University of Würzburg
Institute of Computer Science
Research Report Series

A Performance Evaluation Framework for Network Admission Control Methods

Michael Menth, Stefan Kopf, Jens Milbrandt

Report No. 305 February 2003

Department of Distributed Systems
Institute of Computer Science, University of Würzburg
Am Hubland, D-97074 Würzburg, Germany
phone: (+49) 931-8886644, fax: (+49) 931-8886632
{menth|kopf|milbrandt}@informatik.uni-wuerzburg.de

A Performance Evaluation Framework for Network Admission Control Methods

Michael Menth, Stefan Kopf, Jens Milbrandt
Department of Distributed Systems
Institute of Computer Science, University of Würzburg
Am Hubland, D-97074 Würzburg, Germany
phone: (+49) 931-8886644, fax: (+49) 931-8886632
{menth|kopf|milbrandt}@informatik.uni-wuerzburg.de

Abstract

In this paper we explain the difference between *link* and *network* admission control (LAC, NAC) for wireline networks. We introduce three fundamentally different budget based NAC methods that categorize most of today's implemented NAC approaches and establish a performance evaluation framework for their comparison. We propose a capacity dimensioning method for multi-rate traffic and illustrate its sensitivity on a singe link to various traffic characteristics, leading to a suitable definition of blocking probability. We extend that method to NAC methods that are based on distributed budgets. Finally, we compare the different NAC methods under varying load conditions, using their obtainable resource utilization as performance measure.

Keywords: QoS, Admission Control, Resource Allocation, Performance Evaluation

1 Introduction

In a connection oriented network, admission control (AC) is easily combined with connection state management at each network node. Thus, it is performed link by link like in ATM or in the Integrated Services framework. AC for a single link – we call it *link* admission control (LAC) – can be done by flow descriptor based resource reservation assisted by effective bandwidths or by measurement based AC (MBAC), and it is well understood from research in the ATM context [4]. In contrast, a connectionless network, e.g. IP network, does not deal with connection or resource management at the network nodes. Correspondingly, a *network* admission control (NAC) approach is advisable that admits reservations only at dedicated locations, e.g. the borders of a network, without contacting individual routers for admission decisions. We present three basically different budget based NAC approaches that categorize today's NAC implementations [2, 3] and ease their understanding.

This work was funded by the Bundesministerium für Bildung und Forschung of the Federal Republic of Germany (Förderkennzeichen 01AK045) and Siemens AG, Munich. The authors alone are responsible for the content of the paper.

These NAC approaches have different complexity and resource efficiency but there is no numerical performance comparison in the literature. The presentation of a framework for the performance comparison of budget based NAC approaches is the purpose of this paper. Our work considers different design options for such studies and leads to the following preferred methodology. First, we dimension the NAC budgets based on a given traffic matrix and routing such that desired border-to-border (b2b) blocking probabilities are met in the network. Then, we determine the required capacity for all links in the network and take the resulting resource utilization as performance criterion. The performance depends mainly on the ability of the NAC methods to exploit the economy of scale. For a better understanding, we illustrate the influence of many factors (offered load, request size distribution, definition and size of the required blocking probability) on the resource utilization only on a single link. Finally, we compare the performance of the different NAC approaches in a sample network depending on the offered load.

The paper is structured as follows. Section 2 gives an overview of different existing NAC approaches. Section 3 proposes our framework. The numerical results of Section 4 present a performance comparison for the basic NAC methods. Section 5 summarizes this work and gives an outlook on further research.

2 Methods for Network Admission Control (NAC)

In this section we distinguish between link and network admission control and explain three basically different NAC concepts.

2.1 Link and Network Admission Control

QoS criteria are usually formulated in a probabilistic way, i.e., the packet loss probability and the probability that the transport delay of a packet exceeds a given delay budget must both be lower than certain thresholds. Link admission control (LAC) takes the queuing characteristics of the traffic into account and determines the required bandwidth to carry flows over a single link without QoS violations. This includes two different aspects. First, bursty traffic requires more bandwidth for transmission than its mean rate to keep the queuing delay low which can be predicted by queuing formulae [4]. Secondly, flows usually indicate a larger mean rate than required just to make sure that there is enough bandwidth available when needed. To cope with these problems, overbooking by the provider, measurement based AC (MBAC) [5, 6], or effective bandwidths [7] are used. These mechanisms limit the traffic load primarily on a single link, so we call them LAC.

Network admission control (NAC) needs to protect more than one link with an admission decision. This is a distributed problem with various solutions differing in their degree of storage and processing demands, locality and achievable multiplexing gain due to the partitioning of resources into budgets administered in different locations. Moreover, the solutions have different efficiency, i.e. they require different amounts of network capacity to meet the same b2b flow blocking probability p_{b2b} which affects the network operator's costs.

NAC and LAC can be combined, i.e. flow's required capacity $f.c$ [1] may consist of an effective bandwidth to take some overbooking in the presence of large traffic aggregates into account. In this investigation, we only focus on the combinatoric NAC problem, i.e. we blind out the issues of determining the effective bandwidth for individual reservations or potential MBAC based overbooking.

In general, an AC entity records the demand of the admitted flows $\mathcal{F}_{admitted}$ in place. When a new flow arrives, it checks whether its effective bandwidth together with the demand of already established flows fits within a capacity budget. If so, the flow is accepted, otherwise it is rejected.

2.2 Link Budget Based Network Admission Control (LB NAC)

The link-by-link NAC is probably the most intuitive NAC approach. The capacity $l.c$ of each link l in the network is managed by a single link budget $LB(l)$ (with size $LB(l).c$) that may be administered, e.g., at the ingress router of that link or in a centralized database. A new flow $f_{new}(v, w)$ with ingress router[2] v, egress router w, and bitrate $f_{new}.c$ must pass the AC procedure for the LBs of all links that are traversed in the network by f_{new} (cf. Figure 1). The NAC procedure will be successful if the following inequality holds

$$\begin{aligned}&\forall l \in \mathcal{E} : l.u(v, w) > 0 : f_{new}(v, w).c \cdot l.u(v, w)+\\&\textstyle\sum_{f(x,y)\in\mathcal{F}_{admitted}(l)} f(x, y).c \cdot l.u(x, y) \leq LB(l).c.\end{aligned} \tag{1}$$

There are many systems and protocols working according to that principle. The connection AC in ATM [8] and the Integrated Services [9] architecture in IP technology adopt it in pure form and induce per flow reservation states in the core. Other protocols reveal the same behavior although the mechanism is not implemented as an explicit LB NAC. A bandwidth broker [10, 11, 12] administers the budgets in a central database which represents a single point of failure but behaves the same way from the performance point of view. The stateless core approaches [13, 14, 15] avoid reservation states in the core at the expense of measurements or increased response time. Reservation states in the core, measurements, or increased response times are a drawback if network resilience is required. The following two basic NAC methods manage the network capacity in a distributed way, i.e. all budgets related to a flow can be consulted at its ingress or its egress border router. In a failure scenario, only fast local rerouting of the traffic is required and the QoS is maintained if sufficient backup capacity is available.

2.3 Ingress and Egress Budget Based Network Admission Control (IB/EB NAC)

The IB/EB NAC defines for every ingress node $v \in \mathcal{V}$ an ingress budget $IB(v)$ and for every egress node $w \in \mathcal{V}$ an egress budget $EB(w)$ that must not be exceeded. A new flow $f_{new}(v, w)$

[1] We borrow parts of our notation from the object-oriented programming style: $x.y$ denotes a property y of an object x. We prefer $x.y$ to the conventional y_x since this is hard to read if the name of x is complex.

[2] A networking scenario $\mathcal{N} = (\mathcal{V}, \mathcal{E}, u)$ is given by a set of routers $\mathcal{V}$ and set of links $\mathcal{E}$. The b2b traffic aggregate with ingress router v and egress router w is denoted by $g(v, w)$, the set of all b2b traffic aggregates is $\mathcal{G}$. The function $l.u(v, w)$ with $v, w \in \mathcal{V}$ and $l \in \mathcal{E}$ reflects the routing and it is able to cover both single- and multi-path routing by indicating the percentage of the traffic rate $g(v, w).c$ using link l.

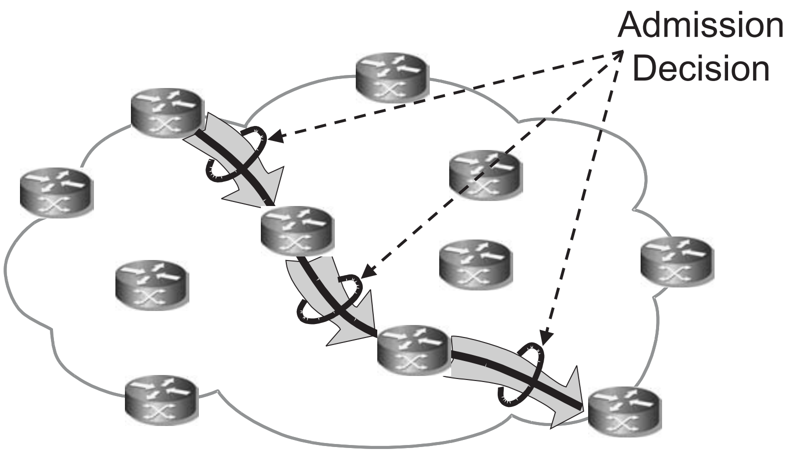

Figure 1: Network admission control based on link budgets.

must pass the AC procedure for $IB(v)$ and $EB(w)$ and it is only admitted if both requests are successful (cf. Figure 2). Hence, the following inequalities must hold

$$f_{new}(v,w).c + \sum_{f \in \mathcal{F}^{ingress}_{admitted}(v)} f.c \leq IB(v).c \quad (2)$$

$$f_{new}(v,w).c + \sum_{f \in \mathcal{F}^{egress}_{admitted}(w)} f.c \leq EB(w).c \quad (3)$$

Flows are admitted at the ingress and the egress irrespective of their egress or ingress routers. This entails that the capacity managed by an IB or EB can be used in a very flexible manner. However, all – also pathological – traffic patterns that are admissible by the IBs and EBs must be carried by the network with the required QoS. Therefore, enough capacity must be allocated on the network links.

If we leave the EBs aside, we get the simple IB NAC, so only Equation (2) must be met for the AC procedure. This idea originates from the DiffServ context [16, 17] where traffic is admitted only at the ingress routers without looking at the destination address of the flows. The QoS should be guaranteed by a sufficiently low utilization of the network resources by high quality traffic. To avoid any confusion: DiffServ is a mechanism for the forwarding differentiation of differently labelled packets while the IB NAC is just one concept among many others for the management of network resources within that context.

2.4 B2B Budget Based Network Admission Control (BBB NAC)

The BBB NAC is able to exclude pathological traffic patterns by taking both the ingress and the egress border router of a flow $f(v,w)$ into account for the AC procedure, i.e. a b2b budget $BBB(v,w)$ manages the capacity of a virtual tunnel between v and w. A new flow $f_{new}(v,w)$ passes only the AC procedure for $BBB(v,w)$ (cf. Figure 3). It is admitted if this request is

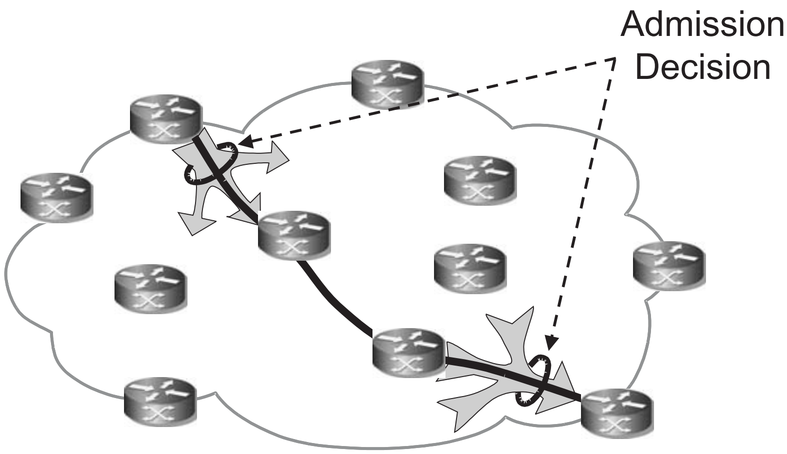

Figure 2: Network admission control based on ingress and egress budgets.

successful, i.e. if the following inequality holds

$$f_{new}(v,w).c + \sum_{f \in \mathcal{F}_{admitted}(v,w)} f.c \leq BBB(v,w).c. \quad (4)$$

The $BBB(v,w)$ may be controlled, e.g., at the ingress router v or at the egress router w, i.e. the BBB NAC also avoids states inside the network. The capacity of a tunnel is bound by the BBB to one specific b2b aggregate and can not be used for other traffic with different source or destination. Hence, there is no flexibility for resource utilization. Therefore, the concept is often realized in a more flexible manner, such that the size of the BBBs can be rearranged [18, 19]. Tunnels may also be used hierarchically [20]. The tunnel capacity may be signaled using explicit reservation states in the network [21, 22], only in logical entities like bandwidth brokers [11], or it may be assigned by a central entity [23].

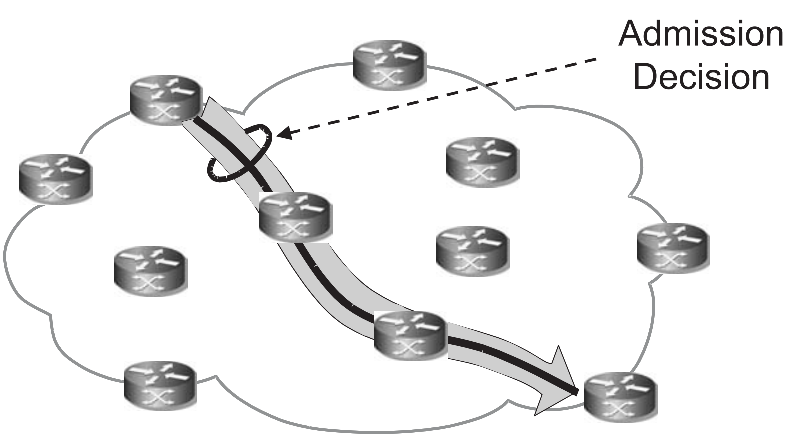

Figure 3: The BBB NAC corresponds to a logical tunnel.

3 A Performance Evaluation Framework for NAC Approaches

In this section we propose a dimensioning method for multi-rate traffic and investigate the influence of offered load, blocking probability, request rate variability on a single link. This illustrates the economy of scale which is the key for understanding NAC performance and leads to an appropriate definition of blocking probability for multi-rate traffic. We point out design options for the assessment of the NAC performance and choose one as our preferred methodology. We explain the budget and link capacity computation for the different NAC types and study the impact of several methods for setting the required budget blocking probabilities. These results lead to our performance evaluation methodology.

3.1 Capacity Dimensioning for a Single Link

NAC is required to protect the network against overload and to guarantee a certain level of QoS in terms of packet loss and delay[3]. This is achieved by flow blocking if the network is highly loaded. Future real-time traffic in the Internet is probably similarly distributed as the traffic in the telephone network since it is triggered and consumed by humans in real-time. Therefore, we use a Poisson traffic model for A although there are more sophisticated traffic models for current Internet applications without real-time requirements. Based on a call inter-arrival time and a call holding time distribution, the mean $E[A]$ and variance $VAR[A]$ of simultaneous calls can be computed. They determine together with the mean $E[C]$ and variance $VAR[C]$ of the request size distribution the mean and the variance of the resulting aggregate reserved rate R:

$$E[R] = E[A] \cdot E[C] \quad (5)$$

$$VAR[R] = E[A] \cdot VAR[C] + VAR[A] \cdot E[C]^2. \quad (6)$$

The large impact of the statistics of C and the fact that the distribution of C is not known for future real-time applications requires parameter studies. Among others, the fluctuation of R ($VAR[R]$) influences the required capacity and the resource efficiency of the NAC methods. However, the relative performance of the NAC methods is independent of the distribution of R. Therefore, the choice of the traffic model is not critical as long as we use the same one for comparisons.

Hence, we assume that flows have a Poisson arrival rate λ and an arbitrarily distributed call holding time with average $\frac{1}{\mu}$. The offered load is the average number of potentially simultaneously active flows $a = \frac{\lambda}{\mu}$. If no flows are blocked, the number of active connections A is a random variable distributed according to a Poisson distribution $\Pr(A = k) = \frac{a^k}{k!} \cdot \exp(-a)$ with a mean of $E[A] = a$ and a variance of $VAR[A] = a$. Unlike in the telephone network, multiple request rates will occur in data networks, e.g. for telephone and video conference applications. Our working assumption is an average request rate of $E[C] = 256$ Kbit/s. We want

[3]Loss and delay probabilities are very sensitive to the traffic characteristics, i.e. they must be taken into account for economic capacity dimensioning if the link bandwidth is small or if overbooking is applied. Since these are LAC issues and not NAC issues, we blind them out and consider only peak rate allocation or effective bandwidths for simplicity reasons.

to conduct parameter studies, so we increase the coefficient of variation linearly by defining interpolated request rate distributions C_t like in Table 1. The rate distribution C_t produces $n_r = 3$ different request types r_i, $0 \leq i < n_r$ with $r_0.c = 64$ Kbit/s, $r_1.c = 256$ Kbit/s, and $r_2.c = 2048$ Kbit/s. We refer to their probabilities by $r_i.prob$, to their rate by $r_i.c$, and to their corresponding offered load by $r_i.a = r_i.prob \cdot a$.

Capacity dimensioning is essentially the computation of a suitable bandwidth c, for which the calculation of flow blocking probabilities yields acceptable results. The bandwidth is modelled by an M/M/n-0 loss model, i.e. several parallel queues with each of them representing a bandwidth portion of 64 Kbit/s, and the probability for the number of occupied queues is determined. The sum of the probabilities of the queue occupation states in which blocking occurs for an arriving request type r_i is its blocking probability $r_i.p$.

Table 1: Statistics of request rate C_t.

$E[C_t]$ = 256 Kbit/s	$r_0.prob = \frac{28}{31} \cdot t^2$
$VAR[C_t]$ = $344064 \cdot t^2$	$r_1.prob = 1 - t^2$
$c_{var}[C_t]$ = $2.291 \cdot t$	$r_2.prob = \frac{3}{31} \cdot t^2$

To that aim, we review the Kaufman & Roberts algorithm [4] for capacity dimensioning if the offered load a and the request rate distribution C_t are given. The Kaufman & Roberts solution requires a maximum capacity unit u_c for scaling, so we choose $u_c = gcd(\{r_i.c : 0 \leq i < n_r\})$[4]. The request rates are converted into multiples of this finest granularity by $r_i.c_u = \frac{r_i.c}{u_c}$ and so is the capacity $c_u = \frac{c}{u_c}$. First, weights are computed

$$w(j) = \begin{cases} 0 & : j < 0, \\ 1 & : j = 0, \\ \frac{1}{j} \cdot \sum_{0 \leq i < n_r} w(j - r_i.c_u) \cdot r_i.c_u \cdot r_i.a & : 0 < j \leq c_u \end{cases} . \qquad (7)$$

The normalization of the weights w yields the probability distribution q for the number of busy servers and their summation yields the distribution function Q

$$q(j) = \frac{w(j)}{\sum_{0 \leq i \leq c_u} w(i)} \text{ and } Q(j) = \sum_{0 \leq i \leq j} q(i). \qquad (8)$$

The blocking probability of a single request type r is $r.p = 1 - Q(c_u - r.c_u)$. Note that the computation of this algorithm needs an intelligent implementation to make it computationally tractable. The blocking probabilities can be calculated either per flow (p_f) or related to the offered transmission rate $a \cdot E[C]$ (p_c). They are computed by the following equations

$$p_f = \sum_{0 \leq i < n_r} r_i.p \cdot r_i.prob \quad \text{and} \qquad (9)$$

$$p_c = \frac{\sum_{0 \leq i < n_r} r_i.p \cdot r_i.prob \cdot r_i.c}{E[C]} \qquad (10)$$

In the following we will see the advantage of p_c over p_f for performance comparison reasons.

[4] gcd denotes the greatest common divisor.

3.2 NAC Performance for a Single Link

We study the impact of the traffic parameters and the blocking probabilities on the required capacity and the resource utilization on a single link. The results help to understand the performance of NAC methods and to reduce the number of parameter studies for their evaluation.

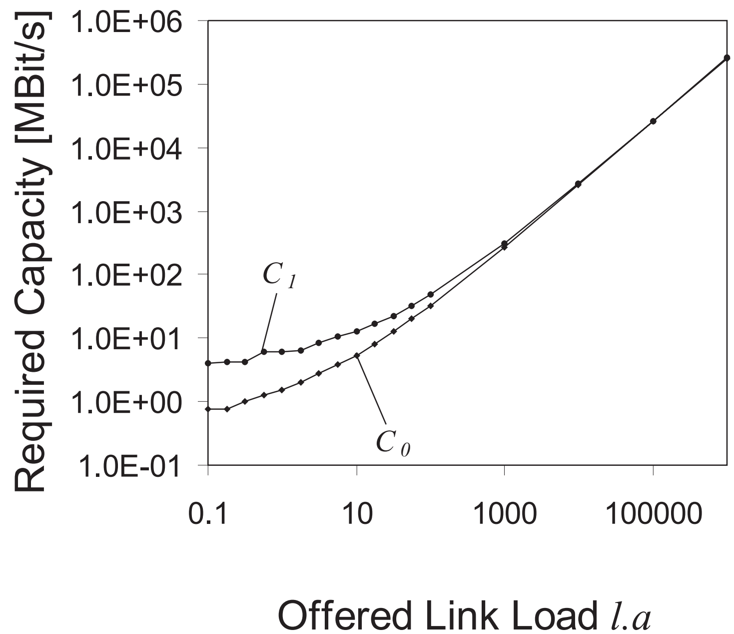

Figure 4: Impact of offered load and request rate variability on the required link capacity.

3.2.1 The Impact of the Rate Variability

Economy of scale or multiplexing gain is the key for understanding the performance behavior of NAC approaches and can be best illustrated on a single link. It is the fact that little offered load leads to low utilization and that large offered load leads to high utilization when a link is dimensioned for that load and for a specific QoS requirement in terms of blocking probability. Figure 4 shows that both for low and highly variable request rate distributions the required link capacity is almost proportional to the offered link load, at least for an offered load of $l.a = 1000$ Erlang or larger. Figure 5 illustrates that the resource utilization increases drastically up to an offered load of $l.a = 1000$ Erlang, hence, the resource utilization depends heavily on the offered link load $l.a$ and it is a good measure for multiplexing gain. The resource utilization for traffic with little or no variance (C_0) is higher than for traffic with large variance (C_1). In the following investigations, we use rate distribution C_1 as default since we expect the traffic in the future Internet to be more variable than in the telephone network. The difference between rate distribution C_0 and C_1 can be better observed with the resource utilization than with the required capacity curves.

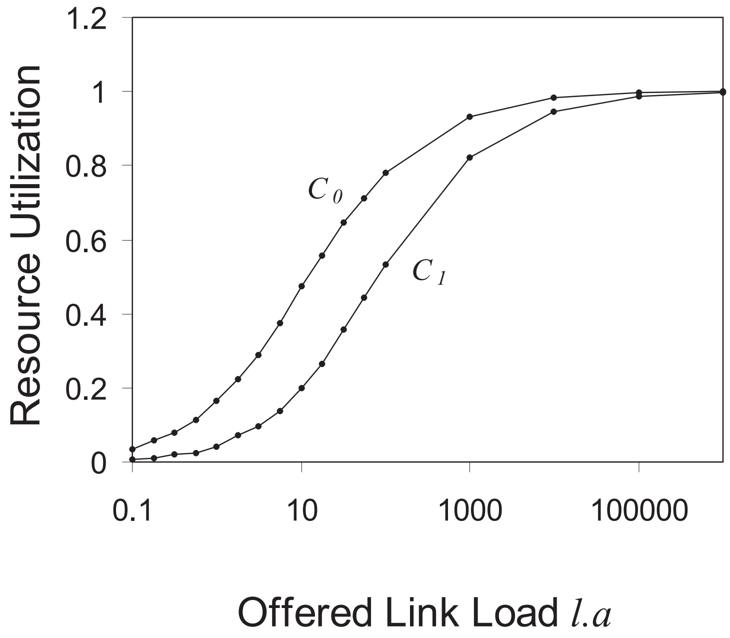

Figure 5: Impact of offered load and request rate variability on the resource utilization.

3.2.2 The Impact of the Link Blocking Probability

Figure 6 illustrates the influence of the offered load and the blocking probability for C_1 on the resource utilization. The resource utilization is mainly ruled by the offered load but it also benefits from large link blocking probabilities. The impact of the blocking probabilities decreases for high offered load. Figure 7 shows the resource utilization for $l.a = 100$ Erlang and for different request type distributions C_t. We observe that the resource utilization decreases with increasing rate variability. The difference between the blocking probability options p_f and p_c is large for large rate variabilities and large link blocking probabilities.

3.2.3 The Impact of the Definition of the Link Blocking Probability

We have pointed out that blocking probabilities can be defined per flow or in relation to the transmission rate. We investigate them for $l.a = 100$ Erlang depending on the request rate variability. Figure 8 shows the request type specific blocking probabilities for blocking probability $p_c = 10^{-3}$ which is related to the overall transmission rate while Figure 9 shows them for the flow specific blocking probability $p_f = 10^{-3}$. In both cases, request type r_0 has a smaller rate and experiences lower blocking probabilities than 10^{-3} whereas request type r_2 with a larger rate experiences higher blocking probabilities than 10^{-3}. This phenomenon can not be avoided without request type specific AC. We observe that the request type specific blocking probabilities are almost constant since dimensioning with p_f does not take the request rate distribution into account. In contrast, the blocking probabilities decrease with increased rate

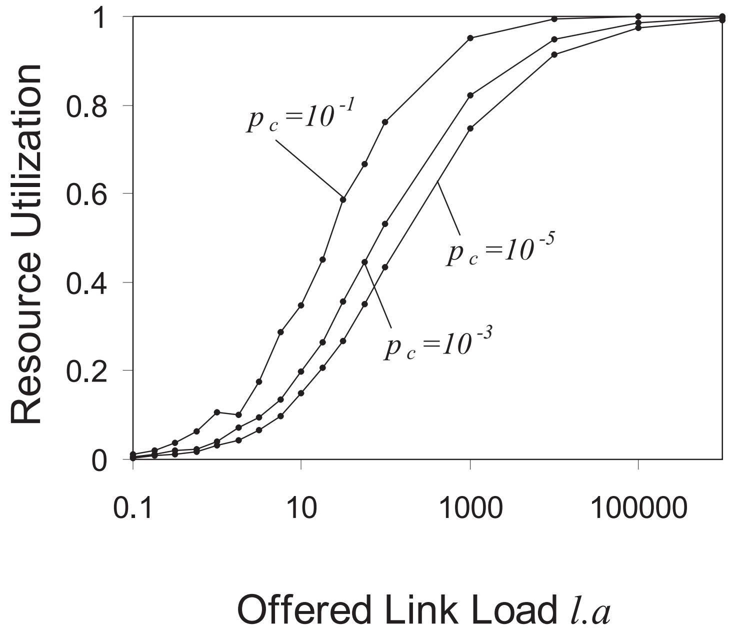

Figure 6: Impact of offered load and the blocking probability on the resource utilization.

distribution variability if capacity dimensioning is based on p_c. This has some impact on the blocked traffic as illustrated in Figure 10. If transmission rate related blocking probabilities are used for link dimensioning, the portion of the lost traffic is per definition constant, regardless of the request rate variability, while with flow related blocking probabilities, the blocked traffic increases significantly, regardless of the absolute value of the blocking probability. If the blocked traffic volume is large enough, this can even influence the required overall capacity. Figure 11 shows that the required capacity increases with the variability of the rate distribution but the values for p_c and p_f deviate. Finally, the curve for $p_f = 10^{-1}$ decreases because mostly large requests are blocked. This is not an intended result. Therefore, we consider the use of p_c more reasonable and apply it in our framework as blocking criterion.

3.3 Design Options for NAC Performance Evaluation

The objective of a network provider is the satisfaction of his customers at minimum expenses. If flows request QoS from the network they should get it instead of being rejected by the NAC, which is the preferred action in QoS networks to avoid congestion in overload situations. Hence, enough capacity must be provided to cover the transmission demand of the flows which is characterized by an average load a and their request size. This should be achieved at least costs. The required link capacities are either capital or operational expenses for the network provider. There are various possibilities for the performance evaluation of NAC approaches that come from the relation among the load, the blocking probability, and the capacity. Two

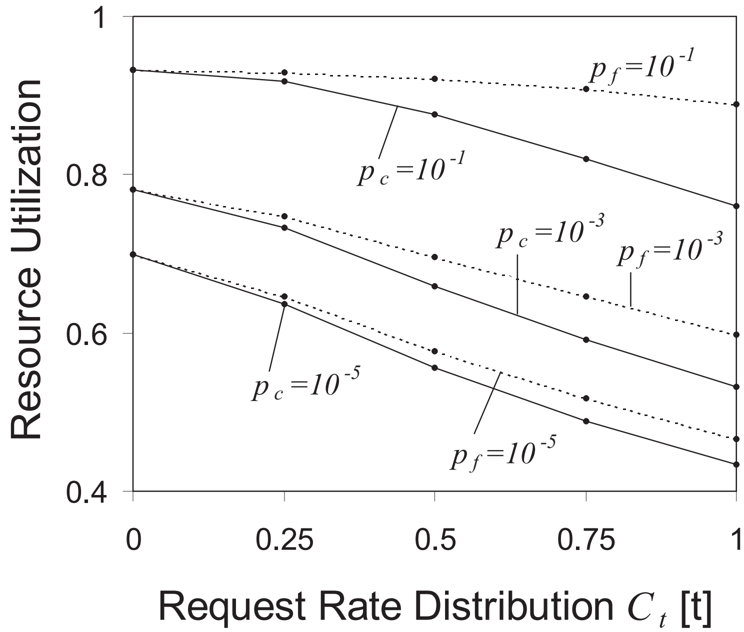

Figure 7: Impact of request rate variability and blocking probability on the resource utilization.

out of them condition the third term. We have listed the resulting experiment designs in Table 2 and discuss them in the following.

Table 2: Three design options for NAC performance investigation.

influencing term	design option 0	design option 1	design option 2
average offered traffic load a	variable	given	given
blocking probabilities $g(v, w).p$	given	variable	given
link capacities	given	given	variable

Design option 0 In design option 0 the network with all its link capacities is given and a given b2b blocking probability p_{b2b} must be met for all traffic aggregates. The offered load in the b2b traffic matrix is the variable parameter being part of the traffic model which determines, e.g., the average path length. The traffic matrix has many degrees of freedom, and so its assignment is difficult. Furthermore, the structure of the traffic matrix influences the potential economy of scale that can be achieved by different NAC methods and it influences herewith the achievable resource utilization of these NAC schemes. For all these reasons, this design option leads to many difficulties and to an unfair NAC comparison. Apart from that,

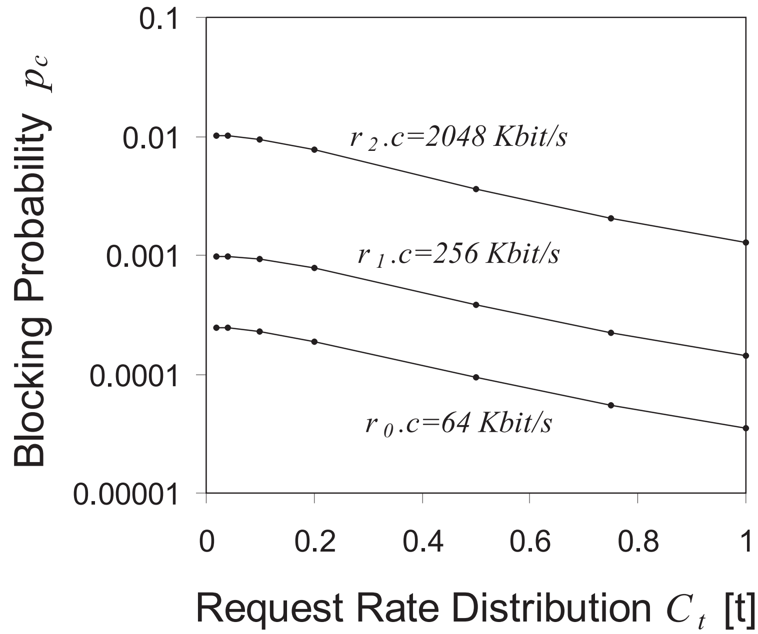

Figure 8: Impact of request rate variability on the rate specific blocking probability for $p_c = 10^{-3}$.

real networks are rather dimensioned to satisfy the offered load than vice versa.

Design option 1 Design option 1 provides the network with all its link capacities and the traffic matrix. Hence, the b2b blocking probabilities $g(v, w).p$ are to be determined. This is the normal case for operational networks. However, appropriate settings for the fixed parameters must be found to achieve reasonable b2b blocking probabilities, which complicates the investigation. Furthermore, an "appropriate" setting depends on the NAC mechanism itself such that the comparability of different NAC methods is not guaranteed by this design option. If different b2b aggregates experience different blocking probabilities, the comparison of different NAC methods becomes even more difficult. If a common minimum blocking probability must be found for all b2b relationships, some link capacities might be partly unused. Hence, there are many obstacles complicating the use of this design option.

Design option 2 In design option 2 the traffic matrix is given and the link capacities are determined to meet a required b2b aggregate blocking probability. With this approach, the above mentioned problems do not exist. Therefore, we use it as the methodology for NAC performance comparison.

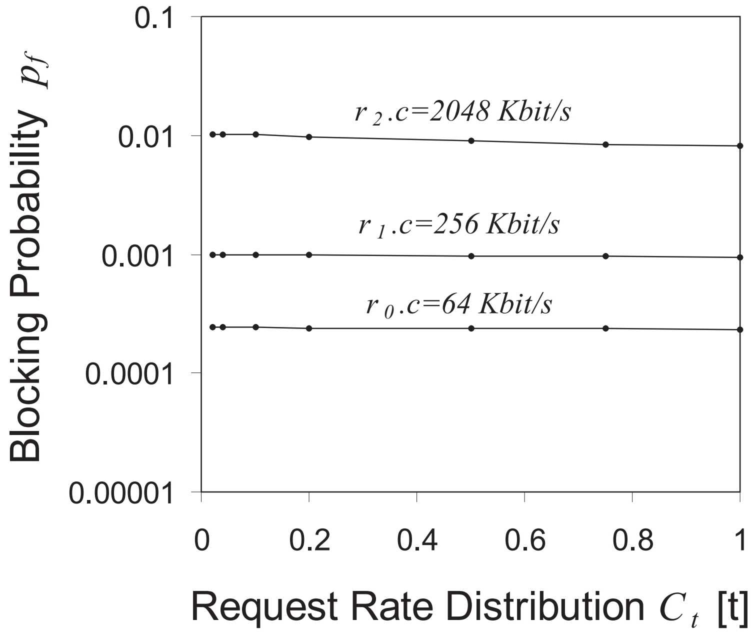

Figure 9: Impact of request rate variability on the rate specific blocking probability for $p_f = 10^{-3}$.

3.4 NAC Specific Capacity Dimensioning for Networks

AC guarantees QoS for admitted flows at the expense of flow blocking if the budget capacity is exhausted. To keep the blocking probability $b.p$ small for the offered traffic load $b.a$, the capacity $b.c$ of a budget b must be dimensioned large enough which influences the required capacity of the links in the network. For a possible traffic pattern[5] $g.c \in \mathbb{R}_0^{+|\mathcal{V}|^2}$ the following formulae hold

$$\begin{aligned} \forall v, w \in \mathcal{V} &: \quad g(v,w).c \geq 0 \\ \forall v \in \mathcal{V} &: \quad g(v,v).c = 0. \end{aligned} \tag{11}$$

If NAC is applied in the network, each traffic pattern $g.c$ satisfies the constraints defined by the NAC budgets. These constraints lead to linear equations, too, serving as side conditions for the worst case scenario in terms of rate maximization on a link l to determine its minimum capacity $l.c$

$$l.c \quad \geq \max_{g.c \in \mathbb{R}_0^{+|\mathcal{V}|^2}} \sum_{v,w \in \mathcal{V}} g(v,w).c \cdot l.u(v,w). \tag{12}$$

[5]We denote the offered load for a b2b aggregate $g(v,w)$ by $g(v,w).a$ and the resulting matrix $g.a = \left(g(v,w).a\right)_{v,w \in \mathcal{V}}$ is the traffic matrix. In contrast, the current requested rate of an aggregate is $g(v,w).c$ and the matrix $g.c = \left(g(v,w).c\right)_{v,w \in \mathcal{V}}$ describes an instantaneous traffic pattern.

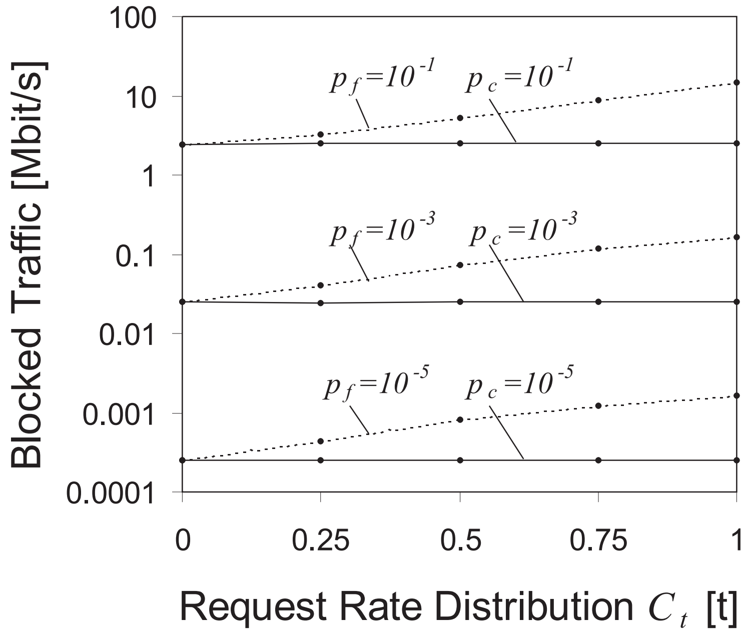

Figure 10: Impact of request rate variability and blocking probability on the blocked traffic.

Since the aggregate rates have real values, the maximization can be performed by the Simplex algorithm [24] in polynomial time. However, for some NACs there are more efficient solutions that we will point out in the following.

3.4.1 LB NAC

All flows traversing link l are covered by the single link budget $LB(l)$. Hence, the expected offered load for budget $LB(l)$ is

$$LB(l).a = \sum_{v,w \in \mathcal{V}} g(v,w).a \cdot l.u(v,w). \tag{13}$$

and the minimum capacity $l.c$ of link l is constrained by

$$l.c \geq LB(l).c. \tag{14}$$

3.4.2 IB/EB NAC

The IB/EB NAC subsumes all flows with the same ingress router v under $IB(v)$ and all flows with the same egress router w under $EB(w)$. The offered load of the respective budgets is

$$IB(v).a = \sum_{w \in \mathcal{V}} g(v,w).a, \text{ and } EB(w).a = \sum_{v \in \mathcal{V}} g(v,w).a. \tag{15}$$

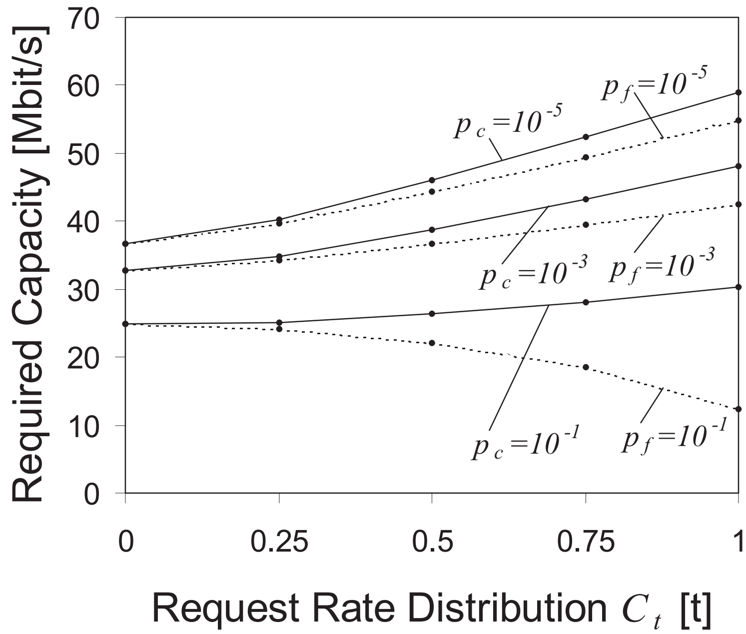

Figure 11: Impact of request rate variability and blocking probability on the required capacity.

Here we use the Simplex method for the computation of the capacity $l.c$ with the side conditions

$$\begin{aligned} \forall v \in \mathcal{V} : \sum_{w \in \mathcal{V}} g(v,w).c &\leq IB(v).c, \text{ and} \\ \forall w \in \mathcal{V} : \sum_{v \in \mathcal{V}} g(v,w).c &\leq EB(w).c. \end{aligned} \tag{16}$$

For the IB NAC, the IBs are computed in the same way like above, however, there is a computational shortcut to the Simplex method for the calculation of the required link capacity $l.c$:

$$l.c \geq \sum_{v \in \mathcal{V}} IB(v).c \cdot \sum_{w \in \mathcal{V}} l.u(v,w) \tag{17}$$

3.4.3 BBB NAC

The BBB NAC subsumes under $BBB(v,w)$ all flows with ingress router v and egress router w. The offered load for $BBB(v,w)$ is simply

$$BBB(v,w).a = g(v,w).a. \tag{18}$$

and the minimum capacity $l.c$ of link l is constrained by

$$l.c \geq \sum_{v,w \in \mathcal{V}} BBB(v,w).c \cdot l.u(v,w) \tag{19}$$

3.5 Performance Measure for NAC Comparison

The link capacities in a network are computed according to the equations above. The required network capacity $\mathcal{N}.c$ is the sum of all link capacities. The overall transmitted traffic rate $\mathcal{N}.c_{trans}$ is the sum of the offered load of all b2b aggregates weighted by their average path lengths $g(v,w).avgPathLen$, their acceptance probability $(1-p_{b2b})$, and the mean request rate $E[r_i.c]_{0\leq i<n_r}$. We can neglect the fact that requests with a larger rate have a higher blocking probability due to the construction in Equation (10).

$$\begin{aligned}\mathcal{N}.c &= \sum_{l\in\mathcal{E}} l.c \\ \mathcal{N}.c_{trans} &= (1-p_{b2b})\cdot E[r_i.c]_{0\leq i<n_r}\cdot \\ &\sum_{\{(v,w):v,w\in\mathcal{V},v\neq w\}} g(v,w).a\cdot g(v,w).avgPathLen \\ \mathcal{N}.\rho &= \frac{\mathcal{N}.c_{trans}}{\mathcal{N}.c}.\end{aligned} \quad (20)$$

The overall resource utilization $\mathcal{N}.\rho$ is the fraction of the transmitted traffic rate and the overall network capacity, and we use it as the performance measure for NAC methods.

3.6 From B2B Blocking Probabilities to Budget Blocking Probabilities

Finally, there is still an open question. How should the budget blocking probabilities $b.p$ be set to meet a desired blocking probability $g.p \leq p_{b2b}$ for a b2b traffic aggregate g? We consider different methods and evaluate them subsequently.

3.6.1 Methods for Setting Budget Blocking Probabilities

Budget sizes are dimensioned for a desired budget blocking probability $b.p$. The set $\mathcal{D}(g)$ consists of the budgets whose capacity needs to be checked for the NAC of an aggregate g. The blocking probability for an aggregate g is then

$$g.p = 1 - \Pi_{b\in\mathcal{D}(g)}(1-b.p) \quad (21)$$

under the assumption that the $b.p$ are independent of each other. Since the blocking probabilities of different budgets tend to be positively correlated if the network is well provisioned, the computation of $g.p$ according to Equation (21) is rather conservative. We further assume that the blocking probabilities of all budgets passed by a flow are equal. Hence, the budget blocking probability $b.p$ at each budget should be

$$b.p = 1 - \sqrt[|\mathcal{D}(g)|]{1-g.p}. \quad (22)$$

In case of the BBB NAC and IB NAC, $\mathcal{D}(g)$ contains a single budget, in case of IB/EB NAC, it contains two budgets, so the computation is easy. However, in case of the LB NAC it contains as many budgets as there are links in the path $g.path$ and this is a number depending on g. Since the number of traversed links is variable, the setting of $LB(l).p$ is difficult. In the following, we derive a simple solution to that problem.

We assume single path routing to avoid complexity. Then, the path $g.path$ of an aggregate is unique and its length $g.len$ is the number of links in the unique path. But still, different aggregates have different path length and, therefore, they require different blocking probabilities $g.p(l)$ on commonly used links. However, the NAC instance for every link l can provide only one link budget blocking probability $LB(l).p$ for all aggregates. We consider now three different methods to set the link blocking probabilities in an adequate way such that $g.p \leq p_{b2b}$ is met.

Method A: Global Minimum of $g.p(l)$ We consider the overall network and take the minimum required blocking probability $\min_{\{g\in\mathcal{G},l\in\mathcal{E}:l\in g.path\}}(g.p(l))$ of all aggregates on all links in the network as a common link blocking probability $LB(l).p$ for all links $l \in \mathcal{E}$. This entails that short distance traffic aggregates have a significantly lower b2b blocking probability $g.p$ than p_{b2b} at the expense of more required network capacity.

Method B: Local Minimum of Independent $g.p(l)$ We consider one link $l^* \in \mathcal{E}$ after another. We take the minimum required blocking probability $\min_{\{g\in\mathcal{G}:l^*\in g.path\}}(g.p(l^*))$ as the link budget blocking probability $LB(l^*).p$. This entails also that short distance traffic aggregates have also a lower b2b blocking probability $g.p$ than p_{b2b} at the expense of more required network capacity but some $LB(l)$ may have a lower blocking probability than with method A. Here, the link blocking probabilities are still set independently of each other.

Method C: Local Minimum of Adapted $g.p(l)$ This method is very similar to method B but we try to take advantage of the following fact. If $LB(l).p < g.p(l)$, $g.p(l^*)$ may be slightly increased on other links l^* which could possibly increase $LB(l^*).p$, too, and reduce the required capacity for those links. Algorithm 1 proposes a method for that.

The set of all links whose link budget blocking probability is not assigned, yet, is initially the set of all links $\mathcal{Q} = \mathcal{E}$. The aggregate blocking probability $g.p$ of all traffic aggregates $g \in \mathcal{G}$ is set to p_{b2b} and their link blocking probabilities are set according to Equation (22). Based on this, the temporary link budget blocking probability $LB(l).p_{tmp}$ is set to the minimum $g.p(l)$ for the link l and $g \in \{g^* \in \mathcal{G} : l \in g^*.path\}$. This is basically the initialization of the $LB(l).p_{tmp}$ that correspond to the $LB(l).p$ in method B. Then, the link budget blocking probabilities $LB(l).p$ for all links are determined one after another until the set of links with unassigned link budget blocking probabilities $\mathcal{Q}$ is empty. A link $l \in \mathcal{Q}$ with the lowest required temporary link blocking probability is chosen, its temporary blocking probability becomes its fixed blocking probability, and it is removed from $\mathcal{Q}$. Since $LB(l).p$ is possibly smaller than $g.p(l)$ for aggregates traversing this link, their remaining required blocking probabilities $g.p$ are adjusted by $g.p = 1 - \frac{1-g.p}{1-LB(l).p}$. As a consequence, their required link blocking probabilities $g.p(l)$ are again set according to Equation (22), which reduces them possibly. Therefore, the corresponding temporary minimum link budget blocking probabilities $LB(l).p_{tmp}$ are adapted, too. The output of the algorithm are the link budget blocking probabilities that support the required aggregate blocking probabilities $g.p$ without wasting too much capacity.

Input: Network $(\mathcal{V}, \mathcal{E})$ with aggregates $\mathcal{G}$ and their routing given by $g.path$, target b2b blocking probability p_{b2b}

```
Q := E {links without fixed blocking probability}
for all {g ∈ G} do   {initial setting of g.p(l)}
  g.p := p_b2b
  for all {l ∈ g.paths} do
    g.p(l) := 1 − (1 − g.p)^(1/g.len)
  end for
end for
for all l ∈ Q do
  LB(l).p_tmp := min_{g∈G: l∈g.path}(g.p(l))
end for
while Q ≠ ∅ do
  choose l ∈ Q : LB(l).p_tmp = min_{l*∈Q}(LB(l*).p_tmp)
  LB(l).p := LB(l).p_tmp
  Q := Q \ l
  for all {g ∈ G : l ∈ g.path} do   {modify g.p(l)}
    g.len := g.len − 1
    g.p := 1 − (1 − g.p)/(1 − LB(l).p)
    for all {l* ∈ Q : l* ∈ g.path} do
      g.p(l*) := 1 − (1 − g.p)^(1/g.len)
      LB(l*).p_tmp := min(g.p(l*), LB(l*).p_tmp)
    end for
  end for
end while
```

Output: $\{LB(l).p : l \in \mathcal{E}\}$

Algorithm 1: Method C: Local Minimum of Adapted $g.p(l)$.

3.6.2 The Impact of the Setting of the Budget Blocking Probabilities

We proposed three different methods for the determination of the required budget blocking probabilities to achieve a desired b2b blocking probability. For a better understanding of the three methods we consider the example network given in Figure 12. The longest path in the network has n hops. Method A requires $LB(l).p = 1 - \sqrt[n]{1 - p_{b2b}}$ for all budgets. The longest path in which l_0 is involved has only $\frac{n}{2} + 1$ hops instead. Therefore, the result of method B deviates in the link budget blocking probability for l_0, which is $LB(l_0).p = 1 - \sqrt[\frac{n}{2}+1]{1 - p_{b2b}}$. Method C also distinguishes therein and yields $LB(l_0).p = 1 - \sqrt[2]{1 - p_{b2b}}$ because the success probability for AC from one far end to l_0 is already $(\sqrt[n]{1 - p_{b2b}})^{\frac{n}{2}} = \sqrt[2]{1 - p_{b2b}}$.

From Figure 6 we know that the blocking probability has only an impact on the resource utilization if the offered aggregate load is small, therefore, we choose an offered load $a = 10$ on the link l_0. Figure 13 shows that the resource utilization for method C is independent of n but for methods A and B it decreases steadily with increasing n. This example illustrates

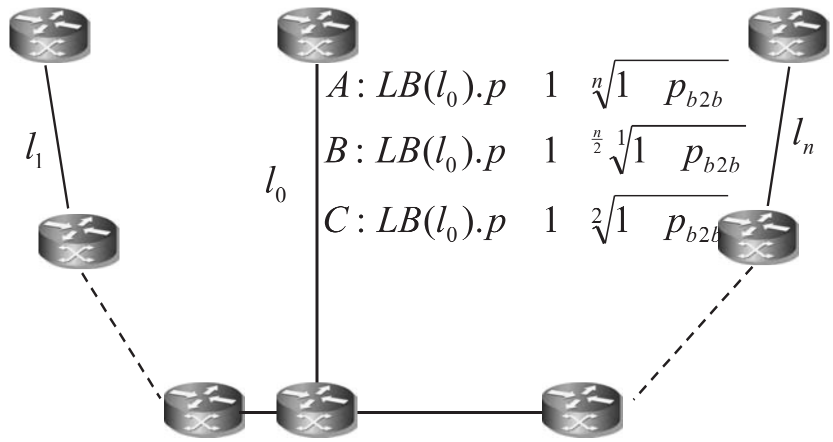

Figure 12: The setting of the link blocking probabilities has a large impact on $LB(l_0).p$.

that the method to set the link blocking probabilities can have an impact on the required link capacity. However, the topology of Figure 12 is very pathologic. Therefore, we construct networks of size 20, 40, 60, 80, and 100, randomly. We set the b2b load $a_{b2b} = 10$ for each $g.a$ and set the link blocking probabilities according to method A, B, and C. The difference between method A and B in the required overall network capacity is at most 0.2% and there is no difference at all between method B and C.

We have proposed different methods to configure the link budget blocking probabilities. For very special topologies their impact is visible but for randomly constructed networks they do not differ at all. Therefore, we use method B in the following since it optimizes the link blocking probabilities with an acceptable computational overhead.

4 Performance Comparison of Basic NAC Approaches

In this section, we give some numerical examples based on our performance evaluation framework.

The topology of our reference network depicted in Figure 14 is based on the UUNET in 1994 [25] where nodes connected by only one or two links were successively removed. The resulting network has $|\mathcal{V}|=20$ routers and $|\mathcal{E}|=51$ links. We assume an average b2b load a_{b2b} between two cities leading to an overall offered load $a_{tot}=a_{b2b}\cdot|\mathcal{V}|\cdot(|\mathcal{V}|-1)$. We construct the traffic matrix $g.a$ in terms of offered load proportionally to the city sizes π which is given in Figure 15

$$g(v,w).a=\begin{cases} a_{tot}\cdot\frac{\pi(v)\cdot\pi(w)}{\sum_{x,y\in\mathcal{V},x\neq y}\pi(x)\cdot\pi(y)} & \text{for } v\neq w,\\ 0 & \text{for } v=w.\end{cases} \quad (23)$$

Based on this traffic matrix and single shortest path routing, the resource utilization $\mathcal{N}.\rho$ is evaluated for a b2b blocking probability $p_{b2b} = 10^{-3}$ and request rate distribution C_1. The

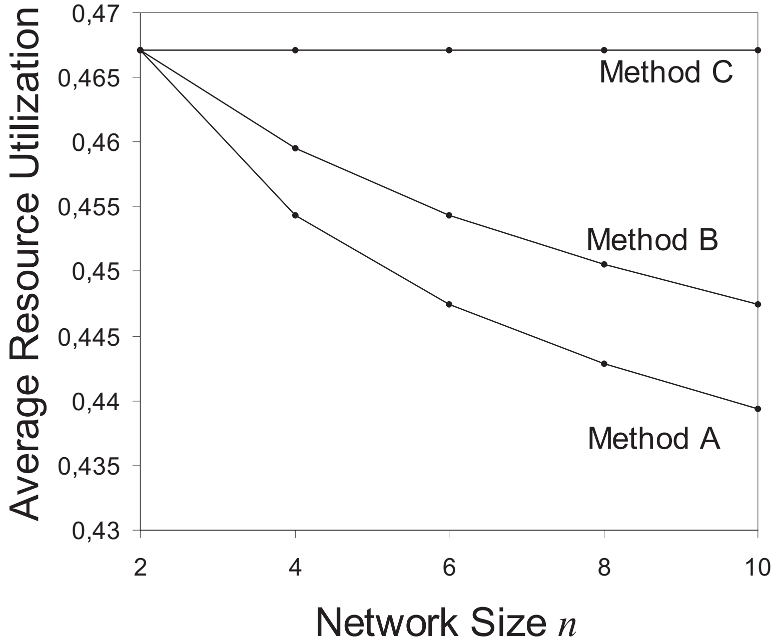

Figure 13: Impact of the setting of $LB(l_0).p$ on the resource utilization for link l_0. ($a_{b2b} = 10$, $p_{b2b} = 10^{-3}$).

results are shown in Figure 16 for all presented NAC methods depending on the offered b2b load a_{b2b}.

The LB NAC uses the network resources most efficiently. A budget $LB(l)$ controls a maximum possible amount of traffic on link l and takes most advantage from economy of scale. The BBB NAC is less efficient because the offered load is partitioned among up to $|\mathcal{V}| \cdot (|\mathcal{V}| - 1)$ different budgets, leading to a larger capacity requirement for the same resource $l.c$. For sufficiently high offered load, the utilization of the LB and the BBB NAC tends towards 100%. The IB NAC has the worst performance (10%) and our IB/EB NAC achieves a three times larger resource utilization (30%) by applying the limitation of the traffic volume in a symmetric way. They both are not able to exclude unlikely traffic patterns which force to allocate high link capacities.

From our experiments on a single link we can conclude that the resource utilization rises for larger blocking probabilities and less variable request size distributions, and in particular, for little offered load. However, this affects all NAC methods and does not change the fundamental differences in resource efficiency.

5 Conclusion

We have introduced the notion of *link* admission control (LAC) and *network* admission control (NAC). LAC limits the number of flows on a link to assure their QoS requirements while

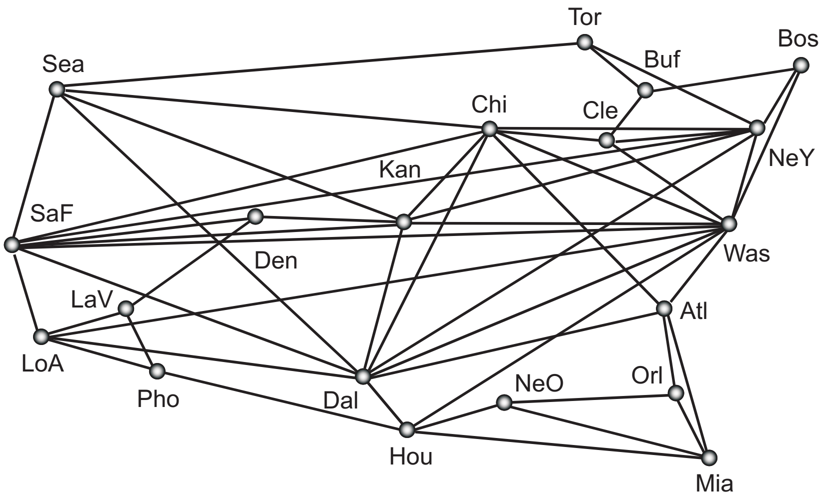

Figure 14: Test network based on the UUNET (1994).

NAC limits the number of flows in a network. We presented three basic NAC methods: the link budget (LB) based NAC, the tunnel budget (BBB) based NAC, which consists of virtual tunnels, and the ingress and egress budget (IB/EB) based NAC, known from the Differentiated Services context. Many research projects implement admission control (AC) schemes that can be classified by these categories.

In this paper we established a framework for a performance comparison of different NAC approaches. The performance measure is the average network resource utilization which is obtained by a suitable network dimensioning for a given traffic matrix and desired border-to-border (b2b) blocking probability. First, we studied the influence of several parameters on the resource utilization of a single link. This is important because the economy of scale on a single link is the key for understanding the performance of NAC methods. The resource utilization mainly depends on the offered load. If the offered load is high, the request rate variability and the blocking probability have a minor impact on the resource utilization. For multi-rate traffic, the blocking probability of a flow depends on its request size. We preferred a transmission rate related definition of the blocking probability because with flow related blocking probabilities the required link capacity can decrease for an increasing request rate variability which is an unintended phenomenon for performance comparisons. We extended the dimensioning process from a single link to distributed NAC budgets in a network and to link capacities. To that aim, we proposed three methods to map b2b blocking probabilities to budget blocking probabilities. Although they differ significantly in complexity, they yield approximately the same results for practical networking scenarios. The numerical comparison of NAC approaches in a given sample network showed that the LB NAC is more efficient than the BBB NAC but for large offered load, they both achieve a resource utilization close to 100%. In contrast, the IB NAC and the IB/EB NAC performance is limited by about 10% and 30% resource utilization.

Currently, we extend this study towards new NAC approaches. In addition, we use our

Name(v)	*(v) [10³]*	*Name(v)*	*(v) [10³]*
Atlanta	4112	Los Angeles	9519
Boston	3407	Miami	2253
Buffalo	1170	New Orleans	1338
Chicago	8273	New York	9314
Cleveland	2250	Orlando	1645
Dallas	3519	Phoenix	3252
Denver	2109	San Francisco	1731
Houston	4177	Seattle	2414
Kansas	1776	Toronto	4680
Las Vegas	1536	Washington	4923

Figure 15: Population of the cities and their surroundings.

framework to test the sensitivity of these methods to network topology [**?**], traffic distribution, and routing [**?**]. For NAC budget configuration, the inverse process of dimensioning is required [**?**]. Resilience mechanisms, e.g. fast rerouting, can detour traffic in a packet-switched network in case of partial network outages. Networks can be dimensioned for normal network operation and partial outages such that the QoS is not affected. Under this aspect, the achievable resource utilization of the NACs will be different [**?**].

References

[1] S. Shenker, “Fundamental Design Issues for the Future Internet,” *IEEE JSAC*, vol. 13, pp. 1176–1188, Sep. 1995.

[2] H. De Meer, P. O’Hanlon, G. Feher, N. Blefari-Melazzi, H. Tschofenig, G. Karagiannis, D. Partain, V. Rexhepi, and L. Westberg, “Analysis of Existing QoS Solutions.” http://www.ietf.org/internet-drafts/draft-demeer-nsis-analysis-02.txt, June 2002.

[3] A. Riedl, T. Bauschert, and J. Frings, “A Framework for Multi-Service IP Network Planning,” in 10^{th} *International Telecommunication Network Strategy and Planning Symposium (Networks2002)*, (Munich, Germany), pp. 183 – 190, June 2002.

[4] J. Roberts, U. Mocci, and J. Virtamo, *Broadband Network Teletraffic - Final Report of Action COST 242*. Berlin, Heidelberg: Springer, 1996.

[5] L. Breslau, E. W. Knightly, S. Shenker, and H. Zhang, “Endpoint Admission Control: Architectural Issues and Performance,” in *ACM SIGCOMM*, Aug. 2000.

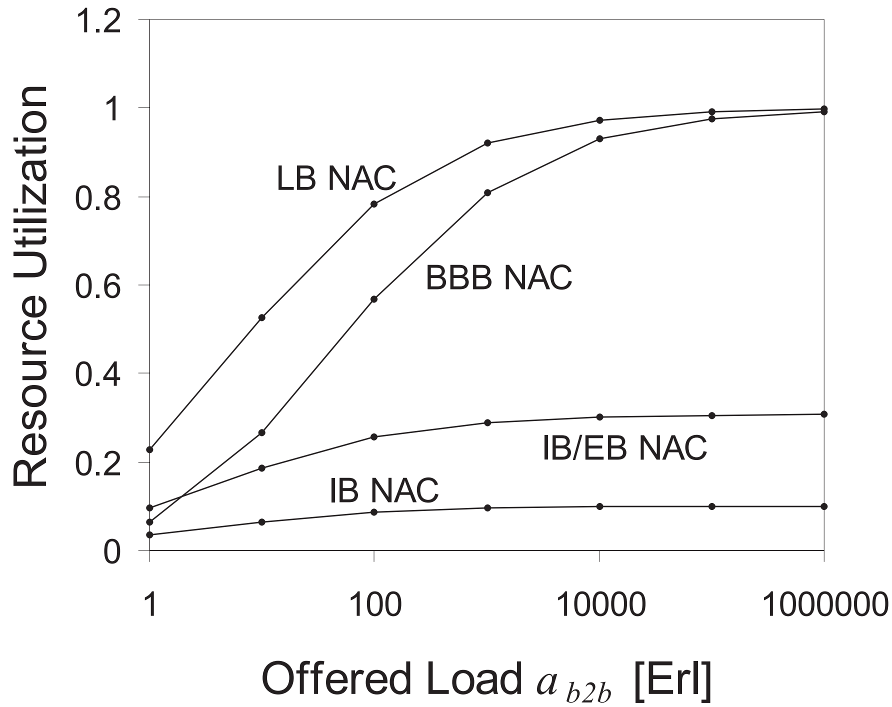

Figure 16: Impact of the offered load and the NAC method on the resource utilization.

[6] L. Breslau, S. Jamin, and S. Shenker, "Comments on the Performance of Measurement-Based Admission Control Algorithms," in *Infocom*, pp. 1233–1242, March 2000. ISBN 0-7803-5880-5.

[7] F. P. Kelly, "Notes on Effective Bandwidths," *Stochastic Networks: Theory and Applications*, vol. 4, pp. 141–168, 1996.

[8] The ATM Forum, *Traffic Management Specification, Version 4.0*, April 1996.

[9] B. Braden, D. Clark, and S. Shenker, "RFC1633: Integrated Services in the Internet Architecture: an Overview." http://www.ietf.org/rfc/rfc1633.txt, June 1994.

[10] A. Terzis, J. Wang, J. Ogawa, and L. Zhang, "A Two-Tier Resource Management Model for the Internet," in *Global Internet Symposium'99*, Dec. 1999.

[11] B. Teitelbaum, S. Hares, L. Dunn, V. Narayan, R. Neilson, and F. Reichmeyer, "Internet2 QBone: Building a Testbed for Differentiated Services," *IEEE Network Magazine*, Sep. 1999.

[12] Z.-L. Z. Zhang, Z. Duan, and Y. T. Hou, "On Scalable Design of Bandwidth Brokers," *IEICE Transaction on Communications*, vol. E84-B, pp. 2011–2025, Aug 2001.

[13] I. Stoica and H. Zhang, "Providing Guaranteed Services Without Per Flow Management," *Computer Communication Review*, vol. 29, October 1999.

[14] S. Bhatnagar and B. Nath, "Distributed Admission Control to Support Guaranteed Services in Core-Stateless Networks," in *IEEE INFOCOM 2003*, (San Francisco, USA), April 2003.

[15] R. Szábó, T. Henk, V. Rexhepi, and G. Karagiannis, "Resource Management in Differentiated Services (RMD) IP Networks," in *International Conference on Emerging Telecommunications Technologies and Applications (ICETA 2001)*, (Kosice, Slovak Republic), Oct. 2001.

[16] S. Blake, D. L. Black, M. A. Carlson, E. Davies, Z. Wang, and W. Weiss, "RFC2475: An Architecture for Differentiated Services." ftp://ftp.isi.edu/in-notes/rfc2475.txt, Dec. 1998.

[17] X. Xiao and L. M. Ni, "Internet QoS: A Big Picture," *IEEE Network Magazine*, vol. 13, pp. 8–18, Mar. 1999.

[18] T. Engel, E. Nikolouzou, F. Ricciato, and P. Sampatakos, "Analysis of Adaptive Resource Distribution Algorithsm in the Framework of a Dynamic DiffServ IP Network," in 8^{th} *International Conference on Advances in Communications and Control (ComCon8)*, (Crete, Greece), June 2001.

[19] H. Fu and E. Knightly, "Aggregation and Scalable QoS: A Performance Study," in *Proceedings of IWQoS 2001*, (Karlsruhe, Germany), June 2001.

[20] K. Kompella and Y. Rekhter, "LSP Hierarchy with Generalized MPLS TE." http://www.ietf.org/internet-drafts/draft-ietf-mpls-lsp-hierarchy-08.txt, March 2002.

[21] F. Baker, C. Iturralde, F. Le Faucheur, and B. Davie, "RFC3175: Aggregation of RSVP for IPv4 and IPv6 Reservations." http://www.ietf.org/rfc/rfc3175.txt, Sept. 2001.

[22] D. O. Awduche, L. Berger, D.-H. Gan, T. Li, V. Srinivasan, and G. Swallow, "RFC3209: RSVP-TE: Extensions to RSVP for LSP Tunnels." http://www.ietf.org/rfc/rfc3209.txt, Dec. 2001.

[23] P. Trimintzios, T. Bauge, G. Pavlou, L. Georgiadis, P. Flegkas, and R. Egan, "Quality of Service Provisioning for Supporting Premium Services in IP Networks," in *IEEE Globecom 2002*, (Taipei, Taiwan), Nov. 2002.

[24] J. Stoer, *Numerische Mathematik 1*. New York, Berlin, Heidelberg: Springer, 5th ed., June 1989.

[25] "Network Maps." http://www.nthelp.com/maps/.