Performance of PCN-Based Admission Control under Challenging Conditions

Michael Menth* and Frank Lehrieder†
*Dept. of Computer Science, University of Tübingen, †Dept. of Computer Science, University of Würzburg

***Abstract*—Pre-congestion notification (PCN) is a packet marking technique for IP networks to notify egress nodes of a so-called PCN domain whether the traffic rate on some links exceeds certain configurable bounds. This feedback is used by decision points for admission control (AC) to block new flows when the traffic load is already high. PCN-based AC is simpler than other AC methods because interior routers do not need to keep per-flow states. Therefore, it is currently being standardized by the IETF. We discuss various realization options and analyze their performance in the presence of flash crowds or with multipath routing by means of simulation and mathematical modeling. Such situations can be aggravated by insufficient flow aggregation, long round-trip times, on/off traffic, delayed media, inappropriate marker configuration, and smoothed feedback.**

***Index Terms*—Admission control, QoS, packet re-marking.**

I. Introduction

To achieve quality of service (QoS) for high-priority traffic in IP networks, its forwarding is prioritized over other traffic using Differentiated Services (DiffServ) [1]. In addition, admission control (AC) may limit the number of high-priority flows on links to avoid overload and to prevent packet loss and delay caused by high-priority traffic. AC for a flow may be performed only once per domain or for every link of a path. If overload occurs in spite of AC, e.g. due to a failure and traffic rerouting, flow termination (FT) can remove some already admitted flows to restore a controlled-load condition [2].

Pre-congestion notification (PCN) is a new per-domain mechanism to facilitate AC and FT, primarily for inelastic realtime flows in wired networks [3]. A DiffServ network using PCN is called PCN domain, and traffic under PCN control is called PCN traffic. The idea of PCN is that routers re-mark PCN packets on an outgoing interface when the PCN traffic rate on that link exceeds the configurable, link-specific admissible or supportable rate. PCN is developed for use in limited domains. PCN ingress nodes color incoming PCN traffic appropriately and egress nodes evaluate the color of outgoing PCN traffic. They communicate the information about differently marked PCN packets, i.e., PCN feedback, to the AC decision points, e.g. ingress nodes, which block admission requests for new PCN flows if needed. Methods for PCN-based AC consist of two components: the packet metering and marking algorithm in the routers [4] and the actual AC algorithms that turn the obtained packet markings into AC decisions. Many PCN algorithms require the notion of ingress-egress aggregates (IEAs) which are the ensemble of all PCN flows between a specific pair of ingress and egress nodes [5].

This work was funded by Deutsche Forschungsgemeinschaft (DFG) under grant TR257/18-2. The authors alone are responsible for the content of the paper.

The Internet Engineering Task Force (IETF) currently standardizes PCN and PCN-based AC and FT for the Internet [6], [7]. This paper studies the performance of the preferred PCN-based AC mechanisms when admission requests arrive more frequently than expected. Then AC should block new flows fast enough to avoid that too many flows are admitted which we call overadmission. This may be challenging in case of insufficient flow aggregation, long round-trip times, delayed media, on/off traffic, inappropriate marker configuration, smoothed feedback, or multipath routing. We investigate how well PCN-based AC can limit the rate of admitted PCN traffic under these conditions using simulation and mathematical analysis, and propose improvements. Our results show that all variants of PCN-based AC can break but to a different extent and some can become inefficient due to early blocking and waste of resources.

The paper is structured as follows. Section II explains PCN, metering and marking algorithms as well as various AC algorithms. Section III reviews related work. Section IV and Section V investigate the performance of "CLE-based" AC and "probe-based" AC methods for threshold and excess traffic marking. They are investigated under challenging conditions to see if they break, if so to what extent, and we propose improvements. Finally, Section VI summarizes our findings and Section VII draws conclusions.

II. Admission Control Based on Pre-Congestion Notification (PCN)

In this section we review the general idea of PCN-based admission control (AC) and flow termination (FT). We summarize PCN's two metering and marking algorithms and how they are used. Finally, we explain different AC algorithms. For detailed information about flow termination algorithms we refer the interested reader to [5] and [8].

A. Pre-Congestion Notification (PCN)

PCN defines a new traffic class that receives preferred treatment by PCN nodes. It provides information to support AC and FT for this traffic type. PCN introduces an admissible and a supportable rate threshold ($AR(l)$, $SR(l)$) for each link l of the network which imply three different load regimes as illustrated in Figure 1. If the PCN traffic rate $r(l)$ is below

$AR(l)$, there is no pre-congestion on the link and further flows may be admitted. If the PCN traffic rate $r(l)$ is above $AR(l)$, the link is AR-pre-congested. In this state, no further flows should be admitted. If the PCN traffic rate $r(l)$ is above $SR(l)$, the link is SR-pre-congested. In this state, some already admitted flows should be terminated to reduce the PCN rate $r(l)$ below $SR(l)$.

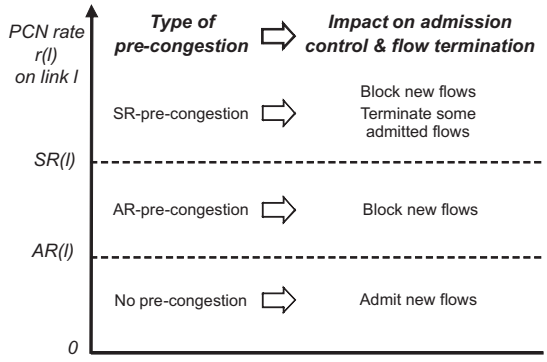


Fig. 1. The admissible and the supportable rate $(AR(l), SR(l))$ define three types of pre-congestion.

B. The Big Picture

PCN-based AC assumes that some end-to-end signalling protocol, e.g. the Session Initiation Protocol (SIP) [9] or the Resource reSerVation Protocol (RSVP) [10], or another mechanism requests admission for a new flow to cross a so-called PCN domain in a similar way as it is done in the IntServ-over-DiffServ proposal [11]. Thus, PCN-based AC is a per-domain QoS mechanism and an alternative to RSVP clouds or extreme capacity over-provisioning. This is illustrated in Figure 2. Traffic enters the PCN domain only through PCN ingress nodes and leaves it only through PCN egress nodes. Ingress nodes set a special header codepoint to make PCN packets distinguishable from non-PCN traffic and the egress nodes clear the codepoint. The nodes within a PCN domain are PCN nodes. They monitor the PCN traffic rate on their links and possibly re-mark PCN packets in case of AR- or SR-pre-congestion. PCN egress nodes evaluate the markings of PCN traffic and send a digest to the AC and FT entities of the PCN domain.

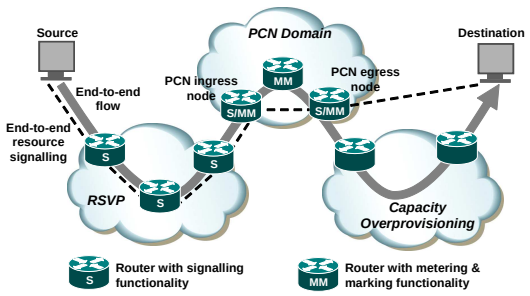


Fig. 2. PCN-based AC is triggered by admission requests from external signalling protocols and guarantees QoS within a single PCN domain.

C. PCN Metering and Marking

Excess traffic marking and threshold marking are defined in [4] and the appropriate encodings are defined in [12]. Packets are "not-marked" (NM) when entering a PCN domain, threshold markers possibly re-mark them to "threshold-marked" (ThM), and excess traffic markers possibly re-mark NM- and ThM-packets to "excess-traffic-marked" (ETM). The behavior of the marking algorithms can be described using a token bucket: a counter F is incremented by a reference rate R over time but cannot exceed F_{max}, i.e., $F(t_0+t) = \min(F_{max}, F(t)+R\cdot t)$. If the token bucket sees a packet of size B, it reduces the counter by that value, but F cannot become negative. If both marking algorithms are implemented, each of them has its own token bucket.

1) Threshold Marking: With threshold marking, F is decremented by any PCN packet. A threshold $F_{threshold}$ is defined so that if $F < F_{threshold}$ holds at the arrival of a NM-packet, this packet is re-marked to ThM. Threshold marking re-marks all NM-packets to ThM if the overall PCN traffic rate exceeds its reference rate.

2) Excess Traffic Marking: With excess traffic marking, F is only decremented by NM- and ThM-packets. If such a packet arrives and $F < F_{min}$ holds, the packet is re-marked to ETM and F is not decremented. F_{min} equals a maximum transfer unit. Excess traffic marking re-marks the rate of all NM- and ThM-packets to ETM that exceed its reference rate R.

D. Two PCN Architectures

There are two different PCN architectures: "Controlled Load" (CL) [6] and "Single Marking" (SM) [7].

1) The CL Architecture: With CL, PCN nodes use both threshold and excess traffic marking to re-mark PCN packets. The reference rate of the threshold marker is set to the admissible rate $AR(l)$ of a link l and the reference rate of the excess traffic marker is set to its supportable rate $SR(l)$. Thus, the threshold marker possibly re-marks NM-packets to ThM and the excess traffic marker possibly re-marks NM- and ThM-packets to ETM.

The CL architecture has two important features. First, in case of any pre-congestion, all packets are re-marked to either ThM or ETM. This is a very clear signal. Second, packets are re-marked to ETM only if they have traversed an SR-pre-congested link so that AR- and SR-pre-congestion can be well differentiated.

2) The SM Architecture: With SM, PCN nodes use only excess traffic marking to re-mark PCN packets. The reference rate of the excess traffic marker is set to the admissible rate $AR(l)$ of a link l. Thus, single marking re-marks only NM-packets to ETM. A domain-wide constant u determines the link-specific supportable rates by

$$SR(l) = u \cdot AR(l). \tag{1}$$

The SM architecture provides only a weak signal for AR-pre-congestion as only a few PCN-packets are re-marked to ETM, and it is hard to differentiate AR-pre-congestion from SR-pre-congestion which makes FT more difficult [8]. However, the SM architecture requires only two codepoints

(NM and ETM) and only a single marking algorithm which makes it easier to deploy as the two-state encoding in [13] is compatible with legacy equipment and excess traffic markers are already available.

E. Algorithms for PCN-Based Admission Control

Various algorithms for PCN-based AC have been summarized in [5]. We present only the two most relevant methods whose performance is investigated in Section IV and Section V.

1) Probe-Based AC (PBAC): With PBAC, n_p unmarked probe packets are sent by the PCN ingress node upon an admission request (probing). They must have the same source and destination address and port as future data packets to guarantee that they are forwarded on the same path as future data packets in case of multipath routing. The probe packets are intercepted by the PCN egress node. If one of them is re-marked to ThM or ETM or is lost, the new flow is blocked, otherwise it is admitted. We review two different implementation options of PBAC [5].

a) Explicit Probing: With explicit probing, the ingress node generates and sends explicit probe packets, and delays the processing of the admission request until it receives feedback from the egress node. As a consequence, PCN ingress nodes must buffer pending admission requests. This increases their complexity and delays call setups.

b) Implicit Probing: Implicit probing avoids the drawbacks of explicit probing by reusing messages of the end-to-end signalling protocol for probing [5]. For instance, RSVP can be modified for that purpose, i.e., the handling of RSVP messages by PCN egress nodes needs to be changed. The first PATH message of a call setup is exploited for probing purposes. If the PCN egress node receives a re-marked (ThM, ETM) PATH message for a new connection, it simply blocks the call by returning a PATH-TEAR message. Otherwise, the PATH message is forwarded downstream and the new call will immediately be accepted by the PCN ingress node when the corresponding RESV message returns upstream to ask for admission. Thus, implicit probing does not introduce additional probing delay. However, its applicability is limited to the CL architecture because implicit probing can usually re-use only a single signalling message as probe packet and, therefore, requires that all PCN packets are re-marked in case of pre-congestion.

2) CLE-Based AC (CLEBAC): CLEBAC is the preferred AC method in the IETF and used in the current official proposals [6] [7]. It requires that ingress and egress nodes classify their flows into ingress-egress aggregates (IEAs) which comprise all flows between specific ingress-egress node pairs. Egress nodes periodically measure the rates of NM-, ThM-, and ETM-traffic (NMR, TMR, EMR) per IEA using measurement intervals of duration D_{MI} and send them to corresponding ingress nodes. The ingress nodes calculate a congestion level estimate (CLE) per IEA as

$$CLE = \begin{cases} \frac{TMR+EMR}{NMR+TMR+EMR} & \text{if } NMR+TMR+EMR > 0 \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

If the CLE value is smaller than a configurable CLE limit L_{CLE}, the ingress node admits further requests for the particular IEA, otherwise it blocks them. CLEBAC can be used in the CL and SM architecture. With CL, admission decisions are rather independent of the CLE limit $0 < L_{CLE} < 1$ while for SM the CLE limit must be set to a small value $0 < L_{CLE} < 0.05$ since excess traffic marking re-marks only a small fraction of the traffic in case of pre-congestion.

3) CLEBAC with Probe Traffic: We propose an addition to CLEBAC. An ingress node sends a PCN probe packet to an egress node if the ingress node has not sent a PCN packet to that egress node for D_{MI}/n_p time. This effects that the egress node receives at least n_p packets within a measurement interval so that it can compute the CLE value based on feedback from the network even if the IEA does not carry any traffic. Extra traffic is generated only if the IEA is empty.

III. Related Work

We first review related work regarding other marking mechanisms and stateless core concepts for AC because they can be viewed as historic roots of PCN. Then, we give a short summary of related PCN studies. Apart from that, there is a large body of literature on other AC mechanisms. However, to keep this section short, we refer to [14] for this purpose.

A. Related Marking Mechanisms

RED and ECN can be seen as precursors of PCN marking.

1) Random Early Detection (RED): RED was originally presented in [15], and in [16] it was recommended for deployment in the Internet. It was designed to detect incipient congestion by measuring a time-dependent average buffer occupation avg in routers and to take appropriate countermeasures. RED drops or marks packets to indicate congestion to TCP senders and the probability for that action increases linearly with the average queue length avg. The value of avg relates to the physical queue size which is unlike PCN metering that relates to a virtual queue size whose server unit is configured with the admissible or supportable rate (technical solution: token bucket).

2) Explicit Congestion Notification (ECN): Explicit congestion notification (ECN) is built on the idea of RED to signal incipient congestion to TCP senders in order to reduce their sending window [17]. Packets of non-ECN-capable flows can be differentiated by a "not-ECN-capable transport" (not-ECT, '00') codepoint from packets of a ECN-capable flow which have an "ECN-capable transport" (ECT) codepoint. In case of incipient congestion, RED gateways possibly drop not-ECT packets while they just switch the codepoint of ECT packets to "congestion experienced" (CE, '11') instead of discarding them. In turn, TCP connections are expected to reduce their sending rate when receiving a CE-marked packet. This variant improves the TCP throughput since ECN saves packet retransmissions. Both the ECN encoding in the packet header and the behavior of ECN-capable senders and receivers after the reception of a CE-marked packet is defined in [17]. ECN comes with two different codepoints for ECT: ECT(0) ('10') and ECT(1) ('01'). They serve as nonces to detect cheating network equipment or receivers [18] that do

not conform to the ECN semantics. The four codepoints are encoded in the (currently unused) bits of the DiffServ codepoint (DSCP) in the IP header which is a redefinition of the type of service octet [19]. The ECN bits can be redefined by other protocols and [20] gives guidelines for that. Current PCN encodings make use of that [12].

B. Admission Control

We briefly review some specific AC methods that can be seen as forerunners of PCN-based AC.

1) Admission Control Based on Reservation Tickets: To maintain a reservation for a flow across a domain, the ingress node sends reservation tickets in regular intervals to the egress node. Intermediate routers estimate the rate of the tickets and thereby estimate the expected load. The ingress node sends probe tickets for a new reservation. Intermediate routers forward them to the egress node if they have still enough capacity to support the new flow and the egress node bounces them back to the ingress node indicating successful reservation; otherwise, the intermediate routers drop the probe tickets and the reservation request is denied. Several stateless core mechanisms work according to this idea [21]–[23].

2) Admission Control Based on Packet Marking: Gibbens and Kelly [24]–[26] theoretically investigated AC based on the feedback of marked packets whereby packets are marked by routers based on a virtual queue with configurable bandwidth. This core idea is adopted by PCN. Marking based on a virtual instead of a physical queue allows to limit the utilization of the link bandwidth by premium traffic to arbitrary values between 0 and 100%. Karsten and Schmitt [27], [28] integrated these ideas into the IntServ framework and implemented a prototype. They point out that the marking can be based also on the CPU usage of routers instead of the link utilization if this turns out to be the limiting resource for packet forwarding.

3) Admission Control Based on Probing: Breslau et al. [29] presented the concept of endpoint AC. Hosts or edge routers send probe traffic into the network for new flows and admit them depending on the fraction of lost and ECN-marked packets. Thus, flows are blocked only if the network is already in a congestion state. This is different with PCN-based AC where admission of further flows is already stopped if the PCN traffic rate on a link exceeds the admissible rate threshold. The advantage of endpoint AC is that it does not need support from the network. However, this is at the expense that a network operator of a core network cannot control or enforce admission decisions. Más et al. studied a PBAC method in [30] which is very similar to endpoint admission control. They proposed an analytical model in [31] and substituted probe traffic with initial voice traffic in [32].

4) Resilient Admission Control: Resilient admission control admits only so much traffic that it still can be carried after rerouting in a protected failure scenario [33]. It is motivated by the fact that overload in wide area networks mostly occurs due to link failures and not due to increased user activity [34]. It can be implemented with PCN by setting the admissible rate thresholds $AR(l)$ low enough so that the PCN rate $r(l)$ on a link l is lower than the supportable rate threshold $SR(l)$ after rerouting.

C. Related Studies in PCN

An overview of PCN including a multitude of AC and FT mechanisms is given in [5]. In [35], a high level summary about a large set of simulation results for PCN-based AC and FT was provided and it was shown that these methods work well in most studied cases. In contrast to that work, we investigate specifically challenging scenarios and point out under which conditions PCN-based AC does not work well. The authors of [36] propose an autonomic PCN-based AC algorithm optimized for video services in multimedia access networks and evaluate it with typical video traffic. They study only the CL architecture and do not look at challenging conditions. The same authors look at alternative metering and marking schemes in [37]. In [38], we studied flow blocking probabilities for single IEAs with static load conditions. In this work, we consider AC for traffic composed of multipe IEAs with flash crowd behavior and provide results about potential overadmission. As the MPLS header provides even fewer codepoints than the IP header, the authors of [39] proposed an encoding scheme for a CL-like architecture in a single codepoint using the frequency of re-marked packets to interpret the case-specific meaning of the codepoint. They compared the perceived quality of experience of voice flows with probe-based AC using standard CL PCN. The study in [40] gives recommendations for the setting of admissible and supportable rate thresholds in the context of resilient networking. Furthermore, it studied how link weights should be set in IP networks to maximize the admissible traffic rates.

IV. Overadmission for PCN-Based AC with Threshold Marking

In this section we consider PCN-based AC with threshold marking as it is used for the CL architecture [6]. We study potential over- and underadmission in the presence of flash crowds and with multipath routing. We first explain the simulation setup. Then, we investigate the admission behavior of PBAC and CLEBAC.

A. Simulation Setup

We consider a PCN domain where pre-congestion is observed only on a single bottleneck link as shown in Figure 3. As packets are re-marked only on this link, we simulate only this link in our experiments and cover the networking aspect by a configurable round-trip time which is mostly set to $RTT = 50$ ms. For ease of implementation, packets take $RTT/2$ to travel from the sender to the bottleneck link while the delay from the bottleneck link to the receiver is negligible.

We use the following default parameters. The admissible rate of the bottleneck link is 8 Mbit/s. The threshold marker of the link is configured with a bucket size of $T_{max} \cdot AR$ with $T_{max} = 100$ ms and the marking threshold of the threshold marker is set to $T_{threshold} \cdot AR$ with $T_{threshold} = T_{max}/2 = 0.05$ ms. We have chosen this value as it is able to accommodate a burst of 50 kB without marking traffic and limits the blocking delay to a relatively low value (see Section IV-B2). We assume that AC-limited PCN traffic does not face congestion even in the presence of moderate overadmission. This is justified by the fact that the physical bandwidth of a link is

usually much larger than its admissible rate AR. Therefore, we are not interested in packet loss and delay and do not need to take packet transmission times and queuing delay into account.

We use constant bit rate (CBR) voice flows that have periodic packet inter-arrival times of $A = 20$ ms and constant IP packet sizes of $B = 200$ bytes resulting in a rate of $c_{flow}^{CBR} = 80$ kbit/s. Those are typical values for the G.711 codec with UDP/IP transport [41]. We also consider on/off voice traffic as produced by the iLBC codec [42] in the XLite tool. It generates 62 bytes large packets that are equipped with 40 bytes RTP/UDP/IP header resulting into $B = 102$ bytes every $A = 30$ ms. However, voice activity detection arranges that no traffic is sent for silence intervals. This leads to contiguous on- and off-periods during which traffic is sent or suppressed. The iLBC codec does not even send infrequent additional control information during off-periods what other on/off codecs may do [43]. In [41] we found that the durations of on- and off-periods are on average 11.00 s and 11.54 s long so that the mean flow rate is $c_{flow}^{iLBC} = \frac{11.00\text{s}}{11.00\text{s}+11.54\text{s}} \cdot \frac{102\text{bytes}}{30\text{ms}} = 13.27$ kbit/s. Moreover, they can be modeled by a geometric distribution.

With these traffic types, the bottleneck link can support $n_{AR}^{CBR} = \frac{AR}{c_{flow}^{CBR}} = 100$ CBR flows or $n_{AR}^{onoff} = \frac{AR}{c_{flow}^{onoff}} = 602$ on/off flows. If we use just n_{AR}, we usually think of CBR flows and denote the number of admitted flows by n. We assume Poisson flow arrivals with a rate of λ and exponentially distributed call holding times with a mean value of $\frac{1}{\mu} = 90$ s. We set the normal flow arrival rate to

$$\lambda = n_{AR} \cdot \mu \tag{3}$$

so that the offered load equals the number of admissible flows n_{AR}. Note that we often use Equation (3) to eliminate λ in equations to make analytical results directly dependent on the flow aggregation on the bottleneck link n_{AR}.

A flash crowd commonly describes an unexpectedly high request rate [44]–[46]. Such situations have been observed in the telephone network, e.g., during voting shows [47] or over Christmas. In our experiments we parameterize the strength of a flash crowd by a factor $f_{crowd}^{flash} > 1$. To test AC during flash crowds, our experiments start with an empty link, new flows arrive with a rate of $f_{crowd}^{flash} \cdot \lambda$, and are continuously admitted until AC starts blocking. As $f_{crowd}^{flash} = 1$ has been used for link dimensioning in our setting, AC will block at least some flows during flash crowd events. If not mentioned differently, we use $f_{crowd}^{flash} = 5$ as default value in our study.

We use a custom-made packet-based simulation tool to simulate the time-dependent PCN traffic rate $r(t)$ on the bottleneck link. We measure it over 500 ms long intervals for illustration purposes in figures. To avoid simulation artifacts due to combinatorial effects caused by overly exact packet arrival times on the bottleneck link, we add some uniformly distributed jitter to the packet arrival times of at most $D_{pkt}^{max} = 5$ ms. We simulate the time-dependent PCN traffic rate $r(t)$ on the link and perform so many runs that 95% confidence intervals are mostly smaller than 1% of the obtained mean values. However, we omit them in the figures for the sake of clarity. We rather show 5%- and 95%-quantiles to illustrate the variation of the PCN traffic rate $r(t)$ over multiple experiments.

The performance metric of interest in this study is overadmission OA. It is the fraction by which the admitted traffic rate $r(t)$ exceeds AR on the simulated link. Thus, a traffic rate of $r(t) = 2 \cdot AR$ corresponds to an overadmission value of $OA = 1$.

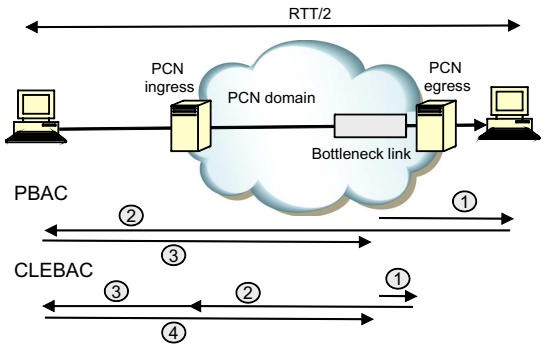


Fig. 3. Analysis of signaling delay for PBAC and CLEBAC. Sequences are explained in the text.

B. Overadmission for PBAC with Threshold Marking

We analyze the reason for overadmission and explain how we measure overadmission in simulations. We model the components of blocking delay, illustrate the impact of several model parameters on blocking delay and overadmission, and validate our model by simulations. We also investigate the effect of on/off traffic and extensive media delay.

1) Analysis of Overadmission: We simulate the time-dependent PCN traffic rate in the presence of flash crowds. Two sample runs are shown in the left part of Figure 4. The PCN traffic rate rises, stops only when it is clearly above the admissible rate of $AR = 8$ Mbit/s, and then oscillates around that value on a relatively short time scale. We briefly explain that phenomenon. Overadmission occurs if new admission requests are admitted although the admitted flows already cause pre-congestion on a bottleneck link. This may happen because AC decisions are taken based on obsolete information. Reasons may be: (1) admitted flows have not yet started sending traffic (media delay D_{media}), (2) packets are not yet re-marked by the threshold marker as its token bucket still contains sufficiently many tokens (marking delay D_{mark}), or (3) positive RESV messages resulting from non-re-marked signalling messages still reach the source node and new traffic reaches the bottleneck link (signalling delay D_{signal}). They add to the blocking delay

$$D_{block} = D_{media} + D_{mark} + D_{signal}. \tag{4}$$

Within that time, additional flows are falsely admitted so that overadmission occurs. If flows finish and the PCN traffic rate falls below the admissible rate AR, the token bucket of the threshold marker is empty and it takes some time D_{mark}^{*} until it contains again sufficiently many tokens so that packet re-marking stops. Together with the signaling delay and the media delay, it composes the admission delay

$$D_{admit} = D_{media} + D_{mark}^{*} + D_{signal}. \tag{5}$$

it takes until the bottleneck link may see traffic from newly admitted PCN flows. Within that time, additional admitted flows terminate so that underadmission occurs and the token

bucket has time to fill up again, at least to some extent. If the flash crowd event still exists, overadmission restarts again after some time. We mostly observe smaller consecutive peaks if the token bucket is not fully re-filled when overadmission occurs again.

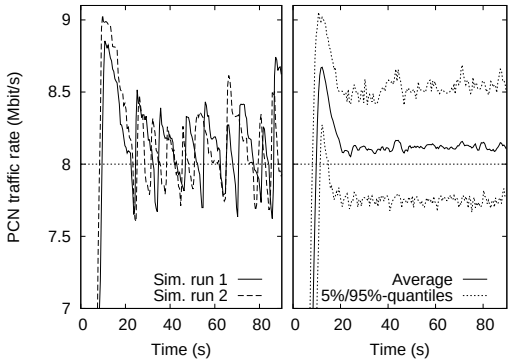


Fig. 4. Time-dependent PCN traffic rates for PBAC with threshold marking in the presence of CBR traffic.

The right part of Figure 4 shows the average, the 5%- and the 95%-quantile of the time-dependent PCN traffic rate over multiple simulation runs. We observe that the overadmission in the averaged curve is clearly smaller than the initial peak overadmission in the two individual runs in the upper part of the figure. They reach their peak after different time so that the peak of the averaged curve is lower than the averaged peak values of individual runs. To get the average peak overadmission, we average the first peak values of the time-dependent PCN traffic rate from multiple simulation runs and call the resulting value average peak overadmission. The 5%- and the 95%-quantiles in the right part of Figure 4 provide a corridor within which we see the PCN traffic rate oscillates for 90% of the time.

2) Modeling Blocking Delay: We analyze the signaling delay and the marking delay, and visualize the blocking delay depending on various parameters.

a) Signaling Delay: We analyze the signalling delay D_{signal} for PBAC using Figure 3. In case of marked signalling, the time a PATH message travels from the bottleneck link via the egress node to the destination (1), the time a corresponding RESV packet travels from the destination via PCN egress and ingress node back to the source (2), and the time data traffic of the newly admitted source takes to reach the bottleneck link (3) contribute to the signalling delay D_{signal} and sum up to one RTT.

b) Marking Delay: We now model the marking delay D_{mark}. Let t_0 be the time when pre-congestion starts on the bottleneck link. We approximate the time-dependent number of admitted PCN flows on the bottleneck link by

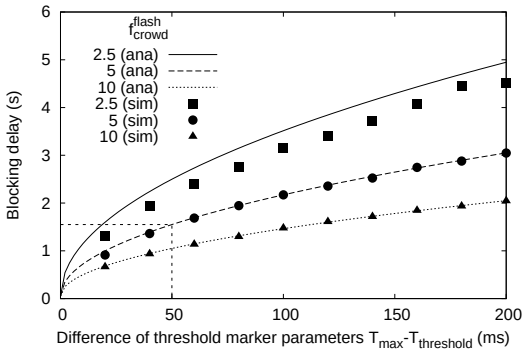
$$\begin{aligned} n(t) &= n_{AR} + (\lambda \cdot f_{crowd}^{flash} - \mu \cdot n_{AR}) \cdot (t - t_0) \\ &= n_{AR} + \mu \cdot n_{AR} \cdot (f_{crowd}^{flash} - 1) \cdot (t - t_0). \end{aligned} \quad (6)$$

Thereby, we have approximated the rate of departing flows $\mu \cdot n(t)$ by $\mu \cdot n_{AR}$ to simplify calculations. As a result of that, D_{mark} and peak overadmission will be slightly underestimated. The threshold marker starts re-marking packets when $(T_{max} - T_{threshold}) \cdot AR$ bytes have been taken from the threshold marker's token bucket which defines D_{mark}:

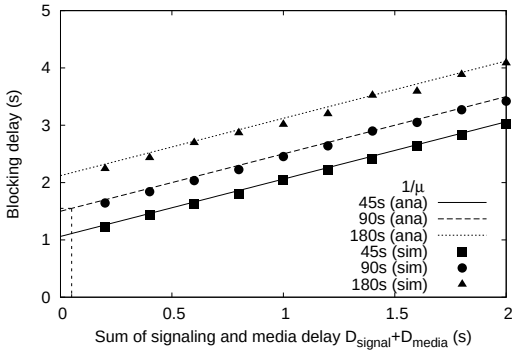
$$\begin{aligned} (T_{max} - T_{threshold}) \cdot AR &= \int_{t_0}^{t_0 + D_{mark}} (n(t) - n_{AR}) \cdot c_{flow}\, dt \\ &= \mu \cdot n_{AR} \cdot (f_{crowd}^{flash} - 1) \cdot c_{flow} \cdot \frac{(D_{mark})^2}{2}. \end{aligned} \quad (7)$$

With $AR = n_{AR} \cdot c_{flow}$, we get $D_{mark} = \sqrt{\frac{2 \cdot (T_{max} - T_{threshold})}{\mu \cdot \left(f_{crowd}^{flash} - 1\right)}}$.

c) Blocking Delay: Figure 5(a) shows the blocking delay calculated according to Equation (4). The blocking delay increases with the difference of the threshold marking parameters $T_{max} - T_{threshold}$ but decreases with increasing flash crowd factor f_{crowd}^{flash}. As we set the sum of the signaling and media delay to a constant value of $D_{signal} + D_{media} = 50$ ms, we observe that the difference of $T_{max} - T_{threshold}$ dominates the blocking delay. Small difference values in the order of hundred milliseconds already lead to long marking delays in the order of seconds. When the flash crowd intensity is stronger, flows are faster admitted, overadmission increases also faster, and tokens are faster removed from the threshold marker's token bucket which leads to shorter marking and blocking delay. For our default setting, we obtain a blocking delay of about $D_{block} = 1.5$ s. We also provide simulation results showing that our model is rather accurate. For $f_{crowd}^{flash} = 2.5$ we see less accuracy because in this case marking delay D_{mark} is hard to measure in simulations due to fluctuations of the PCN traffic rate around the admissible rate AR.



(a) Impact of threshold marker configuration and flash crowd f_{crowd}^{flash}.



(b) Impact of combined signaling and media delay and average flow holding time $\frac{1}{\mu}$.

Fig. 5. Blocking delay depending on various factors. Default settings are indicated by dashed lines.

Figure 5(b) illustrates the impact of combined signaling and media delay as well the one of flow holding times. Again we observe good agreement of analytical and simulation results. The blocking delay scales linearly with the combined signaling and media delay as proposed in Equation (4). Long signaling and media delay can also significantly extend the blocking delay. Assuming that signalling delay is in the order of round-trip times which can amount to no more than 500 ms, then we can conclude that signaling delay has at best secondary influence on blocking delay. This is different for media delay which we set to a default value of $D_{media} = 0$ in our study. But in Section IV-B5 we also investigate what can happen if media delay is large.

3) Impact of Threshold Marker Configuration and Flash Crowd Intensity on Peak Overadmission: We derive a formula to calculate peak overadmission and study it for different marker configurations and flash crowd factors. We also validate the formula with simulation results.

Overadmission occurs because additional PCN flows are falsely admitted for the duration of the blocking delay D_{block}. Within that time, about $(\lambda \cdot f_{crowd}^{flash} - n_{AR} \cdot \mu) \cdot D_{block}$ more flows are addmitted for the bottleneck link, again approximating the number of admitted flows by n_{AR}. Using Equation (3) we derive the average peak overadmission as

$$OA = \frac{D_{block} \cdot n_{AR} \cdot \mu \cdot \left(f_{crowd}^{flash} - 1\right)}{n_{AR}} = D_{block} \cdot \mu \cdot \left(f_{crowd}^{flash} - 1\right). \quad (8)$$

Figure 6 shows a sensitivity analysis of the peak overadmission based on the mathematical model. The peak overadmission increases with the difference of the threshold marker parameters $T_{max} - T_{threshold}$ and with the flash crowd factor f_{crowd}^{flash}. For large parameter values, peak overadmission in the range of 15% - 20% can be observed, but for $(T_{max} - T_{threshold}) \leq$ 100 ms and $f_{crowd}^{flash} \leq 5$ peak overadmission stays below 10%.

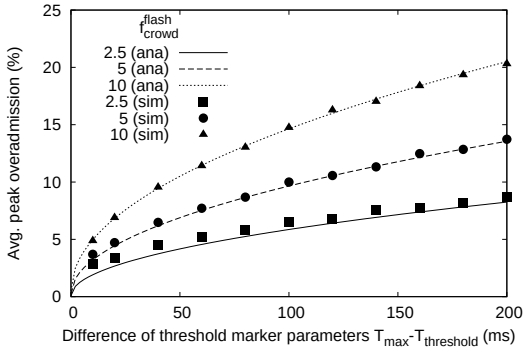
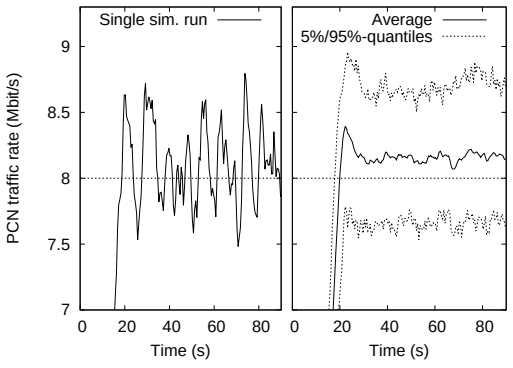
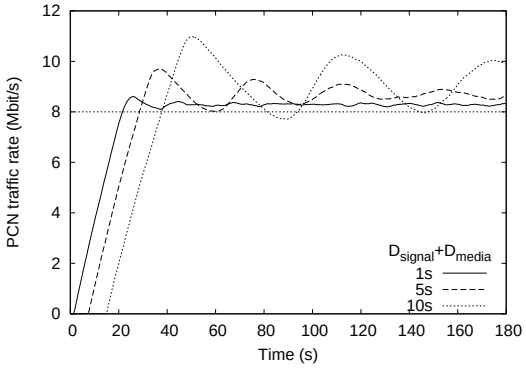


Fig. 6. Peak overadmission for PBAC with threshold marking depending on the difference of the threshold marker parameters $T_{max} - T_{threshold}$ and different flash crowd factors f_{crowd}^{flash}.

We validated our mathematical model by further simulation results and show them in the figure. The fact that the simulated points are rather close to the analytically calculated lines shows that our mathematical model is rather accurate in spite of some approximations and confirms our understanding of overadmission.



(a) In the presence of on/off traffic.



(b) Increased signaling and media delay $D_{signal} + D_{media}$.

Fig. 7. Time-dependent PCN traffic rates for PBAC with threshold marking.

4) Impact of On/Off Traffic on Overadmission: The left part of Figure 7(a) shows the time-dependent PCN traffic rate with on/off traffic instead of CBR traffic. In contrast to CBR traffic in the left part of Figure 4, the curve is a bit more jerky. We observe about the same average overadmission and for on/off traffic as for CBR traffic in the right parts of both figures. However, the interval between the 5%- and 95%-quantiles is larger for on/off traffic than for CBR traffic which means that overadmission is sometimes higher with on/off traffic than with CBR traffic. This is also confirmed in Figure 8 by the complementary cumulative distribution function (CCDF) of the PCN traffic rate: the probability that the PCN traffic rate exceeds 8.5 Mbit/s is larger for on/off traffic than for CBR traffic.

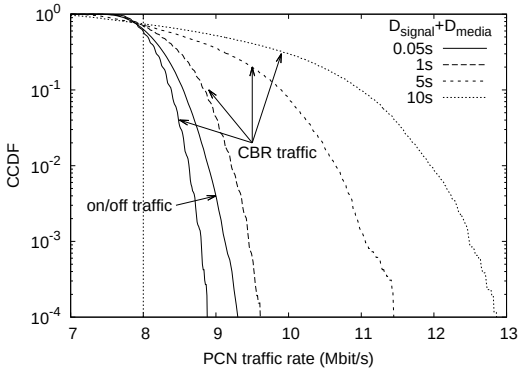


Fig. 8. CCDF of the PCN traffic rate for CBR traffic, for on/off traffic, and for CBR traffic with increased signaling and media delay.

5) Impact of Signaling and Media Delay on Overadmission: We study the impact of long signaling and media delay on overadmission in Figure 7(b). A value of $D_{signal}+D_{media}=1$ s for the combined signaling and media delay leads only to slightly larger PCN traffic rates than a standard value of 50 ms in the right part of Figure 4. However, the CCDF in Figure 8 shows that a slightly increased signalling and media delay leads to higher extreme values for overadmission. A large signalling and media delay of $D_{signal}+D_{media}\in\{5s,10s\}$ causes overadmission of up to 50%. For such values we observe oscillations of the PCN traffic rate in Figure 7(b). This phenomenon can be explained as follows. If the system starts admitting new flows for a link, it takes another $D_{signal}+D_{media}$ time until their effect is visible. Therefore, the PCN traffic rate still decreases due to terminating flow before it exceeds the admissible rate again. Within that time, all flow requests are admitted. Many of them start sending only after admission has stopped and cause overadmission. When the PCN traffic rate again falls below the admissible rate, the whole process repeats so that we can observe oscillations of the PCN traffic rate with periods of multiple tens of seconds.

As there was nothing PCN-specific in the discussed scenario, media delay is challenging for all AC approaches that use feedback from the network. A media delay of 10 s is not so unlikely. The fact that it may lead to 50% overadmission causes concerns. To prevent high overadmission, dummy traffic may be generated as long as the source is inactive. To resolve overload situations, flow termination may be used.

C. Overadmission for CLEBAC with Threshold Marking

From a performance point of view, CLEBAC is more complex than PBAC. We first extend our simulation model. We investigate how overadmission depends on the duration of the measurement interval. We show that standard CLEBAC does not work well for a small number of flows per IEA, especially in the presence of on/off traffic, and that CLEBAC with our proposed modifications in Section II-E3 avoids these problems. CLE smoothing was proposed to yield more stable measurement results, but we show that it can massively contribute to blocking delay and lead to overadmission if not configured carefully. Finally, we study CLEBAC in the presence of multipath routing and analytically show that it may lead to significant underadmission, i.e., to inefficient bandwidth usage.

1) Extending the Simulation Model for CLEBAC: We extend the simulation model in Section IV-A for CLEBAC. As CLEBAC is based on feedback from ingress-egress aggregates (IEAs), we integrate IEAs into our simulation model as shown in Figure 9. The PCN traffic on the simulated bottleneck link belongs to n_{IEA} different IEAs. Each of the IEAs expects an average flow number of $n_{IEA}^{flows}=\frac{n_{AR}}{n_{IEA}}$. The flow arrival rate per IEA is then $\lambda_{IEA}=\frac{\lambda}{n_{IEA}}$ under normal conditions and $\lambda_{IEA}\cdot f_{crowd}^{flash}$ during flash crowds. We configure CLEBAC's CLE-limit with $L_{CLE}=0.5$ and the duration of the measurement interval with $D_{MI}=200$ ms. To avoid potential artifacts due to synchronization, every IEA initially starts measuring after a random time between zero and D_{MI}.

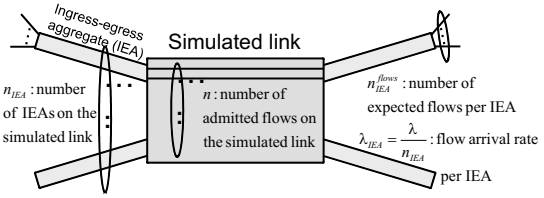


Fig. 9. Ingress-egress aggregates (IEAs): traffic bundling with CLEBAC.

2) Impact of the Duration of the Measurement Interval: With CLEBAC, egress nodes measure rates of differently marked PCN traffic over a duration of D_{MI} time and report them to ingress nodes. We define the measurement delay $D_{measure}$ by the duration of the interval from the arrival of a first re-marked packet at the egress node until the end of that measurement interval whose measured rates lead to blocking at the ingress node. To approximate the average measurement delay, we assume the traffic rate of the considered IEA remains constant during the relevant measurement intervals which is not necessarily the case in practice. If the egress node sees the arrival of re-marked PCN packets at time $0\le t<(1-L_{CLE})\cdot D_{MI}$ within a measurement interval, then the resulting CLE value at the ingress node is at least L_{CLE} so that the ingress node will block when it receives the measured PCN traffic rates. Hence, the measurement delay is $D_{MI}-t$. If the egress node sees the arrival of first re-marked PCN packets at time $(1-L_{CLE})\cdot D_{MI}\le t<D_{MI}$, the CLE resulting from this measurement interval is smaller than L_{CLE} and the ingress node blocks only at the reception of the measured PCN traffic rates from the next interval. Up to then we observe a measurement delay of $2\cdot D_{MI}-t$. Thus, we calculate the average measurement delay by

$$D_{measure}=\int_0^{(1-L_{CLE})\cdot D_{MI}} D_{MI}-t\,dt+\int_{(1-L_{CLE})\cdot D_{MI}}^{D_{MI}} 2\cdot D_{MI}-t\,dt=\left(\frac{1}{2}+L_{CLE}\right)\cdot D_{MI}. \quad (9)$$

It also contributes to the blocking delay for CLEBAC:

$$D_{block}=D_{media}+D_{mark}+D_{signal}+D_{measure}. \quad (10)$$

Figure 3 helps analyzing the signalling delay D_{signal} for CLEBAC. It consists of the time the first marked packet takes to travel from the bottleneck to the egress node (1), the time the measured rate takes to travel from the egress node to the ingress node (2), the time a RESV message takes to travel from the ingress node to the source (3), and the time a data packet takes to travel from a source to the bottleneck link (4). We assume that the destination is close to the egress node so that we can also approximate the signalling delay by $D_{signal}=RTT$ as for PBAC.

We use Equations (8) and (10) to calculate the peak overadmission and illustrate it in Figure 10. Peak overadmission generally increases with increasing duration of the measurement interval D_{MI} as this enlarges the blocking delay. It is also larger for larger flash crowd factors f_{crowd}^{flash}. Moreover, peak overadmission increases only slowly with increasing duration of the measurement interval D_{MI} for a small flash crowd factor

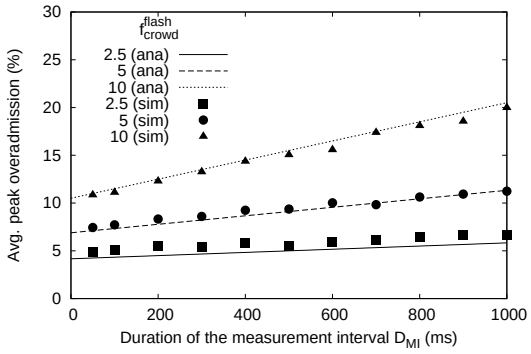
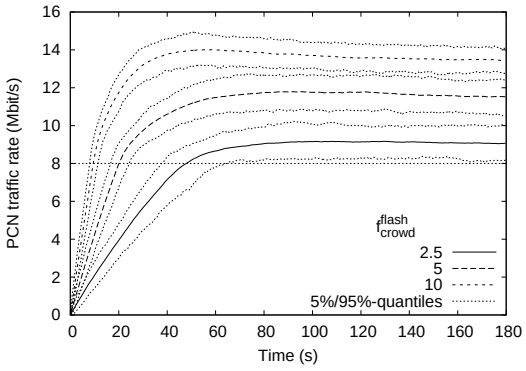


Fig. 10. Peak overadmission for CLEBAC with threshold marking depending on the measurement interval D_{MI} and different flash crowd factors f_{crowd}^{flash} for $n_{IEA}^{flows} = 100$ flows per IEA.

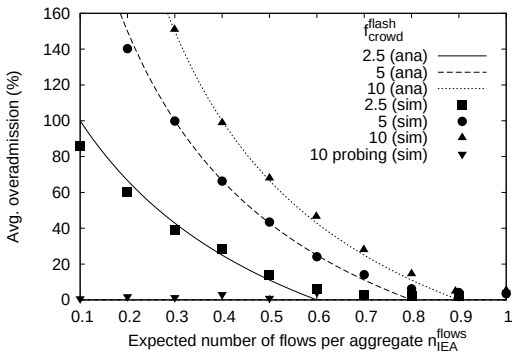
of $f_{crowd}^{flash} = 2.5$, but fast for a large flash crowd factor of $f_{crowd}^{flash} = 10$. We validate the analytically calculated lines in the figure with simulated data points to show that our model is fairly good in spite of approximations.

3) Impact of the Number of Flows per IEA: If an IEA does not carry any flow, it must admit new admission requests to avoid starvation. However, this is a source for overadmission if the admissible rate on the bottleneck link is provisioned for a load composed from many IEAs, each of them carrying only a very low number of flows on average. This is a realistic assumption for large PCN domains with many egde nodes [48]. We investigate such a scenario in the following simulation experiment. We assume a bottleneck link with $n_{AR} = 100$ flows and its load is composed from $n_{IEA} = 200$ different IEAs where every IEA offers a normal load of $n_{IEA}^{flows} = 0.5$ flows over time. Figure 11(a) shows that the time-dependent average PCN traffic rate for flash crowd factors $f_{crowd}^{flash} \in \{2.5, 5, 10\}$ is significantly larger than *AR*. Instead of initial peak overadmission and oscillations of the time-dependent PCN traffic rate around *AR*, we observe permanent overload that depends on the flash crowd factor f_{crowd}^{flash}.

We derive a rough approximation for the expected stationary overadmission. In contrast to previous analyses, it yields average instead of peak overadmission values. When the *AR* of the considered link is configured for only $n_{IEA}^{flows} < 1$ flows per IEA, most IEAs carry at most one admitted flow when the PCN rate on the link has exceeded *AR*. If an IEA carries one flow, it cannot admit further flows if the bottleneck link is pre-congested, which is the case in Figure 11(a). If the IEA carries zero flows, it admits another flow as soon as an admission request arrives. Assuming furthermore that only a single flow is admitted before an IEA blocks (which is only approximative), an IEA carries only one or zero flows. We derive the average number of active flows per IEA over time. Empty IEAs wait on average $\frac{1}{\lambda_{IEA} \cdot f_{crowd}^{flash}}$ time until the next admission request arrives which is then admitted. For IEAs with one admitted flow, it takes about $\frac{1}{\mu}$ time until they become empty. Hence, we calculate the average number of admitted flows per IEA by $\frac{1/\mu}{1/\mu + 1/(\lambda_{IEA} \cdot f_{crowd}^{flash})} = \frac{\lambda_{IEA} \cdot f_{crowd}^{flash}}{\lambda_{IEA} \cdot f_{crowd}^{flash} + \mu}$. Thus, the average number of admitted flows on the link is $n = n_{IEA} \cdot \frac{\lambda_{IEA} \cdot f_{crowd}^{flash}}{\lambda_{IEA} \cdot f_{crowd}^{flash} + \mu}$ while the number of admissible flows



(a) Time-dependent PCN traffic rate CLEBAC with threshold marking for an average number of $n_{IEA}^{flows} = 0.5$ flows per IEA.



(b) Average overadmission for standard CLEBAC and for CLEBAC with probe traffic.

Fig. 11. CLEBAC with threshold marking for a small number of flows per IEA n_{IEA}^{flows} in the presence of different flash crowd factors f_{crowd}^{flash}.

on the link is only $n_{AR} = n_{IEA} \cdot \frac{\lambda_{IEA}}{\mu}$. The corresponding level of overadmission is

$$OA = max\left(0, \frac{n}{n_{AR}} - 1\right) = max\left(0, \frac{\frac{\lambda_{IEA} \cdot f_{crowd}^{flash}}{\lambda_{IEA} \cdot f_{crowd}^{flash} + \mu}}{\frac{\lambda_{IEA}}{\mu}} - 1\right) = max\left(0, \frac{f_{crowd}^{flash}}{n_{IEA}^{flows} \cdot f_{crowd}^{flash} + 1} - 1\right). \quad (11)$$

This equation suggests that the level of stationary overadmission is independent of the number of IEAs n_{IEA} on the bottleneck link. Figure 11(b) illustrates the analytically calculated overadmission depending on the expected number of flows per IEA n_{IEA}^{flows} for different flash crowd factors $f_{crowd}^{flash} \in \{2.5, 5, 10\}$. The validation of the analytical values by simulation results shows that our approximation is rather accurate and can well serve as explanatory model. Additional simulation results for $n_{IEA}^{flows} = 1.0$ or larger (without figure) reveal hardly any stationary overadmission but only peak overadmission, i.e., the PCN traffic rate essentially oscillates around the admissible rate *AR*.

4) Impact of Probe Traffic for Empty IEAs: In Section II-E3 we have suggested that an ingress node should send a small PCN probe packet to the egress node if no other PCN packet has been sent for $\frac{D_{MI}}{n_p}$ time. We added this feature in our

simulation. Figure 11(b) shows that this modification removes all stationary overload that may appear due empty aggregates. Assuming $D_{MI} = 200$ ms, $n_p = 2$, and a probe packet size of 32 bytes including IP and UDP header, the probe traffic rate is 2.56 kbit/s for empty IEAs and zero for IEAs that send PCN traffic. In case of $n_{IEA}^{flows} = 0.1$ and $n_{IEA} = 1000$ IEAs this means that 930 IEAs send probe traffic and 70 IEAs can send voice traffic without exceeding the $AR = 8$ Mbit/s on the bottleneck link. Thus, probe traffic costs equal the capacity requirements of 30 flows in this rather extreme example. For $n_{IEA}^{flows} = 0.2$ and $n_{IEA} = 500$ IEAs the costs for probe traffic quickly reduce to the capacity requirement of 13.2 flows. If all IEAs carry PCN traffic, no probe traffic is sent so that probing overhead is zero.

5) Impact of On/Off Traffic in the Presence of a Small Number of Flows per IEA: The left part of Figure 12 shows the time-dependent traffic rate for $n_{IEA}^{flows} = 1$ flows per IEA and for CBR traffic. We observe some initial overadmission which vanishes after time. Performing the same experiment for on/off traffic leads to tremendous overadmission due to temporarily empty aggregates which is illustrated in the right part of Figure 12. The graph also shows that standard CLEBAC can successfully avoid overadmission for $n_{IEA}^{flows} = 5$ flows per IEA. However, this limits the applicability of CLEBAC in the presence of on/off traffic and low traffic aggregation. CLEBAC with probe traffic causes hardly any overadmission with on/off traffic which is not shown in the figure.

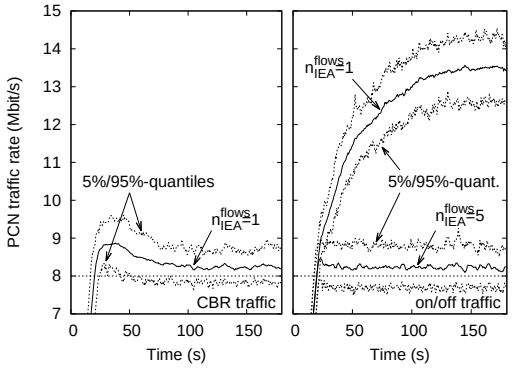


Fig. 12. Time-dependent PCN traffic rate for CLEBAC with threshold marking, CBR and on/off traffic, and for different numbers of flows per IEA n_{IEA}^{flows}.

6) Impact of CLE Smoothing: If the PCN rate on a link is near its *AR*, consecutively calculated CLE values may oscillate around the CLE limit L_{CLE}. To get rid of such oscillations, the authors of [49] proposed to smooth consecutive CLE values using an exponentially weighted moving average (EWMA). The ingress node calculates a smoothed value $CLE_{avg} = (1-w)\cdot CLE_{avg} + w\cdot CLE$ whenever it obtains a new measured CLE and admits or blocks new flows based on CLE_{avg}. The weight parameter w serves to control the dynamics of the averaged values CLE_{avg}. Smoothing lets ingress nodes block later, i.e., a smoothing delay occurs, which also contributes to blocking delay:

$$D_{block} = D_{media} + D_{mark} + D_{measure} + D_{signal} + D_{smooth}. \quad (12)$$

We consider a scenario with sudden pre-congestion, i.e., the measured CLE value suddenly jumps from zero to one and stays on this level. With CLE smoothing, the average value $CLE_{avg}(0)$ is zero at the beginning and slowly approaches one if the ingress node receives multiple CLEs with value one. The averaged CLE value after reception of i CLEs with value one is

$$CLE_{avg}(i) = (1-w)\cdot CLE_{avg}(i-1) + w\cdot CLE. \quad (13)$$

The ingress node stops admitting new flows only if $CLE_{avg}(i)$ is larger than the CLE limit. We calculate the additional smoothing delay as

$$D_{smooth} = D_{MI}\cdot(\min(i : CLE_{avg}(i) > L_{CLE}) - 1). \quad (14)$$

We illustrate it in Figure 13 depending on the weight parameter w. Smoothing delay decreases with increasing weight value and can be in the order of several seconds. The figure also shows the resulting peak overadmission calculated by Equation (8) for different flash crowd factors f_{crowd}^{flash}. Values for peak overadmission without CLE smoothing are visible for $w = 1$. Large values $0.5 \le w \le 1$ increase overadmission only slightly, but small values of w may triple overadmission. Simulation results confirm our findings.

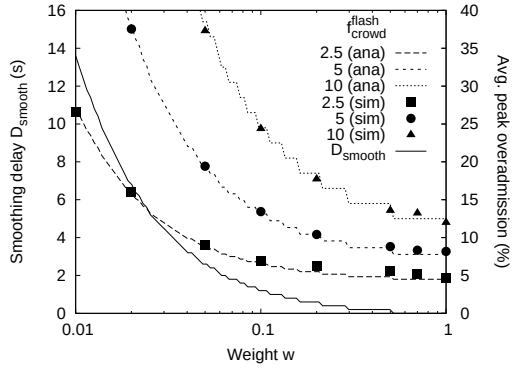


Fig. 13. Smoothing delay D_{smooth} and overadmission due to CLE smoothing depending on the EWMA weight w for $D_{MI} = 200$ ms.

Another option to dampen oscillations at the ingress node between admitting and blocking is the use of a hysteresis which has been investigated in [38]. However, we think that oscillations do not need to be avoided as they are not harmful.

7) Impact of Multipath Routing: With multipath routing, flows of a single IEA may be carried over different paths from the ingress to the egress of a PCN domain. As a result of that, CLEBAC may already stop admission when only one of the parallel paths within a multipath has a pre-congested link. Thus, a flow may be blocked although its prospective path does not contain any pre-congested link. Therefore, CLEBAC sometimes admits less PCN traffic than possible, which we call "underadmission". In that case it is not possible to utilize the entire capacity of all parallel paths. We quantify this effect in the following using a mathematical model.

We consider an IEA whose traffic is carried over a multipath that consists of k partial paths. Usually, only links have an admissible rate. To simplify the analysis, we assume that every partial path $0 \le i < k$ is configured with an admissible rate AR_i in terms of flows, i.e., with an admissible number of flows. The state of the IEA $s = (s_0, ..., s_{k-1})$ indicates the number

of active flows s_i per partial path $0 \leq i < k$. When a flow is admitted, it is randomly assigned to one of the partial paths with probability $p_{path}(i) = \frac{1}{k}$ and the number of admitted flows s_i on the chosen path i is then incremented by one. We model this by a Markov process whose transitions are described by

$$(s_0, ..., s_i, ..., s_{k-1}) \xrightarrow{p_{path}(i)} (s_0, ..., s_i + 1, ..., s_{k-1}). \quad (15)$$

The analysis starts with an empty IEA, i.e., $s_i = 0$ for $0 \leq i < k$. We sequentially add new flows and we do that for all possible transitions in parallel to calculate consecutive state distributions, i.e., the probabilities $p_n(s)$ that we see state s after n flow admissions. Our computation stops after m admissions if the probabilities $p_m(s)$ are zero for all states with $s_i < AR_i$ for $0 \leq i < k$. A "terminating state" has exactly one $s_i = AR_i$ and all other $s_j < A_j$ for all $0 \leq j < k, j \neq i$ and denote the set of all terminating states by $\mathcal{T}$. In these states, the IEA actually blocks so that no further flows can be admitted. The probabilities $p_m(s)$ of all terminating states $s \in \mathcal{T}$ sum up to $\sum_{0 \leq n \leq m, s \in \mathcal{T}} p_n(s) = 1$. They yield a probability distribution that indicates the number of admitted flows m when CLEBAC blocks and their assignment to the k different paths. The utilization of the multipath capacity in state s is given by $U(s) = \frac{\sum_{0 \leq i < k} s_i}{\sum_{0 \leq i < k} AR_i}$. Thus, we calculate the mean utilization of the multipath when the IEA blocks by $U = \sum_{0 \leq n \leq m, s \in \mathcal{T}} U(s) \cdot p_n(s)$.

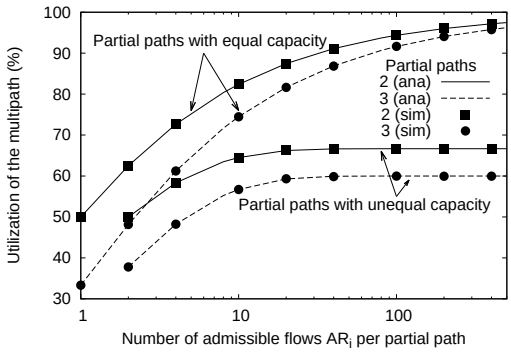


Fig. 14. CLEBAC leads to underadmission in the presence of multipath routing.

We study the average utilization of the multipath capacities when the IEA starts blocking depending on the number of parallel path and the number of admissible flows AR_i. Figure 14 shows that the utilization values increase for increasing numbers of admissible flows AR_i per partial path. This is due to economy of scale. The utilization is lower for $k = 3$ parallel paths than for $k = 2$ parallel paths. We consider two different scenarios: parallel paths with equal capacity and parallel paths with unequal capacity. In the latter case, one path has only half the capacity of the others. With equal-capacity paths a utilization close to 100% is possible for high-capacity paths while the utilization is limited to lower values in case of unequal-capacity paths. For unequal-capacity paths, one can show that the average admitted load on any path is limited by the minimum capacity of all paths. Simulation results show that our analysis is accurate.

V. Overadmission for PCN-Based AC with Excess Traffic Marking

In this section we consider PCN-based AC with excess traffic marking as it is used for the SM architecture [7]. The token bucket size of the excess traffic marker determines the marking delay. To be conform with the experiments in the previous section, we set its value to $T_{max} = 50$ ms, the default value for $T_{max} - T_{threshold}$ for the threshold marker in the previous section.

A. Overadmission for PBAC with Excess Traffic Marking

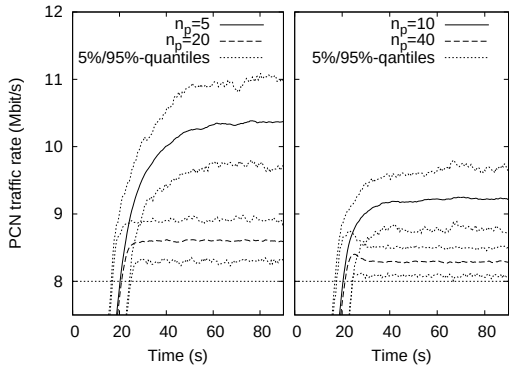
With excess traffic marking, only a fraction of packets is re-marked in case of pre-congestion. Therefore, PBAC requires that the ingress node sends n_p probe packets per flow that all need to arrive correctly at the egress node before the flow can be admitted. With excess traffic marking, only a fraction of packets is re-marked in case of pre-congestion. The probe packets are again 32 bytes large and are issued with exponentially distributed inter-arrival times whose mean value is the same as for future data packets. This is needed to get a reliable estimate of the load condition on the bottleneck link [50].

1) Impact of the Number of Probe Packets: Figure 15(a) illustrates the time-dependent PCN traffic rate averaged over multiple simulation runs for different numbers of probe packets n_p. Significant overadmission occurs because PBAC cannot block reliably with excess traffic marking. With a certain probability all n_p probe packets remain unmarked in spite of pre-congestion. Then, PBAC does not block. Overadmission decreases with an increasing number of probe packets n_p. Only $n_p = 20$ or more probe packets limit the admitted PCN traffic rate to values close to the desired $AR = 8$ Mbit/s. On the one hand, the number of probe packets per admission request n_p should be large for correct admission decisions [38]. On the other hand, n_p should be small to keep the admission delay and probe traffic rate low. Especially during flash crowd events the probe traffic rate can become so significant [29] that other flows cannot be admitted. As PBAC probes the network per admission request, the rate of probe traffic is not bounded which is different for CLEBAC with probe traffic.

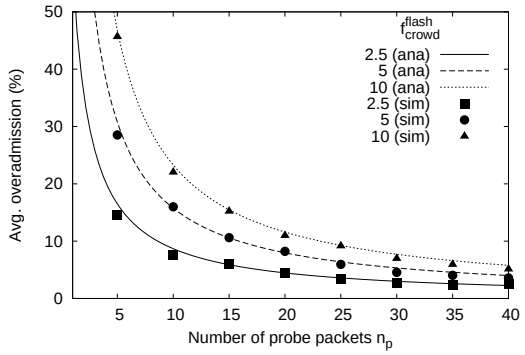
We derive a formula that predicts the expected average PCN traffic rate on the bottleneck link. Let n_{AR} be the number of admissible flows and n the number of admitted flows. PCN packets are not re-marked with a probability of $\frac{n_{AR}}{n}$. Thus, the probability that an admission request is falsely accepted can be computed by $p^{false}_{accept}(n) = \left(\frac{n_{AR}}{n}\right)^{n_p}$. Overadmission on the simulated link reaches an equilibrium when the rate of finishing flows equals the rate of falsely admitted flows in terms of number of admitted flows per second, i.e., $n \cdot \mu = p^{false}_{accept}(n) \cdot f^{flash}_{crowd} \cdot \lambda_{IEA}$. This is met for $\frac{n}{n_{AR}} = \sqrt[(n_p+1)]{f^{flash}_{crowd}}$. Thus, the expected overadmission is

$$OA = \sqrt[n_p+1]{f^{flash}_{crowd}} - 1. \quad (16)$$

Figure 15(b) shows the overadmission depending on the number of probe packets n_p for different flash crowd factors f^{flash}_{crowd}. The points in the figure correspond to the overadmission obtained from simulations and validate our analytical



(a) Time-dependent PCN traffic rate.



(b) Avg. overadmission for different flash crowd factors f_{crowd}^{flash}.

Fig. 15. Impact of the number of probe packets n_p for PBAC with excess traffic marking.

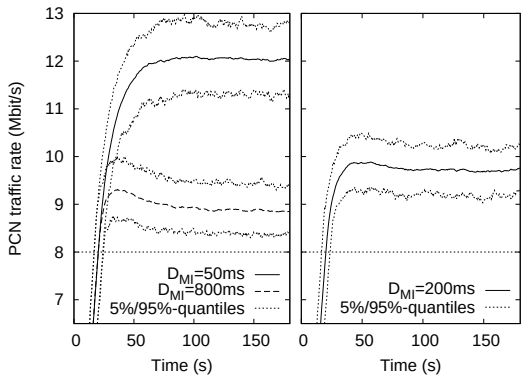
model. With $n_p = 20$ probes per admission request, the average overadmission is lower than 12.5% for flash crowd factors $f_{crowd}^{flash} \in \{2.5, 5, 10\}$. The delay introduced by $n_p = 20$ probe packets is about 400 ms and possibly not tolerable for session setup.

B. Overadmission for CLEBAC with Excess Traffic Marking

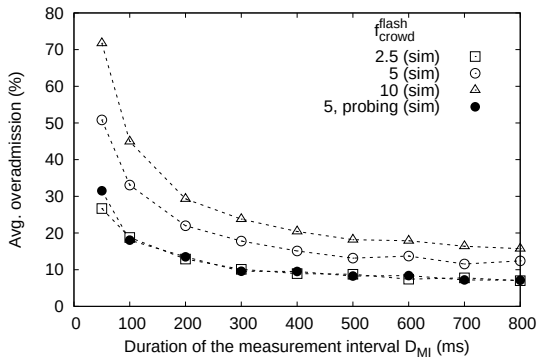
We set CLEBAC's CLE limit to $L_{CLE} = 0.025$ so that ingress nodes can block even if excess traffic marking re-marks only a very small fraction of PCN traffic. However, if the duration of the measurement interval D_{MI} is short and the packet frequency per IEA is low, it is likely that the egress node does not receive any re-marked PCN packets in spite of pre-congestion. Then, CLEBAC cannot block because the CLE value is zero.

We investigate the impact of D_{MI} on potential overadmission for $n_{IEA}^{flows} = 1$ flow per IEA. Figure 16(a) shows the time-dependent PCN traffic rate for different durations D_{MI}. A short $D_{MI} = 50$ ms leads to 50% average permanent overadmission, $D_{MI} = 200$ ms leads to about 20%, and $D_{MI} = 800$ ms leads to 10%. This is quite a lot as the same experiment for CLEBAC with threshold marking causes hardly any average overadmission as can be seen in the left part of Figure 12. Even for larger values of D_{MI}, overadmission cannot fall below $\frac{100\%}{1-L_{CLE}}$ (ca. 0.2 Mbit/s for $L_{CLE} = 0.025$) since admission control blocks only if more than $L_{CLE} \cdot 100\%$ of the PCN traffic is re-marked.

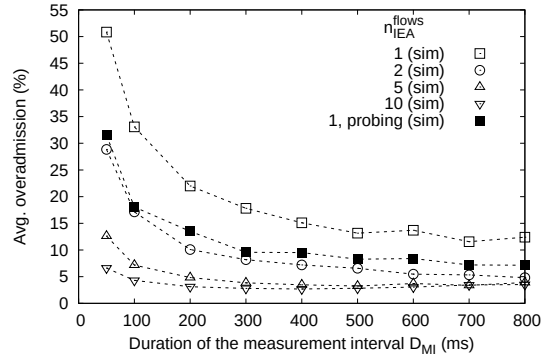
Figure 16(b) shows the average overadmission for different D_{MI} and different flash crowd factors $f_{crowd}^{flash} \in \{2.5, 5, 10\}$.



(a) Time-dependent PCN traffic rate for $f_{crowd}^{flash} = 5$ and $n_{IEA}^{flows} = 1$.



(b) Average overadmission for $n_{IEA}^{flows} = 1$ depending on D_{MI} and the flash crowd factor f_{crowd}^{flash}.



(c) Average overadmission for $f_{crowd}^{flash} = 5$ depending on D_{MI} and the number of flows per IEA n_{IEA}^{flows}.

Fig. 16. Impact of the duration of the measurement interval D_{MI} for CLEBAC with excess traffic marking.

Average overadmission quickly decreases with increasing duration of the measurement interval but we observe no significant reduction for values larger than $D_{MI} = 400$ ms. This is opposite to the finding for CLEBAC with threshold marking in Section IV-C2 where peak overadmission increases with D_{MI}. In both cases, overadmission clearly increases with the flash crowd factor f_{crowd}^{flash}.

Figure 16(c) performs similar experiments for different numbers of flows n_{IEA}^{flows} per IEA. We find extreme average overadmission for $n_{IEA}^{flows} = 1$, high overadmission for $n_{IEA}^{flows} = 2$, but for $n_{IEA}^{flows} = 5$ or larger, average overadmission is below 5% for duration of measurement intervals of $D_{MI} = 200$ ms

or larger. This still differs from CLEBAC with threshold marking which causes mainly peak overadmission, but hardly any average overadmission.

Figures 16(b) and 16(c) also show the average overadmission for CLEBAC with excess traffic marking if probing according to Section II-E3 is applied with $n_p = 10$ for $f_{crowd}^{flash} = 5$. Figure 16(b) illustrates that probing mitigates overadmission, but cannot remove it. This is different in Figure 11(b) for CLEBAC with threshold marking where $n_p = 2$ already avoids permanent overadmission. Figure 16(c) suggests that overadmission caused in spite of probing with $n_p = 10$ is similar to overadmission caused in the presence of $n_{IEA}^{flows} = 2$ flows per IEA, which is still quite a lot. Hence, probing is not effective for CLEBAC with excess traffic marking.

VI. Summary

We have shown that PCN-based AC may lead to temporary or even stationary overadmission in the presence of flash crowds. Its intensity depends on the marking algorithm and on the AC method.

With PBAC and threshold marking, overadmission mainly depends on the blocking delay D_{block} which is the time during which flows are still admitted although they should better be blocked. The blocking delay depends on multiple other parameters such as the marker configuration, round-trip time, and media delay. Especially the effect of media delay can be significant. Other sources of delay that we have not investigated, such as processing delay, can also contribute to blocking delay and overadmission.

With CLEBAC and threshold marking, overadmission depends on the same sources of delay plus an additional measurement delay. To work well, CLEBAC requires $n_{IEA}^{flows} = 1$ flow per IEA in case of CBR voice traffic and about $n_{IEA}^{flows} = 5$ flows per IEA in case of on/off traffic because empty IEAs cannot block new admission request in the presence of pre-congestion. Our proposal to use infrequent probing for empty IEAs solves that problem and extends the applicability of CLEBAC towards very small numbers of flows per IEA. We showed that CLE smoothing contributes to blocking delay and increases overadmission. CLEBAC may lead to underadmission in case of multipath routing as it interprets measured pre-congestion values per IEA and not per path. Flow termination suffers from a similar problem which can be repaired by signalling information about observed re-marked flows from the egress node to the ingress node. However, we do not see a similar method to avoid this inefficiency for CLEBAC. PBAC works well with multipath routing.

PBAC with excess traffic marking suffers from the same sources of overadmission as PBAC with threshold marking. In addition, none of the probe packets may be re-marked in the presence of pre-congestion so that flows are falsely admitted. Therefore, PBAC with excess traffic marking leads to significantly larger overadmission than with threshold marking, it adds additional admission delay and causes more probe traffic in particular during flash crowd events.

CLEBAC with excess traffic marking also faces the same problems as CLEBAC with threshold marking, but we also witness that flows are admitted in spite of long pre-congestion. However, this causes only little additional overadmission if the number of packets per measurement interval is large enough which can be achieved by long measurement intervals or a large number of flows per IEA n_{IEA}^{flows}. The method does not work well with small n_{IEA}^{flows} and probing does not help effectively in this case.

In this study, we have presented experiments for single-link scenarios only. In multiple-link scenarios with cross traffic, multiple bottlenecks can occur. While we have reported interesting effects of overtermination for such scenarios in [51], we could not find analogous results for admission control. We explain why multiple bottlenecks cause neither underadmission nor overadmission. A new flow should be blocked as soon as a single link of the flow's path is pre-congested. That means, if traffic is re-marked by multiple bottleneck links, the admission of new flows is not prevented by mistake because already one of these bottlenecks justifies flow blocking. Conversely, multiple bottlenecks cannot cause overadmission because they can only increase the fraction of re-marked packets which rather leads to faster instead of slower flow blocking.

Furthermore, we have considered mainly rather simple traffic models to keep experiments and explanations simple. More sophisticated traffic models and traffic mixes thereof have been looked at in [35]. However, a systematic study of complex traffic characteristics for different AC methods and marking algorithms under challenging conditions is still to be done.

VII. Conclusion

We investigated various admission control (AC) methods using pre-congestion notification (PCN) during flash crowd events when the rate of admission requests is by a factor f_{crowd}^{flash} larger than the network was designed for. With PCN based on threshold marking, overadmission is caused due to late blocking while with PCN based on excess traffic marking, overadmission is also caused by weak pre-congestion signals. We studied probe-based AC (PBAC) and the more complex congestion-level estimate based AC (CLEBAC) using simulation and mathematical modeling to deepen the understanding of the observed results. We deliberately looked for challenging conditions such as low traffic aggregation, on/off traffic, delayed media, multipath routing, or various network and configuration parameters to provide insight in the applicability of PCN-based AC. To extend the applicability of CLEBAC, we proposed to add probe traffic in case of empty ingress-egress aggregates and showed that this solves observed problems.

Acknowledgements

The authors would like to thank Joe Babiarz, Anna Charny, Michael Hoefling, Alfons Martin, and Xinyang Zhang for their fruitful discussions.

References

[1] S. Blake, D. L. Black, M. A. Carlson, E. Davies, Z. Wang, and W. Weiss, "RFC2475: An Architecture for Differentiated Services," Dec. 1998.

[2] J. Wroclawski, "RFC2211: Specification of the Controlled-Load Network Element Service," Sep. 1997.

[3] P. Eardley (Ed.), "RFC5559: Pre-Congestion Notification (PCN) Architecture," Jun. 2009.

[4] ——, "RFC5670: Metering and Marking Behaviour of PCN Nodes," Nov. 2009.

[5] M. Menth et al., "A Survey of PCN-Based Admission Control and Flow Termination," *IEEE Communications Surveys & Tutorials*, vol. 12, no. 3, 2010.

[6] A. Charny et al., "PCN Boundary Node Behavior for the Controlled Load (CL) Mode of Operation," Internet draft, Jun. 2011.

[7] ——, "PCN Boundary Node Behavior for the Single-Marking (SM) Mode of Operation," Internet draft, Jun. 2011.

[8] M. Menth and F. Lehrieder, "PCN-Based Measured Rate Termination," *Computer Networks*, vol. 54, no. 13, pp. 2099 – 2116, Sep. 2010.

[9] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler, "RFC3261: SIP: Session Initiation Protocol," Jun. 2002.

[10] B. Braden, L. Zhang, S. Berson, S. Herzog, and S. Jamin, "RFC2205: Resource ReSerVation Protocol (RSVP) - Version 1 Functional Specification," Sep. 1997.

[11] Y. Bernet, P. Ford, R. Yavatkar, F. Baker, L. Zhang, M. Speer, R. Braden, B. Davie, J. Wroclawski, and E. Felstaine, "RFC2998: A Framework for Integrated Services Operation over Diffserv Networks," Nov. 2000.

[12] B. Briscoe, T. Moncaster, and M. Menth, "Encoding 3 PCN-States in the IP Header Using a Single DSCP," Internet draft, Aug. 2011.

[13] T. Moncaster, B. Briscoe, and M. Menth, "RFC5696: Baseline Encoding and Transport of Pre-Congestion Information," Nov. 2009.

[14] M. Menth, "Efficient Admission Control and Routing in Resilient Communication Networks," PhD thesis, Univ. of Würzburg, Jul. 2004.

[15] S. Floyd and V. Jacobson, "Random Early Detection Gateways for Congestion Avoidance," *IEEE/ACM Transactions on Networking*, vol. 1, no. 4, pp. 397–413, Aug. 1993.

[16] B. Braden et al., "RFC2309: Recommendations on Queue Management and Congestion Avoidance in the Internet," Apr. 1998.

[17] K. Ramakrishnan, S. Floyd, and D. Black, "RFC3168: The Addition of Explicit Congestion Notification (ECN) to IP," Sep. 2001.

[18] N. Spring, D. Wetherall, and D. Ely, "RFC3540: Robust Explicit Congestion Notification (ECN)," Jun. 2003.

[19] K. Nichols, S. Blake, F. Baker, and D. L. Black, "RFC2474: Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers," Dec. 1998.

[20] S. Floyd, "RFC4774: Specifying Alternate Semantics for the Explicit Congestion Notification (ECN) Field," Feb. 2007.

[21] W. Almesberger, T. Ferrari, and J.-Y. Le Boudec, "SRP: A Scalable Resource Reservation for the Internet," *Computer Communications*, vol. 21, no. 14, pp. 1200–1211, Nov. 1998.

[22] I. Stoica and H. Zhang, "Providing Guaranteed Services without per Flow Management," in *ACM SIGCOMM*, Boston, MA, Sep. 1999.

[23] R. Szábó, T. Henk, V. Rexhepi, and G. Karagiannis, "Resource Management in Differentiated Services (RMD) IP Networks," in *ICETA*, Oct. 2001.

[24] R. J. Gibbens and F. P. Kelly, "Resource Pricing and the Evolution of Congestion Control," *Automatica*, vol. 35, no. 12, pp. 1969–1985, 1999.

[25] ——, "Distributed Connection Acceptance Control for a Connectionless Network," in 16^{th} *ITC*, Edinburgh, UK, Jun. 1999.

[26] F. Kelly, P. Key, and S. Zachary, "Distributed Admission Control," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 12, 2000.

[27] M. Karsten and J. Schmitt, "Admission Control based on Packet Marking and Feedback Signalling – Mechanisms, Implementation and Experiments," Darmstadt University of Technology, Technical Report 03/2002.

[28] ——, "Packet Marking for Integrated Load Control," in *IFIP/IEEE Symposium on Integrated Management (IM)*, 2005.

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

[30] I. Más and G. Karlsson, "PBAC: Probe-Based Admission Control," in *QoFIS*, 2001.

[31] I. Mas and G. Karlsson, "A Model for Endpoint Admission Control Based on Packet Loss," in *IFIP Networking*, Singapore, May 2008.

[32] O. Hagsand, I. Más, I. Marsh, and G. Karlsson, "Self-Admission Control for IP Telephony Using Early Quality Estimation," in *IFIP Networking*, 2004.

[33] M. Menth, R. Martin, and J. Charzinski, "Capacity Overprovisioning for Networks with Resilience Requirements," in *ACM SIGCOMM*, 2006.

[34] S. Iyer, S. Bhattacharyya, N. Taft, and C. Diot, "An Approach to Alleviate Link Overload as Observed on an IP Backbone," in *IEEE Infocom*, San Francisco, CA, April 2003.

[35] X. Zhang and A. Charny, "Performance Evaluation of Pre-Congestion Notification," in *IWQoS*, 2008.

[36] S. Latre, B. De Vleeschauwer, W. Van de Meerssche, S. Perrault, F. De Turck, P. Demeester, K. De Schepper, C. Hublet, W. Rogiest, S. Custers, and W. Van Leekwijck, "An Autonomic PCN based Admission Control Mechanism for Video Services in Access Networks," in *IEEE ACNM*, 2009.

[37] S. Latre, B. De Vleeschauwer, W. Van de Meerssche, F. De Turck, and P. Demeester, "Design and Configuration of PCN Based Admission Control in Multimedia Aggregation Networks," in *IEEE Globecom*, 2009.

[38] M. Menth and F. Lehrieder, "Performance Evaluation of PCN-Based Admission Control," in *IWQoS*, 2008.

[39] M. Arumaithurai, R. Geib, R. Rex, and X. Fu, "Pre-Congestion Notification-based Flow Management in MPLS-based DiffServ Networks," in *IEEE IPCCC*, 2009.

[40] M. Menth and M. Hartmann, "Threshold Configuration and Routing Optimization for PCN-Based Resilient Admission Control," *Computer Networks*, vol. 53, no. 11, pp. 1771 – 1783, Jul. 2009.

[41] M. Menth, A. Binzenhöfer, and S. Mühleck, "Source Models for Speech Traffic Revisited," *IEEE/ACM Transactions on Networking*, vol. 17, no. 4, pp. 1042–1051, Aug. 2009.

[42] S. Andersen, A. Duric, H. Astrom, R. Hagen, W. Kleijn, and J. Linden, "RFC3951: Internet Low Bit Rate Codec (iLBC)," Dec. 2004.

[43] S. Mühleck, "Modelling and Performance Evaluation of Periodic Real-time Traffics," Master's thesis, Univ. of Würzburg, 2007.

[44] J. Jung, B. Krishnamurthy, and M. Rabinovich, "Flash Crowds and Denial of Service Attacks: Characterization and Implications for CDNs and Web Sites," in *WWW*, 2002.

[45] I. Ari, B. Hong, E. L. Miller, S. A. Brandt, and D. D. E. Long, "Managing Flash Crowds on the Internet," in *MASCOTS*, 2003.

[46] X. Chen and J. Heidemann, "Flash Crowd Mitigation via Adaptive Admission Control Based on Application-Level Observation," *ACM Transactions on Internet Technology*, vol. 5, no. 3, pp. 532–562, Aug. 2005.

[47] D. S. Seibel (Broadcasting & Cable), "American Idol Outrage: Your Vote Doesn't Count," http://www.broadcastingcable.com/article/CA417981.html, May 2004.

[48] P. Eardley, "Traffic Matrix Scenario," http://www.ietf.org/mail-archive/web/pcn/current/msg00831.html, Oct. 2007.

[49] B. Briscoe et al., "An Edge-to-Edge Deployment Model for Pre-Congestion Notification: Admission Control over a DiffServ Region," Internet draft, Oct. 2006.

[50] J. Sommers, P. Barford, N. G. Duffield, and A. Ron, "Improving Accuracy in End-to-End Packet Loss Measurement," in *ACM SIGCOMM*, 2005.

[51] F. Lehrieder and M. Menth, "PCN-Based Flow Termination with Multiple Bottleneck Links," in *IEEE ICC*, 2009.

Michael Menth is a full professor at the Department of Computer Science at the University of Tuebingen/Germany and head of the Communication Networks chair. He received a Diploma and PhD degree in 1998 and 2004 from the University of Wuerzburg/Germany. Prior he was studying computer science at the University of Texas at Austin and worked at the University of Ulm/Germany. His special interests are performance analysis and optimization of communication networks, resource management, resilience issues, and Future Internet. He holds numerous patents and received various scientific awards for innovative work.

Frank Lehrieder studied computer science and business administration at the University of Wi£¡rzburg/Germany and University Antonio de Nebrija in Madrid/Spain. In May 2008, he received his Master degree in computer science from University of Wi£¡rzburg. Now, he is a researcher at the Chair of Communication Networks in Wi£¡rzburg and pursuing his PhD. His special interests are performance evaluation and resource management for peer-to-peer networks and Future Internet infrastructures.