Report on the 5th IFIP International Workshop on Quality of Service (IWQOS'97)

May 21-23, 1997.
Center for Telecommunications Research
Columbia University, New York City

http://comet.ctr.columbia.edu/iwqos97/

Oguz Angin, Andrew T. Campbell, Lai-Tee Cheok, Raymond R-F Liao, Koon-Seng Lim and Klara Nahrstedt

workshop@ctr.columbia.edu

Abstract

This paper presents a summary of the fifth International Workshop on Quality of Service (IWQOS) which was held at Columbia University in May 1997. The goal of this three-day meeting was to foster interaction between researchers active in the area of Quality of Service (QOS) research, to reflect on past experiences and lessons learnt, and to discuss future QOS challenges. To reflect this goal, this year's workshop included a hot program made up of (i) a keynote address on "Programming Telecommunications Networks"; (ii) panels addressing "QOS for Distributed Object Computing Middleware – Fact or Fiction?" and "Reservations about Reservations"; (iii) a workshop invited paper entitled, "Quality of Service – Where are we?" and (iv) ten technical sessions that included new topics for IWQOS such as mobile communications, QOS routing and QOS-based transport systems. This report summarizes the technical program and captures the main themes and major areas of discussion that emerged during IWQOS'97.

1. Introduction

Over the past several years, there has been a considerable amount of research within the field of *Quality of Service (QOS).* Much of the work has taken place within the context of QOS support for distributed multimedia systems, operating systems, transport subsystems, networks, devices and formal languages. The objective of the *International Workshop on Quality of Service (IWQOS)* is to bring together researchers, developers and practitioners working in all these facets of QOS research. While many conferences and workshops offer technical sessions on the topic QOS, none other than IWQOS, provides a single-track workshop dedicated to the broad subject of QOS research.

The 5th IFIP International Workshop on Quality of Service was held at the Center for Telecommunications Research, Columbia University, and is the latest in a series of continuing workshops. The first workshop, held in May 1993 in Montreal, Canada, was supported by the European RACE project QOS TOPIC and the Canadian CITR project Broadband Services. The second workshop, organized within the European RACE Conference on Integrated Services and Networks (IS&N), took place in September 1994 in Aachen, Germany. The third workshop was held in February 1995 in Brisbane, Australia, in conjunction with the International IFIP Conference on Open Distributed Processing. In March 1996, the workshop was held in Paris in conjunction with the IFIP/IEEE International Conference on Distributed Platforms.

The theme for IWQOS'97 was "Building QOS into Distributed Systems". Implicit in the theme is the notion that the QOS community should focus on discussing results from prototype implementations of their ideas. We were naturally interested in assessing the impact of ideas discussed at previous meetings on future products as QOS ideas move from research to development. While IWQOS is interested in experimental results, it remains a forum for the discussion of fresh and innovative ideas. As a result of this, authors were solicited to provide experimental research (long) papers and more speculative position (short) statements. The technical program successfully reflected the organizers desire to hear about experiment results, controversial ideas, retrospectives and future directions.

2. IWQOS'97 Program Outline

This year's workshop included an invited program, which comprised a keynote address, panels and a workshop invited paper. The keynote address given by Aurel A. Lazar (Columbia University) was entitled

"Programming Telecommunications Networks". The invited panels addressed two aspects of delivering QOS in distributed computing environments and the Internet. The first panel, chaired by Douglas Schmidt (Washington University), addressed the question: "QOS for Distributed Object Computing Middleware – Fact or Fiction?". The second workshop panel entitled "Reservations about Reservations" and chaired by Henning Schulzrinne (Columbia University), discussed the topic of QOS provision in the next generation Internet. To complete the IWQOS'97 invited program, we had a reality check in the form of the workshop invited paper by Ralf Steinmetz and Lars Wolf (Darmstadt University) on "Quality of Service – Where are we?".

A strong technical program that included twenty long papers and twenty short papers complemented the IWQOS invited program, which were chosen from more than seventy papers from nine countries.

3. Programming Telecommunications Networks

Keynote Speaker: *Aurel A. Lazar, Columbia University*

Recent moves toward market deregulation and open competition have sparked a wave of serious introspection in the telecommunications service industry. Telecom providers and operators are now required to open up their primary revenue channels to competing industries. In the keynote address, Aurel Lazar [Lazar,97] focussed on the problem of programmability of telecommunications networks for new services. The speaker outlined an agenda for realizing an open programmable networking environment based on the concept of a *broadband kernel* that better reflects the service creation and deployment environment and economic concerns of future telecommunications systems.

The address began with an examination of the service structure of two major global communication networks (i.e., the Telephone Network and the Internet) exploring their relative strengths and weaknesses. Lazar proposed a three-tiered *open service model* that reflects the economic market structure of the future telecommunications service industry. He considered that the lowest layer of the model reflected a hardware market where numerous equipment manufacturers and vendors offer hardware and firmware solutions for building the basic communication infrastructure. The customers of this market are typically network carriers, third party software developers who specialize in developing software for service providers and a handful of service providers themselves. The Application Programmer Interfaces (APIs) provided by vendors in this market would allow users to write basic communication services and *middleware* components. He considered that the second layer of the model reflects a middleware service market where carriers, software developers and middleware service providers offer middleware service products to customers who are in the user service provisioning business. The APIs provided in the middleware service market are suitable for development of consumer level services. Finally, at the highest layer he considered that the model reflects a consumer services market where consumer service providers compete to bundle, integrate and customize their wares in the most appealing form for mass market consumption. Within each market there may exist brokers whose role is to mediate the interaction and dealings between buyers and sellers who, because of regulatory and business policies, cannot transact directly.

The keynote speaker commented that this service model falls somewhere between the Internet's peer-to-peer model and the Telephone Network's strict provider-customer model. In essence, it allows for cooperation between any number of entities in the network for realizing a common service as well as the competition among services for network resources. The corresponding engineering model can be parameterized in such a way that the basic characteristics of the peer-to-peer model as well as the characteristics of the provider and consumer model can be accommodated. Within each layer (which models a particular market), players are free to enter and buy, sell or re-bundle each other's services. Across layers, the relationship reflects the traditional provider-customer model.

Lazar argued that investigating such a model would help clarify some important issues facing the telecommunications service industry as it deals with changes in service needs. An engineering model for realizing the open service market model was then presented as a vehicle for creating multimedia services on broadband networks. An example of engineering some of the components of the open service model was then presented from an implementation viewpoint.

In an insightful presentation the speaker connected abstractions with real-life implementation reinforcing major themes of the keynote. At one point, Lazar presented a model of a *schedulable region* (which represents the resource capacity of a switch multiplexer) as a live feed from one of the ATM switches in Columbia's broadband network. The schedulable region was configured and managed using a QOS extended version of Ipislon's General Switch Management Protocol (GSMP) called *qGSMP* developed by the

COMET Group at Columbia. To bring the audience closer to the problem of engineering a solution, the speaker incorporated Java applets to illustrate live feeds from switches in the COMET laboratory. The audience clearly saw a representation of a schedulable region and its operational points changing dynamically as the traffic through the switches in the testbed varied over time. Such illustrations helped to showcase several important points of this spirited and informative keynote address.

4. Technical Sessions

IWQOS'97 was a truly interactive event. The general format of each of the ten technical sessions included long and short papers followed by a thirty minute panel discussion on topics raised during the session and in response to questions from audience.

The ten technical sessions comprised: mobile communication, QOS routing, advanced reservation, traffic management, QOS and video systems, distributed object computing, QOS management, QOS-based transport protocols, QOS mapping and QOS adaptation.

4.1 Mobile Communications

Chair: *Mahmoud Naghshineh, IBM*

The opening session of the workshop investigated the feasibility of delivering continuous media with QOS guarantees in mobile networks. A key observation of the session was that providing QOS support in a wireless and mobile environment requires a fundamentally different approach from that found in wireline networks. The session included two long and four short papers.

Presentations

The first talk was given by Bharghavan [Lu,97], University of Illinois at Urbana-Champaign. Bharghavan proposed a number of solutions to the problem of providing sustained QOS to mobile applications. Limited and varying resources availability, stringent application requirements and user mobility make providing QOS guarantees in mobile and wireless environments challenging. He introduced an adaptive service model that enables the network and mobile applications to renegotiate QOS depending on dynamic network conditions. Following this he described an algorithmic framework that provides cost-effective resource adaptation in the presence of resource dynamics. Bharghavan concluded by arguing for a unified architecture for QOS adaptation.

The next presentation, by Stephen Wade [Blair,97], Lancaster University, addressed the design of a distributed systems platform and algorithms for mobile computing environments. The platform, called Limbo, aims to support the development of demanding mobile-aware distributed applications in a heterogeneous networking environment. Limbo is based on the concept of tuple spaces, which has been extended for the mobile computing environment to support QOS management. Limbo places emphasis on QOS monitoring and adaptation. The speaker argued that the tuple space paradigm was particularly useful in modeling adaptation to changes in network connectivity in mobile networking environments.

The third talk of the session, by Badrinath [Badrinath,97], Rutgers University, focused on architectural support for Internet cellular telephony. It was evident to Badrinath that those who designed RSVP and the integrated service architecture had not looked at mobility issues. In contrast, those that developed mobile-IP had not addressed QOS issues. The speaker made an argument to unify some of these disparate pieces in support of cellular phone services. The proposed service was based on an IP network capable of delivering packetized voice to moving users. Badrinath calls the solution "RSVP+mobileIP+QOS".

Next, Andrew Campbell [Campbell,97], Columbia University, discussed a number of QOS challenges for next generation mobile middleware. The speaker reflected that the area of QOS and mobility was in its infancy. The wireless media systems project at Columbia was attempting to shine some light on the subject by building a prototype QOS-aware middleware platform for mobile multimedia networking called mobiware. The platform is programmable and runs on mobile devices, base stations and mobile-capable switches. Mobiware includes a new active and adaptive transport, QOS controlled handoff algorithms and an adaptive network QOS model.

The next speaker, Javier Gomez-Castellanos [Gomez,97], Columbia University, presented results showing the effect of transmission errors on MPEG streams over wireless links. The speaker proposed several algorithms which improve the perceptible quality of MPEG stream during periods of fast and slow fading. Gomez-Castellanos suggested an algorithm based on a combination of Forward Error Correction (FEC) and Automatic Repeat Request (ARQ) as a way to minimize the impact of error characteristics on video in this instance. He introduced a packet tagging technique that takes into account the particular semantics of MPEG flows and the relative importance

of different packets (e.g., GOP, headers, IPB, scalable profiles) as they traverse a mobile network.

In the final presentation of the session, Steven Pope [Pope,97], Olivetti and Oracle Research Laboratory, discussed QOS support for mobile computing environments. There is a demand for completely portable computers (which Pope called walkstations) to access the network while traversing both indoor wireless LAN networks and the outdoor mobile radio network infrastructure. The speaker introduced a traded handoff where connections are rebuilt during a handoff to the most appropriate service, taking into account the properties required by the application and locally available, replicated or compatible services.

Panel Discussion

At the end of this session, the question of how to provide QOS in wireless and mobile environments was raised. The session chair presented two opposing views for consideration by the panel where the mobile network provided:

- hard guarantees, in which case there was no need for adaptation mechanisms in applications – as in the case of future wireless ATM networks; and
- no guarantees, in which case there was a strong need for highly adaptive mechanisms in the applications – as in the case of today's best effort mobile IP network.

Many of the panelists thought that QOS offered by future mobile multimedia networks would lie between the two extremes. The panel considered supporting hard guarantees for a wide range of multimedia traffic unrealistic given the nature of the wireless medium and mobility requirements. Bhraghavan stated that there seems to be universal acceptance that an intermediate service be based around an adaptive resource management model. This approach benefits by reducing handoff dropping probability and increasing the utilization of the network.

Mahmoud Nagshsineh followed up with another question to the panel: if there was agreement on an adaptive resource management then where should the adaptive algorithms reside: at the physical, MAC, network, transport or application layers? If applied solely to application layer, this would result in low cost and minimal complexity in the network. On the other hand, if the network explicitly supports adaptation then this would result in high cost and increased network complexity. Weighing these arguments the panel agreed that there was need for research into adaptation support in the network and the end-system.

Toward the end of this lively discussion, there was a question from the audience concerning the relationship between pricing models and adaptive mobile multimedia environments. Badrinath explained that clearly there would be a premium cost for mobile over mobile incapable users – as exists today in cellular telephony systems. The issue is what would the policy be when a user service is forced to degrade to a lower quality; how do we cost that likelihood? Badrinath speculated that while users currently pay a premium for mobility they would naturally expect a discount from the provider as a consequence of network initiated QOS degradation.

4.2 Traffic Management

***Chair**: Ed Knightly, Rice University*

The four long papers presented in the traffic management session approached the QOS traffic management problem from various angles and proposed widely differing solutions. The first paper considered the worst case traffic pattern for source policing as an integral part of the traffic model. The second and third papers focused on Markovian modeling techniques, while the fourth advocated a measurement-based approach.

Presentations

Philippe Oechslin [Oechslin,97], University College London, presented the first talk of the traffic management session on the topic of myths of on-off sources. This was in relation to the worst case arrivals of leaky bucket constrained sources. Simulation results from a set of independent connections limited by leaky bucket shapers and fed into a buffered multiplexer were presented. Oechslin indicated that this scenario was typical of an ATM switch or in a looser sense typical of an RSVP capable router. Results from the analysis found periodic traffic patterns resulted in poorer loss rates over the on-off or tri-state patterns models. The results invalidate the widespread belief that on-off patterns are the worst case traffic of independent leaky bucket constrained sources.

The second paper, given by Hoon Lee [Lee,97], Korea Telecom, presented a cell access control scheme for guaranteeing multiple classes of cell loss QOS requirements in an output buffer of an ATM switch. Lee proposed a class acceptance controller, which regulates the acceptance of the cells of QOS classes, based on the dynamic state of the queues. He considered decision functions for the class acceptance controller with a view to comparing their effects to the QOS performance. Queueing analysis of the scheme derived a number of performance measures. The implications of the work

were further illustrated using a number of numerical experiments.

The third paper, by Dietmar Becker [Becker,97], Aachen University of Technology, reported on queueing analysis for a partial buffer system with discrete Markovian arrival processes. An evaluation of the performance of the partial buffer system with finite capacity, deterministic service time and multiple sources was presented. A discrete Markovian arrival process modeled each source. The queueing system was evaluated for several traffic compositions and different sizes of the shared buffer area. Becker considered a number of traffic compositions including VBR sources with periodical or negative exponential correlation functions and CBR traffic with fixed interarrival cell emission. The probability distribution of the cell loss of each source was presented.

The final talk of the session, on real-time estimation of the link capacity in multimedia networks, was presented by Piergiulio Maryni [Maryni,97], DIST-University of Genoa. Maryni suggested that simple but powerful abstractions that represented the capacity of multimedia networks are needed. In order to guarantee QOS, the link capacity must first be calculated, i.e. the total number of calls of different types that can be admitted on a single link at a given time. The speaker used the notion of a schedulable region to represent the link capacity for an ATM multiplexer. Maryni presented a new approach for computing the schedulable region in real-time which could be used as input to admission controllers. The methodology relies on real-time QOS measurements to dynamically compute the size and the shape of the region. No assumptions about traffic resource models or scheduler operations were needed for its construction.

Panel Discussion

The panel discussion began by addressing the role of traffic models in delivering QOS guarantees. The panel recognized the difficulty in traffic modeling and the complexity of meeting traffic characteristics and adapting to a wide variety of user applications needs. Most of them agreed that the measurement based approach could significantly relieve some of the difficulty inherent in many traffic models proposed in the literature.

Ed Knightly commented on the weakness of measurement-based resource allocation, whose inaccuracy lies in the measurement procedure that averages out the heterogeneous behavior of traffic flows. The panelists believed that Markovian modeling and worst-case analysis can still play roles that complement measurement-based approaches; that is, in supporting new types of traffic where measurement data is not available, or in planning backbone networks where the requirement on the modeling accuracy is relaxed due to the multiplexing of a large number of flows.

Lee indicated that in the backbone networks, resources are conservatively dimensioned so that the control algorithms are not sensitive to the inaccuracy caused by measurement algorithms. This led to a debate initiated by Maryni on whether traffic modeling is needed at all. In future, he rhetorically argued, why use traffic modeling techniques when network capacity may be infinite? Since traffic modeling aims at increasing bandwidth utilization, with the abundance of bandwidth this may disappear like the technique for reducing the number of transistors in circuit design. Oechslin added that abundant bandwidth availability would be a reality now if the right pricing model were in place to generate enough revenue to turn on more bandwidth.

4.3 QOS Routing

Chair: *Mischa Schwartz, Columbia University*

The basic goal of QOS routing is to select a path through the network that satisfies a set of QOS constraints while achieving some level of network resource utilization. The first two long papers in this session presented differing approaches to achieving this goal. Following this, Scott Corson, Maryland University, joined the panel discussion and raised a number of open issues in providing QOS support when routing flows through mobile ad-hoc networks.

Presentations

The first talk given by Qingming Ma [Ma,97], Carnegie Mellon University, addressed the issue of QOS routing for traffic with performance guarantees. The speaker presented some initial results on QOS path selection for traffic flows requiring bandwidth, delay and jitter guarantees. For traffic that required bandwidth guarantees, it was found that several routing algorithms that favored paths with fewer hops perform well. Ma argued that a modified version of the Bellman-Ford shortest-path algorithm in polynomial time was sufficient for traffic with delay guarantees. He showed that the problem of finding a path that can satisfy bandwidth, delay, jitter, and/or buffer could be solved while at the same time deriving the bandwidth that has to be reserved to meet these constraints.

S. Verma [Verma,97], University of Toronto, reported on QOS based routing in support of emerging multicast multimedia communications. Verma proposed a routing metric that could be used in combination with heuristic

algorithms to find the multicast tree for guaranteed QOS services. The optimum tree used in the formulation was a delay-bounded minimum Steiner tree. Simulation results showed a marked improvement in network utilization expressed in terms of cost over other proposed schemes such as QOS path first routing.

Panel Discussion

The panelists were asked by the audience to clarify the relationship between routing and resource reservation for end-to-end QOS guarantees. Ma stated that the first goal is to discover a path and then reserve the resources along the path or reject the flow/call if resources are not available. Other approaches handle routing and resource reservation at the same time – connection oriented systems couple routing and resource reservation.

The issue of state management was debated. An objective of QOS routing is to distribute state information that accurately reflects available resources on a particular path. Probabilistically resources may not be available at a certain switch/router if the state is old. The panel discussed the implication of QOS state management and concluded that it needs to realistically represent available resources. In the past, call setup used crank-back techniques to resolve inconsistencies in the QOS routing state representation.

Scott Corson [Corson.97] described the generic character of routing in mobile ad-hoc networks. He highlighted some of the differences between ad-hoc and wireline multihop networks and the impact these differences had on QOS-based delivery systems. Corson suggested that supporting QOS in ad-hoc networks was an extreme challenge. Wireless QOS architectures must provide a balance between network and application-level QOS adaptation. This, he argued, would help minimize QOS-related signaling which has traditionally been integrated with routing state to support a given QOS. Corson concluded by saying that two approaches seemed to warrant further investigation for ad-hoc services; namely, minimal guaranteed QOS and probabilistic QOS.

4.4 QOS and Video Systems

Chair: *Alexandros Eleftheriadis, Columbia University*

There are two schools of thought related to QOS and video systems. One school argues that resource reservation is essential for video services. The other argues that the network need only be engineered to support best effort services. In this case networked video systems need only estimate resource availability and intelligently adapt to the observed state. This session included two long and two short papers that reported on adaptive QOS driven video systems.

Presentations

Sanjay Jha [Jha,97], University of Technology, Sydney, proposed a set of playout management algorithms for interactive video. The work examined problems associated with display of live continuous media. Under the assumption that the network cannot guarantee the required bounds on delay and jitter and the operating system scheduling is non-real time, there is a need to accommodate the delay and jitter in the end systems in order to maintain a desirable QOS. Jha proposed a method of video playback that requires accurate estimation of display cycle time of video frames and the delay suffered by frames in packet networks. Deterministic forecasting methods used in time series analysis were applied to experimental data collected from video transmission.

The second paper by Rajeev Koodli [Koodli,97] proposed the notion of noticeable loss, which directly relates loss pattern to the perceived QOS for an application. Noticeable loss was used to evaluate resource management algorithms that provided QOS to individual adaptive applications. Simulation results illustrated the performance of the algorithm.

D. Bourges Waldegg [Waldegg,97], ENST, presented a temporal QOS based CPU scheduling model for multimedia. The model is based on a playout specification and a runtime application structure that allows workahead processing and quality degradation for delay during overload conditions. Policy is used by a scheduler to degrade the service in a meaningful way with the goal of supporting better resource utilization. Waldegg added that the run time structure supported workahead processing which had shown benefit particularly in the case of overload.

The last talk by Steven Jacobs [Jacobs,97], Columbia University, proposed that networked video systems could operate sufficiently well over best effort networks. Architecting support for video service in packet networks required the addition of a number of valued added algorithms. Jacobs stated that there was great demand for such services today. Furthermore, he speculated that the demand would persist in the future even when many networks supported multi-level QOS assurances. An Internet-based video delivery system was presented which combines both image processing and networking techniques. Jacobs presented the results from several experiments that utilize a combination of bandwidth estimation and dynamic rate shaping of video sources.

Panel Discussion

Alexandros Eleftheriadis led a lively discussion which addressed a number of issues raised in the session and then opened the floor for questions. The first question concerned media scaling techniques. Eleftheriadis commented that the choice of either frame dropping or quality reduction within a frame (i.e. dropping transform domain coefficients) is strongly content dependent, which can vary along the wide spectrum of applications. Some applications, such as movies, have scenes that tolerate little content loss and require real-time delivery, thus they would have difficulty with best-effort delivery.

The panelists discussed whether hardware solutions for video can alleviate many of the problems experienced using software implementations. Jacobs argued that hardware support for video is improving and one should address other problems instead of duplicating hardware solutions in software. Eleftheriadis pointed out that the types of hardware assists are converging as content providers only use limited set of encoding standards. However, Jha argued that even for the audio codec, many interoperability problems remain. In today's Internet, lack of QOS support means that researchers have little choice but to use flexible software solutions.

The last question from the audience addressed the pros and cons of dynamic rate shaping and quantization. Jacobs pointed out that dynamic rate shaping is better than the quantization technique when the required rate reduction is small. Dynamic rate shaping simply drops some coefficients, requires less state information to be propagated, and offers better quality under this scenario. In other words, quantization pays the penalty for SNR and speed as long as the required reduction ratio is not too large. When a large reduction is needed, however, quantization techniques are superior.

4.5 QOS Management

Chair: *Andreas Vogel, Visigenic Corp.*

The next session, chaired by Andreas Vogel from Visigenic, included two long and three short papers. Each of these papers focussed on the issue of QOS management techniques. The presentations covered both architecture and implementation research, with topics ranging from QOS support in operating systems and file systems, end-to-end specification, adaptive QOS architecture and QOS management agents.

Presentations

The first paper, presented by Lakshman [Lakshman,97], Intel, reported on an integrated approach to CPU and network IO QOS management. Lakshman illustrated the challenge of proving predictable QOS for multimedia applications. In such an environment applications may not know their exact resource requirements in advance and resource requirements and resource availability may be time-varying. To address these challenges, Lakshman proposed a resource management architecture in which applications and the operating system cooperate to dynamically adapt to variations in the resource requirements and availability.

The topic of QOS management of integrated services communication systems was addressed next by Roland Bless [Bless,97], University of Karlsruhe. It is likely that such networks will support a wide variety of applications that will be multiplexed into different service classes. Bless reported on a flexible QOS management scheme for such communications environment. The approach provided flexibility in tailoring the QOS management to suit specific service profiles. Bless outlined the implementation of the QOS management system and presented performance measurements to illustrate the approach taken.

Next, Stefan Fischer [Fisher,97], University of Montreal, presented a decentralized scheme for cooperative QOS management. The system supports QOS management functions such as QOS negotiation and adaptation over the network. The speaker suggested that the work was especially suitable for multicast multimedia communications. Fisher introduced the notion of QOS agents, installed throughout the system where QOS negotiations occurred. QOS agents communicate with their neighboring agents. This communication is application oriented, i.e. the agents know about QOS requirements and negotiated values on behalf of the users. Fischer concluded by discussing the construction of applications based on the notion of cooperative QOS management.

Jan de Meer [de Meer,97], GMD Fokus, reported on the specification of end-to-end QOS control. The speaker distinguished between the QOS control demands for continuous and discrete media. Classic control theory seemed an appropriate vehicle to monitor and respond to time varying quality as flows progress through the communications system. De Meer introduced a "water-level monitoring" paradigm which consists of four components, namely, a source, container, monitor and sink, to assist in the QOS management process.

The final talk of the session, presented by M. Spasojevic [Borowsky,97], HP Labs, on using attribute-managed storage to achieve QOS. The speaker argued that specification of storage systems by means of user-oriented QOS attributes is the key to ease of use and efficient resource utilization. Currently, most existing storage management is too low level requiring the user

to allocate and configure arrays of disks. However, little help is provided in helping the user do this. Attribute-managed storage systems hide details of the underlying storage systems through virtual storage abstractions, units of storage with QOS guarantees. The speaker described a prototype matching engine called Forum currently under development.

Panel Discussion

The first question from the audience addressed existing operating systems: are they sufficient to support integrated QOS? Many panelists thought that work needs to be done before a QOS-driven operating system would be available on the market. Lakshman, from his experience with using Solaris, commented that the hardware already has a high-resolution clock, whereas he considered his enhancements to the operating system targeted faster preemption and better estimation of computing time operations.

A member of the audience asked whether the whole area of QOS management research is well understood. The panel observed that there is still a lack of experience with QOS management systems, with QOS mapping across layers, and with QOS specifications. They believe that the research would shift to toward building large-scale prototypes. This would depend on the availability of a large-scale network testbed.

Finally, Fisher commented that the IP networking view of QOS management architecture is that it was far too complex. It may be best to revisit fundamental control theory, focus on different time-scales and develop simple solutions.

4.6 Distributed Object Computing

Chair: *Douglas Schmidt, Washington University at St. Louis*

Distributed object computing technology has been widely applied to resolve problems stemming from the complexities of developing large heterogeneous software systems. This session focuses on some of the QOS issues that surround the use of distributed object-computing technologies. Issues addressed ranged from QOS parameterized trading services to QOS meta-data management for middleware. This session included two long and three short papers.

Presentations

The first presentation of the session, by Claudia Linnhoff-Popien [Linnhoff-Popien,97], Aachen University of Technology, addressed the integration of QOS constraints into the service selection process. The major challenges addressed by the work were the formulation and evaluation of customized CORBA trader. A service distance is computed between the client and the service offers taking into account QOS properties. The modified trader selects the service with the minimal distance to the service request. The Orbix trader was used to evaluate this QOS metric. Claudia Linnhoff-Popien concluded her talk by describing some implementation results in comparison to an unmodified Orbix trader.

The next paper, presented by W. Almesberger [Almesberger,97], EPFL, surveyed QOS in communication APIs. The speaker first contrasted the RSVP and ATM UNI APIs and distinguished how QOS was exposed to the applications using WinSock 2, X/Open and Arequpa. A natural consideration is the mapping of these APIs to local operating system and network resources. Almesberger went on to summarize how these APIs enable applications to control QOS for their connections/flows. Each API has its own idiosyncrasies (e.g., native ATM APIs deal in cells and RSVP APIs in bytes) and supports a different set of QOS parameters and traffic characterizations. The speaker proposed the unification of QOS description using better abstractions to resolve idiosyncrasies found in existing APIs.

In the third talk, John Zinky [Zinky,97], BBN, reported on managing systemic meta-data for creating QOS-adaptive CORBA applications. Zinky made the point that distributed applications must be able to adapt to quite diverse operating conditions. The speaker introduced the notion of *systematic meta-data*, which captures how applications utilize distributed systems technology such as CORBA, adapt their QOS requirements, use of resources, and allocation policies. In this position statement, Zinky discussed some possible solutions. Meta-data required support from the network, distributed system and applications. The network needs to support explicit mechanism for moving and storing meta-data. Application programmers need APIs for QOS at the client/object boundary rather than at the socket level.

The fourth paper in this session was presented by D. Reed [Reed,97], Stirling University, on supporting QOS components in distributed environments. A key component required in such an environment is QOS monitoring services. In his talk Reed proposed a generic distributed monitoring service which adapts to suit particular applications as a means of overcoming the complexity of specialized monitoring solutions. A number of examples highlighted the flexibility of the monitoring service.

In the final talk of the session, Andrew Grace [Smith,97], British Telecom, reported on a QOS

configuration tool for distributed applications. The range of distributed applications has increased dramatically over the past several years fuelled by the growth of the Internet. Many applications tend to require user level knowledge of low-level technical parameters requiring an appreciation of system heterogeneity issues rather than simply stating what service they require. In this presentation Grace described a working system for QOS configuring for Mbone conferencing applications. QOS profiles were adopted as a means of specifying resources and requirements in the end-system and networks.

4.7 Advanced Reservation

Chair: *Hideyuki Tokuda, Keio University*

Is there a need to make reservations in advance? In this session, a number of speakers argued that there are sets of applications that require a high degree of resources availability in advance. The first paper takes an empirical look at advance reservation from the network viewpoint. The second and final paper in this session reports on a scalable video-on-demand system which utilizes advanced reservation techniques.

Presentations

Olov Schelen [Schelen,97], Lulea University, began this session with a presentation on sharing resources through advanced reservation agents. The speaker proposed an architecture where clients make advance reservations through agents responsible for advance admission control. The agents allocate resources in the routers just before they are needed for packet forwarding. Schelen illustrated that network resources can be shared between immediate and advance reservations applications without pre-partitioning. Admission control decisions for immediate reservations use information about resources to be allocated for advance reservations in the near future. The speaker introduced a new parameter in the admission control algorithm for the lookahead.

Next, Abdelhakim Hafid [Hafid,97], Computer Research Institute of Montreal, proposed a scalable video-on-demand system that uses advanced reservation techniques to support services. Typically, video-on-demand systems check whether there are enough available resources to deliver the requested movie to the user's host. Given sufficient resources, the movie presentation will commence, otherwise, a rejection is sent back to the user. Hafid described an advanced reservation signaling system called NAFUR. If a QOS request for a video stream cannot be immediately supported at the desired rate NAFUR determines at which point in advance of the current request time the video can playout at the desired rate. If the user wishes to only accept the desired rate, the system makes use of advanced reservation to book resources ahead of time for the duration of the video-on-demand.

Panel Discussion

Steve Pink and Lars Wolf joined the presenters and Hideyuki Tokuda for the panel discussion on advanced reservation. The session chair raised a number of interesting questions and directed them at the panel. Tokuda asked if we really need advanced reservations. If so, what type of quality demanding services shall we use advanced reservation for? How can we maintain reservation state and how would failure be handled?

Steve Pink was unsure of the demand for advanced reservation services. State management of reservations looks troublesome since the models introduce a lot of state and require switches/routers with plenty of memory/storage to maintain it. Pink suggested that existing reservation architectures might be too complex for the network at the moment. Therefore, he proposed to separate control functions from packet forwarding. Pink also argued that RSVP is not the most appropriate choice for advanced reservations signaling. For RSVP to function properly, the senders should be present in advance. This cannot always be guaranteed for advanced reservation applications. Therefore, advanced reservations are a function of the management layer and not the packet-forwarding network.

Lars Wolf summarized what he considered to be the open issues on the topic. These included the duration of reservation, stacks, failures, distribution of announcement information, management of resource and required protocol changes. The most challenging problems are state maintenance and failure handling. All systems performing advance reservation must keep the associated state information for potentially long periods of time. This must be stored in non-volatile memory to survive system shutdown. Wolf described this as the hard-state approach. Alternatively, similar to the approach followed by RSVP, the reservation may be refreshed from time-to-time – he called this the soft-state approach. Wolf believes advanced reservations could be useful for several application classes. However, advance reservation raises difficult questions that need to be resolved.

4.8 QOS-based Transport Protocols

Chair: *Steven Pink, Swedish Institute of Computer Science*

Historically, transport protocols have been a hotly researched topic in computer networking. With the

advent of multimedia, there has been a move away from designing reliable, high performance data transports to transports that support end-to-end QOS guarantees. This session consisted of two long and two short paper presentations on issues surrounding the development of QOS-based transport systems.

Presentations

The first paper presented. by K. Fukuda [Fukuda,97], investigated QOS mapping issues between a user and a video transport system. Fukuda described a QOS mapping method between user preference for video quality and the required bandwidth to transport the resulting video flows. This work assumed that the underlying network is capable of supporting a bandwidth allocation mechanism such as deterministic bit rate service class in ATM, RSVP, IPv6, etc. Based on spatial, SNR and time resolutions QOS parameters, the QOS mapping function derives the required bandwidth to support MPEG-2. The mapping between QOS parameters and user perceived video quality is then calculated using classic mean opinion score evaluation testing.

Jean-François Huard [Huard,97], Columbia University, reported on end-to-end QOS mapping. A simple mathematical formulation for mapping QOS parameters between application and transport was derived. A platform was developed for evaluating end-to-end QOS by performing concurrent network, transport and application level measurements. The loss bound empirically obtained under the assumption of uniformly distributed cell losses within a video frame is too conservative. Early results suggest that the existing literature on loss mapping is typically too conservative by a factor of three. Huard concluded that in order to obtain better empirical QOS mapping rules between the application, transport and network, more data needs to be collected and analyzed.

Next, P. Conrad [Conrad,97], University of Delaware made the following position statement: “Transport QOS over Unreliable Networks: No Guarantees, No Free Lunch!”. The talk presented an approach to transport in unreliable networks, investigating trade offs between qualitative QOS parameters (e.g., order and reliability), and quantitative parameters (e.g., delay and throughput). Conrad focused on partially ordered and partially reliable transport services. The key results are that both sender-based and received-based reliability schemes for providing partial reliability achieve almost identical reliability and delay. On the other hand, a sender-based approach provides better throughput than a receiver-based approach at higher loss rates.

In the final talk of this session, Glenford Mapp [Mapp,97] Olivetti & Oracle Research Lab, reported on development of a QOS-based transport protocol called A1. The transport was designed to provide QOS trade offs rather than strong guarantees. Mapp discussed the trade offs between qualitative QOS such as order and reliability and quantitative QOS such as delay and throughput. The transport service supported the notion of a QOS vector to specify all transport requirements at the API. Preliminary performance results for A1 running over ATM were presented and compared with an efficient kernel implementation of TCP/IP.

Panel Discussion

In the panel discussion that followed, the Chair, Steve Pink initiated a discussion by making the observation that two opposing trends seem to be emerging within transport design. On one hand, as applications become increasingly sophisticated in their requirements, newer transport protocols should be developed to support the required functionality. On the other hand, it is argued that since only applications can truly understand their data semantics, traditional transport functions should be removed from the transport to the applications layer. Pink then solicited questions from the audience.

A member of the audience asked that if trends were toward thinner and thinner transports then why have a transport layer in the first place? The panel's response to this question was mixed. Glenford Mapp agreed with this comment and was of the opinion that the transport in its traditional form was on its way out! Jean-Francois Huard, however, took a different line. In his opinion, the classical algorithms for providing reliable, flow-control and sequenced delivery are too complex to be left to the average application programmer. Pink agreed with this and noted that the recent move toward multicast communications may redefine the role of current transport protocols. Here, the difficulty lies in providing scalable reliable multicast flows. Many of the well-understood techniques including positive acknowledgements do not work in this environment.

Pink followed by leading a discussion on user versus kernel level transports. He noted that there are primarily two types of processing activity in any transport protocol: the more expensive per-byte processing for computing packet checksums and the less expensive per-packet processing for flow-control, acknowledgment, etc. The classic argument for kernel-level transport protocols has been to perform efficient per-byte processing. However, as network speeds increase to gigabits, it seems sensible to delegate much of this computer intensive functionality to specialized hardware. This leaves the task of per-packet level

processing to software. The consensus among the panelists was that user-level transport implementations could be as efficient as kernel level implementations. This, plus the added flexibility of being able to decouple the transport from the operating system makes user-level transport protocols ideal for the increasingly specialized needs of today's sophisticated applications.

4.9 QOS Mapping

Chair: *Jean-Pierre Hubaux, EPFL*

QOS mapping performs the function of automatic translation between representations of QOS at different system levels (i.e., application, operating system, transport and network, etc.) and thus relieves the user of the necessity of thinking in terms of lower level specifications. This session included two long and two short papers which addressed the translation between application QOS specification, and the operating system and network QOS.

Presentations

In the first talk, N. Nishio [Nishio,97], Keio University, presented a simplified method for session coordination using a three-level QOS specification and translation scheme. It may be unrealistic to expect the application to specify its QOS requirements using operating system/network specific language, e.g., memory size in Kbytes or bandwidth in Mbps, etc. However, such information needs to be distilled from the application specification for admission control and resource reservation. To address this need, Nishio introduced a QOS architecture which presented the user-level specification in terms of application program QOS, middleware QOS, and system-level QOS. Nishio described a conductor/ performer paradigm used to handle the translation function between these system components.

In the next presentation, H. Knoche [Knoche,97], University of Hamburg, reported on a quantitative QOS mapping approach. Knoche identified QOS mapping as the process of translating QOS parameter bounds from layer to layer and finally to specific resources. General mapping between video frame service data units and network quality requirements such as delay jitter, throughput and reliability were presented. QOS mapping took into account the cases of common service functions such as segmentation/reassembly, blocking, playout buffer, interleaving, coding, or peak smoothing.

Klara Nahrstedt [Kim,97], University of Illinois, Urbana-Champaign, presented an integrated view of QOS translation and admission control. Nahrstedt discussed a translation between the MPEG-2 video quality representation and the underlying operating system resource, namely the CPU. A communications model was analyzed for different MPEG grouping schemes and communications paradigms. Nahrstedt commented that a middleware level seemed like a natural point where the user can specify QOS requirements using application language and the operating system/network derives QOS parameters expressed in its own language

In the final paper of the session, Valerie Issarny [Issarny,97], INRIA, discussed the translation of QOS specifications in a QOS architecture. Issarny discussed the development of customized software architectures. These include the specification of execution properties such as interaction properties, that capture the communication patterns, and QOS properties, which represent resource management policies implemented by middleware. The main challenge of QOS translation is to correctly specify the interaction and QOS properties so that a proper customization of the resulting QOS architecture may be provided.

Panel Discussion

Jean-Pierre Hubaux highlighted the layers between the application and operating/network for which QOS mapping was needed. He then posed a general question to the panel: given the diversity of application and underlying networks and operating systems, is it feasible to try to formulate a generic framework for QOS mapping and, if so, how far are we from it? The response from the panel was unanimous in that they believe that QOS mapping is an area that is just beginning to be understood.

However, the reasons given by each member of the panel varied widely. Klara Narhstedt reasoned that any form of generic framework would have to be qualified given the diversity. The typical technique used today is through some form of application profile characterization. She believes that the whole field of QOS mapping is still in its infancy given that the scope of mapping considered in most schemes is still fairly simplistic and static in nature: focusing on a certain application, operating system and network.

Another point raised by the floor was that most QOS mapping schemes address only continuous media services avoiding other communications services, e.g., transactions. Hendrick Knoche felt that the primary difficulty in formulating good QOS mapping stems from the fact that we do not yet understand the perceptual effect of QOS. Nobuhiko Nishio noted that the interaction problem among the many layers in a system complicates the issue of QOS mapping since the

dynamics are usually non-linear and difficult to characterize.

Jim Van Loo, Sun Microsystems, asked the panel if there was a good abstraction or metaphor for expressing QOS mappings? This question cut to the heart of the problem. Nobuhiko reflected that whatever the abstraction might be, it would likely be based on economic concepts and be closely indicative of the cost involved of rendering the service to be mapped.

4.10 QOS Adaptation

Chair: *Klara Nahrstedt, University of Illinois at Urbana-Champaign*

Many distributed multimedia applications are adaptive in nature and exhibit flexibility in dealing with fluctuations in network conditions. QOS adaptation algorithms can, for example, trade temporal and spatial quality to available bandwidth or manipulate the playout time of continuous media in response to variations in delay. This session comprised two long and three short papers on the topic.

Presentations

In the first talk of the session, Pratyush Moghe [Moghe,97] Bell Labs, addressed what he described as "terminal QOS" for adaptive applications. Next generation terminals are expected to support sophisticated adaptive applications. Since some terminals have limited power (e.g., personal digital assistants and network computers), the application processing delay can be a significant component of end-to-end delay. Moghe called this the terminal QOS measure. Currently, each adaptive application has its native adaptation algorithm that operates independently of other applications, their adaptation algorithms, or scheduler. Moghe presented an analytical relationship between network feedback and the level of adaptation. This theoretical framework is useful in understanding the relationship and interaction between the adaptive application, end-system and network. The speaker concluded by saying that the notion of terminal QOS may be used to tune adaptation algorithms.

Next, Dorgham Sisalem [Sisalem,97], GMD Focus, presented an approach for dynamically adjusting the sending rate of applications to the congestion level observed in the network. The speaker discussed how senders could increase their sending rate during lightly loaded situations and reduce it during overload periods. Sisalem presented results which illustrated the efficiency of a direct adjustment algorithm responding to fluctuations in available bandwidth while maintaining low loss rates. Currently, the work does not address the issue of fairness nor interaction between adaptive traffic, e.g., TCP. Sisalem concluded by indicating that adaptive schemes can suffer from fairness problems potentially causing starvation of reactive flows such as TCP.

The third paper, by Max Ott [Ott,97], NEC, reported on adaptive QOS in multimedia systems. Ott made the point that most QOS architectures present QOS-aware APIs to the applications. This is either achieved by adding QOS parameters to standard system calls, or by raising the system abstraction to a higher level, filling the gap with what is often referred to as middleware. The speaker reflected that in either case there seems to be a serious desire to draw a strict line between "us", the QOS provider, and "them" the elusive application. Ott argued that it seemed natural to define an architecture that allows the introduction of QOS at any level: from the CPU and network resources to the user's "satisfaction". He reported on an architecture that has evolved over the past few years at NEC that has introduced the concept of QOS adaptation at each level of the architecture. QOS is specified by contracts which are established between clients and service providers using a single generic programmable API.

Oh [Oh,97], Osaka University, presented the next paper on a dynamic QOS adaptation mechanism for a networked virtual reality system. The motivation behind the work is to maintain acceptable user presentation quality when resources fluctuate in a networked virtual reality system. Oh introduced the notion of "importance of presence" which when applied to objects in a virtual reality system can be used to trade off available resources and objects. Importance of presence is based on the maximum visible distance and angle of incidence between the users and an object in the user's virtual space. The speaker discussed an adaptive algorithm that reduced the QOS of an object based on its importance of presence indication. At the end of his presentation the speaker demonstrated the technique using a videotape of the networked virtual reality system using adaptation based on the importance of presence.

In the final talk of this session Dan Revel [Revel,97], Oregon Graduate Institute of Science and Technology, discussed predictable file access latency for multimedia. The speaker asserted that multimedia applications are sensitive to IO latency and jitter when accessing data from secondary storage. To address this challenge, Revel introduced the concept of transparent adaptive prefetching which uses software feedback to provide multimedia applications with file system QOS guarantees: predictable low-latency access to data on secondary storage. The research at Oregon Graduate Institute is strongly focused on adaptive software

feedback algorithms for Internet video, mobile systems and adaptive QOS access to storage. Currently, the QOS interface to transparent adaptive prefetching allows applications to express adaptive needs in a vocabulary that is meaningful to them.

Panel Discussion

Klara Nahrstedt commented that QOS adaptation techniques, while complex in nature help to efficiently share resources in the end-systems and network for a class of applications that can accommodate varying resource availability. Nahrstedt followed up by asking the panel whether they now considered QOS adaptation as a replacement for reservation? A wide variety of opinions were articulated. Dorgham Sisalem used TCP as an example to argue that TCP's congestion adaptation mechanisms did very well without any explicit call setup and resource reservation; therefore why is reservation needed?

Max Ott argued that if strong QOS guarantees were required (e.g., timing guarantees) then reservation was unavoidable. Nahrstedt said that QOS adaptation filled the middle ground providing a poor person's QOS guarantee. Pratyush Moghe thought that many continuous media applications could operate at a minimum level guarantee and use QOS adaptation techniques to achieve better quality when additional resources became available. Such a hybrid approach grew out of the consensus that many applications can not degrade below a certain level.

5. Workshop Panels

The workshop included two panels. The first panel [Schmidt,97] looked at the emerging field of QOS in distributed object computing environments. The second panel [Schulzrinne,97] highlighted some of the concerns researchers have about a reservation driven Internet.

5.1 QOS for Distributed Object Computing Middleware - Fact or Fiction?

Chair and Organizer: *Douglas C. Schmidt, Washington University*

Panelists: *Max Ott, NEC, Guru Parulkar, Washington U., Rolf Stadler, Columbia U., Andreas Vogel, Visigenic*

The panel discussion started with a statement by the Chair that nobody uses a distributed middleware with QOS support, therefore, do we need such a layer? If yes, what does QOS-based middleware look like? Is QOS-based distributed middleware a fact or a fiction? Currently, it is a fiction since no such product or public domain platform exists. It will become a fact when researchers and developer are using QOS-based middleware as they do C++ or JAVA. Clearly, we need to ask ourselves what does such a middleware bring? Yet another layer?

The panel argued that distributed object computing was essential for architecturing complex software systems, ease of programmability and interworking across heterogeneous systems. A number of research groups are building real-time ORBs and exposing QOS at the API for user programmability. Freeware is becoming available (e.g., Glenford Mapp, Olivetti & Oracle Research Lab, announced that ORL had put their ORB 2.0 on the web). Andreas Vogel stressed that real-time CORBA products will become a reality when there is demand from the customer. When the demand transpires the industry would react positively.

A member of the audience asked why CORBA has become the common language, ORB of choice and subject of this panel? Guru Parulkar pointed out that CORBA is an existing standard and has been widely implemented by multiple vendors. It has a number of deficiencies, he asserted, many ORB implementations are not highly inefficient and support of QOS is missing. However, Guru Parulkar felt optimistic that these issues could be amended and were not "show stoppers".

Each of the panelists provided a list of open issues which they believed must be addressed before QOS-based middleware became a reality:

- Schmidt maintained that the ORB needed a real-time capability and middleware, such as CORBA, is being modified to incorporate such a feature;
- Ott argued that any emerging platform needs to take seriously the user perspective to middleware to ease programmability and simplify QOS abstractions;
- Parulkar stressed that we need to fix CORBA to support QOS in the end-system and at the network access points in order to provide service guarantees at the I/O level, appropriate packet and process scheduling was needed;
- Stadler remained upbeat about the use of the technology in the network arguing for distributed interactive resource controllers with global control to deliver services with a minimal set of guarantees and, in case of higher quality, they should support adaptive behaviors;

- Vogel reflected that ORB vendors (e.g., Visibroker) will eventually incorporate new CORBA services and QOS bindings for multimedia streams and control when the market demands it. He predicted that it will happen soon.

Toward the end of the session questions of whether ORBs should explicitly support multimedia streaming were raised. There was clear disagreement about this issue. The question is whether the ORB is used to set up streams and then moves out of the way or, conversely, it can explicitly support isochronous communications via RPC. Some panelists stated that this should not be part of CORBA. Others argued that for continuity the ORB should support streaming. Andreas Vogel mentioned that OMG would not support streaming through the ORB.

5.1 Reservations about Reservations

Chair and Organizer: *Henning Schulzrinne, Columbia University*

Panelists: *Fred Baker, CISCO, Andrew T. Campbell, Columbia U., Jon Crowcroft, UCL, Roch Guerin, IBM and Dilip Kandlur, IBM*

The panel discussed the current state and future developments of QOS support in an Internet and the impact of a reservation driven network. Each member of the panel started off by presenting a short position statement of their vision of the rollout of QOS in the Internet. This was followed by heated debate on concerns about reservations.

Fred Baker, Cisco, stated that QOS routing, line protocols and queueing management can improve QOS. He emphasized that for queuing, congestion management algorithms, such as weighted fair queuing, serve as good tools for low speed links when there is a limited number of flows. However, on the average OC-3 link where there are hundreds and thousands of simultaneous flows (e.g., in a backbone network), statistical approaches like random early detection prove to be very useful. In this case, the binding is based on IP precedence. This precedence level can be set either by a traffic originator or via administrative controls in the routers. However, like random early detection, being a FIFO queueing algorithm, is not predictable and is not deterministic since it depends on host behavior. Moreover, it is also dependent on bulk of traffic being TCP. Besides these approaches, Baker suggested that RSVP is a reasonable answer to QOS-enhanced traffic, particularly for edge networks and large flows.

Next, Roch Guerin, IBM, asked what the driving force behind reservations was. He mentioned that there is currently no application that is so critical that it can only function by making reservations. Alternatively, there are so many of these applications that it is impossible to define a generic reservation framework to satisfy them all. Guerin argued that the main drivers for reservations are the economic and contractual factors. As the Internet is moving towards a commercial network, people would like to know what they are paying for – and ISP's also need a pricing model. This requires enforceable and observable service contracts. These contracts should be simple and deterministic to avoid adding additional complexity to the infrastructure. Guerin also pointed out that there are many conflicting forces to economic factors such as the cost of resources and their exploitation. In particular, cheap resources lead to simple signaling, where value is added at end-systems. In contrast, if resources remain expensive, a network will need complex signaling, thus leading to providers and equipment vendors adding most of the value.

Dilip Kandlur, IBM, expressed some reservations about RSVP scalability. He cautioned that RSVP is able to scale to a large number of receivers in a session, but not a large number of unicast sessions. He provided possible solutions to achieving the latter, which included aggregation of sessions, flow state management and path management.

Henning Schulzrinne, Columbia University, discussed several points: the need for a basic service with call admission control; at what point is reservation the best option; flow aggregation and RSVP issues. The existing consumer ISP model is based on a multiplexing model that supports 200-300 concurrent users with 10-15 customers sharing an ISP line. With Internet telephony, radio-like services and content pushing such a model is not sustainable. This implies that volume-based charging is required rather than reservations. Schulzrinne mentioned that reservations are still necessary for guaranteeing special purpose QOS, for services that cannot tolerate disruption. RSVP as a reservation protocol for the Internet introduces unnecessary complications such as flow merging, delay guarantees and receiver diversity.

The fourth presentation in this panel session focused on the use of a simple approach to QOS provisioning. Jon Crowcroft had arranged to come in live over the Internet. However, moments before the panel was to begin, UCL experienced a complete brown out and Jon Crowcroft was source squenched! Andrew Campbell stepped in to present the main points of Crowcroft's

position. Crowcroft suggested that were three simple choices – each with their pros and cons:

- *over-engineering selected paths*, where a particular route could be over provisioned. That route might be shared by multiple networks as well as users, hosts or applications,
- *subscription for selected terminals or addresses*, where a virtual private Internet offering improved QOS can be created by assigning resources based on address prefixes or IP network numbers; and
- *on demand service*, where a virtual circuit is set up using a signaling protocol, call admission control, traffic policing, traffic accounting, and service discrimination through scheduling.

Crowcroft preferred the first choice of over-engineering selected paths but cautioned that it might be difficult to support new services due to such an over-simplistic view of resources. However, his contention was that the cost of providing new services outweighs the benefit. The balance, he suggested, should be redressed toward simpler end of the QOS provisioning spectrum: that is, over-engineered selected paths.

Andrew Campbell presented his own reservations about reservations, pointing out that both ATM and RSVP signaling are too heavy weight and complex. He also revealed his concerns about scalability and stability of RSVP as a widely deployed signaling system for the Internet. Out-of-band signaling systems, he asserted, need to be sized, engineered and their stability thoroughly investigated.

The presentations sparked off an interesting debate on reservations about reservations. The panelists seem to agree that reservations are at times inappropriate. Guerin, who is less comfortable with the need for reservations, felt that reservations would only add complexity in the end.

6. QOS : Where are we ?

***IWQOS'97 Invited Workshop Paper**: by Ralf Steinmetz and Lars Wolf, Darmstadt University*

Lars Wolf [Steinmetz,97] presented the invited workshop paper on topic of "QOS: Where are we?". He began with an overview of terminology, issues and trends in the provisioning of QOS concentrating on QOS principles and the architectural issues addressed by a number of research groups and standards bodies (e.g., IETF's int-serv work). Wolf reflected that over the past several years, QOS has evolved as a major field of research to support new applications such as real-time distributed multimedia applications, which very often are based on networked computer systems. These applications require time-dependent data processing and place huge processing demands on distributed computer systems.

The goal of QOS architecture is to ensure the overall presentation of multimedia data with respect to the end-to-end QOS requirements of applications. Because QOS is end-to-end, he argued, end-systems, servers, networks, system software and applications must handle the data accordingly. QOS model generally have a number of components/viewpoints (e.g., user, application, system and network QOS) which address various approaches to QOS provisioning, both pessimistically (e.g., worst-case assumptions based on peak rate resource allocation) and/or optimistically (e.g., resource reservation and adaptive mechanisms). Among the optimistic approaches, resource reservation is preferred since adaptive methods like media scaling and filtering cannot offer hard QOS.

Wolf proceeded to describe the fundamental steps in provisioning QOS in the end-system and network based on signaling, resource reservation and scheduling mechanisms. These steps can be divided into the QOS negotiation phase (including QOS specification, capacity testing and QOS calculation, resource reservation) and the data transmission phase, where the negotiated QOS is enforced by appropriate resource scheduling mechanisms. Several resource management components need to interact to provide end-to-end assurances : applications, QOS translators, admission controllers, resource schedulers and resource monitors. Resource reservation protocols serve as a means to transfer information about reservations and to participate in the negotiation of QOS values.

7. Closing Remarks

Andrew Campbell provided the closing remarks. He announced that the next workshop would be held in Napa Valley, California in May 1998 and would be hosted by Ed Knightly from Rice University and Rich Friedrich from HP Labs.

The theme of IWQOS'97 - building QOS into distributed systems - begs the question: in today's world is QOS a fact or a fiction? For those that attended the workshop, with its excellent technical sessions and panels, it is indeed, a fact. During the three day meeting, participants had heard reports about actual implementation results and more speculative work. While this looks very promising, some practitioners have reservations about QOS research. Perhaps, this was best typified during the panel on reservations about

reservations. The catchy title of the panel captures that feeling of concern. For example, the entire reservations panel agreed that it was very unlikely that a single approach (e.g., RSVP) would fit all applications, with different trade offs between complexity, level of guarantees and scaling issues.

Returning to the theme of whether QOS was a fact or fiction, Campbell asserted that QOS is fiction in the sense that it is not a concrete fixture in our lives, e.g., no network exists today that lets the user configure QOS on-demand. So how long do we have to wait until fiction becomes fact? Campbell concluded by telling the audience to "stay tuned" for Napa '98.

For details on IWQOS'98, Napa Valley, see:

- `http://www-ece.rice.edu/conf/iwqos98/`

8. Acknowledgements

We would like to thank the Program Committee for all their efforts to make this a successful workshop, the Center for Telecommunications Research at Columbia University for hosting this year's workshop and to the local Organization Committee (particularly, Koon-Seng Lim and Oguz Angin) for all their hard work. Finally, any errors or omissions are ours and for these, our apologies in advance.

ORGANIZED BY

Center for Telecommunications Research, Columbia University

IN ASSOCIATION WITH

ACM SIGCOMM
IEEE Communications Society
IEEE Signal Processing Society
IFIP Joint International Conference on Open

Distributed Processing and Distributed Platforms

IWQOS'97 WORKSHOP PATRON

Hewlett-Packard Company

PROGRAM CO-CHAIRS

Andrew T. Campbell, Columbia University
Klara Nahrstedt, University of Illinois at Urbana Champaign

PROGRAM COMMITTE

Gordon Blair, University of Lancaster
Simon Crosby, Cambridge University
Jon Crowcroft, UCL
Hermann de Meer, University of Hamburg
Jan de Meer, GMD Fokus
Alexandros Eleftheriadis, Columbia University
Richard Friedrich, Hewlett-Packard Company
Michael Fry, University of Technology
Andrew Herbert, APM Ltd., UK
Jean-Peirre Hubaux, EPFL
Kevin Jeffay, UNC at Chapel Hill
Brigitte Kerhervé, U. du Q. à Montréal,
S. Keshav, Cornell University
Glenford Mapp, Olivetti Research Ltd
Steve McCanne, UCB
Mahmoud Nagshineh, IBM
Elie Najm, ENST
Max Ott, C&C Research, NEC
Guru Parulkar, Washington University
Jerry Rolia, Carleton University
Cormac Sreenan, Lucent Technologies
Chris Sluman, OpenIT Ltd
Hideyuki Tokuda, Keio University
James VanLoo, Sun Microsystems
Andreas Vogel, Visigenics Corp
Andrew Watson, OMG
Hartmut Wittig, Multimedia Software
Lixia Zhang, UCLA
Hui Zhang, Carnegie Mellon University
Martina Zitterbart, TU Braunschweig

LOCAL ORGANISATION COMMITTEE

Oguz Angin , Laitee Cheok and Koonseng Lim

9. References

[Almasberger, 97] Almasberger A., Sasyan S. and Wright S., "Quality of Service (QoS) in Communication APIs", Proc. 5th International Workshop on Quality of Service (IWQOS'97), Columbia University, New York, USA, Pages 237-248.

[Badrinath, 97] Badrinath B.R. and Talukdar A.K., "IPv6 + Mobile-IP + MRSVP = Internet Cellular Phone ?", Proc. 5th International Workshop on Quality of Service (IWQOS'97), Columbia University, New York, USA, Pages 49-52.

[Becker, 97] Becker D., "Comprehensive Queueing Analysis for a Partial Buffer System with Discrete Markovian Arrival Processes", Proc. 5th International Workshop on Quality of Service (IWQOS'97), Columbia University, New York, USA, Pages 89-100.

[Blair, 97] Blair. G.S., Davies N., Friday A. and Wade S.P., "Quality of Service Support in a Mobile Environment: An Approach Based on Tuple Spaces", Proc. 5th International Workshop on Quality of Service (IWQOS'97), Columbia University, New York, USA, Pages 37-48.

[Bless, 97] Bless R., Jacob M. and Schmidt C., "Service-Tailored QoS Management in High Performance Networks", Proc. 5th International Workshop on Quality of Service (IWQOS'97), Columbia University, New York, USA, Pages 179-190.

[Borowsky, 97] Borowsky E., Golding R., Merchant A., Scheier L., Shriver E., Spasojevic M., Wilkes J., "Using attribute-managed storage to achieve QoS", Proc. 5th International Workshop on Quality of Service (IWQOS'97), Columbia University, New York, USA, Pages 199-202.

[Campbell, 97] Campbell A.T., "QoS Challenges for Next Generation Mobile Middleware", Proc. 5th International Workshop on Quality of Service (IWQOS'97), Columbia University, New York, USA, Pages 53-56.

[Castellanos, 97] Castellanos, J.G. and Naghshineh M., "Link Error Impact on MPEG Video over Wireless Broadband Networks", Proc. 5th International Workshop on Quality of Service (IWQOS'97), Columbia University, New York, USA, Pages 57-60.

[Conrad, 97] Conrad P., Amer P., Golden E., Iren S., Marasli R. and Caro A., "Transport QoS over Unreliable Networks: No Guarantees, No Free Lunch", Proc. 5th International Workshop on Quality of Service (IWQOS'97), Columbia University, New York, USA, Pages 315-318.

[Corson, 97] Corson M.S., "Issues in Supporting Quality of Service in Mobile Ad Hoc Networks", Proc. 5th International Workshop on Quality of Service (IWQOS'97), Columbia University, New York, USA.

[de Meer, 97] de Meer J., "On the Specification of End-to-End QoS Control", Proc. 5th International Workshop on Quality of Service (IWQOS'97), Columbia University, New York, USA, Pages 347-358.

[Fischer, 97] Fischer S., Salem M.-V.O.M. and Bochmann G.v., "Application Design for Cooperative QoS Management", Proc. 5th International Workshop on Quality of Service (IWQOS'97), Columbia University, New York, USA, Pages 191-194.

[Fukuda, 97] Fukuda K., Wakamiya N., Murata M., and Miyahara H., "QoS Mapping between User's Preference and Bandwidth Control for Video Transport", Proc. 5th International Workshop on Quality of Service (IWQOS'97), Columbia University, New York, USA, Pages 291-302.

[Hafid, 97] Hafid A., "Providing a Scalable Video-on-Demand System using Future Reservation of Resources and Multicast Communications", Proc. 5th International Workshop on Quality of Service (IWQOS'97), Columbia University, New York, USA, Pages 277-288.

[Huard, 97] Huard J.F. and Lazar A.A., "On End-to-End QoS Mapping", Proc. 5th

International Workshop on Quality of Service (IWQOS'97), Columbia University, New York, USA, Pages 303-314.

[Issarny, 97] Issarny V., Bidan C., Leleu F. and Saridakis T., "Towards Specifying QoS-Enabling Software Architectures", Proc. 5th International Workshop on Quality of Service (IWQOS'97), Columbia University, New York, USA, Pages 363-366.

[Jacobs, 97] Jacobs S., Eleftheriadis A., "Adaptive Video Applications for Non-QoS Networks", Proc. 5th International Workshop on Quality of Service (IWQOS'97), Columbia University, New York, USA, Pages 161-165.

[Jha, 97] Jha S.K., Wright P.A. and Fry M., "Playout Management of Interactive Video - an Adaptive Approach", Proc. 5th International Workshop on Quality of Service (IWQOS'97), Columbia University, New York, USA, Pages 145-156.

[Kim, 97] Kim K. and Nahrstedt K.," QoS Translation and Admission Control for MPEG Video", Proc. 5th International Workshop on Quality of Service (IWQOS'97), Columbia University, New York, USA, Pages 359-362.

[Knoche, 97] Knoche H. and de Meer H., "Quantitative QoS-Mapping: A Unifying Approach", Proc. 5th International Workshop on Quality of Service (IWQOS'97), Columbia University, New York, USA, Pages 347-358.

[Koodli, 97] Koodli R. and Krishna C.M., "Supporting Multiple-tier QoS in a Video Bridging Application", Proc. 5th International Workshop on Quality of Service (IWQOS'97), Columbia University, New York, USA, Pages 133-144.

[Lakshman, 97] Lakshman K., Yavaykar R. and Finkel R., "Integrated CPU and Network I/O QoS Management in an Endsystem",Proc. 5th International Workshop on Quality of Service (IWQOS'97), Columbia University, New York, USA,

Pages167-178.

[Lazar, 97] Lazar A.A., "Programming Telecommunication Networks", Proc. 5th International Workshop on Quality of Service (IWQOS'97), Columbia University, New York, USA, Pages 3-24.

[Lee, 97] Lee H., "Guaranteeing Multiple Cell Loss Classes in Shared ATM Output Buffer", Proc. 5th International Workshop on Quality of Service (IWQOS'97), Columbia University, New York, USA, Pages 77-88.

[Linnhoff-Popien, 97] Linnhoff-Popien C. and Thissen D., "Integrating QoS Restrictions into the Process of Service Selection", Proc. 5th International Workshop on Quality of Service (IWQOS'97), Columbia University, New York, USA, Pages 225-236.

[Lu, 97] Lu S., Lee K.-W. and Bhargavan V., "Adaptive Service in Mobile Computing Environments", Proc. 5th International Workshop on Quality of Service (IWQOS'97), Columbia University, New York, USA, Pages 25-36.

[Ma, 97] Ma Q. and Steenkiste P., "Quality of Service Routing for Traffic with Performance Guarantees", Proc. 5th International Workshop on Quality of Service (IWQOS'97), Columbia University, New York, USA, Pages 115-126.

[Mapp, 97] Mapp G. and Hodges S., "QoS-Based Transport", Proc. 5th International Workshop on Quality of Service (IWQOS'97), Columbia University, New York, USA, Pages 319-322.

[Maryni, 97] Maryni P. and Pacifici G., "Real-Time Estimation of the Link Capacity in Multimedia Networks", Proc. 5th International Workshop on Quality of Service (IWQOS'97), Columbia University, New York, USA, Pages 101-112.

[Moghe, 97] Moghe P. and Kalavade A., "Terminal QoS of Adaptive Applications and its Analytical Computation", Proc. 5th International Workshop on Quality of Service (IWQOS'97), Columbia University, New York, USA, Pages 369-380.

[Nishio, 97] Nishio N. and Tokuda H., "Simplified Method for Session Coordination Using Multi-level QOS Specification and Translation", Proc. 5th International Workshop on Quality of Service (IWQOS'97), Columbia University, New York, USA, Pages 335-346.

[Oechslin, 97] Oechslin O., "Worst Case Arrivals of Leaky Bucket Constrained Sources: The Myth of the On-Off Source", Proc. 5th International Workshop on Quality of Service (IWQOS'97), Columbia University, New York, USA, Pages 67-76.

[Oh, 97] Oh. S., Sugano H., Shimojo S., Miyahara H., Fujikawa K., Matsuura T., and Arikawa M., "A Dynamic QoS Adaptation for Networked Virtual Reality', Proc. 5th International Workshop on Quality of Service (IWQOS'97), Columbia University, New York, USA, Pages 397-400.

[Ott, 97] Ott M., Michelitsch G., Reininger D. and Welling G., "Adaptive QoS and in Multimedia Systems", Proc. 5th International Workshop on Quality of Service (IWQOS'97), Columbia University, New York, USA, Pages 393-396.

[Pope, 97] Pope S. and Webster P., "QoS Support for Mobile Computing", Proc. 5th International Workshop on Quality of Service (IWQOS'97), Columbia University, New York, USA, Pages 61-64.

[Reed, 97] Reed D.A. and Turner K.J., "Support Components for Quality of Service in Distributed Environments: Monitoring Service", Proc. 5th International Workshop on Quality of Service (IWQOS'97), Columbia University, New York, USA, Pages 255-258.

[Revel, 97] Revel D., Cowan C., McNamee D., Pu C. and Walpole J., "Predictable File Access Latency for Multimedia", Proc. 5th International Workshop on Quality of Service (IWQOS'97), Columbia University, New York, USA, Pages 401-404.

[Schelen, 97] Schelen O. and Pink S., "Sharing Resources through Advance Reservation Agents", Proc. 5th International Workshop on Quality of Service (IWQOS'97), Columbia University, New York, USA, Pages 265-276.

[Schmidt, 97] Schmidt D., Ott M., Parulkar G., Stadler R. and Vogel A., "QOS for Distributed Object Computing Middleware", Proc. 5th International Workshop on Quality of Service (IWQOS'97), Columbia University, New York, USA, Pages 205-207.

[Schulzrinne, 97] Schulzrinne H., Baker F., Campbell A.T. and Kandur G., "Reservations about Reservations", Proc. 5th International Workshop on Quality of Service (IWQOS'97), Columbia University, New York, USA, Pages 325-331.

[Smith, 97] Smith S. and Grace A., "A QoS Configuration System for Distributed Applications", Proc. 5th International Workshop on Quality of Service (IWQOS'97), Columbia University, New York, USA, Pages 259-262.

[Steinmetz, 97] Steinmetz R. and Wolf L., "Quality of Service: Where are we?", Proc. 5th International Workshop on Quality of Service (IWQOS'97), Columbia University, New York, USA, Pages 211-222.

[Verma, 97] Verma S., Pankaj R.K. and Leon-Garcia A., "QoS Based Multicast Routing for Multimedia Communications", Proc. 5th International Workshop on Quality of Service (IWQOS'97), Columbia University, New York, USA, Pages 127-130.

[Waldegg, 97] Waldegg D.B., "A Temporal QoS based CPU Scheduling Model for Multimedia Applications in General Purpose Operating Systems', Proc. 5th International Workshop on Quality of Service (IWQOS'97), Columbia University, New York, USA, Pages 157-160.

[Zinky, 97] Zinky J.A. and Bakken D.E., "Managing Systemic Meta-Data for Creating QoS-Adaptive CORBA Applications", Proc. 5th International Workshop on Quality of Service (IWQOS'97), Columbia University, New York, USA, Pages 251-254.