PCN-Based Flow Termination with Multiple Bottleneck Links

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

***Abstract*—Pre-congestion notification (PCN) is a new packet marking scheme based on which simple measurement-based admission control (AC) and flow termination (FT) are implemented. FT is useful for traffic management in unexpected events, e.g., when admitted flows lead to overload on a link after rerouting which may be due to a link or node failure. While AC is a classic flow control function, FT is new and only little understood so far. The limited literature on FT focuses mainly on a single overloaded link. However, when a link or node fails, redirected traffic is likely to cause overload on multiple backup links (bottlenecks) at the same time. As the packet marking probability for flows traversing multiple bottlenecks is larger than for flows traversing only the most severe bottleneck, more traffic is possibly terminated than needed, i.e. overtermination occurs. This paper quantifies potential overtermination in case of multiple bottlenecks for different FT mechanisms which are currently discussed by the IETF.**

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 so 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].

Flow termination has been investigated only little in the past [4], [5] and only with a single overloaded link. Overload often appears due to redirected traffic [6] and when traffic is rerouted over a backup path consisting of several links, possibly multiple bottlenecks occur. It is not clear how FT behaves in such a case and we investigate that in this paper for various FT methods which are currently under discussion in the IETF. We show that overtermination appears, i.e. more traffic is terminated than necessary. This happens with all FT methods but to a different degree. We investigate this problem by packet-based simulation and model it mathematically to provide a better understanding of the phenomenon.

The paper is structured as follows. Sect. II explains PCN, metering and marking algorithms as well as various FT algorithms. Sect. III reviews related work. Sect. IV studies potential overtermination in multiple bottleneck scenarios. Finally, Sect. V summarizes this work and draws conclusions.

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 explain excess marking and review algorithms for FT which are the mechanisms relevant to this study.

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 that crosses a so-called PCN domain. This is similar to the IntServ-over-DiffServ concept [7]. 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 a special header codepoint to make

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

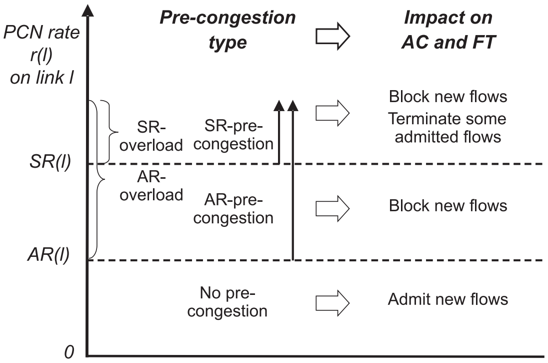

Fig. 1. The admissible and the supportable rate $(AR(l), SR(l))$ define three types of pre-congestion.

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.

C. Excess Marking

PCN nodes use metering and marking algorithms to control the current PCN traffic rates on their links and to mark packets if these rates exceed the admissible or supportable rates of theses links. We briefly review excess marking as it is used to support FT. It has a reference rate which may be set to the admissible or supportable rate depending on the FT algorithm. The excess meter controls the rate of unmarked PCN traffic and marks those unmarked PCN packets that exceed its reference rate. Details including pseudocode can be found in [3]. In case that all PCN traffic is unmarked when it enters the excess marker, the resulting 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 on the basis of a token bucket marker with only few modifications of existing hardware.

D. Flow Termination

FT mechanisms evaluate packet markings at the boundary of the PCN domain, detect potential *SR*-pre-congestion, and determine how much traffic should be terminated in that case. We describe *measured rate termination (MRT) based on SR- and AR-overload* as well as *marked flow termination (MFT)*. They use marking feedback from so-called ingress-egress aggrates (IEAs) which is the ensemble of flows between a specific pair of ingress and egress nodes of a PCN domain.

1) Measured Rate Termination based on SR-overload (MRT-SR): To support MRT-SR, PCN nodes perform excess marking based on the supportable rate *SR* on each link. PCN egress nodes measure the rate of unmarked traffic U per IEA using measurement intervals of duration D_{MI}. This is used as the so-called edge-to-edge supportable rate *eSR* and termination is triggered for the received PCN traffic above that rate. Consecutive termination steps require a minimum inter-termination intervals. When overload starts only in the middle of a measurement interval, the *SR*-overload is underestimated in the first termination step and two termination steps are needed.

2) Measured Rate Termination based on AR-overload (MRT-AR): With MRT-AR, PCN nodes perform excess marking based on the admissible rate *AR* on each link. These markings can be used to support AC and MRT-AR. The advantage of this approach is obvious: PCN nodes need to run only a single metering and marking scheme and only a single codepoint for this marking is needed. Potential drawbacks of this solution are documented in [5]. MRT-AR is similar to MRT-SR, but it requires a domain-wide parameter u to control the ratio between the admissible and supportable rate on each link: $SR = u \cdot AR$. Like with MRT-SR, PCN egress nodes measure the rate of unmarked traffic U per IEA. This rate is multiplied with u to calculate the edge-to-edge supportable rate $eSR = u \cdot U$ and termination is triggered for the received PCN traffic above that rate. A minimum inter-termination time also applies to MRT-AR.

The presented MRT algorithms are the simple direct MRT methods (DMRT). More complex indirect MRT (IMRT) methods have advantages in case of traffic loss which is not considered in this paper.

3) Marked Flow Termination: MFT terminates flows only if at least one of their packets was marked. The advantage of MFT is that it works well with any number of flows per IEA and with multipath routing. MRT-AR and MRT-SR fail under these conditions. Various MFT methods have been proposed in [4], they all exhibit the same termination behavior, but in this work we focus only on MFT for IEAs. MFT assumes that PCN nodes perform excess marking based on the supportable rate just like for MRT-SR. Each egress node maintains a credit counter c_g for each IEA g. When a marked packet of 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. We choose $\alpha = 1$. Larger values of α lead to faster termination but also to overtermination, smaller values lead to slower termination [4]. The credit counter c_g is randomly initialized 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.

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 [8], 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 [9].

A two-layer architecture for PCN-based AC and FT was presented in [10] and flow blocking probabilities have been studied for single aggregates and static load conditions. The work presented in [4] 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. One of these MFT mechanisms is used in [11] and adapted to work with PCN marking based on *AR*-overload. Measured rate termination (MRT) based on *AR*- and *SR*-overload is investigated in [5]. All aforementioned studies regarding PCN-based flow termination consider for their performance evaluation only a single bottleneck link, i.e., the supportable rate is exceeded on only one link in the PCN domain. In this paper we extend the investigation to scenarios with multiple bottleneck links.

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 [12]. An additional investigation about how *AR* and *SR* thresholds should be set in PCN domains with resilience requirements is contained in [13]. Furthermore, it studies how link weights should be set in IP networks in order to maximize the admissible traffic rates. The authors of [14] investigate the impact of admissible and supportable rate thresholds on the admission and termination of on/off traffic.

IV. Flow Termination in Multiple Bottleneck Scenarios

In this section we investigate the termination behavior of various FT methods in multiple bottleneck scenarios. First, we present the experiment setup including a simulation environment and a mathematical model to quantify potential overtermination. Then, we investigate potential overtermination of MRT-SR, MRT-AR, and MFT under various conditions.

A. Experiment Setup

We consider m serial links which are numbered $1 \leq i \leq m$ and have an admissible and supportable rate of AR_i and SR_i. We assume that all links have sufficient capacity so that no packet loss occurs even in case of *SR*-overload. In addition, there are $m+1$ traffic aggregates which are numbered by $0 \leq i \leq m$ and have an initial traffic rate of R_i. Fig. 2 illustrates that aggregate 0 passes all links while aggregate $i > 0$ contributes only *cross traffic (CT)* for link i. Such scenarios can occur in PCN domains after rerouting due to link or node failures. Before the failure, aggregate 0 is carried over a different path and the considered links i carry only traffic of aggregate i. After the reroute event, aggregate 0 is redirected to links $1 \leq i \leq m$ which carry then additional traffic. This may cause *SR*-pre-congestion on some of them and trigger flow termination. In the following we call aggregate 0 backup traffic and aggregates $i > 0$ cross traffic.

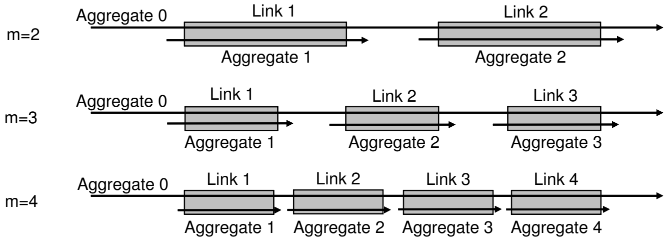

Fig. 2. Experiment setup for $m \in \{2,3,4\}$ bottleneck links: any link is *SR*-pre-congested; Aggregate 0 is backup traffic and aggregates $i > 0$ contributes cross traffic.

In this work we focus on multiple bottleneck scenarios. We assume that all links on the backup path are *SR*-pre-congested after the reroute. Possibly additional non-pre-congested links are disregarded. To simplify our study, we use a symmetric experiment setup, i.e. equal rate thresholds $AR_i = AR$ and $SR_i = SR$ and equal cross traffic $R_i = R_{CT}$ on all links $1 \leq i < m$, so that all links experience the same *SR*-overload immediately after the reroute of aggregate 0. If not mentioned differently, the reroute instant of aggregate 0 coincides with the start of its measurement interval at the egress node. Furthermore, the measurement intervals of all aggregates are synchronized. When PCN detects *SR*-overload, flow termination is triggered. We study the traffic rates of all aggregates on all links and are especially interested in the remaining overall traffic rate on link 1 after the termination process has completed. In particular, we determine the overtermination which is the relative difference between its remaining overall traffic rate and its supportable rate. Although the simulation model is simplified, the results can be transfered to more complex network topologies when traffic demands and paths are given.

1) Simulation Environment: We use a custom-made packet-based simulator written in Java. We simulate homogeneous connections with realtime characteristics having an inter-arrival time of 20 ms and a constant packet size of 200 bytes yielding flows with 80 kbit/s. To avoid simulation artifacts, we add a uniformly distributed jitter of up to 1 ms to the theoretical arrival instants of the packets and average results from multiple simulation runs. In each simulation, a different arrival pattern of the flows is generated which is important for the simulation of periodic traffic. We run so many simulations that confidence intervals are small and omit them in our figures for the sake of clarity. We simulate the above scenario where rerouting causes a multiple bottleneck scenario with consecutive flow termination. To that end, we assume a supportable rate of $SR = 80$ Mbit/s per link and set $AR = \frac{SR}{u}$ accordingly. We implement MRT-SR, MRT-AR, and MFT according to the description in Sect. II-D. For MRT, we use measurement intervals of length $D_{MI} = 200$ ms. For MFT, we set the termination aggressiveness $\alpha = 1$. Furthermore, we set the termination delay $D_T = 50$ ms and use this value also for the configuration of MFT (cf. Sect. II-D). According to [4], overtermination does not occur with these values.

We simulate the time dependent termination behavior. Fig. 3

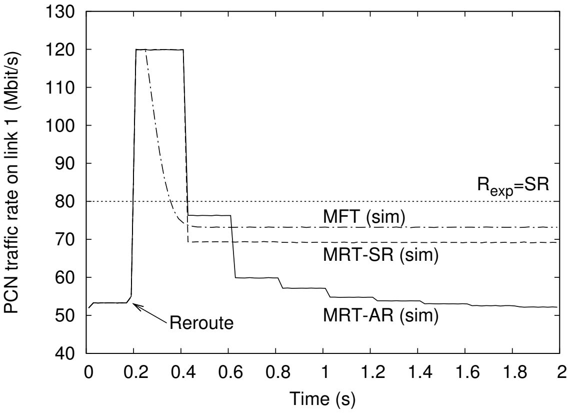

Fig. 3. Termination behavior of MRT-AR, MRT-SR, and MFT in an example setup with $m = 3$ links and $u = 1.5$.

illustrates the PCN traffic rate on link 1 for $m = 3$ links and the three different termination methods. The initial load is produced only by the cross traffic of aggregate 1 with rate $R_{ct} = 53.36$ Mbit/s. After 0.2 s, a reroute occurs and link 1 is suddenly faced with additional traffic of the rerouted aggregate 0 with rate $R_0 = 66.64$ MBit/s. With MRT-SR, only 69 Mbit/s remain on link 1. This is significantly lower than the expected rate $R_{exp} = min(SR, R_0 + R_{ct})$ which the link can carry. Thus, MRT-SR yields roughly 11 Mbit/s overtermination on link 1. This phenomenon is due to multiple bottlenecks and cannot be observed on a single *SR*-pre-congested link. The reason is that packets of aggregate 0 are marked on link 1 and consecutive links. As a result, the traffic rate to be terminated is overestimated. In this example MRT-SR needs only one termination step because the reroute coincides with the start of the measurement intervals and the whole overload can be captured in the first interval. If the reroute occurs during a measurement interval (not shown), MRT-SR requires two termination steps to remove *SR*-overload. This is different with MRT-AR. Traffic is terminated in several steps (cf. Fig. 3) whose duration corresponds to the minimum inter-termination time. Overtermination is about 30 Mbit/s which is larger than the one for MRT-SR. The reason is that MRT-AR marks a larger traffic fraction than MRT-SR since traffic is marked both in case of *SR*- and *AR*-pre-congestion. Even if *SR*-overload is removed, further packets are marked and termination possibly continues for aggregate 0 since its traffic is marked by all links and may trigger termination of further flows. MFT removes the load gradually and about 73 Mbit/s of the overall traffic remain on link 1 after termination has completed.

Overtermination possibly also occurs on the $m-1$ subsequent links of link 1. However, it is less serious than on link 1 because the meter does not count traffic of aggregate 0 which arrives already marked. Therefore, more cross traffic remains on the links $i > 1$. We validated this by simulations and the following mathematical analysis, but omit the figures and do not consider this further due to space limitations. Instead, we focus on the investigation of the overtermination on link 1.

2) Mathematical Analysis for MRT: We derive a mathematical model describing a single termination step by MRT-AR in multiple bottleneck scenarios. With MRT-AR, the egress node measures the rate of unmarked traffic U_i and multiplies it by u to obtain the edge-to-edge supportable rate $eSR_i = u \cdot U_i$. Thus, $max(0, R_i - eSR_i)$ traffic is terminated. R_i and SR are known and we derive now the value of the unmarked traffic U_i for all aggregates i. The unmarked traffic rate of aggregate i is denoted by U_i^{j-1} before it is marked on link j and by U_i^j afterwards. When traffic enters its first bottleneck, it is not yet marked and hence we have $U_i^{i-1} = R_i$ for cross traffic aggregates $1 \leq i \leq m$. For the rerouted aggregate holds $U_0^0 = R_0$ when it enters link 1. We denote the probability that an unmarked packet is not marked by the meter and marker of link j by p_u^j. Hence, the rate of unmarked traffic U_i^j at the end of link j can be computed by

$$U_i^j = U_i^{j-1} \cdot p_u^j. \tag{1}$$

The probability p_u^j that a packet remains unmarked on link j depends on the rate of unmarked traffic $U_0^{j-1} + U_j^{j-1} = U_0^{j-1} + R_j$ and the configured rate for the excess marker $AR = \frac{SR}{u}$ and can be calculated by

$$p_u^j = \begin{cases} 1 & \text{if } U_0^{j-1} + R_j \leq \frac{SR}{u}, \\ \frac{SR/u}{U_0^{j-1} + R_j} & \text{otherwise.} \end{cases} \tag{2}$$

When the rate of unmarked traffic at link j is smaller than the reference rate $U_0^{j-1} + R_j \leq \frac{SR}{u}$, no packets are marked ($p_u^j = 1$). If the reference rate is exceeded, the unmarked excess traffic is marked. Traffic is terminated only if $R_i > eSR_i$. Therefore, the remaining traffic rate of aggregate i is $\hat{R}_i = min(R_i, eSR_i)$ after termination. It can be computed by $eSR_i = U_i \cdot u$ with $U_0 = U_0^m$ for the redirected aggregate and $U_i = U_i^i$ for cross traffic aggregates $1 \leq i \leq m$. The presented analysis determines the remaining traffic rate $\hat{R}_i$ for all aggregates i after one termination step. However, MRT-AR terminates traffic in several steps (cf. Fig. 3). Thus, the analysis is applied iteratively by using the remaining traffic rates $\hat{R}_i$ as input rates R_i for the next iteration step. In this study, we use 25 iteration steps and the remaining rates seem to have converged after this number which corresponds to 10 s termination in real life.

We derive the overtermination on link 1. After the termination process has completed, the remaining overall traffic rate on link 1 is $\hat{R}_0 + \hat{R}_1$, but its expected traffic rate after termination is $R_{exp} = \min(SR, R_0 + R_{CT})$. Thus, the overtermination is

$$OT = \frac{R_{exp} - (\hat{R}_0 + \hat{R}_1)}{SR}. \tag{3}$$

B. Numerical Results

We illustrate the overtermination observed on link 1 in Fig. 2 by various experiments. We first elaborate on realistic experimental parameters. Then we investigate the impact of the u parameter of MRT-AR, the relative *SR*-overload, and the relative cross traffic. Finally, we drop the assumption of synchronized reroute times and measurement intervals and show that our findings are still valid.

1) Parameters for Realistic Experiments: We define the *relative load* on the links immediately after the reroute and before termination by $\rho_{SR} = \frac{R_{CT}+R_0}{SR}$ and the *relative cross traffic* by $\gamma = \frac{R_{CT}}{R_0}$. The values are the same for all links due to the symmetry assumption. These definitions are useful for systematic parameter studies, but a realistic choice of ρ_{SR} and γ must respect some constraints. Combining both definitions we follow that the cross traffic rate is $R_{CT} = \frac{\rho_{SR}}{1+\frac{1}{\gamma}} \cdot SR$. It can be at most as large as the admissible rate on the considered links. For MRT-SR, AR can be set to any value smaller than SR but for MRT-AR, AR must be set to $\frac{SR}{u}$. Hence, for MRT-SR ρ_{SR} and γ must fulfill $\frac{\rho_{SR}}{1+\frac{1}{\gamma}} \leq 1$ and for MRT-AR they must fulfill $\frac{\rho_{SR}}{1+\frac{1}{\gamma}} \leq \frac{1}{u}$. In our study we use $\rho_{SR} = 1.5$, $u = 1.5$, and $\gamma = 0.8$ as default values which give enough freedom to keep two parameters constant and vary the third one within realistic bounds. Given ρ_{SR}, γ, u, and $SR = 80$ Mbit/s, all other parameters relevant to the analysis or simulation can be derived.

2) Impact of the Number of Bottleneck Links and the u-Parameter of MRT-AR: Fig. 4 shows simulated and analytical overtermination for MRT-AR depending on the u-parameter and the number of bottlenecks m. Overtermination increases with increasing u-parameter and with an increasing number of bottlenecks m.

Overtermination occurs because traffic of aggregate 0 is subject to marking on every link on the backup path. Therefore, the unmarked traffic rate of aggregate 0 is larger after the first link (U_0^1) than after the last link ($U_0^m = U_0$) based on which termination is triggered. Thus, more traffic than necessary is terminated from aggregate 0. As a result, overtermination occurs on link 1 and possibly also on some other links. For increasing values of u, the marking probability increases on every link. Consequently, the fraction of marked packets of aggregate 0 increases more than linearly with u because traffic of aggregate 0 can be marked on all of the m links. Hence, with increasing u, the unmarked traffic U_i decreases and overtermination increases. For $u = 1$ we get MRT-SR which exhibits considerably less overtermination than MRT-AR with large u parameters. In the extreme case of $u = 1.5$, we have 44% overtermination, i.e. only 56% of the supportable rate on link 1 are used after termination because 63% of the traffic that was on the link immediately after the reroute was terminated.

The fact that overtermination increases with the number of bottleneck links m is rather obvious as additional marking stages reduce the rate of unmarked traffic U_0 of aggregate 0. Consequently, more traffic is terminated. The figure compares our analytical results with those from simulation. They are rather accurate and the difference can be explained by the fact that simulation deals with integral quantities such as discrete flows and implements timing constraints while our analysis is based on real values and no timing constraints. For instance, the simulation respects the time until markers react after aggregate 0 is rerouted. Thus, our mathematical model correctly describes the overtermination process, contributes to the understanding of overtermination due to multiple bottlenecks, and may be used to quickly predict potential overtermination under various conditions.

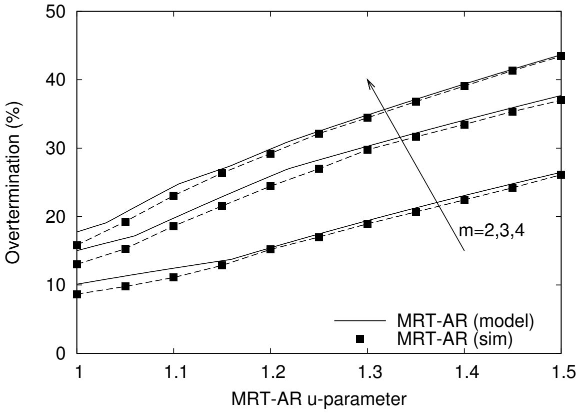

Fig. 4. Overtermination on link 1 depending on the number of bottleneck links m and the u-parameter of MRT-AR ($\gamma = 0.8$, $\rho_{SR} = 1.5$).

3) Impact of the Relative Load ρ_{SR}: We compare the overtermination caused by MRT-SR, MRT-AR, and MFT. We present only simulation results because we have no analytical description for the overtermination caused by MFT. Fig. 5 shows the impact of the relative load ρ_{SR} on the overtermination for $m = 3$ bottlenecks.

MRT-AR causes more overtermination than MRT-SR and MFT and its curve already ends at $\rho_{SR} = 1.5$ since larger values are not feasible for the given parameters (cf. Sect. IV-B1). MRT-AR yields overtermination of up to 18% even without SR-pre-congestion ($\rho_{SR} \leq 1$) on any link. The reason is that traffic marking starts at a relative load of $\rho_{SR} = \frac{1}{u} = 67\%$. Traffic marking on consecutive links leads possibly to such a high fraction of marked traffic at the egress node that it is interpreted as SR-overload by the MRT-AR algorithm and flow termination is triggered. This phenomenon starts at a relative load of $\rho_{SR} = 80\%$. For MRT-SR and MFT, overtermination occurs only for $\rho_{SR} > 1$. MFT yields slightly less overtermination than MRT-SR because it terminates traffic gradually an not in one or two shots like MRT-SR.

4) Impact of the Relative Cross Traffic γ: In theory, overtermination is small for very small and large relative cross

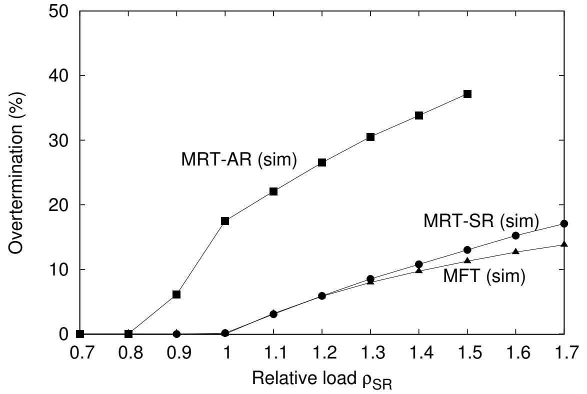

Fig. 5. Overtermination on link 1 depending on the relative load ρ_{SR} ($m = 3$, $\gamma = 0.8$, $u = 1.5$).

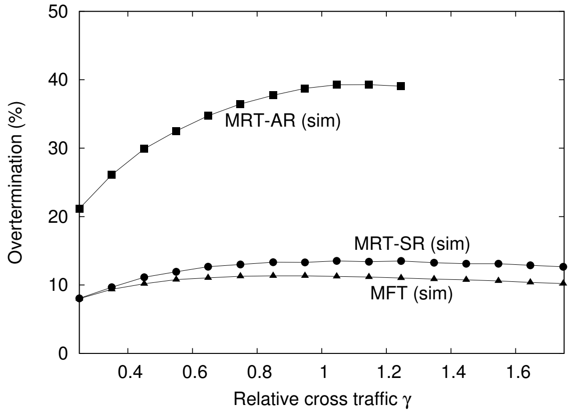

Fig. 6. Overtermination on link 1 depending on the relative cross traffic γ ($m = 3$, $\rho_{SR} = 1.5$, $u = 1.5$.)

traffic $\gamma = \frac{R_{CT}}{R_0}$ and has a maximum in between. Fig. 6 shows the overtermination depending on the relative cross traffic for $m = 3$ bottlenecks. For MRT-SR and MFT, the overtermination is almost constant around 10% for the most relevant parameter range of γ. For MRT-AR, it increases significantly from 20% to 40% between $\gamma = 0.2$ and $\gamma = 1.25$. Larger values than $\gamma = 1.25$ are not feasible for the given parameters (cf. Sect. IV-B1).

5) Impact of Synchronization: Previous work has shown that the start of the *SR*-overload relative to the start of PCN's measurement interval influences the termination behavior of MRT [5]. Therefore, we investigate its impact also in multiple bottleneck scenarios. We vary the parameter $0 \leq T_R < D_{MI}$ which describes the time between the start of a measurement interval of aggregate 0 and its reroute.

Fig. 7 shows the overtermination for MRT-AR with $u = 1.5$ an for MRT-SR. In addition, we consider the case that the measurement intervals of all aggregates are synchronized or that the measurement intervals for the cross traffic are maximally shifted by $\frac{D_{MI}}{2} = 100$ ms relative to the measurement interval for the backup traffic. The results show that synchronized measurement intervals and reroute times have some impact on the overtermination, but they do not change it fundamentally. It is still in the same order of magnitude. Thus, the results from our mathematical model and our simulation results are still approximatively valid although they are based on the simplifying assumption of synchronized measurement intervals and reroute time.

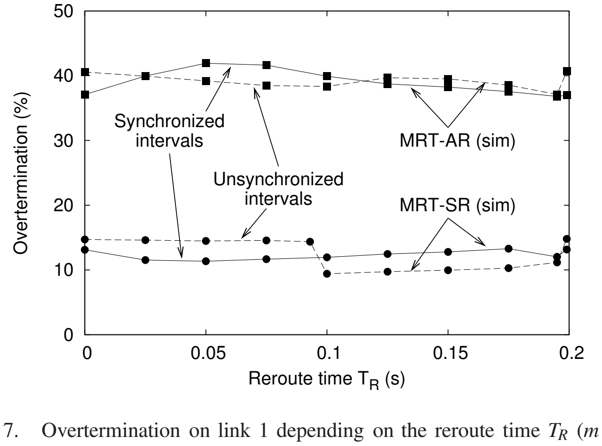

Fig. 7. Overtermination on link 1 depending on the reroute time T_R ($m = 3$, $\rho_{SR} = 1.5$, $\gamma = 0.8$, $u = 1.5$).

V. Conclusion

In this paper we compared the termination behavior of measured rate termination based on *SR*-overload (MRT-SR), measured rate termination based on *AR*-overload (MRT-AR), and marked flow termination (MFT) in multiple bottleneck scenarios. Multiple bottlenecks can occur when a link or node fails and traffic is rerouted which possibly causes *SR*-pre-congestion on several links of the backup paths. Our results show that overtermination can occur on some links, i.e., too much traffic is terminated so that some of their supportable PCN rate is not used after the termination has completed. We quantified the amount of overtermination and argue that this issue should be taken into account in the standardization process in the IETF. The observed overtermination increases with the *SR*-overload and depends on the rates of the primary and backup traffic on the bottlenecks. While MRT-SR led to overtermination of up to 10% in our experiments, MRT-AR caused overtermination of up to 40%. Overtermination in case of MRT-AR depends on the *u*-parameter and can also occur when no link is *SR*-pre-congested. We obtained our results by packet-based simulation and by mathematical analysis. The mathematical model improves the understanding of the observed phenomenon and provides a means to quickly predict potential overtermination for specific multiple bottleneck scenarios.

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 et al., "PCN-Based Admission Control and Flow Termination," in *to be published in IEEE Communications Surveys & Tutorials (COMST)*, 2009.

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

[5] ——, "PCN-Based Measured Rate Termination," *currently under submission*, 2008.

[6] 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.

[7] Y. Bernet et. al., "RFC2998: A Framework for Integrated Services Operation over Diffserv Networks," Nov. 2000.

[8] X. Zhang and A. Charny, "Performance Evaluation of Pre-Congestion Notification," in *IWQoS*, Enschede, The Netherlands, Jun. 2008.

[9] 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.

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

[11] F. Lehrieder and M. Menth, "Marking Conversion for Pre-Congestion Notification," in *IEEE International Conference on Communications (ICC)*, Dresden, Germany, Jun. 2009.

[12] 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.

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

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