©2009 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. The definitive version of this paper has been published in IEEE International Conference on Communications (ICC) 2009 10.1109/icc.2009.5199164.

Marking Conversion for Pre-Congestion Notification

Frank Lehrieder and Michael Menth
University of Würzburg, Institute of Computer Science, Germany

***Abstract*—Pre-congestion notification (PCN) defines admissible rates (AR) and supportable rates (SR) per link and marks the PCN traffic rate above these thresholds as AR- or SR-overload. The IETF standardizes simple mechanisms for admission control (AC) and flow termination (FT) based on this PCN-feedback for high-priority DiffServ traffic. While admission control (AC) has been extensively discussed in the literature, flow termination (FT) is a new control function. In this paper we propose an algorithm that converts marked *AR*-overload into marked *SR*-overload by unmarking appropriate packets. Classic marked flow termination (MFT) is based on marked *SR*-overload and works well even with a small number of PCN flows per ingress-egress aggregate and in case of multipath routing. Thanks to the new marking converter MFT also works with marked *AR*-overload so that a single marking scheme suffices to support AC and FT. We investigate whether MFT with marking conversion based on *AR*-overload retains the benefits classic MFT.**

I. Introduction

Pre-congestion notification (PCN) is a new mechanism currently developed by the IETF to facilitate PCN-based admission control (AC) and flow termination (FT) primarily for wired networks and inelastic realtime flows [1]. Traffic belonging to the PCN service class is prioritized over non-PCN traffic, which is essentially the DiffServ principle, and hence PCN traffic does not suffer from packet loss or delay when overload occurs in a network. In addition, the rate of admitted PCN traffic is controlled such that overload cannot evolve within the PCN traffic class under normal operation. If the rate of PCN traffic becomes too large in case of a failure with subsequent rerouting, FT can remove some of the admitted traffic to restore a controlled load condition [2] on the overloaded link. The idea of PCN is that routers mark PCN packets on outgoing links when their PCN traffic rates exceed their configured admissible or supportable rates. Currently, PCN-based AC and FT is developed for a domain concept. That means egress nodes evaluate the PCN packet markings and communicate the information about marked packets to ingress nodes which block admission requests for new PCN flows or terminate already admitted flows if required. An overview of existing techniques is provided in [3].

PCN uses excess marking on a link to mark PCN packets that exceed the admissible rate (*AR*) or supportable rate (*SR*) of that link, i.e. the so-called *AR*- or *SR*-overload. Classic marked flow termination (MFT) is based on marked *SR*-overload and works well even with a small number of PCN flows per ingress-egress aggregate and in case of multipath routing. In this work, we propose a new algorithm that converts marked *AR*-overload into marked *SR*-overload. Thanks to this algorithm, MFT also works with marked *AR*-overload so that a single marking scheme in a network suffices to support AC and FT. However, it is not clear whether MFT with marking conversion based on *AR*-overload retains the benefits classic MFT. Our performance evaluation investigates this issue.

The paper is structured as follows. Sect. II explains PCN, metering and marking algorithms, various FT algorithms, and our new marking conversion algorithm. Sect. III reviews related work. Sect. IV studies the applicability of the marking conversion algorithm under various conditions. Finally, Sect. V summarizes this work and draws conclusions.

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

II. Pre-Congestion Notification (PCN)

In this section we review the general idea of PCN-based admission control (AC) and flow termination (FT) and illustrate their application in a domain context in the Internet. We revise the metering and marking algorithms and marked flow termination (MFT) for so-called ingress-egress aggregates. Finally, we present our new marking conversion algorithm and explain its application with MFT.

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. This implies three different load regimes as illustrated in Fig. 1. If the PCN traffic rate $r(l)$ is below $AR(l)$, there is no pre-congestion and further flows may be admitted. If the PCN traffic rate $r(l)$ is above $AR(l)$, the link is *AR*-pre-congested and the rate above $AR(l)$ is *AR*-overload. 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 and the rate above $SR(l)$ is *SR*-overload. In this state, some already admitted flows should be terminated to reduce the PCN rate $r(l)$ below $SR(l)$.

B. Edge-to-Edge PCN

Edge-to-edge PCN assumes that some end-to-end signalling protocol (e.g. SIP or RSVP) or a similar mechanism requests admission for a new flow to cross a so-called PCN domain similar to the IntServ-over-DiffServ concept [4]. Thus, edge-to-edge PCN is a per-domain QoS mechanism and presents an alternative to RSVP clouds or extreme capacity overprovisioning. Traffic enters a PCN domain only through PCN ingress nodes and leaves it only through PCN egress nodes. Ingress nodes set

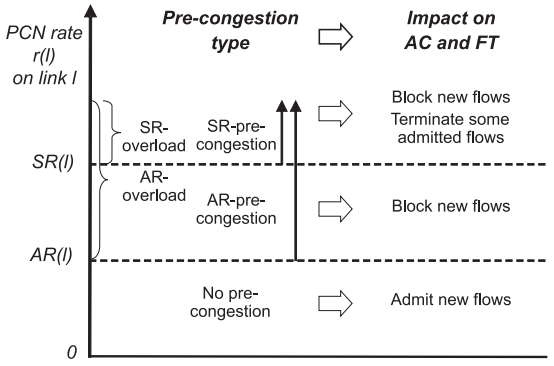

Fig. 1. The admissible and the supportable rate $(AR(l), SR(l))$ define three types of pre-congestion.

a special header codepoint to make the packets distinguishable from other 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 remark the traffic in case of *AR*- or *SR*-pre-congestion. PCN egress nodes evaluate the markings of the traffic and send a digest to the AC and FT entities of the PCN domain. The overview in [3] presents different algorithms for these purposes which are not necessarily compatible with each other. Many of them require the notion of an ingress-egress aggregate (IEA) which is the ensemble of all PCN flows between a specific ingress and egress node of a PCN domain. In the following, we review only the metering and marking algorithm and flow termination algorithm that are relevant to our study.

C. Excess Marking

Excess marking uses a token bucket based meter that tracks whether a certain reference rate is exceeded and marks only those packets that exceed the reference rate. The rate of marked packets provides an estimate of the rate by which the reference rate was exceeded while the rate of unmarked packets corresponds to the reference rate. Excess marking can be implemented with only few modifications of existing hardware.

When PCN nodes perform excess marking based on the admissible rate, the markings can be used for AC decisions. As soon as packets are marked, the PCN traffic rate has exceeded the admissible rate on some link in the network and requests for flows using this link are rejected. We call this kind of marking "admission-stop" (AS) marking and packets are AS-marked.

When PCN nodes perform excess marking based on the supportable rate, the markings can be used for FT decisions. The marked traffic rate corresponds to the *SR*-overload which is the traffic rate to be terminated. We call this kind of marking "excess-traffic" (ET) marking and packets are ET-marked. However, some FT algorithms also work with AS-marking when the supportable rate $SR(l) = u \cdot AR(l)$ is a fixed multiple u of the admissible rate on all links l of the PCN domain. Then, the termination rate can be derived from the unmarked traffic rate, the amount of marked *AR*-overload, and the parameter u. The advantage of these FT algorithms is that AS-marking alone possibly suffices to facilitate AC and FT which simplifies the operation of the PCN nodes.

D. Marked Flow Termination for Ingress-Egress Aggregates

The principle of marked flow termination (MFT) is to terminate flows only if at least one of their packets was marked. Various MFT methods have been proposed in [5], but in this work we focus on MFT for IEAs and describe it in the following. MFT assumes that PCN nodes perform ET-marking. Each egress node maintains a credit counter c_g for each IEA g, i.e. for each ingress node. When a marked packet for a specific IEA g arrives and the credit counter c_g is not negative, c_g is decremented by the size of the marked packet; if c_g is negative, the egress node triggers the termination of a recently marked flow f of the IEA g and c_g is incremented by $\frac{2 \cdot E[D_T] \cdot R_f}{\alpha}$. R_f is the rate of the terminated flow f. $E[D_T]$ is a preconfigured value that estimates the delay from the termination trigger by the egress node until the termination becomes visible at the egress node. The termination aggressiveness α controls the termination speed. Larger values lead to faster termination and values larger than $\alpha > 1$ may lead to overtermination. Therefore, $\alpha = 1$ is recommended in [5] for most cases. Initially, c_g is initialized randomly according to an exponential distribution with a mean value of $\frac{2 \cdot E[D_T] \cdot R_f}{\alpha}$ when the first flow of that IEA is admitted. MFT reduces the load on the bottleneck link gradually, i.e. one flow after another. If the *SR*-overload is large, flows are quickly terminated while flows are slowly terminated when the *SR*-overload is small. However, the overall termination process is fast as most of the *SR*-overload is usually removed within $3 \cdot E[D_T]$. The advantage of MFT is that it works well with any number of flows per IEA and with multipath routing. Most other FT methods [3] fail under these circumstances.

E. Marking Conversion

We present an algorithm that converts a stream with AS- and non-AS-marked packets into a stream with ET- and non-ET-marked packets by deleting some of the markings. To that end, we use the assumption $SR = u \cdot AR$. The packets of the input stream are consecutively feeded into the algorithm. The algorithm is based on a token bucket (TB) with size S and fill state F. It differs from conventional TB implementations as it does not have a constant fill rate R. Its operation is explained in Algorithm 1. The number of tokens in the bucket F indicates how many AS-marked bytes can be re-marked to unmarked. For each non-AS-marked byte, the fill state F is incremented by $u-1$ tokens. When a packet is AS-marked and if the fill state F is not negative, the packet is re-marked to unmarked and the fill state of the TB is reduced by the packet size B. Otherwise, the packet remains marked which is then interpreted as ET-marking.

A sufficiently large TB size S is needed to tolerate short-term variations, i.e. a burst of S AS-marked bytes should not be ET-marked when the overall fraction of AS-marked bytes is low. However, this tolerance also delays initial re-marking.

F. MFT Based on Converted AR-Overload

PCN nodes running excess marking based on the admissible rate generate marked *AR*-overload. We propose that egress

Input: token bucket parameters S and F, packet size B and marking M

if $(M ==$ unmarked$)$ **then**
  $F = \min(S, F + (u-1) \cdot B)$;
else if $(F \geq 0)$ **then** $\{(M == \text{AS})\}$
  $F = F - B$;
  $M =$ unmarked;
else
  $M =$ ET;
end if

Algorithm 1: MARKING CONVERSION: converts a stream with AS- and non-AS-marked packets into a stream with ET- and non-ET-marked packets.

nodes evaluate the original marking to support AC decisions. In addition, the packet stream of the IEA is passed to a marking converter such that an equivalent marked *SR*-overload is also available. Based on this converted ET-marked packet stream MFT may be applied. The advantage of this approach is obvious: only a single metering and marking scheme is needed to support both AC and MFT. However, it is not clear whether the benefits of MFT can be retained.

III. Related Work

An overview of PCN including a multitude of AC and FT mechanisms is given in [3]. It also reviews related work regarding the historical roots of PCN. In [6], a high level summary is provided about a large set of simulation results regarding PCN-based AC and FT which shows that these methods work well in most studied cases.

In contrast to excess marking, exhaustive marking is intended to mark all packets if a given reference rate is exceeded. Ramp marking and threshold marking are two different implementation options for that purpose. Their impact on packet marking probabilities has been investigated in [7]. It turned out that threshold marking is as good as ramp marking which excluded ramp marking from further consideration because it is more complex than threshold marking.

A two-layer architecture for PCN-based AC and FT was presented in [8] and flow blocking probabilities have been studied for single aggregates and static load conditions. The work presented in [5] proposes various algorithms for PCN-based marked flow termination (MFT) and gives recommendations for their configuration. It assumes that PCN marking is based on *SR*-overload. In this paper, we use one of the mechanisms proposed in [5], adapt it to PCN marking based on *AR*-overload, and evaluate the performance. Overtermination due to multiple bottlenecks is studied in [9].

The efficiency of resilient PCN-based AC with flow termination and other resilient AC methods without flow termination in optimally dimensioned networks is evaluated in [10]. An additional investigation about how *AR* and *SR* thresholds should be set in PCN domains with resilience requirements is contained in [11]. Furthermore, it studies how link weights should be set in IP networks in order to maximize the admissible traffic rates. The authors of [12] investigate the impact of admissible and supportable rate thresholds on the admission and termination of on/off traffic.

IV. Termination Behavior of MFT with Marking Conversion

We have simulated the termination process of MFT with marking conversion when *SR*-overload occurs. The time-dependent PCN traffic rate behaves like with classic MFT in [5]. Thus, it works as intended. We validated that for several different scenarios and parameter settings, but we do not show these results here. Instead, we study in this section whether MFT with marking conversion leads to terminated flows without *SR*-pre-congestion. This possibly happens in case of a small number of flows per IEA and in case of multipath routing. We use a packet-based simulation to investigate the first issue and a mathematical analysis based on a discrete-time Markov chain to clarify the second issue.

A. Simulation Setup

We simulate PCN flows that are homogeneous and periodic with a deterministic packet inter-arrival time $A = 20$ ms and packet size $B = 200$ byte. Thus, their rate is $E[R_f] = 80$ kbit/s. To avoid simulation artifacts due to marking synchronization for periodic traffic, we add an equally distributed random delay of up to 1 ms to the theoretic arrival instant of every packet. This assumption is realistic because realtime applications send traffic periodically, but packets arrive at the bottleneck link with a small jitter.

We simulate the time-dependent PCN traffic rate $r(t)$ on a bottleneck link that is shared by n_{IEA} IEAs, each of which carries n_{IEA}^{flows} flows. Its supportable rate is $SR = n_{IEA} \cdot n_{IEA}^{flows} \cdot E[R_f]$. Thus, the load on the bottleneck link is exactly the supportable rate so that no traffic should be terminated. We use $u = 2$, i.e. $SR = 2 \cdot AR$, and set the bucket size of the marker and meter on the bottleneck link to $0.05\ \text{s} \cdot AR$. Marking converters and subsequent MFT are applied per IEA. The fill state F of the bucket of each marking converter is initialized with the size S, i.e., all buckets are full at simulation start. For the configuration of MFT, we set the termination aggressiveness to $\alpha = 1$. Furthermore, we set the termination delay to $D_T = 50$ ms and use this value also for the configuration of MFT (cf. Sect. II-F). According to [5], overtermination does not occur with these values when MFT is based on *SR*-overload.

Our simulator is a custom-made Java tool. The presented time-dependent PCN rate is calculated based on 100 ms long measurement intervals. We perform multiple experiments and report average results in our figures. In particular, the initial arrival pattern of the flows is different for every simulation run. We run so many simulations that the 95% confidence intervals are small. However, we omit them in the figures for the sake of easier readability.

B. Impact of the Bucket Size S of the Marking Converter on the Termination Behavior

We analyze the impact of the bucket size S of the marking converter. The bottleneck link l is shared by $n_{IEA} = 100$ IEAs

and each IEA is dimensioned for $n_{IEA}^{flows} = 10$ flows. The *SR* is set exactly to the overall rate of these flows so that *SR*-pre-congestion does not occur. The results are shown in Fig. 2. For small bucket sizes of 0.6, 2, and 6 kb (3, 10, and 30 packets) we observe a significant amount of falsely terminated traffic although the supportable rate was not exceeded on any link. Although the load on the bottleneck link is configured so that it is *AR*-pre-congested, i.e. some packets are AS-marked, but not *SR*-pre-congested, MFT obviously detects *SR*-pre-congestion on an IEA basis and terminates flows. The reason is that the AS-marked packets on the bottleneck link randomly belong to IEAs. Therefore, the short-term fraction of AS-marked packets may be larger than $\frac{u-1}{u}$ for an IEA so that *SR*-pre-congestion is recognized, packets are ET-marked, and flows are possibly terminated. Larger bucket sizes can tolerate larger bursts of AS-marked packets without ET-marking them in spite of missing *SR*-pre-congestion on the bottleneck link. Thus, large bucke sizes of 20, 60, and 200 kb (100, 300, and 1000 packets) lead to only little overtermination. Bucket sizes larger than 300 packets do not seem to reduce overtermination any further.

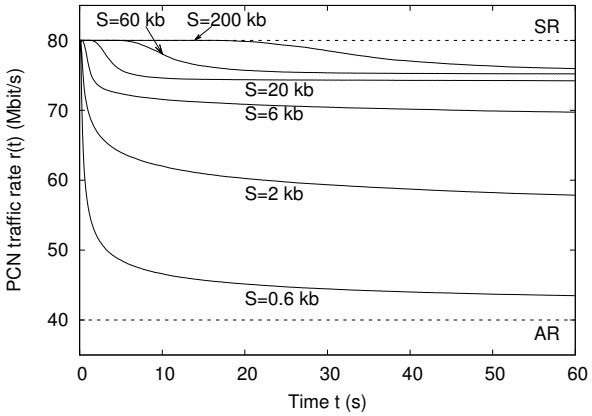

Fig. 2. Impact of the bucket size S on the amount of falsely terminated flows ($n_{IEA} = 100$, $n_{IEA}^{flows} = 10$).

C. Impact of the Number of Supportable Flows per IEA n_{IEA}^{flows}

We study the impact of the number of supportable flows per IEA n_{IEA}^{flows} while keeping the number of flows on the bottleneck link constant at $n = 1000$. Thus, we repeat the experiment of the previous section with different values for $n_{IEA}^{flows} \in \{1, 4, 10, 40, 100\}$ and the number of IEAs on the bottleneck link is $n_{IEA} = \frac{1000}{n_{IEA}^{flows}}$. We report the results without figures as they are similar to those in Fig. 2. For a given bucket size S, the amount of overtermination is nearly the same for all studied values of n_{IEA}^{flows}. Thus, the bucket size can be set independently of the number of flows per IEAs n_{IEA}^{flows}. This experiment also shows that the number of IEAs sharing the bottleneck link has no impact on the amount of overtermination for $n_{IEA} = 10$ or larger.

D. Impact of the Number of Flows on the Bottleneck Link

To study the impact of the number of flows on the bottleneck link, we vary the number of IEAs each of which carries $n_{IEA}^{flows} =$ 3 flows. Again, the supportable rate is $SR = n_{IEA} \cdot n_{IEA}^{flows} \cdot E[R_f]$. We set the bucket size to $S = 6$ kb which is so small that about 12% overtermination occurs in Fig. 2. Fig. 3 shows the time-dependent PCN traffic rate $r(t)$ normalized by the supportable rate *SR*. Overtermination decreases with an increasing number of IEAs n_{IEA} on the bottleneck link which seems to be a contradiction to the findings in Sect. IV-C. The reason for this phenomenon is that a sufficiently large packet rate is required on the bottleneck link to produce random marks. If the number of IEAs on the bottleneck and the number of flows per IEA are small, the packet rate on the bottleneck link is also small. This leads to an almost deterministic system where packets of a single IEA are AS-marked with a higher probability than packets of another IEA. Therefore, the marking converter is likely to ET-mark packets of that IEA so that a flow of that IEA is possibly terminated. After the termination of that flow, the same may happen to another IEA. Increasing the number of IEAs on the bottleneck link n_{IEA}, but also increasing the number of flows per IEA n_{IEA}^{flows} (not shown), decreases the overtermination. Thus, the findings in the sections above are valid only if the packet marking process is sufficiently random from an IEA point of view.

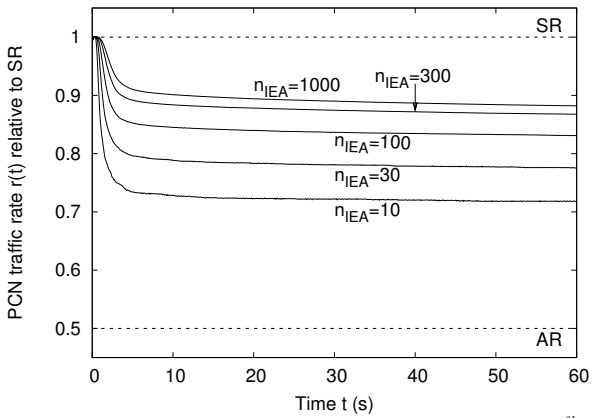

Fig. 3. Impact of the number of IEAs n_{IEA} on the overtermination ($n_{IEA}^{flows} = 3$, $S = 6$ kb).

E. Impact of the Mean Packet Size $E[B]$

We explain the impact of the mean packet size $E[B]$ for constant bucket sizes S. The marking converter is able to tolerate bursts of $\frac{S}{E[B]}$ AS-marked packets. Thus, increasing the packet size decreases the size of tolerable bursts of AS-marked packets without ET-marking outgoing packets. Conversely, marking converters require larger bucket sizes S for traffic with larger packet sizes $E[B]$ to avoid more overtermination. We validated this hypothesis by simulation experiments but omit the figures due to the page limit.

F. Impact of the Bucket Size S on the Conversion Delay D_c

We derive the conversion delay D_c which is induced by the marking converter from the arrival of the first *AS*-marked packets due to sudden *SR*-overload until the marking converter ET-marks the first packets. To that end, we assume that the token bucket of the marking converter is fully filled with S

bytes. Furthermore, the rate on the bottleneck link suddenly increases to $f_{OL}^{SR} \cdot SR$. Thus, we observe an *SR*-overload of $SRO = (f_{OL}^{SR} - 1) \cdot n_{IEA}^{flows} \cdot E[R_f]$ per IEA. Therefore, the fill state F decreases with rate *SRO* and it takes

$$D_c = \frac{S}{SRO} = \frac{S}{(f_{OL}^{SR} - 1) \cdot n_{IEA}^{flows} \cdot E[R_f]} \quad (1)$$

until the first AS-marking is converted to an ET-marking. We call this time conversion delay D_c. Larger bucket sizes S lead to longer D_c. D_c also depends on the overload factor f_{OL}^{SR}: in case of severe overload, it reacts faster than in case of moderate overload. Furthermore, the reaction time decreases with an increasing number of supportable flows per IEA n_{IEA}^{flows}. Hence, we conclude that an appropriate value of S is a trade-off between fast termination in case of *SR*-overload and avoiding unintended termination without *SR*-pre-congestion.

G. Configuration of the Bucket Size S for a Fixed Conversion Delay D_c

We configure the bucket size of the marking converter S so that the conversion delay D_c is limited for a given overload factor f_{OL}^{SR}:

$$S = (f_{OL}^{SR} - 1) \cdot n_{IEA}^{flows} \cdot E[R_f] \cdot D_c. \quad (2)$$

Thus, the bucket size S scales with the number of flows per IEA n_{IEA}^{flows}. The conversion delay should be small for the sake of a fast termination time. We postulate that it should be $D_c = 0.1$ s for an overload factor of $f_{OL}^{SR} = 2$, i.e., $S = n_{IEA}^{flows} \cdot 1$kb which corresponds to 5 packets of 200 bytes per flow. We study the impact of this dimensioning rule on the expected overtermination without *SR*-overload. To that end, we consider a bottleneck link with $n = 1000$ flows that are grouped into IEAs with $n_{IEA}^{flows} \in \{1, 4, 10, 40, 100\}$ flows per IEA and the number of IEAs on the bottleneck link is $n_{IEA} = \frac{1000}{n_{IEA}^{flows}}$. The bucket size for the converter is set according to Eqn. (2).

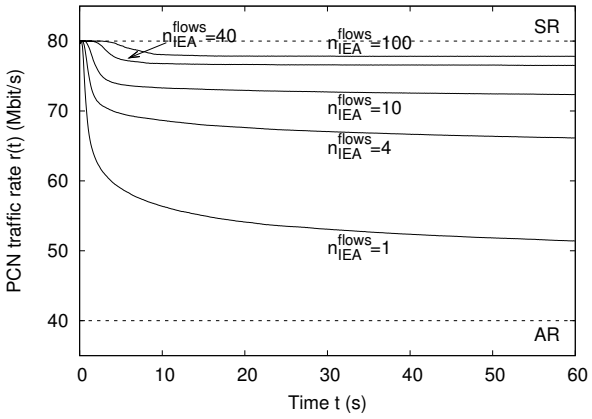

Fig. 4. Termination behavior on a bottleneck link with $n = 1000$ flows for a conversion delay of $D_c = 0.1$ s at an overload factor of $f_{OL}^{SR} = 2$.

Fig. 4 shows that significant overtermination occurs when the number of flows per IEA is small, i.e. at most $n_{IEA}^{flows} = 10$. For such applications a larger bucket size is required to avoid overtermination which essentially increases the conversion delay D_c. Then, MFT with marking conversion based on marked *AR*-overload becomes slow which is not acceptable in practice. For IEAs with $n_{IEA}^{flows} = 40$ or more, the observed overtermination seems to be small. Hence, for these scenarios MFT with marking conversion based on marked *AR*-overload is a feasible solution.

H. Impact of Multipath Routing

In this section we study the termination behavior of MFT with marking conversion based on marked *AR*-overload in case of multipath routing. The marking conversion depends on the rate of AS- and non-AS-marked traffic per IEA and the resulting ET-marked packets possibly trigger the termination of ET-marked flows. However, if a partial path of the multipath is *SR*-pre-congested, traffic from other non-*SR*-pre-congested paths may dilute the fraction of AS-marked packets and hence the marking converter does not produce any ET-markings. As a consequence, undertermination occurs, i.e., *SR*-pre-congestion is possibly not detected or not fully removed. This cannot happen with classic MFT which is based on *SR*-overload. In addition, with MFT with marking conversion based on *AR*-overload, AS- and ET-marked packets can stem from *AR*- or *SR*-pre-congested paths so that flows from possibly non-*SR*-pre-congested paths are terminated which causes overterminaiton. In the following, we derive a mathematical model to quantify this under- and overtermination.

1) Analytical Model: We model the termination process assuming equal flow rates and hence denote the admitted traffic by the number of flows. The model state $s = (s_0, ..., s_{k-1})$ $(0 \le i < k)$ indicates the number of admitted flows on the k partial paths of an IEA. Admissible rates are assigned to links within a PCN domain, but in our analysis AR_i indicates the number of admissible flows on each partial path i. Furthermore, we assume a constant u which allows to calculate the implicit supportable rates by $SR_i = AR_i \cdot u$. Our model neglects the time-component, i.e., it assumes that the removal of a flow is immediately reflected in the received packet markings. The probability $p(s,i)$ for the removal of a flow from path i depends on the *AR*-overload on this path $\max(0, s_i - AR_i)$ and the *AR*-overload on the entire IEA:

$$p(s,i) = \frac{\max(0, s_i - AR_i)}{\sum_{0 \le j < k} \max(0, s_j - AR_j)}. \quad (3)$$

This yields the probability for the transition steps of a simple death process:

$$(s_0, ..., s_i, ..., s_{k-1}) \xrightarrow{p(s,i)} (s_0, ..., s_i - 1, ..., s_{k-1}). \quad (4)$$

The process starts with $s_i = n_i$ flows on partial path i. We compute the probability $p(s)$ of all states s with $0 \le s_i \le n_i$ by an iterative algorithm. We do not explain further details due to page limitations. The termination process stops if the overall received traffic rate is at most the rate of the unmarked traffic multiplied by u, i.e., if the condition

$$\sum_{0 \le i < k} s_i \le u \cdot \sum_{0 \le i < k} \min(s_i, AR_i) \quad (5)$$

is met because then the marking converter stops generating ET-markings. The set $\mathcal{T}$ contains all states s in which the iterative calculation terminates because the stop condition is met. The probability of the states in the terminating set $\mathcal{T}$ sums up to 1. Hence, we can calculate the average relative amount of overtermination and undertermination by

$$OT = \frac{\sum_{s\in\mathcal{T}}\sum_{0\le i<k}\max(0,\min(n_i,SR_i)-s_i)\cdot p(s)}{\sum_{0\le i<k}\min(n_i,SR_i)} \quad (6)$$

$$UT = \frac{\sum_{s\in\mathcal{T}}\sum_{0\le i<k}\max(0,s_i-SR_i)\cdot p(s)}{\sum_{0\le i<k}\min(n_i,SR_i)}. \quad (7)$$

2) Numerical Results: We perform some experiments that show how different but also how large the amount of over- and undertermination can be. We consider a single IEA with two parallel paths, each of them having an admissible rate of $AR_i = 20$ flows, and $u = 2$. Thus, each partial path can carry up to 40 flows without being *SR*-pre-congested. We set the number of flows on the first partial path to $n_0 \in \{20, 40, 60\}$. Fig. 5 shows the average relative over- and undertermination as well as their sum. The values are observed when the termination process stops and depend on the number of flows n_1 on the second partial path.

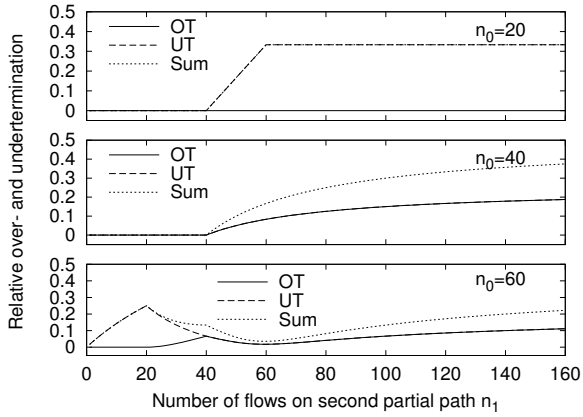

Fig. 5. Average relative overtermination, undertermination, and their sum for MFT with marking conversion based on marked *AR*-overload; $AR_0 = 20$, $AR_1 = 20$, $u = 2$.

For $n_0 = 20$ flows on the first partial path, flows are not terminated for $n_1 \le 60$ flows on the second partial path although the second partial path is already *SR*-pre-congested for more than $40 < n_1$ flows. Therefore, we observe up to 33% undertermination. We observe no overtermination because the non-congested path does not AS-mark any packets. Thus, none of its flows it terminated. For $n_0 = 40$, none of the partial paths is *SR*-pre-congested for $n_1 \le 40$ and flows are not terminated. From $n_1 > 40$ on, *SR*-pre-congestion is indicated for the IEA and flows are terminated. The amount of over- and undertermination is the same and increases with the number of flows n_1 on the second path. For $n_0 = 60$, the IEA indicates *SR*-pre-congestion for $n_1 < 20$ and $n_1 > 20$ and hence flows are terminated in these ranges from the *SR*-pre-congested path. However, they are not sufficiently many and so undertermination occurs. For $n_1 = 20$ the IEA does not indicate *SR*-pre-congestion although the first partial path is *SR*-pre-congested. For small values of $n_1 < 40$ there is more under- than overtermination. For $n_1 \ge 40$, the amount of over- and undertermination is the same. This is of course not an in-depth analysis, but the experiments show that over- and undertermination can be quite large and they are very sensitive to the load on the partial paths of a multipath. Hence, MFT with marking conversion based on marked *AR*-overload does not work with multipath routing.

V. Conclusion

In this paper we have proposed a new algorithm that converts PCN markings caused by excess marking based on the admissible rate into markings based on the supportable rate. We have shown that the new algorithm allows to use marked flow termination (MFT) when excess marking based on the admissible rate is the only marking scheme in the network. We investigated the application of MFT in combination with the new marking converter and proposed guidelines for its configuration. We showed that overtermination occurs for a small number of flows per ingress-egress aggregate and that both over- and undertermination occur in case of multipath routing. Hence, it cannot be used under these circumstances. Thus, it loses the main benefits of classic MFT that uses original packet markings based on the supportable rate. This is an important insight for the standardization process as two separate marking algorithms still seem to be required in order to cope with challenging conditions.

References

[1] P. Eardley (ed.), "Pre-Congestion Notification Architecture," http://tools.ietf.org/id/draft-ietf-pcn-architecture-05.txt, Aug. 2008.

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

[3] M. Menth, F. Lehrieder, B. Briscoe, P. Eardley, T. Moncaster, J. Babiarz, K.-H. Chan, A. Charny, G. Karagiannis, , and X. J. Zhang, "PCN-Based Admission Control and Flow Termination," in *to be published in IEEE Communications Surveys & Tutorials (COMST)*, 2009.

[4] 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.

[5] M. Menth and F. Lehrieder, "PCN-Based Marked Flow Termination," in *currently under submission*, 2008.

[6] X. Zhang and A. Charny, "Performance Evaluation of Pre-Congestion Notification," in *International Workshop on Quality of Service (IWQoS)*, Enschede, The Netherlands, Jun. 2008.

[7] M. Menth and F. Lehrieder, "Comparison of Marking Algorithms for PCN-Based Admission Control," in 14^{th} *GI/ITG Conference on Measuring, Modelling and Evaluation of Computer and Communication Systems (MMB)*, Dortmund, Germany, Mar. 2008.

[8] ——, "Performance Evaluation of PCN-Based Admission Control," in *International Workshop on Quality of Service (IWQoS)*, Enschede, The Netherlands, Jun. 2008.

[9] F. Lehrieder and M. Menth, "PCN-Based Flow Termination with Multiple Bottleneck Links," in *IEEE International Conference on Communications (ICC)*, Dresden, Germany, Jun. 2009.

[10] M. Menth, "Efficiency of PCN-Based Network Admission Control with Flow Termination," *Praxis der Informationsverarbeitung und Kommunikation (PIK)*, vol. 30, no. 2, pp. 82 – 87, Apr. 2007.

[11] M. Menth and M. Hartmann, "Threshold Configuration and Routing Optimization for PCN-Based Resilient Admission Control," *to appear in Computer Networks*, 2009.

[12] J. Jiang and R. Jain, "A Simple Analytical Model of Pre-Congestion Notification," in *currently under submission*, 2008.