The IP Multimedia Subsystem in Next Generation Networks

Gilles Bertrand

May 30, 2007

Abstract

The IP Multimedia Subsystem (IMS) is a network functional architecture that is seen as a promising solution for facilitating multimedia service creation and deployment, as well as supporting interoperability and network convergence. IMS allows network operators to play a central role in traffic distribution, therefore being more than “bit pipes”. For all these reasons, IMS has generated intense research and standardization efforts. The aim of this paper is to present the overall IMS architecture and protocols, as well as the related stakes.

Keywords: IP Multimedia Subsystem, Quality of Service

1 Introduction

The Internet Protocol (IP) is ubiquitous: according to the Internet Society it is used to interconnect more than 1 billion people all over the world. More than 15% of the world population now has access to the Internet and this penetration rate has doubled between 2000 and 2006. The Internet provides interoperability at a very large scale, enabling people using different terminals to communicate.

While the first generation of the Internet was mostly dedicated to the transport of non real time data, services with stringent Quality of Service (QoS) requirements are now largely adopted (e.g. Telephony over IP (ToIP), Video-conferencing). Moreover the share of the multimedia services in the operators revenue is expected to grow in the next few years [1]. The move toward an all IP architecture for service delivery appears to be a strong trend. In this context, customers seem to desire an access to personalized interactive, multimedia services, on any device, and anywhere. This trend introduces new requirements for network infrastructures. The IP Multimedia Subsystem (IMS) is seen as a promising solution for fulfilling these expectations.

IMS refers to a functional architecture for multimedia service delivery, based upon Internet protocols. Its aim is to merge Internet and cellular worlds, in order to enable rich multimedia communications [2, 3]. It is specified in the 3rd Generation Partnership Project (3GPP).

IMS was introduced in UMTS release 5 (March 2003) and 6 [4]. In its first version, it focused on facilitating the development and deployment of new services in mobile networks [4]. It was later extended by the European Telecommunication Standards Institute (ETSI), in the scope of its work on Next Generation Networks (NGNs)². A standardization body of ETSI, called Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN) standardizes IMS as a subsystem of NGNs. TISPAN has published a first release of ETSI IMS standards and is currently working on a second release. We could say that 3GPP describes the point of view of mobile operators (support of new applications), while TISPAN adds the wireline operators specifications (convergence). TISPAN makes specifications for several non IMS subsystems like Network Attachment Subsystem (NASS) and the Resource Admission Control Subsystem (RACS) (see section 3) [2]. Most of the IMS protocols are standardized by the Internet Engineering Task Force (IETF) (e.g. the Session Initiation Protocol (SIP)). Other standardization bodies are involved in the development of IMS. For example, the Open Mobile Alliance publishes additional service related requirements (e.g. Push to talk over Cellular (PoC) [6]) and leads interoperability related operations.

We should distinguish between the IMS core (TISPAN vocabulary) and IMS (3GPP vocabulary). It is to be noted that this document focuses on ETSI TISPAN standards. Therefore, in the rest of this document the term IMS architecture refers to an “IMS core” plus several non IMS subsystems such as the NASS and the RACS. Stated differently, we use the term IMS architecture to refer to a NGN architecture featuring an IMS core.

It is difficult to get a complete view of the IMS architecture, because IMS is still being defined³, by several standardization bodies using different terminologies. In addition, IMSs standards describe several interfaces and functional entities having complex relationships. Moreover, among the several good papers on IMS, some are partially obsolete because of the fast evolution of

---

¹Network, Multimedia and Security department (RSM) – GET/ENST Bretagne – gilles.bertrand@ieee.org
²3GPP Release 7 (March 2007) provide a unified IMS supporting heterogeneous network access technologies (e.g. DSL, WLAN). [5]
³TISPAN Release 2 is under way
the standards [7] [8], and very few give an overview of the complete NGN/IMS architecture (including non IMS subsystems) [9]. Therefore, this paper aims at depicting the overall IMS architecture and protocols, as well as the related motivation.

This paper is divided in four sections. After a short introduction to IMS, we introduce its basic principles and purposes in section 2. The IMS architecture is described in section 3. The IMS protocols are described in the subsequent section. In the last section, the mechanisms for QoS support are explored. To the best of our knowledge, no other short paper summarizes all these aspects of IMS.

2 Motivation for the use of IMS

2.1 Basic principles

One aim of IMS is to make the network management easier. Therefore, it separates control and bearer functions. This means that IMS features an overlay service delivery network on top of a packet switched infrastructure. Moreover, IMS should allow the migration of Circuit Switched services like voice telephony to the Packet Switched domain. As a result, IMS should lead to network administration savings, because an all-IP integrated network is easier to manage.

IMS is an end-to-end architecture that must support several kinds of equipments. In addition, IMS is intended to be “access agnostic”, which means that service delivery should be independent of the underlying access technology. Thus, the use of open Internet Protocols is specified in IMS for better interoperability. IMS supports roaming between different networks (3GPP Release 6).

The level of QoS that can be provided in IMS networks determines the services that can be deployed in such networks. QoS delivery is therefore critical in IMS networks. As a result, QoS management functionalities are integrated in the IMS architecture.

IMS is a horizontal architecture: it provides a set of common functions called service enablers that can be used by several services (e.g. group/list management, presence, provisioning, operation and management, billing...). This makes service implementation much easier and faster. Moreover, it allows a tight interaction between several services. This is an appreciable progress compared to most of currently used architectures, that feature vertical “stovepipe” service implementation (see Figure 1) [12].

2.2 Business and technical motivations

While the average revenue per user (ARPU) is decreasing for several network operators, IMS is seen by many

\[\text{Figure 1: Horizontal vs vertical services integration}\]

(UE), Access Network, NGN core, NASS (see section 3.2) and RACS (see section 3.3).

The Resource Admission Control Subsystem (RACS) and the Network Attachment Subsystem (NASS) are two important non IMS subsystems standardized by ETSI TISPAN.

Several service subsystems can coexist in an IMS architecture, for example the IMS core and the PSTN emulation function (Figure 3). Because IMS is still being defined (TISPAN Release 2 is under way), and because it describes several interfaces and functional entities, a complete IMS system is quite difficult to represent. It should nevertheless be kept in mind that IMS is a part of a functional architecture, and several of its components can be implemented in a single hardware.

3.2 The Network Attachment Subsystem (NASS)

NASS is specified in [13]. It provides IP addresses and other user equipment configuration parameters dynamically. Its functionalities may be roughly summarized in the following way: the NASS play the role of a DHCP server, a RADIUS client and provides location management function [13]. More precisely it provides:

- IP addresses and configuration parameters
- User authentication
- Authorization of network access, based on user profile.
- Access network configuration, based on user profile.
- Location management.

The NASS can be divided in several functional entities [13], as depicted in Figure 4.

- The Network Access Configuration Function (NACF) is responsible for the IP address allocation to the User Equipment and it may provide some additional parameters. It could be implemented by a Dynamic Host Configuration Protocol (DHCP) server.
  - The Access Management Function (AMF) is mostly an interface equipment (translation, forwarding of user requests) between the Access network and the NACF.
  - The Connectivity Session Location and Repository Function (CLF) is used to associate the user IP address to his location information. In addition, it may store further information about the user (profile, preferences...).
  - The User Access Authorization Function (UAAF) performs authentication for network access, based on the user profile stored in the PDBF.
  - The Profile Data Base Function (PDBF) stores the user profiles and authentication data.
  - The CNG Configuration Function (CNGCF) is used to configure the Customer Network Gateway (CNG) when necessary.

3.3 The Resource Admission Control Subsystem (RACS)

The RACS is specified in [14]. It is used to perform admission control. It can be divided into two functional blocks: the Serving Policy Decision Function (S-PDF) and the Access Resource and Admission Control Function (A-RACF), as described in [14].

The S-PDF performs policy decisions and is able to send resource requests to a A-RACF and/or Border Gateway Function (BGF). It communicates the policy decisions back to the Application Function (e.g. the P-CSCF) [14].

The A-RACF performs Admission control which means that it checks if the requested resources may be allocated for the involved access. It returns the result of the admission control to the S-PDF [14].

3.4 The IMS core

The IMS core is used for session and media control. Its structure is depicted in Figure 6 [15] and its functional
components are the following:

- **The Call Session Control Function (CSCF)** establishes, monitors, supports and releases multimedia sessions and manages the user’s service interactions [16]. It can play three different roles: Serving-, Proxy- or Interrogating- Call Session Control Function (S-, P- and I-CSCF) [17]. The S-CSCF is the proxy server controlling the communication session. It invokes the Applications Servers related to the requested services. It is always located in the home network [11]. The P-CSCF is the IMS contact point for the SIP user agents. According to 3GPP, it may include a Policy Decision Function (PDF) that manages the QoS over the media plane. The policy Decision Function can be integrated in the P-CSCF\(^5\) or implemented as a separated entity\(^6\) (see Figure 7) [11, 18]. The PDF is part of the RACS in ETSI TISPAN standards. The I-CSCF provides a gateway to other domains. It is used essentially for topology hiding or if several S-CSCF are located in the same domain [9].

- **The Multimedia Resource Function Controller (MRFC)** is used for controlling a Multimedia Resource Function Processor (MRFP) that essentially provides transcoding and content adaptation functionalities [9].

- **The Breakout Gateway Control Function (BGCF)** "selects the network in which PSTN breakout is to occur and - within the network where the breakout is to occur - selects the MGCF" [15]. This means that it is used for interworking with the Circuit Switched domain.

- **The Media Gateway Controller Function (MGCF)** is, as its name indicates, used to control a Media Gateway.

3.5 Additional components

The application servers are SIP entities that execute the services. They can have an interface to the Open Service Access (OSA) framework or to the GSM Service Control Function [19].

The Home Subscriber Server (HSS) or User Profile Server Function (UPSF) is defined in is a secure data-based storing user profile information [20]. It can be accessed by the S-CSCF using Diameter protocol. It is to be noted that the HSS can be seen as an evolution of the former Home Location Register (HLR). In case several HSS are used in a domain, a Subscriber Location Function (SLF) is required. The SLF is simple database indicating in which HSS a user profile is located [11].

Several components are used as gateways for interworking with legacy circuit switched networks (e.g. SGW, MGCF, MGW, BGCF) [15]. The Signaling Gateway (SGW) interfaces the signaling plane of the circuit switched (CS) network [11]. In addition to controlling the Media Gateway, the Media Gateway Control Function (MGCF) is used for protocol conversion between SIP and ISUP\(^7\) or BICC\(^8\). The Media Gateway (MGW) is able to send and receive media over Packet Switched (PS) and CS protocols [15].

4 Protocols used in IMS

As mentioned before, most of the protocols used in IMS are standardized by the IETF. They are briefly described in the sections below.

4.1 Signaling and media flow description

The main signaling protocol used in IMS is called the Session Initiation Protocol (SIP). It was primarily defined in RFC 2543 (obsolete) and later RFC 3261 [21, 22]. SIP borrows some principles of HTTP and SMTP, the two most successful Internet protocols. SIP has been selected in IMS essentially because it complies with IMS requirements and it is considered flexible (several extensions are standardized) and secure. In fact the IMS SIP is an enhanced version of SIP including several extensions as described in the 3GPP TS.24.229 standard [23, 18]. The main purpose of SIP is the establishment, modification and termination of multimedia sessions between two terminals.

The body of SIP messages is described using the Session Description Protocol (SDP). The SDP is a syntax for describing media flows (address, port, media type, encoding, etc.) standardized in RFC 2327 (obsolete) and later RFC 3264 [24, 25].

SIP is probably the key protocol in the IMS architecture. In IMS, in addition to its functionalities defined

---

\(^5\) mandatory in 3GPP Release 5  
\(^6\) 3GPP Release 6 gives the choice between these two options

---

\(^7\) Integrated Services Digital Network User Part  
\(^8\) Bearer Independent Call Control protocol
in RFC 3261, SIP is able to handle subscriber management, service control, Single Sign On, QoS authorization, billing, resource management, etc. Most of these numerous extensions have led to IETF RFCs (e.g. RFC 3312, 3262, 3313, 3455, etc.) [22].

4.2 Authentication, Authorization, and Accounting
Diameter is a recent Authentication, Authorization, and Accounting (AAA) protocol replacing the RADIUS protocol [24]. It is defined in RFC 3588 [26]. Diameter security is provided by IPSEC or TLS. Diameter is used in the IMS service framework by the I-CSCF, S-CSCF and the Application Servers (ASs) in their exchanges with the HSS containing the user profiles. It is also used in the exchanges between the RACS and the AS and CLF.

4.3 Policy support
The Common Open Policy Service (COPS) is a protocol defined in RFC 2748 [27]. It supports policy control over Quality of Service (QoS) signaling protocols (e.g. RSVP) [27]. It is used to convey policy requests and decisions between Policy Decision Points (PDPs) and Policy Enforcement Points (PEP) (see Figure 8).

COPS supports two policy management models:

- The **Outsourcing** model specifies that PDP-PEP exchanges occur for each policy decisions (e.g. COPS for RSVP, RFC 2748 and 2749) [27, 28].
- In the **Configuration or Provisioning** model the PEP stores policy rules defined by the PDP and use them for policy decisions. This provides excellent scalability for the related protocol (COPS-PR) [29].

In the configuration model, the policy rules need to be stored in the PEP. The corresponding data structure is called the Policy Information Base (PIB). It describes the configurable mechanisms for implementing policies in the PEP as well as the events that can trigger the exchange of policy information [8]. It adopts a tree-like structure described as the “Structure of Policy Provisioning Information” (SPPI) [30]. The PIBs are composed of several meta rules for policy called Provisioning classes (PRCs) (see Figure 9). The PRCs consist themselves of several provisioning instances (PRIs) that correspond for example to individual access control filters that must be applied.

4.4 Additional protocols
MeGaCo, also called H248, is a successor of the Media Gateway Control Protocol (MGCP) used for controlling media serving functions in an IMS environment. It is specified in RFC 3015 [31].

The Real Time Protocol (RTP) provides transport functions for transmitting real time data. It is specified in RFC 3550 [32]. It is used in conjunction with a control protocol called Real Time Control Protocol (RTCP) in order to allow monitoring of the data delivery and to provide minimal control and identification functionality.

It is to be noted that the use of IPv6 is mandatory in IMS networks complying with 3GPP release 5, but several equipment vendor implementations support both IPv4 and IPv6 [33].

5 QoS support in IMS
The idea of transforming a best effort IP network by introducing end-to-end QoS guarantees is an important driver for the development of IMS. This is a key consideration because the level of QoS that the IMS architecture is able to provide determines the services that can be deployed on it, and the value is assumed to reside in real time multimedia services.

5.1 QoS management principle
Two strategies are usually associated for providing a good level of QoS in packet networks. The first involves avoiding congestion phenomena. This can be done by implementing Connection Admission Control (CAC), resource reservation or simply by over-dimensioning the network (over-provisioning). A famous example of a QoS framework based on resource reservation is “Integrated Services” (IntServ) [34]. This strategy can also be used in Multiprotocol label switching (MPLS) networks using the Resource ReserVation Protocol (RSVP) [35, 36]. The second method focuses on managing congestion. It usually relies on traffic differentiation for providing better QoS to most important flows. Several standards are related to this idea, the most well known being Diffserv [37, 38].

We can distinguish between two types of QoS management schemes. The first aims at providing guaranteed QoS while the other is focused on Relative QoS. QoS guarantees like delay or loss rate bounds can be provided by resource reservation schemes. Relative QoS can
be implemented by traffic differentiation [38].

Most QoS management schemes require network hardware to be configured. For example in the Diff-Serv framework the data packets have to be marked with a suitable Differentiated Services Code Point (DSCP) according to network policies [37]. Policy based QoS control allows a network operator to easily configure the network equipments. It is essentially used for defining admission control rules and facilitates the translation of business level agreements like Service Level Specifications (SLSs) and Service Level Agreements (SLAs) into network level policies. Such translation mechanisms are represented in Figure 10. The network policy rules are defined by the operator in the Policy Decision Point (PDP). This network element is used for taking policy decisions. It answers the requests emitted by a Policy Enforcement Point (PEP). As was mentioned in section 4.3, COPS can be used to carry admission control related information between the PDP and the PEP [27].

5.2 QoS architecture

IMS networks support both admission control and QoS differentiation (see Table 1) [39]. The main network functions involved in QoS provision in IMS network are depicted in Figure 11. In this Figure can be seen the following architecture elements:

- The P-CSCF, that is aware of the SDP context of the session, and thus of the resources required for this session.
- The RACS, that takes policy decisions and interfaces with transfer functions.
- The transport layer components that actually apply the policy decisions.

The main transport layer functions are listed below:

- The Resource Control Enforcement Function (RCEF) enforces policies under the control of the

A-RACF. It opens and closes unidirectional filters called gates or pinholes, polices traffic and marks IP packets [14].

- The Border Gateway Function (BGF) performs policy enforcement and Network Address Translation (NAT) functions under the control of the S-PDF. It operates on unidirectional flows related to a particular session (micro-flows) [14].

- The Layer 2 Termination Point (L2TP) terminates the Layer 2 procedures of the access network [14].

Their QoS capabilities are summarized in Table 2 [14]. The admission control usually follows a three step procedure:

1. Authorization of resources (e.g. by the A-RACF)
2. Resource reservation (e.g. by the BGF)
3. Resource commitment (e.g. by the RCEF)

### Table 1: IP Bearer Services Manager capability in the UE and GGSN

<table>
<thead>
<tr>
<th>Capability</th>
<th>UE</th>
<th>GGSN</th>
</tr>
</thead>
<tbody>
<tr>
<td>DiffServ Edge Function</td>
<td>Optional</td>
<td>Required</td>
</tr>
<tr>
<td>RSVP/IntServ</td>
<td>Optional</td>
<td>Optional</td>
</tr>
<tr>
<td>IP Policy Enforcement Point</td>
<td>Optional</td>
<td>Required</td>
</tr>
</tbody>
</table>

### Table 2: QoS capabilities of transport layer components

<table>
<thead>
<tr>
<th>Function</th>
<th>RCEF</th>
<th>BGF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open/close gate</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Mark packets</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Police traffic</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Resource allocation (per flow)</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Meter usage</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Support NAT</td>
<td>x</td>
<td></td>
</tr>
</tbody>
</table>

6 Conclusion

As discussed in previous sections, IMS opens up new perspectives for network operators. But several technical and business challenges have to be faced in order to enable the wide adoption of this promising technology. Moreover, IMS has to solve its inherent contradictions: it relies on IP technologies allowing free communication but aims at controlling IP services.

6.1 Business issues

IMS leads networks operators to play a central role in service distribution. This involves that carriers will have to obtain content [3]. The role of the operators in the
billing of services provided by third parties has also to be clarified.

With IMS a single customer may subscribe to services from several providers. IMS therefore leads network operators into a competition with players of the Internet world. The decision to deploy IMS is strategic. Network operators may choose an early deployment scheme in order to take advantage of the higher prices charged to early adopters. In this pioneering strategy they would open the way for their competitors and take significant risks. Alternatively, a network operator may wait in order to reduce his investment costs and learn from his competitors failures. But he would face several competitors having better market experience [40].

As a conclusion, the decision to deploy IMS is more a strategic decision than a technological decision.

6.2 Technical issues

The end-to-end model adopted in IMS introduces several technical challenges, for example concerning QoS, privacy and billing [41].

The main technological issue is related to interoperability. IMS mixes the points of view of IP, wireline telephony and mobile network operators. Moreover, it introduces new networking paradigms and provides specifications, not implementation-ready solutions. Finally, it uses some recent protocols like Diameter that have not been widely deployed. For all these reasons, interoperability may be difficult to achieve in IMS networks.

The TISPAN IMS is designed to be access agnostic. One of the main motivations for IMS is to enable the delivery of real time multimedia services using IP related technologies, but IMS has to manage the different access related constraints imposed by heterogeneous access technologies (e.g. handover in radio access networks). In particular, this makes the establishment of end-to-end QoS guarantees quite difficult.

Moreover, mobile wireless devices have limited functions, are usually related to a single user and are controlled by the network operator, whereas fixed wired terminals are powerful and controlled by the user. These differences have to be taken into account for authentication (e.g. using SIM-based secure access) and security related operations, for example.

IMS aims at providing multimedia services with a unified network architecture. However, IMS does not yet include several promising technologies (e.g. P2P, VPNs, SMS and IPTV). IPTV in particular is outside the IMS domain and is not yet standardized[9].

Additional issues are currently or need to be addressed in the standards, for example: universal service obligation, number planning, lawful intercept, number portability, reliability and voice quality, emergency services, inter-carrier compensation and data protection [12].

Last but not least, there is a risk that consumers will not readily accept the privacy model related to IMS.

Acknowledgment

The author would like to thank Dr. Julio Orozco (Orange group) for his valuable advice on writing of scientific documents and Dr. Janet Ormrod for the correction of this paper.

References

[41] S. Hayes, “IP based Multimedia Services Plat-