PCN
Internet-Draft
Intended status: Informational
Expires: May 22, 2008

K. Chan
Nortel
G. Karagiannis
University of Twente
November 19, 2007

Pre-Congestion Notification Encoding Comparison
draft-chan-pcn-encoding-comparison-01

Status of this Memo

Copyright Notice

Abstract

DiffServ mechanisms have been developed to support Quality of Service (QoS). However, the level of assurance that can be provided with DiffServ without substantial over-provisioning is limited. Pre-Congestion Notification (PCN) investigates the use of per-flow admission control to provide the required service guarantees for the admitted traffic. While admission control will protect the QoS under normal operating conditions, an additional flow termination mechanism

is necessary in the times of heavy congestion (e.g. caused by route changes due to link or node failure).

Encoding and their transport are required to carry the congestion and pre-congestion information from the congestion and pre-congestion points to the decision points. This document provides a survey of several encoding methods, using comparisons amongst them as a way to explain their strengths and weaknesses.

Table of Contents

1. Introduction

The main goal of this document is a survey and comparison of several encoding and transport methods that are required to encode the pre-congestion information and to transport it from the PCN interior nodes to the decision PCN egress nodes. In order to accomplish this comparison, a number of criteria are developed. For transporting using data packet (IP) header, the encoding methods investigated are:

1. Encoding using the combination of the ECN and DSCP bits of a data packet header

2. Encoding using the ECN bits of a data packet header

3. Encoding using the DSCP bits of a data packet header

We have also considered:

1. Encoding and transport using a different channel than data packets

The rest of this document is organized as follows:

- Section 2 describes the encoding requirements indicated by currently known detection and marking mechanisms that can be used within the PCN-domain.

- Section 3 describes the encoding and transport methods.

- Section 4 provides the comparison of these methods.

- Section 5 provides the conclusion.

2. Encoding Requirements

The internal PCN encoding requirements are based on the functionality of PCN, and possibly how the PCN Marking Algorithms achieve the functionality. There may be external requirements depending on the environment in which PCN operates, for example co-existence with ECN. These are discussed secondary to the internal PCN encoding requirements because we have limited the PCN operational environment in the PCN WG's first phase charter.

The authors of the different PCN Algorithm documents have agreed to use the notion of Encoding States to represent the information each algorithm wants to export, and hence to be carried from the interior nodes to the edge nodes for flow admission control and flow

termination decisions. These Encoding States form the fundamental functional requirements for the encoding choices.

The encoding states required are

- o PCN Capable Transport Marking, for separation from None PCN Capable Transport
- o Not congested Marking, for indication of No Congestion Indication
- o Admission Marking, for indication of Flow Admission Information
- o Termination Marking, for indication of Flow Termination Information
- o Nonce Marking, for cheater detection
- o Affected Marking mainly for ECMP indication. Note however, that i can be used in combination with the Termination Marking for indication of Flow Termination.

A total of six required encoding states.

3. Encoding Options

There are couple of methods for transporting the encoding states. The method used affects the encoding options. Hence when we describe the different encoding options in this section, we group them based on their transport method.

The encoding transport methods considered are:

- o using the combination of the ECN and DSCP bits of a data packet header
- o using the ECN bits of a data packet header
- o using the DSCP bits of a data packet header
- o using a different channel (e.g., IPFIX, see RFC 3955 [18]) than the IP header of the data packets

We discuss the encoding options for each of the encoding transport methods separately in their own subsections.

In each of the subsections, we use tables to organize the different encoding options within each transport method. The tables contain

abbreviations of terms, their meaning are as follows:

- ECN Bits: This refers to the two bit field in the IP header defined by RFC 3168 [15].
- DSCP: DiffServ Code Point. This refers to the six bit field in the IP header defined by RFC 2474 [10].
- Not-ECT: Not ECN Capable Transport. Defined in RFC 3168 [15].
- ECT(0), ECT(1): ECN Capable Transport. Defined in RFC 3168 [15].
- CE: Congestion Experienced. Defined in RFC 3168 [15].
- NA: Not Applicable. Meaning this field is not used for this encoding choice.
- AM: Admission Marked.
- TM: Termination Marked.
- AFM: Affected Marked.
- PCN: The DSCP field uses a specific code point for PCN traffic.
- Not-CE: Not experiencing congestion. This have the same meaning as ECT(0) and ECT(1), but without the cheater detection functionality.
- NDS-CE: Not DiffServ capable traffic with congestion experienced.

3.1. ECN and DSCP Fields as Encoding Transport

IP header real estate have always been expensive, it is no exception here. With the six required encoding states, we need to be frugal with IP header bit usage. The use of both DSCP and ECN fields allow a clean traffic treatment separation of PCN Capable traffic and None PCN Capable traffic. This natural use of the DSCP field, to provide treatment differentiation of packets using different DSCP encoding, is one way of providing the required "PCN Capable Transport Marking" encoding state.

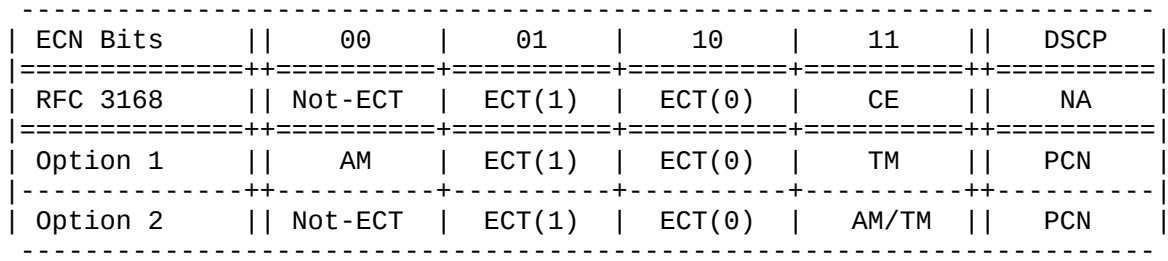


Figure 1: Encoding of PCN Information Using DSCP and ECN Fields

In Figure 1, we listed the fundamental options when both DSCP and ECN fields are used. There are couple of variations of the theme provided by these options. For example, the "01" and "10" encoding can be interpreted as ECT(A) and ECT(T) instead of just ECT(1) and ECT(0) respectively. Using the ECT(A) and ECT(T) variation provides the additional information of the ratio of packets AM marked to packets Not AM marked, and the ratio of packets TM marked to packets Not TM marked. Having these ratios being independent from one another. Another variation on the theme is the use of an extra DSCP value to represent the TM encoding state for Option 2. Doing so will eliminate the need to modulate both AM and TM using the single "11" encoding.

3.1.1. Benefits of Using DSCP and ECN Fields

A major feature of using both DSCP and ECN fields is the ability to use the inherent nature of DiffServ for traffic class separation to allow PCN treatment be applied to PCN traffic, without concerns of applying PCN treatment to none PCN traffic and vise versa. This feature frees this approach for PCN encoding from some of the concerns raised by RFC 4774 [20]. This feature will also keep none PCN Capable traffic out of the PCN treatment mechanisms, allowing the PCN treatment mechanisms focus on their respective PCN tasks.

This approach also leaves the ECN field available totally for PCN encoding states purposes.

3.1.1.1. Concerns on Alternate Semantics for the ECN Field

Section 2 of RFC 4774 [20] raised couple of issues for usage of alternate semantics for the ECN field. We try to address each of the issues in this section.

Section 3.1 of RFC 4774 [20] clarifies Issue 1: "How routers know which ECN semantics to use with which packets." by following the recommendation of RFC 4774 [20] on using a diffserv codepoint to

identify the packets using the alternate ECN semantics. This diffserv codepoint may possibly be a new diffserv codepoint to minimize the possible confusion between using the old per hop behavior of the codepoint and the using of the alternate ECN semantics per hop behavior of the codepoint.

Section 4 of RFC 4774 [20] discusses Issue 2: "How does the possible presence of old routers affect the performance of the alternate ECN connections." and Issue 3: "How does the possible presence of old routers affect the coexistence of the alternate ECN traffic with competing traffic on the path."

The environment using the alternate ECN semantics is envisioned to be within a single administrative domain. With the ability to ensure that all routers along the path understand and agree to the use of the alternate ECN semantics for the traffic identified by the use of a diffserv codepoint. This uses option 2 indicated in section 4.2 of RFC 4774 [20].

Issue 4: "How well does the alternate ECN traffic perform."

The performance of the different proposed alternate ECN (PCN) metering and marking algorithms are currently under study with their simulation and study results described by their respective documentations. Hence not in the scope of this document.

3.1.2. Drawbacks of Using DSCP and ECN Fields

In many cases, a method can provide both benefits and drawbacks. It is just a matter of placement of preference and priority on how one may out weight the other. This is also the case with the use of both DSCP and ECN fields. The use of DSCP will require the setting aside of one DSCP for use by PCN. This may add complexity to the PCN encoding standardization effort and possibly adding complexity when tunneling of the PCN encoding is required.

3.1.3. Comparing DSCP and ECN Fields Encoding Options

Here we discuss the differences between the different encoding options when both DSCP and ECN fields are used. As indicated earlier, when both DSCP and ECN fields are used, there are many encoding options. But we observed they are variations of two themes, indicated by Options 1 and 2 in Figure 1.

When DSCP is used to differentiate between PCN capable and Not-PCN capable traffic, the encoding of "Not-ECT" in the ECN field is not required. This is the motivation for Option 1 in Figure 1, where the encoding "00" for "Not-ECT" is being used for "AM" (Admission

Marking) encoding state. The encodings "01" and "10" for "ECT(1)" and "ECT(0)" supports the required encoding states for "Not Congested Marking" and "Nonce Marking". With the possible additional encoding of "ECT(A)" and "ECT(T)" in place of "ECT(1)" and "ECT(0)" for indicating percentage of Admission Marked traffic and percentage of Termination Marked traffic when the algorithm benefits from such additional information.

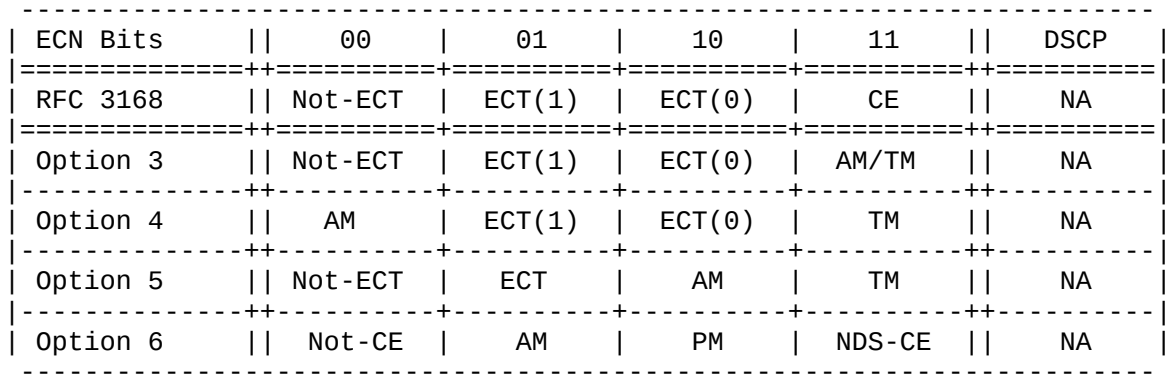
Option 2 in Figure 1 kept the "00" encoding for "Not-ECT". This allows Option 2 to be more compatible with the ECN encoding indicated in RFC 3168 [15], but sacrificed a valueable encoding. This requires the use of "11" encoding for both "AM" (Admission Mark) and "TM" (Termination Mark) states or requiring the allocation of a DSCP for encoding the "TM" state.

With the current PCN working environment of a single administrative domain and the use of diffserv for separation of PCN capable and none-PCN capable traffic, it is clear Option 1 is a better choice because it provides a needed valuable code point of "00". If ECN field code point syntax compatibility with RFC 3168 [15] is required, then Option 2 will be a better choice. But if code point syntax compatibility with RFC 3168 [15] is required for mixing of PCN and none-PCN traffic, then the concerns raised in RFC 4774 [20] will need to addressed differently.

Please notice neither Option 1 nor Option 2 provide encoding for the Affected Marking state, which is one deficiency of these two options and hence of using the combined DSCP and ECN Fields as the transport. Unless Affected Marking is somehow supported by the algorithms with another mean.

3.2. ECN Field as Encoding Transport

This section describes the encoding options that uses only the ECN field (without the DSCP field) available in the IP header of the data packets to encode the PCN states.



ECN Bits	00	01	10	11	DSCP
RFC 3168	Not-ECT	ECT(1)	ECT(0)	CE	NA
Option 3	Not-ECT	ECT(1)	ECT(0)	AM/TM	NA
Option 4	AM	ECT(1)	ECT(0)	TM	NA
Option 5	Not-ECT	ECT	AM	TM	NA
Option 6	Not-CE	AM	PM	NDS-CE	NA

Figure 2: Encoding of PCN Information Using ECN Field

In Figure 2, we listed the fundamental options when ECN field is used. Like in Figure 1, there are variations of the theme provided by these options. For example, the "01" and "10" encoding can be interpreted as ECT(A) and ECT(T) instead of just ECT(1) and ECT(0) respectively. Using the ECT(A) and ECT(T) variation provides the additional information of the ratio of packets AM marked to packets Not AM marked, and the ratio of packets TM marked to packets Not TM marked. Having these ratios being independent from one another.

3.2.1. Benefits of Using ECN Field

When the same treatment can be provided to both ECN and PCN traffic to achieve each of ECN and PCN purpose, then not having DiffServ as separation between ECN and PCN traffic may be a benefit. Under such circumstances, having the same encoding between ECN and PCN may be desireable (Option 3). But this can only be true if the requirement set forth in RFC 4774 [20] for alternate ECN semantics can be satisfied.

If the same treatment can be applied to both ECN and PCN traffic, then:

- The first issue of RFC 4774 [20]: "How routers know which ECN semantics to use with which packets." may be solved because there are no difference in the treatments of ECN and PCN packets, hence they can use the same semanics.

- The second and third issues of RFC 4774 [20]: "How does the possible presence of old routers affect the performance of the alternate ECN connections." and "How does the possible presence of old routers affect the coexistence of the alternate ECN traffic with competing traffic on the path." are also solved because there

are no difference in the treatment of ECN and PCN packets.

o The forth issue of RFC 4774 [20]: "How well does the alternate ECN traffic perform." are dependent on the algorithm used, and should be provided by the respective algorithm document, and not in the scope of this document.

3.2.2. Drawbacks of Using ECN Field

Notice this group of encoding options does not use DiffServ at all. Hence there are no separation of traffic based on their DSCP values and DiffServ Classes. With this group of encoding options, the required states of "PCN Capable Transport"/"None PCN Capable Transport" must be encoded using the ECN field. A bigger drawback is without the protection/separation capability provided by DiffServ, it is typically harder to satisfy the requirement set forth in RFC 4774 [20] for alternate ECN semantics.

3.3. DSCP Field as Encoding Transport

In this type of encoding and transport method the congestion and precongestion information is encoded into the 6 DSCP bits that are transported in the IP header of the data packets. Four possible alternatives can be distinguished, as can be seen in Figure 3. Options 7, 8 and 9 need three different DSCP values, while Option 10 needs four different DSCP values. Note that all DSCP values are representing and are associated with the same PHB. The 1st, 2nd and 3rd DSCP values are representing DSCP values that are assigned by IANA as DSCP experimental values, see RFC 2211 [8].

DSCP Bits	Original	Experimental 1	Experimental 2	Experimental 3
Option 7	Not-CE	PCN/AM/TM	NA	NA
Option 8	Not-CE	PCN/AM/TM	PCN/AFM/TM	NA
Option 9	Not-CE	PCN/AM	PCN/TM	NA
Option 10	Not-CE	PCN/AM	PCN/TM	PCN/AFM/TM

Figure 3: Encoding of PCN Information Using DSCP Field

3.3.1. Benefits of Using DSCP Field

The main benefit of using the DSCP field is that it is not affecting the end-to-end ECN semantics and therefore the issues and concerns raised in RFC 4774 [20] are not applicable for this encoding scheme. Another benefit is related to the fact that all 4 DSCP encoding options depicted in Figure 3 can support the not congested indication, PCN capable transport marking, the admission control and flow termination encoding states. In addition Option 8 and 10 can in addition support the ECMP solution.

3.3.2. Drawbacks of Using DSCP Field

This type of encoding needs to use per PHB, in addition to the original DSCP and depending on the encoding option used, one, two or three DSCP values, respectivelly. These additional DSCP values can be taken from the DSCP values that are not defined by standards action, see [8]. Note that all the DSCP values are representing and are associated with one PHB. Furthermore, if the separation between the PCN traffic and non- PCN traffic is required, then an additional DSCP or PHB value is needed for the "Not PCN-capable" encoding mode. The value of this DSCP/PHB can either follow a standards action or use a value that is applied for experimental or local use. It is important to note that the number of the DSCP values used for local or experimental use is restricted and therefore the number of different PHBs supported in the PCN domain will also be restricted.

Another drawback is related to the fact that the co-existence of the PCN and non-PCN traffic is not directly supported, but it can however, be realized by using in addition to the original DSCP value also experimental DSCP values, see RFC 2211 [8], to encode the different PCN encoding states.

3.3.3. Comparing DSCP Field Encoding Options

All four DSCP options can support the four basic encoding values, i.e., Not-CE, PCN, AM and TM encoding. Furthermore, Option 8 and 10 can support in addition to the four encoding values also the AFM encoding value. Option 7 needs 2 DSCP values and Option 9 needs three DSCP values to interpret the four basic encoding values. Option 8 needs three DSCP values and Option 10 needs four DSCP values to interpret the four basic encoding values and the AFM encoding value.

3.4. Out-of-Band Channel as Encoding Transport

In this type of encoding and transport method the congestion and precongestion information can be encoded using the IPFIX protocol RFC

3955 [18], that is normally used to carry flow-based IP traffic measurements from an observation point to a collecting point. Note that this encoding scheme is denoted in this document as "IPFIX channel". An observation point is a location in a network where IP packets can be observed and measured. A collecting point can be a process or a node that receives flow records from one or more observation points. In the PCN case, each PCN-interior-node will be an IPFIX observation point and the PCN-egress-node will be the IPFIX collecting point.

The PCN-interior-node will support the metering process and the flow records. Note that in this case each flow record can be associated with the record of the congestion and pre-congestion metering information associated with each PHB. The PCN-egress-node will then support the IPFIX collecting process, which will receive flow records from one or more congested and pre-congested PCN-interior-nodes. Using this encoding method the encoding modes/states can be aggregated and transported to the egress node by using the flow records at regular intervals or at the moment that a congestion and pre-congestion situation occurs. The used transport channel in this case is not the data path but a signaling protocol.

3.4.1. Benefits of Using Out-Of-Band Channel

This encoding scheme does not use the data path for encoding and transport, but it is able to transport the congestion and pre-congestion information associated with the encoding states by using a separate signaling channel. Another benefit of using this encoding scheme is that it is not affecting the end-to-end ECN semantics and therefore the issues and concerns raised in RFC 4774 are not applicable for this encoding scheme.

3.4.2. Drawbacks of Using Out-Of-Band Channel

The "IPFIX channel" encoding mode needs a separate signaling channel for the transport of the congestion and precongestion information from the PCN-interior-nodes towards the PCN-egress-node. The requirement of using an additional channel increases the complexity and influences negatively the performance of the PCN-interior-nodes since each PCN-interior-node needs to support in addition to the data path a separate channel.

4. Encoding Recommendations

To Be Filled In After PCN List Discussions.

5. Security Implications

Packets from normal precedence and higher precedence sessions [22] aren't distinguishable by PCN Interior Nodes. This prevents an attacker specifically targeting, in the data plane, higher precedence packets (perhaps for DoS or for eavesdropping). However, PCN End Nodes can access this information to help decide whether to admit or terminate a flow. The separation of network information provided by the Interior Nodes and the precedence information at the PCN End Nodes allows simpler, easier and better focused security enforcement.

PCN End Nodes police packets to ensure a flow sticks within its agreed limit. This is similar to the existing IntServ behaviour. Between them the PCN End Nodes must fully encircle the PCN-Region, otherwise packets could enter the PCN-Region without being subject to admission control, which would potentially destroy the QoS of existing flows.

It is assumed that all the Interior Nodes and PCN End Nodes run PCN and trust each other (ie the PCN-enabled Internet Region is a controlled environment). For instance a non-PCN router wouldn't be able to alert that it's suffering pre-congestion, which potentially would lead to too many calls being admitted (or too few being terminated). Worse, a rogue router could perform attacks such as marking all packets so that no flows were admitted.

So security requirements are focussed at specific parts of the PCN-Region:

The PCN End Nodes become the trust points. The degree of trust required depends on the kinds of decisions it has to make and the kinds of information it needs to make them. For example when the PCN End Node needs to know the contents of the sessions for making the decisions, when the contents are highly classified, the security requirements for the PCN End Nodes involved will also need to be high.

PCN-marking by the Interior Nodes along the packet forwarding path needs to be trusted, because the PCN End Nodes rely on this information.

6. IANA Considerations

To be completed.

7. Acknowledgements

To be completed.

Appendix A. Current PCN Detection, Marking and Transport Mechanisms

This appendix indicates the different available PCN based mechanisms that can be used for congestion and pre-congestion detection and marking used at interior nodes. The requirements and characteristics of such algorithms may influence the encoding and transport of the PCN encoding states.

Appendix A.1. Detection, Marking and Transport Mechanisms in CL-PHB

Please see draft-briscoe-tsvwg-cl-phb-03.txt [5] for details on the Controlled-Load PHB Algorithm.

Appendix A.2. Detection, Marking and Transport Mechanisms in Three State Marking

Please see draft-babiarz-pcn-3sm-01.txt [2] for details on the Three State Marking Algorithm.

Appendix A.3. Detection, Marking and Transport Mechanisms in Single Marking

Please see draft-charny-pcn-single-marking-03.txt [3] for details on the Single Marking Algorithm.

Appendix A.4. Detection, Marking and Transport Mechanisms in Load Control Marking

Please see draft-westberg-pcn-load-control-02.txt [4] for details on the Load Control Algorithm.

8. Informative References

[1] Eardley, P., "Pre-Congestion Notification Architecture", draft-ietf-pcn-architecture-01 (work in progress), October 2007.

[2] Babiarz, J., Liu, X., Chan, K., and M. Menth, "Three State PCN Marking", draft-babiarz-pcn-3sm-01 (work in progress), November 2007.

[3] Charny, A., Zhang, X., Faucheur, F., and V. Liatsos, "Pre-

Congestion Notification Using Single Marking for Admission and Termination", draft-charny-pcn-single-marking-03 (work in progress), November 2007.

[4] Westberg, L., "LC-PCN: The Load Control PCN Solution", draft-westberg-pcn-load-control-02 (work in progress), November 2007.

[5] Briscoe, B., "Pre-Congestion Notification marking", draft-briscoe-tsvwg-cl-phb-03 (work in progress), October 2006.

[6] Baker, F. and J. Polk, "MLEF Without Capacity Admission Does Not Satisfy MLPP Requirements", draft-ietf-tsvwg-mlef-concerns-00 (work in progress), February 2005.

[7] Braden, B., Clark, D., and S. Shenker, "Integrated Services in the Internet Architecture: an Overview", RFC 1633, June 1994.

[8] Wroclawski, J., "Specification of the Controlled-Load Network Element Service", RFC 2211, September 1997.

[9] Braden, B., Clark, D., Crowcroft, J., Davie, B., Deering, S., Estrin, D., Floyd, S., Jacobson, V., Minshall, G., Partridge, C., Peterson, L., Ramakrishnan, K., Shenker, S., Wroclawski, J., and L. Zhang, "Recommendations on Queue Management and Congestion Avoidance in the Internet", RFC 2309, April 1998.

[10] Nichols, K., Blake, S., Baker, F., and D. Black, "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers", RFC 2474, December 1998.

[11] Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z., and W. Weiss, "An Architecture for Differentiated Services", RFC 2475, December 1998.

[12] Heinanen, J., Baker, F., Weiss, W., and J. Wroclawski, "Assured Forwarding PHB Group", RFC 2597, June 1999.

[13] Awduche, D., Malcolm, J., Agogbua, J., O'Dell, M., and J. McManus, "Requirements for Traffic Engineering Over MPLS", RFC 2702, September 1999.

[14] Bernet, Y., Ford, P., Yavatkar, R., Baker, F., Zhang, L., Speer, M., Braden, R., Davie, B., Wroclawski, J., and E. Felstaine, "A Framework for Integrated Services Operation over Diffserv Networks", RFC 2998, November 2000.

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

[16] Davie, B., Charny, A., Bennet, J., Benson, K., Le Boudec, J., Courtney, W., Davari, S., Firoiu, V., and D. Stiliadis, "An Expedited Forwarding PHB (Per-Hop Behavior)", RFC 3246, March 2002.

[17] Charny, A., Bennet, J., Benson, K., Boudec, J., Chiu, A., Courtney, W., Davari, S., Firoiu, V., Kalmanek, C., and K. Ramakrishnan, "Supplemental Information for the New Definition of the EF PHB (Expedited Forwarding Per-Hop Behavior)", RFC 3247, March 2002.

[18] Leinen, S., "Evaluation of Candidate Protocols for IP Flow Information Export (IPFIX)", RFC 3955, October 2004.

[19] Babiarz, J., Chan, K., and F. Baker, "Configuration Guidelines for DiffServ Service Classes", RFC 4594, August 2006.

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

[21] "Supporting Real-Time Applications in an Integrated Services Packet Network: Architecture and Mechanisms", Proceedings of SIGCOMM '92 at Baltimore MD, August 1992.

[22] "Multilevel Precedence and Pre-emption Service (MLPP)", ITU-T Recommendation I.255.3, 1990.

[23] "Economics and Scalability of QoS Solutions", BT Technology Journal Vol 23 No 2, April 2005.

Authors' Addresses

Kwok Ho Chan
Nortel
600 Technology Park Drive
Billerica, MA 01821
USA

Email: khchan@nortel.com

Georgios Karagiannis
University of Twente
P.O. Box 217
7500 AE Enschede,
The Netherlands

Email: g.karagiannis@ewi.utwente.nl

Full Copyright Statement

Intellectual Property

Acknowledgment

Funding for the RFC Editor function is provided by the IETF Administrative Support Activity (IASA).