J Netw Syst Manage
DOI 10.1007/s10922-007-9096-3

A Review of Policy-Based Resource and Admission Control Functions in Evolving Access and Next Generation Networks

Christian Esteve Rothenberg · Andreas Roos

Abstract The IP Multimedia Subsystem (IMS) adopted in the core of Next Generation Networks (NGNs) promises to make network management easier by separating the control and the transport planes. Therefore, an interface between applications and the underlying transport network has been defined that offers a dynamic and efficient management of network resources based on a policy-based resource control engine. The resulting resource management framework enables the delivery of both the existing carrier grade existing and the next generation Quality of Service (QoS) sensitive services across operator-controlled networks using heterogeneous transport technologies. This review sheds some light into the policy control layer concept and the extended nomenclature introduced by current standardization works. The approaches of international standards development organizations, such as the Third Generation Partnership Project (3GPP), the International Telecommunication Union (ITU), the European Telecommunications Standards Institute (ETSI), the WiMAX Forum, and CableLabs are reviewed and compared with each other revealing the common architectural trend. Challenges and works in progress of NGN resource management towards Fixed and Mobile Convergence (FMC) are discussed as well.

Keywords NGN · Standardization · Signaling · QoS · Convergence · IP multimedia subsystem · IMS · FMC

C. Esteve Rothenberg (✉)
Centro de Pesquisa e Desenvolvimento (CPqD), Rodovia Campinas, Mogi-Mirim,
km 118,5, CEP 13086-902 Campinas, SP, Brazil
e-mail: chesteve@gmail.com
URL: www.cpqd.com.br

A. Roos
T-Systems Enterprise Services GmbH, Deutsche-Telekom-Allee 7, 64295 Darmstadt, Germany
e-mail: Andreas.Roos@t-systems.com
URL: www.t-systems.com

1 Introduction

Traditionally, the “best-effort” model has characterized service provisioning over Internet Protocol (IP) networks. This approach has worked fine for Internet traffic, but is insufficient when trying to provide carrier grade services with real-time characteristics such as Voice over IP (VoIP) or IP television (IPTV). Furthermore, the delivery of future and converged services to an increasing number of heterogeneous user terminals and with the constraints of multiple business models requires an evolution of network architectures.

This evolution is called Next Generation Network (NGN) and it is based on the IP Multimedia Subsystem (IMS) [1], a framework for managing and controlling multimedia sessions over IP networks that separates service, control and transport planes. The layered architecture promises significant benefits in terms of new service creation and operational savings. However, the coordination between the different layers becomes a challenge and requires a real time resource control engine enabling an access technology–agnostic interaction with the underlying network infrastructure.

Essentially, policy-based resource control provides the network with the required intelligence to manage transport network resources and adapt transparently to the different needs of running services and applications in terms like Quality of Service (QoS), Authentication, Authorization, Accounting and Charging (AAAC). The resource management functions ensure that QoS of the subscribed by users is actually supported at the transport plane. This functionality is considered crucial for telecom operators in order to deliver quadruple play services (voice, data, video and mobility) over Fixed and Mobile Converged (FMC) networks in a profitable way. Consequently, different Standards Development Organizations (SDOs) are specifying policy-based resource management functionalities, including the International Telecommunication Union (ITU) and the European Telecommunications Standards Institute (ETSI) for their NGN architectures. The Third Generation Partnership Project (3GPP), the Worldwide Interoperability for Microwave Access (WiMAX) Forum, Cablelabs and the Digital Subscriber Line (DSL) Forum are working similarly on their new releases of wired and wireless access networks. The related specifications share almost the same principles but introduce a confusing nomenclature when describing the different interfaces and functional elements involved.

This article aims at providing a comprehensive review of ongoing standardization works in the area of policy-based resource management to fully understand the virtues of next generation networking and its impacts on transport network technologies. We describe the approaches of different SDOs and provide a comparison model that outlines the observed commonness and ease the understanding of the acronym-rich standards activities. We explore the challenges and standardization efforts towards a harmonized policy-based resource control framework for operator controlled NGNs in multi-domain environments using varying transport network technologies.

The outline of this work is organized as follows. Section 2 introduces background concepts on policy-based network management, resource management, and the related standardization activities. Section 3 describes the approaches of the different SDOs, whereas Section 4 provides a comparative analysis of the reviewed architectures. The challenges and the related ongoing work are presented in Section 5. Finally, Section 6 provides a conclusion of the review presented in this work.

2 Background

This section introduces basic concepts of policy-based network resource management in the area of NGN and presents the related activities in the standards arena. Policy technology leverages the network and eases the implementation of resource control and management functions in a scalable fashion. The work and interactions between the different standards development organizations involved in the NGN definition aim at achieving the long sought after goal of having standardized technology-independent interfaces to interact with resource control functions of telecommunication operator networks.

2.1 Policies in the Context of NGN Resource Management

Policy-Based Network Management (PBNM) [2] is the ability of a network to provide an automatic response to network conditions according to pre-defined policies and it is considered the best approach to ease the complex network configuration processes involved in the integration of network services into a single large network. The resource control framework of NGN architectures has the challenge to efficiently manage a series of network resources to guarantee the delivery of a wide range of QoS sensitive services over multiple transport technologies. Network resources include QoS, IP address and ports, gate control, security and charging. In order to respond to a real time resource request, the network supports various resource control functions such as admission control, resource reservation and monitoring, firewall, Network Address Translation (NAT) handling, etc.

Policy-based mechanisms leverage the network's capabilities enabling user services to operate seamlessly across varying transport networks. The operator defines network level policy rules to manage resource usage and to set priorities across applications and users in accordance with the contracted business level agreements like the Service Level Agreements (SLAs). A Service Level Specification (SLS) is a subset of an SLA that describes the operational characteristics, but hides the details of the underlying QoS-enabled network. As shown in Fig. 1, the operator translates the SLA description into SLS parameters, which are mapped onto network specific parameters. Under this scheme, changing the underlying network technology should have minimal impact on the service level descriptors.

The *policy function* is referred to as the middleware that controls network resources and enables to dynamically modify the behavior of the network. A Policy Decision Point (PDP) is a functional entity that acts as an asynchronous event handler (e.g., decide on service resource request) and provides the operator with a scalable way for network configuration and SLA monitoring. A network element

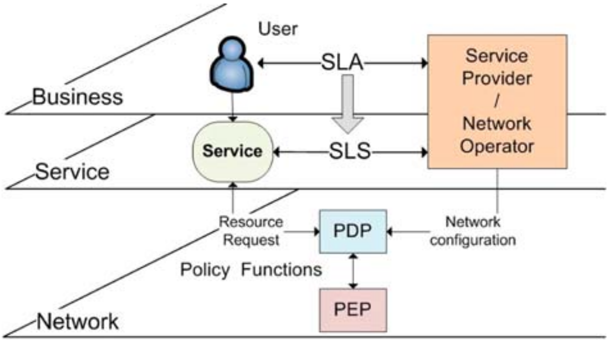


Fig. 1 Complex relations involving the business, service, and network levels can be synthesized as policies enabling the dynamic configuration of network elements to guarantee the delivery of contracted services

interpreting the policies receives the name of Policy Enforcement Point (PEP). *Policy control* is used to describe the process by which a new dynamic service flow or a bearer is created in the transport network upon a resource request. If accepted, a new policy is installed in the PEP.

Basically, a *policy* is a set of rules that governs the behavior of a system. The rules may have a static characteristic (e.g., subscriber's profile) or can be based on dynamic information (e.g., available QoS). Policies are defined by conditions and actions, where conditions are evaluated when triggered by an event. A condition represents the required state that defines whether or not a policy rule should be enforced. For example, a match criterion can be defined on the value of the source or destination IP address, ports, protocol number, link layer information or any data flow characteristics. When the policy conditions are evaluated to true and the other decision strategies or the rule priorities have been considered, the associated actions will be executed, as illustrated by the function diagram presented in Fig. 2.

An action defines what is to be done to enforce a policy rule. In the context of network resource control, an action comprises the execution of one or more operations to manage network traffic and set up network resources. Examples of actions are the configuration of network elements to release user traffic (open gates), guarantee minimum bandwidth or low latency of data flows through the setting of appropriate traffic marking, shaping, dropping, prioritization, etc.

Admission control (AC) ensures the overall performance of the network assuring the QoS of the provided services. A new service provided through the network is only admitted if, based on its requirements and on network resource availability, there is no impact on the QoS of the active services. The *service-based admission control* checks local static policies of the network (e.g., allowed codecs, maximal bandwidth, time limitations) and evaluates also the user profile information (priority, subscribed services and QoS). To guarantee the availability of requested QoS, *resource-based admission control* uses mechanisms such as accounting of active sessions, QoS measurements and reservation methods. QoS resource control can be divided in three logical steps [3]:

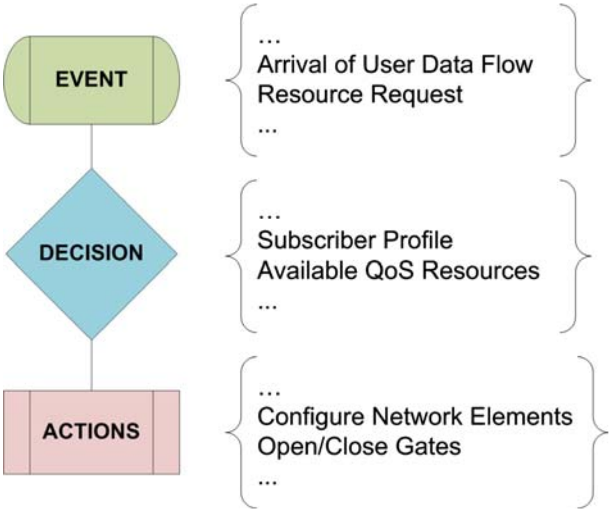


Fig. 2 The abstracted flowchart illustrates how the applied actions upon an event are ruled by a policy decision. This model applies to the field of QoS resource events

- *Authorization*: Service-based admission control performs a checking of static rules defined by the network operator and service provider.
- *Reservation*: Resource-based admission control checks for the availability of the requested resources and, if successful, the QoS resources along the data path are reserved.
- *Commitment*: Reserved resources are enforced and the user traffic is released. This phase ensures that granted QoS is committed and accordingly accounted for charging and usage-metering purposes.

Authorized resources may differ from the effective amount of resources reserved and the ones actually committed that can never exceed the limits set in the reservation and authorization phases. Depending on the service requirements and the capabilities of the transport network, the different phases can be combined into appropriate resource control models.

Summing up, NGN policy-based resource control ensures that different types of traffic (e.g., voice, video, messaging) are given appropriate QoS treatments transparently to the operational aspects of an in-service system easing thus the service set-up and provisioning the network and service delivery elements with the required authorization, billing, metering and control instructions.

2.2 Related Standardization Activities

Back in the late 1990s, the consortium of mobile operators Third Generation Partnership Project (3GPP) started the definition of a new framework called the IMS to deliver IP-based multimedia services. The IMS was based on the standards available from the Internet Engineering Task Force (IETF). By borrowing from the IETF policy model [4], the IMS introduced a new functionality that opened network resource control to applications and was referred to as Service-Based Local Policy (SBLP) [5]. In 2000, Third Generation Partnership Project 2 (3GPP2) adopted the IMS and presented a similar effort on Service-Based Bearer Control (SBBC) [6] for their Code Division Multiple Access (CDMA) domain. The work around the IMS caught attention of the fixed access network standards body of the European Telecommunications Standards Institute (ETSI). ETSI's technical committee Telecoms and Internet converged Services and Protocols for Advanced Networks (TISPAN) adopted the IMS core architecture as the basis for their NGN architecture. The main work was concentrated on the required extensions to convert the IMS into a truly access technology agnostic control architecture. Since 2003, TISPAN has been developing the Resource and Admission Control Subsystem (RACS) [7] that defines a policy control layer to manage resources in the transport network. The International Telecommunication Union Telecommunications Standardization Sector (ITU-T) began in 2004 the development of a Resource Admission Control Function (RACF) [3] for their NGN based on the early work of 3GPP and TISPAN.

Figure 3 illustrates a relevant subset of relationships between SDOs. The mapping of all existing liaisons and collaborations between them would result in a complicated mesh. While 3GPP and 3GPP2 specify the IMS core architecture for respectively the Universal Mobile Telecommunications System (UMTS) and code division multiple access (CDMA) domains, the Open Mobile Alliance (OMA) collaborates with the definition of services and third-party applications (e.g., push-to-talk services). The IP-based interfaces of IMS/NGN are based on the protocol standards supplied by the IETF, a standardization community that is assisting the operator's needs derived from the IMS/NGN adoption. TISPAN and 3GPP are working closely to maintain a single and aligned IMS specification through a

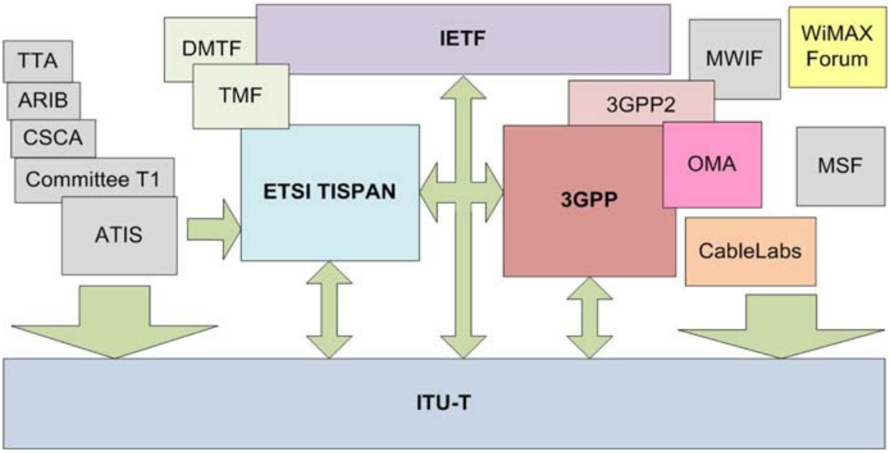


Fig. 3 A close cooperation among the standards development organizations is required to ensure the success of the envisioned NGNs. Only a subset of the involved organizations and its liaisons are pictured in this illustration

combination of liaison statements with other standards bodies and forums. For example, CableLabs received copyright licenses to define its IMS delta specifications for the IP-based standards of the cable television industry. Approved enhancements to IMS are submitted to the 3GPP to be comprised in the new releases.

The Alliance for Telecommunications Industry Solutions (ATIS) contributes with a NGN framework gathering the requirements imposed by North American wired networks. ATIS is working intensively on new NGN topics like IPTV, an area that should be covered in the Release 2 of both TISPAN and ITU-T. Other regional standards organizations passing their requirements to the ITU-T include the Association of Radio Industries and Businesses (ARIB—Japan), the Consortium on Standards & Conformity Assessment (CSCA—China), the Committee T1 (USA) and the Telecommunications Technology Association (TTA—Korea). ITU-T has also liaison statements to a number of standards bodies like the Mobile Wireless Internet Forum (MWIF), Multiservice Switching Forum (MSF), the DSL Forum, the TeleManagement Forum (TMF) and the Distributed Management Task Force (DMTF).

The former ones are working within this area to define a shared understanding of information and data concepts, definitions, and models to enable interoperable policy-based network management solutions. Back in the late 1990s, the IETF defined originally the Directory Enabled Network (DEN) policy model and applied it to the QoS application area. DEN was based on the DMTF Common Information Model (CIM) aiming at providing means of storing information describing the services required by the clients and the capabilities of the devices making up the network. A new generation policy model (DEN-ng) has been worked out by the TMF that defines a Shared Information and Data (SID) model attempting to resolve some of the limitations around DEN and CIM.

The role of the Next Generation Network Global Standards Initiative (NGN-GSI) within ITU-T is to accommodate the international standardization works through liaison discussions with other SDOs looking after a global consensus on the NGN definition. The specifications of the different standardization bodies reference each other when necessary. Therefore, the cooperation among these different standard organizations is a very important issue.

To conclude, the IMS architecture that was originally designed for mobile networks, which introduced Wireless Local Area Network (WLAN) inter-working in Release 6 and incorporated fixed networks support in Release 7, can be considered the trigger of NGN development and the enabler of the Fixed and Mobile Convergence (FMC), one of the key trends of the telecommunications industry today.

3 Policy-Based Resource and Admission Control Architectures

In this section, we describe the principles of the policy-based resource and admission control architecture approaches developed by different SDOs. To simplify the introduction to this field, we shall start by describing the general policy architecture of IETF that underpins the network architectures of the evolving and NGNs standardized by 3GPP, ETSI TISPAN, ITU-T, Cablelabs, the WiMAX and the DSL Forum. These architectures are presented within this section following a chronological and status of completeness order.

3.1 Internet Engineering Task Force (IETF)

RFC 2753 [4] defines a general policy architecture for admission control that includes one Policy Decision Point (PDP) and at least one PEP, depicted in Fig. 4.

The PDP (e.g., a resource manager) is responsible for handling events and making decisions based on those events (e.g., under circumstances x do y) and updating the PEP configuration appropriately. An event can be issued by a PEP under its control or by an external entity (e.g., a network element that queries for network resource availability). As a consequence, the PDP sends a policy-based rule set to be enforced in the PEP (e.g., router, firewall, host). A Local PDP (LPDP) is an optional element that may be used to make decisions based on the policy elements handled locally. A policy repository provides storage and retrieval of policies as well as policy components containing definitional information to be used as part of the policy decision and/or enforcement processes. The framework does not specify any concrete protocol for the interaction between the different elements but rather defines the functional requirements of the policy framework.

The Common Open Policy Service (COPS) [8] protocol was designed to fulfill the requirements of the policy information exchange between the PDP and the PEP. COPS is a request/response protocol that supports both the outsourcing (pull) and the provisioning (push) [9] modes of operation. In the outsourcing mode, the PEP contacts the PDP each time a policy decision is needed. The PDP makes the

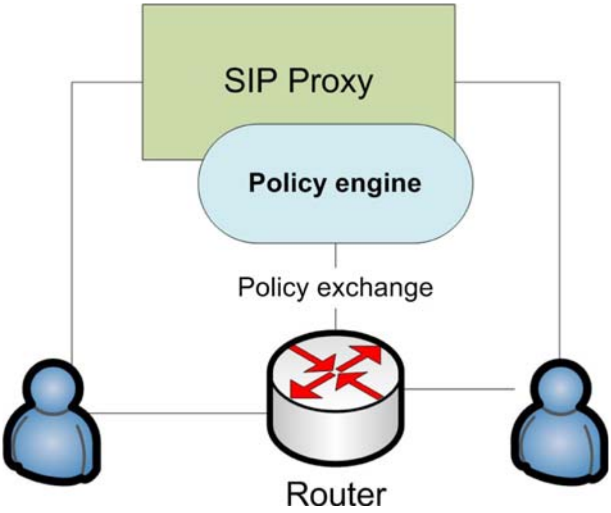


Fig. 4 The IETF general policy architecture [4] specifies the functional requirements of a policy framework for admission control

decision and communicates this information to the PEP for enforcement. In the provisioning mode, the PDP configures the PEP with the enforcement policy, which the PEP stores and uses for current and future decision-making.

3.2 Third Generation Partnership Project (3GPP)

With the definition of the IMS [1], 3GPP was the first standards organization to introduce QoS control policy concepts in the specifications of its evolving access network architecture. The IMS decomposes the network infrastructure into separate functions with standardized “reference points”, that define both the protocol over the interface and the functions amongst which it operates. The IMS architecture defines three main planes or layers, shown in Fig. 5 each of which is described by a number of equivalent names: service or application plane, control or signaling plane, and user, bearer or transport plane. The IMS bases on IETF protocols and it does not standardize network elements but the functionality provided by them. Physical implementations of the functional units are left up to manufacturers. On the same lines, IMS standardizes service enablers (e.g., common session control functions) but not the services themselves.

In a nutshell, at the core of the control plane, the Call Session Control Functions (CSCFs) coordinate with other network elements to control session features like routing, resource allocation or security. The Serving-CSCF (S-CSCF) is essentially

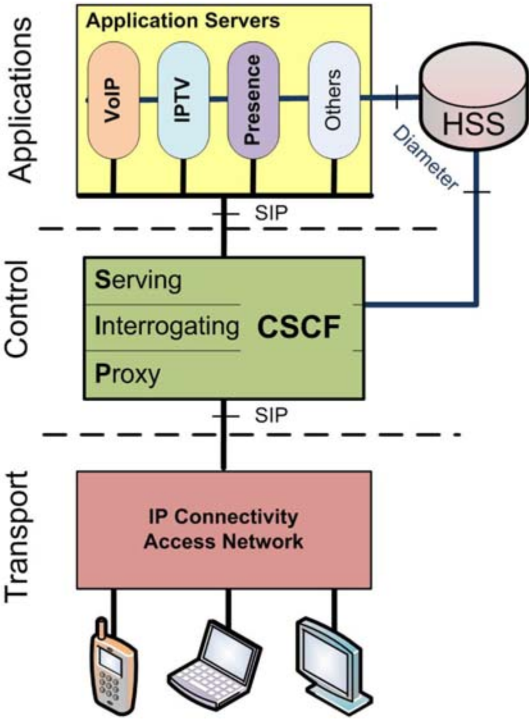


Fig. 5 The IP Multimedia Subsystem operates at the control plane and logically separates services and applications from the IP capable access networks

a Session Initiation Protocol (SIP) [10] server that interacts with the Home Subscriber Server (HSS) to manage user location and for Authentication, Authorization and Accounting (AAA) purposes. The S-CSCF processes the SIP signaling events and forwards them to whatever application servers and/or media gateways are required to complete the service request. Interactions with the transport layer to control the network resources are responsibility of the Proxy-CSCF (P-CSCF).

In general, there are several reasons to include a policy engine for SIP sessions in carrier grade network operator deployments (see simplified policy control for SIP communications in Fig. 6).

First, it is required to deliver services to users without public IP addresses (NAT control). Second, a mechanism is needed to control the transport path between the communicating parties. In absence of any control techniques, end users could terminate a session at the signaling layer (SIP path under operator control) and continue sending media streams directly to each other over their known IP identifiers. Third, it is required to ensure that transport network resources are available before the media session is started. To address these requirements, the IMS specifications extended SIP to integrate policy-based QoS admission control with the IMS session control.

In the IMS architecture, the PDP was instantiated by the so-called Policy Decision Function (PDF) [5]. Basically, the PDF authorizes session resource requests coming from the P-CSCF and issues a media authorization token that is sent back to the user over the SIP signaling path. In turn, the received token is used in the link-layer signaling to authorize the media flows against the Gateway "General Packet Radio Service (GPRS)" Support Node (GGSN) (pull mode). In IMS Release 5 (R5), shown in Fig. 7a, the PDF was co-located with the entry point of the IMS, the P-CSCF, leveraging the SIP proxy to control (open/close gates, NAT) the underlying PEP located at the access network gateway (e.g., GGSN).

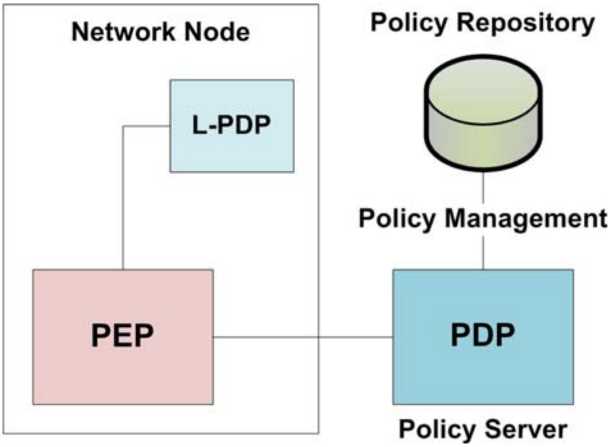


Fig. 6 Policy functions within SIP communications ensure operator control over the multimedia sessions

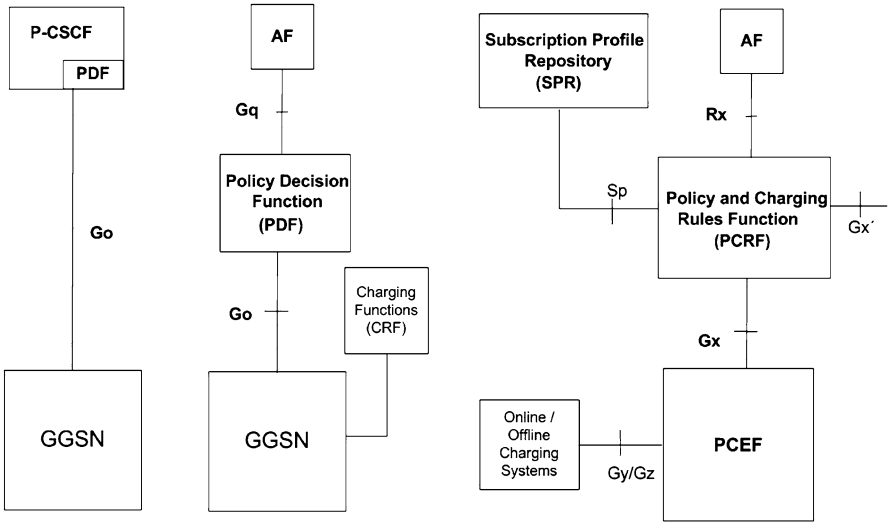

P-CSCF
PDF
Go
GGSN
AF
Gq
Policy Decision Function (PDF)
Go
Charging Functions (CRF)
GGSN
Subscription Profile Repository (SPR)
AF
Rx
Sp
Policy and Charging Rules Function (PCRF)
Gx′
Gx
Online / Offline Charging Systems
Gy/Gz
PCEF

Fig. 7 Evolution of 3GPP policy decision functions: (a) combined P-CSCF/PDF in R5, (b) separated PDF in R6, and (c) enhanced PCRF in R7

Service-Based Local Policy (SBLP) information was transported over the *Go* interface [11] using the COPS protocol. Release 6 of IMS, shown in Fig. 7b, introduced the Diameter-based *Gq* interface that separated the PDF from the P-CSCF enabling one PDF to serve more than one Application Function (AF) and one given AF (e.g., P-CSCF) to interact with different PDFs over a generic IP Connectivity Access Network (IP-CAN). This release also introduced Flow-Based Charging (FBC) to enhance the charging capabilities at the bearer layer. In Release 7, shown in Fig. 7c, the *Go* and *Gq* interfaces and the PDF, respectively, evolved to the *Gx* and *Rx* interfaces and the Policy Charging Rules Function (PCRF) [12]. The PCRF is a functional entity that combines charging rules and policy decision functions using the Diameter protocol. Consequently, the FBC evolved to the Policy and Charging Control (PCC) providing a fine granularity control over charging, QoS policing and firewall functionality on a service flow basis. Service flows are described by an aggregate set of packet flows characterized by identical source and destination IP address and port numbers. The PCC binds transport and service information in such a way that charging and policy are tied together to target heterogeneous transport networks. The Diameter harmonized solution for flow-based PCC is based on the Diameter base protocol [13], the Diameter credit control application and the Network Access Servers Requirements (NASREQ) application. The PCRF authorizes service requests coming from the AF and is responsible for the consistency configuration of the Policy and Charging Enforcement Function (PCEF) nodes allowing for example dynamic selection of charging models even in roaming scenarios (*Gx*′ interface). The PCEF is located in the transport plane and performs service flow detection, charging, gating and QoS management. The Subscription Profile Repository (SPR) is a logical entity that contains subscriber related information needed for access level and charging policing at the PCRF.

The 3GPP2 approach [6] is almost identical to the 3GPP model and is therefore not included in this review since it only differs from 3GPP in some access technology specifics of the CDMA domain.

3.3 ETSI TISPAN

The TISPAN NGN architecture defines a control layer based on the IMS core (among other service subsystems) and introduces at the transport layer the Resource and Admission and Control Subsystem (RACS) [7] and the Network Attachment SubSystem (NASS) [14]. The functional architecture of the RACS approved in Release 1, exhibited in Fig. 8, essentially provides transport control services to higher-level AFs enabling them to request and reserve network resources. The Service-based Policy Decision Function (SPDF) provides AFs with a single point of contact: the inter-domain Diameter-based *Gq′* interface.

Before admitting user traffic, the admission control functionality provided by the RACS performs the following verifications. First, it checks via the Diameter-based *e4* interface the user authentication against the profiles stored in the NASS. Second, the RACS applies operator-specific polices stored in the SPDF, and third it checks resource availability by querying the Access—Resource and Admission and Control Function (A-RACF). The A-RACF bases the admission decision on the QoS resources available in the IP edges under its control. Finally, the policy and admission control decisions are installed to the Resource Control Enforcement Function (RCEF) located at the access network.

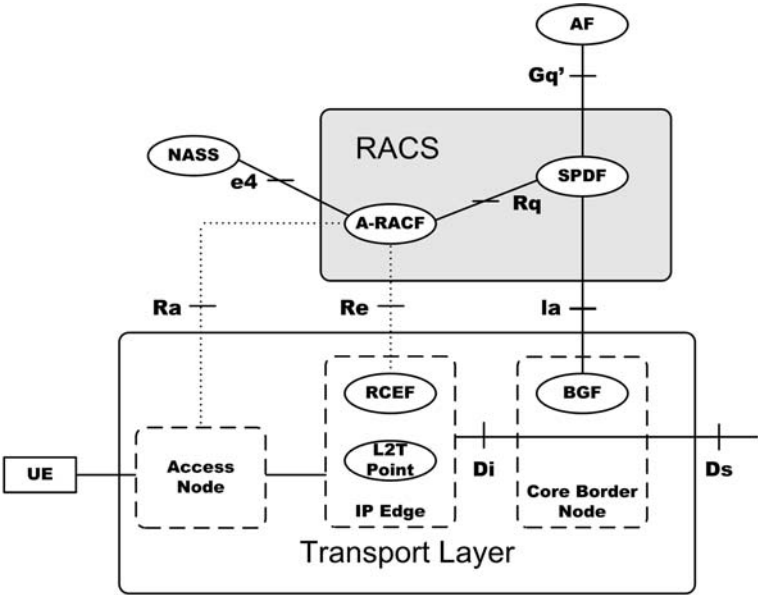


Fig. 8 The functional architecture of ETSI TISPAN RACS comprises resource management and policy decision entities governing transport control functions [7]

The RACS also provides access to services supported by the Border Gateway Function (BGF). The BGF acts as a gateway between different IP transport domains, providing capabilities that include NAT, gate control, marking of outgoing flows, policing of incoming traffic, topology hiding, IPv4/IPv6 interconnection, usage-metering and resource allocation. The SPDF communicates with the BGF using the H.248 protocol [15] over the *Ia* Interface. In turn, the applications may use the transport level information (e.g., NAT bindings) to correct any addressing values embedded within the application signaling, thus enabling Network Address Port Translation (NAPT) traversal for both the service signaling and the media flows. The RACS model provides guaranteed and relative QoS mechanisms. Relative QoS is achieved through packet marking and allows for traffic class differentiation by dynamically applying appropriate QoS at the IP edge and the core border node, while guaranteed QoS defines absolute bounds of QoS parameters (e.g., throughput, jitter, delay) and is achieved via tight traffic control and policing.

3.4 ITU-T

The ITU-T has developed a Resource Admission Control Function (RACF) [3], displayed in Fig. 9, as the central control element to intermediate between the network infrastructure and the Service Control Functions (SCFs), e.g., IMS, IPTV applications, softswitches, etc. The RACF hides service and transport network details from each other and manages QoS resources within access and core networks. The admission control combines transport subscription information of the user, network policy rules, service priority, and transport resource status and utilization information.

The RACF, presented in Fig. 9, consists of a Policy Decision Functional Entity (PD-FE) and a Transport Resource Control Functional Entity (TRC-FE). The PD-FE is application aware and it is responsible for translating upper layer

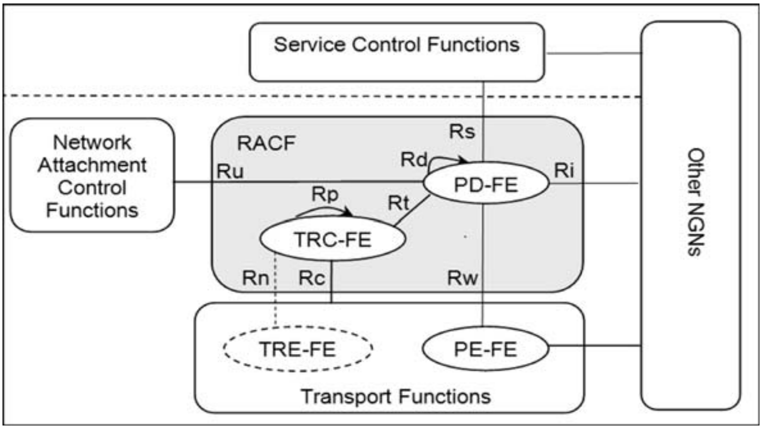


Fig. 9 The ITU RACF [3] arbitrates between the session control functions (e.g. IMS CSCF) and the transport control functions

resources requests into a transport technology independent class of service definition (e.g., BW, delay, jitter, packet loss, etc). The PD-FE queries the TRC-FE to check if the required network resources are available to support the requested service flows. The TRC-FE performs admission control by tracking network resources usage over the network segment topology under its control. The transport resource enforcement function entity (TRE-FE) is dynamically instructed by the TRC-FE to perform polling of network usage, bandwidth reservation and allocation or traffic shaping. To take into account the capabilities of transport networks and the associated transport subscription information of users, the PD-FE communicates with the network attachment control functions (NACF) and checks user related information like network access registration, authentication and authorization, parameters configuration, transport subscription, etc.

The PD-FE pushes service definitions in the form of policy rules to the policy enforcement functional entity (PE-FE) located at the border of the transport elements to accomplish media path functions such as NAT transversal, bandwidth allocation, gate control, QoS marking, rate limiting, usage report, etc.

The RACF supports both the pull (outsourcing) and the push (provisioning) modes of operation, shown in Fig. 10, to control the services provided to a user connected to the network via the customer premise equipment (CPE). The CPE communicates (Step 1) with the service control layer that triggers the resource request (Step 2) to the RACF.

In the push model, the RACF authorizes the resources request and pushes the policy rules down (3) to the transport functions (TRE-FE, PE-FE) that enforce the policy decision (e.g., resource reservation and commitment rules). The push model is suitable for CPE that are unaware of the transport-specific QoS attributes of the network and have only service-layer QoS negotiation capability (e.g., IMS clients with SDP/SIP QoS extensions).

In the pull model, the RACF authorizes the resource request and generates an authorization token that is passed back to the issuer (3). In turn, the user equipment embeds the authorization information in the data path QoS signaling (4). The transport functions use the token (or other unique flow identifier) to request the RACF for resource re-authorization (5). Finally, the RACF responds with the final policy decision (6) for enforcement. This model requires QoS negotiation support at the transport stratum like Resource ReSerVation Protocol (RSVP), or link layer QoS signaling capabilities (e.g., as available in UMTS or IEEE 802.16).

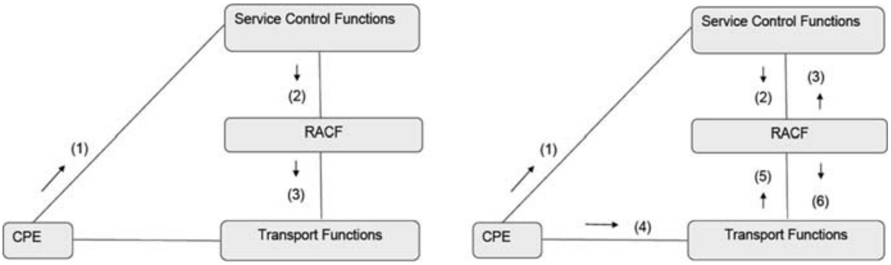


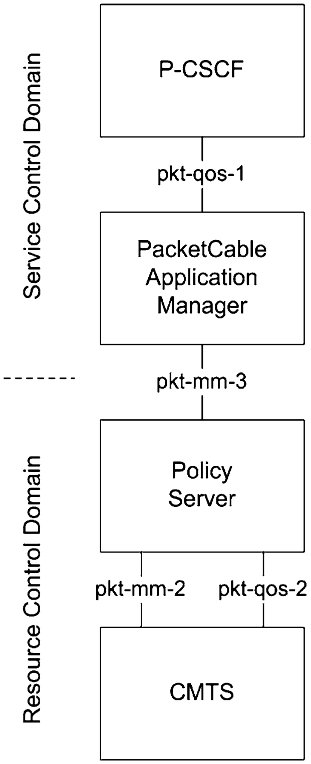
Fig. 10 Flowcharts of the push and pull modes of operations as per ITU-T RACF [3]

3.5 Cable Labs

The PacketCable multimedia architecture [16], shown in Fig. 11, includes a policy-based control layer and support for multiple applications and services including the IMS since Release 2.0. The PacketCable Application Manager (PAM) maintains an application's session-level state and enforces any Service Control Domain (SCD) policies against network resource requests via an application server (e.g., P-CSCF).

The newly defined pkt-qos-1 interface between the IMS P-CSCF and the PAM is based on Web Services [17] and conveys session-level QoS information (e.g., extracted from the SIP signaling). The PAM receives the indication whether to reserve or commit the resources for the session (single-phase and two-phase commit operation modes are supported). The PAM applies SCD policies (e.g., user is authorized for the requested service) and translates the session needs into PacketCable multimedia requests over the COPS profile of the *pkt-mm-3* interface [18] to the Policy Server (PS). The PS performs Resource Control Domain (RCD) policy check (e.g., requested resources within limits, appropriate scheduling type for the service, etc.) ensuring the request meets the network-based policies. After passing PS checks, these requests are forwarded to the Cable Modem Termination System (CMTS) for action via the COPS-based *pkt-mm-2* interface [18]. Upon receipt of the resource request, the CMTS is responsible for admission control and resource allocation. This process ensures that the CMTS has adequate resources to support the service flow request. Finally, the

Fig. 11 The policy-based resource control approach of the PacketCable Multimedia architecture addresses next generation networking requirements



CMTS installs the necessary flows and notifies the cable modem serving the user equipment. The pull mode of operation is also supported, in which policy authorization decisions are requested on demand by the CMTS using COPS.

3.6 WiMAX Forum Network Working Group

In March 2005, the WiMAX Forum formed the Network Working Group (NWG) to develop technical specifications beyond what is defined in the scope of IEEE 802.16. While the IEEE QoS framework focuses on the radio link connection-oriented services (QoS classes, admission policies for new service flows) the NWG is working on the WiMAX network integration in an all-IP (e.g., IMS-based) architecture. As shown in Fig. 12, recent standardization efforts define a service flow QoS framework [19] that is basically composed by a Service Flow Manager (SFM), a Service Flow Authentication (SFA) and a Policy Function (PF).

In the pre-provisioned (push) functional model, a Mobile Station (MS) communicates at application layer (e.g., using SIP) with the AF (e.g., P-CSCF).

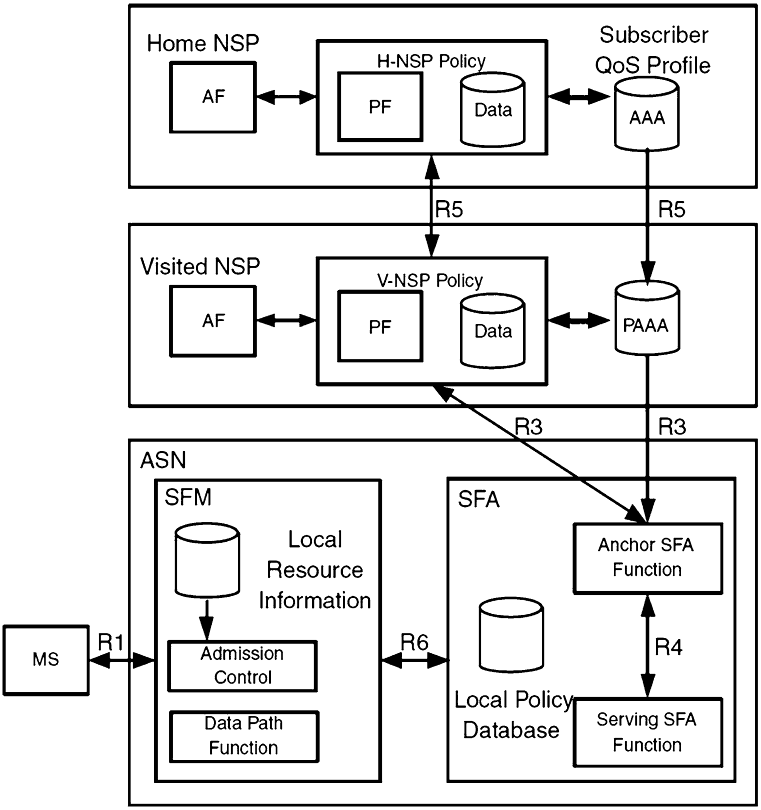


Fig. 12 WiMAX Forum QoS policy framework [19] follows the NGN functional principles for resource management

As a result, the AF triggers the PF that checks the WiMAX service flow request against Network Service Provider (NSP) policies. Managed information in the PF includes NSP's general policy and application dependent rules. User QoS profiles can be provisioned by the AAA infrastructure and may include specific information of the Medium Access Control (MAC) connections at the air interface. Based on SLAs, the provisioned profiles may include user priority to enforce relative precedence in terms of access to radio resources. After policy checks, the PF sends the service flow establishment request to the SFA. The SFA is likely to be deployed on the Access Service Network (ASN) gateway and translates the WiMAX service description into appropriate IEEE 802.16 format and service profiles. Finally, the service request reaches the SFM managing the wireless connections to the MS. The SFM is a logical entity located in the base station of the ASN responsible for the creation, admission, activation, modification and deletion of 802.16 service flows. However, the precise specification of the Admission Control (AC) module functions is left up to implementers.

In the pull model, dynamic service flow triggers can include user-initiated link layer IEEE 802.16 signaling triggers as well as network layer QoS signaling triggers like RSVP. The SFA serving the MS can decide on service requests based on a local PF (pre-provisioned during MS network attachment) or can delegate the admission decision to the PF at NSP level. In roaming scenarios, the visited PF forwards the request to the home network's PF.

The NWG QoS framework is almost at an initial stage. Current release defines mainly high-level requirements and most of the reference points are still unspecified [19].

3.7 DSL Forum

The DSL Forum specifies the required architecture enhancements regarding the evolution of new services (e.g., multicast services for IPTV) and the integration into the NGN architectures. Recently, standardization works started to leverage the DSL architecture by specifying a "Policy Control Framework for DSL" [20] that provides a policy control layer to manage the transport resources. With relation to the IETF approach the Broadband Remote Access Server (BRAS) and the Routing Gateway (RG) are comparable with the PEP, as shown in Fig. 13. The policy-based IP QoS framework aims to provide dynamic session control and real-time network control as well as support for network resource requests coming from application functions, such as IPTV services or the IMS. The specifications [20] are at a very early stage and no explicit protocol or interfaces have been defined. Thus, the DSL architecture is left out of the following comparative analysis.

4 Comparison of the Policy-Based Architectures

The IETF specification of a general framework [4] for policy-based admission control inspired 3GPP's IMS service-based local policy control, which in turn, was the basis for the resource management functions of other SDOs, such as ETSI

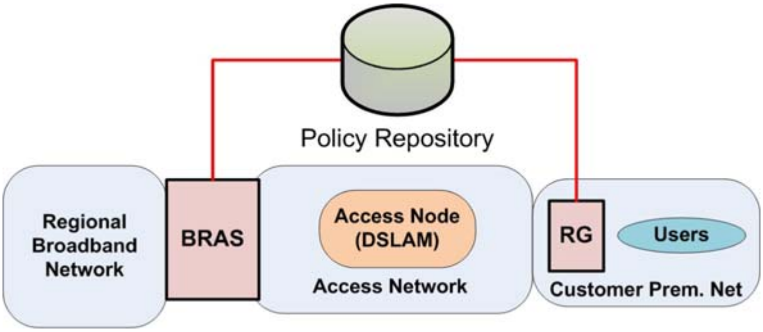


Fig. 13 Policy control functionality under specification leverages DSL-based access networks to conform NGN requirements

TISPAN and ITU-T. Thus, the emerging architectures present significant commonality and share similar objectives in accordance to the principles of next generation networking. First, a provision of a QoS control framework that decides on incoming service requests taking into consideration not only transport resource availability but also policies that combine service, user and network operator requirements. Second, border control (e.g., NAT, gating, traffic policing) is opened to applications and service control functions.

4.1 Architectural Alignment

In an attempt to better present the similarities observed in the reviewed architectures, we propose a simplified PDP-PEP model, presented in Fig. 14, that provides the required level of abstraction when referring to the resource control and transport functions. The policy-based resource management functions of the different SDOs act all as a decision point for resource requests and can be thus grouped into a generalized PDP function. Similarly, the transport related functions are clustered into a generalized PEP controlled by the PDP. The resource and admission control functions of the PDP serve as an arbitrator between application functions (e.g., NGN service control functions) and the transport functions making them to work and evolve independently. Table 1 presents the interface nomenclature of the reviewed SDO architectures mapped to the harmonized PDP-PEP model.

A detailed review of the associated network attachment functions of the different SDOs architectures has been left out of this work. However, a similar alignment is present in the functionalities of the Subscription Profile Repository (SPR) of 3GPP, the NACF of ITU-T and the NASS of TISPAN. Common network attachment functions include subscriber location management, dynamic provision of IP address and user access network authentication and authorization. They also provide means for the configuration of customer equipment and access network elements based on transport related user profiles and other network configuration parameters. These functions are commonly interfaced by the respective resource management functions to complete the resource control and authorization process.

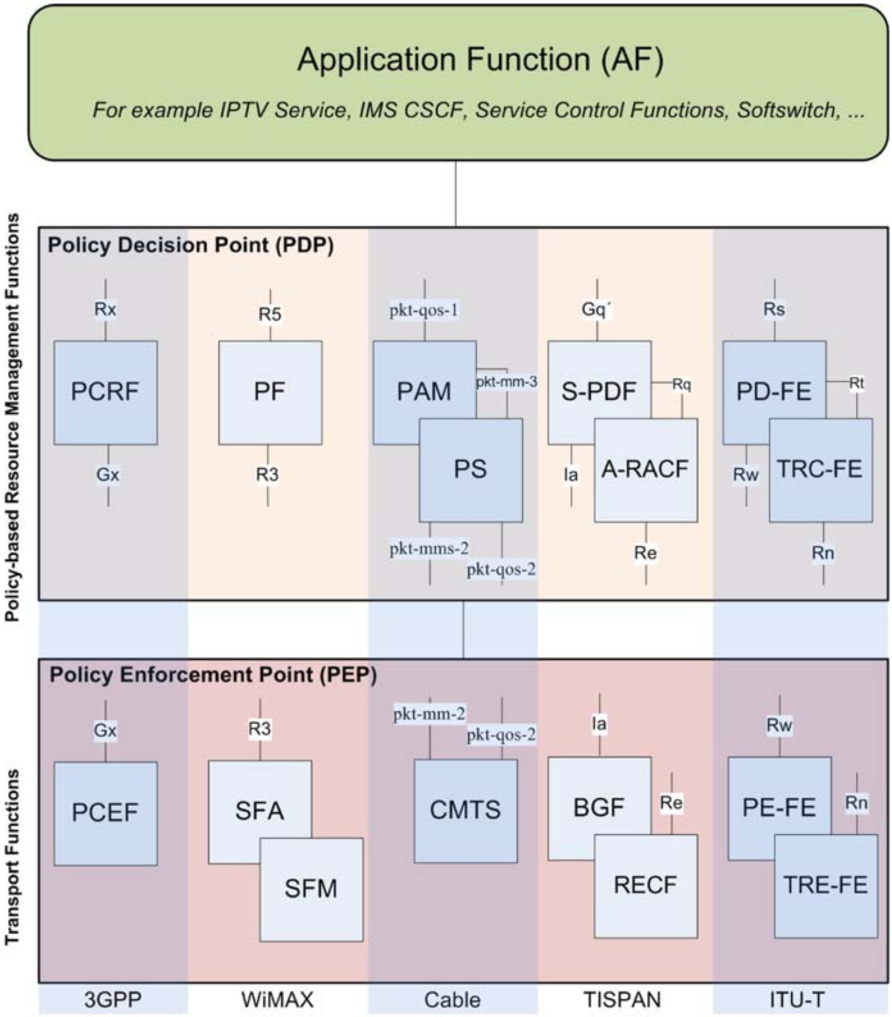

Application Function (AF)
For example IPTV Service, IMS CSCF, Service Control Functions, Softswitch, ...
Policy-based Resource Management Functions
Policy Decision Point (PDP)
Rx
PCRF
Gx
R5
PF
R3
pkt-qos-1
PAM
pkt-mm-3
PS
pkt-mms-2
pkt-qos-2
Gq´
S-PDF
Rq
Ia
A-RACF
Re
Rs
PD-FE
Rt
Rw
TRC-FE
Rn
Transport Functions
Policy Enforcement Point (PEP)
Gx
PCEF
R3
SFA
SFM
pkt-mm-2
pkt-qos-2
CMTS
Ia
BGF
Re
RECF
Rw
PE-FE
Rn
TRE-FE
3GPP
WiMAX
Cable
TISPAN
ITU-T

Fig. 14 The generalized PDP and PEP functional entities illustrate the observed commonalities in the evolving resource management approaches of different SDOs

4.2 Scope and Functional Divergences

While the access network SDOs focus on the control of technology-specific QoS resources within the access network, NGN SDOs consider also the aggregation network and the interconnection to the network core in multiple transport technology deployments. Thus, the use of two policy decision entities in the NGN architectures, namely the SPDF/PD-FE and the A-RACF/TRC-FE is justified by the need of a scalable solution in multi-network and multi-domain scenarios. The intra-domain transport functions can be distributed for efficient resource control over different network segments. In terms of intra-domain policy decision, a

Table 1 Interface description and examples of information exchanges in the simplified model for next generation policy-based resource control

Interface	Description	Information exchanged
AF-PDP	Application level session-based resource authorization requests	Session identifier, e.g., *call-id*; Media flows, e.g., *IPs, ports, protocol, codec, direction, UL/DL*; Service profile, e.g., *conversational, best effort, streaming video*; Priority, e.g., *low, medium, high*; Authorization token, e.g., *128 bit value*, etc.
PDP-PEP	Used to carry transport level information, such as QoS and charging policies, transport service request, or polling of network resources and topology information	Data flow identifier, e.g. *connection id*; Charging policy, e.g., *cost-, unit-, credit-information*; QoS policy, e.g., *marking and shaping rules*; QoS parameters, e.g., *TMOD, DSCP, Y.1541, DS-TE*; Bandwidth, e.g., *authorized/committed/available bps*; Network topology, e.g., *IP routing, MPLS tunnels, layer 2 info*; Gate, e.g., *open/close ports*; NAT control, e.g. *set in/out IP:ports*; etc.

network provider can trust its own policy information delivered from its own network architecture. However, in terms of inter-domain policy decision a network provider should be able to evaluate the policy information delivered by the other provider before affecting its transport plane.

The RACF and the RACS specify similar, but not identical functional entities and reference points. The ITU-T RACF has a wider scope that includes also core network control and inter-domain communications. Going more into detail in the specifications, the RACF supports both the push and pull models for policy installation, whereas the current release of the RACS supports only the push model. Though, the Release 2 of ETSI RACS is expected to support also the pull model and support for inter-domain communications. Furthermore, the ITU RACF contemplates a more flexible firewall working mode selection capability that enables varied levels of security strength. The RACF has the broadest view of the current standards; however, it is still a NGN vision and has yet not been fully realized.

4.3 Protocol Choice

Table 2 gathers the interface names and the functional elements of the different SDO architectures mapped to the simplified AF-PDP-PEP model. Regarding the protocol choice there is an alignment on Diameter for the interface between the AF and the PDP (*Rx*, *Gq′*, *Rs*). The protocol diversity on the PDP-PEP interface is justified by the constraints of already deployed transport network infrastructures and the broad scope of next generation resource management. While the PDP-PEP interfaces of evolving access networks (*Gx*, *pkt-mm-2*) are meant to perform resource control and its associated policies, the additional interfaces (*Re, Rn, Ra,*

Table 2 Interface names and, where already specified, chosen protocols for the resource control functions of the next generation SDO architectures

IETF	3GPP R5/6	3GPP R7	WiMAX	Cable Labs	ETSI TISPAN	ITU-T
PDP	PDF	PCRF	PF	PAM PS	RACS (S-PDF, A-RACF)	RACF (PD-FE, TRC-FE)
PEP	PEP	PCEF	SFA SFM	CMTS	BGF RECF	PE-FE TRE-FE
AF—PDP	Gq <Diameter>	Rx <Diameter>	R5 <unspec.>	pkt-qos-1 <WebServices>	Gq′ <Diameter>	Rs <Diameter>
PDP—PEP	Go <COPS>	Gx <Diameter>	R3 <unspec.>	pkt-qos-2 <CPD> pkt-mm-2 <COPS>	Ia <H.248> Rq <Diameter> (Re, Ra) <unspec.>	Rw <unspec.> Rt <unspec.> (Rn, Rc) <unspec.>
Push mode	No	No	Yes	Yes	Yes	Yes
Pull mode	Yes	Yes	Yes	Yes	No (R1)/Yes (R2)	Yes

Rc) in the NGN architectures are defined to collect transport network information (e.g., network topology, usage, state).

COPS has been the first approach used by 3GPP to exchange policy information, but it showed some limitations in roaming scenarios and regarding its extensibility to carry QoS and charging information. Other protocols like Media Gateway Control (Megaco)/H.248, Remote Authentication Dial-In User Service (RADIUS) [21], Network Management Protocol (SNMP) [22], and Diameter [13] have evolved to be used for the same purposes. Even though RADIUS is limited in its amount of functionalities, it is a widely distributed AAA protocol. Its extended deployment explains the effort and sense of current protocol extensions of RADIUS to enable novel network policy concepts. Different from the traditional framework of network access control and resource management (e.g., RADIUS, SNMP), Web Services [17] uses the Simple Object Access Protocol (SOAP) and the Extensible Markup Language (XML) for communications between different network elements in a flexible way. Diameter is the successor AAA protocol of the RADIUS protocol and enables easy protocol extensions due to its protocol design. Diameter applications can be defined to meet SDO specific requirements such as QoS and charging. In addition to its inherent AAA capabilities, the main advantages of Diameter include domain oriented routing capabilities, which suit it exceptionally for roaming scenarios. The Diameter-based PCC [12] offers an enhanced granularity of the classification and population of policies. However, the actual Diameter implementation in the PCC is basically a 3GPP specific application. A Diameter QoS application is being specified by IETF [23] and could be the harmonizing basis for NGN Diameter-based resource control.

In order to accommodate all the requirements of the ambitious deployment scenarios targeted by the ITU-T NGN, it is expected that more than one protocol will be supported for some reference points. Protocol development is the final stage of standards development after identification of the requirements, architecture, services, etc. Diameter is one of the alternatives for the *Rw* interface among COPS and H.248. Protocol alternatives for the *Rc* interface to collect transport network topology and resource status information include COPS, SNMP and SOAP. Having several protocols for a reference point requires further study on how these protocols will be consistent with each other in terms of information semantics. This and other issues are explored in the next section.

5 Challenges and Related Ongoing Work

Even though the standardization works show a general agreement on the need for some mechanisms to control the allocation of resources, a harmonization on the resource and admission functions is still far from be achieved. This section explores the main issues around the resource management entity (synthesized as PDP to be consistent with the presented generalized model) regarding its use in intra-domain and inter-domain environments over heterogeneous transport technologies. In a convergent environment, the different network segments (access, aggregation, core network, etc.) can be operated by different owned domains and potentially use different transport technologies. The support for such deployment scenarios is called the multi-network/multi-domain awareness. To fulfill this requirement, we have identified two types of coordination that the resource control functions need to achieve, as shown in Fig. 15. First, a vertical coordination is required to reconcile the inherent semantic differences between the service and the transport planes. Second, a horizontal coordination is needed to achieve end-to-end resource control across heterogeneous domains.

While NGN application functions may operate over different domains, resource related functions belong to domains owned and controlled by different network operators. The scenario with transit domains interconnecting destination and terminating domains is not shown in Fig. 15. The coordination challenges and issues explored in this section include inter-domain signaling, admission control performance, support for mobility and the industry initiatives towards inter-operable NGNs.

5.1 Vertical Coordination

By vertical coordination we understand the capability of providing the correct binding of the inherent different service and transport level semantics. To achieve this, a consistent multilayer policy methodology and appropriate QoS translation policies have to be worked out.

5.1.1 Multilayer Policing

In the decoupled NGN architecture policy control can be implemented at several layers and needs to satisfy both the business and QoS needs of the network operators:

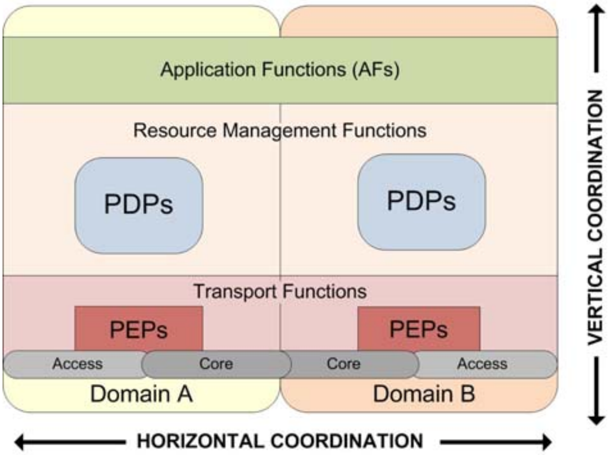


Fig. 15 Suitable horizontal and vertical coordination schemes between the different layers and policy elements are required to provide successfull NGN end-to-end resource management

- *Application*: Services may employ policies to constrain the use of a service based on subscription or other application-level information. Such policies are stored at the service profiles of a user (e.g., in the HSS) and they are accessed by the application functions and configured by the business and operations systems support of the service provider.
- *Control*: Network-based restrictions on the session signaling may be enforced. For example, the service control functions (e.g., CSCFs) can check the media parameters contained in the session signaling (e.g., SDP contents) and modify or deny the session request (e.g., SIP message).
- *Transport*: Different functional entities in the bearer network may perform policy control:
  - The PDP managing the resources may receive requests from multiple sources. Policies are expected to optimize the use of network resources between multiple applications and traffic types.
  - PEPs implement admission control based on policies that take into account the allocation of resources among various types of traffic based on the service class and on the applied authorization model.
  - Network attachment functions manage subscription-based transport related information that includes user access network authentication and authorization policies to control the configuration of user equipment and access network elements.

In some cases, equivalent policy decisions can be made at more than one level in the network. The choice of on which level to implement a given policy can be based on the availability of the required information and the ease of implementing policy control at that level. Moreover, policies need to be consistently distributed among the network elements to achieve a maximal performance when performing policy control particular network deployment. As an example, it is more efficient to let the admission control detect at the control plane (e.g., P-CSCF) that certain media codecs (e.g., high video quality) outbound the capabilities of an access network (e.g., low-band cellular network) than allowing resource authorization requests reach the PDP resulting in signaling overhead to check for resource availability and potential unnecessary resource reservations.

5.1.2 QoS Translation

The PDP performing resource control handles both application-driven (AFs) and transport network-driven (PEPs) QoS resource requests. In order to accommodate the heterogeneous QoS descriptors of each layer, the following QoS mapping schemes should be carefully considered:

- *Technology independent QoS translation (AF-PDP)*: The service descriptors (type, QoS parameters, priority) received from the AF is mapped to a common language network QoS parameters. A common set of network parameters and traffic classes is required to provide a harmonizing basis to perform effective and flexible admission control of QoS resource requests coming from varying sources, e.g., different PEPs and AFs. For example, the QoS specification (QSPEC) [24] traffic model (TMOD) parameter is a mathematically complete way to describe a traffic source (data rate, bucket size, peak rate, minimum policed unit). QSPEC defines additional constraints parameters to fully characterize a network path (e.g., accumulated latency, jitter, packet loss rate, packet error rate).
- *Technology dependent QoS translation (PDP-PEP)*: Common network QoS parameters and classes are mapped to transport specific QoS parameters and traffic classes. A pre-defined static transport policy rule determines how the network QoS parameters are best matched to technology dependent QoS parameters for a given transport technology. Network specific parameters interpreted by the PEP can include traffic classifiers (e.g., RSVP parameters, 3GPP classes, Y.1541 QoS class) or link layer QoS information (e.g., UMTS, 802.1p priority values).

Consistent policy rules defining the QoS mapping are claimed for to accommodate heterogeneous resource requests, including varying applications and types of transport networks. It is for further study, what service and transport QoS related information should be standardized. Standardization efforts are constrained by the tradeoff between simplifying the interfaces to the PDPs and the benefits of providing an accurate QoS template. Furthermore, the translation of complex, highly interactive rich multimedia service requirements (e.g., involving multiple flows and media codecs) into efficient aggregated QoS resource requests is an ongoing research area. A consistent QoS translation scheme that uses general parameters with common definitions across all QoS control services is a prerequisite to achieve high performance of the admission control mechanisms and to enable a transparent interaction between heterogeneous network segments and domains. This horizontal interaction issue is further explored in the following sections.

5.2 Horizontal Coordination

By horizontal coordination we mean the ability to provide seamless end-to-end resource control functionality. Many research efforts have been and are being done to provide end-to-end QoS control. However, only parts of the researched network engineering approaches are applicable to the operator driven NGN environment. SDOs are working on the functional architectures and reference points that ensure the inter-operability of the resource control functions. Current status of these works is the careful evaluation of suitable approaches for end-to-end QoS control constrained by the multiple deployments scenarios being targeted.

When interconnecting operator-controlled domains, the business interests play an important role and often compete against the technical realization of inter-domain resource signaling. End-to-end QoS support is a broad topic that we cannot expect to cover in one section. Thus, we will only explore some of the design principles and challenges to achieve inter-domain (horizontal) coordination in NGN architectures as indicated in Fig. 16. Issues highlighted in the following sections include the decoupled levels of end-to-end signaling, the distribution of admission control functionality and the aggregation techniques.

5.2.1 Multilayer Signaling

Similar to the multilayer policing challenge, the provision of end-to-end aware QoS communications requires an horizontal coordination scheme that consistently handles the signaling capabilities at the decoupled levels:

- *Service control signaling*: The common session control layer provides the required information (end-points identifiers, QoS, etc.) to the appropriate PDPs handling the network resources in each domain. However, resources can be only

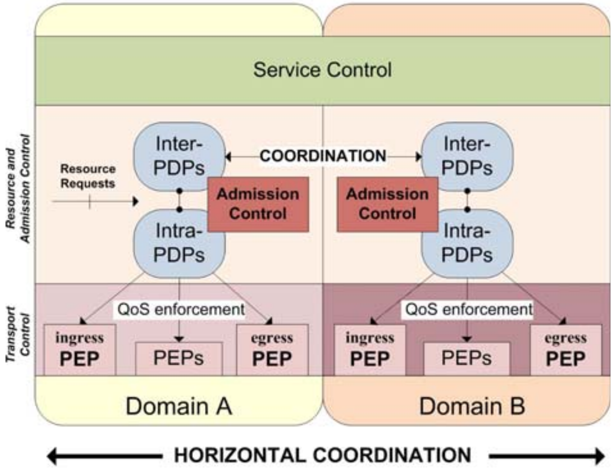


Fig. 16 An efficient horizontal coordination scheme is required to ensure end-to-end resource control across next generation admission-controlled domains using heterogeneous transport technologies

requested at the originating and terminating domains along the session control signaling path. No resource control can be provided on the data plane of transit domains alone from the information available at this level.

- *Transport level signaling*: Pure transport QoS signaling approach requires hop-by-hop support across the data path to dynamically perform explicit QoS resource reservations (e.g., RSVP, Next Steps in Signaling (NSIS)). When interconnecting heterogeneous domains a transport QoS signaling adaptation has to be done probably involving the functions provided by resource control layer.
- *Resource control signaling*: The standardization trends have shown that every domain and even network segments will implement some type of resource control layer. Consequently, it makes sense to exploit the common semantics under specification for the PDP functions and provide end-to-end QoS control at this level.

The vertical coordination between the decoupled layers becomes even more indispensable to ensure a QoS aware end-to-end communication for services provided intra- and inter-domain. Based on both scenarios, we can define two different types of PDPs. First, the intra-domain PDP that manages resources within one single domain and second, the inter-domain PDP that enables resource and admission control across different domains, as shown in Fig. 16.

5.2.2 Distributed Admission Control

A design principle of the PDP functionality in NGN is the ability to be distributed over an arbitrary number of elements (and physical devices) to provide a scalable solution for network resource management (e.g., arbitrary relationship between TRC-FEs and PD-FEs in the ITU model). Admission control can be performed for each of the different network segments forming thus a series of admission-controlled domains. The PDP receiving the admission request returns an answer on behalf of the network topology it controls.

A complex issue is the standardization of the inter-domain signaling protocol to collect the responses from the several admission-controlled domains. Next Steps in Signaling (NSIS) [25] is a candidate signaling protocol for end-to-end QoS control and a first draft on how to fit in the ITU RACF functional architecture has been already proposed in [26]. A more detailed discussion on the different distributed approaches to allocate the resources across different domains and coordinate the inter-domain communications is provided in [27].

5.2.3 Admission-Controlled Domains

Each admission domain is free to decide how to implement the AC related functions:

- Enforce the QoS of the admitted flows by traffic engineering mechanisms of its choice (e.g., as per [28]).
- Implement AC based on the policies and resource management algorithms that ensure the overall performance of the domain.

- Compute the data path across the network topology under its control that has the capacity to carry the requested flow with the required QoS. The Inter-PDP decides and provides the routing information of the ingress/egress PEPs (located at the domain edges) supporting the requested services (e.g., as per [29]).

Preserving internal network details (e.g., topology hiding) is a critical aspect in inter-domain communications. In this context, virtual topology approaches can be used as proposed in [30]. Another challenging task is to decide on the deployment scenario for the Intra-PDP functionality that can follow a fully centralized, fully distributed or hybrid approach. The different approaches present coordination issues and tradeoffs regarding scalability, resilience and signaling overhead. A thorough discussion on this topic is available in [26]. The most suitable AC solution strongly depends on the characteristics of the domain (size, network technology, QoS mechanisms) and the requirements of the services to be offered.

5.2.4 Inter-PDP Path Computation

Before the inter-domain service requests can be triggered and transverse the admission controllers of the different domains, the inter-domain signaling protocol needs to know the Inter-PDP route. It is an open issue to determine how inter-domain PDPs can be actively discovered. Therefore, a data path computation has to be performed and each transited domain has to implement a mechanism to advertise the Inter-PDP performing admission control and the suitable ingress/egress routers (PEPs). Appropriate border gateway selection algorithms have been studied in [29]. The provision of a mechanism for dynamic Inter-PDP discovery is still an open issue and requires interaction with the inter-domain routing protocol (e.g., Boarder Gateway Protocol (BGP)). Additionally, means to exchange capabilities between the Inter-PDPs and to assess the resource control layer semantics are also desirable functionalities.

5.2.5 Aggregation

End-to-end service requests, as discussed before, are carried out in a per-flow (per-session) basis. However, this type of requests do not scale when they reach the PDP controlling the resources at the core network. Thus, PEPs at the aggregation networks should provide means to aggregate the requests in order to reduce signaling and computation overhead at the PDPs. However, aggregating resource requests is a challenging issue. It requires a powerful algorithm to optimize the amount and timing of the resources that should be pre-allocated to perform efficient aggregation. Regarding this aspect, [31] provides an extended discussion on related issues like aggregation areas identification, routing, signaling requirements, marking, etc.

5.3 Admission Control Performance

The overall performance of the admission control mechanisms is highly dependent on the effectiveness of the QoS mapping and the successful coordination among layers and network domains. As already outlined, the distribution and location [32] of the multiple PDPs and the coordination schemes (both intra-domain and inter-domain) heavily affects the admission control performance. Furthermore, the multiple options to combine the resource control logical steps (authorization, reservation and commitment) in the push and pull models require further studies considering individual service requirements (e.g., session setup delay) and the specifics of the transport network and its resource control mechanisms.

The richer and more accurate the QoS description of the requested resource is the better the admission control can derive the transport level requirements to support the demanded service. Under the push model, QoS information at the service layer should be sufficiently detailed to enable an appropriate mapping on final transport technology parameters. In the pull model, the upper level QoS information can be enriched with QoS descriptors within the transport signaling protocol. However, the user equipment can be limited in its capabilities to provide the QoS information. Increasing the space of QoS descriptors complicates the QoS translation and requires additional standardization efforts.

So far, the standardization works on service admission control takes only into consideration media resource availability. Additional measures, such as the overload indication of signaling resources, may be needed to ensure that the desired performance of service control signaling, such as certain session setup or resource priority signaling can be met. As presented next, admission control performance is further challenged when applied under mobility conditions.

5.4 Mobility Support

When mobility comes into play, the horizontal and vertical coordination schemes are further challenged. Up to now, proprietary, usually link layer, solutions such as the employed in cellular networks have provided full mobility support within a network technology. In an NGN scenario, mobility has to be addressed at higher layers common to the involved operator-controlled networks. IETF IP mobility protocols, such as Mobile IP (MIP) have been successfully deployed to handle inter-domain mobility and further enhancements are under development to fit in all-IP scenarios. However, further work is required to deal with the mobility of running policies (network and user oriented) across subscribed networks [33]. A user changing its point of attachment to the network (may be also the change of access technology) within a session further justifies the requirement of efficient coordination schemes (horizontal and vertical) and continuous support for running policies. Moreover, the requirement of fast handover mechanisms between different network domains arises to provide fast user authentication and authorization for network access. Based on this, seamless network handover as well as uninterrupted service provisioning to users require high performance inter-domain signaling and admission control mechanisms that satisfy the user's experience and the provider requirements.

5.5 Advances Towards an Interoperable NGN

A group of five major equipment vendors and one operator announced in July 2006 their common goal in developing enhancements to the IMS architecture called advances to IMS (A-IMS) [34]. A-IMS can be seen as an IMS overlay initiative that tries to speed up real-life development of IMS systems by focusing main issues like support of non-SIP applications and the ease of interconnection to other networks and service operators. One key feature, apart from the simplified architecture, is the inclusion of a Policy Manager (PM) upon the standard IMS policy functional elements. The PM is not limited only to QoS and accounting control, but also scopes packet flow optimization, mobility, access control, binding network conditions to the behavior of applications, etc. The PM is flexible enough to support device-specific or application-specific interfaces. In case of roaming, a better support is achieved through the peering between home and visited policy servers.

On one hand, A-IMS demonstrates the support on the achieved IMS standards but, and on the other hand, this movement violates the standardization practices and can be interpreted as a consequence of uncovered real world deployment issues and the slow standardization processes. However, it is expected that the A-IMS 300-page concept and architecture document to be put forth to the appropriate standards bodies.

Continuing with the focus on pragmatics like interoperability for turning the NGN framework into reality, the physical implementation specifications provided by the MSF [35] enable equipment manufactures to proof its interoperability in real-world trials. The DMTF and TMF are working together towards convergent information modeling proposals to enable interoperable policy-based network management deployments. An additional pragmatic step is the ETSI IMS/NGN performance benchmark [36] that will give the operators one more data point to understand the impacts of an IMS/NGN deployment.

6 Conclusion

The next generation policy practices for resource control layer tie together the subscriber, the services and the network infrastructure and are thus considered an essential step towards true FMC. Policy-based resource management functions provide the network layer with the intelligence to manage resource availability in real time enabling a high quality and efficient roll out of new services over varying underlying transport technologies.

Standards development organizations are all working on policy-based resource control architectures to adapt to the NGN principles of broadband service delivery. The different SDOs participating in the specification and evolution of access and NGNs have introduced extensive proprietary nomenclatures to describe similar functionalities. This review shows the relation and similarities between the functional and architectural approaches of 3GPP, ETSI, ITU-T, WiMAX Forum and CableLabs. We have presented a simplified model that illustrates the observed commonalities and provides the required abstraction to understand the common challenges of next generation resource control. Issues and related work highlighted in this work include the coordination challenges among the decoupled layers (service, resource, transport) in NGN and the interconnection of multi-technology domains.

Besides overcoming the QoS related challenges, policy technology promises a cost-effective and technical efficient solution for next generation networking that optimizes and eases the management of network resources. Operation and management expenses costs are a major concern of operators and service providers who want to optimize the network resource investments. All these factors are expected to push the activities within and between SDOs to continue the network architecture evolution towards interoperable resource control solutions for the NGN. A close cooperation in standards development is claimed to make NGNs a reality. Otherwise, non-interoperability will cause the envisioned FMC to remain unrealized.

References

1. 3GPP: IP Multimedia Subsystem (IMS)—Stage 2. In: 3GPP TS 23.228 V8.0.0, March 2007
2. Strassner, J.: Policy-based network management. Morgan Kaufman Publishers Inc., Los Altos, NJ, ISBN 1-55860-859-1, September 2003
3. ITU-T: Resource and admission control functions in next generation networks. In: ITU-T recommendation Y.2111, October 2006
4. Yavatkar, R., Pendarakis, D., Guerin, R.: A framework for policy-based admission control. In: IETF Standard RFC 2753, January 2000
5. 3GPP: Quality of Service (QoS) concept and architecture. In: 3GPP TS 23.107 V6.4.0, March 2006
6. 3GPP2: Service based bearer control—Stage 2. In: 3GPP2 X.S0013-012-0 v1.0, March 2007
7. ETSI TISPAN: Resource and Admission Control Subsystem (RACS)—functional architecture. In: ETSI ES 282 003 V1.1.1, June 2006
8. Durham, D., Boyle, J., Cohen, R., Herzog, S., Raja, R., Sastry, A.: The COPS (Common Open Policy Service) Protocol. In: IETF Standard RFC 2748, January 2000
9. Chan, K., Seligson, J., Durham, D., Gai, S., McCloghrie, K., Herzog, S., Reichmeyer, F., Yavatkar, R., Smith, A.: COPS usage for policy provisioning (COPS-PR). In: IETF Standard RFC 3084, March 2001
10. Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., Handley, M., Schooler, E.: SIP: session initiation protocol. In: IETF Standard RFC 3261, June 2002
11. 3GPP: Policy control over Go interface. In: 3GPP TS 29.207 V6.5.0, September 2005
12. 3GPP: Policy and charging control architecture. In: 3GPP TS 23.203 V1.2.0, September 2006
13. Calhoun, P., Loughney, J., Guttman, E., Zorn, G., Arkko, J.: Diameter Base Protocol. In: IETF Standard RFC 3588, September 2003
14. ETSI TISPAN: NGN functional architecture—network attachment sub-system (NASS). In ETSI ES 282 004 V1.1.1, June 2006
15. ETSI TISPAN: H.248 Profile for controlling Border Gateway Functions (BGF) in the Resource and Admission Control Subsystem (RACS)—Protocol specification Ia Interface H.248. In: ETSI ES 283 018 V1.1.1, June 2006
16. Cable Television Laboratories, Inc.: PacketCable 2.0, Architecture framework technical report, In: PacketCable 2.0 PKT-TR-ARCH-FRM-V02-061013, October 13th, 2006
17. World Wide Web Consortium (W3C): Web Services Description Language (WSDL) 1.1. In: W3C NOTE-wsdl-20010315, http://www.w3.org/TR/2001/NOTE-wsdl-20010315. Accessed 15 March 2001
18. Cable Television Laboratories, Inc.: PacketCable 2.0, PacketCable Application manager interface specification, In PacketCable 2.0 PKT-SP-PAMI-I02-061013, 13 October 2006

19. WiMAX Forum Network Working Group (NWG): WiMAX End-to-end network systems architecture (stage 2). In Technical Document Release 1 DRAFT, 8 August 2006
20. DSL Forum: Policy control framework for DSL. In: Working text WT-134 Rev. 2, April 2006
21. Rigney, C., Willens, S., Rubens, A., Simpson, W.: Remote authentication dial in user service (RADIUS). In IEFT Standard RFC 2865, June 2000
22. Case, J., McCloghrie, K., Rose, M., Waldbusser, S.: Structure of management information for version 2 of the Simple Network Management Protocol (SNMPv2). In IETF Standard RFC 1442, April 1993
23. Alfano, F., McCann, P., Tschofenig, H., Tsenov, T., Tsou, T.: Diameter quality of service application. In IETF draft-tschofenig-dime-diameter-qos-01, http://www.tools.ietf.org/html/draft-ietf-dime-diameter-qos-05, October 2006
24. Ash, J., Bader, A., Kappler, C., Oran, D.: QoS-NSLP QSPEC template. In: IEFT draft-ietf-nsis-qspec-17, http://www.ietf.org/internet-drafts/draft-ietf-nsis-qspec-19.txt, June 2007
25. Hancock, R., Karagiannis, G., Loughney, J., van den Bosch, S.: Next steps in signaling (NSIS): framework. In: IETF RFC 4080, June 2005
26. Zhang, J., Monteiro, E., Mendes, P., Karagiannis, G., Karagiannis, G., Andres-Colas, J.: InterDomain-QOSM: The NSIS QOS Model to fulfill the E2E QoS control in the ITU-T RACF functional architecture. In: IETF draft-zhang-nsis-interdomain-qosm-04.txt, http://www.potaroo.net/ietf/all-ids/draft-zhang-nsis-interdomain-qosm-04.txt, March 2007
27. Anderson, T., Faynberg, I., Lu, H., Sun, D.: On the mechanisms for real-time application-driven resource management in next generation networks. In: Proceedings of international congress on intelligence in networks, ICIN2006, Bordeaux, France, http://www.tools.ietf.org/html/draft-korhonen-mobopts-mobility-policy-00, May 2006
28. Awduche, D., Chiu, A., Elwalid, A., Widjaja, I., Xiao, X.: Overview and principles of internet traffic engineering. In: IETF RFC 3272, May 2002
29. Ho, K., Wang, N., Trimintzios, P., Pavlou, G., Howarth, M.: On egress router selection for interdomain traffic with bandwidth guarantees. In: Proceedings of the IEEE workshop in high performance switching and routing, HPSR 2004, Phoenix, Arizona, USA, April 2004
30. Verdi, F.L., Magalhães, M.F., Madeira, E.R.M., Welin, A.: The virtual topology service: a mechanism for QoS-enabled interdomain routing. In: Proceedings of the 6th IEEE international workshop on IP operations and management, IPOM 2006, Dublin, Ireland, October 2006
31. Sorniotti, A., Corliano, G., Smith, A.P.: Design guidelines for an internet-scaled QoS framework. In Proceedings of fourth European conference on universal multiservice networks, ECUMN 2007, Toulouse France, February 2007
32. Georgoulas, S., Trimintzios, P., Pavlou, G., Ho, K.: On the location-awareness of bandwidth allocation and admission control for the support of real-time traffic in class-based IP networks. In: Proceedings of IEEE/IFIP management of multimedia and mobile networks and services, MMNS 2006, Dublin, Ireland, October 2006
33. Korhonen, J., Devarapalli, V., Giaretta, G., Koodli, R.: IP mobility and policy control. In: IETF draft-korhonen-mobopts-mobility-policy (work-in-progress), October 2006
34. ITU-T: Introduction to A-IMS (Advances to IMS). In: ITU-T NGN GSI Study Period 2005 Contribution 53, October 2006
35. MultiService Forum: GMI 2006 Physical scenarios. In: MultiService Forum MSF-TR-SCN06.001v2-FINAL, June 2006
36. ETSI TISPAN: IMS/NGN performance benchmark. In: Draft ETSI TS 186 008 V 0.0.98, January 2007

Author Biographies

Christian Esteve Rothenberg holds a 5-year Engineering degree (integrated master) in Telecommunications Engineering from the Technical University of Madrid (UPM), Spain. Within the framework of the T.I.M.E. double degree exchange program (2004/2006), he received the Dipl. Ing. degree (M.Sc.) in Electrical Engineering and Information Technology from the Darmstadt University of Technology (TUD), Germany. He wrote his diploma thesis at Deutsche Telekom/T-Systems on IMS-based fixed mobile convergence and mobility management and worked as a research intern in R&D projects. Currently, he is a research consultant at Centro de Pesquisa e Desenvolvimento (CPqD) and works towards a Ph.D. degree at School of Electrical and Computer Engineering (FEEC), State University of Campinas

(Unicamp), Brazil. His research interests include content and media delivery in next-generation networks and Internet architectures.

Andreas Roos received his diploma degree in Electrical Engineering and Information Technology from the University of Applied Sciences Darmstadt, Germany, in 2004. Currently he is a Ph.D. candidate in School of Electronic and Communication Engineering at Dublin Institute of Technology, Dublin, Ireland. In context of different research projects of German Federal Ministry of Education and Research, he is working for University of Applied Sciences Leipzig, Germany and T-Systems Enterprise Services GmbH in Darmstadt, Germany. His research interests are authentication, authorization, and accounting functionalities and security in next generation networks with focus on wireless mesh networks.