Route Selection Impacts on Achieving Enhanced IMS QoS

Vitalis G. Ozianyi♀, Gabriel Andrews, Neco Ventura
University of Cape Town,
Electrical Engineering
7700 – Rondebosch
South Africa
Tel + Fax: 0216505296
vitozy♀, gandrews, neco@{crg.ee.uct.ac.za}

Abstract-**The different planes in the IMS interact via specific reference points to deliver multimedia services to the user. QoS provisioning for IMS communications has been standardized for access networks only, with the assumption of an over provisioned IP core. Effective provisioning of multimedia services requires performance guarantee along the complete path of the sessions. End-to-end QoS in IP networks is affected by the route traversed by the user traffic. Moreover QoS guarantees in one ISP domain are not effective for transit traffic exiting the domain. QoS extensions to exterior gateway routing protocols have been proposed to transfer route QoS information beyond one autonomous system (domain). This paper explores options for mapping inter-domain QoS information learnt on the media plane into control plane session information for IMS QoS control. Through testbed evaluations we show the effect of routing on delays experienced in IMS communications.**

Index Terms—**E2E, IMS, ISP, QoS, SIP.**

I. INTRODUCTION

Global IP communications are affected by the existence of different domains along the end-to-end (E2E) path between communicating nodes. Different network domains offering Quality of Service (QoS) guarantees would often use different definitions for QoS classes, for example the DiffServ code points used between adjacent DiffServ domains may translate to dissimilar QoS metrics in terms of guaranteed minimum delay or bandwidth across the domain for the same traffic class. This creates challenges for mapping of QoS classes to actual packet handling procedures, referred to as the Per Hop Behaviors (PHB), in inter-domain DiffServ scenarios. Service Level Agreements (SLA) are used in commercial deployments to specify expected network performance levels for traffic entering or leaving a domain. SLAs are only applicable to client – provider scenarios, where the provider agrees to handle the client's traffic by meeting defined QoS parameter tolerances. Thus the SLA may be defined in terms of bandwidth and one-way delay and jitter etc. It should be noted that SLA are non-transitive; hence QoS guarantees do not hold beyond the provider's domain [1].

This feature can lead to service discontinuity due to lack of E2E performance guarantee. The IP Multimedia System (IMS) was developed to facilitate delivery of real-time and non-real-time (streaming) multimedia services over merged cellular and IP networks [2]. IMS relies on the Session Initiation Protocol (SIP) for session establishment and control. SIP uses the session description protocol to convey useful information about sessions, e.g., the media codec and bandwidth requirements between IMS User Agent (UA) and the network. Delays experienced in IMS communications accrue from signaling and media traffic delays, and this would affect the user experience.

The route traversed by IMS (signaling and media) traffic would affect the achieved QoS. End-to-end path delay, which is additive along the route, would adversely impact real-time communications, e.g., VoIP telephony and IPTV. Thus it is essential for IMS operators to route traffic via ISPs that meet the required performance metrics. The common business models for Internet communications are peer, wholesale and retail. In wholesale and retail models peering Internet Service Providers (ISP) charge for traffic that is relayed through their networks [1]; the retail pricing model accounts for the level of quality offered to traffic. Internet users would subscribe to IMS services from various operators who may also provision the access network for user traffic.

Operators may present users with differentiated levels of network performance that are accompanied with corresponding pricing options. This would translate to the QoS guarantee that is negotiated and applied to traffic that is sent to other networks via various upstream ISPs. Traffic from premium pricing profiles would be routed via ISPs that meet the required QoS guarantees – for example high availability and 1+1 optical and IP protection, whereas flat-rate priced traffic would be routed via the cheapest wholesale priced ISP – say providing a 1+N optical protection only. By using exterior gateway routing protocols like the Border Gateway Routing Protocol (BGP), the domains along the E2E path do not reveal the internal structure of their networks. Only reachability information is advertised; however, more information is required to decide the likeliness of the path available via a peering ISP to meet the E2E QoS requirements. This information would be useful in updating the IMS network QoS resource manager on the availability of resources along the path for various applications and user pricing profiles.

Accurate information on the E2E status of the available communication paths would be used by various application functions to perform intelligent operations and make decisions to accept or reject user service requests. In this paper we explore schemes for obtaining E2E QoS information across multiple domains and propose the integration of the learnt QoS information into IMS QoS control. We present testbed results showing the affect of path selection on E2E delays experienced by IMS procedures. The rest of the paper is structured as follows: section II presents a review of background information; section III presents a proposal for the integration of QoS information learnt on the IMS media plane with IMS QoS control; section IV presents a lab test platform for IMS QoS performance evaluation by emulating global routing; section V presents evaluation results from the lab prototype; section VI presents discussions; section VII concludes the paper; and section VIII points out some future work.

II. SCALABLE INTER-DOMAIN IP COMMUNICATIONS

Supporting inter-domain QoS is necessary for achieving continuity of traffic flow as network conditions change across domains from the source to the destination of communicating nodes. Continuity is defined in terms of QoS parameters, i.e., maintaining the negotiated bandwidth level, packet delays, packet loss ratio etc. in the course of an active communication. Service Level Agreements that govern QoS guarantees are normally negotiated between neighboring domains (or ISPs) only, yet effective QoS guarantees are required along the complete path of the communication. Each domain defines the QoS guarantee using specific values for selected QoS parameters, e.g., bandwidth and delay. If E2E bandwidth is the desired network performance parameter, the effective bandwidth is constrained by the minimum bandwidth offered by a domain along the E2E path; whereas the resultant path delay is an additive function of path delays across all domains that are traversed [1][3]. It is thus essential for an IMS provider to select routing paths that meet E2E QoS constraints for user and signaling traffic. Multimedia traffic, e.g., VoIP is specifically sensitive to E2E delay and jitter.

In addition to using QoS metrics for defining QoS guarantees, domains define classes of service (CoS) for traffic belonging to different applications – this is inherently a differentiated services (DiffServ) characteristic [4]. For example a Premium CoS would offer better network transport characteristics than an Olympic CoS; however, the values of different QoS metrics for a given CoS may differ between adjacent domains. Moreover, SLAs between adjacent domains are non-transitive [1]. Achieving inter-domain QoS has been investigated through proposals for adopting common Service Level Specification (SLS) templates [5] [6]. Despite these efforts, each domain retains autonomy on its QoS guarantee levels. Mismatch in QoS support mechanisms can lead to service discontinuity for multi-media applications that negotiate QoS at session establishment.

Consistent SLS and QoS class definitions would facilitate E2E QoS negotiations for IMS services. In the IMS QoS provisioning makes the assumption that the IP core network is over-provisioned and would meet required QoS guarantees for multimedia traffic. Thus it is assumed that the QoS constraint lies in the IP connection access networks (IP-CAN). Generally any IP-based access network can be used to access IMS services; however, the 3GPP is standardizing different IP-CANs for inter-working with the 3G UTRAN [7]. These IP-CANs exhibit characteristic QoS support; however, some access technologies, e.g., the popular 802.11(a,b,g) do not support traffic quality differentiation. IMS session information transported in SIP messages is used in session management including QoS control. These messages are formatted according to the session description protocol, and are extracted at the Proxy Call State Control Function (P-CSCF) for use in policy control [2].

SIP messages are accessible only by authorized entities, i.e., CSCFs and the communicating terminals where the IMS UA is running. The terminals negotiate QoS parameters using pre-defined information for required QoS parameters of the service. This defines the QoS guarantee required of the network, and is essentially as a result of operational policies using best practice performance for the requested service [6]. Once the QoS request is made the application functions, e.g., P-CSCF, and the Policy Decision Function (PDF) that are in charge of the user's access domain perform admission control and QoS resource reservation. It should be noted that admission control and resource reservation in the IMS is performed with respect to individual IP-CANs only. As mentioned earlier the assumption made is that the IP core network is over provisioned to accommodate all offered traffic. In practice IP core networks have limited capacity; hence QoS degradation will be experienced during congestion periods. To ensure better performance for some traffic types network operators negotiate SLAs with neighboring ISPs.

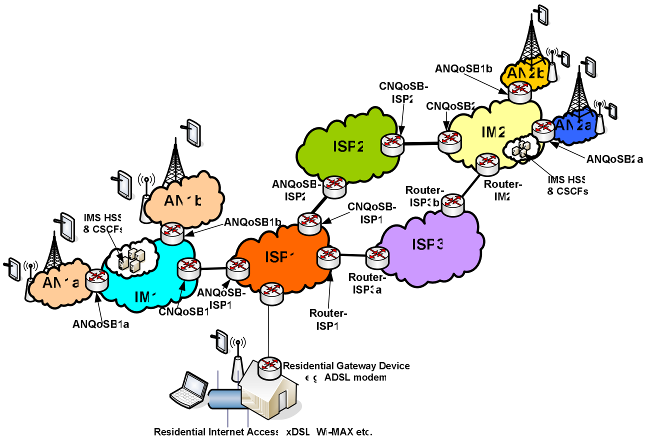


Fig.1: Internet Communications with E2E QoS Scenario

As mentioned earlier, SLAs are non-transitive; thus the guaranteed QoS as specified in an SLA will not be effective when traffic exits the ISP's network. Figures 1 depicts this scenario - where an SLA between IM1 and ISP1 is not valid when traffic exits ISP1 into ISP2 or ISP3. To extend the QoS guarantee beyond the adjacent ISP would involve E2E SLA negotiation mechanisms. This is where the serving domain (IM1) negotiates with neighboring domains (ISP1) for QoS resources; the negotiation proceeds along the E2E path (through ISP2 into IM2)[1]. Aggregation of QoS guarantees for traffic in QoS classes [6] [4] would maintain the scalability level required when speedy session establishment and re-establishment is imminent. Otherwise QoS negotiation along full communication path would affect service continuity and the perceived quality, an aspect that makes the schemes discussed in [1] not feasible since they should be enforced for each session establishment request; however, the relation between the cost of service provision and the SLA enforcement scheme highlights information that is beneficial to operators and ISPs.

Maintaining of QoS state information on the negotiated paths for every session in core networks that are of the magnitude global IMS will take faces scalability problems. Handling of QoS resource reservations in aggregate blocks according to the DiffServ architecture in core networks achieves scalability, and when combined with per-session resource allocation in access networks better QoS guarantee may be achieved. The aggregate resource reservations should be designed carefully to maintain high efficiency of resource utilization, as well as admit the maximum sustainable number of user sessions.

III. END-TO-END QOS CONTROL FOR THE IMS

ISPs and networks deploying various service delivery platforms (SDP) with QoS support enforce schemes for managing local and Inter-domain resources. In the release 7 of 3GPP standardization of the IMS [9], QoS policy control was defined as part of the Policy Control and Charging (PCC) architecture. Details of IMS policy control can be found in [10] and [11]. In attempt to extend the QoS guarantee to the IP core network, domains that adopt DiffServ-based QoS provisioning proposals would define a platform with resource managers to manage the domain's local resources and resources provisioned from external domains – inter-domain resources. With respect to IMS scenarios, an access domain would consist of IP-CAN and IP core networks, whose resources are under the same administrative control [12], see also Fig. 1. In a network enforcing QoS resource control for an access domain with multiple access networks, e.g., the DAIDALOS framework, access networks receive QoS resources from the core network. QoS managers (QoS brokers) in the access networks and the core network control resource allocation and utilization [13]. With respect to the IMS the Access Network QoS broker (ANQoSBr) is actually the PDF; there is no QoS management entity would map to the Core Network QoS broker (CNQoSB). Inter-domain resources are acquired when the CNQoSB requests for aggregate resources from the ANQoSB in charge of the ingress point of the neighboring domain – in Fig. 1 CNQoSB1 would request for inter-domain resources from ANQoSB-ISP1. These resources would be provisioned at the direction of the CNQoSB of the neighboring domain, i.e., CNQoSB-ISP1.

The aggregate resources provisioned by CNQoSB-ISP1 above will be used to convey traffic via ISP2 into IM2 where the destination node is located. The destination node could be an application server (e.g., IPTV server) or a user agent (mobile or fixed node).

Since ISP2 provides QoS guarantees for traffic from ISP1 the achieved network performance for traffic on that route will be high, or at least satisfactory. Alternatively traffic destined to an access network in IM2 may be routed via ISP3. As seen in Fig. 1 the routers at the borders of ISP3 and ISP1 and IM2 do not have QoS negotiation capabilities. They simply handle traffic on a best effort basis, thus ISP3 doesn't provide any QoS guarantee. ISP3 would normally use a wholesale pricing scheme [1] to sell aggregate bandwidth blocks to ISP1. The strategy used by ISP1 in traffic handling involves routing traffic with tight QoS requirements, and for which a premium CoS has been set via ISP2. Traffic with less strict QoS requirements and/or traffic requiring high network performance guarantees but for which a non-premium CoS[2] has been set would be routed via ISP3.

In terms of admission control, if a feedback method (as discussed in [6] and [8]) is used to communicate the congestion and resource availability state on the QoS enabled path to IM1 and its IP-CANs, informed QoS policy control can be achieved. In the TEQUILA approach this would be triggered in the normalization phase of admission control management; as part of the severe admission control strategy to ensure on-going services receive an almost satisfied performance. Interactions between the PDF in charge of an IP-CAN, the P-CSCF or a media application server and the IMS client will facilitate the establishment of a media session with appropriate QoS support. SIP based applications, which negotiate QoS at session establishment or re-establishment, can adapt to the available E2E resources by using codec settings that perform satisfactorily with the available resources. Services set by users to use a premium CoS (or profile [14]) would be allocated the requested bandwidth in the IP-CAN and their traffic would be marked for routing via ISP2 when forwarded to ISP1. Services in other CoS or profiles would be allocated the required bandwidth in the IP-CAN, but would be marked for possible routing via ISP3 if congestion thresholds in ISP2 are exceeded.

Since inter-domain resources from ISP1 will be bought in aggregate blocks, economic and network efficiency should be maintained to ensure IM1 doesn't pay for more resources than required by its users. Through class promotion [15], services using non-premium CoS and profiles can be granted better QoS, i.e., be routed via ISP2 when ISP3 reports congestion during periods when the subscribed capacity via ISP2 is underutilized. This will boost overall user satisfaction and benefit the operator. One of the main performance differences between traffic routed via ISP2 and ISP3 is the delay likely to be experienced on the ISP3 route during congestion periods.

E2E packet delay and jitter are common challenges to the delivery of real-time multimedia services. Modern real-time applications like VoIP may tolerate some bandwidth fluctuations through adaptation procedures involving the use of low bandwidth codecs; however, excessive delay and jitter can lead to session failure. In the next section we present a laboratory prototype of an IMS network with which we investigate the

---

[1] ISP3 is left out since we assume the lack of SLAs between ISP1 and ISP3

[2] The selected CoS would determine the price the user of IM1 pays.

delay experienced in performing IMS registration and session setup when traffic is routed via a congested route (ISP3) and when direct routing is enforced (ISP2).

IV. IMS TESTBED ARCHITECTURE FOR OPTIMIZED ROUTING TESTS

A lab prototype for network performance evaluations was laid out as shown in Fig. 2. It illustrates a scenario with three IMS domains that are interconnected. Each IMS network allows users to register and setup voice calls to other users registered and connected to the same domain or other domains. The IMS networks are deployed using the Open Source IMS (OSIMS) developed by the Fraunhofer Fokus institute in Germany [16]. The clients used in the tests utilize the UCT IMS client[3] developed by the Communications Research Group at the University of Cape Town in South Africa [17]. The challenge of achieving E2E QoS for multimedia communications is investigated by routing IMS traffic via a less congested network and also via a congested network. The network in Fig. 2 corresponds to Fig. 1; hence the discussions given above apply in a one-to-one fashion.

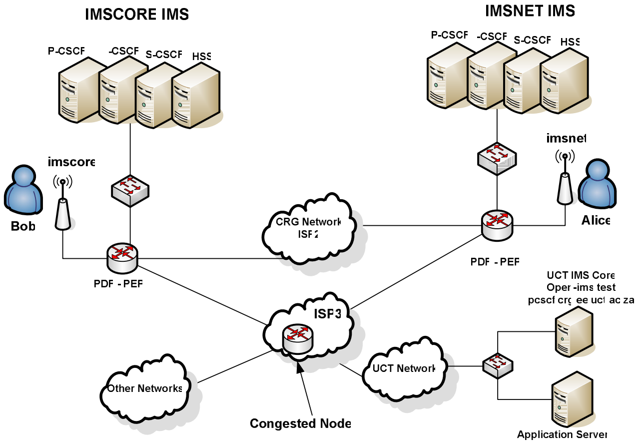


Fig. 2: Evaluation Testbed

We investigate the delay incurred in performing IMS registration using a wireless client accessing the network via the *imsnet* access point (AP) and another client via fixed Ethernet on *imscore.ims*. Delay analysis for media session establishment between two IMS clients is performed by connecting one client to the *imsnet* AP and another client to fixed Ethernet on *imscore.ims*. By this procedure it is also possible to account for delays that may result from accessing the network via the air interface (WLAN). Using a defined baseline, delays due to the selected route are determined.

To register with an IMS domain the user connected to the *imsnet* AP sends a *register* request to the P-CSCF of the associated realm (e.g., a request is sent to pcscf.imscore.ims), and subsequently to pcscf.imsnet.ims and pcscf.crg.ee.uct.ac.za, which is the P-CSCF of the open-ims.test realm. Voice call sessions are established by sending an *invite* request to a callee by specifying the IMS public user identifier (IMPU). For example a user registered as *bob@imscore.ims* tries to call another user of IMPU *alice@imsnet.ims*.

Registration and session establishment transaction use SIP signaling. SIP is heavy in message size, due to its text-based nature [18]. Thus, the amount of network bandwidth consumed by SIP control messages increase with the number of users. But the delivery time of the messages should be kept low, regardless. In the next section we detail the tests conducted on the lab testbed to measure the performance delay as a QoS constraint.

V. EVALUATIONS TESTS AND RESULTS

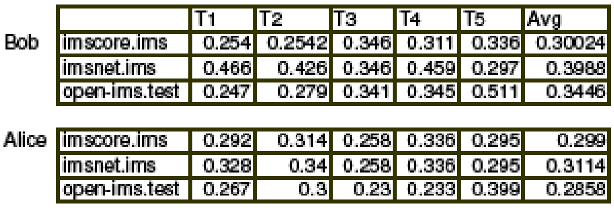
These are initial test results demonstrating the functioning of the testbed as a test tool for network performance measurements. The performance tests were conducted as detailed below; delay measurements were recorded for each procedure.

A. *Registration delay*

1) *Routing via ISP2*

Bob connects to *imscore* and sends a registration request to imscore.ims, imsnet.ims and open-ims.test in succession. Alice then connects to *imsnet* and registers with imsnet.ims, imscore.ims and open-ims.test. Each procedure is repeated 5 times and averages are recorded; the measured registration delay is given in table 1, and graphically in Fig. 3.

TABLE 1: ISP2 REGISTRATION DELAY



		T1	T2	T3	T4	T5	Avg
Bob	imscore.ims	0.254	0.2542	0.346	0.311	0.336	0.30024
	imsnet.ims	0.466	0.426	0.346	0.459	0.297	0.3988
	open-ims.test	0.247	0.279	0.341	0.345	0.511	0.3446
Alice	imscore.ims	0.292	0.314	0.258	0.336	0.295	0.299
	imsnet.ims	0.328	0.34	0.258	0.336	0.295	0.3114
	open-ims.test	0.267	0.3	0.23	0.233	0.399	0.2858

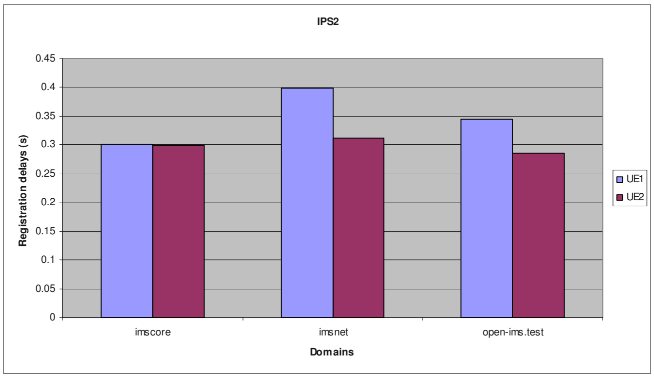


Fig.3: ISP2 Registration delays

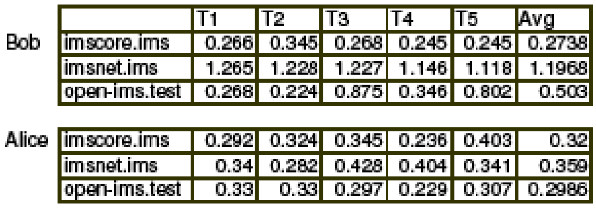
2) *Routing via ISP3*

All traffic between imscore.ims and imsnet.ims is routed via ISP2 and that between open-ims.test and the other domains is routed via ISP3. Bob and Alice register to the IMS domains as described in A..1) above. Table 2 shows the registration delay, and Fig. 4 graphically represents the results.

---

[3] The UCT IMS client has an inbuilt timer to record the time delay for various events, e.g., registration and session setup.

TABLE 2: ISP3 REGISTRATION DELAYS



		T1	T2	T3	T4	T5	Avg
Bob	imscore.ims	0.266	0.345	0.268	0.245	0.245	0.2738
	imsnet.ims	1.265	1.228	1.227	1.146	1.118	1.1968
	open-ims.test	0.268	0.224	0.875	0.346	0.802	0.503
Alice	imscore.ims	0.292	0.324	0.345	0.236	0.403	0.32
	imsnet.ims	0.34	0.282	0.428	0.404	0.341	0.359
	open-ims.test	0.33	0.33	0.297	0.229	0.307	0.2986

B. *Call setup*

1) Routing via ISP2

After Bob registers on imscore he then calls Alice, who is first registered on imscore, then imsnet. Session establishment delay is measured. Each step is repeated 5 times; the results are given in table 3 (for ISP2 and ISP3).

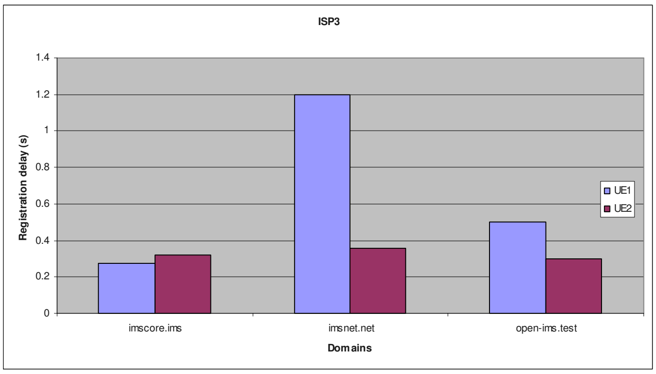
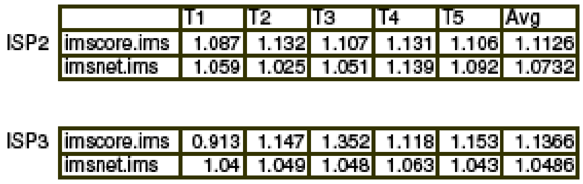


Fig.4: ISP3 Registration delays

2) Routing via ISP3

The steps in C.1 above are repeated with routing done via ISP3. Figure 5 graphically depicts the call setup delay.

TABLE 3: CALL SETUP DELAYS



		T1	T2	T3	T4	T5	Avg
ISP2	imscore.ims	1.087	1.132	1.107	1.131	1.106	1.1126
	imsnet.ims	1.059	1.025	1.051	1.139	1.092	1.0732
ISP3	imscore.ims	0.913	1.147	1.352	1.118	1.153	1.1366
	imsnet.ims	1.04	1.049	1.048	1.063	1.043	1.0486

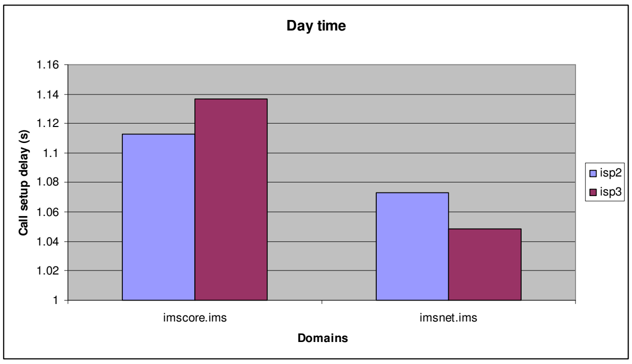


Fig. 5: Call setup delays

VI. DISCUSSIONS

The results indicate a considerably large[4] increase in registration and call setup delay for IMS procedures when traffic traverses a non-optimized path. In these tests the optimization was in terms of network congestion in the IP core. Although traffic to open-ims.test traverses ISP3, lower registration delays are seen in Fig. 3 and Fig. 4. This is attributed to the layout of IMS entities in each domain. All entities in open-ims.test are co-located in one box, thus SIP and diameter interactions are localized in the system and don't incur delays in the protocol stack or DNS query related delays. However, it is evident that lower registration delays are incurred when traffic is routed via ISP2. The same trend applies to the call setup transactions for imscore; imsnet depicts an interesting trend. This will be subjected to further analysis to determine in which part of the network was most call setup delay incurred. Depending on the IMS layer where the bulk of call setup signaling occurs, the reason why ISP3 records lower delays would be identified.

Although it is assumed that IP core networks would generally be over-provisioned, we have reviewed the effect a congested route can cause to the timely delivery of high quality multimedia services. Large delay values would have adverse effects on mobility procedures, e.g., session mobility that is handled through refreshing of registrations and re-invites. It is necessary that session re-establishment during mobility be completed in a very short time so that users do not notice disruption in service – achieving this is termed as seam-less mobility. Thus it would be important to use traffic engineering procedures, as reviewed in this paper to find and enforce routes that provide the minimum possible E2E delay.

VII. CONCLUSIONS

Heterogeneous NGNs serve users with diverse needs of QoS and using different pricing schemes that are enforced by different business models. Users who pay fees for premium IP multimedia services will expect to access those services with good network performance. The IP-CANs used to access the services may be provisioned by the users on networks without active SLAs defining delay bounds; however, if the network access is provisioned by the IMS operator it would be imperative to extend QoS guarantee beyond the IP-CAN into the IP core network and across multiple domains. In this paper we presented a discussion of E2E QoS provisioning mechanisms, and applied this to IMS frameworks. Through initial evaluations on a testbed we showed how the use of a congested ISP can impact IMS procedures like registration and session setup. Still more conclusive evaluations need to be done as discussed below.

VIII. FUTURE WORK

More evaluation tests using scenarios that increase the network load will be done to determine the scalability and validity of the work discussed in this paper. Further tests on the

---

[4] The introduction of one additional hop (congested domain) along the E2E path results in upto 800ms registration delay for Bob on imsnet; thus the existence of more domains without QoS guarantee would adversely affect IMS performance.

signaling load between different IMS entities will be conducted to account for their contribution to the measured delay.

REFERENCES

[1] Panita Pongpaibool and Hyong S. Kim, “Guaranteed Service Level Agreements across Multiple ISP Networks”, IEEE DRCN, pp. 325-332, Oct. 2003.

[2] G. Camarillo, Miguel A, “The 3G IP Multimedia Subsystem (IMS): Merging the Internet and the Cellular worlds” 2nd. Ed. John Wiley & Sons, 2006.

[3] Prior R, Sargento S, “Inter-Domain QoS Routing with Virtual Trunks”, IEEE ICC, pp. 139 – 146, June 2007.

[4] S. Blake et al, “An Architecture for Differentiated Services”, IETF RFC 2475, Dec. 1998.

[5] Goderis D, “Internet design for SLA delivery from service level agreement to per-hop behavior: a report on the 1st TEQUILA workshop”, IEEE Comm, vol. 36 no. 6, pp. 70-72, June 2001.

[6] Mykoniati E et al, “Admission control for providing QoS in DiffServ IP networks: the TEQUILA approach”, IEEE Comm, vol. 41 no. 1, pp. 38-44, Jan. 2003.

[7] F. G. Márquez et al, “Interworking of IP Multimedia Core Networks Between 3gpp and Wlan”,

[8] Prior R, Sargento S, “Inter-Domain QoS Routing with Virtual Trunks”, IEEE ICC, pp. 139 – 146, June 2007.

[9] 3GPP, “Policy and Charging Control Architecture”, 3GPP TS 23.203, vol. 7.1.0, Dec. 2006.

[10] Richard Good, Neco Ventura, “An end to end QoS management framework for the 3GPP IP multimedia subsystem”, IEEE ICT-MICC, pp. 605 – 610, May 2007.

[11] Fabricio C. de Gouveia, Thomas Magedanz, Richard Good, Neco Ventura, “The role of open ims testbeds in complex service delivery platforms”, AFRICON, pp. 1-7, Oct. 2007.

[12] Vitalis G. Ozianyi, Vitor Jesus, Susana Sargento, Rui Aguiar, Neco Ventura, “Virtual Network Capacity Expansion through Service Outsourcing”, IEEE WCNC, pp. 3208 – 3213, Mar – Apr. 2008.

[13] Azevedo R et al, “End-to-end QoS implementation in a B3G network”, IEEE AICT/SAPIR/ELETE, pp. 122 – 127, July 2005.

[14] Vitalis G. Ozianyi, Neco Ventura, Eugene Golovins, “A novel pricing approach to support QoS in 3G networks”, Computer Networks, Volume 52, Issue 7, pp, 1433-1450, May 2008.

[15] E.W. Fulp and D.S. Reeves, “Optimal provisioning and Pricing of Differentiated Services using QoS Class Promotion”, Proc. INFORMATIK: Workshop on Advanced Internet Charging and QoS Technology, Sep. 2001.

[16] OPENIMS, http://www.openimscore.org/

[17] UCT IMS CLIENT, http://coe.uct.ac.za/index.php?pid=193

[18] Dario S. Tonesi et al, “Evaluation of signaling loads in 3GPP networks”, IEEE Wireless Communications, vol. 15, no. 1, pp. 92-100, Feb. 2008.

IX. BIOGRAPHY

Vitalis G. Ozianyi graduated with a Msc. degree in Electrical Eng. at the University of Cape Town in 2006. He is currently undertaking a Phd in Electrical Eng. at the University of Cape Town.