The Role of Log Entries in the Quality Control of Video Distribution

Ismo Hakala, Sanna Laine, Mikko Myllymäki, and Jari Penttilä

University of Jyväskylä, Kokkola University Consortium Chydenius
P.O. Box 567, FIN-67701, Kokkola, Finland
{ismo.hakala,sanna.laine,mikko.myllymaki,
jari.penttila}@chydenius.fi

Abstract. Diversification of university teaching with the help of video lectures has become much more common during the past few years. Once videos have become an essential part of teaching arrangements, whoever organizes the teaching must also pay attention to factors related to videos in quality system work for teaching. In the video production process it is the factors related to video transmission that exert influence on the usability of videos and set limitations for their production. A lot of information about those factors can be obtained from the media server log files. The particular focus of this paper is on the functionality of the connection between a server and a client and its effect on users. The paper deals with information obtained from a media server's log file, describes the activities around collection and handling of log data, and introduces a preliminary classification for monitoring video transmissions. The results obtained from the media server's log files are presented in accordance with that classification system at the end of the paper.

Keywords: Quality of Service, log file, streaming video lecture.

1 Introduction

During the last few years there has been a strong increase in the use of video lectures in many universities. The motivation behind this increased use of videos has typically been the desire to make studying more flexible as regards time and place. The idea is to provide students who live in places that are geographically distant better chances for studying and, at the same time, ensure that the lectures are available also for temporarily absent students. These arrangements also allow students to revise any difficult or complicated lecture items and topics. Diversification of face-to-face teaching, by delivering it as real time video and by providing recorded (on-demand) videos alongside it, makes it possible for a student to study solely with the help of videos without a need to participate in any traditional type of contact teaching. In teaching arrangements, the organizer of the education must then, in addition to video production, pay attention to the possibilities that the student has for viewing videos and to the factors related to video transmission.

L. Atzori, J. Delgado, and D. Giusto (Eds.): MOBIMEDIA 2011, LNICST 79, pp. 168–179, 2012.

Internet connections in homes in recent years have speeded up considerably and the problems related to video transmissions have become less common. However, portable computers and new mobile devices, that are suitable for study, have increased the number of mobile broadband connections which are slower and more unstable in congested networks than traditional broadband connections. Evaluation of the factors related video distribution and habits and activities formed around them is an important part of a quality system for teaching, especially in blended teaching arrangements, where the role of videos alongside face-to-face teaching is important.

The education organizer must, among other things, know the problems associated with video transmission and use, how to detect those problems, how to react on them, the reasons for them and how to influence them. Regardless of the quality of a produced video, the student becomes aware of the problems in video transmission, for example, if the video keeps on stopping during its viewing or in the worst of cases if there is an abnormal break during it [1],[5],[6]. In a quality system framework, these related factors can be taken into account to some extent when soliciting course feedback from the students. Information received through course feedback, however, is subjective and general. The media server's log files provide, for the organizer, an information source that is adequate and accurate but also fairly difficult and little used. The log files provide information, among other things, about server work load, the system that the students use and bandwidth employed as well as about possible problems in transmission or abnormal breaks that are independent of viewers. Partial answers are obtained as to the causes of the problems found, which makes it possible to take anticipatory measures to those problems.

Log data can be used for example to solve problem situations and to monitor the amount of usage and availability of service. Statistical summaries based on log entries can be helpful when deciding about development directions related to the use of videos in teaching and to their production process. For example, information obtained from the log data about the bandwidth used in video transmissions can be useful when considering questions related to image quality and size. The log data can also be used when drawing up guidelines for video use.

This paper deals with information obtained from a server's log file, describes the activities around collection and handling of log data, and introduces a preliminary classification for monitoring the development of video transmissions. The results obtained from analyses of Windows Media Server's (WMS) log files of 2008-2010 for Master Studies in Mathematical Information Technology at the Kokkola University Consortium of the University of Jyväskylä are presented in accordance with a preliminary classification at the end of the paper.

2 Log File Handling

Windows Media Server's log file comprises information that is obtained from the server as well as from its clients. The server collects some of the information to be entered in the log immediately when a client contacts the server. The client collects information throughout the duration of the connection, and only sends small amounts of information to the server while a video is being viewed. During the viewing, the client sends information related to the user's activities on the media player (forward, rewind, pause). On the other hand, the server constantly collects data needed for the log. At the end of a successful connection, the client sends the data it has collected to the server. The server combines this data with the data it itself has collected during the connection. In situations where the client cannot enter data in the log, the server simulates the log entry and enters there the information that it has collected dynamically, in addition to the information it collected when forming the connection. In cases where streaming does not start at all (for example, when there is a wrong user name or incorrect file name), streaming related data naturally is missing and only connection time data is entered in the log. [4]

2.1 WMS Log File

WMS records data on 52 fields in its log (Table 1). With the help of these, it is possible to monitor, in multiple ways, the use of video files distributed from the media server [4].The fields give information about the client, server, data network, video transmission, video to be transmitted, and the use of the transmitted video.

Table 1. Windows Media Server's log file fields

c-ip	c-cpu	c-pkts-recovered-resent
date	filelength	c-buffercount
time	filesize	c-totalbuffertime
c-dns	avgbandwidth	c-quality
cs-uri-stem	protocol	s-ip
c-starttime	transport	s-dns
x-duration	audiocodec	s-totalclients
c-rate	videocodec	s-cpu-util
c-status	channelURL	cs-user-name
c-playerid	sc-bytes	s-session-id
c-playerversion	c-bytes	s-content-path
c-playerlanguage	s-pkts-sent	cs-url
cs(User-Agent)	c-pkts-received	cs-media-name
cs(Referer)	c-pkts-lost-client	c-max-bandwidth
c-hostexe	c-pkts-lost-net	cs-media-role
c-hostexever	c-pkts-lost-cont-net	s-proxied
c-os	c-resendreqs	
c-osversion	c-pkts-recovered-ECC	

When the factors related to video transmission are being examined, the information obtained from the media server's log entries that is usually made use of in these circumstances includes the duration of connection, status, bandwidth used for video transmission, number of rebufferings and the time used for buffering, as well as the number of packets sent and received ([1], [5], [6]).

The X-duration field contains information about the client's viewing time. Information about network connection used for the viewing is given by the *avgbandwidth* field, which indicate the average bandwidth used by the client during the connection. Success of the connection is indicated by the value of the *c-status* field. The values in this field reveal whether the connection was a complete success or whether there were breaks and reconnections during the connection, or whether the viewing was terminated as a result of a connection breakdown or a server internal error. The field also tells about unauthorized attempts to view videos and of attempts to view videos not on the server at all.

The *transport* field tells about the transport protocol used. The log has several fields related to data transmission. Most of these fields are given a value only when the UDP protocol is in use. This means that some of the functions related to transmission reliability are moved to the application layer. For example, the *c-pkts-lost-client* field, which tells about the packets lost by the client, is given a value only when the UDP protocol is used. The same applies to the *c-pkts-lost-net* and *c-pkts-recovered-resent* fields. The *s-pkts-sent* field contains the number of packets the server has sent to the client and the *c-pkts-received* field, in turn, the number of packets correctly received at first attempt by the client. Both of these fields are given a value whether the UDP or TCP protocol is used.

Fields *c-buffercount* and *c-totalbuffertime* give information about the number of bufferings and the duration of time used for buffering. The media player normally uses buffering as its default to reduce the effect of lags in the data network between it and the media server. Changes along the network, such as server overload or bandwidth changes, may regardless cause emptying of the buffer, the effect of which is to stop playback on the viewer's computer for the time it takes to rebuffer.

2.2 Handling and Parsing of Log Files

There exist suitable commercial programs to analyze log files. The programs enable, in a limited manner, filtering of log files, but they do not fully meet needs for joining log entries or removing or combining certain data. For this reason preparatory joining of the log files and their filtering was done with a separate parser. The transmission of videos was then examined with the help of parsed log files by analyzing the data with software dedicated for statistical analysis.

The problem with Windows Media Server's logs is that they contain a lots of lines about client activities and every line has 52 fields. Thus a log file may contain tens of thousands of lines, leaving its structure quite unorganized. As each movement (forward, rewind) along the timeline leaves its mark on the log, the use of log data as such with an analysis software is often not very practical. For this reason, it seemed a reasonable idea to assemble a file on the basis of the log file. In that new file, all the user operations during a session were combined into a single entry. This joining of data was realized with a simple PHP parser, which automatically combines log files that are input as parameters into a single file in a manner desired.

During parsing, only operations related to successful viewings, wrong logins, and unsuccessful viewing requests were combined into one log entry. In case of an unsuccessful viewing request, of the attempts directed to the same file only one attempt per hour was left on the parsed file. In case of a forgotten password or missing user rights, it was also removed all consecutive unsuccessful login attempts except one. Viewing sessions that ended abnormally were kept as such also in the parsed log. During parsing, any entries by our own personnel were removed from the log data, too.

When combining successful operations related to the same session, the fields were subjected to various operations in accordance with the data contained in them. For example, the fields describing transmitted data (among them s-pkts-sent, c-pkts-received, c-pkts-lost-client, c-pkts-lost-net and c-pkts-recovered-resent) were summed up, as was the field defining the duration of viewing time (x-duration). When parsing, the weighted average of the avgbandwidth field was calculated in relation to the values of the x-duration field, to get the average bandwidth truthfully into the parsed file. Depending on the contents, the maximum (e.g., s-totalclients) and minimum (e.g., c-quality) values of some of the fields were also taken.

3 Analysis of Log Data

3.1 Classification of Log Data

When assessing the success of video transmissions we have first roughly classified the log data, from the users' viewpoint, into successful and unsuccessful viewing sessions as well as unsuccessful connection attempts. In [5], successful connections have been classified by their quality into three different groups: good, average or low. In line with the criteria in [5], we have classified successful connections as follows:

A. During a connection, at most 10% packets are retransmitted and there is no rebuffering at all.
B. During a connection, more than 10% packets are retransmitted and there is no rebuffering at all.
C. During a connection, rebuffering takes place.

Unsuccessful connections, in their turn, belong to class

D. The connection terminates abnormally with status 408 or 500.

and unsuccessful connection attempts belong to class

E. No connection is created and status 401 or 404 is entered in the log.

In a class A type connection the log file status is 200 and the portion of retransmitted packets of all packets (packet loss rate) transmitted remains below 10%. The number of packets that are retransmitted is calculated from the log files as the difference between the s-pkts-sent and c-pkts-received field values. This means that the information can be obtained regardless of whether the protocol used is UDP or TCP. Thus the new field gives only mediated information about the number of the packets lost by the client. We define a packet as lost by the client if it cannot arrive at the client by the playback time. According to [6], only when the packet loss rate is above 10% it is shown clearly as packet loss by the client. Therefore, on the basis of log data, in the class A type connection video is transmitted to the viewer without problems.

In a class B type connection the log file status 200 tells that the connection is successful and without problems, but that more than 10% of all the packets are retransmitted. If the number of retransmitted packets is great, it indicates problems with the connection as the number of packets lost by the client starts increasing. For the user, loss of packets may appear as malfunctioning while watching videos, for example, as dropping of video frames [5], [6].

In a class C type connection the status of the log file is either 200 or 210 and there is buffering during the connection. Buffering presents itself as jitteriness in the data transfer connection, and when watching videos these problems become manifest as a halting video image [5], [6]. If the data transfer connection is jittery, the size of the buffer is not necessary big enough to smooth down the variations in the connection. As a result the buffer might become empty. The time taken for rebuffering is usually very short, but the user’s video stops for the duration of the operation. In some cases the player itself can create a connection after a short connection break and continue playback from the same point (status 210). For the user, however, it appears that the video stops once the buffer is emptied.

The quality of a class D type connection before the connection terminated abnormally might have been anywhere between A and C. If the log ends up with status 408 or 500, much less information about the connection quality remains than would remain were the session to end with status 200 or 210. If the log file ends with status 408, it means that the client’s connection to the server has been broken or slowed down to the extent that it makes watching videos impossible. In that case the client cannot create a log entry. Instead the server will create a 408 entry in the log on behalf of the client. There are many reasons for a connection to slow down or break, among them congestion in data communications, problems related to network devices (switch, router, firewall, proxy server, etc.) or a network cable becoming disconnected. Also, a media server can become overloaded, which may lead to a situation where data communications slow down to such an extent that the media server may even have to drop some clients. In that case the log will end up with status 500.

In a class E type connection there has been an error when trying to create a connection and the connection has not been created. If the viewer tries to watch a non-existent recording the log file will have 404 entered, and if the user has not been authorized to watch a video, 401 will be entered.

3.2 Statistics That Can Be Derived from Log Entries

All lecture teaching for the students in Mathematical Information Technology at the Kokkola University Consortium is recorded on video, which is offered to the students both as real time video and on-demand. During 2008-2010 44 courses recorded on video were offered. Of these 11 were recorded in 2008, 15 in 2009, 13 in 2010 and the rest were from 2007. The bandwidth allocated for the users was at least 512 Kbps, also mobile users. The minimum bandwidth required by the videos was 340-380 Kbps. Altogether 187 students watched real time or on-demand videos. Of them, 86% watched in real time and 61% watched on-demand recordings. The average length of on-demand session was 42 min and the average length of real time session was 81min.

Fig. 1 shows the number of sessions viewed in 2008-2010. The videos were viewed altogether 12 673 times (9959 hrs) in this period. The on-demand recordings were viewed altogether 10 850 times (7507 hrs) and real time video 1823 times (2452 hrs). The number of all viewed sessions almost doubled from 2008 to 2010.

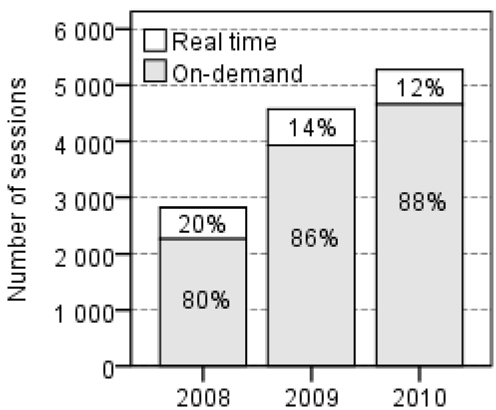

Fig. 1. Number of sessions viewed in 2008-2010

In 2008-2010, there were a total of 1555 class E type unsuccessful connection attempts entered in the parsed log. Status 401 came up in 890 of these and status 404 occurred 665 times. These unsuccessful connection attempts consist of human errors committed by the service provider or by the user. Typical errors made by the service provider include writing mistakes (a link on the learning environment or the name of a video file on the server misspelled) and links written beforehand (a link that is added on the learning environment before the video is moved to the server). Also the user errors consist typically of writing mistakes, such as a misspelled user name or password. In the following statistical evaluation, class E type unsuccessful connection attempts have been left out.

Fig. 2 shows both real time and on-demand connections in the master studies in 2008-2010. The figure is based on the classification discussed earlier. Fig. 2A shows all the connections formed. Fig. 2B shows only the connections that were created successfully.

Of all these connections, approximately 23% (2956 times) terminate abnormally, and belong to class D. One exception notwithstanding, all the terminated connections ended with status 408. Of the on-demand recordings 24% (2581 times) and of the real time videos 21% terminate abnormally. In these cases the average session time was 47 minutes for the on-demand videos and 29 minutes for the real time videos. Most of the students (74%) that used on-demand videos have experienced an abnormal termination of an on-demand recording at least once. The corresponding figure for the real time videos is 47%.

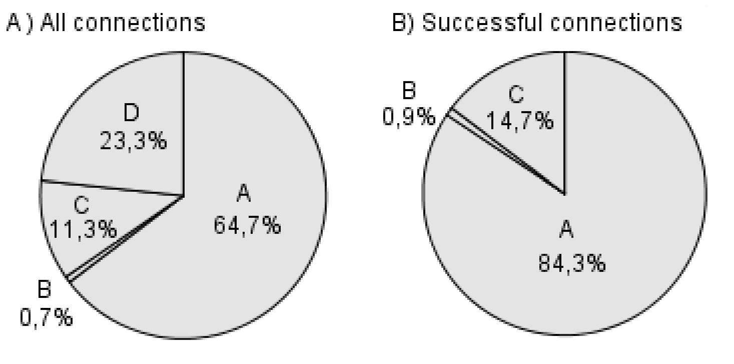

Fig. 2. Real time and on-demand video connections in 2008-2010

The data from 2008-2010 shows malfunctioning due to buffering having occurred in 11% of all connections and in 15% of successful connections. According to Fig. 3 class C can further be subdivided into three subclasses based on how common the malfunction is. A malfunction was regarded as occasional and almost unperceivable for the user if rebuffering took place at most twice. The malfunction was regarded as occasionally repeated if rebuffering took place 3-5 times and often repeated if it happened at least 6 times. In the successful connections most (8.5%) of the malfunctions were occasional, 2.9% occasionally repeated and 3.4% often repeated. During successful viewing sessions, malfunctioning due to buffering occurred slightly less in real time videos (13.6%) than in on-demand videos (14.9%).

Numerous retransmittings of packets may appear to the user as different kinds of faults in the videos. Based on the data from 2008-2010, there were only a few cases of this kind belonging to class B, a mere 0.7% of all connections and 0.9% of successful connections. Of the successful connections, the real time videos belonging to class B formed 0.1% while the corresponding figure among the on-demand recordings was 1.1%.

The portion of all connections that belonged to class A was 65%, and the portion of successful connections belonging to that same class was 84%. According to the log data, these connections appeared faultless to the users. Of the successful connections, on-demand recordings belonging to class A formed 84% and real time videos about 86%.

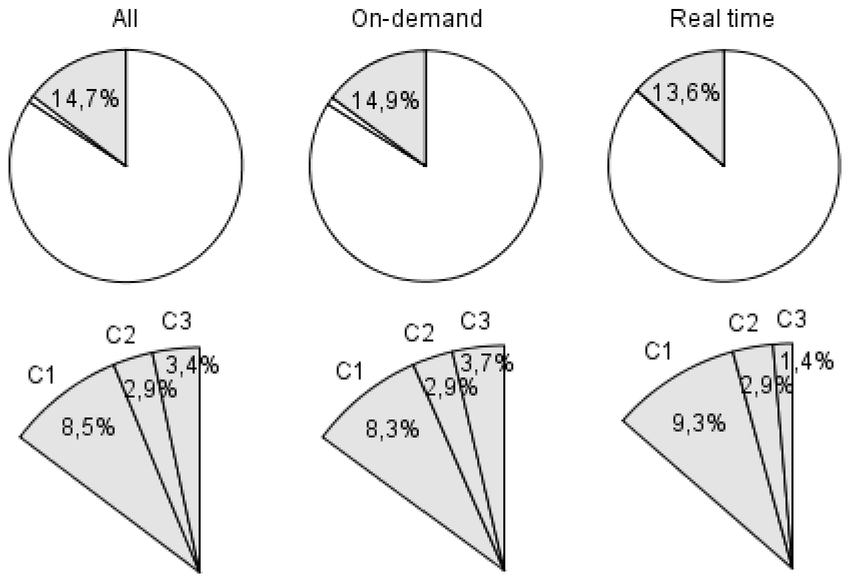

Fig. 3. Class C type connections of all successfully created real time and on-demand video connections in 2008-2010. *(C1=occasional, C2=occasionally repeated, C3=often repeated).*

3.3 Discussion

The classification above enables us to examine video transmission on the basis of the media server's log file more accurately and reliably than before. Judging from the results of the current log data analysis, information about connection quality obtained with the help of student questionnaires remained clearly inadequate. We were surprised, for example, of the relatively large portion of abnormally terminated and faulty connections, which had not been brought to our attention by earlier student questionnaires or other feedback. Based on our earlier research work, only about 3% of our students regarded problems with video image as disturbing [2].

The relative share (23%) of the abnormally terminated connections is fairly large. One cannot obtain any explicit reason for these connection breakdowns because very little information related to data connections is entered to the log. A partial explanation can be found in the transport protocols used for data communications. When the UDP protocol is used, 21% of on-demand video connections terminate abnormally. For the TCP protocol the same figure is 24%. Moreover, according to Fig. 4 the relative percentages for connections lasting less than 5mins are 23% (TCP) and 45% (UDP), and for connections lasting more than 40mins 47% (TCP) and 16% (UDP).

Another possible explanation regarding the on-demand recordings is that the errors might have accumulated for a small number of students. In case of those recordings, about 40% of all abnormal terminations of the recordings were caused by 10 students.

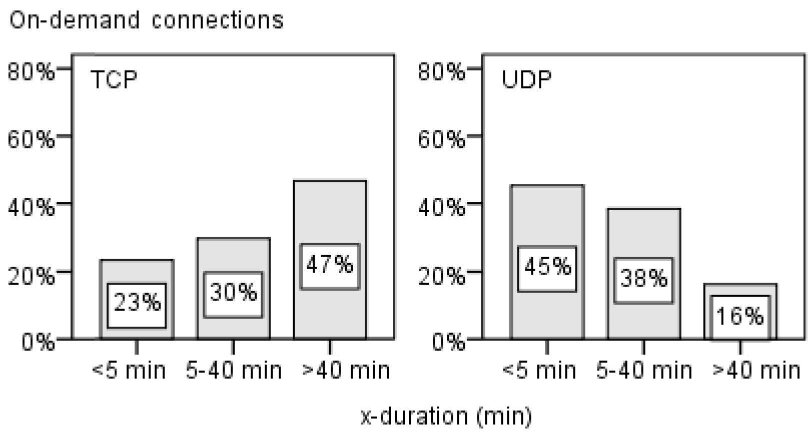

Fig. 4. Abnormally terminated connections for the on-demand recordings

The total share of faulty connections is 12% overall. Depending upon video material used, faulty connections may appear as malfunctions to the user also. In video lectures, the relative portion of static images is predominant. It is therefore likely that, to a large extent, faults in these go unnoticed by the user. The biggest source of the malfunctions is rebuffering. Nevertheless, it is quite rare to find that it happens repeatedly. Increasing the size of the buffer reduces the number of malfunctions, but on the other hand it also increases the delay in playback. Here the latter has no great importance, however.

Some of the factors that affect connection quality and which the user can influence are buffer size and transport protocol. The increase of the buffer size can be done in the media player. In Windows Media Player, the size of the buffer can be selected between 0 and 60 secs, the default value being 5 secs. The user also has a say in what transport protocol to use. Depending on the server settings, the server offers either the UDP or TCP protocol as a default, but the user can force the player or computer to use any of the two. By instructing the students in how to increase the adjustable buffer size in the media player and by making TCP the obligatory transport protocol, some of the problems mentioned above can be eliminated to some extent though not entirely.

The log file also provides information which has more to do with the user or the organizer of the education than with the network connection. In some cases, entries of this type convey useful information that can be used in the development of teaching arrangements. For example, if the user is not authorized to watch a video, a separate entry will be left in the log file about the matter. Usually this is due to that the user has misspelled his/her user name or password or that the user profile has expired. Because these types of entries are quite common in the log file, a facility has been created for the students to allow them to check the functionality of their user ID and also to remotely change their password if needed.

4 Conclusions

The use of lecture videos is already an essential part of educational arrangements in many universities. For this reason, the role of quality control of video transmissions has become more important. Regular monitoring of video distribution quality should form a part of the quality system work. When assessing the quality of operations related to streaming video, the analysis of log data collected by the media server has proved a very useful tool. Analyzing log data helps, above all, in directing development work associated with video technologies. It makes possible also to instruct students in making selections that help one succeed in video viewing.

One of the aims of the quality monitoring is to create indicators for measuring video transmission quality. With the help of the kind of classification that measures quality it is possible to compare the quality of video transmission at different times. This makes it possible to follow the productivity of the development work in video transmission. In addition, these indicators provide us with a tool that enables us to react whenever the quality of transmission for one reason or another may have deteriorated.

If the videos have been made available through Windows Media Server, it is possible to analyze the log data in many different ways. The log data alone, however, is not enough for analyzing how the students have experienced the quality of the videos. To complete the picture, we need to resort to questionnaires directed to the students. More useful material could be collected by mapping students' experiences of watching videos immediately after watching the video. This, however, requires a development of suitable video sharing application.

To ensure that maximum benefit is derived from monitoring log data, that monitoring must be regular. Further automation in future should make it easier to arrange regular monitoring of large log files and would help in the detection of problem situations. With the help of automated monitoring, the education organizer could, for example, be notified daily about problems occurred.

References

1. Chang, C.: Constructing a Streaming Video-Based Learning Forum for Colloborative Learning. Journal of Educational Multimedia and Hypermedia 13(3), 245–263 (2004)
2. Hakala, I., Myllymäki, M.: Video lectures alongside with contact teaching. In: Proceedings of the 18th EAEEIE Annual Conference on Innovation in Education for Electrical and Information Engineering, Praha, Czech Rebublic (2007)
3. Hakala, I., Myllymäki, M.: Quality Management for Mathematical Information Technology Teaching at the Kokkola University Consortium Chydenius. In: Proceedings of the 19th EAEEIE Annual Conference on Innovation in Education for Electrical and Information Engineering, Tallinn, Estonia (2008)
4. Koyun, H.: Logging Model for Windows Media Services (2007), http://www.microsoft.com/windows/windowsmedia/howto/articles/LoggingModel.aspx

5. Terada, N., Kawai, E., Sunahara, H.: Extracting Client-Side Streaming QoS Information from Server Logs. In: Proceedings of IEEE Pacific Rim Conference on Communications, Computers and Signal Processing, Victoria, B.C., Canada (2005)
6. Wang, Z., Banerjee, S., Jamin, S.: Studying Streaming Video Quality: From An Application Point of View. In: Proceedings of ACM MultiMedia Conferenc 2003, Berkeley, CA (2003)