IMPROVEMENT OF VOIP QUALITY BY PACKET DROPPING IN ADSL ROUTERS

Qin Dai, Matthias Baumann
Technische Universität Dresden, Dresden, Germany

Ralf Lehnert
Technische Universität Dresden, Dresden, Germany

Keywords: VoIP, QoS, ADSL Downlink, Queueing Delay, End-to-End Delay, Loss Bursts.

Abstract: Packet dropping is known as a simple mechanism to control TCP traffic. In this paper, TCP packet dropping is introduced in the egress router of an ADSL downlink. The aim is to improve the quality of VoIP connections that compete with TCP applications in downlink direction. The ADSL downlink buffer is assumed to operate as simple FCFS queue. Different simulations have been conducted that evaluate the mechanism in two scenarios. Firstly, the long-term impact of the mechanism both on VoIP application and TCP applications is investigated. Secondly, with more realistic network settings, the effectiveness of the mechanism for a short-time real speech is evaluated. The speech's PESQ estimate is used to assess the service quality. The results indicate that in both cases packet dropping can improve the VoIP quality. However, the required high dropping ratio can result in TCP traffic bursts and therefore unstable VoIP quality as well as bad TCP performance.

1 INTRODUCTION

As one kind of real-time applications, VoIP (voice over IP) has its QoS (quality of service) requirements on delay, delay variation and packet loss. Among them, delay related measurements are most critical. The first big problem caused by delay is the talker overlap. In a two-way conversation it is hard to continue when both sides start talking at the same time. The presence of voice echo also has a significant impact on delay sensitivity. Moreover those packets which have been delayed for "too" long time will be discarded by the de-jittering buffer in the receiver.

VoIP packet delay, which can be caused by different reasons, can be characterized into two basic types: the fixed delay and the variable delay. Fixed delay includes coding / decoding delay, processing delay, propagation delay, serialization delay etc. All of these delays are fixed according to the applied technologies, e.g. the algorithms for coding / decoding, the bit rate for the packet serialization, and the propagation delay itself. Variable delay mainly is caused by varying queue occupations in network nodes which largely depend on the situation of competing traffic streams. The queueing delay of a VoIP packet especially can become serious when entering a bottleneck link with fast packet stream aggregation at the input side and slow release at the output side.

In recent years, ADSL (Asymmetric Digital Subscriber Line) technology has become one of the most popular access solutions for broadband networks. In ADSL access network, voice and data traffic will be multiplexed / demultiplexed in its edge node to the IP/ATM network named DSLAM (Digital Subscriber Line Access Multiplexer). From the comparatively low speed in the uplink (typically 128Kbit/s), one can expect that QoS problems mainly arise in this direction, and most of the research focuses on the uplink problem, e.g. (Orozco_barbosa, Siddiqui, Yongacoglu, 2002). Although the downlink speed of ADSL may be quite large (up to 8 Mbit/s), the considerable bandwidth difference between core and access network leads to queueing of downstream packets in the DSLAM. If the IP network operator does not support different service types for VoIP and data applications, FCFS

Dai Q., Baumann M. and Lehnert R. (2007).
IMPROVEMENT OF VOIP QUALITY BY PACKET DROPPING IN ADSL ROUTERS.
In *Proceedings of the Second International Conference on Signal Processing and Multimedia Applications*, pages 265-272
DOI: 10.5220/0002135502650272

(First Come First Serve) queueing may be necessary. In this paper, we consider such a scenario: different traffic streams including VoIP, TCP (Web and FTP traffic) go through the core IP network and pass the access node DSLAM. The link between DSLAM and ADSL modem becomes the bottleneck in the downlink of the communication. The FCFS queueing in DSLAM leads to unacceptable delay for VoIP packets.

ITU-T recommendation G.114 states that the one-way VoIP communication delay should not exceed 150 ms. We assume a voice communication with ADSL access on both end points and consider the downlink performance in one of the points. We neglect the delay in the Internet and assume that a non-adaptive de-jittering buffer with the size of 150 ms is used in the receiver side. For ADSL links with 1 MBit/s downlink and 128 kBit/s uplink bit rate, up to 100 ms of the total delay budget should be reserved for the remote uplink part of the VoIP connection. In order to restrict the downlink delay to the remaining budget, the number of packets queued by TCP connections has to be controlled.

As one kind of TCP traffic control mechanisms, packet dropping can effectively limit the TCP sending rate and has been implemented with different intentions. To avoid network congestion and improve the overall performance, Random Early Detection (RED) uses a linearly increasing function of the queue length to probabilistically drop incoming packets with the expectation that the remote sender can slow down its transmission rate (Floyd, Jacobson, 1993). (Shyu, Chen, Luo, 2002) introduces an adaptive dropping mechanism in the router for Internet congestion control. In this paper, we implement packet dropping in the ADSL router at the user's side and expect that the number of TCP packets delaying VoIP packets in the DSLAM queue can be controlled. We investigate the influence of the mechanism both on the behaviour of the TCP control loop and on the VoIP application. In our simulations, we firstly investigate both VoIP and TCP quality on packet level by using a long time VoIP session. In the second simulation, we use a short-time real speech as the VoIP streaming since the real conversation consists of short speech sections and silence periods. Additionally, higher bandwidths both in up and down links are reserved for the ADSL user in this simulation. A perceptual evaluation of the VoIP quality will be conducted by using the PESQ model.

The remainder of this paper is organized as follows. In Section 2, the packet dropping mechanism is analyzed and discussed. Later the PESQ model is introduced. In Section 3, the simulation models and the metrics for performance evaluation in two simulations are introduced. The numerical results are presented in Section 4. Finally the conclusions are drawn.

2 BACKGROUND

2.1 TCP Packet Dropping

As already mentioned in the last section, it is necessary to limit the number of packets a TCP connection can place in the DSLAM buffer. TCP is a sliding window-based protocol. The TCP sender restricts its window size according to the minimum of the receiver's buffer size and the congestion window size. In TCP, packet loss is considered as a sign of network congestion. Reacting on packet loss, the sender reduces its congestion window to half of the old value during "Fast Retransmission and Recovery" phase to reduce network congestion. In case of retransmission time-out, the window is even closed to one packet. Thus it seems feasible to avoid large window sizes by introducing "artificial" packet losses in the downstream of active TCP connections. If packets of a single connection are dropped periodically, the TCP window size cannot exceed a certain threshold, thus effectively limiting the maximum number of packets that are placed in the DSLAM buffer. One drawback of this approach is that the dropping ratio should depend on the network parameters, e.g. the number of TCP connections and the lowest round-trip time (RTT), which can lead to very low link utilization and bad perceived performance e.g. for Web applications. Our simulation results will further illustrate the problem. Besides it might appear counterintuitive to drop packets already transmitted successfully over the ADSL link. In spite of the above disadvantages, the approach‘s simplicity could lead to an almost gratis implementation still offering a sensible QoS advantage.

2.2 PESQ Model

As one objective method for speech quality assessment, the approach Perceptual Evaluation of Speech Quality (PESQ) is standardised by ITU-T as recommendation P.862. Figure 1 shows the basic philosophy used in PESQ (ITU-T, 2001). Firstly, a series of delays between original input and degraded output are computed by the time alignment module. Secondly, based on this set of delays PESQ compares the original signal with the aligned degraded one by using a perceptual model. The key to this process is transformation of both the original and degraded signals to an internal representation that is analogous to the psychophysical representation of audio signals in the human auditory system. The differences in internal representation between the original signal and the degraded one are computed in the cognitive model. Finally, an objective listening quality MOS (mean opinion score) is given. PESQ is believed to be able to predict subjective quality with good correlation in a very wide range of conditions, that may include coding distortions, errors, noise, filtering, delay and variable delay (Rix, Hollier, Hekstra, Beerends, 2002). The average correlation coefficient between PESQ and the subjective scores is indicated as 0.935 in (ITU-T, 2001) from 22 known ITU benchmark experiments. (ITU-T, 2001)

Table 1 shows the ACR (Absolute Category Rating) opinion scale used in the development of PESQ (Beerends, Hekstra, Rix, Hollier, 2002). The speech quality is graded from bad to excellent with the PESQ score increases from 1 to 5. Since a comparable voice quality is always expected when the user switches from the traditional telephone to VoIP, a speech with low PESQ value can not be acceptable. Different codecs can introduce different level of degrading to the voice. (Markopoulou, Tobagi, Karam, 2003) indicates that the ideal MOS value of a speech coded with G.729 is around 4. The PESQ value of 3.5 is taken as the criterion for judging the quality of speech in our second simulation, with which the quality of the speech is considered between fair to good as shown in the table.

Table 1: ACR listening quality opinion scale used in the development of PESQ.

Quality of the speech	Score
Excellent	5
Good	4
Fair	3
Poor	2
Bad	1

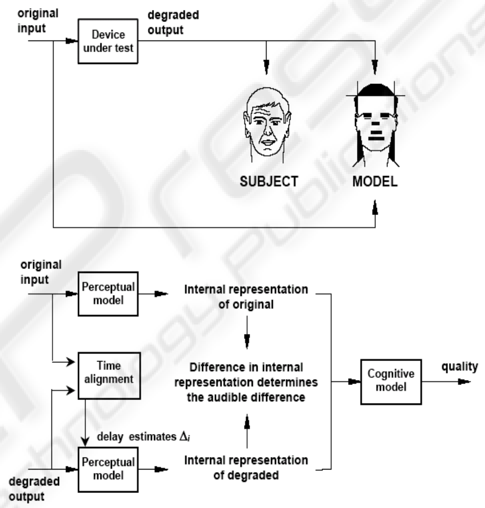

Figure 1: Overview of the basic philosophy used in PESQ (ITU-T, 2001).

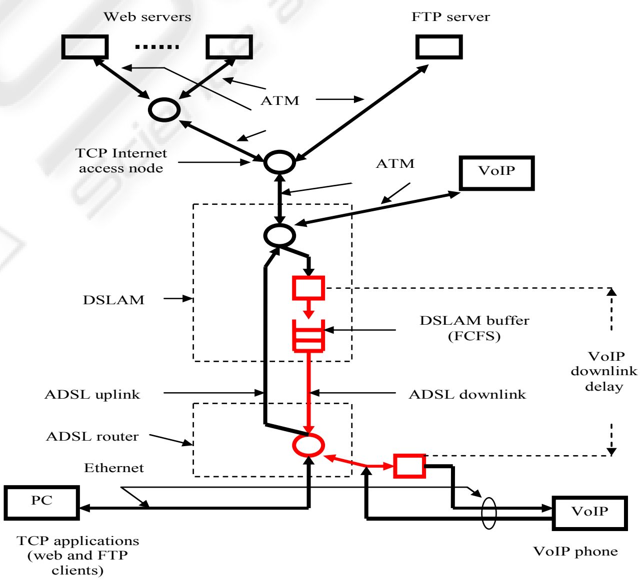

Figure 2: System model (ATM links have a bit rate of 150 Mbit/s).

3 SIMULATIONS AND PERFORMANCE METRICS

3.1 Simulation I

Figure 2 shows the system model adopted in this simulation. Firstly, we investigate the VoIP quality using only Web traffic as background load. In a Web application, usually several short-lived TCP connections run simultaneously. This leads to bursty traffic, since the slow start at the beginning of a TCP connection results in steep packet rate increase. In our simulation, we use two concurrent Web sessions to intensify the traffic burstiness. Within each session, documents from a server farm of 100 Web servers are downloaded in a random fashion. These servers are connected to the Internet part of the network model with link delays ranging from 15 to 35 ms. The traffic model for the Web applications is taken from (Mah, 1997). It specifies distributions for the distance between Web requests (exponential), for the number of objects per main page (Pareto), for the distance between object requests in a page (exponential), and for the object sizes (Pareto). In order to assess the influence of the approach, we measure mean response times for Web pages with a total size of less than 30 Kbytes (the mean bit rate of the Web applications is rather small).

Secondly, only a greedy FTP connection is combined with the VoIP traffic. This allows us to evaluate the possible download rate for a bulk data transfer. Different network delays of 1, 10 and 100 ms for the FTP connection shall yield further insight into the performance of the investigated approach.

The VoIP connection is modelled as bidirectional flow with isochronous payload of 20 bytes every 20 ms (corresponding to G.729A). Including all overheads on the ADSL link (RTP, UDP, IP, PPPoE, AAL, ATM), this leads to an effective amount of 159 bytes per voice packet. In the uplink queue of the ADSL router, scheduling with absolute priority of VoIP packets is assumed.

For this simulation, only packet-level performance metrics have been used to yield an estimate of the eventual speech quality. Recent investigations, e.g. (Markopoulou, Tobagi, Karam, 2003), (James, Chen, Garrison, 2004), find that the mean packet loss ratio should not exceed 1%, and that packet losses in clusters or bursts can be compensated to different degrees by PLC (packet loss concealment) algorithms of the voice decoder. According to this, we assess the QoS of VoIP using the average packet loss ratio as well as the distribution of loss burst lengths. Two losses are considered to be part of the same burst if less than 10 packets are transmitted successfully in between. As the robustness of PLC algorithms may vary, we plot the frequency of loss bursts longer then 2, 5 and 8 packets in the Figures 4 and 5 below.

In the applied network model, all buffer sizes are set to very large values. Thus, packet losses for the VoIP connection only are caused by excessive delay in the DSLAM queue and subsequent packet dropping in the de-jittering buffer of the VoIP phone. In this paper, we only consider a very simple algorithm for the de-jittering buffer. All packets arriving "in time" will be buffered till a threshold time before being played out. Conversely, any packets that arrive late will be dropped by the buffer. The numerical results below assume a delay budget of 50 ms between arrival at the DSLAM buffer and packet delivery to the decoder.

All simulations have been performed for 100 sub-runs each of 10 minutes model time, the confidence level is 95%. The simulations are realized in ns-2 (http://www.isi.edu/nsnam/ns/).

3.2 Simulation II

Figure 3 illustrates the experimental structure. The source signal holding about 20 seconds speech will be coded and packtized into VoIP packets in the VoIP sender. After transversing through the network introduced in Simualtion I, the speech signal will be recovered from those packets with the help of the data extractor and the decoder. Finally, the PMOS (PESQ MOS) value will be computed in PESQ by comparing the degraded speech file with the original one.

The increasing interest on applications like VoD (Video on Demand) and interactive Internet gaming leads to rising bandwidth demand. Correspondingly the ISPs also offer higher bandwidth to their customers. This higher bandwidth might also solve the QoS problem of some applications having comparatively low bandwidth requirements like VoIP. In this simulation, besides the normal rate as set in simulation I, doubled bandwidths are set to both of ADSL up and down links for investigation. The network model introduced in simulation I is adopted. Two Web sessions running simultaneously are set to the background load. The coding rate of G.729 is set to 8Kbit/s which leads to 20 Bytes payload for every 20 ms. We turn off the VAD (Voice Activity Detection) function in codec to maximize the bandwidth requirement of VoIP application. For estimating voice quality, the mean

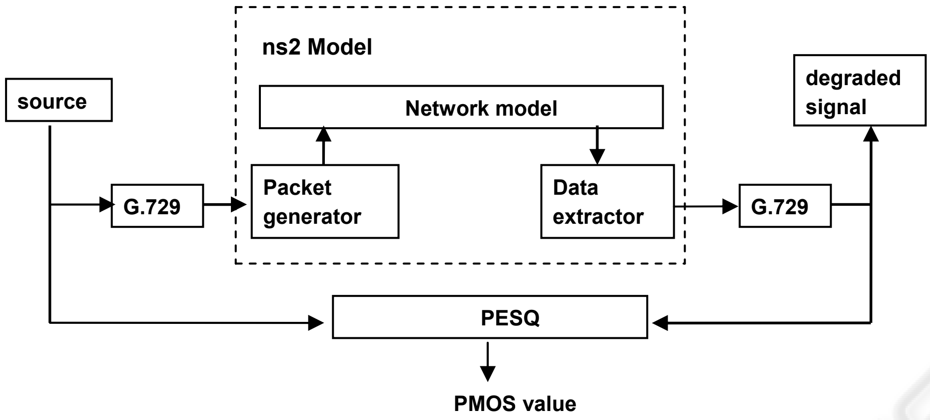

Figure 3: Structure in simulation II.

packet loss ratio and the number of the loss burst in 20 s, as well as the mean PESQ value and the bad PESQ ratio are used. The latter measure indicates the relative frequency of PESQ values below 3.5. In simulation I, the loss distance (the difference in sequence numbers between two successively lost packets) is defined as 10 under the assumption that PLC is disabled in the decoder. According to (Levy, Zlatokrilov, 2006), small values of loss distance can be used when FEC (Forward Error Correction) or PLC mechanisms are enabled. Based on the assumption that the codec G.729 only can repair 30 to 40 ms losses (Perkins, Hodson, Hardman, 1998), the number of loss bursts with length great than 2 (20 ms per packet) is depicted in Figure 6 below. In both cases we assume that 100 ms delay budget should be reserved for the uplink of the other side and use a non-adaptive de-jittering buffer with size of 150 ms in the receiver. The simulations have been performed for 10x40 times; within each time the whole transmission of the test speech file is completed. In order to clearly separate to sub-runs, a silence period of 300 ms is introduced between two speech transmissions.

4 SIMULATION RESULTS AND COMPARISONS

4.1 Simulation I

Figure 4 shows the VoIP quality in ADSL downlinks with different delay budgets when competing with 2 Web sessions. To ensure a VoIP packet loss ratio below the tolerable value of 1%, about 500 ms delay budget should be reserved for the VoIP downlink (between DSLAM and the receiver) which is far above the limitation.

Figure 5 shows the performance of VoIP and TCP with packet dropping in ADSL router. As discussed in Section 2, the introduced dropping ratio should increase with the number of TCP connections. In order to reduce the queueing delay of VoIP packet in DSLAM to an acceptable level of e.g. 50 ms, the queue length at arrival of a VoIP packet should not exceed 4 packets (TCP packet size: 1500 bytes, downlink bit rate: 1 Mbit/s). In the case of only one connection, the TCP window size therefore should be limited to 4 packets. With Web sessions in the network, TCP packets from different connections in a session can arrive simultaneously at the DSLAM queue and provoke a larger queue occupation. To ensure that the accumulated queue length will not exceed 4, the window size for each TCP connection should be further limited. Hence, a very high drop ratio should be introduced which may lead to low link utilization. High packet loss ratios cause poor TCP behaviour during the start of the connection and thus result in bad perceived performance especially for Web applications. Additionally, the dropping ratio should be adjusted to a value that ensures the VoIP quality for the lowest possible packet round-trip time (RTT) of all TCP connections. In case of actually larger RTTs, a certain fraction of the TCP packets will be in transit on links or waiting in queues other than the DSLAM. Thus, the TCP performance will be worse than necessary.

As shown in Figure 5 (left side), discarding more than 5% TCP packets can push the VoIP packet loss ratio (fraction of late packets) below 1%, while the frequency of packet loss bursts still can be serious. We consider that the number of loss bursts with length greater than 5 packets should not exceed 5 within 10 minutes. To satisfy this requirement, it follows from the figure that more than 10% dropping ratio should be introduced. The right side of Figure 5 demonstrates that then both FTP throughput and mean Web response time deteriorate considerably.

4.2 Simulation II

Figure 6 shows the VoIP performance in ADSL downlinks when competing with 2 Web sessions.

In the case of 1 Mbit/s downlink bit rate, one may find that our algorithm performances much better on a short-time speech than on a long speech (simulation I). The fraction of late VoIP packet can be controlled below our criteria of 1% with a drop period of 40 packets, while simulation I led to a dropping ratio of more than 1/25. As shown in Figure 6, the *mean* PESQ value can be quite good with only 1/90 drop ratio. At the same time, the mean number of loss bursts in 20 seconds is only about 1. Nevertheless, more than 10% drop ratio is required to hold the relative frequency of bad PESQ values below 1%. This means that the user will not more than once experience a bad VoIP quality during 30 minutes.

Table 2 depicts the VoIP parameters with 2 MBit/s downlink bit rate on the ADSL link. Packet dropping is not applied. The table shows that the packet-level performance metrics do not indicate QoS problems. The fraction of late VoIP packets in 20 s is only 1% which is almost within our criterion. The mean number of loss bursts with burst length greater than two is 0.4 times in 20 s. But the serious situation of VoIP application under this scenario can still be seen from the bad PMOS fraction which is indicated as 15%. This means that the user will be dissatisfied with the conversation quality more than 15 times within 33 minutes. The green curve in Figure 7 addresses this problem more clearly. According to the bursty background traffic, the VoIP quality can be very different over time. In the total 100 simulation runs, there are 15 PESQ value drops below 3.5, two estimates even fall below 3.0. The worst PESQ value is indicated as 2.75 in Table 3.

Figure 6 indicates that dropping every 30th packet could be a good operation point in this scenario. With this setting, only 0.06% of the VoIP packets are lost, and the occurrence of bad PESQ values can be almost suppressed (indicated to be under 0.8%). However, the QoS can not be guaranteed, since dropping of TCP packets can trigger loss bursts of VoIP. When the lost TCP segment is successfully retransmitted, the receiver will send back a packet that acknowledges the lost together with subsequent correctly transmitted packets. This can lead to a sudden shift of the transmission window and therefore cause a burst of packets. This TCP burst can delay a sequence of VoIP packets, leading to a VoIP loss burst and thus degrading the voice quality. The sample curve of PESQ values in Figure 7 highlights this aspect. Although the PESQ estimate can be improved for most of the speech intervals, the approach sometimes can even make worst the speech quality. This problem seems to be solved by further increasing the TCP packet dropping ratio and therefore limiting the TCP window size furtherly. With a dropping ratio extended to e.g. 1/15, sustained good PESQ values can be obtained, varying from 3.71 to 3.75. As seen from simulation I, this, however, will lead to severely reduced TCP performance.

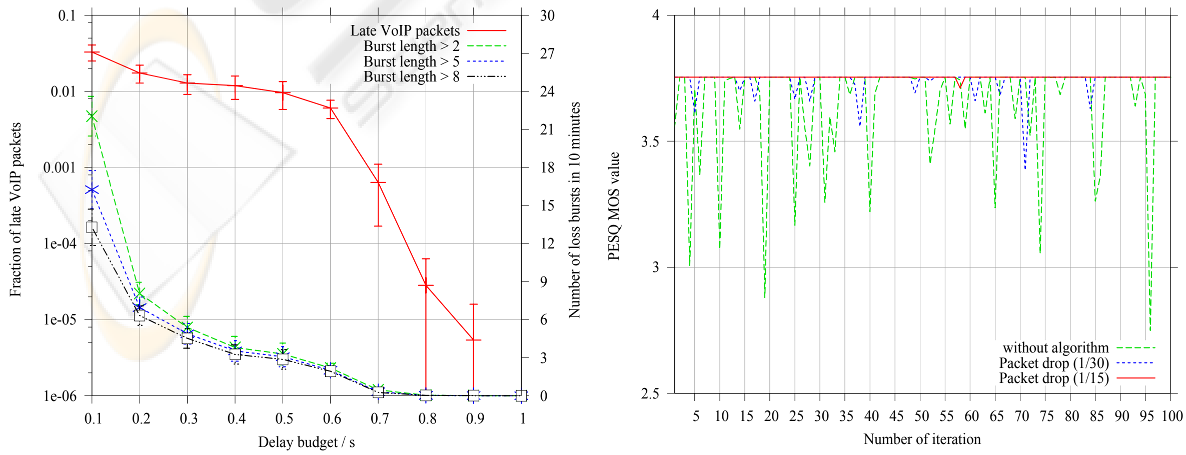

Figure 4: Performance of VoIP in ADSL downlinks in simulation I & Figure 7: PESQ values with different algorithms.

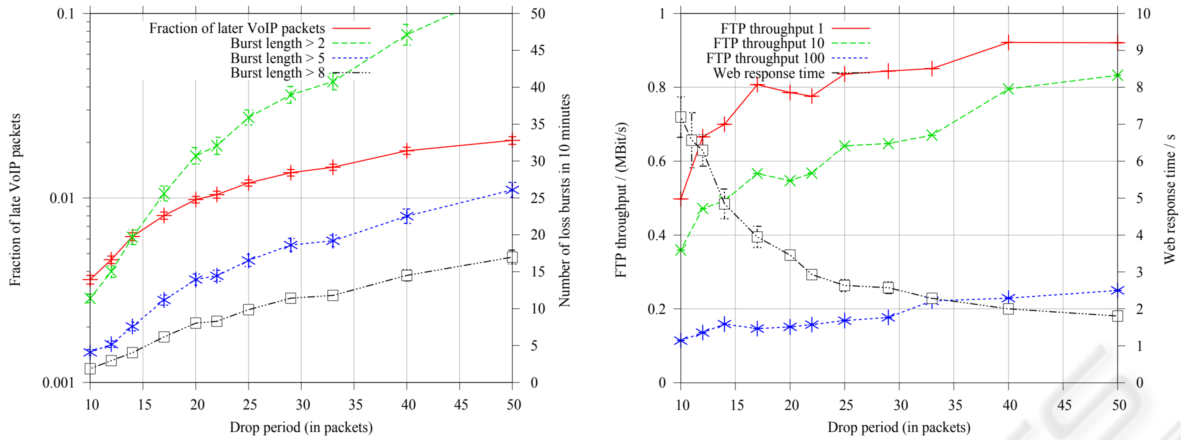

Figure 5: Performance of VoIP and TCP with packet dropping in simulation I.

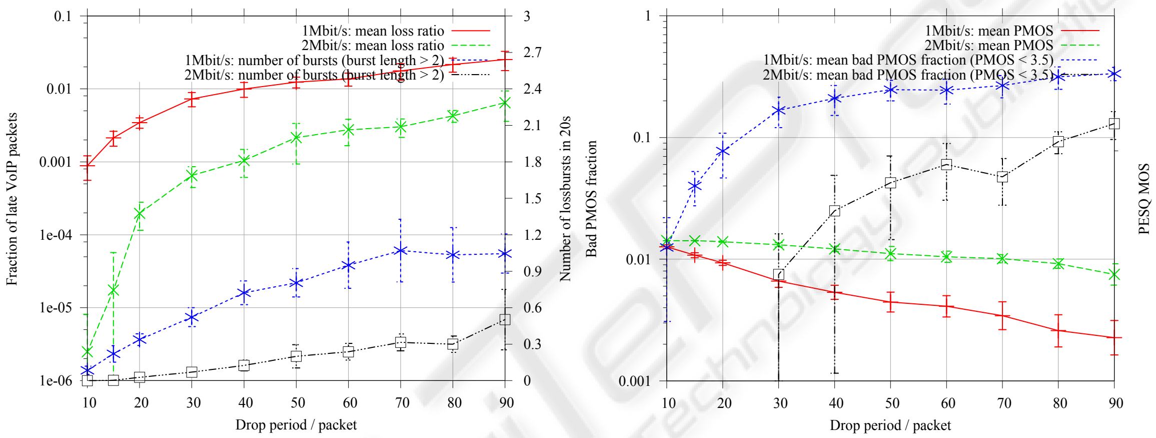

Figure 6: Performance of VoIP with packet dropping in simulation II (single bandwidth).

Table 2: VoIP parameters with double bandwidth without packet dropping.

Fraction of VoIP packets	Number of loss bursts (Burst length > 2)	Bad PMOS (PMOS < 3.5)
1.01%	0.407	15.8%

Table 3: PESQ values from different algorithms (double bandwidth).

	Mean	Best	Worst	Bad fraction (<3.5)
No algorithm	3.654	3.754	2.750	0.15
Packet drop (1/30)	3.740	3.754	3.385	0.01
Packet drop (1/15)	3.754	3.754	3.711	0

5 CONCLUSION

In this paper we investigate the simple approach of TCP packet dropping to reduce the queueing delay of VoIP packets in the ADSL downlink under different TCP traffic conditions. We assess the performance of the algorithm both in a critical and a more practical scenario. In the former case VoIP has to compete for the limited bandwidth with TCP traffics during a long time. In the latter one, a short-time speech is investigated with single and double bandwidth in the ADSL links. Our simulation results indicate that VoIP suffers from QoS problems even with high ADSL downlink bandwidth. In both cases, the VoIP quality can be improved by TCP packet dropping. The required high dropping ratio can, however, lead to poor TCP behaviour, e.g. low FTP throughput which yields low link utilization, and long Web response time. Furthermore, the TCP packet bursts indirectly caused by the dropping algorithm can result in unstable VoIP quality.

ACKNOWLEDGEMENTS

This work was supported by the Sphairon Access Systems GmbH.

REFERENCES

Beerends J. G., Hekstra A. P., Rix A. W., Hollier M. P., 2002. *Perceptual Evaluation of Speech Quality (PESQ), the new ITU standard for end-to-end speech quality assessment. Part II – Psychoacoustic model.* Journal of the Audio Engineering Society (J. Audio Eng. Soc.), 2002.

Floyd S., Jacobson V., 1993. *Random Early Detection gateways for congestion Avoidance*. IEEE/ACM Transaction on Networking, Vol. 1 No. 4, Aug. 1993, pp.397-413.

ITU-T, 2001. ITU-T Recommendation P.862, 2001.

James J. H.., Chen B., Garrison L., 2004. *Implementing o f VoIP: A Voice Transmission Performance Progress Report*, IEEE Communication Magazine, pp. 36-41, July 2004.

Levy H., Zlatokrilov H., 2006. *The effect of packet dispersion on voice applications in IP networks*. IEEE/ACM transaction on networking, Vol. 14, No. 2, April 2006.

Mah B. A., 1997. *An Empirical Model of HTTP Network Traffict*, Proc. of IEEE Infocom 1997, April 1997.

Markopoulou A.P., Tobagi F.A., Karam M.J., 2003. *Assessing the Quality of Voice Communications Over Internet Backbones.* IEEE/ACM Transactions on Networking, vol. 11, no. 5, October 2003, pp. 747-760.

Orozco-Barbosa L., Siddiqui A., Yongacoglu A., 2002. *Efficient integration of voice and data for transmission over Digtal Subscriber Line (DSL)*, Proc. of the 2002 IEEE Canadian Coference on Electrical & *Computer Engineering, 2002.*

Perkins C., Hodson O., Hardman V., 1998. *A survey of packet loss recovery techniques for streaming audio*. IEEE Network, September/October 1998.

Rix A. W., Hollier M. P., Hekstra A. P., Beerends J. G., 2002. *Perceptual Evaluation of Speech Quality (PESQ), the new ITU standard for end-to-end speech quality assessment. Part I – Time alignmentl.* Journal of the Audio Engineering Society (J. Audio Eng. Soc.), 2002.

Shyu Mei-Ling, Chen Shu-Ching, Luo Hongli, 2003. *Per-class queue management and adaptive packet drop mechanism for multimedia networking.* Proceedings of the 2003 International Conference on Multimedia and Expo - Volume 3, 6-9 July 2003.

Spring N. T., Chesire M., Berryman M., Sahasranaman V., Anderson T., Breshad B., 2000. *Receiver Based Management of Low Bandwidth Access Links.* IEEE INFOCOM'00, Tel Aviv, Israel, Apr. 2000.

The Network Simulator- ns-2, from http://www.isi.edu/nsnam/ns/.