Rate Controlled Adaptive Resource Coordination Framework for Wireless Aware Multimedia Applications

N. Kumar* **and Y. D. S. Arya****,

*Dept. of CSE, Invertis Inst. of Engg. & Tech., Bareilly, UP, INDIA,
**Dept. of CSE, Indian Institute of Technology (IIT), Kanpur, UP, INDIA

Summary
Modern networks provide QoS model to go beyond best-effort services, but current QoS models are oriented towards low level network parameters (e.g. Bandwidth, Latency, and Jitter). Application developers are interested in quality models that are meaningful to the end-user and therefore struggle to bridge the gap between networks and application QoS models. Network-aware applications are promising new concepts, in which the applications are aware of network conditions and thus can adapt to the varying environment to achieve acceptable and predictable performance. In this paper we have proposed a Relaxed Layer Architecture and a protocol framework to satisfy users' specific QoS need, which implements QoS based adaptive resource management through control algorithms. The framework facilitates QoS provisioning through a semantically rich interface that enables the end users to get the services according their requirements and they are willing to make under potential resource shortages. Simulation results show that the resources are allocated to most of the accepted requests within a short span through negotiation.
Keywords:
QoS, resource management, renegotiation, SLA, Relaxed layer.

1. Introduction

Current computer networks and the Internet are becoming more and more heterogeneous. The applications operate across different types of users, wired or wireless networks, with the availability of different type of resources [2]. Mobile multimedia application is a typical example of such applications that run in a heterogeneous environment. IP-based network technology has shown tremendous growth in recent years, and is becoming the backbone of the next generation data network. QoS issues have been widely studied for conventional IP networks, and almost all-necessary elements now exist for providing QoS support in conventional IP networks. However, the conventional IP network architecture was originally designed for fixed nodes connected by wired links. Mobile wireless networks have a few fundamentally different characteristics from conventional wired networks (See [1] for a tutorial on wireless errors and their models). These include:

- Low bandwidth wireless links.
- High link error rate of wireless links.
- Mobility of end hosts resulting in hand-offs between access points.

The objective of network-awareness is to allow an application to be sensitive to change in the network environment with the goal of maximizing user-perceived quality. Basically there are two ways to provide QoS for applications in wireless environment. One is by resource reservation from networks point of view, the other is by behavior adaption from the application's point of view. Some network architectures may not support reservations at all or may support them only to a limited degree, and although future versions of popular protocol suites may support reservations, not all sites will run the most recent software. Furthermore, as network providers attempt to develop usage-based charging schemes, there will be financial incentives to restrain applications from uncontrolled use of network resources. Today's networks have really two aspects that make adaptivity unattractive: there is almost no usage-based charging, and what is worse; the most aggressive applications are often rewarded with the largest share of the bandwidth pie [3].

In a reservation-based approach an application must address the two issues one is (i) how to find out what and how much to reserve (e.g., given some limit on the costs) and other is (ii) how to adjust to meet the confirmed reservation, which may be less than the application has asked for. These involve continuous negotiation depending on resource availability and constant monitoring by the service provider and the client leads to QoS tradeoff and resource tradeoff. So there is a need for network-aware architecture that can understand heterogeneous environment and can allocate the resources to the users with negotiation capability, if sufficient resources are not available. So we present a Relaxed Layer Architecture for QoS trade-off with network aware capability. In this architecture the apportioned system resources are optimized when quality of service requirements are met such that the service layer requirements for a particular application is satisfied and the net utility accrues to the user.

Multimedia applications and their users can differ enormously in their requirements for service quality and

Manuscript received December 5, 2010
Manuscript revised December 20, 2010

the resources available to them at the time of application use. Therefore, there is an increasing need for customizable services that can be tailored for the end users' specific requirements at the minimum cost with optimum utilization of the system resources. However, these QoS requirements can be objective in some aspects and subjective in others. Moreover, because of the manifold and subjective nature of user quality demands, it is very hard to determine the optimal combination of resources that provide the quality needed to fulfill the stated SLA [4]. The SLA defines the QoS assured service to the client in a distributed multimedia system. To give customizable services, resource allocation has to be done dynamically on the basis of the QoS need, resource availability, and adopt the negotiation. In this paper, we presents a framework for quality of service (QoS) centric service level agreements (SLA) between service providers, who also own servers and network for their clients. The SLA may specify certain SLA parameters but the system may be able to accommodating some but not all. So it is important for the service provider to provide customized SLA and dynamic negotiation/renegotiation. Relaxed layer architecture can negotiate resources to satisfy the users' QoS requirements.

The rest of the paper is organized as follows. In Section II we present the wireless awareness architecture for QoS awareness with experimental flowchart diagram. In Section III, the proposed Relaxed Layer Architecture and E2E tradeoff protocol is developed, with class diagram designed in UML is also present. End system and APN's interaction with sequence diagram as described in section-IV. The simulation results are shown in section V. Finally, conclusions are drawn in Section VI.

2. Wireless Awareness Archiecture

There are concepts as network-aware [5], system-aware and corresponding researches. The model of a network-aware application emphasizes that the performance problems in the network are the causes of the degradation of application performance while the system-aware applications focus on the system that is bottlenecked by the components other than networks such as the local bus, disk, CPU or memory. The above system awareness is called end-host awareness in this paper, because the system awareness here is more general, which refers to both the end-host awareness and the network awareness. In fact two approaches to make the applications wireless-aware, i.e. the end-host approach by accessing the device driver of the network interface card (NIC) and the networks' approach by instrumenting the round-trip time (RTT) of the communication entities.

The block diagram of the wireless-awareness architecture is given in Figure 1. The component to fulfil the wireless awareness has been designed as an agent that can be used by multimedia applications directly or to be embedded into a middleware layer. The advantage of using the middleware is its programmability, adaptability, and extensibility. The wireless-awareness agent can communicate with the other QoS-middleware components and QoS-related agent such as SNMP (Simple Network Management Protocol) agent via the inference engine so that the awareness becomes more accurate and informative. This open architecture makes the wireless awareness extensible to system awareness, which is defined as the ability of software to deal with the variations of the characteristics of the end-systems. The communication networks achieve the best service quality specified for applications and the maximum resource utilization constrained by the system features.

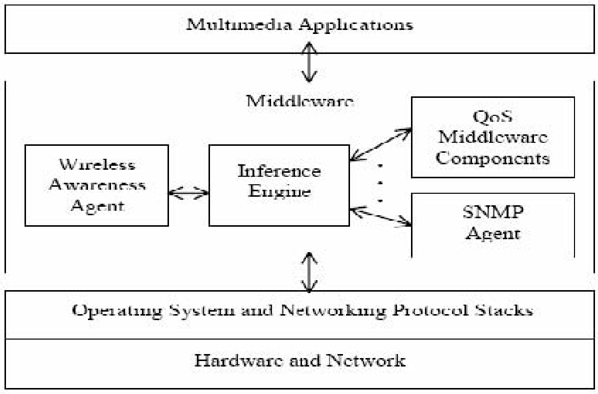


Figure 1: Architecture of Wireless Awareness

According to the architecture of wireless-awareness, as shown in figure 1, the wireless awareness agent is implemented in the middleware layer. The UDP packets are used to measure RTT values. From the experiments it is established the RTT value of the communication sessions in the wired case has different statistical patterns from that in the wireless case. Based on the results, the existence of the wireless links in the communication path can be detected in the wired and wireless LAN environment by applying suitable thresholds to the mean value and variance of RTT values to differentiate the wired and wireless case. The flowchart of the implementation of wireless awareness is shown in Figure 2.

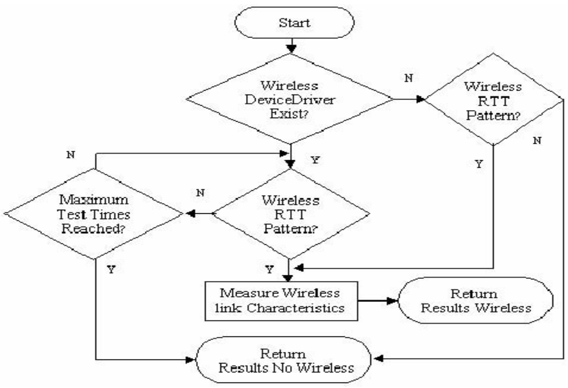


Figure 2. Implementation Flowchart of Wireless Awareness

Noted that even after the wireless device driver is detected, there is a possibility that the communication applications transmit packet only via wired links because the local host is multi-homed, i.e. the computer connected to the network via multiple NICs. Moreover, sometimes there are conflicts occur between the awareness result of accessing device driver and that of measuring RTT patterns because of the networks' transient dynamic behaviour. Therefore, several RTT tests are needed when there are conflicts.

A generic framework for developing network aware applications is proposed by Bolliger in [5]. The core of this framework is a feedback closed-loop that controls adjustment of an application to network properties as shown in Figure 3. This feedback loop is designed as an adaptation layer sitting between the application layer and the lower layers in a common network model (e.g. TCP/IP). Monitor and react phase in this loop obtains information about the network status such as available bandwidth, decides which object needs to be adapted and how to adapt, and determines a QoS goal for the object. The prepare phase then applies a transformation strategy to the objects to realize the QoS goal as determined in the monitor and react phase. Finally, the prepared objects are transmitted in a transmit phase.

In this three-phase feedback loop, the monitor and react phase is the key phase to determine the required quality and the adaptation to be applied. If the desired quality of the application is significantly different from the possible quality of the network, the sender must find objects to transform. However, the transformation algorithms are heavily application dependent, and therefore cannot be specified in a general framework.

Bolliger's framework provides a good abstraction of network-aware applications. It consists of some novel concepts like application/network QoS mapping. However, any application developer who wants to apply this framework, still needs to specify the concrete functions use in each phase. This specification process will be domain and application dependent. We propose to add feedback control loop policy which will also consider network state.

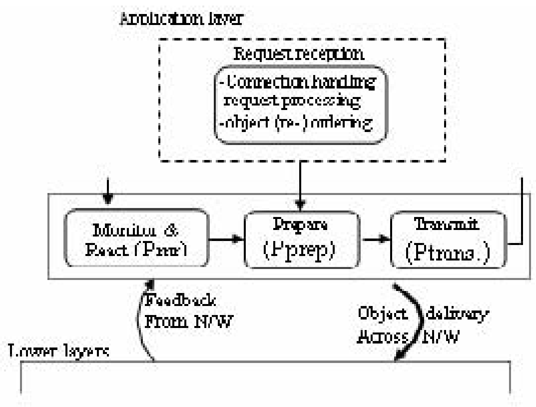


Figure 3. Feedback control loop

3. Relaxed Layer Architecture E2E Tradeoff Protocol

*A. **Observations:***

The QoS architecture concentrates on the resource-oriented mechanisms like flow control and admission control in the transport layer while putting the service oriented mechanisms in the application layer. Resource oriented architectures like Telnet, Lancaster, and Heidelberg are almost opaque and are not amenable to resource control mechanism needed for adaptive QoS provisioning. Object Oriented Architectures like TINA provides fine-grained structure by applying the principle of separation at the application level. By allowing interaction between the components, it is possible to assign QoS tasks to management objects and to identify access points and signalling paths. The values added services in TINA are usually restricted to signalling paths. Darwin provides specific resource mechanism that can be customized to meet service specific needs. However, Darwin's proposal is only semi open assumes heavily loaded end-points to provide appropriate service to users, treats all the applications a like and presents a common interface to all. However, the lack of openness, excludes the possibility of client resources for better QoS management, openness provides the required visibility needed for observation, prediction and thereby control. By controlling and the relationship between different internal variables, different QoS activities like controlling, filtering, allocation etc can be adopted, in accordance to the policy, service characteristics, or state of the shared resources. To fulfil the constraints of this adaptability, there must be a mediator layer between application and resource management. Sensors must be used a mediator layer between application and resource manager. Sensors must be attached to resource management entities with actuators plugged to the application and resource management entities, so that resources at clients, servers and network, can be rearranged through mediation layers whenever need arises. All this requires the QoS architecture to be able to enlist cooperation of all resources in a seamless transparent manner and warrants the need for a unified framework for QoS provisioning.

Our proposed model is modeled as four layers, named as Relaxed Layer Architecture; it is used to analyze different roles of QoS-based application management within global network architecture. It defines a strict differentiation between QoS-relevant system facets. This model emphasizes three important features: latency feedback, packet forwarding, and service provisioning to optimize user preferences. Here, we also introduce the E2E tradeoff protocol concept for optimizing the session description format and the negotiations (tradeoffs) of the multimedia session parameters. In order to enable optimized QoS tradeoffs of resources for different QoS abstraction levels and for various types of networks, a well structured description model and QoS based architecture for streaming multimedia services in wireless access networks for QoS is necessary. A critical issue of the design of QoS architectures is the deployment of various kinds of mechanisms to observe and to control stream continuity, buffer capacities, transmission delays, and integrity of data.

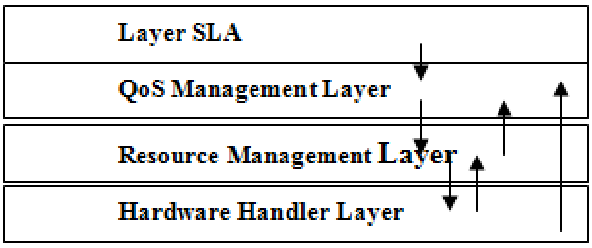


Figure4. Relaxed Layer Architecture

We propose the Relaxed Layer Architecture for quality of service (QoS) centric service level agreements (SLA) between service providers, who also own servers, networks and their clients, to provide customizable services. Resources allocated will dynamically on the basis of the QoS needs, resource availability and negotiation. In the Relaxed Layered Architecture, each layer may use the services of all layers below it as shown in figure 4.

A layer is translucent and not opaque allowing cooperation between the layers for optimal performance. This SLA framework relaxes the layered architecture by allowing QoS management layer to work in collaboration with the resource management and hardware handler layer for resource allocation. This architecture is modeled in UML [6], [7] to describe service creation and structured requirements of the system for desired QoS.

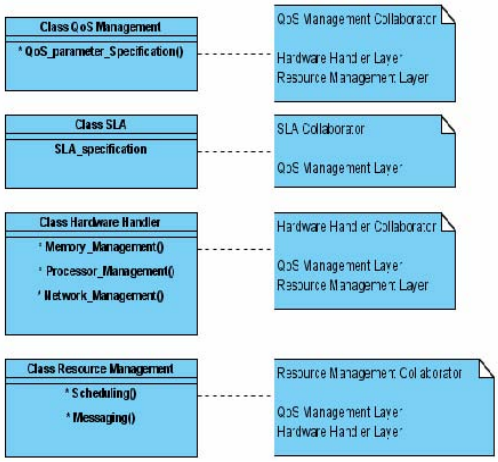


Figure 5. Layers-Functions & Collaborations

The class diagram of application Model, for the Architecture is shown in figure 5 defines the functions and collaborations of different layers. The System States mention continuously and then characterize in QoS space partitioning to upgrade or downgrade SLA during the entire lifetime of the application and its termination.

4. End-System and Service Provider Interaction

QoS aware services to some extent are also related to A4C services [11] [12], because one has to pay more for a premium service. Hence service/session/QoS signalling can interact be integrated in the A4C and security signalling. The interactions with A4C and security services within a QoS-aware architecture occur at two levels i.e. Transport Domain interactions & Service Domain interactions. Figure 6 depicts a general view of a negotiation procedure with E2ENP Protocol (End-to-End Negotiation Protocol) in consideration with internal network services (Only the logical meaning of the interactions). The exact negotiation with SIP- specific methods in accordance with different negotiation scenarios and alternative reservation possibilities can be found in [10]. During a registration procedure (Actions 1 - 6), APN(s) could provide end-systems with validation policies so that end-systems can validate QoS configurations by themselves. In this example, however, the APN(s) validates later. The registration procedure can

be applied in two steps – first: registration to gain network access over pure DIAMETER services [13] and second: registration to gain details on the service support over SIP REGISTER [14] or over a combination of hybrid approaches (e.g. DIAMETER/SIP [15]). APN's (Access Provider Networks) register End-systems with their validation policies in two steps: First registration to gain network access over pure DIAMETER services and second: registration to gain details on the service support over SIP REGISTER [14] or over a combination of hybrid approaches (e.g. DIAMETER/SIP [15]).
The end-system exchange QoS relevant information in a negotiation procedure (Message 1-10). The respective QoS-configuration requests and responses are validated and enhanced with transport QoS (e.g. bandwidth, delay, etc.) and policy parameters (e.g. max. bandwidth, max. number of streams per applications, etc.) by the corresponding APN (s) (Actions 11, 12, 13, 14 and 15). When the response is received, the initiator of the signaling phase can calculate the final QoS configuration (Action 16), and it generates a general system configuration (during the pre-negotiation phase) or a particular session configuration (during negotiation phase). During session setup (negotiation) or adoption (re-negotiation), the notification of the final QoS configuration (Message 17) can also serve as a trigger for network resource-reservation (e.g. via RSVP).

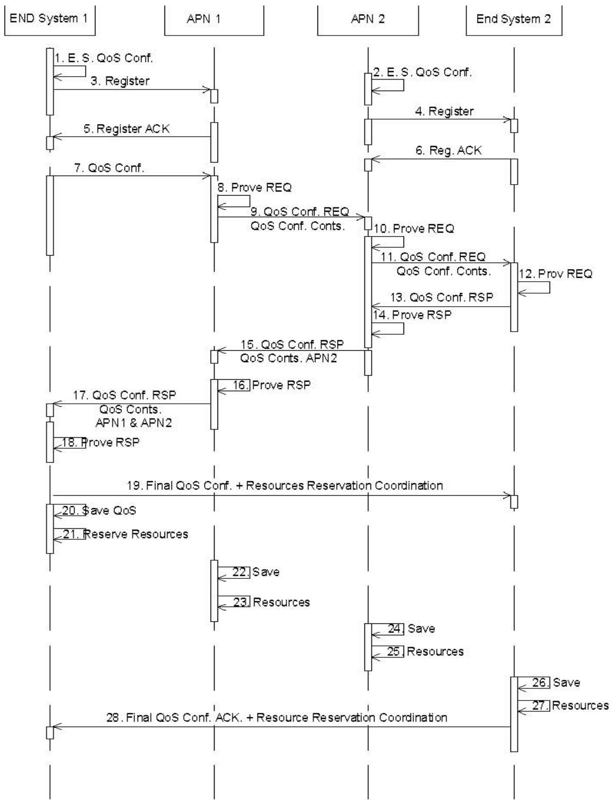


Figure 6. Sequence diagram for End system and APN's interaction

End-systems need to explicitly coordinate their resource reservations, as different underlying reservation mechanisms might be in use (e.g. via RSVP [16], might be terminated at a network gateway within APN1 and/or APN2 might be pure DiffServ [17]). Therefore E2ENP informs the peers about the state of network resource-reservation (Message 13-22). During this resource-coordination process both end-systems and the APN(s) save of the final QoS configuration in order to understand the keys the QoS descriptions that the end-systems exchange during the next E2ENP phases. Thus the APN(s) are also able to assist the end-systems in their network reservations. When listening for the Resource-Reservation-Coordination message as well as interpreting this message, the APN(s) can reserve resources on the network on behalf of that end-system, which cannot do this by them. When the answer to the Resource-Reservation-Coordination message is interpreted, the APN(s) can start the billing process for the reserved resources within the network. Due to the flexibility of the E2ENP description model, it is possible to define and control different signaling scenarios, e.g. with end-system reservation signaling, with provider-initiated reservations, with specific scenarios for initiating metering and charging processes, etc. (see also [10] [9] [8]).

5. SIMULATION RESULTS

The performance of the proposed framework is evaluated using simulation experiments with OMNET++ simulator with Mobility Framework [18]. As illustrated in Figure 7, we have considered a simple topology consisting of two end-systems named as client & endsys and two APN named as APN1 and APN2 through two switches (apnq1 & apnq2 respectively). First end system client(0) registers its request at APN1 with its QoS requirements (in this simulation we had taken bandwidth) to communicate with endsys(0) through APN2 and if resources are available than allocate the resources to the request if sufficient resources are not available than negotiation process apply and degrades the resource of other existing users at the certain level which enables allocation of resources to incoming request.
The simulation statistics is shown in figure 8 and processing time at APN-1 and APN-2 in figure 9 & figure 10. Throughput and performance of the proposed framework is given in figure 11, which shows that the throughput of the end-systems remain in permissible limits, if number of request increases and it may allocate the resources to every request according to simulation results. The uniqueness of this research is that the existence wireless awareness and characteristics of the wireless links

are investigated from both the end-hosts' and the networks' point of view.

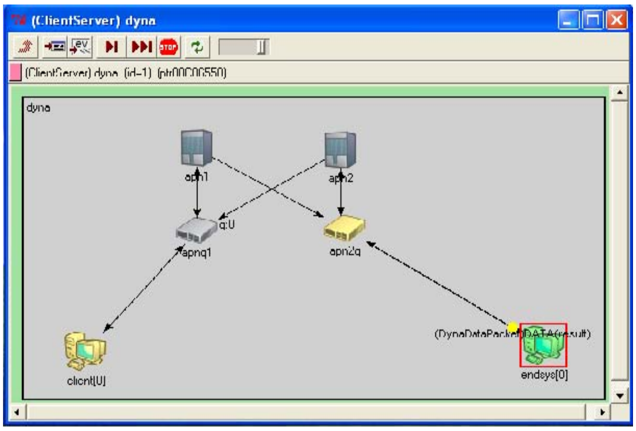


Figure 7. Simulation model of End System and APN's Interaction.

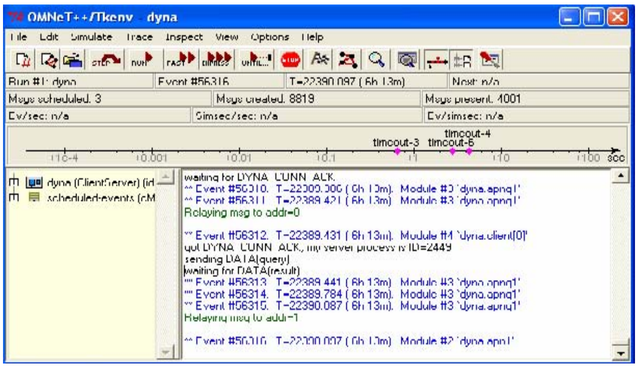


Figure 8. Statistics of proposed model

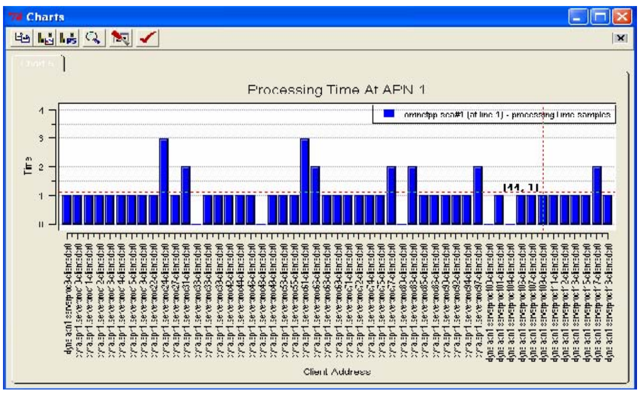


Figure 9. Processing Time at APN-1

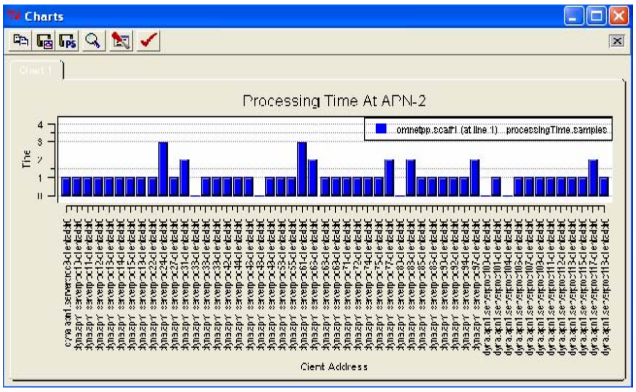


Figure 10. Processing Time at APN-2

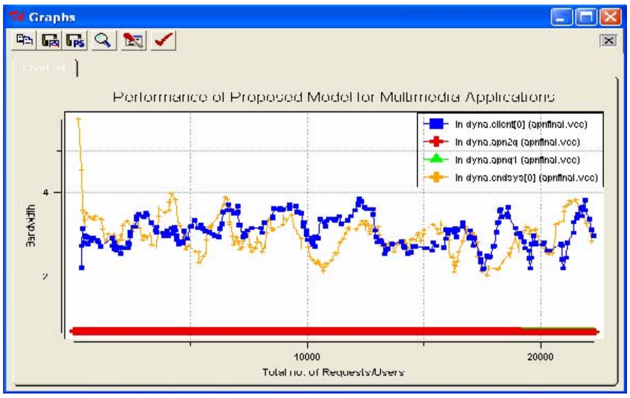


Figure 11. Performance of proposed model

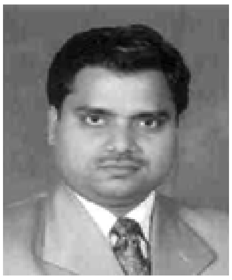
In this paper, we have also highlighted research studies about network-aware applications, especially network aware mobile multimedia applications. As a result of this review, we show that a feedback loop concept is critical to any network-aware application. It must have two critical components: One is for monitoring networks and obtaining information of changes in networks. The other is to select the right object to adapt and select the adaptation methods appropriate to the applications based on the feedback from networks.

6. Conclusions

The proposed frameworks consist of Flow, rate and admission control algorithms that work in conjunction with an application-aware QoS based pro-active middleware to implement a congestion avoidance mechanism for network-aware applications, while maximizing resource utilization and at the same time ensuring the system and application QoS. Affordable QoS provisioning application-layer softwares' are expected to be portable and reusable on top of different commercial off-the-self lower-layer protocols and hardware. All of these engineering requirements can be satisfied by wireless QoS enabled middleware, due to its unique ability of abstracting the heterogeneity of operating systems, networks, hardware, and even programming languages. We would also like to extend and evaluate our framework for other types of applications in the future.

References

[1] H. Bai, M. Atiquzzaman, “Error modeling schemes for fading channels in wireless communications: Asurvey,” IEEE Communications Surveys and Tutorials, vol no. 2, pp. 2-9, Oct.-2003.
[2] C. Hesselman, H. Eertink, "A Scalable QoS Adaptation Service for Mobile Multimedia Applications," presented in Proceedings of the ‘6th Open European Summer school (EUNICE2000)’, Enschede, The Netherlands, 2000.
[3] S. Floyd, K. Fall, “Router mechanisms to support end-to-end congestion control,” Technical report, Lawrence Berkeley National Laboratory, Berkeley, February 1997.
[4] D Verma, “Supporting Service Level Agreements on IP Networks,” Macmillan Technology Series (MTS).
[5] J. Bolliger, T. Gross, “A Framework-Based Approach to the Development of Network-Aware Applications,” IEEE Transactions on Software Engineering, Vol. 24, No. 5, pp. 376-390, May 1998.
[6] ISO, “IT ODP-RM – Quality of Service,” ISO/IEC 1999.
[7] Campbell A., Coulson G. and Hutchison D., “A Quality of Service Architecture,” ACM Communication Review, 24(2), pp. 6-27, 1994.
[8] T. Guenkova-Luy , A. Kassler, J. Eisl, D. Mandato, “End-to-End Quality of Service Coordination for Mobile Multimedia Applications in Heterogeneous Mobile Networks,” IEEE International Conference on Communications, June-2004.
[9] T. Guenkova-Luy, A. Kassler, J. Eisl, D. Mandato, “End-to-End Quality of Service Coordination for Mobile Multimedia Applications,” IEEE Journal on Selected Area in Communications (Issue on advanced mobility management and QoS protocol wireless Internet)-2004.
[10] MIND project, “Top level architecture for providing seamless QoS, security, accounting and mobility to the applications and services,” Deliverable D1.2, November-2002.
[11] G. Huston, “Next steps for the IP QoS Architecture,” IETF RFC 2990, November 2000.
[12] C. Mills et al., “Internet Accounting Background,” IETF RFC 1272, November 1991.
[13] 3GPP TS 24.228 V5.1.0, “Signaling flows for the IP multimedia call control based on SIP and SDP,” Technical Specification, June 2002.
[14] G. Camarillo et. Al., “Integration of Resource Management and Session Initiation Protocol (SIP),” IETF RFC 3312, October 2002.
[15] M. Garcia-Martin et al., “Diameter Session Initiation Protocol (SIP) Application,” IETF Internet-Draft: draft-ietf-aaa-diameter-sip-app-00.txt, October 2003.
[16] H. Schulzrinne et.al, “RTP: A transport Protocol for Real-Time Application,” IETF RFC 1889, January 1996.
[17] S. Blake et. al., “An Architecture for Differentiated Services,” IETF RFC 2475, December 1998.
[18] OMNET++ Simulator, with mobility framework (mobility-fw2.0p2) http://www.omnetpp.org



N. Kumar ((Dr. Narander Kumar) received his Post Graduate degree and Ph. D. in CS & IT, from the Department of Computer Science and Information Technology, Faculty of Engineering and Technology, M.J.P. Rohilkhand University, Bareilly, Uttar Pradesh, INDIA in 2002 and 2009 respectively. His research interest includes Quality of Service (QoS), Computer Networks, resource management mechanism, in the networks for multimedia applications, performance evaluation.

Y. D. S. Arya (Dr. YDS Arya) received the M. Tech. and Ph. D. from IIT Kanpur in Computer Science & Engineering has extensive experience in teaching UG/PG courses and has worked and programmed all kinds of computers starting from IInd generation (IMB 1620) to latest high performance computing clusters from HP and sun Microsystems in the Department of Computer Science and Engineering, Indian Institute of Technology, Kanpur. His current area of interest is Networking and Natural Language Processing.