Quality of Service, Authorization and Accounting for Internet Telephony

by Henry Sinnreich, Steve Donovan, Diana Rawlins, and Stephen Thomas[1]

Summary

Commercial grade Internet telephony may act as a trigger for service providers to offer quality of service on the Internet. Payments and settlements may be enablers for QoS based services for real time IP communications. Call signaling and QoS parameters have to be passed between various protocol engines such as SIP, OSP, COPS, RSVP and SBM for IP telephony with QoS. Example call flows using these protocols are shown for QoS assured telephony and QoS enabled telephony. Sample mechanisms for internal accounting, inter service provider payments and settlements are explored for two possible scenarios. First, a PSTN oriented style, where QoS assurance is established during call setup is presented. We refer to these calls as QoS assured calls. Then, we explore a more Internet oriented style where QoS support is established separate from the call setup. We refer to these calls as QoS enabled calls. We discuss the options of both service models and compare complexity, call setup delays and other service features.

1. Introduction

End-to-end QoS on the public Internet does not exist at present to our knowledge, but may be implemented, as soon there are economic incentives to do so and suitable IETF standards are fully developed. Commercial grade IP telephony is arguably the lynchpin for real time IP communications and we believe will gain universal acceptance only when it will have a quality comparable or better than the existing digital circuit switched telephone system. This requires end-to-end quality of service (QoS) across all administrative domains in the path of voice packets. Service providers will however deploy QoS and authorize its usage only if they will be assured that it will be profitable. This shows the dependency of IP telephony on QoS delivery and payments. There seems to be a circular dependency chain in the industry: Carriers need a viable business case to deploy QoS, while commercial IP telephony needs QoS to become viable. Telephony may prove to be one of the enablers for large scale QoS deployment on the Internet. This paper explores two possible scenarios for QoS enhanced IP telephony and is targeted on protocol interactions necessary to establish QoS.

PSTN customers have shown marked preference for flat rate pricing plans, but at the same time put high value on the availability of detailed call history logs. For this reason, besides QoS comparable to the PSTN, commercial IP telephony will also require similar sophisticated tools for usage reporting and payments.

2. IP Telephony with QoS

Service plans for circuit switched telephony may be flat rate or usage based. A common approach to meter usage in present telephone networks is based on usage of circuits by time and distance. Usage is metered and recorded irrespective of the charging policy. We believe Internet telephony could change the underlying business approach, as presented here.

[1] Steve Donovan, Diana Rawlins and Henry Sinnreich are with MCI WorldCom. Stephen Thomas is CTO of TransNexus, LLC.

If there is no pre-arranged business agreement between service providers, a service provider may place a call or accept an incoming QoS enabled call from another service provider only if there is authorization of payment for the call.

IP QoS Billing Metrics: For telephony, QoS may be the main reason to apply premium charges, since the basic IP telephony call can occur without additional charges to the basic Internet service, if both parties know each others IP address, have the appropriate telephony clients and are willing to use best effort service. QoS however, when used for a specific bandwidth over a specified time will use valuable network resources, for which service providers may track usage. Thus, one can imagine premium QoS "minutes" on the Internet to replace, maybe in part, the familiar call minutes for call accounting that we know from the public switched telephone network (PSTN), though this remains to be validated by commercial offerings. Given user preferences for flat rates, QoS minutes may be *recorded only* for various purposes, but not necessarily used as a billing metric. Metrics for the accounting of IP QoS can be time, bandwidth and distance (such as resulting from country codes) or at least geographic region, since these are valid measurements of usage of network resources for input to accounting systems.

In summary, we need not assume that QoS based services will necessarily be charged by duration, since using PSTN style billing by the minute may add additional PSTN style billing costs – an undesirable feature for Internet services. However, even if customers are not charged by the minute, it may be desirable to have records of individual calls and their metrics. For this reason we will explore here the hard problem of providing such records.

Clearing House Services and Protocols: Since the global number of Internet Service Providers (ISPs) is very high, it is not practical to have bilateral service agreements between all ISPs. As a consequence, clearing houses for IP telephony have emerged that provide (1) the trust service and (2) call routing support between ISPs. Since there are quite a number of clearing houses and their affiliates, often with incompatible protocols, it is difficult to insure payments between ISPs that are affiliates of different clearing houses. Thus standard protocols for settlement services have emerged for global IP telephony service across the Internet.

The use of clearinghouses in our general model does not exclude bilateral agreements between service providers, since such agreements will always coexist with clearing houses. As a matter of protocol design, messages for direct authorization of QoS between service providers should use the same protocol design as used with a clearinghouse, presuming the existence of appropriate trust relationships.

3. End-to-End QoS Model

There are probably 10's of very large global transit or "backbone" networks and 10,000's of access networks on the Internet, such as enterprise networks, ISPs, school, academic and government networks. The term "backbone" used in the literature refers mainly to size, since a large network can also serve for access to the Internet. Fig. 1 shows the end-to-end interdomain model with a transit or backbone IP network between two access networks. The model has to accommodate various options for QoS on the IP transit network(s) as discussed in paragraph 5.

4. QoS in Networks

Access networks have historically been the main problem for QoS, but also offer the means to provide QoS assured micro flows, such as for single phone calls. For simplicity, we will examine in this paper QoS based on the Resource Reservation Protocol (RSVP) in the networks, though other QoS mechanisms are also possible. The model for QoS in an access network is shown in Fig. 2.

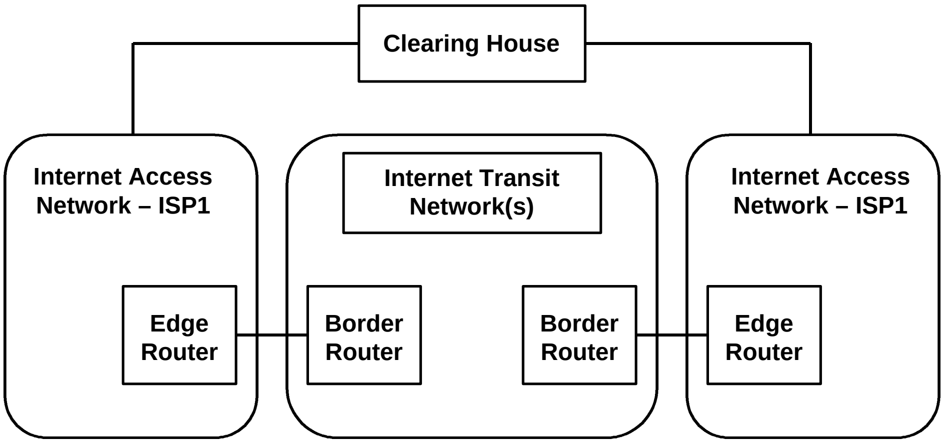

Fig. 1 Internet access and transit networks

The access network has the following network elements:

1. Telephony end-points for real time applications. IP telephones, telephony gateways connected to the PSTN or PBX and various computers (PCs, palmtops) are considered telephony end-points in this model. Telephony end-points have two protocol clients of interest here:
   - SIP client for call setup,
   - RSVP client for QoS signaling.
     In QoS assured calls, SIP call setup depends on successful RSVP completion. In QoS enabled calls, SIP and RSVP completion is decoupled.

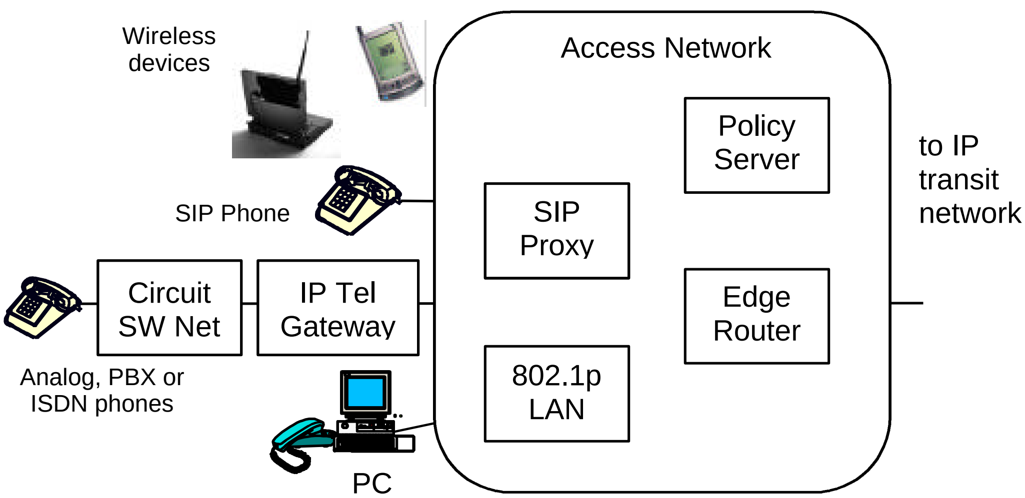

Fig. 2 Access network with QoS

   One or more SIP proxy call setup servers to set up QoS enabled or QoS assured phone calls. SIP proxy servers act as the control element for delivery of enhanced communication services that include telephony with QoS. [2]
2. One or more policy servers to control the access of various hosts to QoS transport, irrespective of the application.

---

[2] IP telephony clients can freely establish calls to any other clients on the Internet or use any service provider of their choice. However, adding the use of the SIP proxy server supports the provisioning of QoS for telephony calls, along with other value added services.

3. Edge and border routers that enforce the policy for admission of QoS requests for internal and for interdomain transport. The edge and border router is also metering the QoS usage, since it is the designated network element for delivering QoS. These routers may also aggregate and deaggregate the RSVP flows.
4. Layer 2 network elements, such as 802.1 style Ethernet switches that deliver QoS at the LAN level subnets.

Network elements communicate using the following protocols:

Session Initiation Protocol (SIP): Signaling between clients and servers to set up sessions, such as for media streams in a telephone call between IP hosts [1], [2], [3].

Open Settlement Protocol (OSP): Requesting and providing authorization from a server, such as from a clearinghouse[3], to fulfill a request to set up a telephone call with QoS. Is also used for pricing and to record usage for payments and settlements between service providers [4], [5].

Common Open Policy Protocol (COPS): Used to install policy in edge routers for particular flows identified by the address of the SIP client making the RSVP request. The Policy server acts as the Policy Decision Point (PDP) and the edge router acts as the Policy Enforcement Point (PEP) [6]. Policy and the protocol choice must be flexible and extensible to meet evolving needs. COPS is well suited for this. It is an object oriented, Header, Length, Content query/response protocol that is designed for Client Type Specific information exchanges. It is easy to understand and straightforward to extend.

We see the SIP-COPS-OSP exchanges as being defined as new classes of Internet objects which might need to be registered and catalogued as have other similar protocol components in the past.

Resource Reservation Protocol (RSVP): Signaling between IP endpoints and intermediate edge routers to provide QoS with certain characteristics such as premium service or guaranteed QoS [7], [8].

Subnet Bandwidth Manager (SBM): Protocol between policy servers or routers and layer 2 Ethernet switches to map RSVP parameters into IEEE 802.1p classes of service [10]. Other types of layer two networks compliant with the Internet Integrated Services Architecture [7], [11] are also possible by using other suitable Layer 2 control protocols.

In the examples provided here it is assumed that the edge router acts as the SBM control element, though it is also possible, depending on the implementation of the access network, to have dedicated SBM servers.

Several delivery service options for networks are possible, since each domain will probably choose its own mechanism to manage its resources.

a. Best effort service only, with enough bandwidth to support adequate QoS on a statistical basis. Adequate is not well defined in this case and voice calls or other applications with guaranteed QoS cannot be supported. Statistically adequate QoS can be provided, but not guaranteed for each individual micro flow, such as for a commercial grade phone call.

[3] The Internet standards community is pursuing work on the Telephony Routing Information Protocol (TRIP). TRIP is a policy driven interdomain protocol for advertising the reachability of telephony destinations between location servers, and for advertising attributes of the routes to those destinations. TRIP may replace some of the functions of presently used clearinghouses [14].

b. Premium service, possibly based on Differentiated Services (DS) [9]. Service level agreements specify the amount of premium traffic that the access network can exchange with the transit network. Edge and border routers on the network classify and police the DS flows from the access networks to assure they stay within the limits of the Service Level Specification (SLS).

c. Premium services with per flow QoS guarantees assured QoS for the individual micro-flows. Such service may be suitable for QoS assured telephony and other applications, such as demanding real time communications.

For both types of premium services (b) statistical QoS and (c) per flow QoS, the edge and border routers for the network may aggregate the flows into DS classes. [11], [12], [13]. Phone calls with guaranteed QoS can be provided in this manner.

In the following, we will discuss only the protocols necessary to exchange end-to-end call setup, QoS and payment information and leave aside the particulars of the respective SLS between networks .

5. Passing of Parameters Between Protocols

We will address two scenarios:

PSTN style fully assured QoS calls: Remote ringing is heard only after there is QoS established in both directions. A busy line tone is returned when QoS cannot be provided.

QoS enabled calls: Telephony call setup and QoS setup proceed in parallel and the conversation can start before QoS is fully established in both directions. If QoS cannot be provided, the call can nevertheless continue at the users choice.

5.1 Fully QoS Assured SIP Call Setup

Complete duplication of PSTN QoS on the Internet is a difficult problem. We will illustrate here the option to provide a fully QoS assured telephone call between two domains of access service providers across a common IP transit network [15]. Fig. 3 shows the call flows for a successful call setup. The call flows include messages from SIP, OSP, COPS and RSVP. SBM messages are not shown, since they are not part of interdomain signaling. The fully assured QoS call setup model is deterministic in the sense that the edge router acts as a gateway for every single phone call, and upon admission of the respective media flow, there is an individual guarantee for the RSVP parameters for the call.

A somewhat similar call flow exists for tearing down the call but is not shown here. To tear down a call one of the SIP clients initiates a BYE message that ripples to the other protocol engines so as to de-install policy (COPS), release the network resources (RSVP and SBM) and report usage (OSP).

The detailed call flows and message exchange for the example in Fig. 3 have been published in [15] and need not be repeated here.

Final acknowledgement messages (ACKs) to complete the call cannot happen until successful RSVP based QoS has been established in both directions. In case of failure of RSVP in either direction or in one of the two access networks, the call will not be completed and one of the possible failure scenarios will result in a SIP 6xx message.

SIP Global Failure messages of the 6xx type for failure to set up QoS may need to be defined. The parameter passing between the RSVP and SBM protocol entities is not detailed here any further, since SBM is of local nature are not necessarily of interest outside the respective domains. A point of interest are failure scenarios for SBM, such as in congested LANs. If a LAN is congested, the edge router will refuse the RSVP request from the telephony endpoint and not forward it further to the other domain.

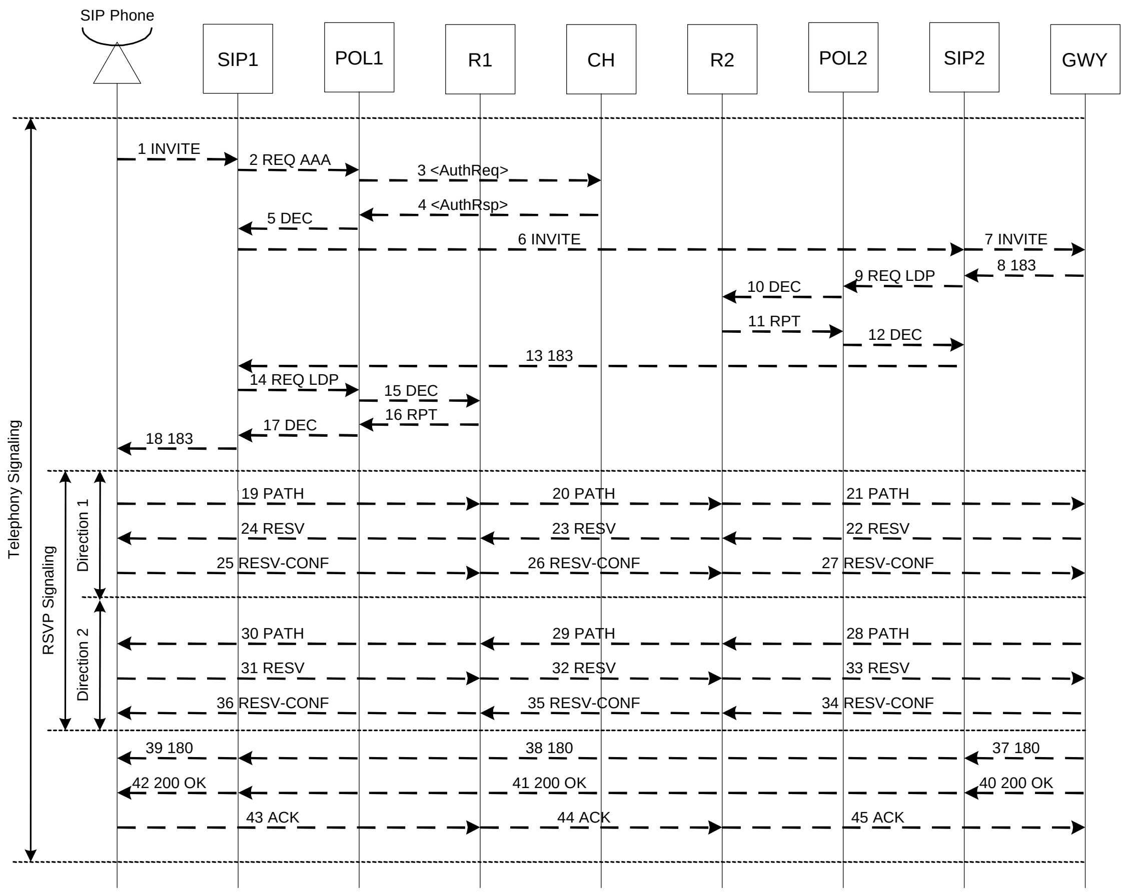

Fig. 3 Fully QoS assured interdomain SIP call setup

A brief examination of call flows for fully assured QoS between domains across the public Internet shows the following drawbacks:

- Complexity, comparable to the PSTN,
- Potential for noticeable call setup delay due to the large number of messages traversing IP networks. It should be noted that QoS setup in the access networks and in both directions can proceed at the same time, though the call flow diagram may convey the message of a rigid sequence. Post dial delay is however determined by completion of QoS setup.
- It is not possible to set up calls to remote ISPs that have not implemented QoS in a generic compatible way. There is no flexibility, since the communicating ISPs have to respect the same SIP and RSVP call flow sequences.

A sometimes-desirable feature is traffic congestion in any of the participating networks will result in call setup failure and denial of service, similar to PSTN.

In spite of complexity and possibly noticeable call setup delay, we do not dismiss the premium service option of providing PSTN-style fully QoS assured call setup.

5.2 Decoupled Application and Transport Signaling

An alternative to the fully deterministic end-to-end QoS assured call setup, as shown in figure 3 is the case of Internet style loosely coupled SIP, OSP and RSVP processes in the domains traversed by the call. The statistical model for QoS delivery and admission control means that the edge router is ignorant of packet flows for individual calls or in general of the nature of the application to which the flow belongs. Rather, their respective servers control various applications. The SIP proxy server of the access network domain exercises telephony policy control but does not trigger QoS policy decisions like in the QoS assured case. This frees the access router from considerable admission policy processing, since it will only keep tab if the aggregate QoS flows are within the limits of the service level specification.

Whenever this limit is exceeded, a ResvErr (reservation error) message is returned to the requesting host, irrespective of the application.

Figure 4 shows the call flows for a successful SIP call setup. The application layer signaling and transport layer signaling processes are decoupled.

We have assumed here a simple scenario where the edge router has policy installed only for the maximum value of aggregate RSVP flows. All RSVP requests that fall below this maximum threshold will be honored, requests falling above the threshold will be refused and the RSVP client will generate an appropriate alert.

The advantages are:

- Clean separation between the application of SIP call setup and transport QoS using COPS and RSVP. QoS setup is kept application independent.
- Simplicity due to decoupling of telephony and QoS,
- The edge router does not need to support per call dynamic policy establishment,
- Smaller call setup delay due to fewer messages and completely decoupled processing at the application layer (SIP, OSP) and transport layer (COPS, RSVP),

ISPs with QoS can complete calls to remote ISPs, irrespective of QoS service at the other end. If the remote ISP has no QoS capability, best effort service is still possible.

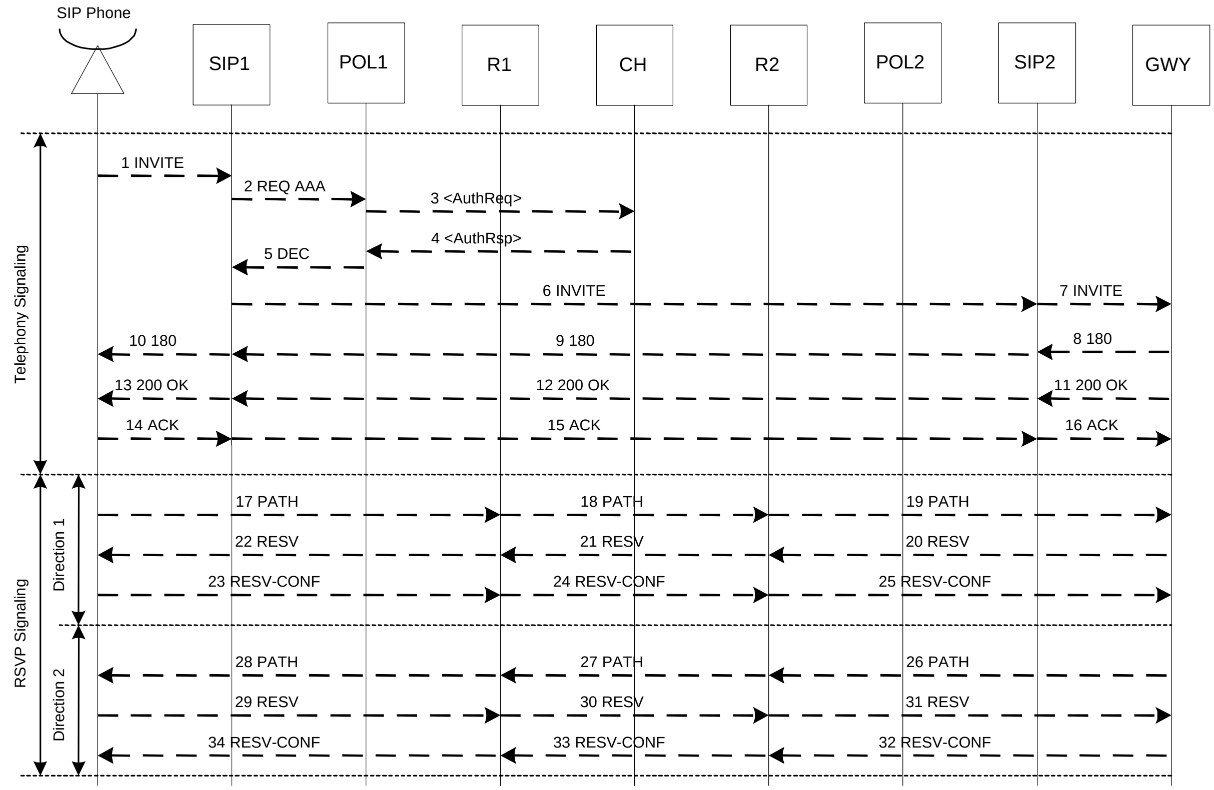

Fig. 4 Decoupled SIP, OSP and COPS, RSVP signaling.

The disadvantages of decoupling call setup and QoS setup are:

- Fully assured QoS calls are no longer possible,
- QoS setup may be complete only after the call is already in progress, leaving the first seconds vulnerable to quality impairments.

Calls with decoupled QoS setup can be said to be *"QoS enabled",* versus the previous example of *"QoS assured"* calls.

Several QoS failure scenarios are possible in both models:

- The transit or remote access networks do not provide QoS,
- QoS setup failure in one of the networks

.
It is the responsibility of the caller's RSVP client to report RSVP failure and to provide the reason code why and where the failure has occurred. No special SIP 6xx Global Failure messages are necessary since the call may proceed without QoS in one or both directions.

Traffic congestion will not result in call setup failure, but give the caller the option to continue with the call. IP telephony client implementations can provide the caller with the appropriate QoS failure notifications and depending on the type of service give the caller the option to proceed with the call at lower cost or even for free, such as" Sorry, we cannot assure quality for this call. Would you like to proceed at half price (or with a free call)?"

Depending on the business purpose, both QoS assured and QoS enabled IP telephony will be valid services. The drawbacks shown for not completely assuring QOS will be balanced by the larger functionality of IP communications. This is an analogy to the acceptance of lower QoS for wireless phones, where other useful functions make up for occasional lower quality compared to wireline calls.

6. Accounting for QoS IP Services

QoS for IP telephony can be accounted for in the same way as any other QoS enhanced premium Internet service. An accounting architecture for generic, application independent pricing and billing models has been reported in [16]. Given the experience of the high cost of PSTN accounting systems, detailed charging for usage may not be optimal for Internet services. Using a common set of premium Internet rates for any application requiring QoS, including telephony, will enable lowering the overall costs to network providers to deliver such services.

Besides QoS many other SIP based chargeable services are possible, such as:

- Interaction of PSTN and Internet services [17],
- Advanced Intelligent Network (AIN) style services as defined in CS-1 and CS-2 [18],
- Caller and called user preferences [19],
- Gateway services to connect to/from the PSTN [20],
- IP Centrex type services and desktop services [21],
- Presence and instant voice/text chat [22],
- Local Number Portability [23],
- Global device and network application level mobility [24]. etc.

Flat rate charges for such services are independent of the charges for QoS and may well be in addition to charging for QoS. We believe however, advanced services will be enhanced by QoS support.

7. Future Work

QoS assured IP telephony may require the definition of the SIP type 6xx Global Failure messages to convey the reason for failing to set up calls due to missing QoS.

QoS enabled IP telephony requires SIP extensions to carry a token from the clearinghouse to the remote SIP server. See in Fig. 4 the message number 6 INVITE that has to carry the authorization token.

Interdomain payments and accounting across the Internet will also require the definition of standard QoS "rate" metrics in terms of IETF Integrated Services and Differentiated Services interworking.

The authors are also aware of other future work items related to QoS policy architecture that go beyond the space for this article.

8. Acknowledgements

The authors would like to thank Teresa Hastings, John Gallant, Patrice Carroll and Vinton Cerf from MCI WorldCom for helpful comments to this paper.

9. Disclaimer

The positions and views expressed in this paper are the opinions of the authors and do not necessarily reflect the views of MCI WorldCom or TransNexus.

10. References

[1] RFC 2543: SIP: Session Initiation Protocol, IETF, March 1999. ftp://ftp.isi.edu/in-notes/rfc2543.txt

[2] J. Rosenberg, H. Schulzrinne, S. Donovan: Establishing QoS and Security Preconditions for SDP Sessions. Internet-Draft, work in progress, January 1999. http://ietf.org/internet-drafts/draft-ietf-mmusic-sdp-qos-00.txt

[3] S. Donovan, H. Schulzrinne, J. Rosenberg, M. Cannon, A. Roach: SIP 183 Session Progress Message. Internet-Draft, work in progress, October 1999.

[4] European Telecommunications Standards Institute (ETSI): Telecommunications and Internet Protocol Harmonization Over Networks (TIPHON); Inter-domain pricing, authorization, and usage exchange. [Technical Specification TS 101 321]. Version 1.4.2. December 1998 [Available at www.etsi.org].

[5] S. Thoma, R. Brennan, B. Anton, D. Oran: The application/osp-token MIME type. Internet-Draft, work in progress, October 1999. http://www.ietf.org/internet-drafts/draft-thomas-mime-osp-token-00.txt

[6] J. Boyle, R. Cohen, D. Durham, Shai Herzog, R. Rajan, A. Sastry: The COPS (Common Open Policy Service) Protocol. Work in progress, IETF, November 1999.

[7] RFC 2210: The Use of RSVP with IETF Integrated Services by J. Wroclawski, September 1997.

[8] RFC 2205: Resource ReSerVation Protocol (RSVP)Version 1 Functional Specification. IETF, September 1997. ftp://ftp.isi.edu/in-notes/rfc2205.txt

[9] SBM (Subnet Bandwidth Manager): A Protocol for RSVP-based Admission Control over IEEE 802-style networks by R. Yavatkar, D. Hoffman, Y. Bernet, F. Baker and M. Speer. Internet-Draft, work in progress, October 1999. http://ietf.org/internet-drafts/draft-ietf-issll-is802-sbm-10.txt

[10] RFC 2430: A Provider Architecture for Differentiated Services and Traffic Engineering (PASTE). October 1998. ftp://ftp.isi.edu/in-notes/rfc2430.txt

[11] A Framework For Integrated Services Operation Over Diffserv Networks by Y. Bernet, R. Yavatkar, P. Ford, F. Baker, L. Zhang, M. Speer, R. Braden, B. Davie, J. Wroclawski, and E. Felstaine. Internet-Draft, work in progress, September 1999. http://ietf.org/internet-drafts/draft-ietf-issll-diffserv-rsvp-03.txt

[12] Fred Baker, Bruce Davie, Francois Le Faucheur, Carol Iturralde : Aggregation of RSVP for Ipv4 and Opv6 Reservations. Internet-Draft, September 1999. < draft-ietf-issll-rsvp-aggr-00.txt>

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

[14] J. Rosenberg, H. Salama, M. Squire: Telephony Routing Information Protocol (TRIP). IETF work in progress, October 1999.

[15] H. Sinnreich, S. Donovan, D. Rawlins, S. Thomas: Interdomain IP Communications with QoS, Authorization and Usage Reporting. Work in progress, IETF, October 1999. http://www.ietf.org/internet-drafts/draft-sinnreich-interdomain-sip-qos-osp-00.txt

[16] F. Hartanto and G. Carle: Policy-based Billing Architecture for Internet Differentiated Services, Proceedings of IFIP Fifth International Conference on Broadband Communications (BC'99), Hong Kong, 10-12 November 1999. http://www.fokus.gmd.de/research/cc/glone/research/genca/

[17] RFC 2458: Toward the PSTN/Internet Inter-Networking. November 1998. ftp://ftp.isi.edu/in-notes/rfc2458.txt

[18] Jonathan Lennox, Henning Schulzrinne and Thomas F. La Porta: Implementing Intelligent Network Services with the Session Initiation Protocol. Columbia University Computer Science Technical Report CUCS-002-99, January 1999.

[19] H. Schulzrinne and J. Rosenberg: SIP Caller Preferences and Callee Capabilities. IETF work in progress, April 1999. http://www.ietf.org/internet-drafts/draft-ietf-sip-callerprefs-00.txt

[20] A. Johnston, S. Donovan, R. Sparks, C. Cunningham, K. Summers: SIP Telephony Call Flow Examples. Work in progress, IETF, October 1999. http://www.ietf.org/internet-drafts/draft-johnston-sip-call-flows-00.txt

[21] A. Johnston, S. Donovan, R. Sparks, C. Cunningham, K. Summers: SIP Telephony Service Examples With Call Flows. Work in progress, IETF, October 1999. http://www.ietf.org/internet-drafts/draft-sparks-sip-service-examples-00.txt

[22] M. Day, J. Rosenberg and H. Sugano: A Model for Presence and Instant Messaging. IETF work in progress, August 1999. http://www.ietf.org/internet-drafts/draft-ietf-impp-model-03.txt

[23] I. Dalgic, M. Borella, R. Dean, J. Grabiec, J. Mahler, G. Schuster, and I.:True Number Portability and Advanced Call Screening in a SIP-Based IP Telephony System. IEEE Communications Magazine, Vol. 37, No. 7, July 1999.

[24] Elin Wedlund and Henning Schulzrinne: Mobility Support using SIP. Second ACM/IEEE International Conference on Wireless and Mobile Multimedia (WoWMoM'99), Seattle, Washington, August, 1999.