

# Transport resources reservation in IMS frameworks: Terminal vs. PDF driven

Antonio Cuevas<sup>1</sup>, Jose I. Moreno<sup>1,2</sup>, Hans Einsiedler<sup>2</sup>

<sup>1</sup> Universidad Carlos III de Madrid, Dpto. Ing. Telemática  
Avda. Universidad 30,  
28911 Leganés, Madrid, Spain  
{antonio.cuevas, joseignacio.moreno}@uc3m.es  
Deutsche Telekom Laboratories  
Ernst Reuter Platz 7  
D-10587 Berlin, Germany  
{joseignacio.moreno, hans.einsiedler}@telekom.de

**Abstract.** IMS is a good candidate to become the service platform for next generation networks (NGN). Among one of its key characteristics is the ability to keep the Internet paradigm of application and transport separation while designing interfaces between the two layers. In IMS there is an interaction between the CSCFs (call session control function), SIP proxies managing the application setup, and the PDF (Policy Decision Function), controlling the transport network. Still, the terminals have to perform the transport resource allocation (activating the PDP context). In future networks, a similar behavior is possible but also another approach can be followed: the PDF is in charge of allocating the transport resources between the terminals, allowing to reinforce the network control. In this paper we analyze and compare both approaches, including by resorting to simulation.

**Keywords:** IMS, PDP, RSVP, ns2, simulations, resource reservation, transport.

## 1 Introduction

It is word of mouth that telecommunications are suffering a major revolution and that, like all major changes, lots of uncertainties exist. However some trends can be identified: technologically, a migration to a universal IP network and, in the commercial field, the raise of new business models and relationships between the parties.

Migrating to a universal IP network gives birth to many opportunities: the use of any access technology, wired or wireless, of any device, any application and under all circumstances (fixed, mobile) is a promise of these new networks, that are being designed to be user centric. In fact, most of the telecom operators and service providers have internal roadmaps to migrate to IP technology completely within the next 5 years trying to reduce OPEX (OPERational EXpenditures) and CAPEX (CAPital EXpenditures). This can be done by migrating all their networks (e.g. SDH,

ATM, FR, X.25) to a universal IP network [6]. The technological challenges are also important, struggling mainly in integrating different functionalities and solutions, for instance, IP mobility and enabling QoS in data transport

In the business arena, tensions emerge and the directions to be taken by the telecommunications world are not so clearly identified. However, the raise of new business models, like the semi-walled garden one [18], is identified along with the necessity for all the possible models to coexist. The semi-walled garden model takes advantage of the openness of the Internet to build services but, still, lets the network operators keep a central role in the business value chain, making them “service brokers”.

Service platforms are a key part in next generation networks. They enable new business models and commercial relationships between the parties. NTTs’ i-mode [18], and 3GPP Open Service Access (OSA) and IMS systems are known service platforms based on the semi-walled business model [8].

The technological challenges implied the need of supporting and coordinating several business models, entities and services, requires large research developments. It is thus no surprise that aspects that have “long ago” attracted attention and reached satisfactory solutions are again “hot topics” striving to cope with these challenges, bringing technology and business concepts together. One of these topics is QoS (Quality of Service). And one of the biggest concerns is how to integrate and manage the QoS-enabled data transport in next generation networks (NGN) and service platforms.

IMS service platform directly targets the operator’s traditional business of multimedia communications (phone calls). IMS “core” applications demand strong QoS requirements and, thus, the QoS interfaces to the network are an important part of the IMS design. While i-mode or, even OSA, still adopt “traditional cellular networks” protocols, IMS is based on IPv6 and other open IETF protocols. IMS respects the Internet paradigm of transport and application layers separation, yet building the interfaces between them. This paper embraces IMS’ approach in managing QoS-enabled data transport in next generation networks and service platforms.

IMS key element in dealing with QoS is the PDF (policy decision function) that is like a Broker between the QoS defined at the application level and its actual enforcement at the network level. Still, in IMS, the terminals must perform the transport resources reservation. This paper explores the possibility of the PDF performing the transport reservation and not the terminals. It will analyse and compare both approaches in an IMS-like framework.

To perform our analysis the paper is structured as follows: in the next section we provide the reader with the basic IMS knowledge so that it can understand the discussion. Section 3 is the core of this paper: it analyses and compares two strategies to perform transport resource allocation in IMS-like frameworks. An important part of this analysis is done by resorting to simulation and this is dealt with in section 4. The conclusion gathers the main results, tries to answer which strategy is better and opens for future works.

## 2 IMS overview

This section will present what we consider will be the core of NGN service platforms, the IMS.

The IMS goal is not to deliver any service; rather IMS is designed to assist peers in establishing, managing and tearing down their (multimedia) sessions. Roughly, we can say that IMS is a SIP proxies infrastructure –termed CSCFs-. CSCF are ready to be used by costumers and their devices' SIP User Agents (SIP-UA). Figure 1 presents the IMS architecture. Some of the main IMS features are:

- IMS uses IPv6 as network layer and its interfaces employ open IETF standardized protocols.
- Like any SIP based infrastructure, it assists peers in setting up, controlling and tearing down their sessions (although IMS is not a session participant).
- It interacts with the network operator infrastructure to fit the QoS given to transport the session flows.
- IMS informs the network operator about the sessions so that it can do service bundling and the “semi-walled garden” business model is enabled.
- IMS respects the Internet paradigm of transport and application separation.

In the following discussion, for ease of understanding, we will consider the IMS infrastructure as an entity independent of the network operator. This is not really so, indeed, IMS is owned by the network provider, but this “ownership” does not reduce the validity of our exposition.

The IMS platform works at the service level. IMS follows the Internet paradigm and, as such, it is completely independent of the data transport level which is handled by the UMTS network. However, a fundamental IMS feature is to achieve negotiation with the network operator, on a per session basis, of the transport-level parameters so that they fit the application-layer requirements. The IMS' PDF is the key entity: It is contacted using Diameter protocol [3] by the IMS's CSCFs nodes. Using another IETF standardized protocol, COPS [9], the PDF contacts the network routers (or the GGSN in UMTS networks). The GGSN (gateway GPRS support node) connects the UMST network to other networks, such as IP based ones like the Internet. Instructed by the IMS's PDF, the GGSN will enforce the QoS at network level.

Following the semi walled garden business model, the user's profile, his Authentication, the Authorization to consume resources and the accompanying Accounting (AAA) and posterior billing are handled by the network operator. IMS depends on the network operator's AAA services to control their users. Users trust and pay the network operator who, in its turn, trusts and pays (retaining, for example, 10 percent of the total amount) the IMS.

When a user gains access to UMTS networks, an authentication process is performed with the network operator's home subscriber server (HSS), which holds user's data and credentials. When IMS (which does not hold the user credentials) needs to authenticate a user, it delegates and depends on the HSS and the previous authentication of the user to this system. The IMS CSCF interfaces with the HSS following the IETF Diameter protocol.

Concerning authorization, this is done by the IMS/CSCF itself but is based on the user profile and context obtained, also using the Diameter protocol, from the HSS.

Accounting (and charging) follow the same philosophy as QoS: they are done at two separate levels (application and data transport) and an entity, this time part of the UMTS network infrastructure, processes, correlates and consolidates the relevant parts of the information and generates call detail records for the UMTS billing system. Thus, richer charging schemes (and tariffs) than the current ones are possible. The central entities are the charging data function (CDF) and the charging gateway function (CGF). IMS entities (namely CSCFs) communicate with the CCF following the Diameter base protocol. CSCFs can instruct the CCF about the type of session (e.g. audio or video call), its duration, or the number of participants. The interaction between the GGSN and the CGF is also defined to receive session information in the network plane (e.g., number of bytes sent and received).

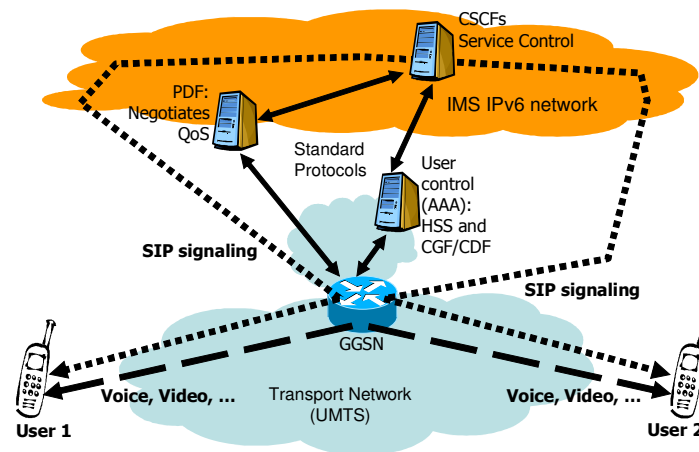


Figure 1 IMS architecture

A short usage scenario will help the reader to better understand IMS. A user powers on his mobile. A registration is done with the network, following a pure 3G, non-IMS-related process. The central node in this process is the HSS. Latter on, this user wants to setup a call using the IMS infrastructure. He sends SIP signaling to the callee. This SIP signaling traverses IMS CSCFs. CSCFs, before letting the call setup proceed, consult the HSS. The HSS “knows” the user (who registered before) and gives a positive answer to CSCFs. CSCFs then let the signaling proceed. CSCFs inform the PDF that network resources will be needed to accommodate the voice transport between the users. Once this is done, the users allocate the transport resources in the GGSN. The GGSN, before allowing this allocation consults the PDF. The PDF configures the GGSN to allocate the needed resources for the voice call. CSCFs may account aspects such as quality of the voice call and its duration. GGSN, on its side, can account for bytes sent and received. All the accounting information is gathered and correlated by the CGF/CDF. Depending the network policies the user

will (or will not) be billed both by the call service and by network resources consumed. The user will finally receive one bill from his network operator with all the services he enjoyed, including this voice call.

IMS has other features, for example, content adaptation capabilities. Besides, some details are still to be completely solved like the integration with Mobile IPv6. For the purposes of our discussion, we do not need to describe them further; the interested reader can check, for instance, [16] or [4]. The reader now understands the main IMS characteristics and we are ready to analyze and compare the IMS way of allocating transport resources (letting the terminal allocate the PDP context) to the approach where the PDF is in charge of such task.

### **3 Terminal vs. PDF/SIP Proxy driven resource reservation**

The IMS service platform presented in the previous section provides a good solution for managing and coordinating the different aspects of service provision, including the various issues affecting QoS, for instance, matching the codecs chosen and the transport resources allocated. In IMS, the terminal is responsible for the transport resource reservation. It does so by activating a PDP context, a well-known 3G networks procedure. In NGN, only IP mechanisms, not tied to any network technology, should be employed and the PDP context activation should be replaced by protocols like RSVP [2]. We point out that, for scalability reasons, the most accepted approach ([10], [14]) is to design IP networks with RSVP and IntServ [1] capabilities only in the edges; the core network is endowed with less accurate, more scalable QoS capabilities. This discussion is completely out of this paper's objectives. Regaining our main thread, we consider two models for resource reservations:

- Terminal driven, in which terminals allocate transport resources using RSVP. A PDF "policing" this resource allocation is a possible scenario (see [11]).
- Network driven, in which the PDF is responsible for allocation of transport resources between the communication endpoints. This resource allocation will be done following the "guidelines" that the PDF is given by the CSCFs (the IMS' SIP proxies) to police the resource allocation (see section 2). COPS [9] and its COPS-PR [5] extension are the protocols that the PDF could employ to configure the routers (or GGSN in UMTS) to accommodate the needed resources.

Let's recall that the main goal of this paper is comparing these two approaches: the terminal vs. the PDF doing the transport resource allocation in IMS-like frameworks.

Salsano and others ([15], [17]) have already considered and compared these two possibilities in a SIP proxy infrastructure (CSCFs in IMS) interacting with bandwidth brokers. Bandwidth brokers (or QoS brokers) are nodes comparable to the PDF in IMS, taking policy decisions affecting the transport resources but also managing the whole IP network. Salsano et al. do not consider other elements present in IMS and NGN service platforms like the ones performing accounting (CDF/CGF) or user control (HSS). Thus, their analysis and comparison of the two transport reservation techniques has to be broadened to IMS like frameworks; this is the goal of this paper.

Let's note that Salsano's proposed interaction follows this sequence: SIP proxy, routers, bandwidth broker and routers again, while we profit from the IMS defined PDF to CSCF interaction to follow a simpler sequence: CSCF, PDF, routers (or GGSN). For reader's comfort we recall that SIP proxy is assimilated to IMS' CSCFs and bandwidth brokers to IMS PDF.

In the PDF driven transport resource reservation, terminals will need to support only basic SIP, they do not need to implement any resource reservation mechanism like RSVP, since the appropriate inter-working of CSCF, PDF and routers (or GGSNs) is able to "deduce" from the SIP message exchange (e.g. looking at the codecs) the required transport resources and allocate them in the transport network. Terminals, thus, can be simpler. Besides, heterogeneous networks may employ different resource allocation mechanisms, provide that the SIP signaling and the correct interpretation of the transport resources needed by the codecs is kept end-to-end.

Next generation networks and service platforms will enable complex "ecosystems" where different parties interact making profit and joining their efforts to offer richer customized services to the users. QoS enabled data transport will not be an exception. It will evolve (keeping the "user interface" simple) from current flat rates and best-effort characteristics to differentiated services with various performances and will become a tool to develop different products and tariffs for different market segments, boosting the competition (as stated e.g. in [19]). Pushing the resource reservation process from the terminals to networks nodes like the PDF (besides letting those nodes take policy decisions) will improve the integration and management of the transport service into the overall service provisioning chain. The network could provide, for instance, multicast capabilities with content adaptation for the branches of the multicast tree directed to terminals with little performance and slow connections.

Orchestration of all the nodes in these rich ecosystems will be a key aspect in next generation networks and service platforms. Each node will have different interfaces. For instance, IMS' CSCFs interact with many more nodes (e.g. PDF, HSS, CDF). To avoid overloading the nodes and making them potential bottlenecks, the functionalities they assume must be chosen carefully. And this includes letting the CSCFs and PDF doing the transport resources allocations instead of assigning this task to the terminals. This may specially impact the SIP proxies (in IMS actually the P-CSCFs) since they would need to keep state of the transport reservation and of the call (SIP session) associated to it. Note that, in IMS, P-CSCFs are not stateful, this is let to the S-CSCFs. We will focus on the processing costs for the different nodes of the two proposed resources reservation strategies (transport resource reservation driven by the terminals or by the PDF/CSCFs), by resorting to simulations as we will discuss and describe in section 4 of this paper.

Mobility also impacts on QoS model, choosing one the two transport reservation strategies. In large networks, it is sensible to have several SIP proxies scattered along the network (like the IMS' P-CSCFs). The scenario that is relevant to us is when the terminal hands over between parts of the network assigned to different P-CSCFs. Mobility in NGN seems to be supported by two mechanisms: make-before-break and context transfer [13]. Performing a context transfer in the handover process implies gathering state information from all stateful entities involved in a session. In case of

transport QoS configurations, the context transfer must be immediate. The parts of the context to be transferred immediately during a handover should be kept to the minimum. From this point of view, it is interesting to place a large fraction of the context in the common element during handover: the terminal. Due to this reason it is better to let the terminals doing the QoS-enabled transport reservation rather than the PDF instructed by the CSCFs.

Another point to consider is that, if the terminal needs to handle the transport resource reservation process, more messages need to be sent over the access link, which may have scarce bandwidth. Thus, the transmission time of these messages may be significant and may slow the session setup process.

#### **4 Simulation based analysis and comparison**

Previous section analyzed the pros and cons of letting the terminal or the PDF - instructed by the CSCFs- doing the transport resource reservation. We saw the processing load of the different nodes was an aspect to be analyzed. There are essentially three ways to do so: creating and analyzing a mathematical model, the implementation of a prototype where a number of tests are executed and resorting to a simulation model. Due to the many nodes and protocols involved in next generation networks and service platforms, the development of a mathematical model that is simultaneously faithful and tractable is difficult. Same happens with deploying a test-bed. Besides, in lab prototypes it is not possible to perform some kind of tests like overloading the network with thousands of calls. Simulation is, therefore, a valuable tool in obtaining indicators early in the engineering process and refining the system as it is being implemented. Simulation is our choice for comparing the processing load of the two resource reservation strategies studied in this paper.

We employ ns2 [12] to build our simulation model. We extended ns2 to support the protocols present in an IMS like scenario [7]. We developed a “two level” SIP proxy infrastructure emulating the IMS’ P-CSCF and S-CSCF+I-CSCF SIP proxies. The P-CSCFs can interact with a PDF and accounting nodes (the CDF in IMS). The S-CSCF + I-CSCFs can interact with an authentication and authorization infrastructure (like the HSS in IMS). Roaming was supported. Our main goal was to evaluate the processing delays in the nodes. Each message, before exiting a node, is put in a FIFO queue where it has to wait for the other messages to be processed and an own processing time. User terminals do not have any processing delay. This reflects the fact that terminals have to handle few messages while “network nodes” have to process several hundreds of them, corresponding to the terminals they serve. To our knowledge, the novel “IMS like” model we developed can be a valuable tool to asses in the engineering and deployment of the complex IMS infrastructures.

The developed IMS model will serve us to evaluate and discuss the impact of the different transport resource strategies. We used a scenario (Figure 2) with 4 domains, each domain divided in one “Core Network” with a S+I CSCF and a HSS and 8 “Access Networks” each with one P-CSCF, a PDF and the accounting infrastructure (CDF). The “networks” are connected by ERs (edge routers). The processing delays of the nodes were chosen proportionally to the number of messages they had to

process. For the P-CSCFs the calculation was done for the case they do not have to handle the transport reservation messages.

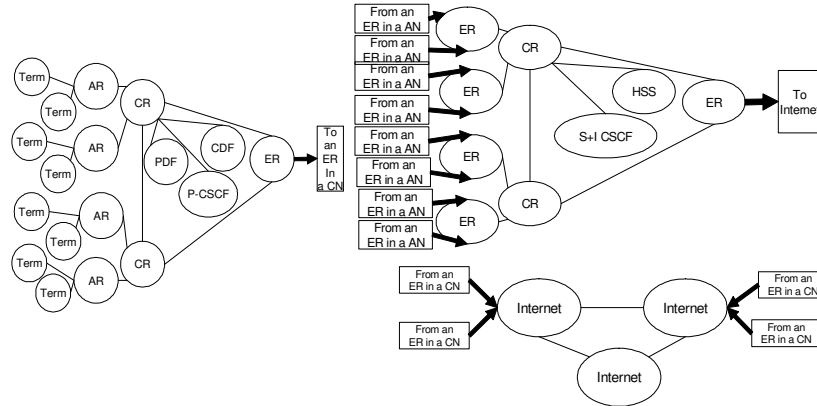


Figure 2 Simulation scenario: Terminals, an “access network” (AN), a “core network” (CN) and the connection between the four domains

Under a low load of sessions (calls) setup attempts to be handled by the IMS infrastructure, the difference in session setup time between the two strategies (terminal vs. PDF/P-CSCF driven transport reservation) matches the extra processing time needed to be done by the P-CSCF in the second strategy. But when the network load increases this time difference increases (Figure 3).

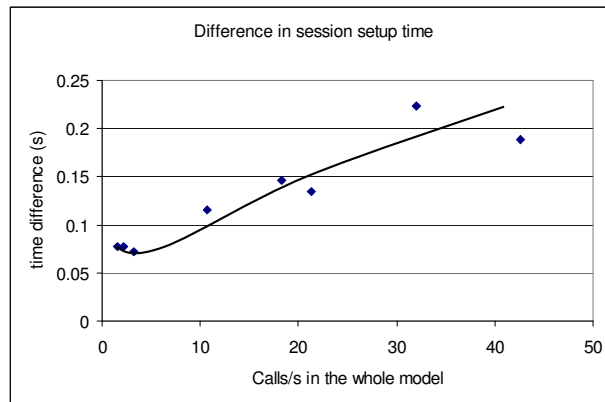


Figure 3 Difference in average session setup between the two strategies for reserving transport resources. This difference increases when the IMS infrastructure starts to be highly loaded. Points represent the difference in average session setup time for the PDF driven and the terminal driven strategies. Measures are done for different call loads. Each point is a measure.

To try to explain this difference we analysed individually the behaviour of the different nodes (HSS, PDF, CDF and the two kinds of CSCFs). We saw that none of the nodes reached an overloaded state even for the highest call load. By overloaded



sate we mean that the rate of input messages into a node is constantly higher than the rate of messages exiting this node, thus, due to the FIFO characteristic of the message processing queues, the processing time of the messages increases steadily. We stress that the overloaded state is not reached for any of the nodes and that the processing time of the messages has a similar behaviour in both strategies. But, for the P-CSCF/PDF transport resource driven reservation strategy, the processing time in the P-CSCF suffers big fluctuations and so do the other nodes. Since in this strategy this node has to process more messages, it is more exposed to randomly distributed instants of higher processing load. This node interacts with all the others (but with the HSS) thus this phenomenon affects the whole system. This results in a higher session setup time in this mentioned strategy.

An important conclusion here is that in complex scenarios, like the IMS, with many nodes interacting, the whole system performance can be influenced by the behaviour of one central node and this node is the P-CSCF.

We performed another set of tests, also for both strategies. We exposed the system to a constant and low load, introduced a short peak of calls and, then, injected again the same low and constant load. In the transport resource allocation driven by the terminal strategy, during this peak of calls, the session setup time increases fewer than in the P-CSCF/PDF driven strategy. Besides, the lapse to “recover” the “normal” session setup time in stable conditions is shorter in the P-CSCF/PDF driven strategy. The tests were repeated but increasing the processing power of the P-CSCFs in P-CSCF/PDF driven strategy, proportionally to the extra number of messages it had to manage respect to the terminal driven strategy. Results were similar. This stresses the conclusions we presented before: the strong influence in complex systems in the behaviour of a node (and of the whole system) of the many other nodes inter-working with it.

## 4 Conclusion

We analyzed two possible strategies for performing transport resource allocation in IMS and next generation network (NGN) scenarios. Those were letting the terminals perform this process (as it is done in IMS) or assigning this task to the PDF (or bandwidth broker in NGN). As usual, both possibilities have pros and cons. Making the PDF deal with this task presents many benefits allowing simpler terminals and making easier integrating the QoS enabled transport into the whole service delivery chain. However, as we saw thanks to simulation, if this strategy is chosen, the SIP proxies interacting with the PDF need to be carefully dimensioned since they become a central point interfacing with many nodes and may be potential bottlenecks. Mobility and context transfer issues are also an important part to analyze in future works.

**Acknowledgments.** This work was partly funded by Spanish Minister of Education (MEC) under CASERTEL-NGN project (TSI2005-07306-C02-02) and Madrid National Research Program e-Magerit (S-0505/TIC/000251).

## References

1. Braden, R. et al. RFC 1663, "Integrated Services in the Internet Architecture: an Overview", June 1994
2. Braden R. et al. RFC 2205, "Resource ReSerVation Protocol (RSVP) -- Version 1 Functional Specification" September 1997
3. Calhoun P. et al. RFC 3588, "Diameter Base Protocol" September 2003
4. Camarillo, G. and García Martín, M.A.. "The 3G IP Multimedia Subsystem (IMS): Merging the Internet and the Cellular Worlds", Wiley, 2004.
5. Chan, K. et al., RFC 3084 "COPS Usage for Policy Provisioning (COPS-PR)". March 2001
6. Cherry, S. "Nothing but Net: Britain switches its entire phone network to the Internet Protocol." IEEE Spectrum. January 2007.
7. Cuevas. A. "Contribution to Design Suitable Session Setup Solutions in 4G Networks". Ph.D. Thesis. Universidad Carlos III de Madrid. 2006
8. Cuevas, A. et al., "The IMS Service Platform, the Key for Next- Generation Network Operators to Be More than Bit Pipes", IEEE Comm. Mag., vol. 44, no. 8, Aug. 2006
9. Dirham, D. et al., RFC 2748 "The COPS (Common Open Policy Service) Protocol" January 2000
10. García G. et al. "Soporte de QoS en Redes de 4ª Generación" Revista IEEE America Latina, Volume 4, Issue 1, March 2006
11. Herzog, S., RFC 2749 "COPS usage for RSVP", January 2000
12. Information Sciences Institute, University of Southern California, Network Simulator 2 <http://www.isi.edu/nsnam/ns/>
13. Jähnert, J. et al. "The pure-IP Moby Dick 4G architecture" Computer Communications Vol 28/9 pp 1014-1027
14. Marques, V. et al. "An IP-based QoS architecture for 4G operator scenarios" IEEE Wireless Communications, June 2003
15. Papalilo, D., Salsano, S., Veltri, L. "Extending SIP for QoS support", Joint Planet-IP NEBULA workshop, Courmayeur, 2002.
16. Poikselkä, M. "The IMS : IP multimedia concepts and services in the mobile domain" Wiley & Sons ISBN: 047087113X (2004)
17. Salsano, S. Veltri, L. "QoS Control by Means of COPS to Support SIP-Based Applications" IEEE Network, March/April 2002
18. Scott-Joynt, J. "The secret of NTT's i-mode success" BBC news available at <http://news.bbc.co.uk/1/hi/business/1835821.stm>
19. WirtschaftsWoche, "Google & Co. müssen zahlen" WirtschaftsWoche, 23<sup>rd</sup> february 2006, Germany