Applicability Statement for the Use of Pre-Congestion Notification in a Resource-Controlled Network
draft-tsou-pcn-racf-applic-01.txt

Status of this Memo

Abstract

This document is written to help coordinate work on pre-congestion notification (PCN) between the IETF PCN Working Group and the ITU-T. It maps the use of PCN into the ITU-T transport control architecture. It examines three scenarios, showing in each, where new requirements are placed on the ITU-T architecture. In each case, the ITU-T functional entity known as the Transport Resource Control Functional Entity (TRC-FE) is seen as the logical destination for PCN congestion reports and PCN flow termination reports, which it uses to keep track of network status. As logical entities, instances of the TRC-FE can be present in the ingress nodes, in one or more centralized devices, or in both. These alternatives define the scenarios examined.

Table of Contents

1. Terminology   TOC

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 RFC 2119.

2. Introduction   TOC

2.1. Pre-Congestion Notification (PCN)   TOC

The PCN Working Group is working on a new approach for maintenance of
quality of service (QoS) within Diffserv-controlled domains. This
approach is called "pre-congestion notification" (PCN). The principles
and associated architecture are documented in [I-D.PCNarch] (Eardley,
P., "Pre-Congestion Notification Architecture," October 2008.).
PCN distinguishes and assigns roles to ingress nodes, interior nodes,
and egress nodes relative to a given PCN domain. Ingress nodes mark

admitted packets to indicate that they should be PCN-metered. Interior nodes check the next-hop interface traffic status for each PCN-marked packet before routing it. Marking behaviour is described in detail in [I-D.PCNmark] (Eardley, P., "Marking behaviour of PCN-nodes," October 2008.) and depends in fact on the encoding scheme being used. The current view of this encoding scheme is provided in [I-D.PCNencod] (Moncaster, T., Briscoe, B., and M. Menth, "Baseline Encoding and Transport of Pre-Congestion Information," October 2008.), but other schemes are possible. In particular, note [I-D.PCN3in1] (Briscoe, B., "PCN 3-State Encoding Extension in a single DSCP," October 2008.). If three encoding states are available, it is possible to have the following marking behaviour:

*If the traffic status exceeds a lower "pre-congestion" threshold (but not the upper threshold described next), the packet is marked to indicate that it has encountered pre-congestion.

*If the traffic status exceeds an upper "termination" threshold, the packet is marked to indicate that it has encountered a flow termination condition.

The egress nodes relate the packets they receive to the aggregate flows they receive from individual ingress nodes. Statistics on packet marking are reported to a decision point (possibly within the egress node itself), which makes two decisions:

*whether one or more flows should be terminated immediately to preserve QoS for the remainder;

*whether new flows can be admitted without degrading the QoS for existing flows to an unacceptable level.

The architecture on which the IETF work has focussed assumes that egress nodes communicate directly to ingress nodes to effect the termination and admission decisions.

2.2. Admission Control System

A number of standards bodies (such as the ITU-T and ETSI TISPAN) have defined admission control systems (ITU-T RACF [Y.2111] (International Telecommunications Union, "Resource and admission control functions in Next Generation Networks," September 2006.), TISPAN RACS [ES283003] (ETSI Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN), "ETSI ES 282 003 V2.0.0, Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN); Resource and Admission Control Sub-System (RACS): Functional Architecture," May 2008.)) to allow per-flow

imposition of policy based on subscriptions across a variety of
transport technologies.
This section provides some general guidance how the admission control
system is used in the PCN environment.

2.2.1. Basic Flows of the Admission Control System

The admission control system generally performs both admission control
functions and flow termination functions based on the network
congestion status information collected from the PCN-egress-node.
The PCN-egress-node measures the traffic from a particular PCN-ingress-
node, and calculates the congestion value at the ingress-egress-
aggregate level. The PCN-egress-node may compare the current congestion
value with the previous congestion value, if the difference of the two
congestion value exceeds a preset range, the PCN-egress-node sends the
PCN-feedback-information (e.g. the current congestion value for a
particular pair of PCN-boundary-nodes) to the admission control system.
After receiving the PCN-feedback-information, the admission control
system saves the information for admission control and flow
termination. Based on this information, the admission control system
can determine whether to admit a service flow request or not after
receiving the flow request, and can also determine whether some of the
admitted flows need to be terminated. The admission control system also
needs to signal to the PCN-ingress-node the decision about admission or
termination.

2.2.2. Addressing

Generally, there is a mapping between the PCN-ingress-node and the
addresses which are outside the PCN domain and accessible by the PCN-
ingress-node, i.e. the source addresses of the incoming flows to the
PCN domain. The PCN-ingress-node can abstract this mapping from for
example the routing table or the configuration data. The PCN-ingress-
node can send the mapping to the admission control system and the
admission control system can determine the PCN-ingress-node for each
flow based on the mapping and the address information of the flow. The
admission control system may also forward the mapping to the PCN-
egress-node, so that the PCN-egress-node can determine the PCN-ingress-
node of a flow based on the mapping and the address information of the
flow.

2.3. Combining the PCN and ITU-T RACF Architectures

The ITU-T RACF architecture is a generalization of the architecture
originally developed by 3GPP for cellular radio networks connected to
an IP core. The architecture features a top-level control function
which transmits requests (push mode) or provides policy responses (pull
mode) to edge nodes where enforcement takes place. This top-level
control function provides an abstract view of network infrastructure to
the session control layer (SIP servers and the like). A second-level
control function is aware of the specific transport technology in use
in the network, tracks network topology and resource availability, and
indicates to the top-level function whether resources are available to
serve a specific flow request.
In the ITU-T architecture in particular [Y.2111] (International
Telecommunications Union, "Resource and admission control functions in
Next Generation Networks," September 2006.), the top-level function is
called the Policy Decision Functional Entity (PD-FE-FE), and the
second-level function is called the Transport Resource Control
Functional Entity (TRC-FE). The complete transport control subsystem is
called the Resource and Admission Control Function (RACF). The function
within the edge node at which decisions are enforced is called the
Policy Enforcement Functional Entity (PE-FE). Interested parties will
find a complete description of the first release of this architecture
in [ITU-T Y.2111] , but a partial diagram labelling the reference
points of interest in this document is shown in Figure 1 (A Portion of
the ITU-T-Defined Transport Control Architecture).

```
                       Session Control

                   _____
                          |
                          + Rs
                     -----+-----
                     |  PD-FE  |------
                     -----+-----     |
                          |          |
                          + Rt       |
     -----------   Rp    -----+-----     + Rw
     |  TRC-FE |---+----|  TRC-FE |     |
     -----+-----         -----+-----     |
          + Rc                + Rc       |
     ------+----------------+---------|-------
     |                          -----+----- |
     |        . . . .           |  PE-FE  | |
     |     Transport Processing  ----------- |
     -----------------------------------------
```
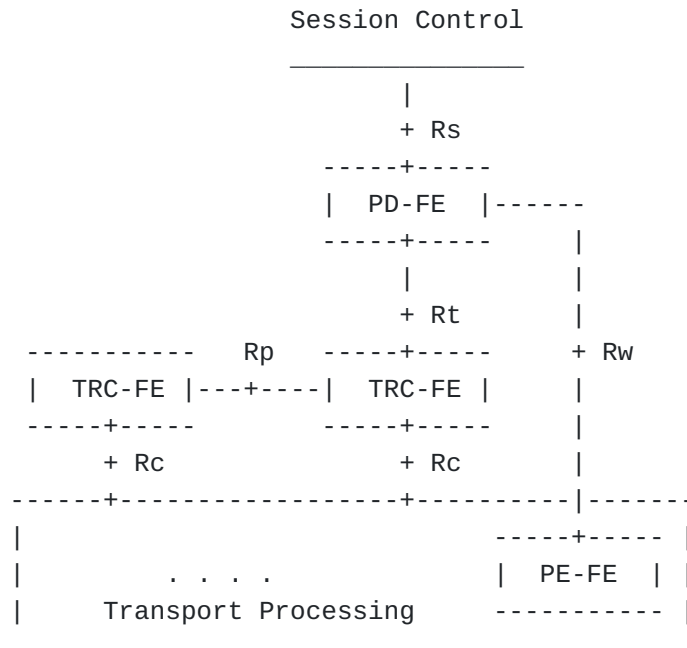
Figure 1: A Portion of the ITU-T-Defined Transport Control Architecture

---

The typical message flow associated with a flow admission request in
the ITU-T architecture is as follows:

1. In the general case, the TRC-FE gathers topology and status
    information from the transport processing layer across the Rc
    reference point. The use of PCN offers the possibility that the
    TRC-FE no longer has to deal with detailed topology.

2. The PD-FE receives a request to admit a flow. In push mode this
    comes from a P-CSCF, for example, across the Rs reference
    point; in pull mode it comes from the PE-FE across the Rw
    reference point.

3. The PD-FE checks subscription data (provided by another
    subsystem) to see if the subscriber is permitted to add the
    requested flow. If not, it rejects the request immediately.

4. The PD-FE contacts an instance of the TRC-FE by way of the Rt
    reference point to determine if the network resources are
    available to allocate to the requested flow. The selected
    instance may locate other TRC-FE instances along the flow path
    via the Rp reference point to make this determination.

5. If the response from the TRC-FE is negative, the PD-FE rejects the request. Otherwise it downloads the required policy to the PE-FE across the Rw reference point. In push mode it also responds positively to the original request across the Rs reference point.

The PD-FE and TRC-FE actually track sessions, which typically consist of multiple flows in both directions. The TRC-FE can tell the PD-FE to terminate a flow or to abort a complete session due to, for instance, loss of the supporting resources. The PD-FE from its side can add or terminate individual flows or terminate a complete session as required by the user.
This document considers how PCN would interact with the RACF architecture just described. Since that architecture is functional, various implementations are possible, depending on what elements are combined in the same physical entities. This document looks at three possible deployment scenarios. In all three cases, the PD-FE is implemented in a centralized device and the PE-FE is a functional component of the ingress nodes. The scenarios vary in where the TRC-FE is implemented.
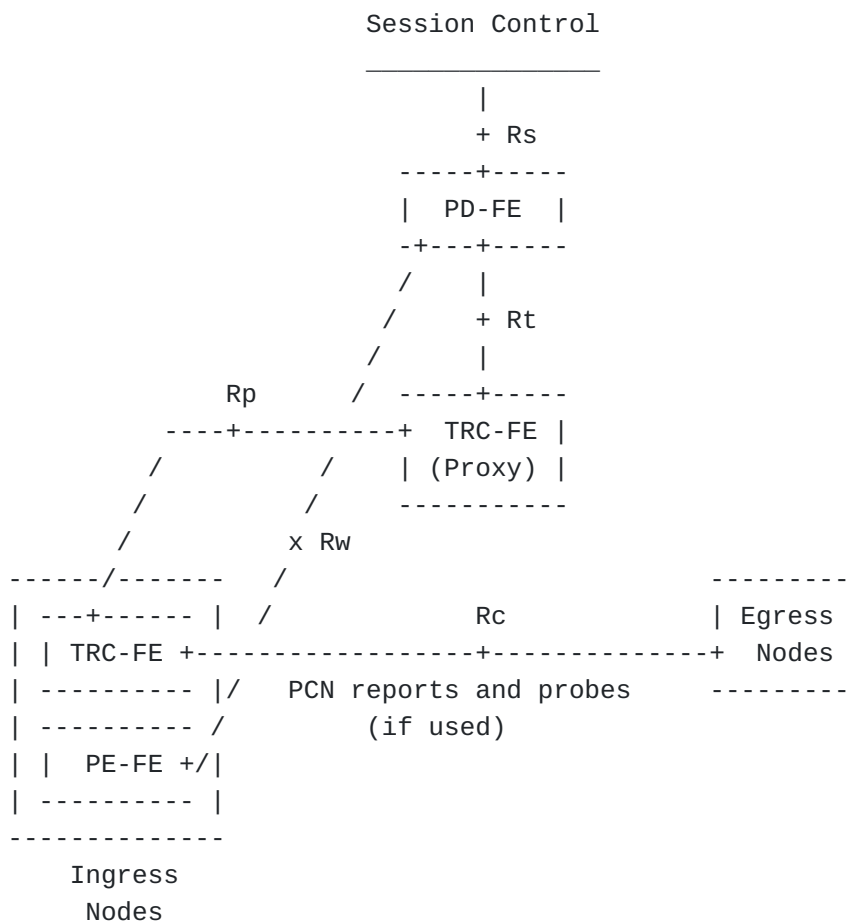
1. The TRC-FE is implemented both in one or more centralized devices (possibly combined with the PD-FE) and as a logical component of the ingress nodes. Congestion level and flow termination reports from the egress nodes pass directly to the ingress nodes. The TRC-FE in the centralized device is just a proxy that forwards requests and responses between the PD-FE and the ingress nodes. The TRC-FE instances in the ingress nodes respond to flow admission requests from the PD-FE with indications of resource status derived from the PCN reports. They process flow termination reports based on the priority parameters for existing flows that were supplied by the PD-FE during resource reservation, and send reports to the PD-FE indicating which flows or sessions must be terminated to protect the QoS of the remaining flows

2. The TRC-FE is absorbed into the ingress nodes and no centralized instance exists. Congestion level and flow termination reports from the egress nodes pass directly to the ingress nodes. The TRC-FE instances in the ingress nodes respond to flow admission requests and process flow termination reports in the same way as in Scenario 1.

3. The TRC-FE is implemented in one or more centralized devices only (possibly combined with the PD-FE). It receives and processes the congestion level and flow termination reports and responds to flow admission requests in the same way as the distributed TRC-FE instances in the previous two scenarios.

The following sections examine each of these scenarios in turn. Amongst
other things, they will consider whether the message flow described in
the previous section is appropriate or whether it should be modified.

---

3. Scenario 1: Proxy and Ingress TRC-FE Instances

The functional layout for Scenario 1 is shown in Figure 2 (Functional
Architecture For Scenario 1) below. The PCN reports pass across the
existing Rc reference point between the embedded TRC-FE and the
transport processing layer. If probes are used, they also cross the Rc
reference point. The figure assumes that the PD-FE uses the proxy TRC-
FE to locate the ingress node for a given user flow, and places the Rt
and Rp reference points accordingly.

```
                          Session Control
                        _____
                                |
                                + Rs
                          -----+-----
                          |  PD-FE  |
                          -+---+-----
                          /    |
                         /     + Rt
                        /      |
            Rp         /  -----+-----
          ----+----------+  TRC-FE |
         /           /    | (Proxy) |
        /           /     ----------
       /           x Rw
  ------/-------   /                            ----------
  | ---+------ |  /              Rc             | Egress |
  | | TRC-FE +-----------------+--------------+  Nodes |
  | ---------- |/   PCN reports and probes     ----------
  | ---------- /          (if used)
  | |  PE-FE +/|
  | ---------- |
  --------------
       Ingress
        Nodes
```

Figure 2: Functional Architecture For Scenario 1

If we follow through the message flows specified in [Section 2.2 (Admission Control System)](#), we find that they now run as follows:

1. Network status, but not network topology, is now determined from the PCN report message flows across Rc, as already noted.

2. Step 2 is unchanged.

3. Step 3 is unchanged.

4. Step 4 involves messaging through the Rt reference point to the proxy TRC-FE instance, and from there through the Rp reference point to the TRC-FE instance embedded in the desired ingress node. If probing is used to determine congestion status, the probing messages are another increment to the messaging through the Rc reference point. The responses to the probes pass back through Rc.

5. Step 5 is unchanged: a request from the PD-FE directly to the PE-FE instance in the ingress node, across Rw.

Note that the messaging flow appears to be less than optimal, in that the PD-FE is really communicating twice with the ingress node - once to the embedded TRC-FE instance in step 4, then, based on a positive response, to the PE-FE instance in step 5. Messaging could be reduced to one exchange if the PD-FE were aware in advance that it was communicating with a combined TRC-FE + PE-FE entity. It is an open issue whether such a special-case modification of the general control architecture should be developed.
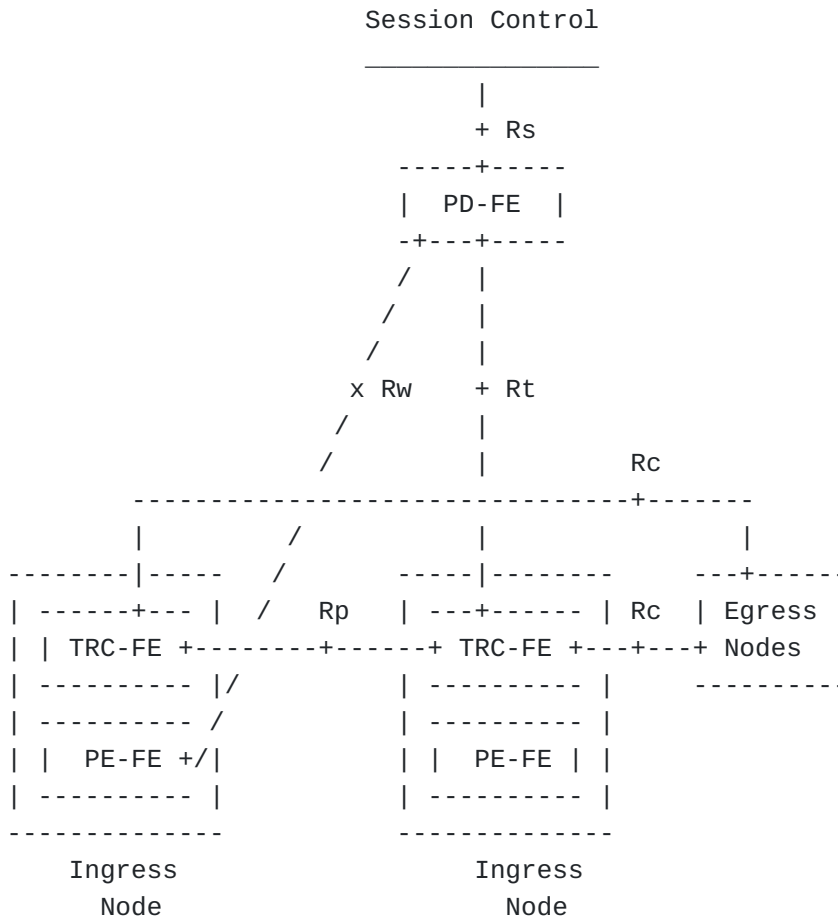
---

4. **Scenario 2: TRC-FE Present In Ingress Nodes Only**

The functional architecture corresponding to this scenario is shown in [Figure 3 (Functional Architecture For Scenario 2)](#). The figure assumes that the PD-FE does not know which ingress node to contact until the TRC-FE instance it reached in the first place locates it. It is plausible that the TRC-FE instance has the necessary knowledge simply because ingress nodes tend also to be egress nodes, and in this scenario the egress nodes have to know where to send their PCN reports. As an alternative, the PD-FE could be required to have the necessary knowledge to locate the right ingress node directly. However, this seems contrary to the spirit of the existing ITU-T architecture.

```
                         Session Control
                       _____
                              |
                              + Rs
                         -----+-----
                         |  PD-FE  |
                         -+---+-----
                         /    |
                        /     |
                       /      |
                      x Rw    + Rt
                     /        |
                    /         |          Rc
          --------------------------------+-------
          |          /           |              |
     --------|-----    /      -----|--------     ---+------
     | ------+--- |  /   Rp   | ---+------ | Rc  | Egress |
     | | TRC-FE +--------+------+ TRC-FE +---+---+ Nodes  |
     | ---------- |/         | --------- |     ----------
     | ---------- /          | --------- |
     | |  PE-FE +/|          | |  PE-FE | |
     | ---------- |          | --------- |
     --------------          --------------
          Ingress                Ingress
           Node                   Node
```

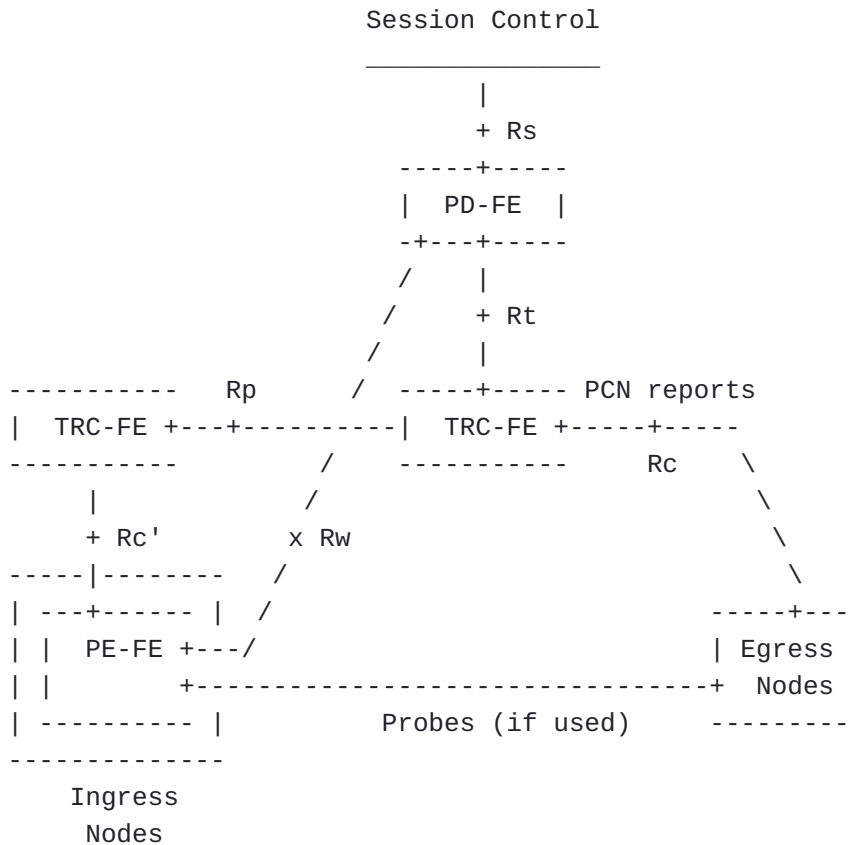**Figure 3: Functional Architecture For Scenario 2**

Aside from the routing of messages between the PD-FE and the TRC-FE, this scenario is the same as Scenario 1.

---

5. Scenario 3: Centralized TRC-FE Only

In this scenario centralized TRC-FE instances collect PCN reports from egress nodes within their scope of responsibility. As in the other scenarios, the TRC-FE instances track resource status, respond to inquiries from the PD-FE, and notify the PD-FE if flows have to be terminated. However, since they are physically separate from the ingress nodes, there is no possibility of message optimization. It is assumed in the figure that a given egress node sends all of its PCN reports to the same TRC-FE instance, which must then distribute them to the TRC-FE instances in charge of the various ingress nodes. Thus the

PCN reports appear as incremental information flows Rc' and Rp' across
the Rc and Rp reference points respectively.

```
                        Session Control
                       _____
                              |
                            + Rs
                       -----+-----
                       |  PD-FE  |
                       -+---+-----
                       /    |
                      /     + Rt
                     /      |
    -----------   Rp    /  -----+----- PCN reports
    |  TRC-FE +---+---------|  TRC-FE +-----+-----
    -----------         /   -----------      Rc    \
        |              /                             \
       + Rc'         x Rw                             \
    -----|--------    /                                \
    | ---+------ |  /                          -----+----
    | | PE-FE +---/                            | Egress |
    | |       +-------------------------------+  Nodes |
    | ---------- |        Probes (if used)     ----------
    --------------
        Ingress
         Nodes
```

Figure 4: Functional Architecture For Scenario 3

The first thing to note in Figure 4 (Functional Architecture For
Scenario 3) is that the centralized TRC-FE reduces the scale of a
problem already encountered in PCN operation: where to send the PCN
reports? In the straightforward case initially dealt with by the PCN
Working Group, each egress node must know to which ingress node to send
each of its flow reports. Now the centralized TRC-FE has that problem,
but the egress nodes only have to discover or be configured with the
address of a single collection point.
The information exchanges needed to admit a flow are pretty well those
described in Section 2.2 (Admission Control System). This scenario has
one interesting point not seen in the other scenarios. If probes are
used, they must be sent from the ingress nodes rather than the TRC-FE
instances. This raises the question of how they are triggered. It seems
inappropriate to require the PD-FE to be aware of the use of PCN in the

domain, since it is supposed to operate in technology-independent fashion. It makes more sense for the TRC-FE to trigger a probe at the PE-FE in response to a query from the PD-FE. In terms of the ITU-T architecture, the trigger and response from the PE-FE would pass via the Rc reference point, but this is shown as Rc' in the figure because it represents an incremental requirement.

6. Security Considerations

The security considerations for the different scenarios described in this document are not fundamentally different from those already applying for PCN and for the ITU-T architecture respectively. A more detailed analysis of specific issues may be appropriate as this document is further developed.

7. IANA Considerations

This memo presents no IANA considerations.

8. Acknowledgements

Thanks to Gabriele Corliano for ideas contributed in preliminary discussion.

9. References

9.1. Normative References

[I-D.PCNarch]	Eardley, P., "Pre-Congestion Notification Architecture," October 2008.
[I-D.PCNencod]	Moncaster, T., Briscoe, B., and M. Menth, "Baseline Encoding and Transport of Pre-Congestion Information," October 2008.
[I-D.PCNmark]	Eardley, P., "Marking behaviour of PCN-nodes," October 2008.
[RFC2119]	

	Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels," BCP 14, RFC 2119, March 1997 (TXT, HTML, XML).
[Y.2111]	International Telecommunications Union, "Resource and admission control functions in Next Generation Networks," September 2006.

9.2. Informative References

[ES283003]	ETSI Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN), "ETSI ES 282 003 V2.0.0, Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN); Resource and Admission Control Sub-System (RACS): Functional Architecture," May 2008.
[I-D.PCN3in1]	Briscoe, B., "PCN 3-State Encoding Extension in a single DSCP," October 2008.

Authors' Addresses

	Tina Tsou
	Huawei Technologies
	F3-5-089S, R&D Center,
	Longgang District
	Shenzhen 518129
	China
Email:	tena@huawei.com
	Fortune Huang
	Huawei Technologies
	F3-5-089S, R&D Center,
	Longgang District
	Shenzhen 518129
	China
Email:	fqhuang@huawei.com
	Tom Taylor
	Huawei Technologies
	1852 Lorraine Ave
	Ottawa, Ontario K1H 6Z8
	Canada
Phone:	+1 613 680 2675
Email:	tom.taylor@rogers.com

Full Copyright Statement

Intellectual Property