Telephony Over IP: A QoS Measurement-Based End to End Control Algorithm and a Queue Schedulers Comparison

Luigi Alcuri
Department of Electric Engineering – University of Palermo
Viale delle Scienze 9, 90128 Palermo
Italy

Francesco Saitta
Department of Electric Engineering – University of Palermo
Viale delle Scienze 9, 90128 Palermo
Italy

ABSTRACT

This paper presents a method for admitting voice calls in Telephony over IP (ToIP) scenarios. This method, called QoS-Weighted CAC, aims to guarantee Quality of Service to telephony applications. We use a measurement-based call admission control algorithm, which detects network congested links through a feedback on overall link utilization. This feedback is based on the measures of packet delivery latencies related to voice over IP connections at the edges of the transport network. In this way we introduce a close loop control method, which is able to auto-adapt the quality margin on the basis of network load and specific service level requirements. Moreover we evaluate the difference in performance achieved by different Queue management configurations to guarantee Quality of Service to telephony applications, in which our goal was to evaluate the weight of edge router queue configuration in complex and real-like telephony over IP scenario. We want to compare many well-know queue scheduling algorithms, such as SFQ, WRR, RR, WIRR, and Priority. This comparison aims to locate queue schedulers in a more general control scheme context where different elements such as DiffServ marking and Admission control algorithms contribute to the overall Quality of Service required by real-time voice conversations. By means of software simulations we want to compare this solution with other call admission methods already described in scientific literature in order to locate this proposed method in a more general control scheme context. On the basis of the results we try to evidence the possible advantages of this QoS-Weighted solution in comparison with other similar CAC solutions (in particular Measured Sum, Bandwidth Equivalent with Hoeffding Bounds, and Simple Measure CAC), on the planes of complexity, stability, management, tune-ability to service level requirements, and compatibility with actual network implementation.

Keywords: VoIP, Admission Control, Quality of Service, DiffServ

1. INTRODUCTION

A lot of work has been done studying the performance problems affecting the development of telephony applications over IP (ToIP) networks [1, 2, 3, and 4].
An essential component for the success of ToIP consists in achieving the same Quality of Service (QoS) of PSTN and ISDN networks with as few changes as possible to the actual IP network implementation. For this reason we looked for a solution that would require very few changes in the existing backbone routing network. In order to achieve this goal, we retain that an end-to-end measurement based access control strategy should be sufficient for succeeding in guaranteeing QoS to voice calls. So we have experimented a solution that works without introducing new protocols or new network elements for monitoring and avoiding network congestion as proposed by other similar studies [6].
As our first step, we individuated the network parameters, which are indicative of QoS for phone calls over IP networks; they could be summarized as bandwidth, packet loss-rate, latency and jitter [7]. As our second step, we investigated the possible solutions for guaranteeing to these parameters values that are indicative of a good QoS without reducing excessively the efficiency of network resources. Finally, we tested and analyzed this new end-to-end measurement based Call Admission Control (CAC) for avoiding network congestion and consequent impairments on the QoS perceived by the end user. In order to estimate the validity of the solution presented, we confronted the new CAC algorithm with other different CAC already analyzed by scientific literature [8], such as Simple Sum, Measured Sum, and Equivalent Bandwidth using on Hoeffding Bounds.

As an addon to this work we tested and analyzed the impact of queue scheduler on the QoS perceived by the end user. The results partially achieved evidence a significant influence of queue scheduler on the performance of the network quality of service parameters, and the good behaviour of Weighted Interleaved Round Robin (WIRR) scheduler versus the others solution tested. At the end in order to give a significant contribute to the ToIP research we present a complete and configured ToIP architecture and its related performance.

2. DESCRIPTION OF THE TOIP SCENARIO

Service Provider structure

Today development of ToIP architecture is based on the distribution of Points of Presence POP on the territory. Each POP encloses the access interfaces for users of those services offered by Service Providers. In the case of telephony over IP, each POP is responsible for accepting new calls from the users and addressing them toward the correct POP destination with the minimum required service level. For this reason, in every POP there are: one or more Media Gateway Controllers (MGC), which control and allocate the resources to each call and set the connections with the other POP; Media Gateways (MG), which convert the voice communications from the PSTN/ISDN network into IP packets and collect statistics for each call received; and Internet Gateway (IG), with the functions of monitoring and marking the IP packets received by users on a local LAN, which already adopts Voice/IP devices inside its organization.

Description of the services

One of the main advantages of using ToIP is the possibility of implementing different telephony services, which have different costs in relation with the effective network resources requirement and utilization. Each SP has to characterize its services with well-defined Service Level Agreements (SLA).

We chose to implement several services with different specifications; each service is focused on a specific feature that could be considered relevant for the development of future ToIP applications.

The services proposed to the user are:

1) A voice over IP service, called "**Premium**". Mainly focused on guaranteeing latency requirements of service flows.
2) A voice over IP, called "**Basic**". Mainly focused on guaranteeing an upper bound for packet loss rate.
3) A File transfer service, called "**Data**", which could be used for some non real-time telephony applications such as fax, sms, e-mail or file transfer. The main requirement of this service is a minimum guaranteed bandwidth.

The Data service is intended to study the effects of aggressive UDP privileged streams versus connection-oriented stream with auto-control congestion avoidance mechanisms, such as TCP connections.

Queues were managed using a Random Early Drop (RED) algorithm, in order to reduce the mean queue length and to avoid burst losses during network overloads.

Queue schedulers

At this point we introduce the queue scheduler which were configured to manage two physical queues: one for UDP voice packets and one for Best Effort data traffic, in order to manage different queue priorities and minimum throughput requirements for services level agreements. The algorithms tested in this configuration were:

1) **(RR) Round Robin**: this algorithm is the most simple for the management of multiple queues, and represent a bound on the advantages of queue scheduling into the scenario considered.
2) **(WRR) Weighted Round Robin:** This algorithm gives a weight to the different queues in order to give proportional service time at each queue.
3) **(WIRR) Weighted Interleaved Round Robin:** This algorithm works like WRR but introduce an Interleaved behaviour, which permit the service of queues out of their service time, when there are no packet to serve in the actual serving queue.
4) **(PRI) Priority Queue:** In this case, it is possible to configure a service priority to the different queues according to their specific kind of service
5) **(SFQ) Stochastic Fairness Queuing:** This algorithm works like round robin introducing a stochastic perturbation in the service of the queues maintaining the fairness between the queues

Network topology

We will conduct several simulations by means of NS-2 Network Simulator software [21] in order to tune the solutions adopted with a likely real network scenario.

First, we implement the simple network topology illustrated in Fig. 1. So we can test the control algorithm in a classical configuration with a single bottleneck and two way traffic. The links between POPs and ER are configured to be a bottleneck for the connections between POPs. The Edge Router has also the function of managing the queue of each different class of service by means of dropping packets which are out of the admissible traffic profile for class of service. At this stage we adopted Weighted Interleaved Round Robin (WIRR) scheduler with two physical queues: one for UDP voice

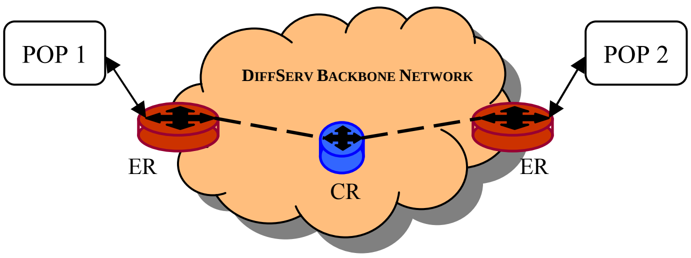

Fig. 1 Network topology of the simulated scenario for TOIP QoS controlled phone calls

packets and one for Best Effort data traffic, in order to manage different queue priorities and minimum throughput requirements for each service.
Then we complicate this simple network configuration introducing more POPs and MPLS-capable router in the backbone network in order to verify the scalability of the different CAC schemes.
For each POP, in order to study the behaviour of the control algorithm under several traffic load conditions, we generate data transport FTP connections towards the other POP, Poisson traffic for voice calls, and Pareto on/off sources which produce long range dependent traffic. A new request to set a voice call is modelled by a Poisson arrival process with lifetimes taken from a lognormal distribution, which changes on the basis of the offered load for each particular simulation.

3. CONFIGURATION AND CALL ACCESS CONTROL

Differentiated Services code-point assignment

Obviously, if we want offer and guaranteeing different services to the users, we need a way for differentiating the packets stream, and so we configured the backbone network to support DiffServ [12].
Moreover we adopted a token bucket policer for marking the packets before they reached the backbone network. Each service has a minimum bandwidth virtually reserved, which could be used by other service like out-profile premium or basic packets, if not used by the assigned service.
In more details, the in-profile packets of Premium and Basic service were marked as AF12 and AF11 respectively, while the out-profile packets of both services were marked as AF13, like described in the following table.

Table I
Associations of Services to DiffServ Class

DiffServ Class	Drop Precedence level		
	Gold (1)	Silver (2)	Bronze (3)
EF	UDP Control Packet		
AF1	Basic in-profile	Premium in-profile	Out-profile
AF2	Not used	Not used	Not used
AF3	Not used	Not used	Not used
AF4	Not used	Not used	Not used
Best Effort	Data traffic		

QoS-Weighted Bandwidth CAC algorithm

The call admission was managed by a self adaptive control mechanism presented in Fig. 2, which pointed to reach the following aims:

1) To guarantee that every service was not affected by the traffic generated by other service.
2) To guarantee that the final QoS of voice calls was better than a minimum level, which is determined by the following parameters: latency, jitter, packet loss rate and maximum burst loss.
3) To guarantee that the throughput of the bottleneck was very close to the theoretical maximum of the link.

In order to reach this goal a measured CAC needs some information on the congestion state of the backbone network. Because we wanted the advantage of a solution independent from the backbone network manager, we chose to collect some statistics for each service at the POP receiver at the end of the network and transmit them periodically to the POP sender at the other end by UDP control packets.
In this scheme the MGC of the destination POP collects the following information by means of MG interfaces for each call received of both voice services:
1) Number of bytes received B
2) Number of packets received N_P
3) Time of transmission of each packet l_P
These statistics on single call connections could be easily collected by using specific functions already implemented in Megaco [19] and RTP [20] protocols.
At every T interval of time, the MGC sends to the others ones from which is receiving voice calls an UDP packet with the following indexes:
Average Latency:

$$\frac{\sum_{i=0}^{N_P} l_i}{N_P} = AL$$

QoS-Weighted Bandwidth:

$$\frac{B}{T - \alpha * AL} = QWB, \qquad 0 \le \alpha < \frac{T}{AL}$$

Where T is chosen as un upper bound for round trip time of real-time applications, in particular we consider during this study equals to 500 ms. The introduction of the QoSWeighted Bandwidth is fundamental for having a simple index, which is able to summarize the effective level of utilization of network resources during congestion period. When the network is experiencing a

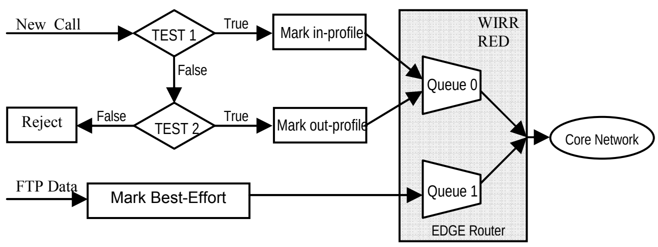

Fig. 2: Block-diagram, which illustrates the admitting and marking of new calls.

period of heavy load, the packets have to wait more time in the router queues before being served, so the AL increase and in a similar way the QWB. The parameter α was introduced for compensating the distance in propagation time between the two POPs, and for managing the trade-off between the importance of QoS for voice calls and efficiency of network utilization. In fact, higher values mean QWB values higher than real Bandwidth utilization, so the congestion avoidance mechanism of CAC method start to work at lower bandwidth utilization thresholds increasing the probability of obtaining satisfactory QoS performance. In more details α is the weight of QoS, summed up by mean average packet latency time, in the evaluation of QWB index.

When a POP receives a new call for service i, it first check with TEST 1, if there is available bandwidth for the requested service using the QWB index, which enclose information on the congestion state of the network by using the estimate of AL and the weight assigned to QoS through the α index.

TEST 1:

$$(QWB_i < BW_{Xi}) \qquad (1)$$

where BWxi is the max Bandwidth virtually reserved to the service i, with i equal to P or B according to the kind of service Premium or Basic to which the new call belongs.

If TEST1 fails it means that the new call does not have a reserved bandwidth in its service category (i.e. to much call already active or bad latency performance suffered by the service). However it could utilize the available bandwidth of other services like an out-profile call, if there is voice reserved bandwidth and no network congestion. So in this case it will be checked the possibility of accepting the new call on the basis of the following

TEST2:

$$(QWB_P + QWB_B) < (BW_{XP} + BW_{XB}) \qquad (2)$$

Note that we do not need to know the bandwidth occupancy of the new call because we can chose the margin of bandwidth greater or equal to the maximum call bit rate of the service.

Others CAC algorithms

On the same network scenario, we compared other well know CAC algorithms to this new proposed QWB CAC, in order to evidence the difference and to locate the new algorithm in a more general literature context. We take in exam the following algorithm for admission control:

1) NO CAC: **No Admission** control, all the incoming calls are admitted without any check. This case was considered to get a bound on the effects of admission control in the network scenario considered

2) SM CAC: **Simple Measured** admission control algorithm, this algorithm simply ensures that the sum of requested resources does not exceed link capacity. Let ν be the sum of reserved rates, μ the link bandwidth reserved to the service, and r the rate requested by a new call flow. This algorithm accepts the new flow if the following check succeeds:

$$\nu + r < \mu \qquad (3)$$

3) MS CAC: **Measured Sum** admission control, this algorithm uses measurement to estimate the load of existing traffic. This algorithm admits the new flow if the following test succeeds:

$$\nu + r < \alpha \cdot \mu \qquad (4)$$

where α is a parameter used to set the utilization target. In our experiments we consider three different values of α in order to evaluate the tune-ability to service level requirements and sensibility to management parameters

4) EB CAC: **Equivalent Bandwidth** admission control. The equivalent bandwidth[10] of a set of flows is defined as the bandwidth such that the stationary bandwidth requirement of the set of flows exceeds this value with probability at most α. The measurement-based equivalent bandwidth based on Hoeffding bounds (C_H) assuming peak rate (p) policing of n flows is given by:

$$C_H(\hat{\nu};\{p_i\}1 \le i \le n;\alpha) = \hat{\nu} + \sqrt{\frac{\ln(1/\alpha)\sum_i^n (p_i)^2}{2}}$$

where ν is the measured average arrival rate of existing traffic and α is the probability that arrival rate exceeds the link capacity.

The admission control in this case check whenever a new flow requests admission if the following test is verified:

$$C_H + p \le \mu \qquad (5)$$

Upon admission of a new flow, the load estimate is increased using $\hat{\nu}' = \hat{\nu} + p$. If a flow's peak rate is unknown, according to [10] the peak rate is derived by the token bucket parameter using the equation: $p = r + b/U$ where U is a user-defined averaging period.

4. Results and Comments

Call Admission Algorithms

In order to evaluate the performance of the proposed CAC method, we present the result of simulations with several configurations of QWB, MS and EB CAC algorithm. This means that each configuration had set a different target of bandwidth utilization and service level quality to reach. Obviously, the higher is the utilization target the less is the guarantee of satisfying voice Service Level Specifications with a resulting poor QoS. Moreover different values of user-defined alpha parameter are

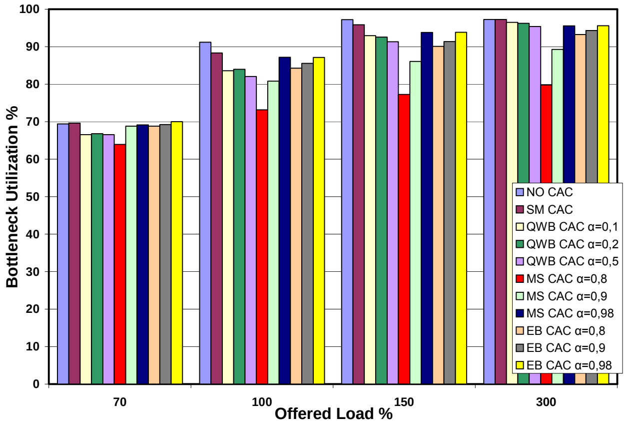

Fig. 3 Bottleneck utilization of virtually voice-only reserved bandwidth versus voice offered load

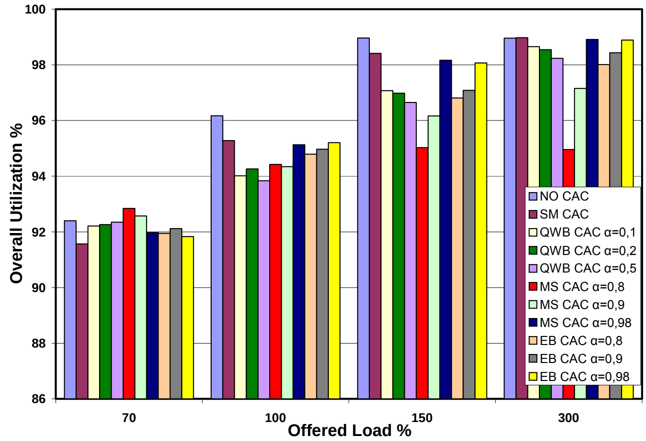

Fig. 4 Data plus Voice services bandwidth utilization versus voice offered load

useful to evaluate the sensitivity of the CAC algorithm to control parameters related to target utilization and QoS performance.

The bar-graph in Fig.3 collect the result achieved during the simulation by all the configuration of admission control tested relatively to the utilization of the bottleneck link reserved to the service of voice flow versus different voice offered load. Obviously we are interested on the effects of admission control algorithm on overloaded link, but light load condition are useful to evaluate the loss of admission control due to parameter configuration.

In Fig. 3 it is possible to note the effect of α user-defined tuning parameter against bandwidth utilization. The QWB and EB CAC presents a more fine control on block probability than MS CAC, this is a relevant advantage in the management of network resources.

Fig. 4 evidence the overall bottleneck link utilization including the Data best effort traffic to Voice flows, in order to evidence the behaviour for call admission control versus different kind of connections characterized by different service requirements.

Fig. 5-6 plot on a log scale packet loss rate results obtained by the different CAC algorithms in Premium and Basic services; the results obtained by QWB algorithm are able to permit an understanding voice call conversation between two end-users. We have configured the Basic voice service as a reliable connection, which is characterized by a reduced packet loss rate. In this configuration all the CAC were able to obtain a better performance in loss rate for basic service than premium service until 90% of bottleneck utilization. However in general QWB admission control was able to get the best behaviour both for basic and premium service. It is interesting to note the behaviour of MS CAC curves for premium service, that at high bottleneck utilization reduce packet loss rate due to a more aggressive control on call admission.

Fig. 7-8 show the mean time spent in border router queue results by packets of Premium and Basic services. We have configured Premium service for having best real-time connections with reduced trip time, so in this graphs it is important to note the difference between premium

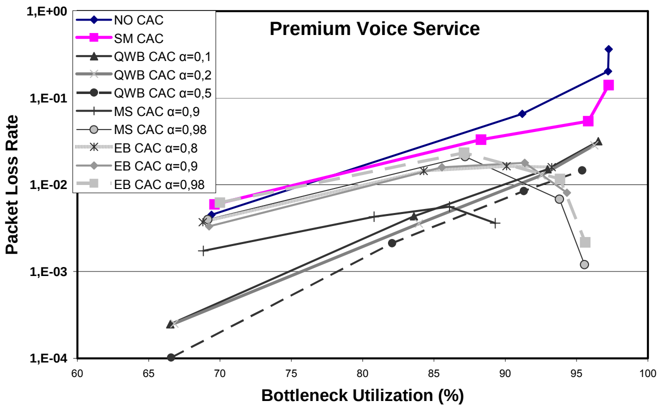

Fig. 5 Premium packet loss rate versus link utilization on a log scale

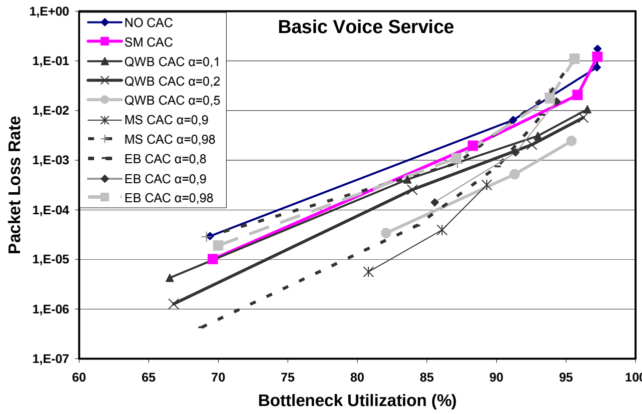

Fig. 6 Basic packet loss rate versus link utilization on a log scale

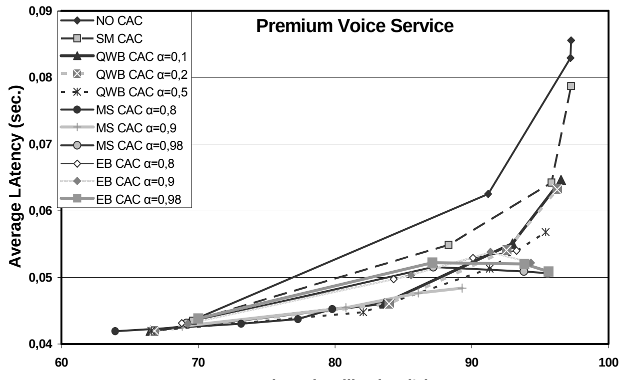

Fig. 7 Premium service latency time results versus offered load and effective link utilization

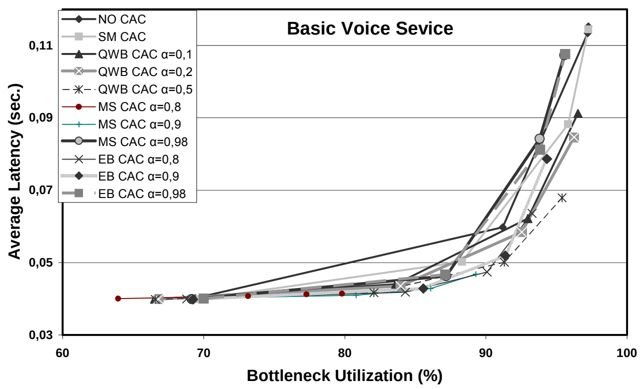

Fig. 8 Basic service latency time results versus offered load and effective link utilization

and basic results. In the main case of premium voice service we obtained a better behaviour from QWB and EB admission controls. While in the case of Basic service, we can notice an important advantage in adopting QWB algorithm in particular with higher voice target utilizations. This is because this new algorithm can adapt itself bandwidth margin to the congestion level of the network using all the bandwidth possible in a compatible manner with the requirements of QoS class.

In general we observed that all the call admission controls examined, apart for NO admission control and Simple Measured admission control, have performed good results in transmission of real-time udp packets, with a good control of packet loss rate for basic service and latency for premium service. The advantage of QWB is that it does not require any knowledge of incoming flow specifications and perform very close to other more sophisticated admission control such as Equivalent bandwidth with a higher computational cost.

Queue schedulers Comparison

In order to evaluate the impact of the queue schedulers on the performance of the TOIP scenario, we present in the following the result of simulations obtained with the previously described queue-schedulers and the QWB CAC algorithm. For each simulation we have used the same traffic parameters and the same configuration of the edge-routers with the appropriate weights of queue servicing in order to get the desired Service Level Agreements. The performances of the queue mechanism were evaluated on the basis of overall throughput achieved on the bottleneck link, and the singular throughput achieved by voice services.

The bar-graph in Fig.9 shows the result achieved during the simulation by all the queue schedulers tested relatively to the utilization of the bottleneck link versus voice offered load. Obviously we are interested on the effects of admission control algorithm and queue schedulers on overloaded link, but light load condition are useful to evaluate the loss of admission control and fairness of behaviour on queue-management due to parameter configuration.

Moreover, in Fig. 9, we can see an homogenous behaviour of queue algorithms with the exception of SFQ, that tends to penalize excessively the data service in

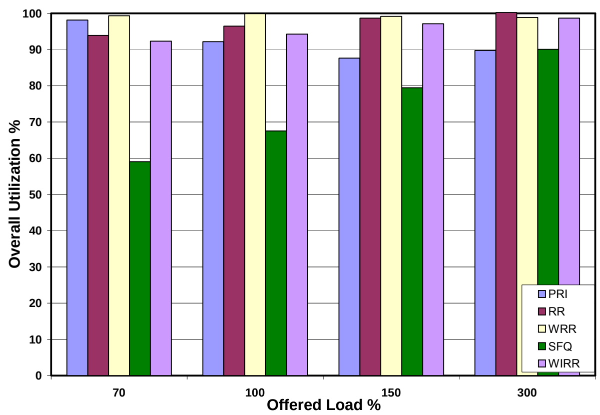

Fig. 9 Data plus Voice services bandwidth utilization versus voice offered load

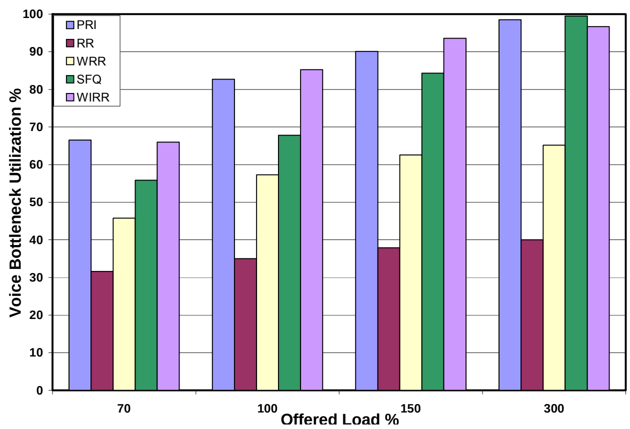

Fig. 10 Bottleneck utilization of virtually voice-only reserved bandwidth versus voice offered load

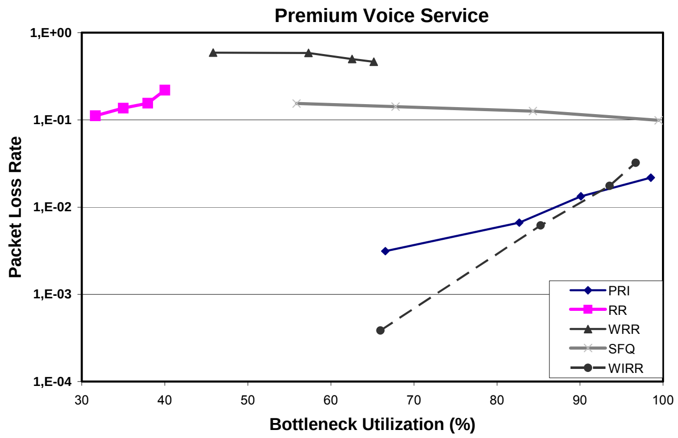

Fig. 11 Premium packet loss rate versus link utilization on a log scale

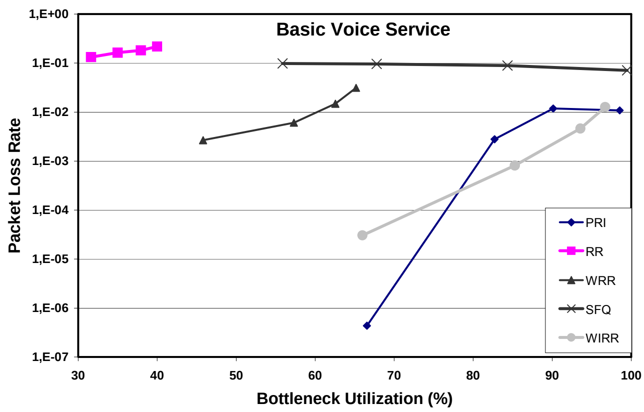

Fig. 12 Basic packet loss rate versus link utilization on a log scale

the case of low voice offered load. On the other hand RR and WRR does not penalize at all data traffic obtaining the best performance in any condition of offered load.

Fig. 10 evidences the utilization of voice reserved part in the bottleneck link versus Voice offered load. So we have subtracted the contribute of TCP best effort data link to the previous considered overall link utilization putting in evidence the effects of queue scheduler on the voice services, which are more relevant in TOIP applications. In this case, we have evidenced that RR and WRR have penalized the voice service obtaining very poor throughput in any condition of offered load. Meanwhile PRI and WIRR were able to give an adequate service to Voice traffic obtaining the best performance in voice throughput. In conclusion on the basis of the simple analysis of bottleneck bandwidth utilization we have put in evidence that RR and WRR are the best indicated queuing algorithm in order to get the maximum link utilization, but are absolutely inadequate for the service of privileged traffic such the voice services considered in our TOIP scenario. While PRI and WIRR are the best solution, as foreseen, for serving the queue characterized by different kind of requirements.

Obviously, these results are still incomplete, because does not have considered the specific service level requirements of the TOIP services implemented in the scenario, but only the aspect of optimization of network resources utilization. It is a matter of fact that in order to get any level of QoS, we need to sacrifice in same way the efficiency of network in order to gain a margin of network resources speedily available for privileged real-time services, such as that considered in TOIP.

So in the following figures we have considered the performances obtained by queue schedulers in QoS indicative parameters for real-time voice services, aiming to optimize the Packet Loss Rate in Basic Service and the network Latency in the Premium Service.

Fig. 11-12 plot on a log scale packet loss rate results obtained by the different queue algorithms in Premium and Basic services; the results obtained under the threshold of 10^{-2} are able to permit an understanding voice call conversation between two end-users. We have configured the Basic voice service as a reliable connection, which is characterized by a reduced packet loss rate. In this configuration almost all the queue solution adopted were able to obtain a better performance in loss rate for basic service than premium service until 90% of bottleneck utilization. However in general PRI and WIRR schedulers were able to get the best behaviour both for basic and premium service. It is interesting to note the behaviour of WIRR curves for basic service, that at high bottleneck utilization has a reduced packet loss rate in confront of PRI ones.

Fig. 13-14 show the mean time spent in border router queue results by packets of Premium and Basic services. We have configured Premium service for having best real-time connections with reduced trip time, so in this graphs it is important to note the difference between premium and basic results. In the main case of premium voice service we obtained a better behaviour from PRI and WIRR queue schedulers. While in the case of Basic service, we can notice an improvement in the performance achieved by WRR algorithm even tough only limited to lower bandwidth utilization.

In general we observed that all the queuing management algorithms examined have performed good results in bottleneck bandwidth utilization, but only PRI and WIRR schedulers have introduced a satisfactory QoS in the service of real-time udp packets, with a good control of packet loss rate for basic service and latency for premium service. The advantage of PRI is that it does not require any knowledge of incoming flow rate specifications and works only considering the grade of real time priority of specific service traffic. Meanwhile the WIRR algorithm need a more complicated configuration with the assignment of the weights to the different service queues but the possibility of guaranteeing a better fairness to the queues with less priority such as Basic Service in the

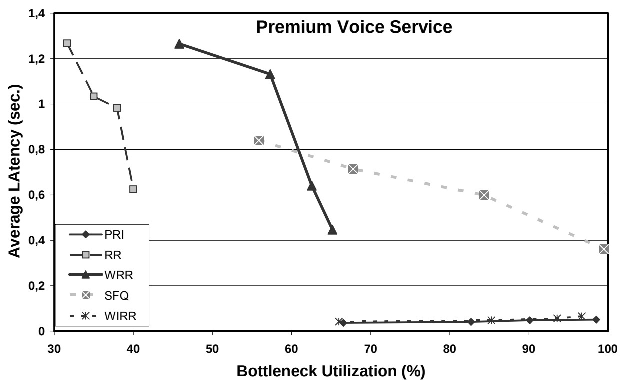

Fig. 13 Premium service latency time results versus offered load and effective link utilization

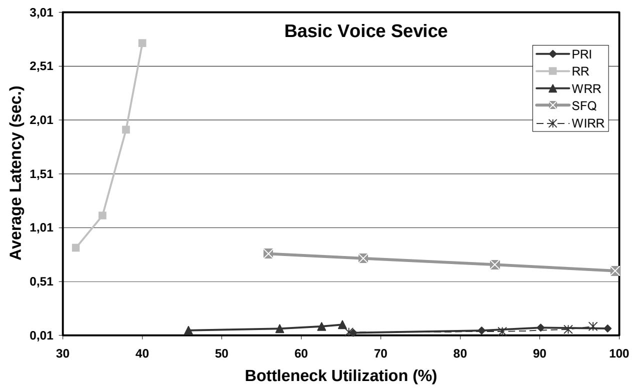

Fig. 14 Basic service latency time results versus offered load and effective link utilization

TOIP scenario considered as verified by the results previously showed.

5. CONCLUSION

The main contribute of this paper consist in introducing a new QoS index-based CAC algorithm and comparing it with other well-know proposed algorithm. The QWB algorithm reaches QoS-parameter results very close to other well-know admission control algorithms, this means that extending the results presented in [8] QWB could be considered on the same level of many other CAC algorithms, but it presents the fundamental advantage of using in a more efficient and adaptive way the available bandwidth. Moreover, the QWB algorithm can work without any a priori knowledge of the incoming traffic characteristics and it could be implemented into ToIP architectures in a very simple way without substantial changes. The results achieved show that it is possible to achieve a sufficient QoS for voice calls with an accurate shaping of traffic offered to the backbone network. In particular we used an end-to-end measured bandwidth and QoS index access control. The differences about QoS-parameters results obtained with the introduction of a CAC algorithm in confront of NO CAC solution are so relevant that the use of these algorithms is highly recommended. Moreover, the introduction of the QoS-Weighted Bandwidth index gives to the CAC a valuable estimate on the congestion state of the network with regard to voice calls, so it avoids that admitting new calls will decrease the quality reached by already established calls. The index α could be efficiently used with a fine granular tuning for improving the parameters indicative of the quality of services, without affecting the efficiency thanks to a dynamic management of bandwidth reservation between the services. The QoS results are quite close to that obtained by Measured Sum and Equivalent Bandwidth with Hoeffding Bounds algorithms with the significant advantage that Qos-Weighted Bandwidth does not require any information about the characteristic of new incoming flows. This fact guarantees a more efficient behavior of the QWB algorithm in a wider range of situations and an easier introduction of this algorithm into the existing network ToIP architecture. Inside this architecture we experimented the impact of queue schedulers in order to find the advantages and disadvantages of well-know algorithm in this TOIP scenario. The behavior of the scheduler was showed to be a decisive element on the overall performance of the system and could influence the end-to-end QoS perceived by the users. Even though many multiple queues scheduler could be configured and work well into the simulated scenario, the PRI and WIRR scheduler show the significant advantage of being easy configuring and with good performance in the management of different classes of services. While other schedulers such as RR and WRR are not indicated to differentiate between real-time voice service obtaining optimal performance in overall bottleneck utilization but not in QoS indicative parameters.

The results achieved show that it is possible to achieve a sufficient QoS for voice calls with an accurate shaping of traffic offered to the backbone network. In conclusion this comparison between queue-scheduler algorithms aimed to focus on the relevant characteristics of the queue service mechanism for TOIP applications. The WIRR algorithm adopt a mechanism of queue choice that is able to offer the best solution in differentiating the service quality performances on the basis of the kind of service even at high offered load for both Premium and Basic, reaching at the same time high values of bottleneck utilization in both overall and voice service cases.

REFERENCES

[1] CSELT DPC 1999.03108, "**Telefonia su IP: problematiche e soluzioni per una rete di transito in tecnologia IP**".

[2] P. Senesi, P. Ferrabone, G. Gritella, R. Rinaldi, M. Siviero, "**Telephony over IP: theoretical modelling and lab experiments**", [Universal Multiservice Networks, 2000.

ECUMN 2000. 1st European Conference on , 2000Page(s): 262 -271]
[3] Computer Networks, **Special Issue on Internet Telephony**, Vol. 31, No. 3, Feb 1999.
[4] IEEE Network, **Special Issue on Internet Telephony**, Vol. 13, No. 3, May/Jun. 1999
[5] A. Vugrinec, S. Tomažič, "**IP telephony from a user perspective**", 10th Mediterranean Electrotechnical Conference, MEleCon 2000, vol. I
[6] D. Houck and G. Meempat, "**Centralized Call Admission Control and Load Balancing for Voice Over IP,**" Pert, and Cont. of Network Sys., SPIE 2000.
[7] H. Knoche and H. de Meer, **QoS Parameters: A Comparative Study for Mapping Purposes**, Tech. REpt.,Computer Science Department, University of Hamburg, August 1998.
[8] S. Jamin, S. Shenker, and P. Danzig. **Comparison of Measurement-based Admission Control Algorithms for Controlled-Load Service**. In Proceedings of IEEE INFOCOM '97, Kobe, Japan, April 1997.
[9] L. Breslau, S. Jamin, and S. Shenker. **Comments on the performance of measurement-based admission control algorithms**. In Proceedings of IEEE INFOCOM 2000, Tel Aviv, Israel, Mar. 2000.
[10] S. Floyd. "**Comments on Measurement-based AdmissionsControl for Controlled-Load Service**". *Computer Communication Review*, 1996.
[11] B. Ahlgren, A. Andersson, O. Hagsand, I. Marsh, "**Dimensioning links for IP telephony**"
[12] D.Black et al., **An architecture for Differntiated Services**, IETF RFC 2475, Dec. 1998
[13] V. Jacobson, K. Nichols and K. Poduri, **An Expedited Forwarding PHB**, IETF RFC 2598, Jun. 1999.
[14] A. Tyagi, J. K. Muppala and H. de Meer, **VoIP Support on Differentiated Services using Expediting Forwarding**, Proc. IPCCC 2000, Phoenix, AZ, USA, Feb. 2000, pp. 574-580
[15] P. Busschbach, D. Houck, and G. Meempat, "**QoS for IP Telephony**", Networks 2000
[16] A. D. Clark, "**Modeling the effects of burst packet loss and recency on subjective voice quality**", IPTEL 2001 IP – Telephony workshop, Columbia University
[17] S. Floyd, V. Jacobson, "**Random Early Detection gateways for congestion avoidance**", IEEE/ACM Transactions on Networking n. 1
[18] S. Blake et a/., "**An Architecture for Differentiated Services**," RFC 2475, Dec. 1998.
[19] N. Greene, M. Ramalho, B. Rosen, "**Media gateway control protocol architecture and requirements**" IETF RFC 2805, April 2001
[20] H. Schulzrinne, S. Casner, R. Frederick, V. Jacobson**,"RTP: a transport protocol for real-time applications**", IETF RFC 1889, Jan. 1996
[21] V. Finenberg, " **A Practical Architecture for Implementing End-to-End QoS in an IP Network**", IEEE Communications Magazine, January 2002
[22] UCB, LBNL, VINT **Network Simulator – NS2**, http://www-mash.cs.berkeley.edu/ns/ns.html