

Internet Engineering Task Force
Internet-Draft
Intended status: Standards Track
Expires: May 11, 2013

Q. Sun
China Telecom
M. Boucadair
X. Deng
France Telecom
C. Zhou
Huawei Technologies
T. Tsou
Huawei Technologies (USA)
S. Perreault
Viagenie
November 7, 2012

**Using PCP To Coordinate Between the CGN and Home Gateway
draft-tsou-pcp-natcoord-09**

Abstract

This document defines an extension to the base PCP. New OpCode is defined to enhance PCP with the ability to reserve port sets for internal hosts.

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of [BCP 78](#) and [BCP 79](#).

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on May 11, 2013.

Copyright Notice

Copyright (c) 2012 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to [BCP 78](#) and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of

publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1.	Application Scenario	3
2.	MAP_PORT_SET Opcode	3
2.1.	MAP_PORT_SET Operation Packet Formats	4
2.2.	MAP_PORT_SET Mapping Table Example	8
3.	MAP_PORT_SET Operation	8
3.1.	Generating a MAP_PORT_SET Request	8
3.2.	Renewing a MAP_PORT_SET Mapping	8
3.3.	Processing a MAP_PORT_SET Request	9
3.4.	Processing a MAP_PORT_SET Response	10
4.	Mapping Lifetime and Deletion	10
5.	PREFER_FAILURE Option for MAP_PORT_SET Opcode	10
6.	Coexistence with MAP OpCode	10
7.	MAP_PORT_SET Failover	11
8.	Security Considerations	11
9.	IANA Considerations	11
10.	Authors List	11
11.	Acknowledgements	12
12.	References	12
12.1.	Normative References	12
12.2.	informative References	13
	Authors' Addresses	13

1. Application Scenario

PCP can be used to control an upstream device to achieve the following goals:

1. A plain IP address (i.e., a non-shared) can be assigned to a given subscriber because it subscribed to a service which uses a protocol that don't embed a transport number or because the NAT is the only deployed platform to manage IP addresses.
2. An application (e.g., sensor) does not need to listen to a whole range of ports available on a given IP address. Only a limited set of ports are used to bind its running services. For such devices, the external port(s) and IP address can be delegated to that application and therefore avoid enforcing NAT in the network side for its associated flows. The NAT in the PCP- controlled device should be bypassed.
3. A device able to restrict its source ports can be delegated an external port restricted IP address. The PCP-controlled device should be instructed to by-pass the NAT when handling flows destined/issued to that device.

This document extends PCP with the ability to reserve port sets instead of individual ports. This is motivated by the need to offload to a port-restricted device in lightweight 4over6 [[I-D.cui-softwire-b4-translated-ds-lite](#)], reduce the logging and enhance the performance of the CGN.

A candidate solution is to define a new Option to request for this feature be enforced by the PCP-controlled device. Nevertheless, this solution is not efficient when large port sets are assigned (e.g., address sharing ratio of 1:2 or 1:8). Another issue, is when no NAT is enforced in the PCP-controlled device but only a Port Range Router (PRR) function, the request has not to indicate the internal ports.

For those reasons, a new PCP OpCode is defined in this document.

2. MAP_PORT_SET Opcode

This section defines a new Opcode to request a port set from a PCP-controlled device.

The format of MAP_PORT_SET is designed to be close to the MAP message format. The port set is encoded using a port mask to convey a contiguous port range.

By analogy, a port set binding can be seen as an aggregate of MAP mappings. When assigning a port set to a PCP Client, the PCP-controlled device maintains a binding between the source IP address of the PCP request, the assigned external IP address and the assigned port set. Allocating port sets can greatly reduce individual MAP requests for a PCP client when requesting a bulk of ports at one time. This mechanism can be applied for lightweight 4over6 [[I-D.cui-softwire-b4-translated-ds-lite](#)] in port-set allocation process. It can also be applied to stateless PCP-controlled device, in which the Internal address, External address and Port set is determined algorithmically.

MAP_PORT_SET: Create an explicit dynamic mapping between an Internal IP Address and an External IP Address + Port set

It is totally up to the PCP server to determine the port-set quota for each PCP client. In addition, when the PCP-controlled device supports multiple port-sets delegation for a given PCP client, the PCP client MAY re-initiate a PCP request to get another port set when it has exhausted all the ports within the port-set.

PCP-controlled device SHOULD provide a configuration option to allow administrators to configure the size of each individual port set (denoted as MAX_REQUEST_QUOTA) to be assigned and the size of the total ports for a PCP client (denoted as MAX_USER_QUOTA).

2.1. MAP_PORT_SET Operation Packet Formats

The MAP_PORT_SET Opcode has a similar packet layout for both requests and response. Figure 1 shows the format of the MAP_PORT_SET request.

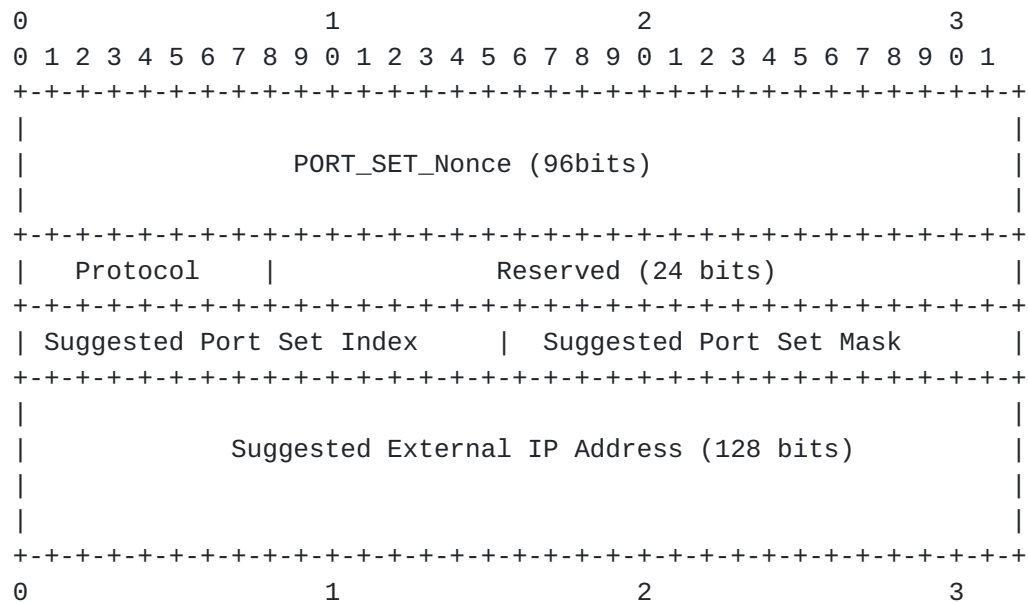


Figure 1: MAP_PORT_SET Opcode Request format

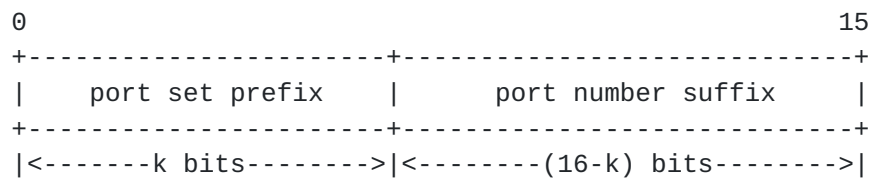
These fields are described below:

- o Requested lifetime (in common header): Requested lifetime of this port set mapping, in seconds. The value 0 indicates "delete".
- o PORT_SET_Nonce: Random value chosen by the PCP client which SHOULD be different for individual PCP requests. But the same value MUST be kept in one request re-transmission. See Section 11.2 of [\[I-D.ietf-pcp-base\]](#).
- o Protocol: the default value is zero (to indicate all transport protocols).
- o Reserved bits: 24 bits MUST be set to 0.
- o Suggested Port Set Index (PSI): The PSI indicates the value of the significant bits of the Port Mask. By default, PSI is set to 0 in a request. It can also convey Suggested Port Set Index if the client has a hint on it. The first k bits on the left of the 2-octet field is the Port Set Index value, with the rest of the field right padding zeros.
- o Suggested Port Set Mask (PSM): The PSM indicates the position of the bits that are used to build the Port Set Index. The 1 values in the Port Set Mask indicate by their position the significant bits of the Port Set Value. By default, PSM is set to 0 in a request. It can also convey Suggested Port Set Mask if the client

has a hint on it. The first k bits on the left is padding ones while the remained $(16-k)$ bits of the 2-octet field on the right is padding zeros.

- o Suggested External IP Address: Suggested external IPv4 or IPv6 address. Same as Section 10.1 of [[I-D.ietf-pcp-base](#)].

In the context of Port Set Option, the port number should consist of port set prefix and port number suffix. The port set prefix can be got from Port Set Index and Port Set Mask, while port number suffix can change continuously. The format of port number is shown below.



In order to exclude the system ports ([[I-D.ietf-tsvwg-iana-ports](#)]) or ports saved by SPs, the former port-sets that contains well-known ports SHOULD NOT be assigned.

Figure 2 shows the format of Opcode-specific information in a response packet for the MAP_PORT_SET Opcode:

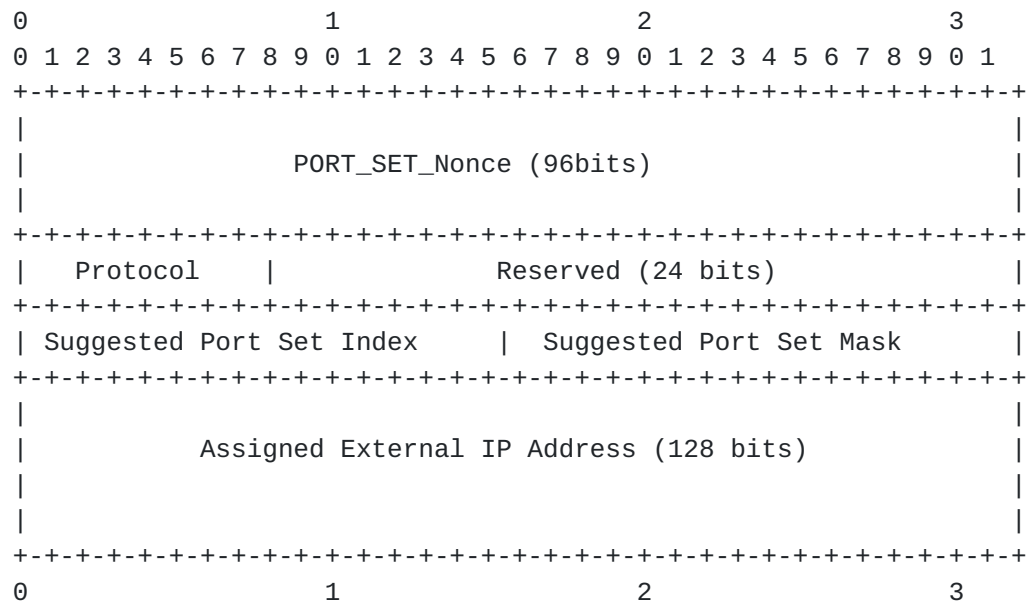


Figure 2: MAP_PORT_SET Opcode format of Response

These fields are described below:

- o Lifetime (in common header): On an error response, this indicates how long clients should assume they'll get the same error response from the PCP server if they repeat the same request. On a success response, this indicates the lifetime for this mapping, in seconds.
- o PORT_SET_Nonce: MUST be copied from the request.
- o Protocol: MUST be copied from the request.
- o Reserved bits: 16 bits MUST be set to 0.
- o Assigned Port Set Index (PSI): The PSI indicates the value of the significant bits of the Port Mask.
- o Assigned Port Set Mask (PSM): The Port Set Mask indicates the position of the bits that are used to build the Port Set Index. The 1 values in the Port Set Mask indicate by their position the significant bits of the Port Range Value.
- o Assigned External IP Address (128 bits): This field conveys the assigned external IPv4 (encoded using IPv4-mapped IPv6 address) or IPv6 address for the mapping. On an error response, the Assigned External IP Address is copied from the request.

2.2. MAP_PORT_SET Mapping Table Example

The following table depicts an example of the mapping table in the PCP Server enabling MAP_PORT_SET OpCode.

Internal Address	External Address	Port Range	Protocol	NONCE
2001:db8::1	192.0.2.33	5120-6143	0	nonce1
2001:db8::2	192.0.2.33	6144-7167	0	nonce2
2001:db8::3	192.0.2.33	7168-8191	0	nonce3
2001:db8::4	192.0.2.33	8192-9215	0	nonce4

Figure 3: Mapping table example in MAP_PORT_SET

3. MAP_PORT_SET Operation

3.1. Generating a MAP_PORT_SET Request

The MAP_PORT_SET request MUST contain values in the Suggested IP Address field, Suggested Port Set Index and Suggested Port Mask. However, this port set indicated in the request of the PCP Client is only a hint; it is up to the PCP Server to assign a port set.

If a PCP Client fails to receive an expected response from a server, the PCP client follows the same retransmission procedure defined for MAP in the base PCP specification (section 8.1.1 of [\[I-D.ietf-pcp-base\]](#)). The PORT_SET_Nonce should be copied from the previous MAP_PORT_SET request.

If a PCP Client uses out all the ports in the current assigned port set, it MAY generate a new MAP_PORT_SET Request to get another delegated port-set. The Client MUST use a different Mapping Nonce for different MAP_PORT_SET mappings. If USER_EX_QUOTA error is received from the server, the PCP client SHOULD NOT request for another new port set.

3.2. Renewing a MAP_PORT_SET Mapping

Port Set mapping renewal for MAP_PORT_SET MUST follow the same procedure for an individual MAP mapping (section 11.2.1 of [\[I-D.ietf-pcp-base\]](#)) except for considerations related to the internal port (which is included in a MAP request but not present in a MAP_PORT_SET).

The MAP_PORT_SET request MUST include the currently assigned IP address and port-set in the Suggested IP address, Suggested Port Set Index and Suggested Port Set Mask.

3.3. Processing a MAP_PORT_SET Request

The PCP server SHOULD take exactly the same order as in (section 11.3 of [[I-D.ietf-pcp-base](#)]). In particular, as there is no Internal Port in MAP_PORT_SET anymore, all the processes regarding to Internal Port should be neglected accordingly.

The PCP Server uses the Internal address as an index to lookup the Mapping table. The PCP Server SHOULD ensure the port sets allocated to different PCP Clients are non-overlap with each other. It SHOULD only assign individual for each MAP_PORT_SET request, rather than an aggregated one.

Considerations related to the assignment of the external IP Address are the same as what is defined in (section 11.3 of [[I-D.ietf-pcp-base](#)]).

The procedures regarding to the port set are similar to the external port processes in MAP Opcode (section 11.3 of [[I-D.ietf-pcp-base](#)]), except that the whole port-set should be treated consistently in MAP_PORT_SET Opcode. The same operations for handling the Suggested external port for a MAP request are applied on the Suggested Port Set.

The procedures for PORT_SET_Nonce is exactly the same as the Mapping Nonce field defined in (section 11.3 of [[I-D.ietf-pcp-base](#)]). The PCP server only needs to remember ONE PORT_SET_Nonce for each mapping (Internal IP Address, External IP address and Port Set).

The error codes in MAP_PORT_SET Response mainly have the following possibilities:

- o If the PCP server or PCP-controlled device does not support MAP_PORT_SET Opcode, the error UNSUPP_OPCODE MUST be returned.
- o If an option does not make sense, (e.g., the PREFER_FAILURE Option is included in a request with lifetime=0, etc.), the request is invalid and generates a MALFORMED_OPTION error. This procedure is the same with section 10.3 of [[I-D.ietf-pcp-base](#)].

If the requested lifetime is zero, it indicates a request to delete an existing mapping.

A PCP server SHOULD maintain MAX_USER_QUOTA and MAX_REQUEST_QUOTA.

MAX_USER_QUOTA is to indicate the maximum number of ports a subscriber may get in total, and MAX_REQUEST_QUOTA is to indicate the maximum number of ports in each request. Therefore, one PCP Client will have up to N mappings, in which N SHOULD NOT be larger than $\text{floor}(\text{MAX_USER_QUOTA}/\text{MAX_REQUEST_QUOTA})$. The specific mechanism to configure the quotas is out of scope.

If the PCP server is configured to allocate multiple port-set allocation for one subscriber, the same External address SHOULD be assigned to one subscriber in multiple port-set requests to guarantee the consistency.

To optimize the number of mapping entries maintained by the PCP server, it is RECOMMENDED to configure the server to assign the maximum allowed port set in a single response. This policy SHOULD be configurable.

When MAP_PORT_SET is applied to stateless PCP-controlled device, the PCP server returns an answer indicating the external IP address and port-set as seen by remote peers.

3.4. Processing a MAP_PORT_SET Response

On receiving a MAP_PORT_SET Response, the same procedure as the one for individual mapping [section 10.4 of [\[I-D.ietf-pcp-base\]](#)] MUST be followed by the PCP Client to validate the response (except the considerations related to the internal port).

4. Mapping Lifetime and Deletion

The procedure for port-set mapping lifetime and deletion is also the same with individual mapping [section 10.5 of [\[I-D.ietf-pcp-base\]](#)].

5. PREFER_FAILURE Option for MAP_PORT_SET Opcode

This option [section 10.2 of [\[I-D.ietf-pcp-base\]](#)] can be applied to MAP_PORT_SET Opcode indicating that if the PCP server cannot map the suggested External Address and port-set, the PCP server should not create a mapping.

6. Coexistence with MAP OpCode

Normally, the PCP server for MAP_PORT_SET will not run NAT. So there is no NAT binding in PCP and the PCP server will not run MAP OpCode for the same subscriber. In the case when the PCP client is embedded

in the host and the PCP server keeps the NAT bindings for some special-purpose applications, the external address and the port allocated to the subscriber should be consistent with the ones in MAP_PORT_SET response.

7. MAP_PORT_SET Failover

The failover mechanism in MAP [section 14 in [[I-D.ietf-pcp-base](#)]] and [[I-D.boucadair-pcp-failure](#)] can also be applied to MAP_PORT_SET.

The only difference compared to MAP is the amount of Mapping entries in MAP_PORT_SET PCP server is much less than MAP. Therefore, the cost of state synchronization has been greatly reduced in MAP_PORT_SET.

8. Security Considerations

The same security considerations discussed in [[I-D.ietf-pcp-base](#)] have to be taken into account.

9. IANA Considerations

The authors request the following new OpCode: MAP_PORT_SET

10. Authors List

The following are extended authors who contributed to the effort:

Yunqing Chen

China Telecom

Room 502, No.118, Xizhimennei Street

Beijing 100035

P.R.China

Chongfeng Xie

China Telecom

Room 502, No.118, Xizhimennei Street

Beijing 100035

P.R.China

Yong Cui

Tsinghua University

Beijing 100084

P.R.China

Phone: +86-10-62603059

Email: yong@csnet1.cs.tsinghua.edu.cn

Qi Sun

Tsinghua University

Beijing 100084

P.R.China

Phone: +86-10-62785822

Email: sunqibupt@gmail.com

Gabor Bajko

Nokia

Email: gabor.bajko@nokia.com

11. Acknowledgements

The authors would like to show sincerely appreciation to Dan Wing, Simon Perreault, Thmoas and Yoshihiro Ohba for their useful comments and suggestions.

12. References

12.1. Normative References

[I-D.ietf-pcp-base]
Wing, D., "Port Control Protocol (PCP)", October 2012.

12.2. informative References

- [I-D.boucadair-pcp-failure]
Boucadair, M., Dupont, F., and R. Penno, "Port Control Protocol (PCP) Failure Scenarios", August 2012.
- [I-D.cui-software-b4-translated-ds-lite]
Cui, Y., Sun, Q., Boucadair, M., Tsou, T., and Y. Lee, "Lightweight 4over6: An Extension to DS-Lite Architecture", Feb 2012.

Authors' Addresses

Qiong Sun
China Telecom
P.R.China

Phone: 86 10 58552936
Email: sunqiong@ctbri.com.cn

Mohamed Boucadair
France Telecom
Rennes, 35000
France

Email: mohamed.boucadair@orange.com

Xiaohong Deng
France Telecom

Email: xiaohong.deng@orange-ftgroup.com

Cathy Zhou
Huawei Technologies
Bantian, Longgang District
Shenzhen 518129
P.R. China

Phone:
Email: cathy.zhou@huawei.com

Tina Tsou
Huawei Technologies (USA)
2330 Central Expressway
Santa Clara, CA 95050
USA

Phone: +1 408 330 4424
Email: Tina.Tsou.Zouting@huawei.com

Simon Perreault
Viagenie
246 Aberdeen
Quebec, QC G1R 2E1
Canada

Phone: +1 418 656 9254
Email: simon.perreault@viagenie.ca

