TD(07)021

The 9th COST 290 MC Meeting

University of Coimbra, Coimbra, Portugal, May 22-23, 2007

Cross-Layer Architecture for H.264 Video Streaming in Heterogeneous DiffServ Networks

Gabriel Lazar, Virgil Dobrota, Tudor Blaga

Technical University of Cluj-Napoca, Communications Department
26-28 Baritiu Street, 400027 Cluj-Napoca, Romania
Tel/Fax: +40-264-597083
E-mail: virgil.dobrota@com.utcluj.ro

Abstract

High quality video streaming over heterogeneous network segments comprising both wired and wireless links is possible only if every node and every protocol layer add their contribution to this purpose. The paper describes the concept of a reliable multimedia network and it highlights the established or emerging technologies that can ensure robust communication. It also describes a practical implementation of the required QoS mechanisms. In the access WLAN, a cross-layer based solution enables QoS support for H.264 video streams at both Data Link and Network Layers. In the core network, a new DiffServ per-hop behavior (PHB) suitable for real-time traffic ensures selective dropping of packets based on their importance within the video stream.

Keywords

cross-layer optimization, H.264, IEEE 802.11e, quality of service, video streaming

Chapter #2 subtopic Cross-Layer

Cross-Layer Architecture for H.264 Video Streaming in Heterogeneous DiffServ Networks

Gabriel Lazar, Virgil Dobrota, Tudor Blaga

Abstract

High quality video streaming over heterogeneous network segments comprising both wired and wireless links is possible only if every node and every protocol layer add their contribution to this purpose. The paper describes the concept of a reliable multimedia network and it highlights the established or emerging technologies that can ensure robust communication. It also describes a practical implementation of the required QoS mechanisms. In the access WLAN, a cross-layer based solution enables QoS support for H.264 video streams at both Data Link and Network Layers. In the core network, a new DiffServ per-hop behavior (PHB) suitable for real-time traffic ensures selective dropping of packets based on their importance within the video stream.

1. INTRODUCTION

Recent years have seen a proliferation of real-time multimedia traffic over an increasingly heterogeneous Internet. A common case is that of video streaming over hybrid wireless-wired IP networks. The low-cost solution of over-provisioning simply cannot be used reliably in this situation, because real-time traffic has very stringent delay, jitter and bandwidth requirements. These parameters cannot be guaranteed when data is handled in a best effort manner; moreover, wireless environments have much more limited resources (e.g. bandwidth) due to channel characteristics, making over-provisioning a prohibitive solution. On top of that, users expect high service quality independent to the underlying network access technologies [1]. Therefore, real-time media streaming while maintaining a high level of perceived quality for the entire duration of the stream transmission becomes a challenging task. It requires cooperation of all involved "actors": all networks, end nodes and each layer of the protocols stack must collaborate and optimize functionality to achieve this goal. A number of technologies have already been designed to increase the quality of media streaming over the Internet:

a. Latest media technologies improving compression efficiency and error robustness of video/audio data;

b. Scalable, packet-based, QoS frameworks such as Differentiated Services (DiffServ) and traffic control blocks (resource reservation, resource allocation, shaping, admission control, routing, policing, classification and marking, queue management, scheduling);

c. 54 Mbps - IEEE 802.11g and 100 Mbps - IEEE 802.11n for new high-speed wireless networks, IEEE 802.11e for QoS-based MAC layer.

The strong demands imposed on video codecs and the wireless links used to transport media streams gave birth to a new paradigm in network architecture design: the cross-layer design

(CLD). Wireless video streaming must ensure that variable rate data is delivered at the destination despite changing conditions, while maintaining a high user perceived quality. Applying cross-layer optimization to multiple layers (application, network, data link, physical) allows for optimal adaptation of the network [2].

In order to achieve a reliable end-to-end video communication system, the best-effort network should be replaced by a QoS-enabled core network, augmented where necessary with a cross-layer solution.

In this article we propose a practical implementation of an end-to-end H.264 video streaming solution over heterogeneous QoS-enabled networks using robust cross-layer architecture based on application, network and MAC layers. In addition, a new DiffServ PHB is presented, suitable for real-time traffic packets having different drop precedence values.

2. RELIABLE MULTIMEDIA NETWORK

A high-speed, reliable multimedia service can be obtained only if the following conditions are satisfied:

1. *The network must adapt to the application requirements if possible.* As a consequence, the underlying network must be QoS-enabled. Each node is "intelligent", so it can recognize multimedia streams, allocate necessary resources and set the required priority for audio/video packets. As elastic and multimedia applications generate traffic of a different nature, only a QoS network can differentiate the flows and is able to treat each packet according to its service requirements, to obtain a high end-to-end performance.

2. *When first condition cannot be applied, (i.e. in network segments with insufficient physical resources, or not capable to offer QoS guarantees), the application must adapt to the network.* As a consequence, the application must be able to adapt to changes in network characteristics by proper processing, ensuring good perceptual quality for the user and graceful quality degradation in adverse conditions (Fig. 1).

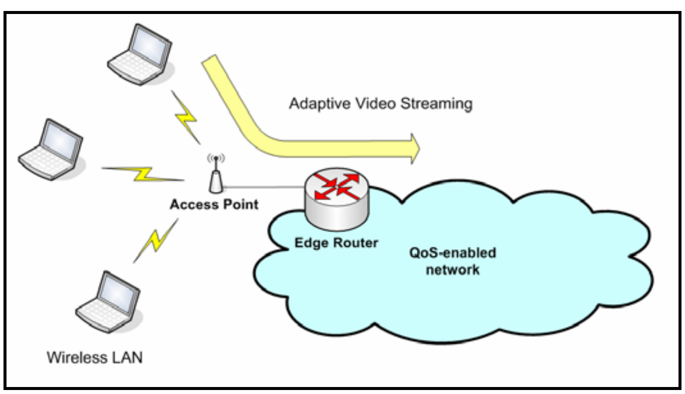

Fig. 1. Heterogeneous multimedia network

An implementation of the high-quality video network depicted above can be obtained by starting with existing, representative technologies and by employing certain modifications or additions relevant for real-time multimedia streaming.

3. QOS-ENABLED NETWORK – DIFFSERV

Inside the core network, Differentiated Services architecture (DiffServ, [3]) can provide class-based QoS guarantees in a simple and scalable manner. Complex decisions are made in the edge routers, and useful edge-to-edge services are built from a set of per-hop behaviors (PHBs) in the core routers. A PHB defines the class of traffic to which a packet belongs, i.e. the treatment of this datagram. This class is recognized by the core routers based on the DiffServ Code Point (DSCP) value, which is stored using the first 6 bits of the old Type-of-Service (TOS) IP header field. The DSCP value is usually set by the edge routers through multi-field (MF) classification and marking, but the bits can be also configured by the source client, e.g. at the application layer.
Most common PHBs implemented in existing networks are:

- the Expedited Forwarding (EF) PHB for low loss, low delay traffic, suitable for real-time multimedia services. Traffic pertaining to this class is using in each node a separate queue with highest priority and fast service rate. To avoid starvation of other flow aggregates from queues having lower service priority, the EF traffic must be limited (policed) at the ingress node.

- the Assured Forwarding (AF) PHB Group, which defines four AF classes of traffic and three possible drop precedence values; the combination yields twelve separate DSCP encodings. Although the PHB definitions do not specify the mechanisms to implement a specific PHB, in practice a flow aggregate pertaining to a specific class goes to a separate router queue, while packets from the same traffic class, with different drop precedence, are discarded according to this value using a form of *random early detection* (RED) [4].

A Voice over IP (VoIP) software application that marks its audio packets with the DSCP value corresponding to the EF PHB, or an edge router that uses the transport port value of the incoming packet to recognize a multimedia flow (and mark it as such) are both performing a basic form of cross-layer optimization: in each case, information from an upper layer is used by the network layer for an appropriate QoS mapping.

4. SERVICE DIFFERENTIATION OVER WLAN

High-speed multimedia communication over wireless links is made possible through the development of the IEEE 802.11g (54 Mbps) and 802.11n (100 Mbps) physical layers. In addition, audio/video data can benefit from the new IEEE 802.11e QoS-based MAC layer to receive a suitable, real-time treatment at the Data Link Layer [5]. The most used access channel mechanism in 802.11 WLANs is the distributed coordination function (DCF). Under DCF, all stations compete to channel access with the same priority, using carrier sense multiple access with collision avoidance (CSMA/CA) as the access method. Prior to any transmission, each station is required to sense the medium and, if the medium is busy, the station must defer its transmission and initiate a backoff timer. Its value is obtained as a random number selected between 0 and the contention window (CW). When the station senses that the medium has been free for the duration of DCF interframe spaces (DIFS), it begins to decrement the backoff timer. The process continues as long as the communication channel is free. If the timer expires and the channel is still idle, the station initiates the transmission. In case of collision, the size of CW is doubled until it reaches a maximum value $CW_{\max}$. If the transmission is successful, CW is reset at the minimum value $CW_{\min}$. The need to provide better, preferential service for multimedia applications has led to the standardization of new 802.11e MAC layer operation; DCF is enhanced through the use of the hybrid coordination function (HCF) and its successor access method becomes the enhanced DCF channel access (EDCA). The alternative, pooling-based, access mechanism is HCF controlled channel access (HCCA).

Service differentiation in EDCA is obtained using access categories (ACs). Traffic assigned to a category is sent to a separate transmission queue and each AC has different channel access parameters: $CW_{\min}$, $CW_{\max}$, arbitrary inter-frame space (AIFS) and transmission opportunity duration limit ($TXOP_{\lim}$). The latter represents the maximum value of an interval used to choose a transmit opportunity time (TXOP) during which a station can send as long as many frames as possible. Traffic from a category with smaller $CW_{\min}$, $CW_{\max}$ or AIFS will wait less, on average, before being sent than traffic from a category where parameters have bigger values.

The standard defines four access categories, from AC0 (lowest priority) to AC3 (highest priority). AC3 and AC2 are usually reserved for multimedia transmission, while AC1 and AC0 should receive best effort and background traffic.

5. REAL-TIME MULTIMEDIA STREAMING

One of the key factors to improve video streaming over networks segments with limited bandwidth (such as wireless links) is employing efficient, error resilient video compression mechanisms. In this respect, the H.264/AVC video standard has achieved a significant improvement relative to existing standards through enhanced compression performance and provision for “network friendly” error resilience tools.

To address the need for flexibility, H.264 specifies a video coding layer (VCL), designed for an efficient representation of the video content. It also defines a network abstraction layer (NAL), which is responsible for the encapsulation of video data into entities suitable for a variety of transport layers or storage media.

A NAL unit (NALU) consists of a one-byte header followed by a bit string that contains fixed sized picture partitions called macro blocks (MBs). An important field in the NALU header is the Nal_Ref_Idc (NRI) field. NRI consists of 2 bits that specify the priority of the payload: from 00 (lowest priority) to 11 (highest priority). The values are set by the encoder and they effectively specify the importance of the video information contained in each NALU payload. This is accomplished through an error resilience technique called data partitioning (DP) that groups video MBs into partitions with similar importance [6].

Three partitions types are defined by H.264:

- Partition A – contains header information (MB types, quantization parameters and motion vectors) relevant for other partitions, which makes this information very important;
- Partition B (intra partition) – contains intra coded block patterns (CBPs) and intra coefficients. Requires the A partition information and it is more important than partition C;
- Partition C (inter partition) – contains intra CBPs and intra coefficients. Requires the A partition information and it is the least important partition type.

In addition to these three partitions, VCL also generates a slice representing the instantaneous decoding refresh (IDR) pictures, which contains information that cannot be included into the A, B and C partitions. The parameter set concept (PSC) is the most important slice type created by VCL, where data is a combination of higher level parameters containing important information relevant for more than one slice (e.g. picture dimension) [7].

6. CROSS-LAYER ARCHITECTURE

Traditional approaches in wireless video transmission rely on application-level QoS control in order to increase the reliability of the streaming. Because the mechanisms employed do not interact with the lower layers transmission parameters or statistics, the transmission cannot adapt to variable channel conditions in an efficient manner.

In [7], the authors propose cross-layer architecture (CLA) for robust video transmission over 802.11e MAC layer using H.264, as an alternative to the above mechanisms. The solution falls into the MAC-centric CLA category, in which the application layer passes relevant information to the MAC layer. The MAC will decide how to prioritize resources for each packet based on its QoS level and other upper layer information. Through the NRI field value set by the H.264 encoder, each NALU containing bits from a specific partition (PSC, IDR, A, B or C) that arrives at the MAC layer is analyzed and the data frame that encapsulates the unit is mapped into an access category in the range AC1-AC3 that corresponds to its importance (e.g. PSC is mapped to AC3, highest priority class). AC0 is used for best-effort traffic.

Our CLA solution extends the above work by additionally taking into consideration the network layer in order to propagate the video related QoS information to the other end of communication (Fig. 2). NRI information extracted from the NALU header is mapped to DSCP field values at the IP network layer and to AC at the MAC layer.

Including the network layer into the architecture is necessary because transmission based on MAC-only CLA has effect at the source node only. In a WLAN, this is not relevant, because the second node is the destination , but in any network containing IP nodes, packets traveling through subsequent routers do not preserve their priority information and they will be consequently treated as best-effort traffic. In addition, due to the variable rate of video traffic caused by the dynamic nature of the captured scene and encoding, the congestion can take place in any intermediate node on the route to destination, not only in the wireless access network.

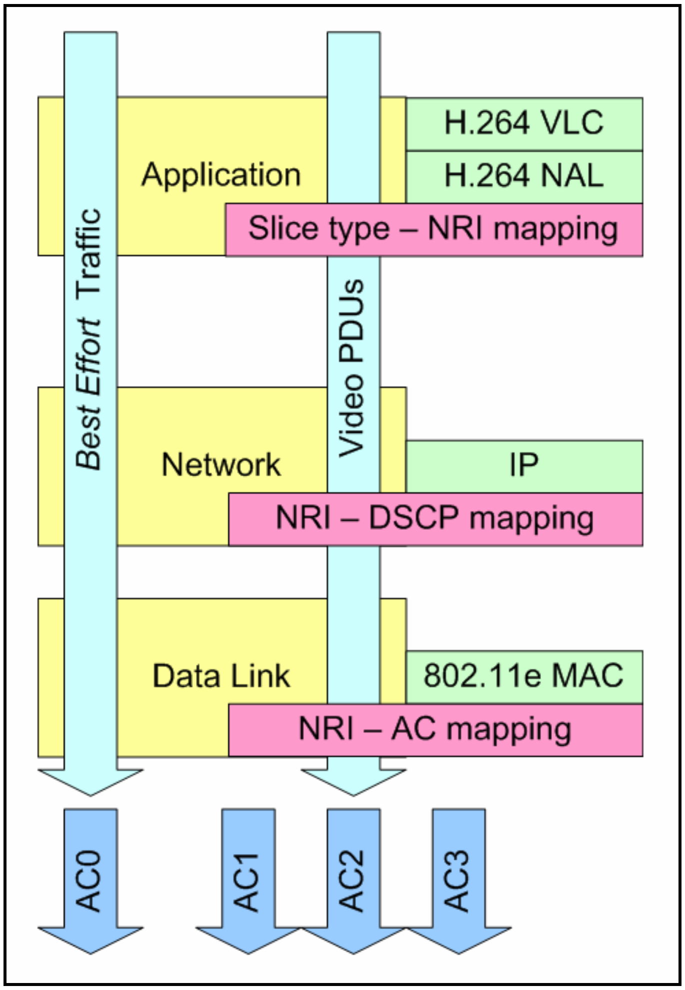

Fig.2. Cross-layer architecture for heterogeneous networks (source node)

7. A DIFFSERV PHB FOR MULTIMEDIA TRAFFIC

The following paragraphs illustrate the shortcomings of the current PHBs with respect to video traffic and propose a new real-time PHB suitable for multimedia streams with multiple drop priorities.

Video traffic has variable data rate due to the dynamic nature of the captured scene and the encoding process. Assigning an EF behavior to video streams in the DiffServ network can pose a policy design problem: because video streams have variable rates and due to the fact that multiple video streams are placed into the same flow aggregate (FA), it is hard to design a maximum limit for traffic policing at the ingress router of a DiffServ domain. If excessive EF traffic enters into the DiffServ domain, the EF PHBs from the core routers can cause starvation of other flow aggregates, by serving continuously EF packets at the highest priority. EF PHBs do not employ large queues, because EF traffic should be served at a rate equal or higher than the incoming rate, and also because waiting in a queue increases delay, which is not desired for real-time traffic. Excessive EF traffic in the core network will therefore lead very fast to full queues and large packet drops. In this case, there is no protection against elimination of important packets. In a similar manner, excessive EF traffic at the border of the DiffServ domain will be policed regardless of the packet importance in the video stream [8].

Using an AF DiffServ class for multimedia traffic raises a different problem. Drop priorities for AF PHBs are usually implemented with a form of RED. Due to the mechanism's statistical approach, in some cases important multimedia packets can be discarded (with lower probability) instead of less important ones. In another words, the priority dropping is not strict [9]. In addition, AF traffic has lower priority than EF traffic.

In order to overcome these limitations, we propose a different PHB for multimedia traffic with drop priorities, called Multimedia (MM) PHB. This class of traffic uses 3 DS code points in order to recognize packets with different importance in the stream. The MM PHB is defined for high-priority, low loss, latency and jitter traffic similar to EF, but additionally employs a strict drop precedence scheme: when resources become insufficient for the MM traffic, the DiffServ core router will always eliminate a packet with highest drop precedence from its queue. In this way, important packets have better chances to survive the end-to-end journey. A MM PHB can be implemented in a straightforward manner using a FIFO queue with strict priority drop policy (Fig. 3).

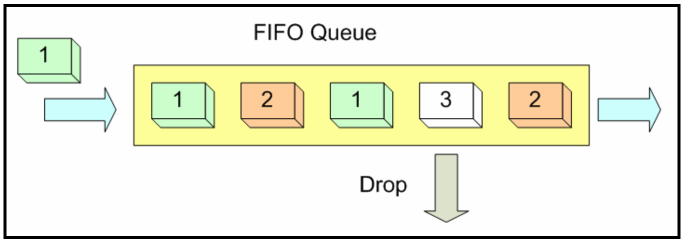

Fig.3. MM PHB Implementation:
1=lowest drop priority, 3=highest drop priority

Policing ingress MM traffic at the edge router can benefit of the 3 DSCP values to selectively drop less important packets first. The same DS code points can also be used at the data layer in each node to prioritize important traffic, if the MAC protocol offers QoS support.

8. IMPLEMENTATION AND PRELIMINARY RESULTS

The CLA at the source node was implemented using a combination of open source software:

- Application Layer: a modified Linux build of VideoLan Client (VLC) [10], capable of H.264 video streaming, analyses each NALU and sets the socket's SO_PRIORITY value for the current packet according to the NRI field.
- Network Layer: Part of the Linux traffic control mechanisms, DSMARK is in charge of marking packets by setting their DSCP at the IP level [11]. A DSMARK queuing discipline was configured at the source node to translate the SO_PRIORITY value to a DS code point. In this way, the NRI value is mapped to a DSCP through the SO_PRIORITY socket parameter sent between layers.

- Data Link Layer: The MadWifi WLAN Linux driver for Atheros chipsets can use TOS or DSCP bits to select the required MAC access class [12]. The driver was modified to implement DSCP to AC mapping as presented in section 6.

In order to test the importance of policing based on drop priorities, a DiffServ edge router was configured using Linux QoS mechanisms to police traffic according to the DSCP value of each packet. Traffic conditioning was implemented with four policy filters combined with a DSMARK queue discipline. Out-of-profile traffic from a class is re-marked and sent to the lower priority class. The fourth filter is associated with best-effort traffic and out-of-profile packets are discarded.

Preliminary tests were performed using the CLA-enabled client as traffic source, a DiffServ edge router for traffic policing and a fast receiver. A H.264 encoded Foreman sequence was sent to the destination through the ingress router. The total rate was limited to 1.1 Mbps to enforce re-classification and dropping. The edge router discarded excess packets according to their DSCP set by the video source. A second experiment was performed with similar setup but, instead of priority dropping, packets were discarded randomly (Fig. 4).

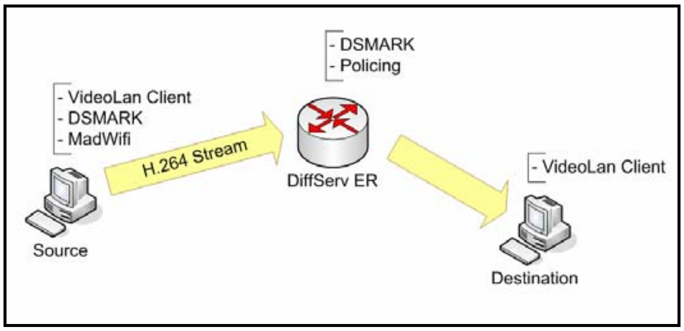

Fig.4. Testbed with CLA client and DiffServ ER

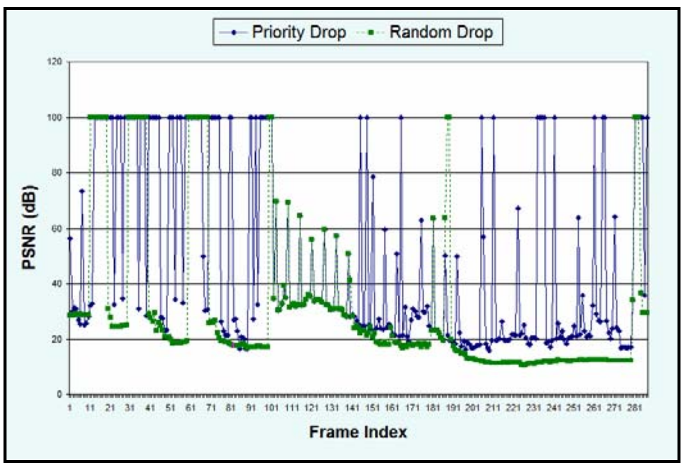

Fig.5. Priority Drop vs. Random Drop at the Edge Router

Image quality was compared using peak SNR (PSNR) for each video frame, and the average PSNR (APSNR) was computed for both video sequences received at the destination node, relative to the source Foreman sequence (Fig. 5). The APSNR value obtained for the first experiment was 48.72 dB, higher than 31.13 dB obtained in the second (random drop) case, which shows the benefic effects of DSCP based policing. Further experiments will include testing of the MM PHB-based Linux implementation and end-to-end video streaming quality measurements through a DiffServ network using CLA-enabled client nodes, wireless access links and MM PHBs in core routers.

9. CONCLUSION

In this article we introduced the concept of a reliable multimedia network, based on established or emerging technologies such as DiffServ or CL design that aims to ensure robust communication for real-time video traffic over heterogeneous network segments. A CLA-based solution at the source node is described, that takes advantage of the new IEEE 802.11e QoS mechanisms, and also preserves the QoS information in the transmitted packets beyond the first node. In the core DiffServ network, a new multimedia PHB is presented together with a simple implementation. Finally, Linux experiments show that ingress video traffic conditioning achieves better results when packets are discarded based on their priorities compared with the random drop case.

References

[1] J.Soldatos, E.Vayias and G. Kormenzatas, "On the Building Blocks of Quality of Service in Heterogeneous IP Networks", *IEEE Communication Surveys & Tutorials*, vol. 7, issue 1, pp. 69-88, First Quarter 2005.

[2] S.Khan, Y.Peng, E.Steinbach, M. Sgroi and W.Keller, "Application-Driven Cross-Layer Optimization for Video Streaming over Wireless Networks", *IEEE Communication Magazine*, vol. 44, no.1, pp. 122-130, January 2006

[3] S.Blake, D.Black, M.Carlson, E.Davies, Z.Wang and W.Weiss, "An Architecture for Differentiated Services", RFC 2475, December 1998.

[4] S. Floyd and V. Jacobson, "Random Early Detection gateways for Congestion Avoidance" *IEEE/ACM Transactions of Networking*, vol. 1, no. 4, pp. 397-413, August 1993.

[5] IEEE 802.11e, "Wireless LAN Medium Access Control (MAC) Enhancements for Quality of Service (QoS)", 802.11e Standard, 2005

[6] ITU-T Recommendation H.264, "Advanced Video Coding for Generic Audiovisual Services", H.264 Standard, 2005.

[7] A. Ksentini, M. Naimi and A. Gueroui, "Toward an Improvement of H.264 Video Transmission over 802.11e through a Cross-Layer Architecture", *IEEE Communication Magazine*, vol. 44, no.1, pp. 107-114, January 2006

[8] B. Davie, A. Charny, J.C.R. Bennet, K. Benson, J.Y. Le Boudec, W. Courtney, S. Davari, V. Firoiu, D. Stiliadis, "An Expedited Forwarding PHB (Per-Hop Behavior)", Internet proposed standard RFC 3246, March 2002.

[9] J. Heinanen, F. Baker, W. Weiss, J. Wroclawski, "Assured Forwarding PHB Group", Internet proposed standard RFC 2597, June 1999

[10] http://www.videolan.org

[11] http://diffserv.sourceforge.net

[12] http://madwifi.org