Congestion and Pre Congestion
Internet-Draft
Intended status: Standards Track
Expires: October 9, 2009

T. Moncaster
BT
B. Briscoe
BT & UCL
M. Menth
University of Wuerzburg
April 7, 2009

Baseline Encoding and Transport of Pre-Congestion Information
draft-ietf-pcn-baseline-encoding-03

Status of This Memo

Copyright Notice

Abstract

The objective of Pre-Congestion Notification (PCN) is to protect the quality of service (QoS) of inelastic flows within a Diffserv domain. The overall rate of the PCN-traffic is metered on every link in the PCN-domain, and PCN-packets are appropriately marked when certain configured rates are exceeded. The level of marking allows the boundary nodes to make decisions about whether t o admit or block a new flow request, and (in abnormal circumstances) whether to terminate some of the existing flows, thereby protecting the QoS of previously admitted flows. This document specifies how such marks are to be encoded into the IP header by re-using the ECN codepoints within this controlled domain. The baseline encoding described here provides for only two PCN encoding states, unmarked and marked.

Table of Contents

1. Introduction

The objective of Pre-Congestion Notification (PCN) is to protect the quality of service (QoS) of inelastic flows within a Diffserv domain, in a simple, scalable and robust fashion. The overall rate of the PCN-traffic is metered on every link in the PCN-domain, and PCN-packets are appropriately marked when certain configured rates are exceeded. These configured rates are below the rate of the link thus providing notification before any congestion occurs (hence "pre-congestion notification"). The level of marking allows the boundary nodes to make decisions about whether to admit or block a new flow request, and (in abnormal circumstances) whether to terminate some of the existing flows, thereby protecting the QoS of previously admitted flows.

This document specifies how these PCN marks are encoded into the IP header by re-using the bits of the ECN field. It also describes how packets are identified as belonging to a PCN flow. Some deployment models require two PCN encoding states, others require more. The baseline encoding described here only provides for two PCN encoding states. However the encoding can be easily extended to provide more states and rules for such extensions are given in this document.

Changes from previous drafts (to be removed by the RFC Editor):

From -02 to -03:

Extensive changes to address comments made by Gorry Fairhurst including:

* Abstract re-written.

* Clarified throughout that this re-uses the ECN bits in the IP header.

* Re-arranged order of terminology section for clarity.

* Table 2 replaced with new table and text.

* Security considerations re-written.

* Appendixes re-written to improve clarity.

* Numerous minor nits and language changes throughout.

Extensive other minor changes throughout.

From -01 to -02:

Removed Appendix A and replaced with reference to [I-D.ietf-tsvwg-ecn-tunnel]

Moved Appendix B into main body of text.

Changed Appendix C to give deployment advice.

Minor changes throughout including checking consistency of capitalisation of defined terms.

Clarified that LU was deliberately excluded from encoding.

From -00 to -01:

Added section on restrictions for extension encoding schemes.

Included table in Appendix showing encoding transitions at different PCN nodes.

Checked for consistency of terminology.

Minor language changes for clarity.

Changes from previous filename

Filename changed from draft-moncaster-pcn-baseline-encoding.

Terminology changed for clarity (PCN-compatible DSCP and PCN-enabled packet).

Minor changes throughout.

Modified meaning of ECT(1) state to EXP.

Moved text relevant to behaviour of nodes into appendix for later transfer to new document on edge behaviours.

From draft-moncaster -01 to -02:

Minor changes throughout including tightening up language to remain consistent with the PCN Architecture terminology

From draft-moncaster -00 to -01:

Change of title from "Encoding and Transport of (Pre-)Congestion Information from within a DiffServ Domain to the Egress"

Extensive changes to Introduction and abstract.

Added a section on the implications of re-using a DSCP.

Added appendix listing possible operator scenarios for using this baseline encoding.

Minor changes throughout.

2. Requirements notation

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

3. Terminology

The following terms are used in this document:

- PCN-compatible Diffserv codepoint - a Diffserv codepoint for which the ECN field is used to carry PCN markings rather than [RFC3168] markings.

- PCN-marked - codepoint indicating packets that have been marked at a PCN-interior-node using some PCN marking behaviour [I-D.ietf-pcn-marking-behaviour]. Abbreviated to PM.

- Not-marked - codepoint indicating packets that are PCN-capable, but are not PCN-marked. Abbreviated to NM.

- PCN-enabled codepoints - collective term for all NM and PM codepoints. By definition, packets carrying such codepoints are PCN-packets.

- not-PCN - packets that are not PCN-enabled.

In addition, the document uses the terminology defined in [I-D.ietf-pcn-architecture].

4. Encoding two PCN States in IP

The PCN encoding states are defined using a combination of the DSCP and ECN fields within the IP header. The baseline PCN encoding

closely follows the semantics of ECN [RFC3168]. It allows the encoding of two PCN states: Not-marked and PCN-marked. It also allows for traffic that is not PCN-capable to be marked as such (not-PCN). Given the scarcity of codepoints within the IP header the baseline encoding leaves one codepoint free for experimental use. The following table defines how to encode these states in IP:

ECN codepoint	Not-ECT (00)	ECT(0) (10)	ECT(1) (01)	CE (11)
DSCP n	not-PCN	NM	EXP	PM

Where DSCP n is a PCN-compatible DiffServ codepoint (see Section 4.3) and EXP means available for Experimental use. N.B. we deliberately reserve this codepoint for experimental use only (and not local use) to prevent future compatability issues.

Table 1: Encoding PCN in IP

The following rules apply to all PCN traffic:

- o PCN-traffic MUST be marked with a PCN-compatible DiffServ Codepoint. To conserve DSCPs, DiffServ Codepoints SHOULD be chosen that are already defined for use with admission controlled traffic, such as the Voice-Admit codepoint defined in [I-D.ietf-tsvwg-admitted-realtime-dscp]. Guidelines for mixing traffic-types within a PCN-domain are given in [I-D.ietf-pcn-marking-behaviour].

- o Any packet that is not-PCN but which shares the same DiffServ codepoint as PCN-enabled traffic MUST have the ECN field equal to 00.

4.1. Valid and Invalid Codepoint Transitions

A PCN-ingress-node MUST set the Not-marked (10) codepoint on any arriving packet that belongs to a PCN-flow. It MUST set the not-PCN (00) codepoint on all other packets.

A PCN-interior-node MUST observe the rules for valid and invalid codepoint transitions as set out in the following table. The precise rules governing which valid transition to use are set out in [I-D.ietf-pcn-marking-behaviour]

	Codepoint Out			
Codepoint in	not-PCN(00)	NM(10)	EXP(01)	PM(11)
not-PCN(00)	Valid	Not valid	Not valid	Not valid
NM(10)	Not valid	Valid	Not valid	Valid
EXP(01)*	Not valid	Not valid	Valid	Valid
PM(11)	Not valid	Not valid	Not valid	Valid

* This SHOULD cause an alarm to be raised at a higher layer. The packet MUST be treated as if it carried the NM codepoint.

Table 2: Valid and Invalid Codepoint Transitions for PCN-packets at PCN-interior-nodes

A PCN-egress-node SHOULD set the not-PCN (00) codepoint on all packets it forwards out of the PCN-domain. The only exception to this is if the PCN-egress-node is certain that revealing other codepoints outside the PCN-domain won't contravene the guidance given in [RFC4774].

4.2. Rationale for Encoding

The exact choice of encoding was dictated by the constraints imposed by existing IETF RFCs, in particular [RFC3168], [RFC4301] and [RFC4774]. One of the tightest constraints was the need for any PCN encoding to survive being tunnelled through either an IP in IP tunnel or an IPSec Tunnel. [I-D.ietf-tsvwg-ecn-tunnel] explains this in more detail. The main effect of this constraint is that any PCN marking has to carry the 11 codepoint in the ECN field since this is the only codepoint that is guaranteeed to be copied down into the inner header upon decapsulation. An additional constraint is the need to minimise the use of DiffServ codepoints as there is a limited supply of standards track codepoints remaining. Section 4.3 explains how we have minimised this still further by reusing pre-existing Diffserv codepoint(s) such that non-PCN traffic can still be distinguished from PCN traffic. There are a number of factors that were considered before deciding to set 10 as the NM state. These included similarity to ECN, presence of tunnels within the domain, leakage into and out of PCN-domain and incremental deployment (see Appendix A.2).

The encoding scheme above seems to meet all these constraints and ends up looking very similar to ECN. This is perhaps not surprising

given the similarity in architectural intent between PCN and ECN.

4.3. PCN-Compatible DiffServ Codepoints

Equipment complying with the baseline PCN encoding MUST allow PCN to be enabled for certain Diffserv codepoints. This document defines the term "PCN-compatible Diffserv codepoint" for such a DSCP. To be clear, any packets with such a DSCP will be PCN enabled only if they also have their ECN field set to indicate a codepoint other than not-PCN.

Enabling PCN marking behaviour disables any other marking behaviour (e.g. enabling PCN disables the default ECN marking behaviour introduced in [RFC3168]). All traffic scheduling and conditioning behaviours are discussed in [I-D.ietf-pcn-marking-behaviour]. This ensures compliance with the BCP guidance set out in [RFC4774].

5. Rules for Experimental Encoding Schemes

Any experimental encoding scheme MUST follow these rules to ensure backward compatibility with this baseline scheme:

- o The 00 codepoint in the ECN field SHALL indicate not-PCN and MUST NOT be changed to any otehr codepoint within a PCN-domain. Therefore an ingress node wishing to disable PCN marking for a packet within a PCN-compatible DiffServ Codepoint MUST set the ECN field to 00.

- o The 11 codepoint in the ECN field SHALL indicate PCN-marked (though this does not exclude the 01 Experimental codepoint from carrying the same meaning).

- o Once set, the 11 codepoint in the ECN field MUST NOT be changed to any other codepoint.

- o Any experimental scheme MUST include details of all valid and invalid codepoint transitions at any PCN nodes.

- o Any experimental scheme MUST NOT update the meaning of the 00 and 11 codepoints defined above.

6. Backwards Compatibility

BCP 124 [RFC4774] gives guidelines for specifying alternative semantics for the ECN field. It sets out a number of factors to be taken into consideration. It also suggests various techniques to allow the co-existence of default ECN and alternative ECN semantics. The baseline encoding specified in this document defines PCN-

compatible DiffServ codepoints as no longer supporting the default ECN semantics. As such this document is compatible with BCP 124. It should be noted that this baseline encoding effectively disables end-to-end ECN except where mechanisms are put in place to tunnel such traffic across the PCN-domain.

7. IANA Considerations

This document makes no request to IANA.

8. Security Considerations

PCN-marking only carries a meaning within the confines of a PCN-domain. Packets wishing to be treated as belonging to a PCN-flow must carry a PCN-Compatible DSCP and a PCN-Enabled ECN codepoint. This encoding document is intended to stand independently of the architecture used to determine whether specific packets are authorised to be PCN-marked, which will be described in separate documents on PCN edge-node behaviour.

This document assumes the PCN-domain to be entirely under the control of a single operator, or a set of operators who trust each other. However future extensions to PCN might include inter-domain versions where trust cannot be assumed between domains. If such schemes are proposed they must ensure that they can operate securely despite the lack of trust but such considerations are beyond the scope of this document.

9. Conclusions

This document defines the baseline PCN encoding utilising a combination of a PCN-enabled DSCP and the ECN field in the IP header. This baseline encoding allows the existence of two PCN encoding states, not-Marked and PCN-marked. It also allows for the co-existence of competing traffic within the same DSCP so long as that traffic does not require ECN support within the PCN-domain. The encoding scheme is conformant with [RFC4774].

10. Acknowledgements

This document builds extensively on work done in the PCN working group by Kwok Ho Chan, Georgios Karagiannis, Philip Eardley, Anna Charny, Joe Babiarz and others. Thanks to Ruediger Geib and Gorry Fairhurst for providing detailed comments on this document.

11. Comments Solicited

(To be removed by the RFC-Editor.) Comments and questions are encouraged and very welcome. They can be addressed to the IETF congestion and pre-congestion working group mailing list <pcn@ietf.org>, and/or to the authors.

12. References

12.1. Normative References

[I-D.ietf-pcn-marking-behaviour] Eardley, P., "Marking behaviour of PCN-nodes", draft-ietf-pcn-marking-behaviour-02 (work in progress), March 2009.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC3168] Ramakrishnan, K., Floyd, S., and D. Black, "The Addition of Explicit Congestion Notification (ECN) to IP", RFC 3168, September 2001.

[RFC4774] Floyd, S., "Specifying Alternate Semantics for the Explicit Congestion Notification (ECN) Field", BCP 124, RFC 4774, November 2006.

12.2. Informative References

[I-D.ietf-pcn-architecture] Eardley, P., "Pre-Congestion Notification (PCN) Architecture", draft-ietf-pcn-architecture-10 (work in progress), March 2009.

[I-D.ietf-tsvwg-admitted-realtime-dscp] Baker, F., Polk, J., and M. Dolly, "DSCP for Capacity-Admitted Traffic", draft-ietf-tsvwg-admitted-realtime-dscp-05 (work in

progress), November 2008.

[I-D.ietf-tsvwg-ecn-tunnel] Briscoe, B., "Tunnelling of Explicit Congestion Notification", draft-ietf-tsvwg-ecn-tunnel-02 (work in progress), March 2009.

[RFC4301] Kent, S. and K. Seo, "Security Architecture for the Internet Protocol", RFC 4301, December 2005.

[RFC5127] Chan, K., Babiarz, J., and F. Baker, "Aggregation of DiffServ Service Classes", RFC 5127, February 2008.

Appendix A. PCN Deployment Considerations

A.1. Choice of Suitable DSCPs

The PCN Working Group chose not to define a single DSCP for use with PCN for several reasons. Firstly the PCN mechanism is applicable to a variety of different traffic classes. Secondly standards track DSCPs are in increasingly short supply. Thirdly PCN should be seen as being essentially a marking behaviour similar to ECN but intended for inelastic traffic. The choice of which DSCP is most suitable for a given PCN-domain is dependant on the nature of the traffic entering that domain and the link rates of all the links making up that domain. In PCN-domains with uniformly high link rates, the appropriate DSCPs would currently be those for the Real Time Traffic Class [RFC5127]. If the PCN domain includes lower speed links it would also be appropriate to use the DSCPs of the other traffic classes that [I-D.ietf-tsvwg-admitted-realtime-dscp] defines for use with admission control, such as the three video classes CS4, CS3 and AF4 and the Admitted Telephony Class.

A.2. Rationale for Using ECT(0) for Not Marked

The choice of which ECT codepoint to use for the Not Marked state was based on the following considerations:

- [RFC3168] full functionality tunnel within the PCN-domain: Either ECT is safe.

- Leakage of traffic into PCN-domain: ECT(1) is slightly less likely to occur so might be considered safer.

- o Leakage of traffic out of PCN-domain: Either ECT is equally unsafe (since this would incorrectly indicate the traffic was ECN-capable outside the controlled PCN-domain).

- o Incremental deployment: Either codepoint is suitable providing that the codepoints are used consistently.

- o Conceptual consistency with other schemes: ECT(0) is conceptually consistent with [RFC3168].

Overall this seemed to suggest ECT(0) was most appropriate to use.

Authors' Addresses

Toby Moncaster
BT
B54/70, Adastral Park
Martlesham Heath
Ipswich IP5 3RE
UK

Phone: +44 1473 648734
EMail: toby.moncaster@bt.com

Bob Briscoe
BT & UCL
B54/77, Adastral Park
Martlesham Heath
Ipswich IP5 3RE
UK

Phone: +44 1473 645196
EMail: bob.briscoe@bt.com

Michael Menth
University of Wuerzburg
room B206, Institute of Computer Science
Am Hubland
Wuerzburg D-97074
Germany

Phone: +49 931 888 6644
EMail: menth@informatik.uni-wuerzburg.de