Linking Session Based Services and Transport Layer Resources in the IP Multimedia Subsystem

Richard Good and Neco Ventura
University of Cape Town, Rondebosch, South Africa
Email: {rgood, neco}@crg.ee.uct.ac.za

***Abstract*—The Web 2.0 Internet revolution has introduced a new communications model, introducing a new age of community based websites with millions of empowered and contributing users exploiting the wisdom of the masses. The majority of these services are available free of charge and revenues are based on the potential for personalised advertising. The new model poses a threat to wide scale IP Multimedia Subsystem (IMS) deployment and a decision needs to be made on the business case for deploying IMS services. To justify charging for services that are typically available free of charge on the Internet, operators must differentiate their services through increased service integration and security, and reliability through efficient management of resources.**

This paper examines standardisation efforts regarding resource management in the IMS environment and highlights important architectural requirements and deployment challenges. A Multilayered Policy Control Architecture is presented that addresses the open issues of policy definition, policy profiling, application-policy interaction and access network policy refinement. The proposed architecture is implemented in a practical IMS test-bed where it is subjected to rigorous validation tests. In particular the effect the architecture has on traffic overhead and session set up delay with single and multiple session requests is examined.

***Index Terms*—IMS, RACS, Policy Control, QoS**

I. Introduction

The Internet has enjoyed massive success as a communications platform largely due to the lack of centralised management and non-proprietary nature of development. Network operators have invested huge amounts in the legacy PSTN networks and have been reluctant to advance from this status quo for fear of losing that investment. However drastically changing market conditions and technological advances are tearing down the borders between industries and platforms. In particular we have seen massive growth in IP Multimedia and Web 2.0 services; because of this, network operators are beginning to recognise the advantages of an IP based infrastructure.

Web 2.0 describes the new generation of Internet services with the defining theme of these services being the harnessing of collective intelligence. This new age of community websites consists of massive numbers of contributing users and exploits the notion that many people sharing a common experience is preferred to the expert judgement of a few. The majority of these web services are available free of charge and revenues are largely based on the potential for personalised advertising.

The IP Multimedia Subsystem (IMS) is a packet switched communications framework currently being developed to provide network operators with a non-disruptive strategy in the evolution to an IP based Next Generation Network. The IMS is a service enabler that allows services to be integrated horizontally as apposed to the traditional stovepipe architecture; this means a single service can be deployed across different architectures and technologies, greatly reducing capital expenditure and time to market. Furthermore the IMS provides common service enablers thereby encouraging the rapid development of innovative new services. This framework was once seen as the saving grace of incumbent Telco operators because it allowed them to combat eroding voice revenues by changing their business focus to service provisioning. However the huge success and subsequent new communications model introduced by the Web 2.0 revolution poses a threat to wide scale IMS deployment but also provides opportunities. Operators can exploit Web 2.0 experiences and use these to develop specialised community based services. However a strategic business case for deploying IMS services needs to be made - If operators are to charge for services that are largely freely available on the Internet they will need to justify this through service differentiation. The main candidates for achieving this differentiation are increased security, greater service integration with a single sign on for all applications and increased reliability through efficient management of resources.

This paper examines architectures for resource management in the IP Multimedia Subsystem. Standardisation efforts and related works are briefly reviewed and architectural requirements and deployment challenges are identified. A Multilayered Policy Control Framework is presented that addresses the challenges of policy representation, application-policy interaction and access network policy refinement. This framework is deployed in a practical IMS test-bed were it is subjected to rigorous validation. The effect on session setup delay, core network overhead and multiple session establishment is presented and analysed.

II. Related Work

The IMS was initially defined by the Third Generation Partnership Project (3GPP) as a service platform for the evolution of GSM systems. It has since been adopted as a service control system by ETSI Telecommunications & Internet converged Services and Protocols for Advanced Networks (TISPAN) and the International Telecommunications Union - Telecommunication Standardisation Sector (ITU-T) both of which are defining NGN architectures. For resource management

3GPP define the Policy Control and Charging Architecture (PCC) [1], ETSI TISPAN define the Resource and Admission Control Subsystem (RACS) [2] and the ITU-T define the Resource and Admission Control Functions (RACF) [3]. While there are subtle differences in these architectures there are no significant conflicts. Each of these architectures has adopted IETF defined Policy Based Network Management as the means of controlling resources. Essentially a service control layer and a resource control layer are defined; the service control layer requests resources on behalf of applications from the resource control layer and based on pre-defined policies the service is assigned resources in the transport layer. For a more detailed explanation the interested reader is directed to [4]. In this paper we provide a high level overview that describes the interaction between the service control functions defined by the IMS and the resource control functions, which encompasses all of the aforementioned architectures; however we limit our scope to IMS session based services. For the sake of simplicity ETSI TISPAN terminology is used throughout the paper although we refer to the general resource management system.

An open area within the standardisation work is policy refinement and particularly the translation of policies between different levels of the management hierarchy. Policy description models have been defined that perform policy refinement [5], however these models are largely limited to specific applications and need extensions to perform general resource management.

In [6] policy enforcement on specific vendor equipment is examined, here simplistic policy refinement is performed but compex translation models are defined to enforce the policies on technology specific (GMPLS) and vendor specific (Juniper) routers. However these models use proprietary management APIs to configure the devices, making wide scale deployment in a multiple technology, multiple vendor environment complex.

III. Architectural Requirements

Apart from the general functional requirements of a resource management framework [7], there are secondary requirements that need to be met to provide the necessary linkage between session based services and transport layer resources. The aforementioned standardised architectures specify logical architectures and not physical implementations. In addition to in-depth protocol and interface definition there are various challenges that need to be addressed before a physical implementation can be realised.

A. Rapid Service Creation

The primary motivation for IMS deployment is to provide an environment to rapidly develop and deploy highly integrated and innovative services. Applications can be re-used as building blocks, allowing developers to concentrate on the logic specific to the service they are creating. Furthermore standardised APIs allow easy service creation through numerous different programming models. A challenge for the resource management framework is to guarantee and commit resources in the transport layer such that the services are correctly handled and a high Quality of Experience is assured for the end user, but at the same time to encourage the innovative development of new services without indepth knowledge of the underlying network topologies. In the current IMS system, service requirements are described using the Session Description Protocol and based on this admission and resource control is performed. This service classification may hinder the creative process and session policies have been proposed to overcome this problem [8]. Essentially greater interaction is needed between the application developers and the policies that govern resource control. This interaction must be on a level that does not require in-depth network level knowledge.

B. Policy Definition and Refinement

The primary challenge facing wide scale deployment of general Policy Based Network Management systems is the myriad of policy representations at every level of the management hierarchy. The levels within the resource control framework range from high level business policies down to technology and device specific enforcement policies. The challenge here is that policies must be simple enough to ensure easy operator configuration, but also able to efficiently manage complex systems. Policy definition has been seen as network operator specific and hence there has been little standardisation, however with no reference model the creation of new policies becomes a complex task and domain interoperability regarding policy interpretation becomes a serious problem.

In a complex management system it is inevitable that policy conflicts will occur, a mechanism to manage this conflict and ensure correct policy enforcement across managed domains is essential.

C. End to End Policy Provisioning

The envisaged NGN architecture raises some interesting roaming requirements because of the complex QoS provisioning and multitude of services that must be supported. Users should be able to access their service profile from any access network at any time, and should be able to establish multimedia sessions with other users in the same or different network. Assuming each network supports the requested service the resource management framework should guarantee resources and perform policy control along all bearer segments involved, regardless of whether either user is located on their home or a visited network.

Additionally the interworking of existing access networks is a central requirements for NGN architectures. The resource management framework needs to support resource and admission control in both the core and access networks. This involves the creation of access specific policies and a specialised function that performs policy refinement specific to that access network. This policy refinement should also be performed on a per vendor basis, an important challenge will be enforcing policies across networks and vendors without having to tailor each implementation to the devices involved.

IV. Multilayered Policy Control Framework

The IMS architecture will see regular deployment of new services, resulting in a constantly changing network dynamic. Hence the resource management framework will need to be flexible and adaptible; specifically network operators should be able to rapidly create and deploy new management policies. Greater interaction between applications, policies and end users is required to cater for the increasing personalisation of multimedia services. Innovative service creation should be in no way hindered, this requires increased interaction between the application developers and the policies that govern resource control.

We define a Multilayered Policy Control Framework that addresses the areas of policy definition, application-policy interaction, policy conflict and access network policy refinement. Three levels of base policies are defined: Domain Policies, Subscriber Policies and Application Policies. It is important to note that these are just base policies - the architecture is flexible and adaptible, and defines a novel procedure for rapidly creating and deploying new policies.

A. Multilayered Policy Provisioning

The architecture defines extensions to the ETSI defined Service - Based Policy Decision Function (S-PDF) and Access - Resource Admission Control Function (A-RACF).

The architecture consists of two kinds of policies; control policies and enforcement policies. The control policy structure is based on RFC 3644 - Policy Quality of Service Information Model [9]; each policy is made up of rules, these rules define conditions and actions, if a condition is met an action is taken. The S-PDF combines these policies with the service information provided by the service control layer to perform admission control and create enforcement policies. These policies are high level and define how the session based service should treated in the transport layer. These enforcement policies are defined by a 5-tuple, bandwidth requirements and a QoS class. All policies are represented using eXtensible Markup Language (XML); this format has been adopted for representing policy information in several IMS services.

For policy storage the Open Mobile Alliance defined XML Document Management (XDM) Server is utilised [10]. The XDM Server is used for the storage and retrieval of XML documents from a centralised server. The specification uses the XML Configuration Access Protocol (XCAP) to transmit the policy information [11]. This protocol defines HTTP PUT, GET and DELETE methods to store, retrieve and remove documents. The information is carried over a secure HTTP connection and all network elements that need access to the policy repository are equipped with XCAP interfaces.

Base control policies are Domain, Subscriber and Application policies. Domain policies will be set by the network operator and represent general network constraints. Subscriber policies refer to the subscription state of each user and define the registered services and service priorities. Application policies are specific to each deployed IMS service, a generic format of this policy is created that essentially defines different quality levels for the service. These policies can be extended using high level management tools (likely a web interface) and it is envisaged that the application developer could create these policies as they deploy their services. Basic Application policies have been defined for standard services including VoIP, IPTv and Video on Demand. End users have controlled input to these policies. Application policies differ from typical service preferences because they provide critical interaction between the applications and the allocation of resources in the transport layer.

Access network specific control policies are defined as an extension to the Domain policies; these policies include more specific network constraints (e.g. bandwidth, latency, etc.). Basic access specific policies have been defined for LAN, EDGE and HSDPA technologies. These policies refine the access and technology independent enforcement policies which are eventually passed down to the transport layer devices where they are converted to vendor specific configuration information and used to configure the network devices.

B. Policy Processing and Profiles

Both the S-PDF and A-RACF receive resource requests from the service control layer, perform admission and resource control, and create enforcement policies. These elements are extended with a modular design where a Decision Engine (DE) carries out the admission control and enforcement policy creation procedures. The DE is made up of Policy Processor Blocks that are defined for each control policy type (Domain, Subscriber, etc.), these modular classes are specific to each policy type and admission control and enforcement policy creation is carried out on a per policy type basis.

This modular design means that deploying a new control policy involves only creating the policy structure and storing it on the XDM Server, and defining a policy specific Policy Processor Block. This flexible system allows network operators to adapt to changing network dynamics by rapidly adding or removing control policies like building blocks. The extended S-PDF logical architecture is shown in Fig. 1.

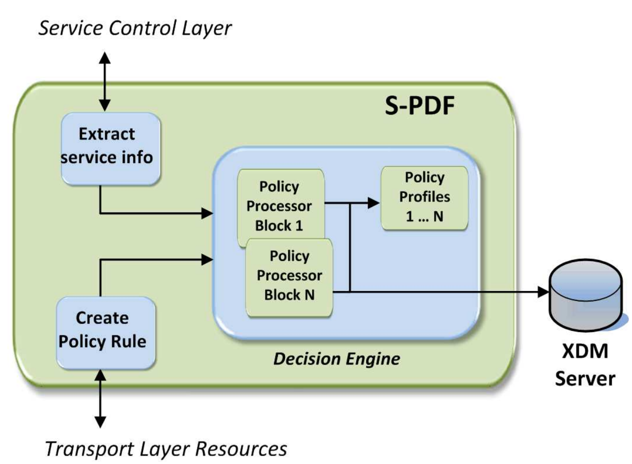

Figure 1. Extended S-PDF architecture.

With multiple levels of policies a deadlock situation is possible where policies specify different constraints. We propose a prioritisation mechanism where for each deployed control policy we define a policy profile. This profile would define the priority of that policy and also filter criteria that describe which kind of resource requests should invoke that policy. Complex filter criteria can be defined that result in policies being invoked based on service parameters, e.g. media type, codec information, etc.

The priority of each control policy is defined by a non zero positive integer where 1 is the highest priority. Control policies are invoked in a serial fashion as shown in Fig. 2, and the higher the priority of the control policy the sooner it is invoked in the resource and admission control procedure. A higher priority policy creates a base enforcement policy on which lower priority policies must build. This presents a compelling business case where 3rd party service providers could pay for higher priority application policies and hence access to more resources.

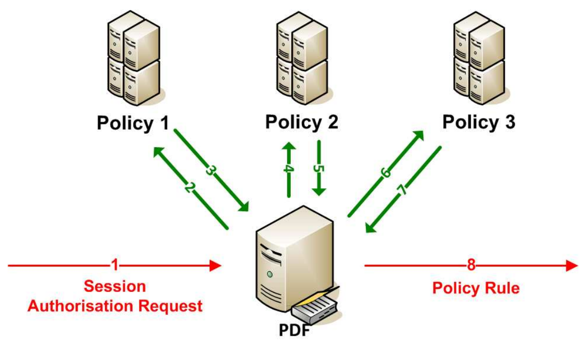

Figure 2. Control policies invoked in a serial fashion based on priority.

C. Architecture Limitations

The proposed architecture introduces policy profiles, multiple levels of policy and a mechanism to increase interaction between session based services and transport layer services. However it does not address the issue of inter domain and end to end QoS reservation. Additionally the architecture controls transport layer resources and does not address the issues of inter-application interaction or service brokering.

V. Testbed Validation

The proposed architecture was implemented in a proof of concept testbed where it was subjected to rigorous validation tests. While the Multilayered Policy Control Framework defines extensions to the general resource management system the interface and protocol definitions defined by the 3GPP PCC were used for implementation purposes. To provide a convenient point of departure for further research and innovation in the field Free and Open Source software is used throughout the implementation.

An extended version of the Open Source IMS (OSIMS) Core was used to facilitate a standards based IMS reference framework [12]. This software package provides implementations of the Proxy/Interrogating/Serving-Call Session Control Functions (CSCF); it is based on 3GPP standards, is widely used and provides a reliable testing environment. These elements comprised the service control layer of the testbed and the P-CSCF was extended to extract service information from the session establishment messages and pass this to the resource control layer via a standardised Diameter interface.

The resource control layer used the UCT Policy Control Framework to define the S-PDF and A-RACF elements [13]. This framework was extended to implement multilayered policy control and policy profiling. A standards based Diameter interface was defined to the service control layer to receive resource requests and to the transport layer to pass enforcement policies. OpenXCAP, an open source implementation of the XDM specification was used for policy storage [14]. The Transport Layer was defined by Linux Software DiffServ routers, the access specific but vendor independent enforcement policies were translated to Linux traffic control parameters to configure the router to mark and queue packets accordingly. The UCT IMS Client, a standards based IMS Client emulation tool was used to represent IMS terminals [15]. For load testing the SIPp protocol testing software package was utilised.

The OSIMS, Resource Control Elements and OpenXCAP server were run on a single high performance Intel Dual Core 3.0Ghz machine. The end user terminals represented by the UCT IMS Client and SIPp package were run on 1.6GHz Intel Centrino laptops. The testbed supported 3 different access technologies: HSDPA, EDGE and LAN. For the HSDPA and EDGE access technologies the machine running the core elements was deployed in the IP Core of a well known South African cellular operator; this prevented public Internet routing delays and allowed for realistic evaluation of a carrier-grade IMS deployment. For the LAN access technology the IMS terminals connected directly to the core machine.

A. Core Signaling Overhead

To analyse the increase in traffic overhead introduced by the Multilayered Policy Control Framework we inspected the traffic volumes at the machine running the core elements both with and without QoS provisioning ranging from a single resource request to 5 simultaneous requests. This test was carried out over the LAN access technology only and it was noted that 5 simultaneous requests was sufficient to establish a trend in the data. Each request represented a typical IMS session with both video and audio components. With QoS provisioning enabled 3 policies were invoked, this involves at least an additional 30 messages within the core if both clients are in the same domain. Traffic was measured on the reference points between all elements running on the core machine and results were averaged over 5 tests runs. Table I shows the results, this analysis gives us an indication of the signaling overhead introduced by such a policy control framework and whether it might overwhelm network elements and decrease network utility.

Table I
AVERAGE SIGNALLING OVERHEAD WITH AND WITHOUT QOS PROVISIONING.

No. Session Requests	1	2	3	4	5
With QoS Provisioning (kB)	693.462	1618.13	2817.19	4085.73	6831.27
Without QoS Provisioning (kB)	47.056	87.323	128.866	170.026	220.864

B. Session Setup Delay

To analyse the effect on session setup delay when incorporating the Multilayered Policy Control Framework we measured the session setup delay with QoS provisioning enabled across different access networks and compared these results to standard IMS session setup results from [13]. The delay was measured using the built in UCT IMS Client timer which started at the first DNS lookup and ended on the receipt of the 180 ringing in response to the initial INVITE. Table II shows the standard session setup results from [13] and table III shows the session setup delays when incorporating the Multilayered Policy Control framework.

Table II
IMS CALL SETUP DELAY RESULTS WITHOUT QOS PROVISIONING [13].

	LAN	HSDPA	EDGE
Minimum (s)	0.82	3.21	20.27
Mean (s)	1.05	3.95	27.56
95th Percentile (s)	1.27	5.25	45.87
Std Dev (s)	0.23	0.75	12.46

Table III
IMS CALL SETUP DELAY RESULTS WITH QOS PROVISIONING.

	LAN	HSDPA	EDGE
Minimum (s)	0.82	3.33	12.60
Mean (s)	1.53	4.03	18.12
95th Percentile (s)	1.95	5.76	32.55
Std Dev (s)	0.39	1.49	6.02

C. Simultaneous Session Requests

In a commercial environment one could expect multiple session requests per second, the increase in signaling could cause unacceptable signaling overheads and session setup delays. To analyse these effect we repeated the session setup delay tests over the LAN access technology. Additionally we used the SIPp protocol testing package to inject session requests into the IMS core at a rate of 8, 10, and 12 requests per minute. The session setup delay was recorded for each scenario both with and without QoS provisioning enabled over 5 test runs. Fig. 3 shows the session setup delay over the LAN access technology under differing core loads.

D. Discussion

The observed increase in traffic between the core elements when utilising the Multilayered Policy Control framework is not insignificant, but it is still within an acceptable range

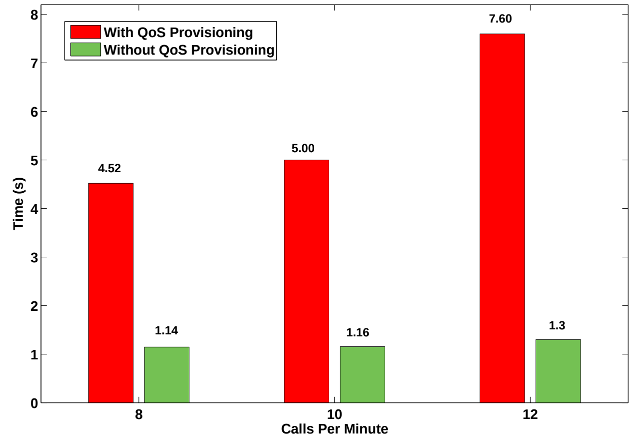

Figure 3. Session setup delay under different loads with and without QoS Provisioning.

when considering the high performance machines that would be deployed in a commercial environment.

The session setup delays for a single request showed unacceptable delays for the EDGE access technology, this is largely the fault of SIP retransmissions being triggered too easily. Once again the increase in session setup delay when incorporating QoS provisioning is not insignificant but for the faster LAN and HSDPA access technologies the delays are still acceptable. When under varying loads we note a sharp increase in session setup time with QoS provisioning while the status quo architecture is not significantly affected. This can possibly be attributed to the limited performance of the server machine. However optimisation techniques need to be investigated to mitigate these effects.

VI. CONCLUSION

This paper has highlighted architectural requirements and deployment challenges for resource control architectures in the IMS, and examined how session based services can be linked to transport layer resources in this environment. A Multilayered Policy Control Framework has been proposed that addresses the open areas of policy representation, policy profiling, application-policy interaction and policy processing. The proposed architecture was implemented in the form of a standards based test-bed where it was subjected to realistic use case scenarios. The increased traffic overhead and session setup delay was observed to be in an acceptable range for a single session request. However when the system was load tested with many simultaneous requests the session setup delay increased dramatically.

Future work includes investigating optimisation techniques to mitigate the effects on session setup delay with many simultaneous requests, and inter domain communication to achieve true end to end policy control.

REFERENCES

[1] 3GPP, “TS 23.203 Policy and Charging Control Architecture,” March 2007.

[2] ETSI TiSPAN, "ES 282 003 Resource and Admission Control Subsystem (RACS) Functional Architecture," 2006.
[3] ITU-T, "Y 2111 Resource and Admission Control Functions in Next Generation Networks," September 2006.
[4] J. Song, M. Chang, and S. Lee, "Overview of ITU-T NGN QoS Control," *IEEE Communications Magazine*, pp. 116–123, September 2007.
[5] C. Kamienski, J. Fidalgo, R. Dantas, D. Sadok, and B. Ohlman, "Design and Implementation of a Policy-based Management Framework for Ambient Networks: Choices and Lessons Learned," *2008 20th IEEE/IFIP Network and Operations Management Symposium*, April 2008.
[6] F. Baroncelli, B. Martini, V. Martini, and P. Castoldi, "Supporting Control Plane-enabled Transport Networks within ITU-T Next Generation Networks (NGN)," *2008 20th IEEE/IFIP Network and Operations Management Symposium*, April 2008.
[7] E. TISPAN, "TS 185 001 QoS Framework and Requirements," 2005.
[8] G. Camarillo, T. Kauppinen, M. Kuparinen, and I. M. Ivars, "Towards an Innovation Oriented IP Multimedia Subsystem," *IEEE Communications Magazine*, pp. 130–135, March 2007.
[9] Y. Snir, Y. Ramberg, J. Strassner, R. Cohen, and B. Moore, "RFC 3644 - Policy Quality of Service Information Model," November 2003.
[10] Open Mobile Alliance, "PoC XDM Specification version 1.0.2," September 2007.
[11] J. Rosenberg, "RFC 4825 - The Extensible Markup Language Configuration Access Protocol," May 2007.
[12] Fraunhofer Institute FOKUS, "Open Source IMS Core," *http://www.openimscore.org/*.
[13] D. Waiting, R. Good, R. Spiers, and N. Ventura, "Open Source Development Tools for IMS Research," *2008 4th International Conference on Testbeds and Research Infrastructures for the Development of Networks and Communities*, March 2008.
[14] M. Amarascu, "OpenXCAP," *http://www.openxcap.org/*.
[15] University of Cape Town, "UCT IMS Client," *http://uctimsclient.berlios.de/*.

Richard Good received his B.Sc. (Hons) degree from the University of Cape Town in 2005. He is currently working towards his Ph.D. at the same insitution. He is an active open source software contributor and has developed various open source IMS tools. His research interests include next generation resources management, service provisioning and QoS in heterogeneous networks.