Internet Engineering Task Force
Internet-Draft
Intended status: Standards Track
Expires: January 17, 2013

Q. Sun
China Telecom
M. Boucadair
X. Deng
France Telecom
C. Zhou
Huawei Technologies
T. Tsou
Huawei Technologies (USA)
July 16, 2012

Lightweight 4over6 Port-set Allocation: Using PCP To Coordinate Between the CGN and Home Gateway

draft-tsou-pcp-natcoord-07

Abstract

This document defines an extension to the base PCP. New OpCode and Options are 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 January 17, 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. Port-Set Options Formats 6
2.2.1. Port_Range_Option 6
2.2.2. Cryptographically_Random_Port_Range_Option 7
2.3. Generating a MAP_PORT_SET Request 8
2.4. Renewing a MAP_PORT_SET Mapping 8
2.5. Processing a MAP_PORT_SET Request 8
2.6. Processing a MAP_PORT_SET Response 11
2.7. Mapping Lifetime and Deletion 11
2.8. PREFER_FAILURE Option for MAP_PORT_SET Opcode 11
3. Security Considerations 11
4. IANA Considerations 11
5. Author List 11
6. Contributor List 12
7. References 12
7.1. Normative References 12
7.2. 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 (i.e., a non-shared) IP address can be assigned to a given subscriber because the subscriber 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 set instead of individual mapping. 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 new PCP OpCode and two new PCP Options are defined in this document.

2. MAP_PORT_SET Opcode

This section defines a new Opcode which requests port set from a PCP-controlled device to a PCP client. 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 port set. It 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.

MAP_PORT_SET: Create an explicit dynamic mapping between an

Internel's IP Address and an External Address + Port set

The format of a port-set can either be contiguous or non-contiguous including a cryptographical assigned port set. The contiguous port-set is simple but since the port space for a subscriber shrinks significantly, the randomness for the port numbers is decreased significantly. This may allow an attacker to guess the port number used. Non-contiguous port-set, e.g., cryptographical algorithm [RFC6431], can be provided to improve the randomness of port number. It may be used as a mitigation tool against blind attacks. Therefore, in MAP_PORT_SET Opcode, it is mandatory to support two port-set options: PORT_MASK Option and Cryptographically_Random_Port_set Option. Besides, PREFERE_FAILURE Option would also apply for MAP_PORT_SET Opcode.

PCP-controlled device SHOULD provide a configuration option to allow administrators to configure the size of the port set to be assigned and whether cryptographical option is supported or not.

2.1. MAP_PORT_SET Operation Packet Formats

The MAP_PORT_SET Opcode has a similar packet layout for both requests and response. The following figure shows the format of the Opcode in a request for the MAP_PORT_SET Opcode.

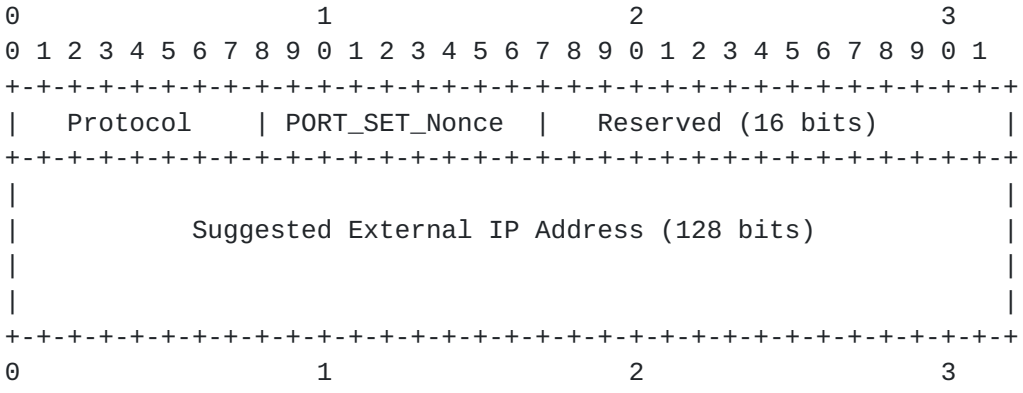

Figure 1: MAP_PORT_SET Opcode format of Request

These fields are described below:

- Protocol: the default value is zero (to indicate all transport protocols).
- PORT_SET_NONCE: Incremental or 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.

- Reserved bits: 16 bits MUST be set to 0.

- Suggested External IP Address: Suggested external IPv4 or IPv6 address. Same as Section 10.1 of [PCP-base].

The following figure shows the format of Opcode-specific information in a response packet for the MAP_PORT_SET Opcode:

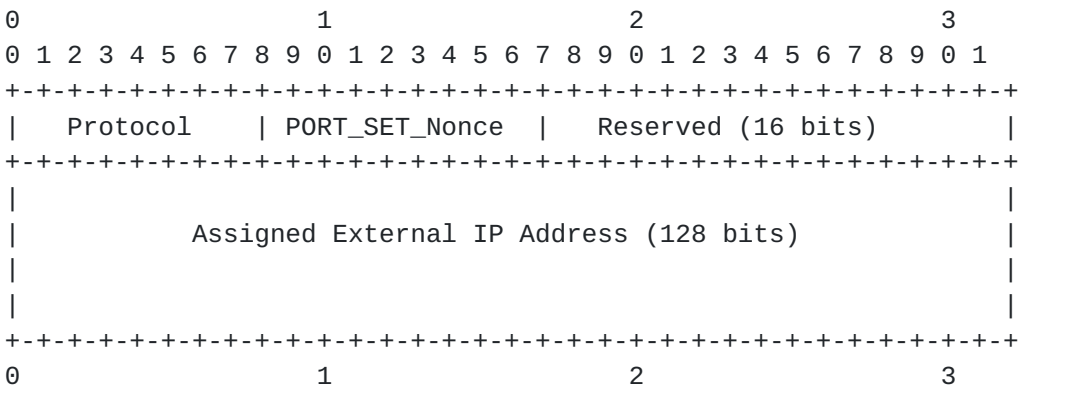

Figure 2: MAP_PORT_SET Opcode format of Response

These fields are described below:

- Protocol: MUST be copied from the request.

- PORT_SET_Nonce: MUST be copied from the request.

- Reserved bits: 16 bits MUST be set to 0.

- Assigned External IP Address (128 bits): This filed 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 coped from the request.

- Requested lifetime (in common header): Requested lifetime for the whole port-set mapping, in seconds. The value 0 also indicates "delete" here.

Discussion note: Assess further whether THIRD_PARTY Option is needed for PORTRANGE OpCode.

2.2. Port-Set Options Formats

The Port_Set options are used to specify one set of ports pertaining to a given IP address. As defined in [RFC6431],there are three kinds of port range: contiguous, non-contiguous and random. A cryptographically random Port Range Option may be used as a mitigation tool against blind attacks. We will describe the two port set PCP options in this section.

2.2.1. Port_Range_Option

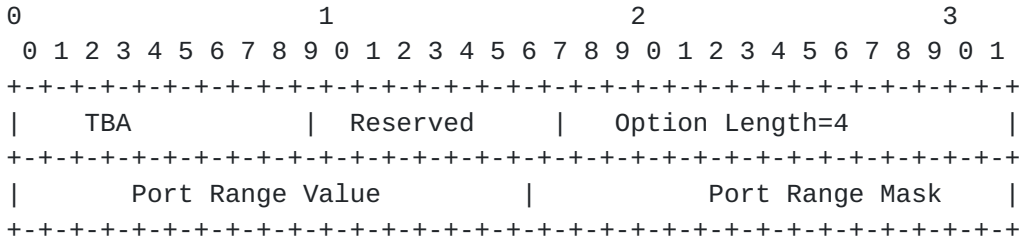

Figure 3: Port_Range_Option

- Port Range Value (PRV): The PRV indicates the value of the significant bits of the Port Mask. By default, no PRM value is assigned. It can also convey Suggested Port Range Value if the client has a hint on it. In MAP_PORT_SET response, it is an Assigned Port Range Value.

- Port Range Mask (PRM): The Port Range Mask indicates the position of the bits that are used to build the Port Range Value. By default, no PRM value is assigned. The 1 values in the Port Range Mask indicate by their position the significant bits of the Port Range Value. It can also convey Suggested Port Range Mask if the client has a hint on it. In MAP_PORT_SET response, it is an Assigned Port Range Mask.

This option:

- name: Port range option

- number: TBA

- purpose: A PCP Client inserts this option in a PCP request to specify one set of ports (contiguous or not contiguous) pertaining to a given IP address.

- is valid for OpCodes:MAP_PORT_SET.
- length:4 octets
- may appear in:request and response
- maximum occurrences:1

2.2.2. Cryptographically_Random_Port_Range_Option

The cryptographically random Port Range PCP Option is formated as below.

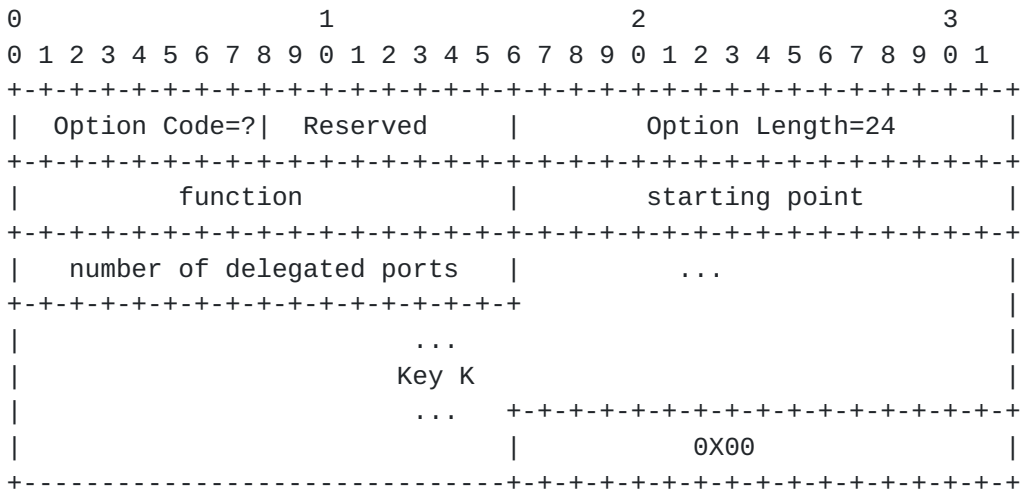

Figure 4: Cryptographically_Random_Port_Range_Option

- function/starting point/number of delegated ports/k: In request packet, it is the suggested function/starting point/number of delegated ports/k which might be helpful for refreshing a mapping after the PCP server loses state. For a success response packet, it is the assigned function/starting point/number of delegated ports/k, while for an error response packet, it is copied from the request.

This option:

- name: Cryptographically Random Port Range Option
- number: TBA
- purpose: A PCP Client inserts this option in a PCP request to specify one set of random ports pertaining to a given IP address. The random ports can be achieved by defining a function that takes

as input a key 'K' and an integer 'x' within the 1024-65535 port range and produces an output 'y' also within the 1024-65535 port range.

- o is valid for OpCodes:MAP_PORT_SET.
- o length: 24 octets.
- o may appear in:request and response
- o maximum occurrences:1

2.3. Generating a MAP_PORT_SET Request

The request MAP_PORT_SET MUST contains one of the port-set options, either PORT_RANGE option or Cryptographically_Random_Port_Set option. The request MAY contain values in the Suggested IP Address field and corresponding parameters in PORT_RANGE option. 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 free port set.

If a client fails to receive an expected response from a server, the client must retransmit its message. The client begins the message exchange by transmitting a message to the server. The PORT_SET_Nonce should be copied from the previous MAP_PORT_SET request.

2.4. Renewing a MAP_PORT_SET Mapping

The similar actions defined in PCP-BASE specification [section 10.2.1 of [ID.ietf-pcp-base]] can be applied to MAP_PORT_SET Opcode to extend the lifetime of a port-set mapping. The MAP_PORT_SET renewal can be regarded as a new PCP request with a different PORT_SET_Nonce. The MAP_PORT_SET request MUST include the currently assigned IP address and port-set in the suggested IP address and port-set options. The PCP-client should renew the port-set mapping before its expiry time.

The PCP client SHOULD renew the mapping before its expiry time, otherwise the port-set binding record will be removed by the PCP server.

2.5. Processing a MAP_PORT_SET Request

The procedures regarding to lifetime is similar to the single port processes in MAP Opcode [section 10.3 of [ID.ietf-pcp-base]], except that the whole port-set should be treated consistently in MAP_PORT_SET Opcode.

It is totally up to the server to determine the port-set quota for each subscriber. 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. The specific mechanism to configure the quotas is out of scope.

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

- If the PCP server or PCP-controlled device does not support MAP_PORT_SET Opcode, the error UNSUPP_OPCODE MUST be returned.

- if the PCP server or PCP-controlled device does not support the port-set option indicated in MAP_PORT_SET request, the error UNSUPP_OPTION MUST be returned.

- If an option does not make sense, (e.g., the PREFER_FAILURE Option is included in a request with lifetime=0, or MAP_PORT_SET Opcode does not include port-set options, etc.,), the request is invalid and generates a MALFORMED_OPTION error. This procedure is the same with section 10.3 of [ID.ietf-pcp-base].

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

The PCP server needs to remember N PORT_SET_Nonces, in which N SHOULD not be larger than floor(MAX_USER_QUOTA/MAX_REQUEST_QUOTA). In order to simplify the implementation, it is recommended that N is equal to ONE so that only one MAP_PORT_SET assignment request is permitted for each subscriber. This policy SHOULD be configurable.

It is possible that a mapping might already exist for a requested Internal address (derived from client's IP address). If so, the PCP server MUST take the following actions:

If the suggested External address and port-set in request packet matches the mapping record (including the Internal address, assigned External address, and the port-set), and the existing mapping is dynamic (created by a previous MAP_PORT_SET), the PCP server MUST update the lifetime of the existing mapping and return the existing External Address and Port in response.

If the suggested External address and port-set in request packet does not match the mapping record for the client, the PCP server SHOULD check whether the PORT_SET_Nonce in the request has a corresponding mapping. If so, it means that this mapping record is created by previous MAP_PORT_SET and request/response might be

discarded for some reason in transimission. Then the PCP server MUST return the existing External Address and Port in its response, regardless of the Suggested External Address and Port in the request. The lifetime of the existing dynamic mapping MUST be updated.

If there is no mapping record in PCP server for the particular PORT_SET_Nonce of MAP_PORT_SET request, it means that the client requires for another delegated set of ports using a new MAP_PORT_SET request. In this case, the PCP server SHOULD check whether the amount of current allocated ports for the client is less than the MAX_USER_QUOTA, and SHOULD assign a new mapping if it does not reach the MAX-USER_QUOTAS and there is no PREFER_FAILURE Option in packet. It is highly suggested that the same external IP address should be assigned for the same subscriber.

If no mapping exists for the requested Internal address (derived from client's IP address), and the PCP server is able to create a mapping using the suggested External Address and Port-set, it Should do so. This is beneficial for re-establishing state lost in the PCP server. If the PCP server cannot assgin the Suggested External Address and Port-set but can assign some other External Address and Port-set (and the request did not contain the PREER_FAILURE Option) the PCP server MUST do so and return newly assigned External Address and Port-set in response.

If the MAP request contains the PREFER_FAILURE Option, but the Suggested External Address and Port is not available, the PCP server MUST return CANNOT_PROVIDE_EXTERNAL.

If the PCP server supports both MAP and MAP_PORT_SET Opcode, the server SHOULD check whether the assigned external address is exactly the same with the one for MAP_PORT_SET , and the external port for MAP is within the range of the port-set for MAP_PORT_SET. Otherwise, the PCP server MUST return NO_RESOURCES.

[Discussion: Should we support MAP_PORT_SET and MAP co-existence scenario? Normally, the PCP server for MAP_PORT_SET will not run NAT. And so, there is no NAT binding in PCP.]

[Discussion note: Do we need to cover the case in which a client MAY send a request to the LSN for another delegated set of ports?]

If all of the preceding operations were successful (did not generate an error response), then the requested port-set mapping is created or refreshed as described in the request and a SUCCESS response is built. The assigned external IPv4 (encoded using IPv4-mapped IPv6

address) or IPv6 address for the mapping should be returned.

2.6. 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 [ID.ietf-pcp-base]] should be followed by the PCP Client to validate the response (except the considerations related to the internal port).

2.7. Mapping Lifetime and Deletion

The procedure for port-set mapping lifetime and deletion is also the same with individual mapping [section 10.5 of [ID.ietf-pcp-base]].

2.8. PREFER_FAILURE Option for MAP_PORT_SET Opcode

This option [section 10.2 of [ID.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.

3. Security Considerations

None.

4. IANA Considerations

The authors request the following new OpCode: MAP_PORT_SET and the following two Options: PORT_RANGE Cryptographically_Random_Port_Set

5. Author 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

6. Contributor List

Gabor Bajko

Nokia

Email: gabor.bajko@nokia.com

7. References

7.1. Normative References

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

[RFC6431] IETF, "Huawei Port Range Configuration Options for PPP IP Control Protocol (IPCP)", November 2011, <http://datatracker.ietf.org/doc/rfc6431/>.

7.2. informative References

[I-D.cui-softwire-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.

[ID.behave-natx4-log-reduction]
Tsou, T., Li, W., and T. Taylor, "Port Management To Reduce Logging In Large-Scale NATs", September 2010.

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-ftgroup.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