satisfactory reaction time.

D. Admission Control Test

The admission control system was validated with 10 test runs. The results of one of the test runs can be seen in Fig. 6. During this test it was discovered that a momentary estimated MOS does not necessarily indicate the general condition of a stream: when the bit rate of a video is dropping, the estimated MOS can be very high for a brief amount of time and vice versa. Therefore, the estimated MOS was averaged over a 30-second window in order to get a better idea on how the stream is maintaining quality. Fig. 6a represents the estimated, time-averaged MOS, and Fig. 6b the throughput. The vertical dashed lines indicate the start times of the streams.

When the management system is off in run #1, and both of the streams enter the AP, the quality degrades to an unacceptable level. The figure shows how the second stream starts to use bandwidth without gaining an acceptable MOS score, thus wasting the bandwidth and dragging the other stream down as well. In run #2, the admission control detects that the first stream is already suffering from low quality and therefore rejects the stream from entering the link. The average MOS of the premium video was increased by over 1.3 points in the test runs when the admission control was on. While this test case bears a resemblance to the client priority test, the major difference is that admission control prioritizes flows according to the initiation time, not by application or subscription type. We are using both methods simultaneously to guarantee a higher QoE to at least one of the flows at any given scenario.

V. CONCLUSION

In this work, we studied a QoE-driven network management approach for OTT multimedia services in a network access point. Based on the acquired data, we demonstrated in the context of RTP videos that the system outperforms an access point without such management system. Not only did this solution yield a better utilization of the AP, but we were also able to free resources which were previously wasted without improving perceived quality. The latter in turn resulted in better QoE for premium users. This goal is especially meaningful from a financial perspective, since the ISP needs to provide the improved quality to premium users without causing a churn of normal customers.

The next step is to broaden the scope of this work by supporting more applications. Currently, there are several challenges which need to be overcome before this can be achieved: the first issue is the single-point measurement setup, which does not allow measuring of all QoS parameters easily, like one-way delay. The traffic classification tool operates mostly fine, but it has difficulties in identifying applications which share similar statistical features (e.g. HTTP file download and HTTP video streaming). Also, as discussed before, applications have unique limits in the tolerance of network disturbances, which means that we would require more results from subjective quality assessments. Currently, the system supports only the assessment of RTP video streams. Finally, given the predominance of HTTP-based video streaming in most web-based scenarios, it would be interesting to augment this work with a suitable quality model for HTTP-based video.

ACKNOWLEDGMENTS

The research behind this paper was conducted within IPNQ-SIS (IP Network Monitoring for Quality of Service Intelligent Support), a Celtic Call 7 project. The authors would like to thank Tekes, the Finnish Funding Agency for Technology and Innovation, for financially supporting this research.

REFERENCES

 [1] Cisco Systems, Inc. (2012, May) Cisco Visual Networking Index: Forecast and Methodology, 2011-2016. [Online]. Available: http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ ns705/ns827/white_paper_c11-481360.pdf [Accessed: 6 Sep. 2012].

- [2] P. Le Callet, S. Möller and A. Perkis, Eds. (2012, Jun.) Qualinet white paper on definitions of quality of experience (2012). [Online]. Available: http://www.qualinet.eu/images/stories/whitepaper_v1. 1_dagstuhl_output_corrected.pdf [Accessed: 6 Sep. 2012].
- [3] K. Kilkki, "Quality of experience in communications ecosystem," *Journal of Universal Computer Science*, vol. 14, no. 5, pp. 615–624, Mar. 2008, [Online] Available: http://www.jucs.org/jucs_14_5/quality_ of_experience_in [Accessed: 6 Sep. 2012].
- [4] K. Kerpez, D. Waring, G. Lapiotis, J. Lyles, and R. Vaidyanathan, "Iptv service assurance," *Communications Magazine*, *IEEE*, vol. 44, no. 9, pp. 166 –172, Sep. 2006.
- [5] J. Lloret, M. Garcia, M. Atenas, and A. Canovas, "A qoe management system to improve the iptv network," *International Journal of Communication Systems*, vol. 24, no. 1, pp. 118–138, 2011.
- [6] M. Zapater and G. Bressan, "A proposed approach for quality of experience assurance of ipty," in *Digital Society*, 2007. ICDS '07. First International Conference on the, Jan. 2007, p. 25.
- [7] K. Haugene and A. Jacobsen, "Network based qoe optimization for "over the top" services," Master's thesis, Norwegian University of Science and Technology, Department of Telematics, 2011.
- [8] B. Lewcio and S. Moller, "A testbed for qoe-based multimedia streaming optimization in heterogeneous wireless networks," in *Signal Processing* and Communication Systems (ICSPCS), 2011 5th International Conference on, Dec. 2011, pp. 1–9.
- [9] M. Varela and J.-P. Laulajainen, "Terminal-side qoe estimations for cross-layer network control," in *Wired/Wireless Internet Communications*, ser. Lecture Notes in Computer Science, X. Masip-Bruin, D. Verchere, V. Tsaoussidis, and M. Yannuzzi, Eds. Springer Berlin Heidelberg, 2011, vol. 6649, pp. 140–149.
- [10] K. Piamrat, A. Ksentini, C. Viho, and J.-M. Bonnin, "Qoe-aware admission control for multimedia applications in ieee 802.11 wireless networks," in *Vehicular Technology Conference*, 2008. VTC 2008-Fall. IEEE 68th, Sep. 2008, pp. 1–5.
- [11] S. Tartarelli and G. Nunzi, "Qos management and congestion control in wireless hotspots," in *Network Operations and Management Symposium*, 2006. NOMS 2006. 10th IEEE/IFIP, apr 2006, pp. 95 –105.
- [12] L. Zheng, D. Hoang, and M. Li, "Applying fair intelligent congestion control in a hybrid qos architecture for wireless environment," in *Information, Communications and Signal Processing*, 2009. *ICICS 2009.* 7th International Conference on, Dec. 2009, pp. 1–6.
- [13] M. Hirvonen, "Two-phased network traffic classification method for quality of service management," Master's thesis, University of Oulu, The Department of Electrical and Information Engineering, 2009.
- [14] Methods for subjective determination of transmission quality, ITU-T Recommendation P.800, Aug. 1996.
- [15] E. Gelenbe, "Random neural networks with negative and positive signals and product form solution," *Neural Computation*, vol. 1, no. 4, pp. 502– 511, 1989.
- [16] M. Brown. Linux traffic control howto. [Online]. Available: http:// linux-ip.net/articles/Traffic-Control-HOWTO/ [Accessed: 7 Sep. 2012].
- [17] B. Hubert. Linux advanced routing & traffic control howto. [Online]. Available: http://lartc.org/lartc.html [Accessed: 7 Sep. 2012].
- [18] M. Devera. Htb home. [Online]. Available: http://luxik.cdi.cz/~devik/ qos/htb/ [Accessed: 7 Sep. 2012].
- [19] Converging Networks Laboratory, VTT. Qosmet enabling passive qos measurements. [Online]. Available: http://www.cnl.fi/qosmet.html [Accessed: 7 Sep. 2012].
- [20] The Linux Foundation. iproute2. [Online]. Available: http: //www.linuxfoundation.org/collaborate/workgroups/networking/iproute2 [Accessed: 7 Sep. 2012].
- [21] Universita' degli Studi di Napoli "Federico II", Dipartimento di Informatica e Sistemistica. D-itg, distributed internet traffic generator. [Online]. Available: http://www.grid.unina.it/software/ITG/ [Accessed: 7 Sep. 2012].
- [22] The Linux Foundation. netem. [Online]. Available: http: //www.linuxfoundation.org/collaborate/workgroups/networking/netem [Accessed: 7 Sep. 2012].
- [23] Amsterdam Blender Institute. Sintel, the durian open movie project. [Online]. Available: http://www.sintel.org [Accessed: 19 Sep. 2012].