A Performance Evaluation Framework for Network Admission Control Methods

Michael Menth, Stefan Kopf, Jens Milbrandt
Dept. of Distributed Systems, Inst. of Computer Science, Univ. of Würzburg, Germany
{*menth,kopf,milbrandt*}*@informatik.uni-wuerzburg.de*

Abstract

In this paper we introduce the notion of *link* and *network* admission control (LAC, NAC) and present three fundamentally different budget based NAC methods which categorize most of today's implemented NAC approaches. We propose a performance evaluation framework for their comparison. The required network capacity for each method is dimensioned for a certain flow blocking probability and the average resource utilization is taken as performance measure. We point out several implementation options and investigate their impact. Based on numerical results we give recommendations for preferred procedures. Finally, we compare different NAC methods under varying load conditions.

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 [17]. 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 and ease their understanding [13].

These NAC approaches have different complexity and resource efficiency but there is no numerical performance comparison in the literature. The framework of this paper closes this gap. Our work considers different design options for such studies and recommends a preferred procedure. First, we dimension the NAC budgets based on a given traffic matrix and routing such that desired border-to-border (b2b) flow blocking prob-

*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.

abilities 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 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.

Network admission control (NAC) needs to protect more than one link with one 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 border-to-border (b2b) flow blocking probability p_{b2b} which affects the network operator's costs.

NAC and LAC can be combined, i.e. a flow's required capacity $c(f)$ may consist of an effective bandwidth to take burstiness and/or 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 work on effective bandwidth budgets and 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 $\mathcal{F}(b)$ in place related to a budget b. When a new flow arrives, it checks whether its effective bandwidth together with the demand of already established flows fits within the capacity budget. If so, the flow is accepted, otherwise it is rejected. This principle is used in link based admission control, controlling one link, as well as as in NAC where a number of network resources are covered by each budget and at the same time the utilization of one resource is affected by a number of budgets.

2.2 Link Budget Based Network Admission Control (LB NAC)

The link-by-link NAC is probably the most intuitive NAC approach. The capacity $c(l)$ of each link l in the network is managed by a single link budget LB_l (with size $c(LB_l)$) that may be administered, e.g., at the router sending over that link or in a centralized database. 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 $u_l(g_{v,w})$ indicates the percentage of the traffic rate $c(g_{v,w})$ using link l. It is able to reflect both single- and multi-path routing. A new flow $f_{v,w}^{new}$ with ingress router v, egress router w, and bitrate $c(f_{v,w}^{new})$ must pass the AC procedure for the LBs of all links that are traversed in the network by $f_{v,w}^{new}$ (cf. Figure 1(a)). The NAC procedure will be successful if the following inequality holds

$$\forall l \in \mathcal{E} : u_l(g_{v,w}) \quad > \quad 0 : c(f_{v,w}^{new}) \cdot u_l(g_{v,w}) + \sum_{f_{x,y} \in \mathcal{F}(LB_l)} c(f_{x,y}) \cdot u_l(g_{x,y}) \leq c(LB_l). \tag{1}$$

There are many systems and protocols working according to that principle. The connection AC in ATM [1] and the Integrated Services [6] architecture proposed for IP adopt it in pure form and induce per flow reservation states in the core. Other architectures reveal the same behavior although the mechanism is not implemented as an explicit LB NAC. A bandwidth broker [21, 20, 24] administers the budgets in a central database. The stateless core approaches [18, 4, 19] 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 three 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 traffic rerouting is required and the QoS is maintained if sufficient backup capacity is available.

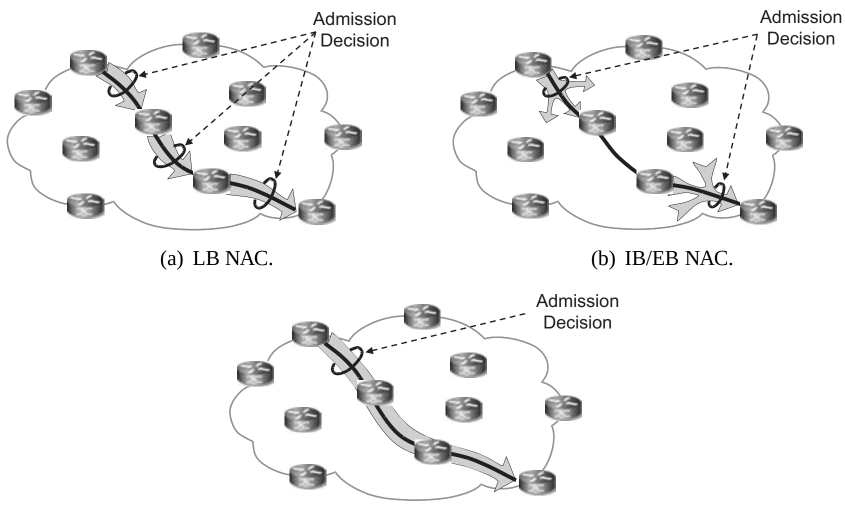
2.3 Ingress and Egress Budget Based Network Admission Control (IB/EB NAC)

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 $f_{v,w}^{new}$ must pass the AC procedure for IB_v and EB_w and it is only admitted if both requests are successful (cf. Figure 1(b)). Hence, the following inequalities must hold

$$c(f_{v,w}^{new}) + \sum_{f \in \mathcal{F}(IB_v)} c(f) \leq c(IB_v) \quad \text{and} \quad c(f_{v,w}^{new}) + \sum_{f \in \mathcal{F}(EB_w)} c(f) \leq c(EB_w) \tag{2}$$

Flows are admitted at the ingress irrespective of their egress router and at their egress router irrespective of their ingress routers, i.e. both AC decisions are decoupled. This entails that the capacity managed by an IB or EB can be used in a very flexible manner. However, the network must be able to carry all – also pathological – combinations of traffic patterns that are admissible by the IBs and EBs with the required QoS. Hence, sufficient capacity must be allocated or the IBs and EBs must be set small enough.

If we leave the EBs aside, we get the simple IB NAC, so only the left part of Equation (2) is checked for the AC procedure. This idea fits within the DiffServ context [5, 23]



(a) LB NAC. (b) IB/EB NAC.

(c) BBB NAC.

Figure 1: Budget based network admission control (NAC) methods.

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.

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 decision, i.e. a b2b budget $BBB_{v,w}$ manages the capacity of a virtual tunnel between v and w. This tunnel can consist of multiple b2b paths if multi-path routing is used. Figure 1(c) illustrates that a new flow $f_{v,w}^{new}$ passes only a single AC procedure for $BBB_{v,w}$. It is admitted if the following inequality holds

$$c(f_{v,w}^{new}) + \sum_{f \in \mathcal{F}(BBB_{v,w})} c(f) \leq c(BBB_{v,w}). \tag{3}$$

The BBB NAC can also avoid states inside the network because the $BBB_{v,w}$ may be controlled at the ingress or egress router. 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 [7, 8]. Tunnels may also be used hierarchically [9]. The tunnel capacity may be signaled using explicit reservation states in the network [3, 2], only in logical entities like bandwidth brokers [20], or it may be assigned by a central entity [22].

3. A Performance Evaluation Framework for NAC Approaches

In this section we propose a dimensioning method for multi-rate traffic and study it under various circumstances on a single link. This illustrates economy of scale which is the key for understanding NAC performance. We point out design options for the assessment of NAC performance and explain the budget and link capacity computation for different NAC types. This constitutes our performance evaluation framework.

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*. This is achieved by flow blocking if the network is highly loaded. Capacity dimensioning is a function calculating the required bandwidth for given traffic characteristics and a desired blocking probability. The specific implementation of that function depends on the underlying traffic model. We assume Poisson arrivals of resource requests and a generally distributed holding time. Although typical Internet traffic has different characteristics on the packet level [16], the Poisson model, which is used in the telephony world, is more realistic for the resource request level of end-user driven real-time applications.

In addition, we are rather interested in a basic performance comparison of the NAC methods than in the capacity dimensioning for a specific network service with known traffic profiles. The offered load a is the mean number of active flows, provided that no flow blocking occurs. In a multi-service world like the Internet, the request profile is multi-rate, so we take n_r different request types r_i, $0 \leq i < n_r$ with a bitrate $c(r_i)$. Given an offered load a, the respective request type specific offered load is $a(r_i) = p_a(r_i) \cdot a$. In our studies, we assume a simplified multimedia real-time communication scenario with $n_r = 3$, $c(r_0) = 64$ Kbit/s and $p_a(r_0) = \frac{28}{31} \cdot t$, $c(r_1) = 256$ Kbit/s and $p_a(r_1) = 1 - t^2$, and $c(r_2) = 2048$ Kbit/s and $p_a(r_2) = \frac{3}{31} \cdot t$. The resulting mean bitrate is $E[C_t] = \sum_{0 \leq i < n_r} c(r_i) \cdot p_a(r_i) = 256$ Kbit/s and the coefficient of variation is $c_{var}[C_t] = 2.29 \cdot t$.

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 $p(r_i)$.

To that aim, we review the Kaufman & Roberts algorithm [17] 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 $u_c = gcd(\{c(r_i) : 0 \leq i < n_r\})^*$. The request rates are converted into multiples of this finest granularity by $c_u(r_i) = \frac{c(r_i)}{u_c}$ and so is the link capacity $c_u(l) = \frac{c(l)}{u_c}$. First, weights

* 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.

* gcd denotes the greatest common divisor.

are computed $w(j) = \begin{cases} 0 & : j < 0, \\ 1 & : j = 0, \\ \frac{1}{j} \cdot \sum_{0 \leq i < n_r} w(j - c_u(r_i)) \cdot c_u(r_i) \cdot a(r_i) & : 0 < j \leq c_u \end{cases}$. 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 $q(j) = \frac{w(j)}{\sum_{0 \leq i \leq c_u} w(i)}$ and $Q(j) = \sum_{0 \leq i \leq j} q(i)$. The blocking probability of a single request type r is $p_b(r) = 1 - Q(c_u - c_u(r))$. 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 $a \cdot E[C_t]$ (p_c). They are computed by the following equations

$$p_f = \sum_{0 \leq i < n_r} p_b(r_i) \cdot p_a(r_i) \quad \text{and} \tag{4}$$

$$p_c = \frac{\sum_{0 \leq i < n_r} p_b(r_i) \cdot p_a(r_i) \cdot c(r_i)}{E[C_t]} \tag{5}$$

In the sequel we show the advantage of p_c over p_f for performance comparison reasons.

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.

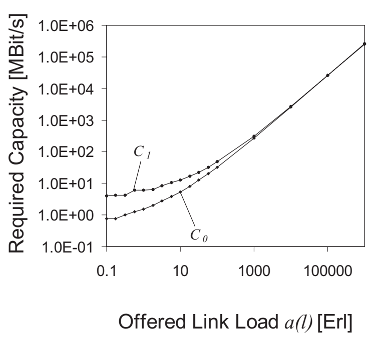


Figure 2: Impact of offered load and request rate variability on the required link capacity.

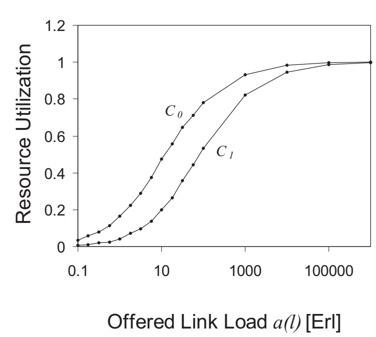


Figure 3: Impact of offered load and request rate variability on the resource utilization.

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 2 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 $a(l) = 1000$ Erlang or larger. Figure 3 illustrates that the resource utilization increases drastically up to an offered load of $a(l)=1000$ Erlang, hence, the resource utilization depends heavily on the offered link load $a(l)$ 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 real-time 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.

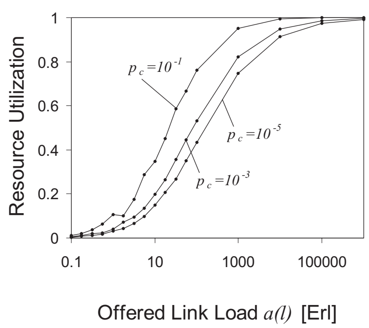


Figure 4: Impact of offered load and the blocking probability on the resource utilization.

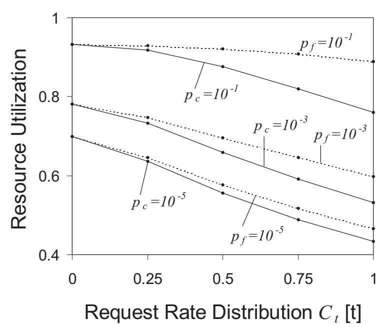


Figure 5: Impact of request rate variability and blocking probability on the resource utilization.

The Impact of the Link Blocking Probability

Figure 4 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 5 shows the resource utilization for $a(l) = 100$ 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.

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 $a(l)=100$ Erlang depending on the request rate variability. Figure 6 shows the request type specific blocking probabilities for blocking probability $p_c=10^{-3}$ which is related to the overall transmission rate while Figure 7 shows them for the flow specific blocking probability $p_f = 10^{-3}$. In both cases, request

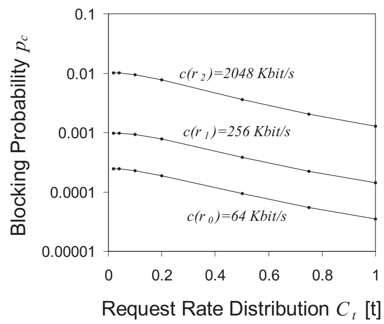


Figure 6: Impact of request rate variability on the rate specific blocking probability for $p_c = 10^{-3}$.

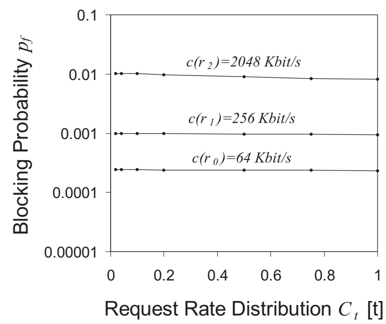


Figure 7: Impact of request rate variability on the rate specific blocking probability for $p_f = 10^{-3}$.

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 distribution variability if capacity dimensioning is based on p_c. This has some impact on the blocked traffic as illustrated in Figure 8. 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 9 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 out of them condition the third term.

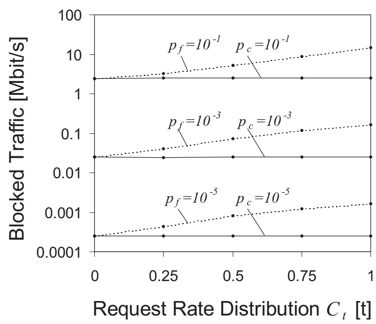


Figure 8: Impact of request rate variability and blocking probability on the blocked traffic.

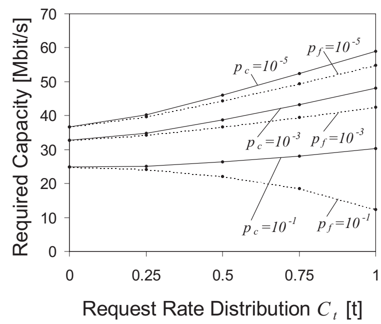


Figure 9: Impact of request rate variability and blocking probability on the required capacity.

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, 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 $p_b(a(g_{),v})w$ 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.

3.4 NAC Specific Capacity Dimensioning for Networks

In the following, we adapt the capacity dimensioning method for a single link to an entire network and respect the NAC method specific requirements.

From B2B Blocking Probabilities to Budget Blocking Probabilities

Budget sizes are dimensioned for a desired budget blocking probability $p_b(b)$. The set $\mathcal{B}_g$ consists of all budgets whose capacity needs to be checked if a flow of the traffic aggregate g asks for admission. The b2b blocking probability associated with this aggregate g is then

$$p_{b2b}(g) \quad = \quad 1 - \Pi_{b \in \mathcal{B}_g}(1 - p_b(b)). \tag{6}$$

under the assumption that flow blocking at different budgets is independent. Since flow blocking at different budgets tends to be positively correlated, the computation of $p_{b2b}(g)$ according to Equation (6) is rather conservative.

In [14] we have proposed three different methods for setting the budget blocking probabilities $p_b(b)$ to achieve a desired b2b flow blocking probability p_{b2b}. They have hardly any effect on the NAC performance, therefore, we stick with the simple approach that all $p_b(b)$ are equal for all budgets $b \in \mathcal{B}_g$. We denote by $m(b)$ the maximum number of budgets to be checked for any flow controlled by b. Then the required $p_b(b)$ is determined by $p_b(b) \leq 1 - \sqrt[m(b)]{1 - p_{b2b}}$.

3.5 Resource Allocation for Budget Based NAC Methods

We denote the offered load for a b2b aggregate $g_{v,w}$ by $a(g_{v,w})$. The resulting matrix $A_{\mathcal{G}} = \left(a(g_{v,w})\right)_{v,w \in \mathcal{V}}$ is the traffic matrix. In contrast, the current requested rate of an aggregate is $c(g_{v,w})$ and the matrix $C_{\mathcal{G}} = \left(c(g_{v,w})\right)_{v,w \in \mathcal{V}}$ describes an instantaneous traffic pattern. For a possible traffic pattern $C_{\mathcal{G}} \in \mathbb{R}_0^{+^{|\mathcal{V}|^2}}$ the formulae $\forall v, w \in \mathcal{V}$: $c(g_{v,w}) \geq 0 \quad$ and $\quad \forall v \in \mathcal{V} : c(g_{v,v}) = 0$ hold. If NAC is applied in the network, each traffic pattern $C_{\mathcal{G}}$ satisfies the constraints defined by the NAC budgets. These constraints lead to linear equations, too, serving as side conditions for the calculation of the worst case scenario on each link $l \in \mathcal{E}$ by the rate maximization $c(l) \geq \max_{C_{\mathcal{G}} \in \mathbb{R}_0^{+^{|\mathcal{V}|^2}}} \sum_{g \in \mathcal{G}} c(g) \cdot u_l(g)$. This determines the minimum required capacity $c(l)$ of link l. Since the aggregate rates have real values, the maximization can be performed by the Simplex algorithm in polynomial time. However, for some NAC methods there are more efficient solutions that we will point out in the following.

LB NAC

A transit flow needs to check a budget LB_l for every link l of its path for admission, hence, the maximum number of passed NAC budgets is $m(LB_l) = max_{\{g \in \mathcal{G}: u_l(g) > 0\}} len_{path}^{max}(g, l)$ whereby $len_{path}^{max}(g, l)$ is the maximum length of a path containing l used by g. As the budget LB_l covers all flows traversing link l, its expected offered load is $a(LB_l) = \sum_{g \in \mathcal{G}} a(g) \cdot u_l(g)$. According to Equation (1) $\forall l \in \mathcal{E} : \sum_{g \in \mathcal{G}} c(g) \cdot u_l(g) \leq c(LB_l)$ must be fulfilled, so the minimum capacity $c(l)$ of link l is constrained by $c(l) \geq c(LB_l)$.

IB/EB NAC

With the IB/EB NAC, a flow is admitted by checking both the ingress and the egress budget. Thus, we get $m(IB_v) = m(EB_w) = 2$. 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 $a(IB_v) = \sum_{w\in\mathcal{V}} a(g_{v,w})$ and $a(EB_w) = \sum_{v\in\mathcal{V}} a(g_{v,w})$. Here we use the inequalities from Equation (2) as side conditions in Simplex method for the computation of the capacity $c(l)$: $\forall v \in \mathcal{V} : \sum_{w\in\mathcal{V}} c(g_{v,w}) \leq c(IB_v)$ and $\forall w \in \mathcal{V} : \sum_{v\in\mathcal{V}} c(g_{v,w}) \leq c(EB_w)$. In case of the mere IB NAC, $m(IB_v) = 1$ holds. 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 $c(l)$: $c(l) \geq \sum_{v\in\mathcal{V}} c(IB_v) \cdot \sum_{w\in\mathcal{V}} u_l(g_{v,w})$.

BBB NAC

With the BBB NAC, only one budget is checked, therefore, $m(BBB_{v,w}) = 1$. 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 $a(BBB_{v,w}) = a(g_{v,w})$. Since Equation (3) is checked for admission $\forall v, w \in \mathcal{V} : c(g_{v,w}) \leq c(BBB_{v,w})$ must be fulfilled and the minimum capacity $c(l)$ of link l is constrained by $c(l) \geq \sum_{v,w\in\mathcal{V}} c(BBB_{v,w}) \cdot u_l(g_{v,w})$.

3.6 Performance Measure for NAC Comparison

We compute the required link capacities for all NAC methods according to the equations above. The required network capacity $c(\mathcal{N}) = \sum_{l\in\mathcal{E}} c(l)$ is the sum of all link capacities in the network. The overall transmitted traffic rate $\hat{c}(\mathcal{N}) = (1-p_{b2b}) \cdot E[C] \sum_{\{g\in\mathcal{G}\}} a(g) \cdot len^{avg}_{path}(g)$ is the sum of the offered load of all b2b aggregates g weighted by their average path lengths $len^{avg}_{path}(g)$, their acceptance probability $(1-p_{b2b})$, and the mean request rate $E[C]$. We can neglect the fact that requests with a larger rate have a higher blocking probability due to the construction in Equation (5). The overall resource utilization $\rho(\mathcal{N}) = \frac{\hat{c}(\mathcal{N})}{c(\mathcal{N})}$ is the fraction of the transmitted traffic rate and the overall network capacity. We use it in the next section as the performance measure for the comparison of NAC methods.

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 test network depicted in Figure 10 is based on the UUNET in 1994 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 11

$$g(v,w).a = \begin{cases} a_{tot} \cdot \frac{\pi(v)\cdot\pi(w)}{\sum_{x,y\in\mathcal{V},x\neq y} \pi(x)\cdot\pi(y)} & \text{for } v \neq w, \\ 0 & \text{for } v = w. \end{cases} \quad (7)$$

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 results are shown in Figure 12 for all presented NAC methods depending on the offered b2b load a_{b2b}.

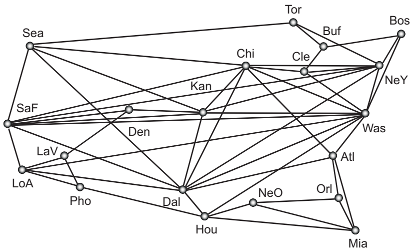


Figure 10: Test network based on the UUNET (1994).

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 11: Population of the cities and their surroundings.

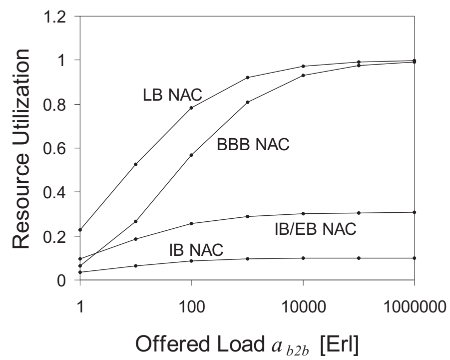


Figure 12: Impact of the offered load and the NAC method on the resource utilization.

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 NAC limits the number of flows in a network. We presented three basic NAC methods: the link budget (LB) based NAC, the t border-to-border (b2b) 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 the 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 b2b blocking probability.

Economy of scale is the key for understanding NAC performance. Therefore, we studied first the influence on the performance on a single link. The free parameters were offered load, desired blocking probability, and request size variability for multi-rate traffic. The offered load has the largest impact on the resource utilization such that the other parameters do not need to be considered in future studies. Due to the multi-rate nature of the traffic, it blocking probabilities related to blocked traffic volume are more reasonable.

We extended the dimensioning process from a single link to distributed NAC budgets in a network and to link capacities. The numerical comparison of NAC approaches in our test 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.

So far, we used our framework to test the sensitivity of the NAC performance to network topology [10], traffic distribution, and routing [15]. For NAC budget configuration, the inverse process of dimensioning is required [11]. 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 performance of the NAC methods differs significantly [12].

References

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

[2] Daniel O. Awduche, Lou Berger, Der-Hwa Gan, Tony Li, Vijay Srinivasan, and George Swallow. RFC3209: RSVP-TE: Extensions to RSVP for LSP Tunnels. http://www.ietf.org/rfc/rfc3209.txt, Dec. 2001.

[3] Fred Baker, Carol Iturralde, Francois Le Faucheur, and Bruce Davie. RFC3175: Aggregation of RSVP for IPv4 and IPv6 Reservations. http://www.ietf.org/rfc/rfc3175.txt, Sept. 2001.

[4] 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.

[5] Steven Blake, David L. Black, Mark A. Carlson, Elwyn Davies, Zheng Wang, and Walter Weiss. RFC2475: An Architecture for Differentiated Services. ftp://ftp.isi.edu/in-notes/rfc2475.txt, Dec. 1998.

[6] Bob Braden, David Clark, and Scott Shenker. RFC1633: Integrated Services in the Internet Architecture: an Overview. http://www.ietf.org/rfc/rfc1633.txt, June 1994.

[7] 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 Conf. on Advances in Commun. and Control (ComCon8)*, Crete, Greece, June 2001.

[8] Huirong Fu and Edward Knightly. Aggregation and Scalable QoS: A Performance Study. In *Proceedings of IWQoS 2001*, Karlsruhe, Germany, June 2001.

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

[10] Michael Menth and Joachim Charzinski. Impact of Network Topology on the Performance of Network Admission Control Methods. In *International Workshop on Multimedia Interactive Protocols and Systems (MIPS2003)*, Napoli, Italy, Nov. 2003.

[11] Michael Menth, Sebastian Gehrsitz, and Jens Milbrandt. Fair Assignment of Efficient Network Admission Control Budgets. In 18^{th} *International Teletraffic Congress*, pages 1121–1130, Berlin, Germany, Sept. 2003.

[12] Michael Menth, Stefan Kopf, and Joachim Charzinski. Impact of Resilience Requirements on the Performance of Network Admission Control Methods. Technical Report, No. 309, University of Würzburg, Institute of Computer Science, Feb. 2003.

[13] Michael Menth, Stefan Kopf, Jens Milbrandt, and Joachim Charzinski. Introduction to Budget Based Network Admission Control Methods. In 28^{th} *Annual IEEE Conference on Local Computer Networs (LCN2003)*, Bonn, Germany, Oct. 2003.

[14] Michael Menth, Stefan Kopf, and Jens Milbrandt. A Performance Evaluation Framework for Network Admission Control Methods. Technical Report, No. 305, University of Würzburg, Institute of Computer Science, Feb. 2003.

[15] Michael Menth, Stefan Kopf, and Jens Milbrandt. Impact of Traffic Matrix and Routing on the Performance of Network Admission Control Methods. Technical Report, No. 307, University of Würzburg, Institute of Computer Science, Feb. 2003.

[16] Vern Paxson and Sally Floyd. Wide-Area Traffic: The Failure of Poisson Modeling. *IEEE/ACM Transactions on Networking*, 3(3):226–244, June 1995.

[17] James Roberts, Ugo Mocci, and Jorma Virtamo. *Broadband Network Teletraffic - Final Report of Action COST 242*. Springer, Berlin, Heidelberg, 1996.

[18] Ion Stoica and Hui Zhang. Providing Guaranteed Services Without Per Flow Management. *Computer Communication Review*, 29(4), October 1999.

[19] Robert Szábó, Tamás Henk, Vlora Rexhepi, and Georgios 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.

[20] Ben 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.

[21] Andreas Terzis, J. Wang, J. Ogawa, and Lixia Zhang. A Two-Tier Resource Management Model for the Internet. In *Global Internet Symposium'99*, Dec. 1999.

[22] 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.

[23] Xipeng Xiao and Lionel M. Ni. Internet QoS: A Big Picture. *IEEE Network Magazine*, 13(2):8–18, March 1999.

[24] Zhi-Li Zhang Zhang, Zhenhai Duan, and Yiwei T. Hou. On Scalable Design of Bandwidth Brokers. *IEICE Transaction on Communications*, E84-B(8):2011–2025, Aug 2001.