Towards Scalable End-to-End QoS Provision for VoIP applications

R. El-Haddadeh, G. A. Taylor and S. J. Watts

Department of Electronic and Computer Engineering, Brunel University
Uxbridge, Middlesex, UB8 3PH, UK
Ramzi.El-Haddadeh@brunel.ac.uk
Tel: 44 1895 274000 Fax: 44 1895 258728

Keywords: IntServ, DiffServ, QoS and VoIP.

Abstract

The growth of the Internet and the development of its new applications have increased the demand for providing a certain level of resource assurance and service support. The concept of ensuring Quality of Service (QoS) has been introduced in order to provide the support and assurance for these services. Different QoS mechanisms have been developed and introduced such as integrated services (IntServ) and differentiated services (DiffServ) to provide different levels of QoS provision. However, IntServ can suffer from scalability issues that make it infeasible for large-scale network implementations. On the other hand, the aggregated-based per-flow technique of DiffServ does not provide such an end-to-end QoS guarantee. Recently, the IETF have proposed a new QoS architecture that implements IntServ over DiffServ in order to provide an end-to-end QoS for scalable networks. Hence, it became possible to provide and support a certain level of QoS for some delay sensitive and bandwidth-demanding applications such as Voice over Internet Protocol (VoIP). With regard to VoIP applications, delay, jitter and packet loss are crucial issues that have to be taken into consideration for any VoIP system design and such parameters need a distinct level of QoS support.

1 Introduction

The global evolution of the Internet and the wide spread growth of networks have made the Internet part of our everyday life. For these reasons, the demand on the various Internet applications has increased. This increase in demand has also produced a number of new applications that did not exist before. Internet users can now access a variety of different multimedia applications such as video conferencing, VoIP, Video on Demand (VoD)...etc.

However, the current Internet infrastructure with its best-effort service provision is unlikely to be able to support such high bandwidth, delay sensitive applications. This is due to the fact that networks may not be able to support such applications using existing network architecture and so, new mechanisms have to be developed in order to adequately support such applications. In addition, the current Internet infrastructure faces a number of challenges and issues that need to be considered in order to enhance network performance so that service providers are able to provide and support the services required by the users [1].

The unpredictable growth of the internet and its usage over the past few years has brought up the issue of scalability. The number of users and hosts has increased tremendously throughout the past few years and service providers have to satisfy the different requirements for this increase. Hence, different QoS schemes need to be implemented that are scalable in order to maintain such growth.

Also, the variety of applications that the Internet can now support has created the need for optimization of such seamless applications and integration within one heterogeneous network. Such applications vary from simple email exchange (a small amount of data that has low profile with regard to QoS requirements) up to streaming video (a large amount of data that has very high profile with respect to QoS requirements) and therefore require service guarantees as they have different traffic characteristics [1, 2].

Moreover, with such increases in network traffic, scale and applications it is important that all these issues have to be resolved between the different service providers. The goal of maintaining some sort of understanding between the service providers is becoming more and more crucial. The Internet is growing and its service providers are growing as well. They need to be able to support each others applications as each one of them has its own network architecture. Therefore, it becomes quite important to have some sort of open and agreed framework for maintaining and guaranteeing the services provided, i.e. the QoS that is supported. Eventually, such understanding between service providers can help provide a controlling mechanism for the internet and with this better administration can be obtained between the different service providers [3].

However, service providers tend to overcome all these limitations by over provisioning network resources. Service providers are also attempting to avoid bottlenecks in order to guarantee the service that they provide for users over existing

best effort IP networks. Although this type of service support can be provided using Dense Wave Division Multiplexing (DWDM) as it is able to provide a very high bandwidth, the usage of the bandwidth is still increasing. Furthermore, this solution can be considered quite costly and non profitable for them as they have to implement and support more devices throughout their backbone. Also, they will not be able to differentiate between the different traffic of the different users. [4].

2. Internet QoS Mechanisms

2.1 Integrated Services Packet Networks

IntServ is a QoS mechanism proposed by the IETF [5] that is characterized by resource reservation. The assumption is that resources (e.g. bandwidth) must be explicitly managed in order to meet application requirements. In addition to the traditional best-effort service, the IntServ architecture proposes two more classes of service; Guaranteed Service (GS) and Controlled-Load Service (CLS).

In order for the applications to communicate their QoS requirements to nodes along the transit path, a signalling mechanism is required. The IETF IntServ Working Group recommends Resource Reservation Protocol (RSVP) [6] as the signalling protocol to reserve resources. The main function of RSVP is to provide QoS requests on behalf of the application traffic to all routers along the transit path and to maintain the state information in the routers for each data flow [6].

2.2 RSVP

As explained above RSVP is a signalling protocol that applications can use to request resources from the network. The network responds by explicitly admitting or rejecting RSVP requests. Certain applications that have quantifiable resource requirements express these requirements using IntServ parameters as defined in the appropriate IntServ service specification. As noted above, RSVP and IntServ are separable. RSVP is a signalling protocol that can carry IntServ information. IntServ defines the models for expressing service types, for quantifying resource requirements and for determining the availability of the requested resources at relevant network elements (admission control) [6].

The prevailing model of RSVP usage is based on a combined RSVP/IntServ architecture. In this model, RSVP signals per-flow resource requirements to network elements using IntServ parameters. These network elements apply IntServ admission control to signalled requests. In addition, traffic control mechanisms on the network element are configured to ensure that each admitted flow receives the service requested in strict isolation from other traffic.

However, the problem with the IntServ/RSVP architecture is of scalability. The model does not scale well in the Internet core primarily because [7]:

1. Huge storage and processing overheads are placed on the routers since the amount of state information in the routers increases proportionally with the number of flows,
2. The requirements on routers are very high as each router must implement RSVP, Admission Control, MF Classification and packet Scheduling.

2.3 Differentiated Services Packet Networks

Due to the difficulties encountered in implementing and deploying the IntServ/RSVP architecture, another QoS mechanism known as DiffServ has been proposed [8]. The goal of DiffServ is to give scalable service discrimination without the need for per-flow state and signaling at every hop or router. The DiffServ architecture achieves its scaling properties by defining a small number of different packet forwarding treatments known as Per-Hop Behaviors (PHB) [9].

DiffServ makes use of the Type of Service (TOS) flag in the IPv4 header, which corresponds to the Traffic Class flag in the IPv6 header. Each packet receives a particular forwarding treatment based on the setting of the IP TOS flag. The packet is treated the same way as others marked the same flag. There is no per-flow state required inside the network; a core device only considers markings and not flows. Per-flow state is kept at the network edge and flows are aggregated based on desired behavior. The Bandwidth Broker (BB) implements dynamic allocation of resources and is also responsible for making sure that network resources, both within the DiffServ domain and on links connecting adjacent domains, are properly provisioned and not over subscribed.

By marking the flags of packets differently and handling packets based on the flags, several differentiated service classes can be created. The IETF DiffServ group has defined two classes of supporting applications; the premium service and the assured service models in addition to the existing best effort service.

3. IntServ over DiffServ Networks

Recent QoS developments have combined both integrated and differentiated services together into a new architecture for an end-to-end service provision [10]. The strength of IntServ is its ability to provide a per-flow QoS guarantee, while it lacks the issue of scalability at its core routers. On the other hand DiffServ enables scalability across large networks but may not be able to support a per-flow QoS guarantee. So, with the new combined architecture it is possible to maintain a scalable QoS service guarantee at the network core using DiffServ, while implementing IntServ at the network edge in order to support a per-flow QoS guarantee [10].
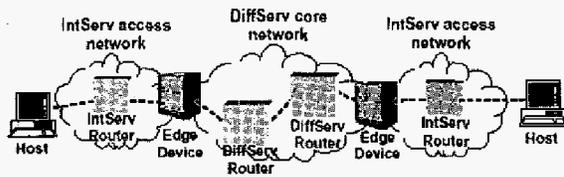
Fig. 1 IntServ implementation over DiffServ

With this new architecture an end-to-end quantitative QoS can be provided. Here, the DiffServ network is not required to participate in the RSVP signaling for the resource allocation and the admission control. In addition, within the DiffServ capable network domain a specific PHB is applied and the total amount of aggregate traffic with the PHB will be limited by policing at the edge of the domain. From the IntServ perspective, the DiffServ network can be viewed as a virtual link that connects the IntServ capable networks with each other.

4 Voice over Internet Protocol (VoIP):

Recently, the integration of voice and data together onto a single network became a priority for many network operators. In VoIP applications a source voice signal is first packetized and then transmitted over an IP network and then when they arrive at their destination they are depacketized in order to retain the voice data. However, due to the nature of voice, VoIP requires timely packet delivery with low delay, jitter and packet loss values. Hence, to achieve that IP networks must be enhanced with certain mechanisms that ensure the required quality of service for such applications [11].

4.1 Variable and Constant Bit Rates (VBR & CBR)

In order to provide and allocate the resources for the requested QoS, a qualitative description is required for the generated traffic pattern. This description will help to allocate the right resources for each request. Hence, based upon their bit rate, applications can be classified into two main categories; variable bit rate (VBR) and constant bit rate (CBR) [12].

CBR describes traffic that is transmitted over a constant rate. For example, voice coded PCM represent such class of traffic generated at 64 kbps. And for VBR, the generated traffic rate is not constant and is always varying as in MPEG video applications.

Due to the difference in these two traffic sources the QoS requirements vary. For admission control, it will be adequate for CBR sources to obtain bandwidth reservation at the peak rate. However, for VBR sources this sort of bandwidth reservation at the peak rate will be wasted, since VBR sources do not send their traffic at such a rate all the time. So, the size of the traffic will vary between the peak and the average rates and therefore a mechanism will be needed to relate a given reservation to the QoS parameters.

Moreover, regardless what policing mechanism applied; simple, token or dual leaky buckets. VBR and CBR traffic sources have a number of issues with regard obtaining the required QoS. This is due to the fact that policing is applied in correspondence to the traffic parameters i.e. peak, average and burst traffic. So for CBR, only the peak rate will be policed, but for VBR the peak, average rates and the burst size have to be policed, which will add more complexity to the system.

For voice traffic, a CBR source can be described by the packet length and fixed inter-packet time, but for a VBR source it is much more complicated. VBR traffic is generated during the voice communication due to the implementation of silence suppression algorithms' in the terminal equipment [13].

5 Network Design and Implementation

In this section a simple network with two end routers is implemented using OPNET. The purpose of this modelling is to look at the individual performance of each QoS mechanism i.e. IntServ and DiffServ. The design of the network is based upon implementing voice traffic only. Throughout the modelling, different QoS queuing techniques are implemented such as WFQ, PQ and CBQ. In addition, other congestion control mechanisms are studied such as RED and ECN [12].
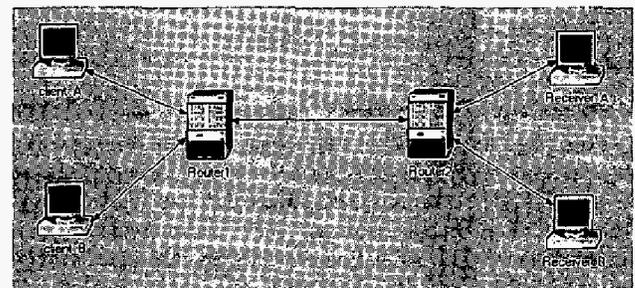

Fig. 2 Simple Network Design

6 Network Performance Analysis

As mentioned before the purpose of this modelling is to study the individual performance of IntServ and DiffServ separately. However, it was expected that the performance of the two different QoS mechanisms with the use of the different queuing and congestion schemes would be similar since the network was not simulated under heavy traffic conditions. The optimal performance of IntServ and DiffServ is not obtained due to the small size of the network. This was clear with the zero traffic loss, constant end- to-end delay value which was less than 60 msec, and with the zero jitter value.
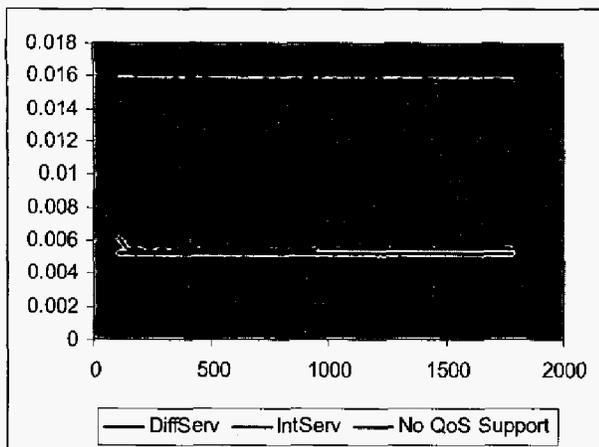
Fig. 3 End-to-End Delay performance

7 Conclusion

The scope of this paper is to provide insight on the implementation of IntServ over Diffserv networks for the provision of an end-to-end scalable QoS for VoIP applications. Future research will examine different traffic characteristics for VoIP over such architecture and will focus on CBR and VBR. However, in this work a simple implementation of a VoIP system has been modelled using OPNET and results show the similarity in performance between IntServ and DiffServ. This basic modelling approach is a key start to further advanced modelling for a VoIP system, which is the goal of this research project. The research examination will take two perspectives; both objective and subjective. The objective perspective will be through the modelling of such architecture using commercial simulation tools. The subjective perspective will be through testing and real-time analysis of an existing campus network.

References

[1] Z. Wang, *Internet Quality of Service: Architectures and Mechanisms: Architectures and Mechanisms*: Morgan Kaufmann, 2001.

[2] P. Ferguson and G. Huston, *Quality of Service: Delivering QoS on the Internet and in Corporate Networks*: John Wiley & Sons Inc, 1998.

[3] "Quality Management and Quality Assurance-Vocabulary," ISO 8402, 1994.

[4] J. Gozdecki, A. Jajszczyk, and R. Stankiewicz, "Quality of service terminology in IP networks," *IEEE Communications Magazine*, vol. 41, pp. 153-159, Mar 2003.

[5] D. Chalmers and M. Sloman, "A Survey of Quality of Service in Mobile Computing Environments," *IEEE Communication Suerveys*, 2nd Quarter 1999.

[6] J. Wroclawski, "Specification of the Controlled-Load Network Element Service," RFC 2211, September 1997.

[7] R. Braden, L. Zhang, S. Berson, S. Herzog, and S. Jamin, "Resource Reservation Protocol (RSVP)," RFC 2205, Version 1, Functional Specification, September 1997.

[8] X. Xiao, A. Hannan, and B. Bailey, "Traffic Engineering with MPLS in the Internet," *IEEE Network Magazine*, vol. 14, Mar/Apr 2000.

[9] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, "An Architecture for Differentiated Services," RFC 2475, December 1998.

[10] Y. Bernet, P. Ford, R. Yavatkar, F. Baker, L. Zhang, M. Speer, R. Braden, B. Davie, J. Wroclawski, and E. Felstaine, "A Framework for Integrated Services Operation over Diffserv Networks," RFC2998, November 2000.

[11] R. B. J. Reynolds and A. W. Rix, "Quality VoIP - an engineering challenge," *BT Technology Journal*, vol. 19, April 2001.

[12] S. Jha and M. Hassan, *Engineering Internet QoS*: Artech House Telecommunications Library, 2002.

[13] V. Matic, A. Bazant, and M. Kos, "Voice traffic performance measurement in packet networks," presented at ITI 2002 - 24th International Conference INFORMATION TECHNOLOGY INTERFACES, 24-27 June 2002.