

Efficiency Issues in MPLS Transport for the UMTS Access Network

Enrique Vázquez, Manuel Álvarez-Campana, and Ana B. García

Dept. of Telematics, Technical University of Madrid,
ETS Ing. Telecomunicación, Ciudad Universitaria, 28040 Madrid, Spain
{enrique, mac, abgarcia}@dit.upm.es

Abstract. Multiprotocol Label Switching (MPLS) offers a simple and flexible transport solution for multiservice networks. Therefore, many UMTS operators are currently considering the use of a MPLS backbone. However, the efficient transport of short voice and data packets in the UMTS access network requires multiplexing and segmentation functions not provided by MPLS. This paper investigates the use of ATM Adaptation Layer 2 (AAL2) over MPLS to perform both functions. The efficiency of AAL2/MPLS is analyzed for different traffic types and compared with other transport options. The results indicate that significant capacity savings can be obtained with this solution.

1 Introduction

The 3G Universal Mobile Telecommunications System (UMTS) is a network designed to support multiple applications (telephony, video-conferencing, audio and video streaming, games, Web access, e-mail, etc.) with very different traffic patterns and quality of service (QoS) requirements. The initial UMTS architecture defined by the Third Generation Partnership Project (3GPP) comprises a Wideband CDMA radio interface, an access network based on Asynchronous Transfer Mode (ATM), and a core network evolved from 2G networks. The core has two domains. The circuit-switched domain, founded on GSM, handles voice and other circuit-mode traffic. The packet-switched domain, derived from GPRS, handles IP traffic. This structure, described in the 3GPP Release 99 specifications, is progressively changing towards a unified model based on IP (Releases 4, 5, and 6).

In contrast with the core network, the UMTS Terrestrial Radio Access Network (UTRAN) relies on an integrated transport infrastructure for all traffic types (voice, data, etc.). The UTRAN protocol architecture is divided into a Radio Network Layer (RNL), designed specifically for UMTS, and a Transport Network Layer (TNL) that reuses existing transport technologies. In Releases 99 and 4 the TNL consists of ATM connections with ATM Adaptation Layer type 2 (AAL2) or 5 (AAL5). See Table 1.

Release 5 specifications [1], [2] allow two TNL alternatives: ATM, as outlined above, or IP with UDP in the user plane, and SCTP (Stream Control Transmission Protocol) for signaling. IP version 6 is mandatory and IP version 4 is optional, although a dual stack is recommended. The default encapsulation for IP packets is PPP

Table 1. UTRAN interfaces and ATM Adaptation Layers used. The air interface (Uu) does not use ATM

Interface	User Plane	Control Plane
Iub: Node B – Radio Network Controller (RNC)	AAL2	AAL5
Iur: Serving RNC – Drift RNC	AAL2	AAL5
Iu-CS: S-RNC – Mobile Switching Center (MSC)	AAL2	AAL5
Iu-PS: S-RNC – Serving GPRS Support Node (SGSN)	AAL5	AAL5

(Point-to-Point Protocol) with HDLC framing, but other layer 2 protocols are also possible.

A typical UTRAN topology will have many Node Bs, often connected by low-capacity links. Therefore, the efficiency of the protocol stack in the Iub interface between Node B and RNC becomes an important design issue. Two main problems must be solved. Firstly, UDP, IP, and layer 2 headers add a large overhead to voice packets, even if header compression procedures are used. This problem can be alleviated by multiplexing voice communications. In this way, several short voice packets are concatenated before transport, so the overhead of the headers added after the concatenation is shared by all the packets in the group. Secondly, long packets may introduce unacceptable delay variations when transmitted over a slow link. To avoid this, packets that exceed a certain maximum size are segmented before transmission. Note that segmentation reduces efficiency, but it may be necessary to meet QoS requirements (delay variation).

The structure of the rest of this paper is as follows. Sect. 2 reviews various solutions proposed to implement concatenation and segmentation in the Iub protocol stack. Sect. 3 presents a new proposal that relies on AAL2 [3] for packet concatenation/segmentation, and MPLS [4] for transport. Sect. 4 evaluates the performance of AAL2/MPLS for voice and data, and compares it with other options. Finally, Sect. 5 summarizes the contributions of the paper.

2 Alternatives for Transport in the UTRAN

ATM in the UTRAN is studied in [5] and [6], focusing on traffic modeling, QoS analysis, and simulation experiments with voice and data traffic. The 3GPP has investigated IP transport in the UTRAN in an ad hoc group created within the Technical Specification Group Radio Access Network, Working Group 3 (TSG RAN WG3), which is responsible for the overall UTRAN design. A technical report [7] presents the conclusions of this work and makes some general recommendations for Release 5 specifications.

Different Iub stacks based on IP are simulated and compared in [8]. The concatenation and segmentation functions mentioned in the previous section may be located above UDP/IP or below it. Protocols such as Composite IP (CIP) and Lightweight IP Encapsulation (LIPE) [9] concatenate and, optionally, segment packets above UDP/IP. Alternatively, both functions may be implemented at layer 2 by using

PPPMux [10] for concatenation, and the multi-class, multi-link extensions to PPP (MC/ML-PPP) [11] for segmentation. This is the option recommended by 3GPP in [7]. The application of the Differentiated Services model in the UTRAN is addressed in several studies. For example, in [12] voice calls are multiplexed with CIP and transferred with the Expedited Forwarding Per-Hop Behavior [13].

While these studies indicate that IP can be a viable alternative to ATM for the UTRAN, it is necessary to implement a complicated protocol stack to achieve the required performance. Firstly, UDP/IP headers have to be compressed to reduce the overhead added to short packets (e.g. voice). Many compression algorithms have been proposed, for example IP Header Compression, Compressed RTP, Enhanced Compressed RTP, Robust Header Compression, and others described in RFCs or Internet drafts. For references and a discussion of the advantages and drawbacks of each method see [7]. PPP encapsulation and framing overhead must also be minimized by using the simplified header formats foreseen in the standard.

Secondly, PPPmux and MC/ML-PPP must be implemented. When there are intermediate IP routers between Node B and RNC, PPPmux and MC/ML-PPP may be terminated at the edge router (ER) next to the Node B or at the RNC. In the former case, header compression, concatenation, and segmentation are applied only in the last hop between the Node B and the ER, where capacity is likely to be more limited. Packets are routed individually between the ER and the RNC. In the latter case, those functions are moved to the RNC, so the efficiency gains are maintained all the way up to the RNC. The disadvantage of this approach is that layer 2 frames must be tunneled between the Node B and the RNC. This implies that another IP header is added to the information transported across the network. See Fig. 1.

The Radio Network Layer (RNL) includes other protocols not shown in Fig. 1, namely Medium Access Control, Radio Link Control, and Packet Data Convergence Protocol. Non-access stratum protocols, which are transparent for the UTRAN, are located on top of the RNL, for example IP packets exchanged between the user equipment and the Gateway GPRS Support Node (GGSN) in the UMTS core network, plus the higher layer protocols required by the user applications (TCP, HTTP,

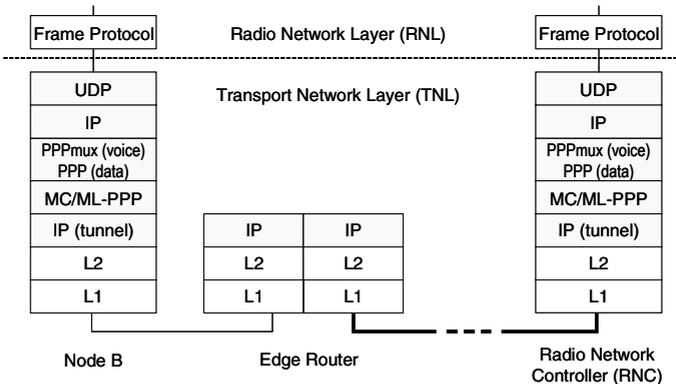


Fig. 1. UDP/IP transport stack in the user plane of the Iub interface, with end-to-end concatenation/segmentation between Node B and RNC

etc.). With the architecture of Fig. 1 these IP packets would be encapsulated with two additional IP headers, resulting in a poor efficiency.

The use of MPLS in the UTRAN has been considered mainly in terms of protocol functionality, for example: mapping of traffic flows to MPLS Label Switched Paths (LSPs) and signaling for LSP establishment [14], combination of MPLS traffic engineering (MPLS-TE) functions with the Differentiated Services model (DiffServ) [15], and mobility support [16], [17]. These references point out potential advantages of MPLS for transport in the UTRAN, but no numerical results are presented.

The MPLS/Frame Relay Alliance (MFA) defined an ad-hoc multiplexing protocol for efficient transport of voice calls over MPLS [18]. More recently, the MFA has proposed to use AAL protocols (without ATM cells) for voice trunking [19] and for TDM circuit emulation [20] over MPLS networks.

3 AAL2/MPLS Transport in the UTRAN

This section proposes a simpler TNL architecture with two main components: AAL2 and MPLS. See Fig. 2. Specifically, we propose to use the Service Specific Segmentation and Reassembly (SSSAR) sublayer [21], which is more adequate for integrated voice and data transport, instead of the convergence sublayer for trunking [22] included in previous proposals that focused on voice traffic only.

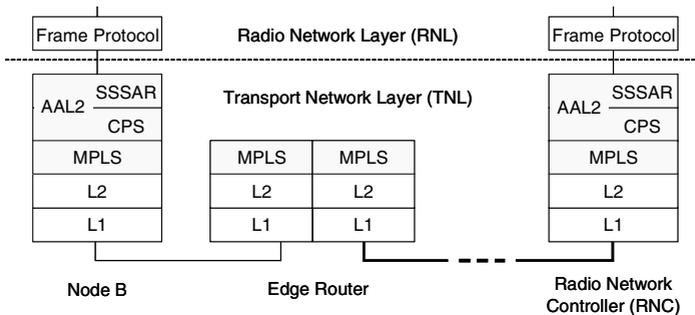


Fig. 2. AAL2/MPLS transport protocol stack in the user plane of the Iub interface

Our solution, explained in more detail below, combines standard protocols in such a way that each one performs the essential function it was designed for, and complements the other. AAL2 is used to multiplex many variable-bit-rate, delay-sensitive traffic flows, and MPLS provides a flexible tunnelling mechanism without the overhead of ATM. The result is a simple protocol architecture that offers significant advantages in comparison with other solutions (see Sect. 3.1.)

The Common Part Sublayer (CPS) [3] of AAL2 concatenates voice and data packets. Each CPS packet has a 3-byte header and a maximum payload length of 45 (default) or 64 bytes. See Fig. 3. The SSSAR sublayer accepts data units up to 65568 bytes and segments them up to the maximum length admitted by CPS. The last segment of each packet is marked in the UII bits of the CPS header, so the SSSAR adds

4 Performance Evaluation

The header formats and procedures of the different protocols mentioned in the preceding sections have been analyzed in order to evaluate the efficiency of alternative transport stacks for the UTRAN. The analysis considers a simple case where every concatenated packet has the same length. The results presented in Sect. 4.1 and 4.2 show the improvements that can be obtained with the AAL2/MPLS solution.

A higher transport efficiency should allow the network to carry more traffic maintaining the QoS levels or, alternatively, to reduce the capacity needed to carry a given amount of traffic with the required QoS. To verify this assumption, we have simulated in more detail each protocol stack, transporting various mixes of voice and Web data traffic in the Iub interface between base stations (Node Bs) and controllers (RNCs). Optimizing this interface is particularly important for the operators due to the high number of Node Bs that must be connected in a typical UMTS network, so we focused the simulation study on it. However, the solution proposed here can be used in other UMTS interfaces as well. Sect. 4.3 gives a sample of the simulation results obtained.

4.1 Analysis of Efficiency for Voice Traffic

In this case, we analyze the transport of voice payloads of constant size $P=40$ bytes, corresponding to 32 bytes generated by the Adaptive Multi Rate codec used in UMTS, plus 8 bytes added by the RNL [5]. The background noise description packets sent by the AMR codec during silence periods are not considered in the analysis. The number of concatenated packets per group (N) is equal to the number of voice connections in the active state. Segmentation is not necessary.

The overhead per packet (H) and the overhead per group (H_g) take different values depending on the transport protocol stacks considered. For example, AAL2/MPLS over PPP with simplified HDLC framing (AAL2/MPLS/PPP/HDLC) gives $H=3$ bytes and $H_g=9$ bytes. UDP/IP with headers compressed to 4 bytes and concatenation at layer 2 (cUDPIP/PPPMux/HDLC) gives $H=H_g=5$ bytes. The details of the model and the complete set of parameter values used can be found in [24].

Fig. 4 shows the efficiency (number of RNL bytes transported divided by the total number of bytes transmitted by the TNL) as a function of the number of active voice connections. The best alternative is AAL2/MPLS, with efficiency above 90% even for moderate values of N . The efficiency of UDP/IP with headers compressed to 4 bytes (cUDPIP) over PPPmux is not as good, because concatenation at layer 2 gives a higher overhead per packet (H). If the end-to-end configuration between Node B and RNC illustrated in Fig. 1 is used, the extra IP tunnel increases the overhead per group (H_g) and reduces the efficiency, especially when there are few packets per group.

The curve labeled MPLSx2/PPP/HDLC corresponds to the case where AAL2 is not used and each packet is transmitted with a stack two MPLS labels: the inner label is used as a channel identifier, and the outer one serves to route the packets to their destination. The two labels (4 bytes each) plus the PPP/HDLC header add a total of 13 bytes to each voice payload of 40 bytes. Therefore, the efficiency is $40/53=75.5\%$,

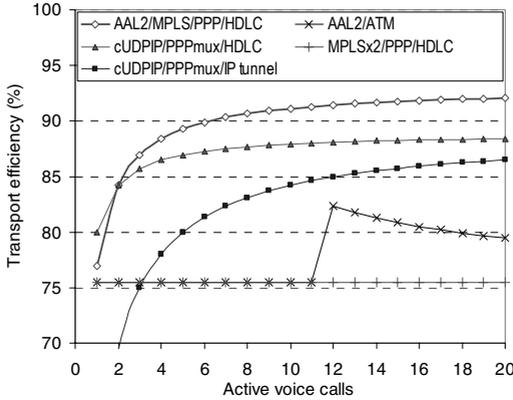


Fig. 4. IP and MPLS transport options vs. AAL2/ATM (voice traffic)

which happens to be the same value obtained with AAL2/ATM when N voice payloads ($N \cdot 40$ bytes) are transmitted in N ATM cells ($N \cdot 53$ bytes). This is true for $N < 12$. When $N = 12$, 12 voice payloads are carried in 12 concatenated CPS packets ($12 \cdot 43 = 516$ bytes), which still fit in 11 cells ($11 \cdot 47 = 517$ bytes), so the efficiency of AAL2/ATM increases to $12 \cdot 40 / 11 \cdot 53 = 82.3\%$ as illustrated in Fig. 4.

Other transport options not shown in the graph give poor results. For example, the efficiency of a simple UDP/IP/PPP/HDLC stack without header compression and no multiplexing is only 55.6%. This value corresponds to IP version 4 headers (20 bytes). With IP version 6 headers (40 bytes) it would be even lower: 43.5%. If IP is sent over AAL5/ATM instead of PPP/HDLC, each voice packet is encapsulated in 2 cells and the efficiency drops further to 37.7%.

4.2 Analysis of Efficiency for Data Traffic

Fig. 5 shows the transport efficiency for data payloads of variable size P up to 1500 bytes. As explained in previous sections, short packets may be concatenated up to a maximum size L_{max} , and packets longer than L_{max} are segmented before transmission. In Fig. 5 L_{max} has been set to 1000 bytes. H is the overhead per packet or segment and H_g is the overhead per group, with different values for each protocol stack as in the previous case.

With compressed UDP/IP (cUDPIP) and MPLS using a 2-label stack (MPLSx2), segmentation is done at layer 2 by MC/ML-PPP. Both exhibit a high efficiency for long data packets, although they were not as good for short voice packets (compare the corresponding curves in Fig. 4 and Fig. 5). In MPLSx2, the internal label identifies each data channel, and the external one is used for routing. AAL2/MPLS reaches approximately the same value, close to 95%, for both voice and data. In this case, MC/ML-PPP is not needed because AAL2 takes care of segmentation at the SSSAR sublayer, so AAL2/MPLS uses the default PPP/HDLC encapsulation.

Fig. 5 indicates that MPLSx2 outperforms AAL2/MPLS for packets longer than 250 bytes approximately. Therefore, in scenarios where the traffic mix includes sig-

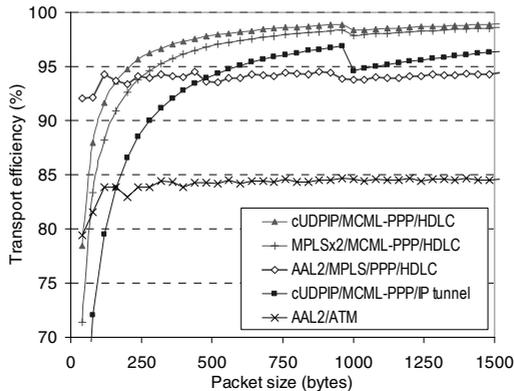


Fig. 5. IP and MPLS transport options vs. AAL2/ATM (data traffic)

nificant amounts of short voice packets (a few tens of bytes) and long data packets (a few hundreds of bytes or more), the overall efficiency can be optimized by separating the traffic of each Node B in two Label Switched Paths: LSP 1 for voice with AAL2, and LSP 2 for data without it. In this case, and assuming that voice packets are always smaller than 65 bytes, the SSSAR part of AAL2 can be removed in LSP 1. Packets sent via LSP 2 are segmented, if necessary, by Multi-Class Multi-Link PPP. Note that the network operator may prefer to use separate LSPs for voice and for data anyway, even if AAL2 is used in both.

4.3 Simulation Results

The Iub transport protocol stacks considered in the preceding sections have been simulated with voice and Web traffic, in order to compare the delay and loss performance of the different options. The simulator was developed in C language, and we are currently porting it to OPNET Modeler [25].

The traffic source modules generate Frame Protocol data units (see Figs. 1 and 2). The data unit sizes and inter arrival times are set taking into account the relevant characteristics of the UMTS radio access bearers used, as well as the overheads added by the RNL protocols. The transmission rate and the Transmission Timing Interval (TTI) are the most decisive parameters. In our simulations, voice is coded at 12.2 kbit/s with TTI=20 ms. Including the RNL overhead, this corresponds to one FP data unit of 40 bytes every 20 ms. During silence periods the data unit size is reduced to 13 bytes. Web pages are downloaded at 64 kbit/s with TTI=40 ms, which corresponds to one data unit of 331 bytes every 40 ms, also including RNL overhead. For AAL2/MPLS, the simulator can be configured to use the same LSP for all traffic, or separate LSPs for voice and for data. In these experiments we chose the latter option.

Fig. 6 is a sample of the results obtained. The curves show the Iub delay (0.95- or 0.99-quantile) vs. the number of active users. The capacity available for user traffic is set to 1.92 Mbit/s, and delays are measured in the RNC-to-Node B direction (Web traffic is higher in this direction). As anticipated by the previous analysis of effi-

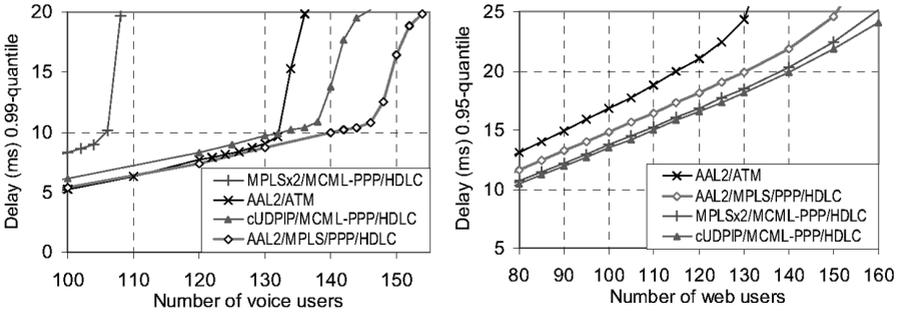


Fig. 6. Simulation of delay at the Iub Interface for different transport options

ciency, AAL2/MPLS gives the lowest delay for voice. MPLS without AAL2 is a good option for data traffic (right) but not for voice (left), because it gives a much higher delay than the other options. For a given maximum delay in the Iub interface, these curves may be used to estimate the number of users that can be served with the available capacity.

A detailed description of the simulated scenarios and additional results, which cannot be presented here due to lack of space, can be found in [5,24].

5 Conclusions

The UMTS transport layer is expected to migrate from ATM to packet-switched architectures that can provide the required quality of service at a lower cost. In this scenario, AAL2 (with SSSAR and CPS) over MPLS is a simple solution that offers a functionality similar to ATM, but in a more flexible and efficient way. This paper has discussed the main issues in transporting voice and data traffic from base stations connected with low capacity links, and has compared the performance of AAL2/MPLS with other proposals, both analytically and by simulation. Although the study focused on the access network, MPLS infrastructure may also be used in the UMTS core, and may even be shared by traffic from external IP networks.

AAL2/MPLS is more efficient than AAL2/ATM (typical differences are between 12% and 20%), so a larger fraction of the available capacity is dedicated to carry user traffic, and more customers can be served with the required QoS. Our simulator can be used to estimate the benefits of AAL2/MPLS in practical scenarios with different traffic loads and multiplexing strategies. AAL2/MPLS is particularly well suited for short packets, which are the most affected by protocol overhead.

References

1. 3GPP Technical Specification 25.426, v5.2.0: UTRAN Iur and Iub interface data transport & transport signalling for DCH data streams, Release 5 (Sep. 2002)
2. 3GPP Technical Specification 25.414, v5.4.0: UTRAN Iu interface data transport and transport signalling, Release 5 (Mar. 2003)

3. ITU-T Rec. I.363.2: B-ISDN ATM Adaptation Layer Specification: Type 2 AAL (Sep. 1997)
4. Rosen, E.: Multiprotocol Label Switching Architecture. RFC 3031 (Jan. 2001)
5. García Hernando, A.B.: Dimensioning and Quality of Service Support in the UMTS Access Network. Ph.D. Thesis. Technical University of Madrid (2002)
6. García Hernando, A.B., García, E., Álvarez-Campana, M., Berrocal, J., Vázquez, E.: A Simulation Tool for Dimensioning and Performance Evaluation of the UMTS Terrestrial Radio Access Network. Lecture Notes in Computer Science, Vol. 2515. Springer-Verlag, Berlin Heidelberg New York (2002)
7. 3GPP Technical Report 25.933, v5.3.0: IP transport in UTRAN (June 2003)
8. Gelonch, A. (ed.): Research results on UTRAN (RRM, QoS and packet data compression) Phase2. Project IST-1999-10699 Wine Glass, deliverable D11 (Dec. 2001)
9. Kempf, J. (ed.): IP in the RAN as a Transport Option in 3rd Generation Mobile Systems. Mobile Wireless Internet Forum, Technical Report MTR-006, release 2.0.0 (June 2001)
10. Pazhyannur, R., Ali, I., Fox, C.: PPP Multiplexing. IETF RFC 3153 (Aug. 2001)
11. Bormann, C. The Multi-Class Extension to Multi-Link PPP. IETF RFC 2686 (Sep. 1999)
12. Venken, K., de Vleeschauwer, D., de Vriendt, J.: Designing a diffserv-capable IP-backbone for the UTRAN. 2nd International Conference on 3G Mobile Communication Technologies. London, UK (Mar. 2001)
13. Davie, B.: An Expedited Forwarding PHB (Per-Hop Behavior). RFC 3246 (Mar. 2002)
14. Guo, Y., Antoniou, Z., Dixit, S.: IP transport in 3G radio access networks: an MPLS-based approach. IEEE Wireless Communications and Networking Conference, WCNC 2002. Orlando, Florida (Mar. 2002)
15. Feng, L.: QoS Support in IP/MPLS-based Radio Access Networks. Institute for Communications Research (ICR) Seminar (Oct. 2002)
16. Jabbari, B., Papneja, R., Dinan, E.: Label Switched Packet Transfer for Wireless Cellular Networks. IEEE Wireless Communications and Networking Conference, WCNC 2000 (Aug. 2000)
17. Chiussi, F., Khotimsky, D., Krishnan, S.: Mobility Management in Third-Generation All-IP Networks. IEEE Communications Magazine, Vol. 40, Num. 9 (Sep. 2002)
18. MPLS Forum Technical Committee: Voice over MPLS - Bearer Transport Implementation Agreement. MPLSF 1.0 (July 2001)
19. MPLS/Frame Relay Alliance Technical Committee: I.366.2 Voice Trunking Format over MPLS. MPLS/FR 5.0.0 (Aug. 2003)
20. MPLS/Frame Relay Alliance Technical Committee: TDM Transport over MPLS using AAL1. MPLS/FR 4.0 (June 2003)
21. ITU-T Recommendation I.366.1: Segmentation and Reassembly Service Specific Convergence Sublayer for the AAL type 2 (June 1998)
22. ITU-T Recommendation I.366.2: AAL type 2 Service Specific Convergence Sublayer for Trunking (Feb. 1999)
23. Fineberg, V.: QoS Support in MPLS Networks. MPLS/FR Alliance (May 2003)
24. Vázquez, E.: Analysis and Simulation of IP and MPLS Transport in the UTRAN. Technical University of Madrid (2003)
25. OPNET Modeler. web page: <http://www.opnet.com/products/modeler/>