On Degrading Best Effort Calls in Future Cellular Mobile Networks

Helmut Hlavacs & Günter Haring

Institute for Computer Science and Business Informatics, University of Vienna, Lenaug. 2/8, 1080 Vienna, Austria
Tel: +43 1 4277 38450, email: { hlavacs | haring }@ani.univie.ac.at

ABSTRACT

In this paper, the simulator CECALL is described. CECALL has been designed to compare various call admission control strategies for future cellular voice and data networks. The model assumed consists of calls of different type arriving at a cell having available a given number of channels. Each arriving call needs at least a minimum number of channels in order to guarantee a certain level of Quality of Service (QoS). Once calls of type guaranteed and handoff are assigned their channels, they are guaranteed to attain this QoS level for their entire hold time. Best effort calls, on the other hand, may be degraded or even dropped by loosing channels to incoming guaranteed or handoff calls. Here, it is desirable to keep the number of degraded or dropped calls as low as possible. In this paper, various schemes for degrading best effort calls are investigated.
Also, the simulation results are validated and explained by using a simpler analytical model showing both quantitative and qualitative similarities to the simulation results.

I. INTRODUCTION

An important class of future generation packet switched data networks is given by cellular radio networks, i.e., the area covered by the network is divided into cells of various sizes and each cell is equipped with a base station transceiver communicating with mobile devices via radio waves. Examples for such networks are given by HSCSD, GPRS and soon UMTS. A user accessing the data network exchanges data with the particular cell he is currently located in. In each cell, a limited number of channels is available for transporting data, each channel being restricted by a certain bandwidth. Network users establishing a connection or call may allocate one or several channels to send or receive their data. The more channels are used for one call, the more data per second can be sent or received, thus increasing the Quality of Service (QoS) level. However, if single calls occupy too many channels, there may not be enough free channels left to be assigned to newly arriving calls, thus either blocking them or even worse dropping already established calls being transferred from a neighbouring cell, the transfer itself being called *handoff* or *handover*. The system managing the assignment of channels is usually called call admission control (CAC). Advanced CACs may, for example, definine different levels of QoS guarantees for new calls. Upon arrival, calls with higher priority may then take away channels from those with lower priority. This taking away of channels will lead to a QoS *degradation* or even termination or *interruption* of the degraded call.
In this paper the simulator CECALL (CEll Channel ALLocator) is used to evaluate different best effort degradation strategies for guaranteeing QoS levels while keeping the number of blocked and dropped calls as low as possible.

II. RELATED WORK

How to guarantee QoS by resource reservation over IP networks is defined in the RSVP protocol [10]. Likewise, an extension of IP called Mobile IP [8] has been already defined to allow nomadic computing in IP networks. A hybrid solution incorporating both nomadic computing and resource reservation has been proposed in [7] and [9]. Previous work about strategies for call degradation in cellular radio networks can be found, for example, in [2,3,4,5], where both analytical modelling and simulation have been applied.

III. THE SIMULATOR CECALL

The simulator CECALL simulates the call admission control of one single cell *S* being part of a larger cellular radio network. Calls originating in this cell may be of type *guaranteed* or *best effort*. Guaranteed calls represent normal voice or video calls having high priority and being guaranteed to maintain their QoS level, i.e., to keep their once assigned channels. Best effort calls denote low priority data connections and will usually try to allocate more channels compared to guaranteed calls, but are tolerant to loosing once allocated channels down to a minimum number of required channels. Additionally, *handoff* calls denote guaranteed calls entering cell *S* from a neighbouring cell. The system structure is shown in Figure 1.
The call arrival rate λ is the sum of the rates of the three different call classes:

$$\lambda = \lambda_b + \lambda_g + \lambda_h . \quad (1)$$

While best effort (rate λ_b) and guaranteed (rate λ_g) calls enter the simulated cell directly (as this is their point of originality), handoff calls (rate λ_h) will first enter a neighbouring cell but will signal their presence to cell *S*. After a so-called *activation time a* has gone by, handoff calls will immediately enter cell *S*, this procedure modelling the very call handoff.

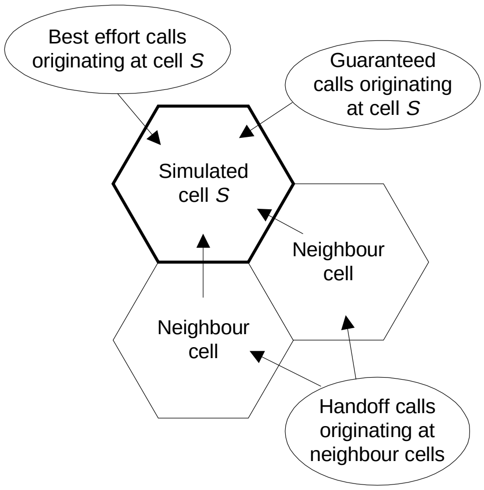

Figure 1: CECALL cell structure.

Each call entering cell *S* must be assigned a minimum number of the *C* channels owned by cell *S*. Best effort calls additionally request more channels up to a certain maximum number, but will work with any number between their minimum and maximum. If there are enough free channels, they are assigned to the newly arrived call and the call may proceed. After the call's *hold time h* has gone by, the call terminates and the allocated channels will be returned to the call admission control where they may be assigned to other calls. On the other hand, if there are not enough channels left, a call originating at cell *S* is said to be *blocked* and will be ended immediately. Likewise, a handoff call being transferred from a neighbour cell is said to be *dropped*. In real life, a dropped handoff call denotes a sudden loss of connection when moving from one cell to another and is considered to be most inconvenient to the network user.

The CECALL model is based on two assumptions. First, each handoff call being generated in a neighbour cell will eventually enter cell *S*. Second, a best effort call may only be degraded or interrupted, but may not be reassigned channels again.

IV. DEGRADATION STRATEGIES

In the current version of CECALL, several different strategies for managing the call access control have been implemented.

The main CAC strategies implemented in CECALL include *complete sharing* (CS), *complete partitioning* (CP) and *dynamic resource partitioning* (DRP). Comparisons of the main features of these strategies have already been presented in [6] and [7]. In this paper, new versions of DRP implementing different ways of degrading best effort calls are be investigated.

In DRP, newly arriving guaranteed and handoff calls may take away channels from already running best effort calls. If due to such a channel loss the number of allocated channels of a best effort call drops below its channel minimum, the call is said to be *interrupted* and terminates, thus deallocating all its still allocated channels. For finding the next channel to be taken away from a best effort call, all DRP versions implement a two-phase approach. In the first phase, channels are taken away only from those best effort calls that can spare channels without being interrupted. If no such calls exist, the CAC enters the second phase, where best effort calls are chosen to be interrupted.

Additionally, if a handoff call is *created*, i.e., it arrives at a neighbour cell, it tries to pre-reserve channels in cell *S*. This procedure, however, is carried out only if enough channels exist that can be reserved. Otherwise, the handoff call will not reserve any channels and will enter cell *S* like a normal guaranteed call. If at this point in time there are still not enough either free or unreserved best effort channels, the handoff call terminates and is counted as being dropped. The channel reservation again follows a two-phase scheme by first reserving only unused and unreserved channels, and, if no such channels are available, by secondly choosing channels currently being used by best effort calls, which have not been pre-reserved by other handoff calls so far.

Each channel thus may be marked as reserved and newly arriving guaranteed and handoff calls may not allocate or reserve such a marked channel. However, if a best effort call does not find enough free channels it may temporarily use reserved channels, risking to be interrupted as soon as the reserving handoff call arrives at cell *S*.

The investigated versions of DRP differ in how best effort calls are chosen next to be either degraded or interrupted. Also, different versions for taking away only one or all channels that can be spared (with names starting with "DRPA") exist. In the standard DRP version, all best effort calls are put into a linear list and newly arriving calls are added to the list end. At all times, a list pointer called *cursor* points at the next best effort call to be degraded or interrupted. In the degradation phase, DRP starts at the cursor and searches for calls that can spare channels without being interrupted. If a call is found, it is degraded by one channel and the cursor is set to the call's successor or the list start in case the list end was reached. If the list is run through once without finding a call being able to spare a channel, DRP first interrupts the call being pointed at by the cursor, then its successors.

In DRP_LT, the best effort calls are ordered according to their lifetime. When DRP_LT enters the degradation phase, it will first degrade the youngest call, then the second youngest and so on. Likewise, in the interrupt phase, first the youngest call will be interrupted, then the second-youngest, and so on. This strategy implements the assumption that interrupting a long-lasting call leaves to a higher customer dissatisfaction than interrupting a young one. Also, it is reasonable to assume that older calls are more likely to reach their call end sooner than younger calls.

Finally, for the degradation phase, strategy DRP_RL orders the best effort calls according to *the relative number of channels they can spare*. If a_b denotes the number of currently allocated and m_b denotes the minimum number of channels of best effort call *b*, then the calls are ordered according to $(a_b - m_b)/m_b$ descending. For the interrupt phase, this time the call with the maximum number of still allocated channels (= min) is chosen to be interrupted next. This is done in the hope that fewer calls will be interrupted, if first the ones with more still allocated channels are terminated. Table 1 shows the investigated degradation strategies.

Table 1: Degradation strategies.

Strategy	Order Phase 1	Taken Away in Phase 1	Order Phase 2
DRP	List	1	List
DRPA	List	All spare	List
DRP_LT	Lifetime	1	Lifetime
DRPA_LT	Lifetime	All spare	Lifetime
DRP_RL	Spare	1	Minimum
DRPA_RL	Spare	All spare	Minimum

V. EXPERIMENTAL SETUP

Although several different distributions are implemented in CECALL, in this investigation, parameters defining the basic system load, i.e., call arrival rates, call hold time and call activation time are assumed to be exponentially distributed, whereas the number of minimum and maximum channels allocated per call are uniformly distributed.

There are two different sets of parameters. The *complex parameter set* is used to derive the main simulation results, whereas the *simple parameter set* defines a simpler model that is mathematically tractable, its sole purpose is for understanding the simulation results. Ranges for the two parameter sets are shown Table 2 and Table 3. In the column denoting the distributions, “C” means a constant, “U” a uniform and “E” an exponential distribution. Calls arrive according to the call arrival rate λ and are partitioned into calls of types best effort, guaranteed and handoff as defined by the parameter “Call partitioning”, in this case having a fixed relation of 4:1:1.

Table 2: Complex parameter set.

Name	Unit	Dist.	Range
Cell bandwidth	Channels	C	10-300
Minimum	Channels	U(1,3)	1-3
Maximum	Channels	U(3,8)	3-8
Arrival rate λ	Calls/Hour	E	200-4000
Hold time *h*	Seconds	E	100/400
Activation time *a*	Seconds	E	50
Call partitioning	b:g:h	C	4:1:1

In the simple parameter set, the handoff activation time is set to zero, thus effectively turning handoff calls into guaranteed calls. Also, each call needs exactly one channel instead of several. This way, the model can be represented by an M/M/C/C loss system with two different classes of calls (guaranteed/handoff and best effort).

Table 3: Simple parameter set.

Name	Unit	Dist.	Range
Cell bandwidth	Channels	C	10-300
Minimum	Channels	C	1
Maximum	Channels	C	1
Arrival rate *l*	Calls/Hour	E	200-4000
Hold time *h*	Seconds	E	400
Activation time *a*	Seconds	C	0
Call partitioning	b:g:h	C	4:1:1

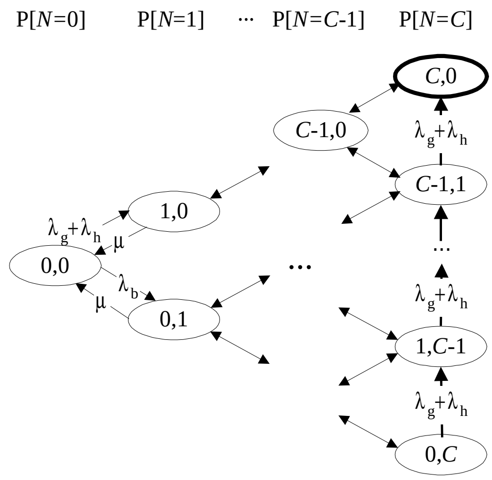

Figure 2: System state transition diagram.

State changes occur upon call arrivals with the rates given in (1) and call departures with rate $\mu = 1/h$. Then, by setting $u = \lambda/\mu$, the probability P[*N*=*n*] that there are exactly $n = 0,1,\ldots,C$ calls in the system is given by [1]

$$\mathrm{P}[N=n] = \frac{u^n / n!}{1+u+u^2/2!+\ldots u^C/C!}. \qquad (2)$$

As there are two different classes of calls, $S = \{(g,b) \mid g+b \le C\}$ denotes the states the system can be in, where *g* is the number of guaranteed and handoff calls and *b* is the number of best effort calls in the system. The state transition diagram of the system can be seen in Figure 2. In the following, the probability of being in state (*g*,*b*) will be denoted by P[*g*,*b*].

As in this simple case, each call allocates only one channel, there are no degradations but only interrupts. Interrupts can only occur in the states $\{(g,b) \mid g+b=C, g \ne C\}$. Here, $\mathrm{P}[N=C] - \mathrm{P}[C,0]$ is the probability of being in *any* of these states, yielding an interrupt rate of $(\lambda_g+\lambda_h)(\mathrm{P}[N=C]-\mathrm{P}[C,0])$ interrupts per second. The number of interrupts occurring in a time period of *t* seconds is thus given by $t(\lambda_g+\lambda_h)(\mathrm{P}[N=C]-\mathrm{P}[C,0])$. Likewise, the probability for a best effort call being *accepted* is given by $1-\mathrm{P}[N=C]$, the number of accepted best effort calls in a time period t thus is $t\left(1-\mathrm{P}[N=C]\right)\lambda_b$. The probability for interrupting an accepted best effort call is then given by

$$\mathrm{PI} = \frac{\left(\lambda_g + \lambda_h\right)\left(\mathrm{P}[N=C] - \mathrm{P}[C,0]\right)}{\lambda_b\left(1-\mathrm{P}[N=C]\right)}. \qquad (3)$$

VI. SIMULATION RESULTS

Each run simulated 150,000 virtual seconds, the CPU time needed on a Pentium II 400 MHz for one run was between 0.5 and fifteen seconds.

Figure 3 to Figure 6 show how the number of degradations per accepted best effort call and the blocking and interrupt probability for best effort calls depend on the call arrival rate λ and the number of available channels C. In these results, the complex parameter set shown in Table 2 is used.

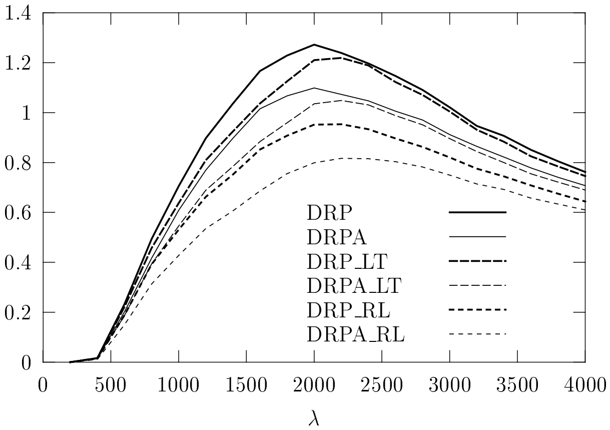

Figure 3: Number of degradations per accepted best effort call depending on the call arrival rate λ. C=250, h=400 and a=50.

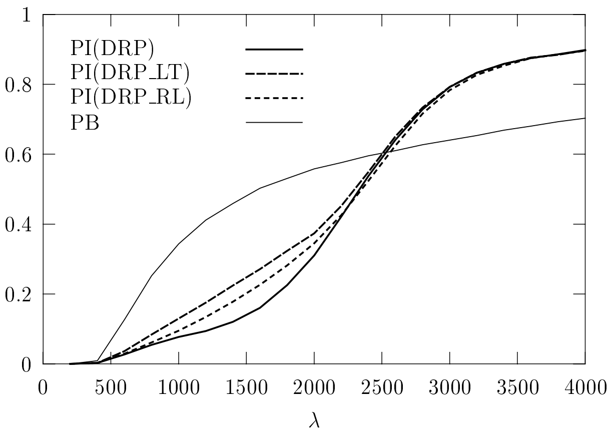

Figure 4: Probability of best effort call blocking (PB) and best effort call interrupt (PI) depending on the call arrival rate λ. C=250, h=400 and a=50.

For the call blocking probability there is no difference between the different strategies, and thus, only one representative result is shown. Likewise, for the interrupt probability, there is no difference between the strategies taking away only one channel or all spareable channels. The results show several peculiarities.

First, when varying both λ and C, the degradations per call first rise but fall again afterwards. This observation can be explained by using the simple parameter set shown in Table 3, as in this case there are only interrupts possible and the interrupt probability is given by (3). In order to compensate the fact that in the reduced model each call allocates less channels (by a factor between 2 and 4), when varying λ, the number of channels C is reduced to 60, whereas when varying C, the hold time h is increased to 400. Both analytical and simulation results for the simple case can be seen in Figure 7 and Figure 8. When comparing the simple model to the complex model, it can be seen that the degradations of the complex model depend on $\mathrm{P}[N=C]-\mathrm{P}[C,0]$, the probability of all channels being allocated, but having one or more channels assigned to best effort calls. Also it can be seen that in both models the probability for interruption can be explained by (3).

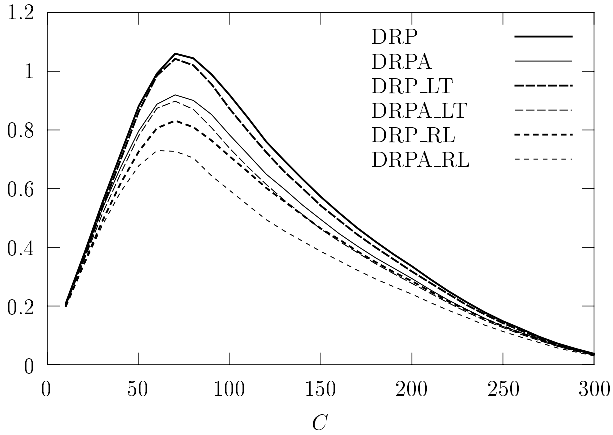

Figure 5: Number of degradations per accepted best effort call depending on the number of channels C. λ=2000, h=100 and a=50.

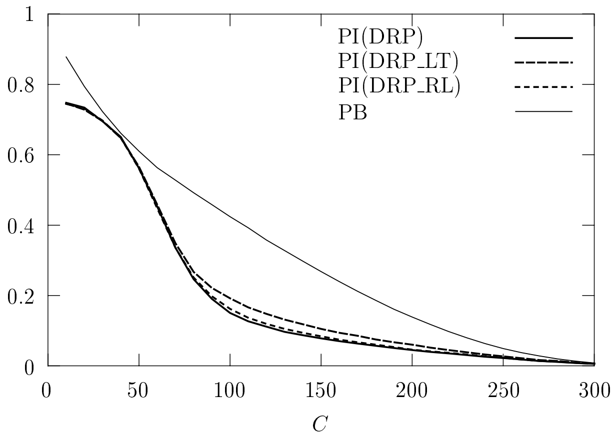

Figure 6: Probability of best effort call blocking (PB) and best effort call interrupt (PI) depending on the number of channels C. λ=2000, h=100 and a=50.

The second observation is that strategy DRP_RL has the same or even higher interrupt probability than DRP. Thus, ordering the calls according to their minimum channels in the interrupt phase does not improve the interrupt probability. This can be explained by the fact (not shown here) that in high load situations, accepted best effort calls are likely to have a minimum of only one channel, thus ordering them according to this minimum does not increase the system information.

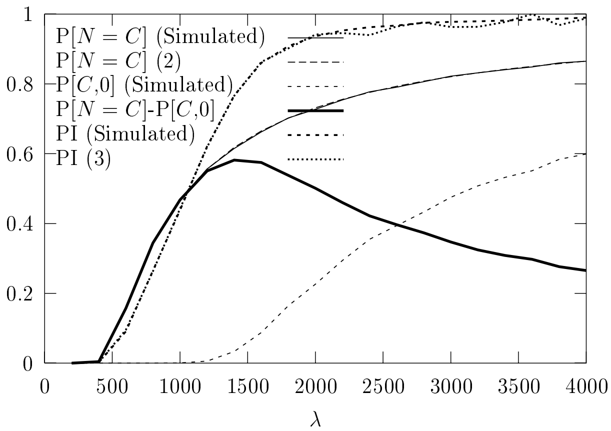

Figure 7: Probabilities for being not in state (*C*,0)) (P[*N*=*C*] - P[*C*,0]) and for best effort call interrupt (PI). *C*=60, *h*=400 and *a*=50.

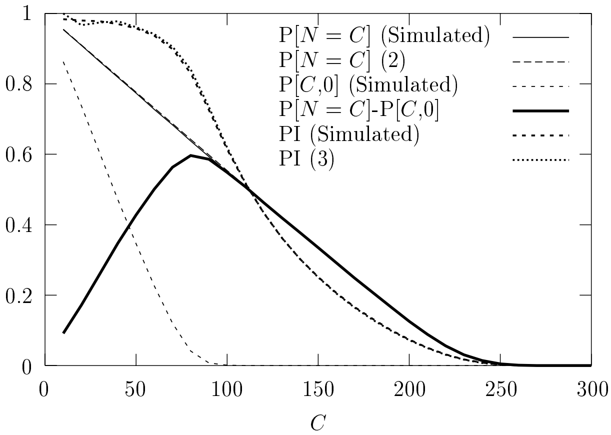

Figure 8: Probability for being not in state (*C*,0) (P[*N*=*C*] - P[*C*,0]) and for best effort call interrupt (PI). λ =2000, *h*=400 and *a*=50.

The next observation shows that in medium load situations there are cases where in the mean, more than one degradation per call is possible. This can be explained by the fact that in such situations, P[*N*=*C*] increases quickly, whereas P[*C*,0] remains zero. Thus, all channels are assigned to calls, but there are always channels assigned to best effort calls. This indicates that the arrival rate of guaranteed and handoff calls is not high enough to occupy all channels themselves. We also see from the results that new best effort calls will be accepted with a probability of 40-60%, thus often not finding free, but pre-reserved channels they can occupy. When the according handoff calls arrive, they claim back their channels, thus degrading the according best effort calls and influencing the degrade statistic accordingly. This coincides with the observation that the number of degradations per call is lowest for DRPA_RL and highest for DRP, as in DRP, newly arriving handoff calls will run through the best effort calls in a round-robin fashion in order to pre-reserve their channels. Here, the number of degraded calls is increased drastically, because only one channel is taken away from a large number of calls, whereas in DRPA_RL, a large number of channels is taken away from a small number of calls. Thus, in DRPA_RL, the number of degradations per accepted best effort call is lowest.

VII. CONCLUSION AND FUTURE OUTLOOK

In this paper, different degradation strategies for best effort calls in future cellular networks have been investigated by applying both simulation and analytical modelling. Whereas simulation is essential for deriving the main results, the modelling of the system by using an extension of an M/M/*C/C* loss system is essential for understanding them.

Future investigations will include an assessment of the satisfaction of network users and how to dynamically adapt the used degradation strategy depending on the currently observed system load.

REFERENCES

[1] A.O. Allen, “Probability, Statistics and Queueing Theory”, Academic Press, New York, 1997.

[2] K. Begain, G. Bolch, M. Telek, “Scalable Schemes for Call Admission and Handover in Cellular Networks with Multiple Services”, Wireless Personal Communications, vol. 15, 2000, pp. 125—144.

[3] V. Bharghavan, K.W. Lee, S. Lu, S. Ha, D. Dwyer, J. Li, "The TIMELY Adaptive Resource Management Architecture," IEEE Personal Communications Magazine, 1998.

[4] S.K. Das, R. Jayaram, N.K. Kakani, S.K. Sen, “A call admission and control scheme for quality-of-service (QoS) provisioning in next generation wireless networks”, Wireless Networks, vol. 6, 2000, pp. 17—30.

[5] K. Lee, “Supporting mobile multimedia in integrated services networks”, Wireless Networks, vol. 2, 1996, pp.205—217.

[6] A. Mahmoodian, G. Haring, ”A Resource Allo-cation Mechanism to Provide Guaranteed Service to Mobile Multimedia Applications”, in Proceedings of First IEEE/POPOV Workshop on Internet Technologies and Services, Oct. 1999, Moscow.

[7] A. Mahmoodian, G. Haring, “Mobile RSVP with Dynamic Resource Sharing”, in Proceedings of Wireless Communications and Networking Conference 2000, Sept. 2000, Chicago.

[8] C.E. Perkins, “Mobile IP”, IEEE Communications Magazine, May 1997, pp. 84—92.

[9] A.K. Talukdar, B.R. Badrinath, A. Acharya, “MRSVP: A Reservation Protocol for an Integrated Services Packet Network with Mobile Hosts”, Technical Report dcs-tr-337, Department of Computer Science, Rutgers University, New Jersey.

[10] L. Zhang, S. Deering, S. Estrin, D. Shenker, D. Zappala, “RSVP: A New Resource Reservation Protocol”, IEEE Network, September 1993, pp. 8—19.