Policy-based QoS management for multimedia communication

Sajjad Ali Mushtaq, Osman Salem, Christophe Lohr, Annie Gravey

Policy-Based QoS Management for Multimedia Communication

Sajjad Ali Musthaq, Osman Salem, Christophe Lohr, Annie Gravey
TELECOM Bretagne Department of Computer Science Brest, France
Email: {sajjad.musthaq, osman.salem, christophe.lohr, annie.gravey}@telecom-bretagne.eu

***Abstract*—NGN is bringing voice, video, data and other triple-play services to a single platform. Policy Based Network Management (PBNM) is becoming indispensable and complicated while converging those services with assured Quality of Service (QoS). Recent work in this area has focused on PBNM considering a particular information domain: e.g. Service Level Agreement (SLA) and/or QoS Routing etc. In this paper we present an architecture with PBNM focussing on access network optimization while taking SLAs, business objectives, routing rules, service Info, user profiles and platform conditions into account. Call setup is based on Session Initiation Protocol (SIP), a text based application layer signaling protocol adopted in NGN for controlling multimedia communications. The service, control and network planes in the proposed architecture are kept isolated. Connection Admission Control (CAC) is correlated with proposed PBNM framework. We also propose to distribute the policy sensitive CAC function between the Call Server (CS) and Session Border Controller (SBC). Two policy enforcement and dissemination modes (Provisioning and Outsourcing) are elaborated.**

I. Introduction

The emergence of multimedia services in IP networks make Quality of Service (QoS) management an unavoidable task. QoS mechanisms impose stringent requirements on service providers, network operators and applications. Dynamic and intelligent Policy Based Network Management (PBNM) has become vital for efficient operation in an enterprise with multiple accesses (to Internet) offering converged services.
The research efforts for QoS provisioning in connection with PBNM has been initiated by the Internet Engineering Task Force (IETF). Integrated Services (IntServ [1]), an extension to the traditional best effort model and Differentiated Services (DiffServ [2]), which works on the basis of Behavior Aggregate (BA) classification are the two QoS mechanisms proposed at IETF. IntServ achieves end-to-end service guarantee on per-flow resource reservation basis, while the DiffServ focuses on traffic aggregation and provides per class prioritization.

Enterprise networks now-a-days involve mutiple media, services, protocols and different platforms. They interconnect with public network via several links/ISPs. There might be multiple providers offering bearer and valus-added services competing to meet customer demands with low cost. Private-public network border have become sensitive in this scenario and require sophisticated management. At the private-public border the task is not just finding the broken link and/or marking good/bad QoS onto the links. The approach must be predictive and proactive taking into account cost, bandwidth utilization, chokes and setting up thresholds etc. These issues are handled partially by conventional devices like Session Border Controller (SBC) and Call Server (CS). Additionally existing traffic classification/marking/tagging mechanisms over some traffic enginnering principal at the private-public border may not function accordingly as both mechanisms are orthogonal. But we are not concentrating on end-to-end issues rather we are emphasizing on the private-public border taffic management issues encompassing high dynamicity and variability in services, QoS requirements of users, tarrifs, and network conditions.

Enterprises want to manage their networks efficiently while fulfilling the SLAs with the peer carriers. They want to achieve their business objectives by using the resources efficiently, and in the same time providing required QoS for applications/users. For example, business objective rules for network management, may dynamically allow low user profile packets to use the costly access if it is under-utilized. Manual configuration of router interfaces (access-list, route-map, static routing, etc.) are not efficient to achieve this goal.

One of the important task is to map user requirements onto the system depending on user profiles, requested QoS (Communication Type) and available resources, while keeping in view business rules, global and local configuration of the platform (e.g. access or billing policy). The selection of a given access for a multimedia flow (identified by the 5-tuple in the IP packet header: the source IP address, destination IP address, protocol number, source port number and destination port number) is controlled by blocking the setup or by selecting an appropriate access depending on specific criteria. The access selection criterion will take into account the company's policy (Routing rules, business objectives, user profiles, network conditions). The enterprise priorities for a certain user type might be: 1) Look for admissible best QoS link first. 2) Select the cheapest admissible access. 3) Select randomly any admissible access. In another case the priorities for a specific group might be: 1) Apply a Multimedia over IP (MoIP) service preference policy on the admissible links. 2) Select an access according to a load balancing (more generally, a traffic engineering policy). 3) Select randomly any admissible access. A fundamental and key characteristics which has been discussed quite a lot in the literature is LCR. The correlation of QoS mechanisms with LCR principles at the borders/edges in an enterprise is indispensable in this NGN and IMS era. Conventional LCR helps to optimize connections between telecommunication operators by minimizing costs

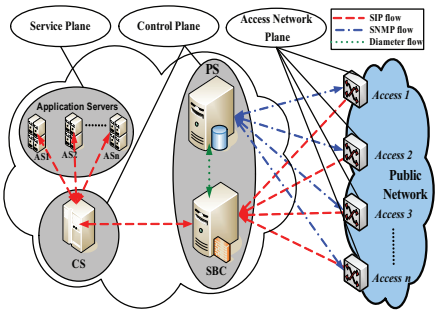

Fig. 1. Proposed Architecture.

for several demands and maximizing the operator income along with efficient use of the existing network infrastructure. We address the generalization of LCR to Session Initiation Protocol (SIP) [3] based multimedia services enveloping a number of constraints and parameters. The fundamental tenet is to present an architecture offering policy-based session management for multimedia communication, by using LCR principles at the network border point which provides public-private accesses to Internet through many Links/SLAs.

Private to public network traffic management problem becomes more complex in the presence of NAT/Firewall enabled endpoints. Additionally with the convergence of voice, video and data with technology over an IP infrastructure require more efficient and sophisticated operations within the conventional devices like SBC and CS. We have emphasized on private-public border traffic management issues in connection with policy based access selection and management. NGN/IMS frameworks might not be effective for our case as firstly they may not provide enough granularity on user profile and/or usage characterization especially when granularity is correlated with the inputs from the platform and secondly we are managing only border traffic management issues rather than taking care of end-to-end QoS which generate lot of QoS signalling and consumes resources with the introduction of delays. These issues has been addressed partially in proprietary devices like SBC and CS but we have to tackle two main issues: 1) We want to propose a method that can account of dynamic variations in resource usage/quality. 2) We want a mechanism that relies on standard methods since the network equipment can be provided be different companies.

The aforementioned QoS mechanisms can not deployed in our case as IntServ relies on signalling messages exchanged between terminals periodically, hence QoS provisioning is not easy in high speed networks on per session basis. DiffServ model can simply provide a coarse end-to-end QoS granularity which may result in a low resource utilization. Real-time data flows in DiffServ network are mapped to Expedited Forwarding (EF). If a real-time flow requires 100 ms delay and another flow needs 500 ms delay bound and they both are mapped to the EF service then there will be wastage of resources by allocating better-than-required QoS. This problem can be solved by defining more classes but the DiffServ code points (DSCP) might not be standardized on the public network. The granularity of defining new classes can not be fixed, so we need dynamicity and flexibility to take variations even in the single parameter (delay) into account. Additionally, inside an enterprise, it is desired to prioritize the multimedia packets to/from some specific user profile along with prioritization on the basis of applications/services. But still there is not enough granularity either in user profile or in usage characterization.

The architecture is proposed in the Companym@ges project which proposes a platform where companies are linked to the rest of the world via different network accesses offering data and multimedia services. The work presented in this paper relates to the sub-project for traffic management issues at the border of the company's network. Components of this platform are provided by partners: SIP CS, SBC,and Policy Server (PS)) constituting the control plane in the proposed architecture (figure 1) are/will be developed by Alcatel-Lucent, Comverse and TELECOM Bretagne respectively. The reader is refereed to section III for more details about these devices. In this paper, we focus on policy server addressing policy based QoS management and control emphasizing on access optimization (Session routing according to the context of external links and user demands along with service requirements).

The remainder of this paper is organized as follows. In the following section, we discuss the related work and previous research related to our work. Section III describes the proposed architecture for PBNM. Section IV elaborates CAC and its distribution. In section V we discuss the policy system functioning, supported policy enforcement modes and the information required for policy computation. Finally, section VI presents concluding remarks and the future work.

II. Related Work

QoS represents the quantitative and qualitative characteristics required to achieve the desired functionality of applications/services. Different services require different QoS. The convergence of technology and services introduces more sensitivity requiring more accuracy in the QoS requirements. Achieving the required QoS with a number of contraints while taking into account **SLA**s, **B**usiness **O**bjectives, **R**outing **R**ules, **S**ervice **I**nfo and **P**rofiles (SLABORRSIP) involve additive complexity. QoS techniques implementation at private-public border to route the incoming/outgoing requests/calls to different accesses with different constraints and associated parameters can not be handled without policy control in all scenarios. SBC controls this access selection while mapping the private addresses to public address space. This process involves the straight forward access selection mechanism but we need to optimize the access selection process under policy control considering the variations and dynamics for efficient resource management and admission control.

Network management is essential to guarantee the communication and services with required QoS. Policy based management has been the subject of extensive research over the last two decade. It has been central part in Next Generation Network (NGN) like 3GPP and TISPAN. PBNM becomes more complex when different paradigms are merged. IETF has been investigating policy based networking as a

means for managing IP-based multi-service networks with QoS guarantees. They have proposed a policy framework [4] for management, representation and control of policies in an independent and interoperable fashion. The Meta Information Base (MIB) [5] concept was replaced by Policy Information Base (PIB [6]), and several protocols were specified for the transfer of policies or network configuration, such as the Common Open Policy Service (COPS [7]) and SNMP [8]. All the frameworks and protocols mentioned above introduce more or less static control/management and most of them are used for device configuration and management with no/little dynamicity along with information retrieval. A framework or protocol might be used for static policy control or information retrieval but they may not fulfill our requirement of private-public boder traffic management issues and control taking dynamic variations and constraints into account.

Voice, video and data require different QoS levels and the situation becomes even more complex when services are merged (e.g video conference). Additionally the heterogeneous technologies with versatile services require adaptability and dynamicity in PBNM. An adaptive, generic and controlled service delivery framework implementation for multimedia applications on IMS architecture has been proposed in [9]. It focusses on session-level, user driven end-to-end QoS emphasizing on terminal and session mobility. The importance of modelling and natural languages in the presence of the policy continuum, resulting in a novel architecture suitable for autonomic computing has been presented in [10]. Derivation of policies from QoS/SLA agreements for dynamic change of resources for QoS-aware applications in heterogeneous network environments (UMTS, WIMAX, WLAN, DVB-T, DVB-H) along with interfaces and data base design for automated policy configuration and adaptation has been discussed in [11]. An architecture and framework for policy based mobility management based on two coordinated decision engines for access detection, evaluation and selection in heterogeneous access networks has been proposed in [12]. In some of the referenced works, authors concentrated on PBNM in the context of access technology (OSI layer 2) while considering specific domain of information (Service plane or control plane or network plane). In some other articles, mobility management in the context of PBNM with adaptation has been addressed.

Policy control is essential for resource management and admission control in VoIP and other IP based multimedia communications. Dynamic QoS and policy enforcement are fundamental entities in multi-service converged networks. Policy control in NGN/IMS is still evolving. The integration of Policy Control Function (PCF) and Flow Based Charging (FBC) to form Policy and Charging Control Function (PCCF) is the current evolution in IMS release 7 [13]. The existing NGN/IMS controls and interfaces can not be deployed directly in our case due to the following reasons: 1) The controls and interfaces are in the process of evolution and some of them are under development and standardization phase. 2) There is not enough granularity in private-public traffic management issues, while in our case the granularity of the QoS resource management and admission control is tied with SLABORRSIP 3) Our platform is a competitive cluster i-e SBC from one partner, CS from another vendor and PS from the third partner. So we need to hanlde the issues with standard methods. 4) In NGN/IMS functionality of the interfaces/devices is important but its location is loosley coupled at the planes interconnections i-e Home Subscriber Server (HSS) can be deployed either in application/service plane or in the Control/Signalling plane. But in our case we are trying to isolate the service, control and transfer palnes as much as possible due to the competitive nature of the architecture.

Finally existing frameworks address PBNM focussing on a particular information domain (e.g. QoS routing and/or service/application level QoS etc). These PBNM paradigms lag dynamicity. The systems maintaining dynamicity fulfils converged services and technology support partially. The frameworks prevailing over those mentioned problems have service, control and transfer planes overlapping issues alongside performance and policy management fall-backs. The motivation of this work is to address PBNM focussing on service, control and management planes isolation taking into account SLABORRSIP.

III. Proposed Architecture

QoS-centered proposed architecture provides linkage between private network (Where companies offering versatile services are linked to the rest of the world via different accesses) and the public network (The Internet). The prime objective of the proposed architecture is accommodation of dynamic modifications/variations by using enhanced general methods and techniques handling service, control and transfer planes separately while supporting the standard protocols. PBNM paradigm in the architecture envisaged an extended IETF framework [14]. SIP, an ASCII based Internet-centric protocol adopted in the 3GPP2 and IMS is the main signalling protocol in the presented architecture. CS, SBC, PS and application servers are the main building blocks of the proposed architecture.

CS is an important component of IP based PBX/Softswitch. It may also support proxy, registrar, redirect and location services. Most of the CS solutions are proprietary and support wide range of services. CS here provide registration, user profile management, service control and user profile CAC functionality. CS is connected directly to SBC and they communicate via SIP protocol. There is no direct communication between CS and PS and it is declared as one of evolutions in the project to manage the profile and service related policy controls directly.

SBC is another significant module of the proposed architecture. It is a session aware device. The primary functionality of the SBC is SIP natting translation and firewall in order to protect private network and to hide the private network topology. The term SBC is not specific since its functionalities are not yet standardized or defined anywhere [15]. SBC provides a variety of functions to enable or enhance session based multimedia services (e.g. VoIP). The key functions of

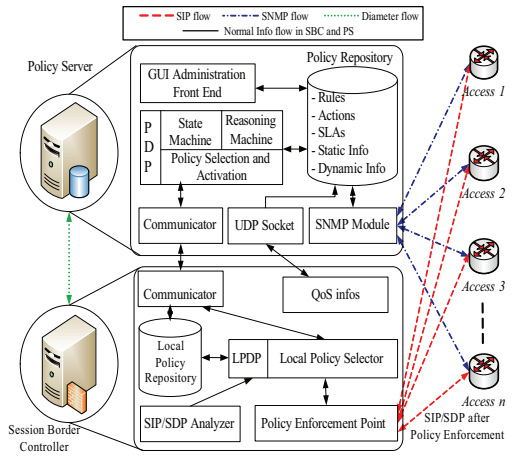

Fig. 2. PS and SBC modules and the their inter-communication

SBC are: Perimeter defence (access control, topology hiding, DOS prevention and detection), functionality not supported at the end points (NAT traversal, protocol interwork, media repair) and network management (traffic monitoring, shaping and QoS). Some of these functions may also get integrated into other SIP elements like 3GPP P-CSCF, 3GPP I-CSCF etc. SBC can handle both signalling and media depending upon its functionality and deployment. In our case, we assume that SBC also embeds: a SIP/SDP (Session Description Protocol [16]) analyzer, a communicator function (used for policy and information exchange between PS and SBC), network QoS monitoring and policy enforcement module. It can also act as a Local Policy Decision Point (LPDP) to support the provisioning mode illustrated in section V-C. SBC is connected to the PS and CS communicating via diameter and SIP protocols respectively. The private-public border is traversed through SBC through its external/internal interfaces to accesses.

PS is the core of our proposed architecture. An abstract modular diagram of PS and its communication with SBC is presented in figure 2. Policy based management system is emerging as the promising technology to address the challenging tasks in the converged NGN [17], [18]. We have tried to extend the IETF policy framework keeping in view integrity and compatibility. Conventional policy based systems do not allow dynamic control, management and extensibility. Although some policy based systems have been proposed for managing telecommunication and enterprise network services [19], [20], but those systems were specifically designed for a certain set of services and environments with limited dynamicity. We have tried to provide a policy system which support converged network services with dynamicity emphasizing on access network optimization. PS is an intelligent and key player in our framework. It has to obey the SLAs and reciprocal agreements. The business objectives of the enterprise must be fulfilled while the users and applications should get assured QoS. We have also addressed the access optimization by taking into account the information narrated along with routing rules and configurations entered by the administrator. In order to provide required QoS for multimedia services, the policy server provides a decision for every request by considering static information and dynamic information (e.g. time of day, context of the external links, Statistical analysis of QoS information or Call Details Records (CDRs), etc.). PS is connected to SBC over IMS defined 'Gq' interface and it is also connected to external links directly. The SNMP flows presented in figure 1 are used by the PS, to query information on the state of the links.

One of the key issues in policy based communications network management and service provisioning is QoS routing. It selects network routes with sufficient resources for the QoS requirements for every admitted QoS flows. QoS routing, if done appropriately, can significantly improve network performance. Conventional QoS policy based routing (Layer 3) provides a mechanism for expressing and implementing forwarding/routing of data packets based on the policies defined by the network administrators. It provides a more flexible mechanism for routing packets through routers, complementing the existing mechanism provided by routing protocols. Policy based routing in the proposed architecture is performed at call/session level on OSI layer 5. The criteria might be based on 5-tuple, source/destination IP address, source/destination port and the protocol field, a URL or the time of day, profile, price and/or security requirements. The routing policies are entered by the administrator of the platform. More details are available in section V-E.

The protocol used for communicating the information/policies between Policy Decision Point (PDP) i-e PS and Policy Enforcement Point (PEP) i-e SBC is Diameter with additional Attribute Value Pairs (AVPs). IP Multimedia Subsystem (IMS) interface Gq has been adopted and modified for policy/information inter-communication between PS and SBC. Diameter is an Authentication, Authorization and Accounting (AAA) protocol being adopted in NGNs for subscription, policy and charging functions. Due to Diameter's AAA characteristics, its enhancement orientations are becoming natural and native choices for PBNM [21]. Diameter has large AVP space and supports large number of pending requests. Reliability is provided by SCTP/TCP with well defined fail over scheme. COPS, a strong candidate for PBNM has not been picked up for policy dissemination, as it is specially designed for device level configuration and management, but we need dynamic session/call management taking into account the variations and latest dynamics. SNMP [8] traps are exploited in the policy system to gauge the QoS parameters of access router interfaces.

The service, control and network planes as shown in figure 1 have been decoupled except some control and network plane functions overlap. Addition of new services without hardware/software upgrade and interruption, automatic discovery of access network failure (topology change) and policy based control at service, user and network levels are the key advantages. Another feature of the architecture is mapping of application level QoS onto network level QoS in consent with active policies. The computation of those active policies take into account the service, control and network planes latest

information.

IV. Connection Admission Control

The primary focus of this paper is PBNM but discussion on CAC can not be avoided as it shares the same framework and goes hands in hands with PBNM. CAC is a mechanism used in networks to administer QoS. CAC is used to limit the number of connections in the network and in some cases it works jointly with bandwidth allocation and QoS routing [22]. It allocates available resources among the outgoing and the incoming connections, for maintaining the QoS performances of both types of connections at the required level. Network accepts or denies data flow on the basis of decision of CAC scheme. Decision (accept/deny) is based on predefined criterion which in turn depends on network environmental conditions in coordination with: rules, business objectives and administrative configurations of the platform. This decision has considerable influence on QoS parameters, which makes CAC an essential tool to guarantee various QoS parameters. Distributing the CAC mechanism on physical, data and network layer have been studied in [23]–[25].

CAC has to be performed with the inter-communication and mutual understanding of control plane (Resource-based CAC) and service plane (Profile-based CAC) and in the same we want to handle those planes separately using standard methods and protocol. The SBC sitting at the edge, as an access point between private and public networks, can handle availability, cost and quality of the accesses, so resource based CAC should be performed at the network border. The CS on the other hand have the latest knowledge of registered users, services, profile related AAA information. Therefore, it seems quite suitable to perform profile based CAC at call server in the proposed architecture (shown in figure 1). Another dimension of splitting the CAC among SBC and CS is the distribution of intensive computational complexity. In our architecture, we propose to distribute the policy sensitive CAC and this distribution stipulates the decoupling of service, control and access network planes. The policy based CAC mechanism has been splitted into user and resource based CAC, and it has been performed at distinct locations (CS & PS). Profile based CAC using *Resource-Priority*' and '*Accept-Resource-Priority* proprietary SIP headers [26] has been addressed in [27] and we will focus on resource based CAC in the present paper.

V. Policy System functioning and policy Enforcement Modes

Today's converged data networks support Internet telephony as well as packetized video and other delay sensitive services in addition to delay insensitive data. Besides rapid technological changes and innovative services require flexible and extensive PBNM paradigms with dynamicity. Policy based network and service management becomes essential when typical best effort data traffic has to be managed with real-time QoS sensitive services.

In this paper we have taken into account the static and dynamic aforementioned information (SLABORRSIP). The significance of considering those different information domains requires a coordinated effort of management, control and adaptation in connection with service, control and network planes in the context of PBNM. We have adopted extended IETF policy framework. Separation of service, control and access network planes, distribution of CAC (between CS and SBC), and dynamic policy computation and enforcement are considered as novel ideas from our view point. KAoS due to its expansion and reasoning support has been chosen as a policy language. Its ontological representation of rules and information facilitates policy computation. KPAT, a graphical interface for creation and edition of policies complements our choice. An administrative API will be used for the specification, addition and edition of rules and relevant information. Two policy dissemination and enforcement modes namely provisioning mode and outsourcing mode have been proposed.

A. Policy System Behavior

Our PS architecture is based on IETF framework with some add-on features. The key elements in IETF policy Framework are Policy Decision Function (PDF), Policy Enforcement Function (PEF), policy management interface and policy repository. Additional modules in the PS in the proposed architecture are Communicator (Diameter Gq interface) for policy and information inter-communication, an SNMP module for QoS information retrieval from routers and from SBC along with network monitoring and the GUI administrative front-end for addition, edition and modification of rules and relevant information. SBC is also enriched with SIP/SDP analyzer, PEP, LPDP, Communicator and SNMP module with the same functionality as described earlier. The PS and SBC block diagram is shown in figure 2.

PS, the central element of our architecture takes environment parameters, administrative configurations, business and routing rules, as input. It then extracts the information and translate it into a certain Database Management System (DBMS) format for further processing. Policy computation is carried out at PDP, the "brain" of PS. Policy instantiation, distribution, disabling, unloading, deletion and conflict detection are the key functions of PDP. PS supports online and off-line policy computation. Off-line computation means that the policy decisions are pre-computed and stored to be used whenever needed, online computation on the other hand encompasses the policy decision computation on the fly for a specific demand. Rise in call/session setup time, increase in latency and performance degradation might be the possible fallbacks of online policy computation. Accommodating the latest and updated static/dynamic information in the online mode leads to refined and efficient resource management. Conflict between policy decision and the platform resources/conditions might be the disadvantage in the off-line policy computation. A trade-off between online policy computation frequency and system performance has to be managed.

B. Policy Enforcement Modes

The PS supports two modes of enforcement operations namely provisioning and outsourcing mode. Whenever a re-

quest arrives, the SIP/SDP analyzer extracts the required information and sends this information to LPDP, which then maps an appropriate policy from the policy base, the process resumes otherwise on mapping failure and the information is sent to the communicator at SBC. The policy enforcement irrespective of the two modes is ultimately done at SBC (PEP). In provisioning mode the pre-computed policies are available in the policy repository at SBC. In outsourcing mode the extracted information from the pending request is fetched to the PDP at PS via communicators. PS, in outsourcing mode is delegated to compute/use online/off-line policy/decision based on the request and the system conditions. Pre-computed policies in the provisioning mode are fetched at SBC, a-priori irrespective of online or off-line computation. The two enforcement mechanisms are in contrast with each other but they are not mutually exclusive. The policy based management system is capable of handling both data and multimedia services. However we are explaining the two policy enforcement modes while considering SIP based multimedia communication. The initial inter-communication of different modules and the information flows in the provisioning and outsourcing modes as shown in figures 3 and 4 respectively are identical till step '3' and it is given in the following.

Call Type	Access
(User Type-1,Communication Type-A)	1
(User Type-1,Communication Type-B)	3
(User Type-3,Communication Type-C)	2

TABLE I
SWITCHING TABLE FOR POLICY-BASED ROUTING

1) This will be a SIP INVITE from CS with resource priority header information as CS is managing registration, user profile management, service control with the underlying profile based CAC mechanism.
2) The SIP/SDP analyzer will then examine the request/session and will send the extracted information to the LPDP. e.g. the communication type can be identified by SBC after analyzing the SDP payload while the user type is computed from proprietary SIP header fields (sent by CS) mentioned earlier. This information is bundled into a pair and is compared with the switching table shown in table I
3) The LPDP will play with this information and sends a request to local policy repository at SBC inquiring an appropriate policy.
   In provisioning mode, upon policy mapping the response is sent back to LPDP otherwise on failure, the system resumes further processing and we are in outsourcing mode now.

C. Provisioning Mode

Pre computed policies in provisioning mode are already available at local repository at SBC before the request/session arrives. LPDP will select an appropriate policy from local policy-base and its enforcement takes place at PEP. The information flow and the inter-communication between different modules is illustrated in figure 3. The tagged information flow

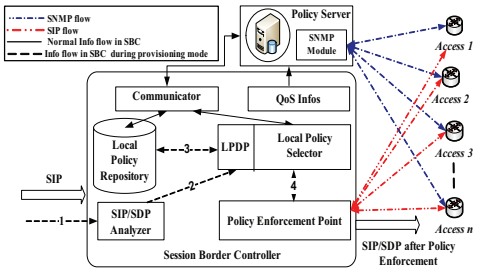

Fig. 3. Information Flow in Provisioning Mode

except the common steps (V-B) in the two modes mentioned earlier is given in the following

(4) LPDP disseminates the policy for enforcement to the enforcement point (PEP), e.g. if the profile type belongs to group XY (User Type-1) and the communication type is A then the policy instructs to route the call/request to link 1 (See table I for more examples).

D. Outsourcing Mode

Unlike provisioning mode, the policy/decision is computed/fetched upon the arrival of request/session in outsourcing mode. When a request arrives at the SBC and it finds no policy for the current request/session. It resumes further processing in outsourcing mode by marking the request as pending and sending the appropriate information to the PS Via Gq interface using the diameter protocol e.g. the communication type, user profile along with required QoS Info may be transmitted. The information flow is given in the figure 4. Sequenced and tagged information except the similar steps in the two modes (V-B) is explained below.

(4a) LPDP asks the communicator at SBC to send request for post policies computation according to the context of the request/session.
(4b) While LPDP is sending request to PS for policy computation, it marks the current request at PEP as pending.
(5) The two communicators at SBC and PS interact via Gq using Diameter protocol and the required information is transferred to PS
(6) The PS server communicator sends this information to the PDP at policy server which compute/use online/off-line policy/decision based on the request and the system conditions.
(7) The PDP at PS then transfers the policy/policies to local communicator.
(8) The PS communicator then transfers the policy/policies to the SBC communicator via Gq..
(9) The communicator at SBC communicates the computed policy/policies to the LDPD.
(10) LPDP transfer the policy/policies to PEP for enforcement onto the pending request/session.

E. Policy information and its Management requirement

The nuts and bolts of our architecture are policies, their representation and management. Despite lot of research on PBNM, the policy definition is still controversial in literature. A policy itself is a rule that specifies under which conditions a resource (or another policy) might be disclosed to a requester.

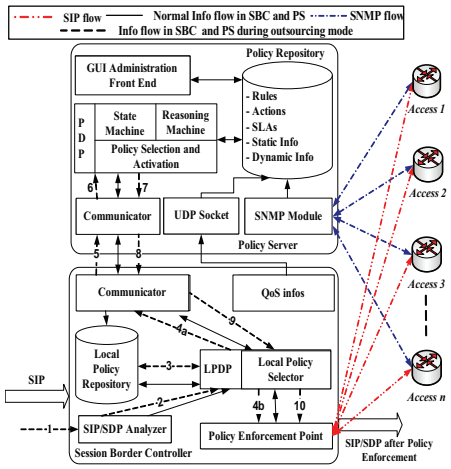

Fig. 4. Information Flow in Outsourcing Mode

A broader definition indicates that policies are a specific type of rules that dictate the behavior of a system. IETF defines policy an event-control-action tuple [28]. From our perspective we have tried to adapt the IETF definition, where the arrival of a request/session is an event. Set of routing rules, business objectives, QoS parameter from the environment and information extracted from the active request/session etc are translated to control. The actions might be: route the request/session with user type 1 and communication type B to the access 3, terminate the session, QoS marking of an access, etc. The policies will be derived from the different domains of information as described earlier, for instance SLABORRSIP and QoS of the external links. From the policy enforcement and dissemination point of view we have a PDP consisting of a state machine and reasoning machine at PS for policy computation/management and conflict detection. PEP, LPDP and local policy repository for policy selection locally in provisioning mode are situated at SBC. The low level details about policy computation and policy activation both at PS and at SBC are outside the primary focus of this work.

The principal objective of our work is policy based network management emphasizing on access network optimization. To achieve this goal we take SLABORRSIP into account. Each access in the architecture has an SLA and an associated identification. SLAs include the definition and the values of some measurable technical parameters defining the quality associated with the negotiated service. This might includes service schedule, latency, jitter, throughput, availability etc. The proposed architecture has different access types and there might be number of SLAs, the local enterprise and the other providers must follow and have to obey. So the methodology for QoS monitoring, marking, under/over subscription of the links, upgrading and downgrading and vice versa of accesses on the basis of QoS and cost has to be handled in a refined and efficient way. Due to the lack of standardization there is no specific SLA for IP services. Most of the providers and vendors have defined their own SLA according to the offered services and technology. An example of IP SLA is given in table II.

Business objectives reflect the business goals, accounting

Parameter	Value
Network Latency	50-100 ms
Packet Loss	1.0 E-09-1.0E-06
Jitter	50-60 ms
Downtime	2 S
Charging	Cumulative (Per byte)

TABLE II
AN IP SLA

and standard operating procedures and rules (e.g. prioritization, accounting principles, access preemption methodology, lawful interruption, resource dedication/allocation to a particular service etc). Those business goals are business terms rather than technical ones. The examples of business objective are: *Allocate 20 percent of the available bandwidth to best effort data traffic*, *give priority to voice calls on a particular time of the day (peak hours)* etc.

Conventional routers implement routing rules at layer 3 in OSI framework according to the configuration and rules. But the session/call level routing rules in the proposed architecture are being implemented at OSI layer 5. The criterion for policy based routing may be routing based on codec used, routing on the basis of price, routing considering the source and/or destination information, routing taking security requirement into account, routing based on time of the day and routing based on profiles.

Service information consists of offered services (e.g. voice, video, data and triple-play of those described services (video call)), and default QoS parameters associated with those services (delay, jitter, packet loss etc). Service information and their default QoS parameters must be embedded so that while taking a decision the required user/application QoS values or parameters should be compared to the threshold levels or default values for the corresponding service. As an example, the upper bound for packet delay, jitter and packet loss for voice service might be 100 milli second (ms), 50 ms and 0.1 respectively.

Profiles contain system, application and user level settings for AAA with prioritization. We have grouped the profiles on macro scale for simplicity (e.g: Gold, Bronze and Silver profile groups). Further subgroups on the microscopic level are also possible depending on the administrative environment, services and charging schemes. Gold, silver and bronze profiles require excellent, good and satisfactory QoS respectively with the corresponding high, intermediate and low priorities. QoS and priority handover may also take place according to administrative policies and state of the network *e.g: bronze class may enjoy gold service if the dedicated resources for that class are under-utilized.*

Routing rules, profile and service information which will be ultimately translated into policies are exemplified as follows

- Gold class profile requests should be routed to the link marked as exceptional good QoS.
- Do not route the video conference request to any of the accesses if the profile belongs to bronze class.
- In case of default policy enforcement, route the request/session to the under utilized link.

- If 95 percent bandwidth of the link is being utilized, do not route any more sessions/calls until an emergency.
- While routing a session/call, look first for the best QoS link for a certain family of profiles then sesrch/find the cheapest access.
- If the load is below a predefined threshold on an access then route sesions/calls to consume the dedicated Best Effort Bandwidth.

In case of conflicts/failures the system have priorities and default policies to enforce for the normal system behavior and functioning. The proposed system's architectural, functional and specification design phase has been completed and at present it is in the initial development phase.

VI. Conclusion

We have proposed a novel QoS-centered architecture in multiservice packet network for PBNM addressing the SIP based communication while emphasizing on access management. This architecture is composed of CS, PS, SBC and application servers. The advantage of the proposed architecture is three fold. The service, control and network planes are kept isolated. Secondly the change in topology likely due to an external link failure, recovery of a broken link and/or addition of a new link will be sensed and accommodated without interruption. Finally the twin drivers of the CAC distribution mechanism (CS and SBC) share and exchange relevant information in control plane and this process has been kept isolated from policy dissemination and enforcement. The ultimate and significant goal is adaptive and dynamic policy based access optimization. Complex and intensive resource consuming CAC mechanism is splitted into profile and resource based functions. The prime objective is dynamic policy management by extracting the relevant information from SLAs, business objectives, routing rules, service Info, profiles and local/global configuration of the platform. Two policy enforcement mechanisms (provisioning and outsourcing) modes have been elaborated. Information required for policy computation and management has been discussed. Some abstract policies and rules have been exemplified. Current work exploits the presentation of a framework for PBNM while considering versatile information domains for dynamic policy computation. Our part of the Companym@ges sub-project, the PS is in initial development/testing phase and the integrated beta version of the three planes (service, control and access network planes) will be tested soon.

VII. Acknowledgment

This work has been partially funded by the DGE (Direction Générale des Entreprises/Ministère des Finances et de l'Industrie) through the Companym@ges project. The authors would like to give many thanks to Antoine Gatineau from Comverse for his support and guidance.

References

[1] J. Wroclawski, "The Use of RSVP with IETF Integrated Services," RFC 2210, Sep. 1997.

[2] K. Nichols, S. Blake, F. Baker, and D. Black, "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers," RFC 2474, Dec. 1998.

[3] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler, "SIP: Session Initiation Protocol," RFC 3261, Jun. 2002.

[4] R. Yavatkar, D. Pendarakis, and R. Guerin, "A Framework for Policy-based Admission Control," RFC 2753, Jan. 2000.

[5] A. Polyrakis and R. Boutaba, "The meta-policy information base," *Network, IEEE*, vol. 16, no. 2, pp. 40–48, Mar/Apr 2002.

[6] R. Sahita, S. Hahn, K. Chan, and K. McCloghrie, "Framework Policy Information Base," RFC 3318, Mar. 2003.

[7] D. Durham, J. Boyle, R. Cohen, S. Herzog, R. Rajan, and A. Sastry, "The COPS (Common Open Policy Service) Protocol," RFC 2748, Jan. 2000.

[8] D. Harrington, R. Presuhn, and B. Wijnen, "An Architecture for Describing Simple Network Management Protocol (SNMP) Management Frameworks," RFC 3411, Dec. 2002.

[9] C. Balakrishna and K. Al-Begain, "Towards a User-Centric and Quality-Aware Multimedia Service Delivery Implementation on IP Multimedia Subsystem," in *NGMAST'07), year = 2007, month = Sept, pages = 36–42,*.

[10] D. Raymer, J. Strassner, E. Lehtihet, and S. van der Meer, "End-to-End Model Driven Policy Based Network Management," in *Seventh IEEE International Workshop on Policies for Distributed Systems and Networks*, 2006, pp. 67–70.

[11] P. A. A. Gutierrez and I. Miloucheva, "Automated QoS policy adaptation for heterogeneous access network environments," in *ICSNC'07*, 2007, pp. 65–72.

[12] C. Fan, M. Schlager, A. Udugama, V. Pangboonyanon, A. Toker, and G. Coskun, "Managing heterogeneous access networks coordinated policy based decision engines for mobility management," in *32nd IEEE LCN'07*, 2007, pp. 651–660.

[13] "3GPP TS 32.260 v7.3.0, IP Multimedia Subsystem (IMS) Charging R7," 2007.

[14] A. Westerinen, J. Schnizlein, J. Strassner, M. Scherling, B. Quinn, S. Herzog, A. Huynh, M. Carlson, J. Perry, and S. Waldbusser, "Terminology for Policy-Based Management," RFC 3198, Nov. 2001.

[15] J. Hautakorpi, G. Camarillo, R. Penfield, A. Hawrylyshen, and M. Bhatia, "Requirements from SIP, Session Border Control Deployments," IETF draft, Mar. 2008.

[16] M. Handley, V. Jacobson, and C. Perkins, "SDP: Session Description Protocol," RFC 4566, 2006.

[17] S. J. Shepard, "Policy-based networks: hype and hope," *IT Professional*, vol. 2, no. 1, pp. 12–16, Jan-Feb 2000.

[18] P. Flegkas, P. Trimintzios, G. Pavlou, I. Adrikopoulos, and C. F. Calvacanti, "On policy-based extensible hierarchical network management in QoS-enabled IP networks," in *POLICY 01), year = 2001, pages = 230–246, publisher = Springer-Verlag, address = London, UK,*.

[19] L. Lymberopoulos, E. Lupu, and M. Sloman, "An adaptive policy based management framework for differentiated services networks," in *Proceedings of the Third International Workshop on Policies for Distributed Systems and Networks*, 2002, pp. 147–158.

[20] D. Verma, M. Beigi, and R. Jennings, "Policy Based SLA Management in Enterprise Networks," *LNCS*, vol. 1995, pp. 137–152, 2001.

[21] M. Brenner, "Diameter Policy Processing Application," RFC 5224, Mar. 2008.

[22] T. Ezaki, H. Kawakami, and K. Asatani, "A new voip call admission control scheme with use of alternate routing for low call loss probability," *2004 IEEE International Conference on Communications*, vol. 4, pp. 2209–2213, 20-24 June 2004.

[23] H. Nan, W. Xiaoxiang, and W. Weiling, "Integrated Cross-Layer Design of Utility-Based Connection Admission Control in Packet-switched OFDM Wireless Networks," in *IWCLD '07.*, 2007, pp. 152–156.

[24] J. Chen and C.-W. Chang, "Traffic-Variation-Aware Connection Admission Control Mechanism for Polling Services in IEEE 802.16 Systems," in *WOCN '07. IFIP*, July 2007, pp. 1–5.

[25] D. Oulai, S.Chamberland, and S. Pierre, "End-to-End Packet Loss Constrained Routing and Admission Control for MPLS Networks," in *CCECE 2007*, April 2007, pp. 341–344.

[26] H. Schulzrinne and J. Polk, "Communications Resource Priority for the Session Initiation Protocol (SIP)," RFC 4412, Feb. 2006.

[27] S. A. Musthaq, O. Salem, C. Lohr, and A. Gravey, ""Distributed Call Admission Control in SIP Based Multimedia Communication" Submitted for Publication."

[28] B. Moore, E. Ellesson, J. Strassner, and A. Westerinen, "Policy Core Information Model – Version 1 Specification," RFC 3060, Feb. 2001.