Threshold Configuration and Routing Optimization for PCN-Based Resilient Admission Control

Michael Menth and Matthias Hartmann

University of Wuerzburg, Institute of Computer Science, Germany

Abstract

Pre-congestion notification (PCN) in IP networks uses packet metering and marking within a PCN domain to notify its egress nodes whether link-specific admissible or supportable rate thresholds have been exceeded by high priority traffic. Based on this information simple admission control and flow termination is implemented. The latter is a new flow control function and useful in case of overload through high priority traffic which can occur in spite of admission control, e.g., when traffic is rerouted in failure cases. Resilient admission control admits only so much traffic that admitted traffic can be rerouted without causing congestion on backup paths in case of a likely failures, e.g., single link failures.

We propose algorithms to configure the link-specific PCN rate thresholds such that resources are utilized efficiently and fairly by competing traffic aggregates while meeting resilience constraints. This is done for the single and dual marking PCN architecture whereby the single marking case is more demanding since it requires that the supportable rate is a fixed multiple of the admissible rate on all links within a single PCN domain. Furthermore, we derive objective functions to optimize the underlying routing system for both cases. Our performance results for various network types show that the dual marking PCN architecture leads to significantly better resource efficiency than the single marking PCN architecture.

Key words: Routing optimization, resilience, admission control, QoS

1 Introduction

Internet service providers (ISPs) recently offer increased access speeds, e.g., by digital subscriber lines (DSL), cable TV (CATV), and fiber to the home (FTTH).

This work is funded by Nortel Networks, Ottawa, and Deutsche Forschungsgemeinschaft (DFG) under grant TR257/18-2. The authors alone are responsible for the content of the paper.

These technologies significantly increased the traffic volume in carrier networks and in 2005, the major traffic in Japan was already produced by residential users [1]. Popular video services like YouTube produce large traffic volumes, but are only weak precursors of high-quality IP-TV services. They present a challenge for ISPs which need to offer triple play, i.e. the integration of the transport of data, voice, and video. However, the resource management for triple play becomes more and more difficult due to the emerging interactive Web 2.0 since residential users also become content providers. In particular, [2] has shown that some normal users get accustomed with new services, change access technologies, and become "heavy hitters" and hence the majority of the overall traffic is produced by a minority of residential users.

Today, ISPs rely on capacity overprovisioning (CO) to support quality of service (QoS) in terms of packet loss and delay. In [3] admission control (AC) was proposed for IP networks, but so far such techniques are applied only locally, they are rarely in use, and not deployed in core networks. However, there is a firm belief that next generation networks require some form of QoS assurance such as AC to enable services that cannot be provided with CO [4]. Conventional AC prevents overload due to increased user activity. If congestion occurs in core networks, it is mainly caused by failures and redirected traffic, and only to a minor degree by increased user activity [5]. Thus resilient AC is required that admitted traffic can be rerouted in likely failure scenarios without causing congestion on backup paths [6]. In other words, both AC and CO require backup capacity to prevent QoS violations due to backup traffic in case of failures. In case of CO, this backup capacity can be used to accommodate both moderate fluctuations of the traffic matrix and backup traffic. As a consequence, there are no significant bandwidth savings when AC is used instead of CO for QoS provisioning [7]. However, the dynamic behavior of users and services sketched above leads to an unpredictability of future demands such that QoS provisioning remains difficult. Therefore, ISPs see the need for AC to offer premium services over integrated IP networks in the future.

The Internet Engineering Task Force (IETF) currently works on "Congestion and Pre-Congestion Notification" (PCN) [8] with the objective to standardize feedback-based admission control (AC) and flow termination (FT) for high-priority PCN traffic for single DiffServ domains [9]. Each link l of a so-called PCN domain is associated with an admissible and a supportable rate threshold ($AR(l)$, $SR(l)$) and the egress nodes of the domain are notified via appropriately marked packets if these thresholds are exceeded by high-priority PCN traffic. This feedback is used to implement AC and FT. Various packet marking schemes as well as AC and FT methods are proposed [10]. Some proposals provide two metering and marking schemes [11, 12], to control the admissible and the supportable rate independently of each other (*dual marking PCN architecture, DM-PCN*). Others provide only a single metering and marking scheme [13, 14] that controls only the admissible rate (*single marking PCN architecture, SM-PCN*). They implicitly assume that the supportable rate $SR(l)$ of a link l is a fixed multiple b of its admissible rate $AR(l)$

in the entire PCN domain

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

As a consequence, egress nodes can infer from the ratio of marked and unmarked traffic whether only the admissible or also the implicit supportable rate is exceeded on some link. We call the parameter *b* the “backup factor” as it controls the relation of primary and backup capacity on the links. The advantage of SM-PCN is that it needs fewer codepoints in the IP header for packet marking and less metering and marking support by routers. Its disadvantage is that Constraint (1) limits traffic engineering capabilities and makes the configuration of the rate thresholds harder when resource efficiency is an objective. In addition, it does not work well with multipath routing and when single edge-to-edge aggregates carry only little traffic [10].

This work investigates the rate threshold setting problem for PCN-based AC and FT. Furthermore, it proposes objective functions for routing optimization in resilient PCN networks. Performance results compare the resource efficiency of DM-PCN and SM-PCN with and without routing optimization for a large set of sample networks. The algorithms presented in this study also serve to configure and optimize PCN networks in practice.

The paper is structured as follows. Section 2 reviews related work showing the historic roots of PCN and similar AC approaches. Section 3 introduces PCN and explains how AC and FT work in the single and dual marking PCN architecture (SM-PCN, DM-PCN). Section 4 proposes algorithms to set the admissible and supportable rate thresholds appropriately for resilient AC. Section 5 provides objective functions to optimize IP routing in order to maximize the admissible protected traffic. Section 6 compares the resource efficiency of SM-PCN and DM-PCN for a large set of networks with different characteristics. Finally, Section 7 summarizes this work and draws conclusions.

2 Related Work

We review related work regarding random early detection (RED), explicit congestion notification (ECN), and stateless core concepts for AC as they can be viewed as historic roots of PCN.

2.1 Random Early Detection (RED)

RED was originally presented in [15], and in [16] it was recommended for deployment in the Internet. It was intended to detect incipient link congestion and to throttle only some TCP flows early in order to avoid severe congestion and to improve the TCP throughput. RED measures the average buffer occupation *avg* in routers and packets are dropped or marked with a probability that increases linearly with the average queue length *avg*.

2.2 Explicit Congestion Notification

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 not-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, ‘10’, ‘01’) 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. This improves the TCP throughput since packet retransmission is no longer needed. Both the ECN encoding in the packet header and the behavior of ECN-capable senders and receivers after the reception of a marked packet is defined in [17]. ECN comes with two different codepoints for ECT: ECT(0) (‘10’) and ECT(1) (‘01’). They help 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 differentiated services 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. They are likely to be reused for encoding of PCN marks.

2.3 Admission Control

We briefly review some specific AC methods that can be seen as forerunners of the PCN principle. They measure the rate of admitted traffic on each link of a network and give feedback to the network boundary if that rate exceeds a pre-configured admissible rate threshold. Thereby, no per-flow reservations need to be kept for a link and the network core remains stateless. This is a key property of PCN-based AC.

2.3.1 Admission Control Based on Reservation Tickets

To keep a reservation for a flow across a network alive, ingress routers send reservation tickets in regular intervals to the egress routers. Intermediate routers estimate the rate of the tickets and can thereby estimate the expected load. If a new reservation sends probe tickets, intermediate routers forward them to the egress router if they have still enough capacity to support the new flow and the egress router bounces them back to the ingress router indicating a successful reservation; otherwise, the intermediate routers discard the probe tickets and the reservation request is denied. Periodic reservation tickets do not need to be sent explicitly, their information can also be conveyed in form of some markings in normal data packets. Several stateless core mechanisms work according to this idea [21–23].

2.3.2 Admission Control Based on Packet Marking

Gibbens and Kelly [24, 25] 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 also allows to limit the utilization of the link bandwidth by premium traffic to arbitrary values between 0 and 100%. Karsten and Schmitt [26, 27] integrated these ideas into the IntServ framework and implemented a prototype. They point out that the marking can also be based on the CPU usage of the routers instead of the link utilization if this turns out to be the limiting resource for packet forwarding. An early version of a PCN-like AC has been reported in [28].

2.3.3 Resilient Admission Control

Resilient admission control admits only as much traffic as still can be carried after rerouting in a protected failure scenario [7, 29]. This is necessary since overload in wide area networks mostly occurs due to link failures and not due to increased user activity [5]. It can be implemented with PCN by setting the admissible rate thresholds low enough such that the rate of PCN traffic on a link is lower than its supportable rate threshold after rerouting.

3 PCN-Based Flow Control

This section illustrates the basic idea of PCN-based admission control (AC) and flow termination (FT) using the nomenclature of [10]. An example illustrates how PCN-based AC and FT fit into the overall Internet structure. We review how AC can be implemented based on appropriate metering and marking schemes. FT methods may reuse the marking scheme for AC or require their own. This leads to the definition of a *single and dual marking PCN architecture (SM-PCN, DM-PCN)*. We show how PCN-based AC and FT can be used to implement conventional and resilient AC. Finally, we explain the threshold setting and routing optimization problem for resilient PCN-based AC and FT which is the focus of this work.

PCN is intended for use in DiffServ networks and 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 link states as illustrated in Figure 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 traffic 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 AR- and SR-pre-congested and the traffic rate above $SR(l)$ is SR-overload. In this state, some already admitted flows should be terminated.

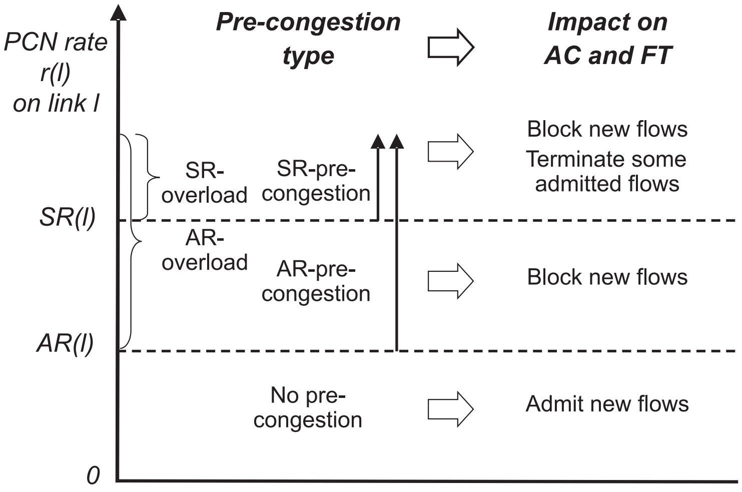


Fig. 1. The admissible and the supportable rate ($AR(l), SR(l)$) define three types of pre–congestion concerning the PCN traffic rate $r(l)$ on a link.

PCN traffic enters a PCN domain with a “no-pre-congestion” (NP) codepoint. PCN nodes monitor the PCN traffic rate on their links and re-mark the codepoints of the packets depending on the pre-congestion states of these links. The PCN egress nodes evaluate the packet markings and their essence is reported to the AC and FT entities of the network such that they can admit or block new flows or even terminate already admitted flows. Therefore, this concept is called pre-congestion notification.

3.2 Application of PCN in the Internet

There are different mechanisms for QoS support in the future Internet. Some domains use extensive capacity overprovisioning for all traffic. Others enable RSVP [30] in all nodes granting prioritized forwarding to flows with individual reservations according to the IntServ principle [31]. DiffServ relies on traffic prioritization for high priority traffic that is identified by an appropriate DiffServ codepoint and hence per-flow reservations are not required at all. To protect a network against overload, AC is required and flows must be individually treated at least at network boundaries. The IntServ-over-DiffServ concept [32] provides a controlled load (CL) service over DiffServ networks using per-flow AC at the ingress nodes of a domain. The CL service offers the same QoS a flow would receive from lightly loaded network elements [33] and is useful for inelastic flows, e.g., realtime media. PCN can be used to implement AC and FT in those networks. A prerequisite is that admission requests for high-priority traffic are triggered by end-to-end signalling protocols such as SIP, RSVP, or similar mechanisms for each flow. Depending on the network-specific QoS support, this signalling is respected or ignored. This is depicted in Figure 2. The PCN ingress node of a PCN region may serve as AC entity and admits or blocks admission requests.

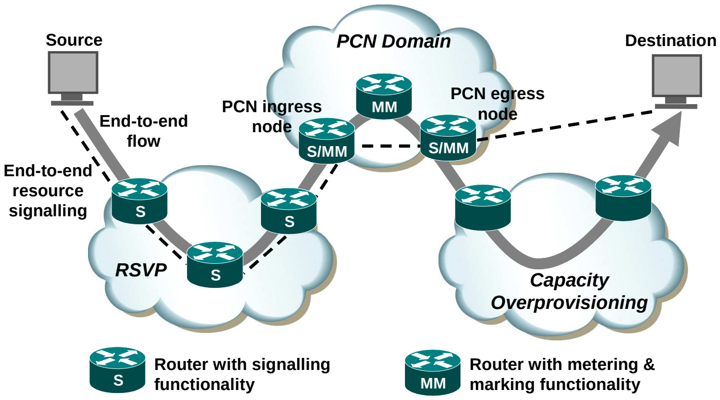

Source
PCN Domain
Destination
End-to-end flow
PCN ingress node
MM
PCN egress node
S/MM
S/MM
End-to-end resource signalling
S
S
S
RSVP
Capacity Overprovisioning
S
Router with signalling functionality
MM
Router with metering & marking functionality

Fig. 2. PCN-based AC guarantees a controlled load (CL) service over a DiffServ region and per flow admission requests to the PCN domain are triggered by external signalling protocols.

3.3 PCN-Based Admission Control

AC methods require that routers mark PCN traffic on links inside a PCN domain when they are *AR*-pre-congested. *Exhaustive marking* marks all PCN packets in that case with “admission-stop” (AS) while *excess marking* marks only those PCN packets that exceed the *AR* of the respective link. The currently preferred AC and FT methods work on aggregated feedback from ingress-egress aggregates (IEAs) [9] and an admission state indicating *admit* or *block* is kept per IEA. If the IEA is in the *admit* state, new flows fitting into this IEA are admitted, otherwise they are blocked. PCN egress nodes classify PCN packets according to their PCN ingress nodes and evaluate their markings per IEA. At the end of a measurement interval, the egress nodes compute the congestion level estimate (CLE), i.e. the fraction of AS-marked packets. If the CLE exceeds an upper threshold T_{CLE}^{AStop}, the admission state is turned to *block* and if it falls below a lower threshold T_{CLE}^{ACont}, the admission state is turned to *admit*. To that end, admission-stop or admission-continue messages are sent to the AC entity of the network. Some proposals use this CLE-based AC (CLEBAC) in combination with exhaustive marking [11, 12] and some others with excess marking [13, 14]. There are other methods for PCN-based AC [34], but they are not needed for this study.

3.4 PCN-Based Flow Termination

FT is a new flow control function protecting the network against congestion caused by already admitted traffic. At first sight, FT does not seem to be necessary when the admission of new PCN flows is controlled. However, admitted traffic can lead to severe overload such that it is beneficial for the network to terminate some flows when the PCN traffic rate exceeds the *SR* of a link. *SR*-overload occurs due to various reasons. (1) In failure cases admitted traffic can be rerouted and cause congestion on the backup path. (2) Already admitted flows may change their typical behavior and switch from low bit rates to high bit rates. (3) Flows are possibly admitted before the effect of previously admitted flows is reflected by the markings and so overload can occur. This is likely in case of flash crowds when lots of flows request admission within short time. For all these reasons it makes sense to deploy FT in a network that already uses AC for the admission of new flows. FT mechanisms may reuse the marking required for AC or they may require their own marking scheme. This leads to *single and dual marking PCN architectures (SM-PCN, DM-PCN)*. Various FT algorithms exist and a survey is given in [10]. In the following we present only two simple examples showing how FT works with SM- and DM-PCN.

3.4.1 Flow Termination with Dual Marking

DM-PCN uses two metering and marking algorithms. The AC method requires exhaustive marking based on the admissible rate as reference rate as described above. The FT method requires excess marking based on the supportable rate as reference rate. As a consequence, *SR*-overload is marked with “excess-traffic” (ET). To implement FT, the egress node determines the rate of ET-marked traffic (ETR) for each IEA and triggers the termination of appropriate flows from the IEA to quickly reduce the PCN traffic rate by ETR in order to remove *SR*-overload. The mechanisms in [11, 12] work similarly.

3.4.2 Flow Termination with Single Marking

SM-PCN uses only a single metering and marking algorithm. The AC method requires excess marking based on the admissible rate as reference rate. The FT method does not need another marking algorithm, it just requires that the supportable rates are fixed multiples of admissible rates (cf. Equation (1)). To implement FT, each egress node determines the rate of AS-marked and non-AS-marked traffic (ASR, $nASR$) per IEA. If the overall PCN traffic rate ($ASR + nASR$) is larger than b times the fraction of non-AS-marked traffic ($b \cdot nASR < ASR + nASR$), some link was SR-pre-congested. Thus, the rate to be terminated from the IEA is

$$TR = \max(0, ASR + nASR - b \cdot nASR) = \max(0, ASR - (b-1) \cdot nASR). \qquad (2)$$

The mechanisms in [13, 35] work similarly.

3.4.3 Pros and Cons of Single and Dual Marking

As mentioned above, SM-PCN requires less support in routers than DM-PCN. Furthermore, SM-PCN re-marks NP-marked packets only to “admission-stop” (AS) while DM-PCN re-marks NP-marked packets to “admission-stop” (AS) and “excess-traffic” (ET). Thus, DM-PCN requires more codepoints in the packet header than SM-PCN and is, therefore, harder to implement in today’s Internet as free codepoints in the IP header are a scarce resource and hardly available. However, SM-PCN does not work well with multipath routing [10] and AC methods do not work well with small IEAs. They react with significant delay when the packet rate of the IEA is small because excess marking AS-marks only a small fraction of the traffic. Small IEAs are not negligible because they are expected to be the majority of IEAs in future core networks [36]. Nevertheless, SM-PCN currently seems to be the preferred option in the standardization process.

3.5 Conventional and Resilient Admission Control with and without Flow Termination

We discuss the use of conventional and resilient AC with and without FT.

3.5.1 Conventional AC

The objective of conventional AC is to block new flows to avoid overload created by users. Almost the full link bandwidth can be used to carry high-priority traffic as long as delay bounds are respected. As a consequence, the admissible rate threshold $AR(l)$ of a link l can be set close to its bandwidth $c(l)$ when the traffic is smooth enough.

3.5.2 Resilient AC

In case of failures, traffic is possibly rerouted and can lead to congestion on backup paths. In fact, this is the major reason for congestion in today's Internet. As shown in [5], only 20% of the congestion observed in core networks are caused by increased user activity, but 80% of the congestion is caused by traffic which is redirected due to failures. Conventional AC cannot guarantee QoS for such cases, but this can be achieved by resilient AC [6, 7]. We call the failures for which no congestion should occur protected failures. Only a fraction of the link bandwidth can be used to carry primary traffic since the remaining fraction is required for backup purposes in case of protected failures. This needs to be respected by AC, and *AR*-thresholds must be set low enough.

3.5.3 Conventional AC with FT

Conventional AC cannot avoid overload situations in case of failures. Therefore, it may be combined with FT. The supportable rates $SR(l)$ are also set close to the link bandwidth $c(l)$, but larger than $AR(l)$. Some safety margin is required between $AR(l)$ and $SR(l)$ to avoid unwanted termination of admitted traffic and between $SR(l)$ and $c(l)$ to avoid slow flow termination. In case of a failure, a large number of admitted flows are possibly terminated. This may be acceptable for some applications and unacceptable for others.

Networks using conventional AC with FT can be provided with sufficient backup capacity. The difference to resilient AC is that almost the entire link bandwidth can be used to admit new traffic. This has two implications. On the one hand, it reduces blocking when more traffic than expected requests admission. On the other hand, if more traffic than expected is admitted, the capacity on backup paths might not suffice in failure cases and hence flows must be terminated. Thus, resilient transport services cannot be provided for admitted traffic. However, they are desirable for demanding applications such as tele-medicine or tele-control of industrial applications.

3.5.4 Resilient AC with FT

Resilient AC admits only as much traffic as can be carried without QoS degradation over the network after rerouting in case of protected failures. However, unlikely failures can happen for which backup capacity does not suffice. Therefore, FT is also a desirable function in combination with resilient AC. Again, the *SR* thresholds may be set close to the link bandwidths with a safety margin towards $c(l)$ in order to guarantee a sufficiently fast termination process. *AR* thresholds are set to lower values. In contrast to conventional AC with FT, a flow is not likely to be terminated once it is admitted such that resilient transport services can be offered.

When PCN-based AC is configured for conventional non-resilient AC, the *AR*-thresholds can be set to almost the link bandwidth and no sophisticated algorithms are required. Resilient AC in general is more difficult. In [6, 37] algorithms are provided to calculate tunnel-specific capacities for a resilient tunnel-based AC. However, this solution cannot be applied for resilient PCN-based AC. Resilient PCN-based AC requires the computation of link-specific *AR*- and *SR*-thresholds. They must be set in such a way that admitted traffic can be accommodated after rerouting in case of protected failure scenarios without being terminated. In case of DM-PCN, only *AR*-thresholds need to be calculated because *SR*-thresholds can be set close to the link bandwidth independently of corresponding *AR*-thresholds. This is different for SM-PCN because the *SR*- and *AR*-thresholds are connected via Equation (1) which makes the threshold assignment problem more complex.

The amount of traffic that can be carried over a network during normal operation and after rerouting in protected failure cases depends on the routing and rerouting function. Moreover, more flows can be carried when they have shorter paths. To be independent of this issue, we consider for throughput maximization problems the fraction or multiple of a traffic matrix that can be supported by a network. In [38], we provided heuristic methods to optimize IP routing to maximize the protected transport capacity for a fraction or multiple of a given traffic matrix. It is applicable in DM-PCN, but not in SM-PCN because SM-PCN requires that the ratio of primary and backup capacity is exactly $\frac{1}{b-1}$. Thus, IP routing optimization is more complex for SM-PCN than for DM-PCN and new objective functions are required.

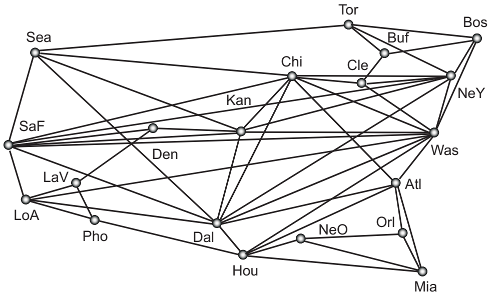
This work develops simple algorithms for the threshold setting and routing optimization problem to provide traffic engineering for resilient PCN-based AC and FT, both for DM-PCN and the more complex SM-PCN. Moreover, a performance study quantifies their difference in the ability to use network resources efficiently.

4 Threshold Configuration for PCN-Based Flow Control

In this section we propose simple and improved algorithms for the configuration of the *AR*- and *SR*-thresholds for SM- and DM-PCN. The simple algorithms set thresholds in such a way that the same fraction of all expected ingress-egress aggregates can be admitted as high priority traffic. This possibly leaves some of the link capacities unused. Therefore, the improved algorithms strive for a higher resource utilization while implementing max-min fairness [39] among ingress-egress aggregates with regard to their admissible rates. This is conceptually similar to the problems treated in [6, 37, 40] but significantly differs by technical constraints. We illustrate the effect of the simple and improved algorithms by numerical results.

4.1 Test Environment and Nomenclature

For our study, we use the Labnet03 network given in Figure 3(a) with equal capacity links. We assume a traffic matrix proportional to the city sizes $\pi(v)$ which are given in Figure 3(b). We use this wide-spread gravity model because of its simplicity although recent research has shown that other models are more realistic [41, 42]. However, our findings do not depend on the accuracy of the traffic matrix. Explicit formulae for the gravity model are given in [6, Equation 3.41].



(a) Topology.

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

(b) Population.

Fig. 3. Labnet03 is the experimental network of the KING project [6] and was inspired by the topology of the former North-American UUNET network.

A network is given by its graph $(\mathcal{V},\mathcal{E})$, where $\mathcal{V}$ is the set of nodes and $\mathcal{E}$ is the set of unidirectional links. The capacity of a link $l \in \mathcal{E}$ that can be used for the transmission of high priority traffic is denoted by $c(l)$. The flows between any two routers $v,w \in \mathcal{V}$ constitute an ingress-egress aggregate (IEA) g whose rate $c(g)$ (or $c(v,w)$) is given by the traffic matrix. $\mathcal{G}$ is the set of all IEAs. We want PCN-based AC to prevent congestion in the presence of a set $\mathcal{S}$ of protected failure scenarios. A failure scenario $s \in \mathcal{S}$ is described by the set of failed network elements and hence $s=\emptyset$ is the failure-free scenario. In our performance studies, $\mathcal{S}$ comprises the failure-free case as well as all single link and router failures.

Routing in IP networks depends on virtual link costs and traffic is forwarded along least-cost paths. We represent the network-wide link costs by a vector $\mathbf{k}$ with one entry for each link $l \in \mathcal{E}$. Standard link costs are defined by the hop count metric, i.e., all link costs are set to one ($\mathbf{k}=\mathbf{1}$). They are default in this section while the link costs are modified in Section 5 to optimize the routing.

Routing also depends on the failure scenario s because network failures lead to rerouting. We describe the routing by the function $u(g,l,s,\mathbf{k})$ indicating the fraction of IEA g being carried over link l in failure scenario s. Throughout our study, we use only single-path routing taking the next hop with the lowest ID in case of equal-cost paths. Thus, the routing function then yields 0 or 1. This is a reasonable decision because some methods for PCN-based AC and FT do not work well with multipath routing. In our computations, we often need the relative load or virtual utilization of a link l in a failure scenario s, its maximum over all links $l \in \mathcal{E}$, its maximum over all failure scenarios $s \in \mathcal{S}$, and its maximum over all links $l \in \mathcal{E}$ and failure scenarios $s \in \mathcal{S}$:

$$\rho(l,s,\mathbf{k}) = \frac{\sum_{g\in\mathcal{G}} c(g)\cdot u(g,l,s,\mathbf{k})}{c(l)}, \tag{3}$$

$$\rho_{max}^{\mathcal{E}}(s,\mathbf{k}) = \max_{l\in\mathcal{E}}\left(\rho(l,s,\mathbf{k})\right), \tag{4}$$

$$\rho_{max}^{\mathcal{S}}(l,\mathbf{k}) = \max_{s\in\mathcal{S}}\left(\rho(l,s,\mathbf{k})\right), \tag{5}$$

$$\rho_{max}^{\mathcal{E},\mathcal{S}}(\mathbf{k}) = \max_{l\in\mathcal{E},s\in\mathcal{S}}\left(\rho(l,s,\mathbf{k})\right). \tag{6}$$

4.2 Simple Assignment of Admissible and Supportable Rate Thresholds

We present a simple, intuitive algorithm to set the admissible and supportable rate thresholds $AR(l)$ and $SR(l)$ for PCN-based AC and FT. The required inputs are the network bandwidths $c(l), l \in \mathcal{E}$, the traffic matrix $c(g), g \in \mathcal{G}$, and the routing $u(g,l,s,\mathbf{k})$. The objective is to set the AR-thresholds in such a way that all IEAs g can send the same maximum multiple $\sigma(\mathbf{k})$ of their expected rates $c(g)$ without causing congestion in protected failure scenarios $s \in \mathcal{S}$ after rerouting. In the following, we call this maximum multiple $\sigma(\mathbf{k})$ the "scaling factor". It is the metric for the performance comparison.

4.2.1 Dual Marking PCN Architecture

The largest link utilization in the network including protected failure scenarios is $\rho_{max}^{\mathcal{E},\mathcal{S}}(\mathbf{k})$. Therefore, scaling the traffic matrix by

$$\sigma_{DM}(\mathbf{k}) = \frac{1.0}{\rho_{max}^{\mathcal{E},\mathcal{S}}(\mathbf{k})} \tag{7}$$

prevents the virtual link utilization $\rho(l,s,\mathbf{k})$ from exceeding 100% in any protected failure scenario $s \in \mathcal{S}$. Therefore, we compute the *AR*- and *SR*-thresholds for DM-PCN

$$AR(l) = \sigma(\mathbf{k}) \cdot \sum_{g \in \mathcal{G}} c(g) \cdot u(g, l, \emptyset, \mathbf{k}) \tag{8}$$

$$SR(l) = \sigma(\mathbf{k}) \cdot \max_{s \in \mathcal{S}} \left(\sum_{g \in \mathcal{G}} c(g) \cdot u(g, l, s, \mathbf{k}) \right) \tag{9}$$

by scaling the expected link loads under failure-free operation and their maximum over all protected failure scenarios with $\sigma(\mathbf{k}) = \sigma_{DM}(\mathbf{k})$. With the proposed thresholds, the traffic fraction $\sigma(\mathbf{k})$ of all IEAs can be admitted and the largest relative link load is at most 1.0 in any protected failure scenario $s \in \mathcal{S}$.

The traffic matrix is only a long-time expectation for planning purposes, but short-time variations can occur. With *AR*-thresholds configured according to Equation (8) AC can admit more traffic for a particular IEA g than $\sigma(\mathbf{k}) \cdot c(g)$ and less of another. If this happens, some traffic is possibly not protected and hence may be terminated in case of a very special failure scenario. This observation holds for PCN-based AC and FT in general and is not specific to our algorithms.

4.2.2 Single Marking PCN Architecture

We set the *AR*- and *SR*-thresholds for the single marking architecture in a similar way. Without AC, the maximum link utilization in all protected failure scenarios is $\rho_{max}^{\mathcal{E},\mathcal{S}}(\mathbf{k})$; a maximum link utilization $\rho_{max}^{\mathcal{E}}(\emptyset, \mathbf{k})$ is observed in the failure-free scenario and Constraint (1) requires that up to the b-multiple of this traffic needs to be accommodated in failure scenarios; otherwise, *SR*-pre-congestion cannot be detected. When scaling the traffic matrix with

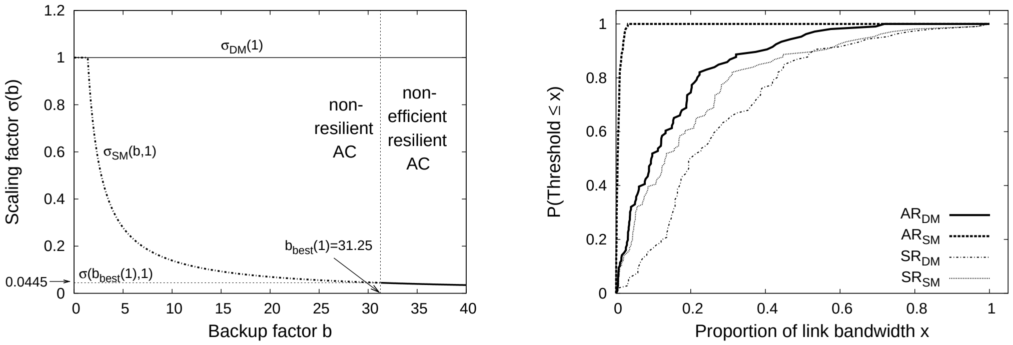
$$\sigma_{SM}(b, \mathbf{k}) = \frac{1.0}{\max\left(\rho_{max}^{\mathcal{E},\mathcal{S}}(\mathbf{k}), b \cdot \rho_{max}^{\mathcal{E}}(\emptyset, \mathbf{k})\right)} \tag{10}$$

neither the virtual link utilizations $\rho(l, s, \mathbf{k})$ nor the expression $b \cdot \rho(l, \emptyset, \mathbf{k})$ exceed 1.0 and at least one of them is exactly 1.0 for at least one link $l \in \mathcal{E}$ and failure scenario $s \in \mathcal{S}$. Finally, the *AR*- and *SR*-thresholds can be set according to Equation (8) using $\sigma(\mathbf{k}) = \sigma_{SM}(b, \mathbf{k})$ and to Equation (1).

4.2.3 Comparison

The scaling factor $\sigma(\mathbf{k})$ expresses the multiple of the traffic matrix that can be admitted as protected priority traffic. Therefore, it is a suitable measure to compare the efficiency of SM- and DM-PCN. Initially we choose the overall traffic load in the network such that we get a scaling factor of $\sigma_{DM}(\mathbf{1}) = 1.0$ for DM-PCN. For SM-PCN the scaling factor $\sigma_{SM}(b, \mathbf{1})$ depends on the backup factor b and Figure 4(a) illustrates that it decreases with increasing b. The optimum backup factor is

$$b_{best}(\mathbf{k}) = \max_{l \in \mathcal{E}} \left(\frac{\rho^{\mathcal{S}}_{max}(l,\mathbf{k})}{\rho(l,\emptyset,\mathbf{k})} \right). \tag{11}$$



(a) Scaling factor $\sigma(\mathbf{k})$ depending on the backup factor b. SM-PCN is resilient only for $b > b_{best}(\mathbf{1})$.

(b) CDF of *AR*- and *SR*-threshold sizes relative to the respective link bandwidths.

Fig. 4. Simple threshold assignment for SM- and DM-PCN. The routing is based on the hop count metric ($\mathbf{k} = \mathbf{1}$).

For backup factors b smaller than $b_{best}(\mathbf{k})$, SM-PCN is not resilient: the *AR*-thresholds are set low enough that the link capacity will suffice to carry rerouted admitted traffic, but the *SR*-thresholds are possibly set to too low values such that some flows will be unnecessarily terminated in protected failure scenarios. For backup factors b larger than $b_{best}(\mathbf{k})$, SM-PCN is resilient, but the large backup factor reserves too much backup capacity resulting in smaller *AR*-values such that AC is less efficient. The best backup factor for the experimental setting in the Labnet03 is $b_{best}(\mathbf{1}) = 31.25$ and leads to a scaling factor of $\sigma_{SM}(b_{best}(\mathbf{1}), \mathbf{1}) = 0.0445$. Thus, SM-PCN can carry only 4.4% of the traffic that can be supported by DM-PCN.

We choose the backup factor b for SM-PCN according to Equation (11) in the following experiment. Figure 4(b) illustrates the impact of SM- and DM-PCN on the *AR*- and *SR*-threshold sizes achieved by the above algorithm. The figure shows the cumulative distribution function (CDF) of the threshold sizes relative to the respective link bandwidths. The x-axis indicates the proportion of link bandwidth x and the y-axis indicates the percentage of links for which the relative *AR*- or *SR*-threshold sizes $\frac{AR(l)}{c(l)}$ and $\frac{SR(l)}{c(l)}$ are smaller than or equal to x. Both SM- and DM-PCN have at least one *SR*-threshold using 100% of the respective link bandwidth. This shows that scaling factors $\sigma_{DM}(\mathbf{1})$ and $\sigma_{SM}(b, \mathbf{1})$ cannot be further increased. The *AR*-thresholds are substantially smaller than the *SR*-thresholds, especially for SM-PCN, which is due to the large backup factor b that cannot be decreased without losing the resilience property of the AC. The average relative size of the *AR*-thresholds is 14.94% for DM-PCN while it is only 0.67% for SM-PCN. Thus, SM-

PCN can admit only very little high-priority traffic with resilience requirements.

4.3 Improved Threshold Assignment of Admissible and Supportable Rates

In the section above, all IEAs were associated with the same scaling factor $\sigma(\mathbf{k})$ which was used to set the *AR*- and *SR*-thresholds based on Equation (8), (9), and (1). We introduce now IEA-specific scaling factors $\sigma(g,\mathbf{k})$, i.e., the scaling factors of some IEAs can be increased if enough resources are available. This leads to larger *AR*- and *SR*-thresholds and allows better usage of link bandwidths.

The basic idea is as follows. The *AR*-threshold $AR(l)$ limits the admissible rate for all IEAs being carried over a specific link l. They can be scaled up to a certain value $\sigma(l,\mathbf{k})$ if their rate is not limited by other thresholds, yet. Thus, $\sigma(l,\mathbf{k})$ indicates the competition for resources on link l: a low value of $\sigma(l,\mathbf{k})$ expresses scarce resources while a large value of $\sigma(l,\mathbf{k})$ expresses abundant resources. Consider an IEA g and its path $p(g)$. The IEA-specific scaling factor $\sigma(g,\mathbf{k}) = \min_{l\in p(g)}(\sigma(l,\mathbf{k}))$ is the minimum scaling factor of the links in the path of g. Limiting the rate of g according to this scaling factor assures that the capacity of the bottleneck link of its path is shared fairly among the flows competing for this link. Conversely, to limit a scaling factor $\sigma(g,\mathbf{k})$ for a certain IEA g, at least one *AR*-threshold of the links along its path $p(g)$ needs to be set to a sufficiently low value.

4.3.1 Dual Marking PCN Architecture

Algorithm 1 determines iteratively the IEA-specific scaling factors $\sigma(g,\mathbf{k})$ for all IEAs $g\in\mathcal{G}$ and sets the link-specific *AR*-thresholds. Before we explain the algorithm, we need some nomenclature and auxiliary functions.

We call an IEA g “fixed” if its scaling factor $\sigma(g,\mathbf{k})$ is already determined; otherwise we call it “free”. The set of all fixed and all free IEAs is denoted by $\mathcal{G}_{fixed}$ and $\mathcal{G}_{free}$. The set of IEAs with traffic routed over a specific link l in a specific failure scenario s is denoted by $\mathcal{G}(l,s) = \{g\in\mathcal{G} : u(g,l,s,\mathbf{k}) > 0\}$. If a link l carries a certain set of fixed and free IEAs in a specific failure scenario s, the capacity left over by the fixed IEAs can be shared among the free IEAs. Thus, we can calculate an upper bound on the link- and failure-scenario-specific scaling factor $\sigma(l,s,\mathbf{k})$ by

$$\sigma(l,s,\mathbf{k}) = \frac{c(l) - \sum_{g\in\mathcal{G}_{fixed}} \sigma(g,\mathbf{k})\cdot c(g)\cdot u(g,l,s,\mathbf{k})}{\sum_{g\in\mathcal{G}_{free}} c(g)\cdot u(g,l,s,\mathbf{k})} \tag{12}$$

if link l carries at least one free IEA. Furthermore, we determine the smallest free scaling factor $\sigma_{min}^{free}(\mathbf{k})$ by

```
Input: $\mathcal{G}, \mathcal{G}(l,s)$

  $\mathcal{G}_{free} = \mathcal{G}, \mathcal{G}_{fixed} = \emptyset, \mathcal{E}^{AR}_{fixed} = \emptyset$
  while $\mathcal{G}_{free} \neq \emptyset$ do
    Calculate $\sigma^{free}_{min}(\mathbf{k})$ according to Equation (13)
    $\mathcal{B} = \emptyset$    {Collect all bottlenecked IEAs in $\mathcal{B}$}
    for all $s \in \mathcal{S}, l \in \mathcal{E}$ do
      if $\sigma(l,s,\mathbf{k}) = \sigma^{free}_{min}(\mathbf{k})$ then
        for all $g \in \mathcal{G}_{free}$ do
          if $u(g,l,s,\mathbf{k}) > 0$ then
            $\mathcal{B} = \mathcal{B} \cup g$
          end if
        end for
      end if
    end for
    while $\mathcal{B} \neq \emptyset$ do    {Enforce scaling factor $\sigma^{free}_{min}(\mathbf{k})$ for bottlenecked IEAs
    by setting AR-thresholds small enough.}
      choose appropriate $g^* \in \mathcal{B}$
      choose appropriate $l^* \in \mathcal{E} \setminus \mathcal{E}^{AR}_{fixed} : u(g^*,l,s,\mathbf{k}) > 0$
      $AR(l^*) = \sum_{g \in \mathcal{G}_{fixed}} \sigma(g,\mathbf{k}) \cdot c(g) \cdot u(g,l^*,\emptyset,\mathbf{k}) +$
                 $\sigma^{free}_{min}(\mathbf{k}) \cdot \sum_{g \in \mathcal{G}_{free}} c(g) \cdot u(g,l^*,\emptyset,\mathbf{k})$
      $\mathcal{E}^{AR}_{fixed} = \mathcal{E}^{AR}_{fixed} \cup l^*$
      for all $g \in \left(\mathcal{G}(l^*,\emptyset) \cap \mathcal{G}_{free}\right)$ do
        $\sigma(g,\mathbf{k}) = \sigma^{free}_{min}(\mathbf{k})$
        $\mathcal{B} = \mathcal{B} \setminus g$
        $\mathcal{G}_{free} = \mathcal{G}_{free} \setminus g$
        $\mathcal{G}_{fixed} = \mathcal{G}_{fixed} \cup g$
      end for
    end while
  end while

Output: Scaling factors $\sigma(g,\mathbf{k})$ for $g \in \mathcal{G}$, threshold sizes $AR(l)$ for
        $l \in \mathcal{E}^{AR}_{fixed}$
```

Algorithm 1: Computation of improved AR-thresholds.

$$\sigma^{free}_{min}(\mathbf{k}) = \min_{\{l \in \mathcal{E}, s \in \mathcal{S} : |\mathcal{G}(l,s) \cap \mathcal{G}_{free}| > 0\}} \left(\sigma(l,s,\mathbf{k})\right) \tag{13}$$

among the combinations of (l,s) with at least one free IEA. Those combinations with $\sigma(l,s,\mathbf{k}) = \sigma^{free}_{min}(\mathbf{k})$ are bottleneck combinations and we call the respective free IEAs “bottlenecked IEAs”.

Algorithm 1 starts with initializing the set of free IEAs by $\mathcal{G}_{free} = \mathcal{G}$, the set of fixed IEAs by $\mathcal{G}_{fixed} = \emptyset$, and the set of links with already assigned AR-thresholds by

$\mathcal{E}^{AR}_{fixed} = \emptyset$. The algorithm repeats the following steps until all IEAs are fixed.

The minimum scaling factor $\sigma^{free}_{min}(\mathbf{k})$ of the free IEAs is calculated and the bottlenecked IEAs are collected in the set $\mathcal{B}$. Their scaling factors $\sigma(g,\mathbf{k})$ need to be limited to $\sigma^{free}_{min}(\mathbf{k})$ by setting at least one *AR*-threshold on their paths small enough. Thus, the algorithm repeats the following steps until the set of bottlenecked IEAs $\mathcal{B}$ is empty.

An appropriate IEA g^* is chosen from the set $\mathcal{B}$. It can be, e.g., an IEA with a shortest (longest) path. Other criteria are possible. To limit the scaling factor $\sigma(g,\mathbf{k})$ of this IEA, a suitable link l^* is chosen from the path $p(g)$ for which the *AR*-threshold is not yet determined. Such a link l^* carries, e.g., the smallest (largest) number of free IEAs, the smallest (largest) rate of free IEAs, or it carries on average the shortest (longest) free IEAs.[1] The correct size of this *AR*-threshold is determined and the link l^* is added to the set $\mathcal{E}^{AR}_{fixed}$. All other free IEAs that are carried over l^* in the failure-free scenario are also limited by this new *AR*-threshold. Therefore, their scaling factor is also set to $\sigma(g,\mathbf{k}) = \sigma^{free}_{min}(\mathbf{k})$, they are removed from the set of bottlenecked IEAs $\mathcal{B}$, and moved from $\mathcal{G}_{free}$ to $\mathcal{G}_{fixed}$.[2]

The algorithm terminates since at least one free IEA becomes fixed in each outer while loop. At program termination, the scaling factors $\sigma(g,\mathbf{k})$ are determined for all IEAs $g \in \mathcal{G}$ as well as the *AR*-thresholds for all links $l \in \mathcal{E}^{AR}_{fixed}$. In pathological scenarios where IEAs with one-link paths are missing, *AR*-thresholds for some links might not be fixed because the scaling factors of all IEAs carried over these links are already limited by the *AR*-thresholds of other links. Then, these *AR*-thresholds can be set to $AR(l) = \sum_{g\in\mathcal{G}} \sigma(g,\mathbf{k}) \cdot c(g) \cdot u(g,l,\emptyset,\mathbf{k})$. The *SR*-thresholds can be set to values of

$$SR(l) = \max_{s\in\mathcal{S}} \left(\sum_{g\in\mathcal{G}} \sigma(g,\mathbf{k}) \cdot c(g) \cdot u(g,l,s,\mathbf{k}) \right) \tag{14}$$

or larger as long as they are smaller than $c(l)$.

4.3.2 Single Marking PCN Architecture

The threshold assignment for SM-PCN works similarly. However, unlike DM-PCN, only $\frac{1}{b}$ of the maximum bandwidth $c(l)$ is available to admit traffic in the failure-free case because of Equation (1). Therefore, we adjust Equation (12) to calculate

[1] In our experiments, the results of this algorithm were rather insensitive towards different policies. Further studies and optimizations are possible but do not change the basic principle.

[2] This part of the algorithm is limited to single path routing for which PCN is currently designed. As soon as the PCN behavior is clear for multipath routing, the above algorithm can be adapted.

$\sigma(l,\emptyset,\mathbf{k})$ for the failure-free scenario by the following equation:

$$\sigma(l,\emptyset,\mathbf{k}) = \frac{\frac{c(l)}{b} - \sum_{g\in\mathcal{G}_{fixed}} \sigma(g,\mathbf{k})\cdot c(g)\cdot u(g,l,\emptyset,\mathbf{k})}{\sum_{g\in\mathcal{G}_{free}} c(g)\cdot u(g,l,\emptyset,\mathbf{k})}. \tag{15}$$

The *AR*- and *SR*-thresholds for SM-PCN are calculated in two steps. In a first step we determine the appropriate backup factor $b_{best}(\mathbf{k})$ based on the expected, unscaled traffic matrix using Equation (11). We calculate the *AR*-thresholds according to Algorithm 1 based on $b_{best}(\mathbf{k})$ and the scaling factors in Equation (15) instead of Equation (12) where applicable. In a second step, we determine again the appropriate backup factor $b^*_{best}(\mathbf{k})$ using Equation (11) but based on the scaled traffic matrix $(c(g)\cdot\sigma(g,\mathbf{k}))_{g\in\mathcal{G}}$. The new value $b^*_{best}(\mathbf{k})$ is possibly smaller than the value $b_{best}(\mathbf{k})$ from the first step. In that case, at most $\frac{b^*_{best}(\mathbf{k})}{b_{best}(\mathbf{k})}$ of any link capacity will be used in any considered failure scenario $s\in\mathcal{S}$. Therefore, we finally multiply the obtained scaling factors $\sigma(g,\mathbf{k})$, $AR(l)$, and $SR(l)$ by $\frac{b_{best}(\mathbf{k})}{b^*_{best}(\mathbf{k})}$ to maximize the rate thresholds without risking overload in any $s\in\mathcal{S}$.

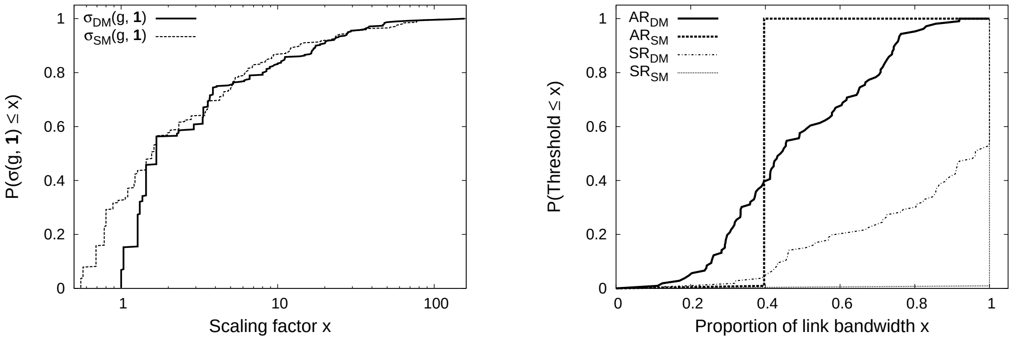
4.3.3 Comparison

We calculate the IEA-specific scaling factors $\sigma(g,\mathbf{k})$ and the *AR*- and *SR*-threshold sizes according to the improved threshold assignment algorithm. Simple threshold assignment leads to a common scaling factor for all IEAs of 1.0 and 0.0445 for DM- and SM-PCN, respectively. Improved threshold assignment increases the scaling factors to average values of 6.90 and 6.51. However, the minimum scaling factor

$$\sigma^{\mathcal{G}}_{min}(\mathbf{k}) = \min_{g\in\mathcal{G}}(\sigma(g,\mathbf{k})) \tag{16}$$

limits the supportable scaling of the entire traffic matrix and the corresponding values are $\sigma^{\mathcal{G}}_{min}(\mathbf{1}) = 1.0$ and 0.5519. Thus, the value for DM-PCN does not change, but SM-PCN benefits a lot from improved threshold assignment. The CDF of individual IEA-specific scaling factors is illustrated in Figure 5(a) both for SM- and DM-PCN. They are distributed over a broad range with maximum values at 156.55 and 107.41. Most of the IEA-specific scaling factors $\sigma(g,\mathbf{k})$ for SM-PCN are significantly smaller than those of DM-PCN. Therefore, DM-PCN is still clearly more efficient than SM-PCN.

We study the impact of the improved threshold assignment on the relative *AR*- and *SR*-threshold sizes. Figure 5(b) illustrates their CDFs and a comparison with Figure 4(b) shows that the threshold sizes are significantly larger with improved threshold assignment than with simple threshold assignment. The average relative *AR*-threshold size increases from 14.94% to 48.75% for DM-PCN and from 0.67% to 39.66% for SM-PCN. We observe the tremendous increase of the *AR*-threshold sizes for SM-PCN because the improved threshold assignment decreases



(a) CDF of IEA-specific scaling factors $\sigma(g,\mathbf{1})$.

(b) CDF of *AR*- and *SR*-threshold sizes relative to the respective link bandwidths.

Fig. 5. Improved threshold assignment for SM- and DM-PCN. The routing is based on the hop count metric $(\mathbf{k}=\mathbf{1})$.

the backup factor from $b_{best}(\mathbf{1})=31.25$ down to $b^{*}_{best}(\mathbf{1})=2.52$. A closer look at the CDF of the *AR*-thresholds for SM-PCN in Figure 5(b) shows that all *AR*-thresholds are set to exactly 39.66% of the respective link bandwidth and the corresponding *SR*-thresholds are set to 100%. This is different for DM-PCN: some *AR*- and *SR*-thresholds use only 20% of the link bandwidth, and some others use more than 80%. Thus, optimum threshold sizes for DM-PCN are more heterogeneous than for SM-PCN.

5 Routing Optimization for PCN-Based Flow Control

In this section we derive objective functions for routing optimization to maximize the protected throughput of high-priority traffic for both SM- and DM-PCN. We illustrate the effect of the algorithms by numerical results.

5.1 Routing Optimization to Increase AR- and SR-Threshold Sizes

The maximum link utilization in the failure-free scenario $\rho^{\mathcal{E}}_{max}(\emptyset)$ can be minimized by routing optimization. In IP networks, the routing depends on the virtual link costs $\mathbf{k}$ whose setting can be optimized such that $\rho^{\mathcal{E}}_{max}(\emptyset)$ is minimized [43]. In a similar way, the maximum link utilization for a set of protected failure scenarios $\mathcal{S}$ can be reduced [44–46]. We adopt and adapt this principle to increase the *AR*- and *SR*-thresholds by increasing the scaling factors $\sigma(\mathbf{k})$. To that end, we developed our own optimization software [38] based on the simulated annealing-like principle “threshold accepting” [47] to find suitable link costs $\mathbf{k_{best}}$ that minimize a given objective function.

5.1.1 Dual Marking PCN Architecture

To maximize for DM-PCN the protected admissible traffic in terms of a proportion of a given traffic matrix, $\sigma_{DM}(\mathbf{k})$ in Equation (7) needs to be maximized. This is achieved by finding a link cost vector $\mathbf{k_{best}}$ that minimizes the following objective function:

$$\rho_{max}^{\mathcal{E},\mathcal{S}}(\mathbf{k}) \to \min\,. \tag{17}$$

5.1.2 Single Marking PCN Architecture

To maximize for SM-PCN the protected admissible traffic in terms of a proportion of a given traffic matrix, $\sigma_{SM}(b_{best}^{*}(\mathbf{k}),\mathbf{k})$ in Equation (10) needs to be maximized. This is achieved by finding a link cost vector $\mathbf{k_{best}}$ that minimizes the following objective function:

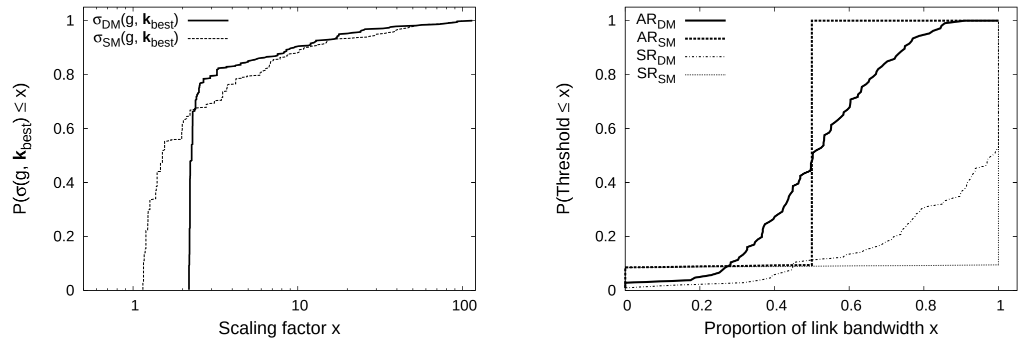
$$\max\left(\rho_{max}^{\mathcal{E},\mathcal{S}}(\mathbf{k}), b_{best}^{*}(\mathbf{k})\cdot\rho_{max}^{\mathcal{E}}(\emptyset,\mathbf{k})\right) \to \min\,. \tag{18}$$

Thereby, the scaling factor $b_{best}^{*}(\mathbf{k})$ is calculated like in Section 4.3.2.

5.1.3 Comparison

We use the routing optimization presented above for SM- and DM-PCN and calculate the scaling factors as well as the *AR*- and *SR*-thresholds by improved threshold assignment. Compared to improved threshold assignment without routing optimization, the minimum scaling factors improve from $\sigma_{min}^{\mathcal{G}}(\mathbf{1}) = 1.0$ to $\sigma_{min}^{\mathcal{G}}(\mathbf{k_{best}}) = 2.1858$ for DM-PCN and from $\sigma_{min}^{\mathcal{G}}(\mathbf{1}) = 0.5519$ to $\sigma_{min}^{\mathcal{G}}(\mathbf{k_{best}}) = 1.1467$ for SM-PCN. Thus, DM-PCN is still about two times more efficient than SM-PCN when routing optimization is applied. The improvement for SM-PCN is partly due to a further reduction of the optimum backup factor from $b_{best}^{*}(\mathbf{1})=2.52$ to $b_{best}(\mathbf{k_{best}})=2.0$. The corresponding CDFs of the IEA-specific scaling factors are illustrated in Figure 6(a). The IEA-specific scaling factors for optimized link costs are more centered around their minimum values than those for the hop count metric (cf. Figure 5(a)). This holds for both SM- and DM-PCN.

After combined routing optimization and improved threshold assignment, the average relative *AR*- and *SR*-threshold sizes are 51.02% and 85.73% for DM-PCN and 45.75% and 91.51% for SM-PCN. However, looking at their CDF in Figure 6(b) we observe that the optimized routing for SM-PCN avoids carrying traffic on a few links. This prevents large backup factors that reduce the scaling factors for SM-PCN. The relative *AR*- and *SR*-threshold sizes of the used links are 50% and 100%, respectively. All used links have the same threshold sizes because of improved threshold assignment.



(a) CDF of the IEA-specific scaling factors $\sigma(g,\mathbf{k_{best}})$.

(b) CDF of the *AR*- and *SR*-threshold sizes relative to the respective link bandwidths.

Fig. 6. Improved threshold assignment for SM- and DM-PCN. The routing is based on optimized link costs $(\mathbf{k}=\mathbf{k_{best}})$.

6 Efficiency of SM- and DM-PCN: A Parametric Study

In this section, we study the ability of SM- and DM-PCN to carry as much protected high-priority traffic as possible. We investigate the impact of simple and improved threshold assignment as well as routing optimization in networks of different size and with different node degree to generalize the results of Sections 4 and 5. We first describe the experiment setup and the exact performance measure and then discuss the results.

6.1 Experiment Setup and Performance Measure

A prerequisite for resilient AC is a resilient network topology which should be at least 2-connected, i.e., any node in the network can fail without partitioning its topology into disconnected subgraphs. Such structures are found in the core of wide area networks, but usually not in access networks. In typical full-fleshed Internet topologies, the number of links connected to a node, i.e. the node degree, usually follows a power law distribution as some few core nodes connect many satellite nodes. This, however, does not lead to a resilient network structure. We use the topology generator of [6] that allows to control the network parameters quite strictly. We randomly generate 15 networks for each combination of size 10, 15, 20, 25, 30, 35, 40, 45, and 50 nodes with an average node degree of 4, 5, and 6 and a maximum deviation from that average of 1. Thus, our experiments comprise altogether 405 different topologies. We use equal link bandwidths and homogenous traffic matrices. As the link bandwidths are not tailored to the traffic matrix, bottlenecks occur on some links.

Our intention is to compare the efficiency of SM- and DM-PCN configured with different threshold assignment algorithms with and without routing optimization.

We want to maximize the multiple of the traffic matrix that can be admitted as protected high-priority traffic. This factor is the minimum scaling factor $\sigma^{\mathcal{G}}_{min}(\mathbf{k})$ (cf. Equation (16)). We calculate the maximum resource utilization of the network

$$\rho(\mathbf{k}) = \sigma^{\mathcal{G}}_{min}(\mathbf{k}) \cdot \frac{\sum_{l\in\mathcal{E}} \left(\sum_{g\in\mathcal{G}} c(g) \cdot u(g,l,\emptyset,\mathbf{k})\right)}{\sum_{l\in\mathcal{E}} c(l)} \qquad (19)$$

based on this scaled traffic matrix and use it as simple performance metric to compare the efficiency of different AC types and configurations in different network topologies.

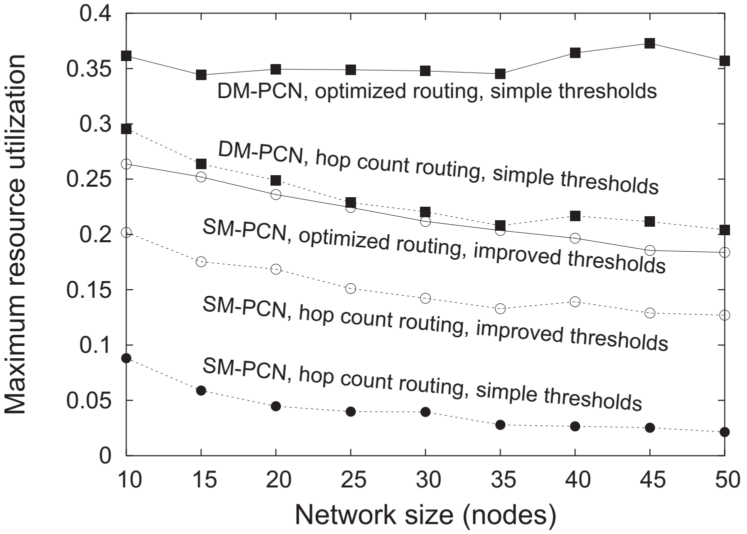


Fig. 7. Maximum resource utilization of SM- and DM-PCN with different configurations depending on the network size (nodes).

6.2 Efficiency of SM- and DM-PCN with Different Configurations

For each network topology we calculate the maximum resource utilization for five different combinations of AC type, threshold assignment, and routing. Figure 7 shows the averaged results depending on the network size. SM-PCN with routing based on the hop count metric and simple threshold assignment is least efficient (4.1% utilization over all experiments), but it can be significantly improved by improved threshold assignment (15.2%). The combination of routing optimization and improved threshold assignment yields a further increase of the AC efficiency (21.7%). DM-PCN with simple threshold assignment and routing based on the hop count metric makes already good use of the network bandwidth (23.3%) and routing optimization further increases its efficiency (35.5%). Improved threshold assignment cannot increase the minimum scaling factor $\sigma^{\mathcal{G}}_{min}(\mathbf{k})$ for DM-PCN, therefore, the corresponding results are missing. For most curves we observe the trend that the maximum resource utilization decreases with increasing network size. This is due to the fact that the probability for strong bottlenecks increases with the network size since the network bandwidths are not tailored to the need of the traffic matrix. Only routing optimization for DM-PCN is able to compensate this structural shortcoming. We also analyzed the impact of the node degree on the the resource utilization, but we have not observed significant dependencies.

After all, improved threshold assignment is crucial to configure SM-PCN for efficient operation. Routing optimization can increase the efficiency of both SM- and DM-PCN. However, DM-PCN can carry significantly more traffic than SM-PCN with and without routing optimization especially in large networks.

7 Conclusion

Pre-congestion notification (PCN) essentially marks packets when PCN traffic exceeds configured admissible or supportable rate thresholds (*AR*, *SR*) on a link of the PCN domain. The IETF attempts to use this feedback for simple and scalable admission control (AC) and flow termination (FT) in IP networks. Currently, there are many different options having benefits and drawbacks [10] that need to be understood. One class of methods requires two different marking mechanisms (dual marking PCN architecture, DM-PCN) and its *AR*- and *SR*-thresholds can be chosen independently of each other. Another class requires only a single marking mechanism (single marking PCN architecture, SM-PCN) and its *SR*-thresholds must be a fixed multiple of the *AR*-thresholds for all links in the PCN domain (cf. Equation (1)).

The objective of this work was to configure the link-specific *AR*- and *SR*-thresholds for PCN domain and to optimize its routing such that the admissible protected high-priority traffic is maximized. This is fairly simple for DM-PCN but more complex for SM-PCN. This is due to the backup factor *b* of Equation (1) and its impact was illustrated in detail for an example network. Our results for a large set of random networks showed that DM-PCN can support 50% more protected traffic than SM-PCN when hop count routing is used. Routing optimization improves the throughput for both SM- and DM-PCN tremendously. With routing optimization, DM-PCN can support even 100% more protected traffic than SM-PCN, at least in large networks.

Finally, this study confirms the initial concern that SM-PCN uses network resources less efficiently for resilient AC than DM-PCN and shows that the difference is significant. This is important information for the standardization process and for ISPs intending to deploy PCN technology in their networks. Moreover, the algorithms presented in this work can be used to configure PCN rate thresholds and to optimize IP routing for PCN networks in practice.

References

[1] K. Fukuda, K. Cho, H. Esaki, A. Kato, The Impact of Residential Broadband Traffic on Japanese ISP Backbones, ACM SIGCOMM Computer Communications Review 35 (1) (2005) 15–22.

[2] K. Cho, K. Fukuda, H. Esaki, A. Kato, The Impact and Implications of the Growth in Residential User-to-User Traffic, in: ACM SIGCOMM, Pisa, Italy, 2006.

[3] S. Shenker, Fundamental Design Issues for the Future Internet, IEEE Journal on Selected Areas in Communications 13 (7) (1995) 1176–1188.

[4] D. M. Johnson, QoS Control versus Generous Dimensioning, British Telecom Technology Journal 23 (2) (2005) 81–96.

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

[6] M. Menth, Efficient Admission Control and Routing in Resilient Communication Networks, PhD thesis, University of Würzburg, Faculty of Computer Science, Am Hubland (July 2004).

[7] M. Menth, R. Martin, J. Charzinski, Capacity Overprovisioning for Networks with Resilience Requirements, in: ACM SIGCOMM, Pisa, Italy, 2006.

[8] IETF Working Group on Congestion and Pre-Congestion Notification (pcn), Description of the Working Group, `http://www.ietf.org/html.charters/pcn-charter.html` (Feb. 2007).

[9] P. Eardley (ed.), Pre-Congestion Notification Architecture, `http://tools.ietf.org/id/draft-ietf-pcn-architecture-08.txt` (Oct. 2008).

[10] M. Menth, F. Lehrieder, B. Briscoe, P. Eardley, T. Moncaster, J. Babiarz, K.-H. Chan, A. Charny, G. Karagiannis, X. J. Zhang, PCN-Based Admission Control and Flow Termination, in: currently under submission, `http://www3.informatik.uni-wuerzburg.de/staff/menth/Publications/Menth08-PCN-Comparison.pdf`, 2008.

[11] B. Briscoe et al., An Edge-to-Edge Deployment Model for Pre-Congestion Notification: Admission Control over a DiffServ Region, `http://tools.ietf.org/id/draft-briscoe-tsvwg-cl-architecture-04.txt` (Oct. 2006).

[12] J. Babiarz, X.-G. Liu, K. Chan, M. Menth, Three State PCN Marking, `http://tools.ietf.org/id/draft-babiarz-pcn-3sm-01.txt` (Nov. 2007).

[13] A. Charny, F. L. Faucheur, V. Liatsos, J. Zhang, Pre-Congestion Notification Using Single Marking for Admission and Pre-emption, `http://tools.ietf.org/id/draft-charny-pcn-single-marking-03.txt` (Nov. 2007).

[14] L. Westberg, A. Bhargava, A. Bader, G. Karagiannis, H. Mekkes, LC-PCN: The Load Control PCN Solution, `http://tools.ietf.org/id/draft-westberg-pcn-load-control-05.txt` (Nov. 2008).

[15] S. Floyd, V. Jacobson, Random Early Detection Gateways for Congestion Avoidance, IEEE/ACM Transactions on Networking 1 (4) (1993) 397–413.

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

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

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

[19] K. Nichols, S. Blake, F. Baker, 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, J.-Y. Le Boudec, SRP: A Scalable Resource Reservation for the Internet, Computer Communications 21 (14) (1998) 1200–1211.

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

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

[24] R. J. Gibbens, F. P. Kelly, Distributed Connection Acceptance Control for a Connectionless Network, in: 16^{th} International Teletraffic Congress (ITC), Edinburgh, UK, 1999, pp. 941 – 952.

[25] F. Kelly, P. Key, S. Zachary, Distributed Admission Control, IEEE Journal on Selected Areas in Communications 18 (12) (2000) 2617–2628.

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

[27] M. Karsten, J. Schmitt, Packet Marking for Integrated Load Control, in: IFIP/IEEE Symposium on Integrated Management (IM), 2005.

[28] D. J. Songhurst, P. Eardley, B. Briscoe, C. di Cairano Gilfedder, J. Tay, Guaranteed QoS Synthesis for Admission Control with Shared Capacity, technical report TR-CXR9-2006-001, BT (Feb. 2006).

[29] M. Menth, S. Kopf, J. Charzinski, K. Schrodi, Resilient Network Admission Control, Computer Networks 52 (14) (2008) 2805–2815.

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

[31] B. Braden, D. Clark, S. Shenker, RFC1633: Integrated Services in the Internet Architecture: an Overview (Jun. 1994).

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

[33] J. Wroclawski, RFC2211: Specification of the Controlled-Load Network Element Service (Sep. 1997).

[34] M. Menth, F. Lehrieder, Performance Evaluation of PCN-Based Admission Control, in: International Workshop on Quality of Service (IWQoS), Enschede, The Netherlands, 2008.

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

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

[37] M. Menth, J. Milbrandt, S. Kopf, Capacity Assignment for NAC Budgets in Resilient Networks, in: International Telecommunication Network Strategy and Planning Symposium (Networks), Vienna, Austria, 2004, pp. 193 – 198.

[38] M. Menth, M. Hartmann, R. Martin, Robust IP Link Costs for Multilayer Resilience, in: IFIP-TC6 Networking Conference (Networking), Atlanta, GA, USA, 2007.

[39] D. Nace, M. Pioro, Max-Min Fairness and Its Applications to Routing and Load-Balancing in Communication Networks: A Tutorial, IEEE Communications Surveys & Tutorials 10 (4).

[40] M. Menth, S. Gehrsitz, J. Milbrandt, Fair Assignment of Efficient Network Admission Control Budgets, in: 18th International Teletraffic Congress (ITC), Berlin, Germany, 2003, pp. 1121–1130.

[41] A. Nucci, A. Sridharan, N. Taft, The Problem of Synthetically Generating IP Traffic Matrices: Initial Recommendations, ACM SIGCOMM Computer Communications Review 35 (3) (2005) 19–32.

[42] V. Erramilli, M. Crovella, N. Taft, An Independent-Connection Model for Traffic Matrices, in: ACM Internet Measurements Conference (IMC), Rio de Janeiro, Brazil, 2006.

[43] B. Fortz, M. Thorup, Internet Traffic Engineering by Optimizing OSPF Weights, in: IEEE Infocom, Tel-Aviv, Israel, 2000, pp. 519–528.

[44] A. Nucci, B. Schroeder, S. Bhattacharyya, N. Taft, C. Diot, IGP Link Weight Assignment for Transient Link Failures, in: 18th International Teletraffic Congress (ITC), Berlin, 2003.

[45] D. Yuan, A Bi-Criteria Optimization Approach for Robust OSPF Routing, in: 3rdIEEE Workshop on IP Operations and Management (IPOM), Kansas City, MO, 2003, pp. 91 – 98.

[46] B. Fortz, M. Thorup, Robust Optimization of OSPF/IS-IS Weights, in: International Network Optimization Conference (INOC), Paris, France, 2003, pp. 225–230.

[47] G. Dueck, T. Scheuer, Threshold Accepting; a General Purpose Optimization Algorithm, Journal of Computational Physics 90 (1990) 161–175.