Network Working Group
Internet-Draft
Intended status: Experimental
Expires: April 15, 2013

X. Li
C. Bao
CERNET Center/Tsinghua University
W. Dec
O. Troan
Cisco Systems
S. Matsushima
SoftBank Telecom
T. Murakami
IP Infusion

October 12, 2012

Mapping of Address and Port using Translation (MAP-T)
draft-ietf-softwire-map-t-00

Abstract

This document specifies the "Mapping of Address and Port" double stateless translation based solution (MAP-T) for providing shared or uniquely addressed IPv4 host connectivity to and across an IPv6 domain,

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 April 15, 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. Introduction

Experiences from initial IPv6 deployments indicate that transitioning a network providers' domain fully to IPv6 requires not only the continued support of legacy IPv4 users connected to the boundary of that domain, allowing IPv4 address sharing, but also the need for carrying out IPv6-only operational practices in that domain [, also for traffic from IPv4 users. The use of an double NAT64 translation based solutions is an optimal way to address these requirements, particularly in combination with stateless translation techniques that seek to minimize challenges outlined in [I-D.ietf-softwire-stateless-4v6-motivation].

The Mapping of Address and Port - Translation (MAP-T) solution defined in this document is such a solution, that builds on existing stateless NAT64 techniques specified in [RFC6145], along with a stateless algorithmic address & port mapping scheme to allow the sharing of IPv4 addresses across an IPv6 network. The MAP-T solution is closely related to MAP-E [I-D.ietf-softwire-map], with both utilizing the same algorithmic method, but differing in their choice of translation [RFC6145] and encapsulation [RFC2473]based IPv6 transports.

A companion draft defines the DHCPv6 options for provisioning of MAP [I-D.mdt-softwire-map-dhcp-option], applicable to both MAP-T and MAP-E.

2. Conventions

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 [RFC2119].

3. Terminology

MAP domain: One or more MAP CEs and BRs connected to the same IPv6 network. A service provider may deploy a single MAP domain, or may utilize multiple MAP domains.

MAP Rule: A set of parameters describing the mapping between an IPv4 prefix, IPv4 address or shared IPv4 address and an IPv6 prefix or address. Each MAP domain uses a different mapping rule set.

MAP node: A device that implements MAP.

MAP Border Relay (BR): A MAP enabled router managed by the service provider at the edge of a MAP domain. A Border Relay router has at least an IPv6-enabled interface and an IPv4 interface connected to the native IPv4 network. A MAP BR may also be referred to simply as a "BR" within the context of MAP.

MAP Customer Edge (CE): A device functioning as a Customer Edge router in a MAP deployment. A typical MAP CE adopting MAP rules will serve a residential site with one WAN side interface, and one or more LAN side interfaces. A MAP CE may also be referred to simply as a "CE" within the context of MAP.

Port-set: Each node has a separate part of the transport layer port space; denoted as a port-set.

Port-set ID (PSID): Algorithmically identifies a set of ports exclusively assigned to the CE.

Shared IPv4 address: An IPv4 address that is shared among multiple CEs. Only ports that belong to the assigned port-set can be used for communication. Also known as a Port-Restricted IPv4 address.

End-user IPv6 prefix: The IPv6 prefix assigned to an End-user CE by other means than MAP itself. E.g. Provisioned using DHCPv6 PD [RFC3633] or configured manually. It is unique for each CE.

MAP IPv6 address: The IPv6 address used to reach the MAP function of a CE from other CEs and from BRs.

Rule IPv6 prefix: An IPv6 prefix assigned by a Service Provider for a mapping rule.

Rule IPv4 prefix: An IPv4 prefix assigned by a Service Provider for a mapping rule.

Embedded Address (EA) bits: The IPv4 EA-bits in the IPv6 address identify an IPv4 prefix/address (or part thereof) or a shared IPv4 address (or part thereof) and a port-set identifier.

MRT: MAP Rule table. Address and Port aware data structure, supporting longest match lookups. The MRT is used by the MAP forwarding function.

4. Architecture

Figure 1 depicts the overall MAP-T architecture with IPv4 users (N and M) networks connected by means of MAP CEs to an IPv6 network that is equipped with one or more MAP BR.

```
         User N
       Private IPv4
      |  Network
      |
   O--+---------------O
   |  | MAP-T CE      |
   | +-----+--------+ |
   | NAPT44|  MAP-T | `-.
   | +-----+      | |  -._   ,-------.                     .------.
   |       +--------+ |   ,-'         `-.               ,-'       `-.
   O------------------O  /               \   O---------O /   Public   \
                         /   IPv6 only   \  |  MAP-T  |/     IPv4     \
                        (    Network      --+  Border +-   Network     )
                         \ (MAP-T Domain)/  |  Relay  |\              /
   O------------------O  \               /   O---------O \             /
   |    MAP-T CE      |   ;".         ,-'                `-.       ,-'
   | +-----+--------+ | ,"   `----+--'                      ------'
   | NAPT44|  MAP-T | | ,"        |
   | +-----+        | |       IPv6 Server(s)
   |   |   +--------+ |         (w/ v4 mapped
   O---.--------------O          address)
       |
         User M
       Private IPv4
         Network
```

Figure 1: Network Topology

The MAP-T CE is responsible for connecting a users' private IPv4, along with any native IPv6 network to the IPv6-only MAP-T domain.

The MAP-T BR is responsible for connecting external IPv4 networks to all devices in one or more MAP-T domains, using stateless NAT64 as extended by the MAP-T rules in this document.

Besides the CE and BR, the MAP-T domain can contain any regular IPv6-only hosts/servers that have an IPv4 mapped IPv6 address (IPv4-translatable address per [RFC6052]) using a prefix assigned to the MAP-T domain. Communication with such devices is naturally possible in the MAP-T architecture from inside or outside the MAP-T domain including from any IPv4-only hosts. In this mode, any IPv6-only servers SHOULD have both A and AAAA DNS server records. DNS64 [RFC6147] becomes required only when IPv6 servers in the MAP-T domain are expected themselves to initiate communication to internal/ external IPv4-only entities.

Functionally the MAP-T CE and BR use existing standard functional building blocks, or extensions to these as follows:

o A standard NAPT [RFC2663] function on a MAP CE is extended with support for restricting the allowable TCP/UDP ports for a given IPv4 address. The enforcement of NAPT function, as well as the port restriction is conditional upon the MAP CE's configuration as applied by the network operator.

o A standard stateless NAT64 function [RFC6145]is extended for translating IPv4 traffic to IPv6, based on a lookup of source/ destination IPv4 address + TCP/UDP port or ICMP id information that is then mapped to the source/destination IPv6 address. This algorithmic mapping is specified in section 5.

Section 6 describes how these functional blocks are combined with the Mapping Rules of Section 4 to enable IPv4-IPv6 communication between the CE and BR and IPv6-only servers in the domain.

5. Mapping Rules

A MAP node is provisioned with one or more mapping rules.

Mapping rules are used differently depending on their function. Every MAP node must be provisioned with a Basic mapping rule. This is used by the node to configure its IPv4 address, IPv4 prefix or shared IPv4 address. This same basic rule can also be used for forwarding, where an IPv4 destination address and optionally a destination port is mapped into an IPv6 address or prefix. Additional mapping rules are specified to allow for e.g. Multiple different IPv4 subnets to exist within the domain and optimize forwarding between them.

Traffic outside of the domain (i.e. When the destination IPv4 address does not match (using longest matching prefix) any Rule IPv4 prefix in the Rules database) will be forward using the Default mapping rule. The Default mapping rule maps outside destinations to the BR's IPv6 address or prefix.

In summary, there are three types of mapping rules:

1. Basic Mapping Rule (BMR) - used for configuring the CE's IPv4 address and/or port set assignment as well as deriving the MAP IPv6 address that the CE is to use. There can only be one Basic Mapping Rule per End-user IPv6 prefix. The BMR is composed of the following parameters:

   * Rule IPv6 prefix (including prefix length)

   * Rule IPv4 prefix (including prefix length)

   * Rule EA-bits length (in bits)

   * Rule Port Parameters (optional)

2. Forwarding Mapping Rule - used for forwarding in the MAP domain. The Basic Mapping Rule is also a Forwarding Mapping Rule. Each Forwarding Mapping Rule will result in an entry in the MRT for the Rule IPv4 prefix + any port range. The FMR consists of the following parameters:

   * Rule IPv6 prefix (including prefix length)

   * Rule IPv4 prefix (including prefix length)

   * Rule EA-bits length (in bits)

   * Rule Port Parameters (optional)

3. Default Mapping Rule - used for destinations outside the MAP domain. An IPv4 0.0.0.0/0 entry is installed in the MRT for this rule.

   * IPv6 prefix of the BR

   * Rule BR IPv4 address (Optional - can be used for testing a BR's reachability)

A MAP node finds its Basic Mapping Rule by doing a longest match between the End-user IPv6 prefix and the Rule IPv6 prefix in the Mapping Rule database. The rule is then used for IPv4 prefix,

address or shared address assignment.

A MAP IPv6 address is formed from the BMR Rule IPv6 prefix. This address MUST be recognized by the MAP node, typically a CE, and used to terminate all MAP traffic received by the node.

Port-aware IPv4 entries in the MRT are installed for all the Forwarding Mapping Rules and an IPv4 default route for the Default Mapping Rule.

In hub and spoke mode, all traffic MUST be forwarded using the Default Mapping Rule.

5.1. Basic mapping rule (BMR)

The BMR is used in combination with the CE's IPv6 prefix to derive the MAP IPv4 address, port-set and also the MAP IPv6 address. The structure of a MAP CE's IPv6 address is shown below, along with the Interface-identifer that is defined in Section 5.4.

```
                | p bits   |         |   q bits   |
                +----------+         +------------+
                |IPv4  sufx|         |Port-Set ID |
                +----------+         +------------+
                \          /    ____/    ________/
                  \       :  __/   _____/
                    \     : /     /
|     n bits         |  o bits   | s bits  |   128-n-o-s bits      |
+--------------------+-----------+---------+------------+----------+
|  Rule IPv6 prefix  |  EA bits  |subnet ID|     interface ID      |
+--------------------+-----------+---------+-----------------------+
|<---  End-user IPv6 prefix  --->|
```

Figure 2: IPv6 address format

The Rule IPv6 prefix is the part of the End-user IPv6 prefix (i.e. the regular IPv6 prefix or address that is assigned to any IPv6 device) that is common among all CEs within the MAP domain. The Embedded Address bits (EA bits) are the unique per end user within that Rule IPv6 prefix (some readers may want to think of these as "MAP customer identifier" bits for a given MAP domain covered by the Rule IPv6 prefix). When present, the EA bits encode the CE specific full or part of an IPv4 prefix or address, and in the shared IPv4 address case contain a Port-Set Identifier (PSID) that ultimately determines the allowed port-range for the CE. An EA-bit length of 0

signifies that all relevant MAP IPv4 addressing information is passed directly in the BMR rule

The MAP subnet ID is defined to be the first subnet (all bits set to zero) of length s required to make n+o+s=64. A MAP CE node MUST reserve the first IPv6 prefix in a End-user IPv6 prefix for the purpose of MAP.

CE's MAP IPv6 address is created by concatenating the End-user IPv6 prefix with the all zeros MAP subnet-id and the interface-id as specified in Section 5.4.

The CE's IPv4 address is synthesized from the CE's IPv6 address and the parameters obtained in the BMR, namely by extracting the IPv4 suffix from the EA-bits, if any, and combining it with the Rule's IPv4 prefix. Figures below show the synthesis for cases of a Shared IPv4 address, and a non-shared, complete, IPv4 address:

```
                        Shared IPv4 address:


     |   r bits    |        p bits       |         |   q bits   |
     +-------------+---------------------+         +------------+
     |  Rule IPv4  | IPv4 Address suffix |         |Port-Set ID |
     +-------------+---------------------+         +------------+
     |            32 bits                |
```

Figure 3: Shared IPv4 address

```
                       Complete IPv4 address:


                |   r bits    |        p bits       |
                +-------------+---------------------+
                |  Rule IPv4  | IPv4 Address suffix |
                +-------------+---------------------+
                |            32 bits                |
```

Figure 4: Complete IPv4 address

IPv4 prefix:

```
|   r bits    |        p bits       |
+-------------+---------------------+
|  Rule IPv4  | IPv4 Address suffix |
+-------------+---------------------+
|           < 32 bits               |
```

Figure 5: IPv4 prefix

The length of r MAY be zero, in which case the complete IPv4 address or prefix is encoded in the EA bits. If only a part of the IPv4 address/prefix is encoded in the EA bits, the Rule IPv4 prefix is provisioned to the CE by other means (e.g. a DHCPv6 option). To create a complete IPv4 address (or prefix), the IPv4 address suffix (p) from the EA bits, are concatenated with the Rule IPv4 prefix (r bits).

The offset of the EA bits field in the IPv6 address is equal to the BMR Rule IPv6 prefix length. The length of the EA bits field (o) is given by the BMR Rule EA-bits length. The sum of the Rule IPv6 Prefix length and the Rule EA-bits length MUST be less or equal than the End-user IPv6 prefix length.

If o + r < 32 (length of the IPv4 address in bits), then an IPv4 prefix is assigned.

If o + r is equal to 32, then a full IPv4 address is to be assigned. The address is created by concatenating the Rule IPv4 prefix and the EA-bits.

If o + r is > 32, then a shared IPv4 address is to be assigned. The number of IPv4 address suffix bits (p) in the EA bits is given by 32 - r bits. The PSID bits are used to create a port-set. The length of the PSID bit field within EA bits is: o - p.

The length of r MAY be 32, with no part of the IPv4 address embedded in the EA bits. This results in a mapping with no dependence between the IPv4 address and the IPv6 address. In addition the length of o MAY be zero (no EA bits embedded in the End-User IPv6 prefix), meaning that also the PSID is provisioned using e.g. The DHCP option.

In the following examples, only the suffix (last 8 bits) of the IPv4 address is embedded in the EA bits (r = 24), while the IPv4 prefix (first 24 bits) is given in the BMR Rule IPv4 prefix.

```
Example:

  Given:
   End-user IPv6 prefix:  2001:db8:0012:3400::/56
   Basic Mapping Rule:    {2001:db8:0000::/40 (Rule IPv6 prefix),
                          192.0.2.0/24 (Rule IPv4 prefix),
                          16 (Rule EA-bits length)}
   Sharing ratio:         256 (16 - (32 - 24) = 8. 2^8 = 256)
   PSID offset:           4 (default value as per section 5.1.3)

  We get IPv4 address and port-set:
   EA bits offset:       40
   IPv4 suffix bits (p): Length of IPv4 address (32) -
                         IPv4 prefix length (24) = 8
   IPv4 address:         192.0.2.18 (0x12)

   PSID start:           40 + p = 40 + 8 = 48
   PSID length:          o - p = 16 (56 - 40) - 8 = 8
   PSID:                 0x34
   Port-set-1:        4928, 4929, 4930, 4931, 4932, 4933, 4934, 4935,
                      4936, 4937, 4938, 4939, 4940, 4941, 4942, 4943
   Port-set-2:        9024, 9025, 9026, 9027, 9028, 9029, 9030, 9031,
                      9032, 9033, 9034, 9035, 9036, 9037, 9038, 9039
   ...
   Port-set-15:          62272, 62273, 62274, 62275,
                         62276, 62277, 62278, 62279,
                         62280, 62281, 62282, 62283,
                         62284, 62285, 62286, 62287,
```

5.2. Forwarding mapping rule (FMR)

On adding an FMR rule, an IPv4 route is installed in the MRT for the Rule IPv4 prefix.

On forwarding an IPv4 packet, a best matching prefix lookup is done in the MRT and the correct FMR is chosen.

```
|        32 bits           |         |    16 bits        |
+--------------------------+         +-------------------+
| IPv4 destination address |         |  IPv4 dest port   |
+--------------------------+         +-------------------+
                :          :           ___/       :
                | p bits   |          /  q bits   :
                +----------+         +------------+
                |IPv4  sufx|         |Port-Set ID |
                +----------+         +------------+
                \          /    ____/    ________/
                  \       :  __/   _____/
                    \     : /     /
|     n bits         |  o bits   | s bits  |   128-n-o-s bits      |
+--------------------+-----------+---------+------------+----------+
|  Rule IPv6 prefix  |  EA bits  |subnet ID|     interface ID      |
+--------------------+-----------+---------+-----------------------+
|<---  End-user IPv6 prefix  --->|


  Given:
   IPv4 destination address: 192.0.2.18
   IPv4 destination port:    9030
   Forwarding Mapping Rule:  {2001:db8:0000::/40 (Rule IPv6 prefix),
                              192.0.2.0/24 (Rule IPv4 prefix),
                              16 (Rule EA-bits length)}
   PSID offset:              4 (default value as per section 5.1.3)

  We get IPv6 address:
   IPv4 suffix bits (p): 32 - 24 = 8 (18 (0x12))
   PSID length:          8
   PSID:                 0x34 (9030 (0x2346))
   EA bits:              0x1234
   MAP IPv6 address:     2001:db8:0012:3400:00c0:0002:1200:3400
```

Figure 6: Deriving of MAP IPv6 address

5.3. Default mapping rule (DMR)

The Default Mapping rule is used to represent as IPv6 destinations all IPv4 destinations outside of the MAP IPv4 domain. For MAP-T, the DMR is specified in terms of the BR IPv6 prefix. The Rule IPv4 prefix in the MRT is: 0.0.0.0/0, i.e. the default IPv4 route.

There MUST be only one Default Mapping Rule within a MAP domain.

```
 Default Mapping Rule:
      {2001:db8:0001::/Prefix-length (Rule IPv6 prefix),
```

```
          0.0.0.0/0 (Rule IPv4 prefix)}

                       Example: Default Mapping Rule
```

The Deafult Rule's IPv6 prefix is combined by a MAP-T CE with the IPv4 destination addresses, using RFC6052, to form a full IPv6 destination address for any IPv4 destination following the IPv4 default route. Figure 7 below shows such an address format. Note that the BR prefix-length is variable and can be both shorter or longer than 64 bits, up to 96 bits. In the respective cases the IPv4 address and the BR prefix are shifted and "bit spread" across the fixed u-octet boundary as per [RFC6052]. All trailing bits after the IPv4 address are set to 0x0.

```
<---------- 64 ------------>< 8 ><----- 32 ----->&lt;--- 24 --->
+--------------------------+----+---------------+-----------+
|      BR IPv6 prefix      | u  | IPv4 address  |     0     |
+--------------------------+----+---------------+-----------+
```

Figure 7: IPv6-IPv4 Mapped Address

5.4. MAP IPv6 Interface Identifier

The IPv6 Interface identifier format of a MAP node is based on the format specified in section 2.2 of [RFC6052], with the added PSID field if present, as shown in Figure 8.

```
+--+---+---+---+---+---+---+---+---+
|PL|   8  16  24  32  40  48  56   |
+--+---+---+---+---+---+---+---+---+
|64| u | IPv4 address  |  PSID | 0 |
+--+---+---+---+---+---+---+---+---+
```

Figure 8: MAP IPv6 Interface Identifier

The encoding of the full IPv4 address into the interface identifier into the IPv6 addresses have been shown to be useful for troubleshooting.

In the case of an IPv4 prefix, the IPv4 address field is right-padded with zeros up to 32 bits. The PSID field is left-padded to create a 16 bit field. For an IPv4 prefix or a complete IPv4 address, the

PSID field is zero.

5.5. Port mapping algorithm

The Generalized Modulus Algorithm (GMA) is used in MAP domains whose rules allow IPv4 address sharing. Each CE in such a domain has an IPv4 address and a unique Port-Set Identifier (PSID), that is derived by means of the BMR. For a given IPv4 address, the algorithm allows each PSID to be processed to reveal a set of unique non-overlapping ports, or alternatively for any given port to derive the PSID it corresponds to. Two extreme cases supported by algorithm are: (1) the port numbers are not contiguous for each PSID, but uniformly distributed across the port range (0-65535); (2) the port numbers are contiguous in a single range for each PSID.

For a given sharing ratio (R) and the maximum number of contiguous ports (M), the GMA algorithm is defined as:

1. The port (P) of a given PSID (K) is composed of:

P = R * M * j + M * K + i

Where:

* PSID: K = 0 to R - 1

* Port range index: j = (4096 / M) / R to ((65536 / M) / R) - 1, if the port numbers (0 - 4095) are excluded.

* Contiguous Port index: i = 0 to M - 1

2. The PSID (K) of a given port number (P) is determined by:

K = (floor(P/M)) % R

Where:

* % is the modulus operator

* floor(arg) is a function that returns the largest integer not greater than arg.

5.5.1. Bit Representation of the Algorithm

Given a sharing ratio (R=2^k), the maximum number of contiguous ports (M=2^m), for any PSID (K) and available ports (P) can be represented as:

```
0                           8                         15
+---------------+----------+------+------------------+
|                        Port  (P)                    |
----------------+----------------+------------------+
|        j      |   PSID (K)      |         M  (i)     |
+---------------+----------+------+------------------+
|<----a bits--->|<-----k bits---->|<------m bits----->|
```

Figure 9: Bit representation

Where j and i are the same indexes defined in the port mapping algorithm.

For any port number, the PSID can be obtained by a bit mask operation.

For a > 0, j MUST be larger than 0. This ensures that the algorithm excludes the system ports ([I-D.ietf-tsvwg-iana-ports]). For a = 0, j MAY be 0 to allow for the provisioning of the system ports.

5.5.2. GMA examples

For example, for R = 1024, PSID offset: a = 4 and PSID length: k = 10 bits

```
          Port-set-1                 Port-set-2
PSID=0   | 4096, 4097, 4098, 4099, | 8192,  8193,  8194,  8195, | ...
PSID=1   | 4100, 4101, 4102, 4103, | 8196,  8197,  8198,  8199, | ...
PSID=2   | 4104, 4105, 4106, 4107, | 8200,  8201,  8202,  8203, | ...
PSID=3   | 4108, 4109, 4110, 4111, | 8204,  8205,  8206,  8207, | ...
...
PSID=1023| 8188, 8189, 8190, 8191, | 12284, 12285, 12286, 12287,| ...
```

Example 1: with offset = 4 (a = 4)

For example, for R = 64, a = 0 (PSID offset = 0 and PSID length = 6 bits):

```
          Port-set
PSID=0   | [   0 - 1023]
PSID=1   | [1024 - 2047]
PSID=2   | [2048 - 3071]
PSID=3   | [3072 - 4095]
...
PSID=63  | [64512 - 65535]
```

Example 2: with offset = 0 (a = 0)

5.5.3. GMA Excluded Ports

By default the GMA ensures that a number of "well known" ports are excluded from use by the algorithm. This number is determined by the number of offset bits (a), in the figure above. This value can be optionally provisioned via the "Rule Port Mapping Parameters" in the Basic Mapping Rule. In the absence of such provisioning, the defaults are:

- Excluded ports : 0-4095

- Offset bits (a) : 4

For (a) offset bits, the range of excluded ports is 0 to 2 ^ (16-a) - 1.

6. Packet Forwarding

The mapping rules and architectual building blocks are combined at the CE and BR to enable IPv4-IPv6 communication as follows.

6.1. IPv4 to IPv6 at the CE

A MAP-T CE receiving IPv4 packets SHOULD perform NAT44 function first and create appropriate NAT44 stateful bindings. The resulting IPv4 packets MUST contain the source IPv4 address and source transport port number assigned to the CE by means of the MAP Basic Mapping Rule (BMR).

The IPv4 traffic is subject to a longest IPv4 address + port match MAP rule selection using the MRT, which then determines the subsequent NAT64 operation. By default, all traffic is matched to default mapping rule (DMR), and subject to the stateless NAT64 operation using the DMR parameters for the MAP algorithm and NAT64.

An optional mapping rule, known as a forward mapping rule (FMR), can be used when forwarding to destinations that correspond to a specific IPv4+port range in the MAP-T domain i.e. Typically the IPv4 address and port range of another MAP-T CE, aka mesh-mode. Traffic that is matched to such a rule is subject to the stateless NAT64 operation using the FMR parameters for the MAP algorithm and stateless NAT64.

A MAP-T CE MUST support a default mapping rule and SHOULD support one or more forward mapping rules.

6.2. IPv6 to IPv4 at the CE

A MAP-T CE receiving an IPv6 packet performs its regular IPv6 operations, whereby only packets that are addressed to the MAP-T BMR addresses are forwarded to the CE's stateless NAT64 function. All other IPv6 traffic SHOULD be forwarded as per the CE's IPv6 routing rules. The CE SHOULD check that MAP-T received packets' transport-layer destination port number is in the range configured by MAP for the CE and the CE SHOULD drop any non conforming packet and respond with an ICMPv6 "Address Unreachable" (Type 1, Code 3).

The CE's stateless NAT64 function MUST derive the IPv4 source and destination addresses as per Section 5 of this document and MUST replace the IPv6 header with an IPv4 header in accordance with [RFC6145]. The resulting IPv4 packet is then forwarded to the CE's NAPT function, when this is enabled, where the destination IPv4 address and port number MUST be mapped to their original value, before being forwarded according to the CE's regular IPv4 rules. When the NAPT function is not enabled, the traffic from the stateless NAT64 function is directly forwarded according to the CE's IPv4 rules.

6.3. IPv6 to IPv4 at the BR

A MAP-T BR receiving IPv6 packets MUST select a best matching MAP rule based on a longest address match of the packets' source address against the BR's configured MAP BMR prefix(es), as well as a match of the packet destination address against the configured FMR prefix(es). The selected MAP rule allows the BR to determine the CE's range from the port-set-id contained in the source IPv6 address. The BR MUST perform a validation of the consistency of the source against the allowed values from the identified port-range port. If the packets source port number is found to be outside the range allowed for this CE-index and the BMR, the BR MUST drop the packet and respond with an ICMPv6 "Destination Unreachable, Source address failed ingress/egress policy" (Type 1, Code 5).

The BR MUST derive the source and destination IPv4 addresses as per Section 5 of this document and translate the IPv6 to IPv4 headers following [RFC6145]. The resulting IPv4 packets are then passed to regular IPv4 forwarding by the BR.

6.4. IPv4 to IPv6 at the BR

A MAP-T BR receiving IPv4 packets uses a longest match IPv4 + port lookup to select the target MAP-T domain and rule. The BR MUST then derive the IPv6 source and destination addresses from the IPv4 source and destination address and port as per Section 5 of this document.

Following this, the BR MUST translate the IPv4 to IPv6 headers following [RFC6145]. The resulting IPv6 packets are then passed to regular IPv6 forwarding.

Note that the operation of a BR when forwarding to MAP-T domains that do not utilize IPv4 address sharing, is the same as stateless IPv4/ IPv6 translation.

7. ICMP Handling

ICMP messages need to be supported in MAP-T domain and also across it, taking into consideration also the NAPT component and best current practice documented in [RFC5508] along with some additional specific considerations.

MAP-T CEs and BRs MUST follow ICMP/ICMPv6 translation as per [RFC6145], with the following extension to cover the address sharing/ port-range feature.

Unlike TCP and UDP, which each provide two port fields to represent both source and destination, the ICMP/ICMPv6 [RFC0792], [RFC4443] Query message header has only one ID field which needs to be used to identify a sending IPv4 host.

When receiving IPv4 ICMP messages, the MAP-T CE SHOULD rewrite the ID field to a port value derived from the Port-set-id. A BR MUST translate the resulting ICMPv6 packets back to ICMP preserving the ID field on its way to an IPv4 destination.

In the return path, when MAP-T BR receives an ICMP packet containing an ID field which is bound for a shared address in the MAP-T domain, the MAP-T BR SHOULD use the ID value as a substitute for the destination port in determining the IPv6 destination address. In all other cases, the MAP-T BR MUST derive the destination IPv6 address by simply mapping the destination IPv4 address without additional port info.

If a MAP BR receives an ICMP error message on its IPv4 interface, the MAP BR should translate the ICMP message to an appropriate ICMPv6 message, as per [RFC6145] and forward it to the intended MAP CE with the following considerations. If IPv4 address is not shared, the MAP BR generates a CE IPv6 address from the IPv4 destination address in the ICMP error message and encapsulates the ICMP message in IPv6. If the IPv4 address is shared, the MAP BR derives an original IPv4 packet from the ICMP payload and generates a CE IPv6 address from the source address and the source port in the original IPv4 packet.

8. Fragmentation and Path MTU Discovery

Due to the different sizes of the IPv4 and IPv6 header, handling the maximum packet size is relevant for the operation of any system connecting the two address families. There are three mechanisms to handle this issue: Path MTU discovery (PMTUD), fragmentation, and transport-layer negotiation such as the TCP Maximum Segment Size (MSS) option [RFC0897]. MAP uses all three mechanisms to deal with different cases.

8.1. Fragmentation in the MAP domain

Translating an IPv4 packet to carry it across the MAP domain will increase its size by 20 bytes respectively. It is strongly recommended that the MTU in the MAP domain is well managed and that the IPv6 MTU on the CE WAN side interface is set so that no fragmentation occurs within the boundary of the MAP domain.

Fragmentation in MAP-T domain is to be handled as described in section 4 and 5 of [RFC6145].

8.2. Receiving IPv4 Fragments on the MAP domain borders

Forwarding of an IPv4 packet received from the outside of the MAP domain requires the IPv4 destination address and the transport protocol destination port. The transport protocol information is only available in the first fragment received. As described in section 5.3.3 of [RFC6346] a MAP node receiving an IPv4 fragmented packet from outside has to reassemble the packet before sending the packet onto the MAP link. If the first packet received contains the transport protocol information, it is possible to optimize this behavior by using a cache and forwarding the fragments unchanged. A description of this algorithm is outside the scope of this document.

8.3. Sending IPv4 fragments to the outside

If two IPv4 host behind two different MAP CE's with the same IPv4 address sends fragments to an IPv4 destination host outside the domain. Those hosts may use the same IPv4 fragmentation identifier, resulting in incorrect reassembly of the fragments at the destination host. Given that the IPv4 fragmentation identifier is a 16 bit field, it could be used similarly to port ranges. A MAP CE SHOULD rewrite the IPv4 fragmentation identifier to be within its allocated port set.

9. Usage Considerations

9.1. Hub and spoke with per subscriber rules

Existing IPv4 service can be realized with MAP using a mapping rule per subscriber. By embedding no part of the IPv4 address in the IPv6 prefix, no dependency between the two address families is created. This may be useful in cases where the IPv6 address allocation is sparse, or for other reasons it is difficult to create efficient mapping rules.

The operator has to the choice of provisioning a full IPv4 address to the end-user, or a shared IPv4 address by also provisioning the PSID in the DHCPv6 option. A hybrid of this use case is to provision the full IPv4 address in the DHCPv6 option, while embedding the PSID in the IPv6 prefix. That will result in one mapping rule per IPv4 address, e.g. With a sharing ratio of 64, one rule per 64 customers.

9.2. Communication with IPv6 servers in the MAP-T domain

MAP-T allows communication between both IPv4-only and any IPv6 enabled end hosts, with native IPv6-only servers which are using IPv4-mapped IPv6 address based on DMR in the MAP-T domain. In this mode, the IPv6-only servers SHOULD have both A and AAAA records in DNS [RFC6219]. DNS64 [RFC6147] become required only when IPv6 servers in the MAP-T domain are expected themselves to initiate communication to external IPv4-only hosts.

9.3. Backwards compatibility

A MAP-T CE, in all configuration modes, is by default compatible with regular [RFC6146] stateful NAT64 devices that are configured to use/ advertise BR prefixes. This allows the use of MAP-T CEs in environments that require statistical multiplexing of IPv4 addresses while being able to compromise on the stateful nature. Furthermore, a MAP-T CE configured to operate without address sharing (no PSID) is compatible with any stateless NAT64 [RFC6146] devices positioned as BRs.

10. IANA Considerations

This specification does not require any IANA actions.

11. Security Considerations

Spoofing attacks: With consistency checks between IPv4 and IPv6 sources that are performed on IPv4/IPv6 packets received by MAP nodes, MAP does not introduce any new opportunity for spoofing attacks that would not already exist in IPv6.

Denial-of-service attacks: In MAP domains where IPv4 addresses are shared, the fact that IPv4 datagram reassembly may be necessary introduces an opportunity for DOS attacks. This is inherent to address sharing, and is common with other address sharing approaches such as DS-Lite and NAT64/DNS64. The best protection against such attacks is to accelerate IPv6 enablement in both clients and servers so that, where MAP is supported, it is less and less used.

Routing-loop attacks: This attack may exist in some automatic tunneling scenarios are documented in [RFC6324]. They cannot exist with MAP because each BRs checks that the IPv6 source address of a received IPv6 packet is a CE address based on Forwarding Mapping Rule.

Attacks facilitated by restricted port set: From hosts that are not subject to ingress filtering of [RFC2827], some attacks are possible by an attacker injecting spoofed packets during ongoing transport connections ([RFC4953], [RFC5961], [RFC6056]. The attacks depend on guessing which ports are currently used by target hosts, and using an unrestricted port set is preferable, i.e. Using native IPv6 connections that are not subject to MAP port range restrictions. To minimize this type of attacks when using a restricted port set, the MAP CE's NAT44 filtering behavior SHOULD be "Address-Dependent Filtering". Furthermore, the MAP CEs SHOULD use a DNS transport proxy function to handle DNS traffic, and source such traffic from IPv6 interfaces not assigned to MAP-T. Practicalities of these methods are discussed in Section 5.9 of [I-D.dec-stateless-4v6].

[RFC6269] outlines general issues with IPv4 address sharing.

12. Contributors

Mohamed Boucadair, Gang Chen, Maoke Chen, Wojciech Dec, Xiaohong Deng, Jouni Korhonen, Tomasz Mrugalski, Jacni Qin, Chunfa Sun, Qiong Sun, Leaf Yeh.

The following are the authors who provided a major contribution to this document:

Chongfeng Xie (China Telecom)

Room 708, No.118, Xizhimennei Street Beijing 100035 CN

Phone: +86-10-58552116

Email: xiechf@ctbri.com.cn

Qiong Sun (China Telecom)

Room 708, No.118, Xizhimennei Street Beijing 100035 CN

Phone: +86-10-58552936

Email: sunqiong@ctbri.com.cn

Rajiv Asati (Cisco Systems)

7025-6 Kit Creek Road Research Triangle Park NC 27709 USA

Email: rajiva@cisco.com

Gang Chen (China Mobile)

53A,Xibianmennei Ave. Beijing 100053 P.R.China

Email: chengang@chinamobile.com

Wentao Shang (CERNET Center/Tsinghua University)

Room 225, Main Building, Tsinghua University Beijing 100084 CN

Email: wentaoshang@gmail.com

Guoliang Han (CERNET Center/Tsinghua University)

Room 225, Main Building, Tsinghua University Beijing 100084 CN

Email: bupthgl@gmail.com

Yu Zhai CERNET Center/Tsinghua University

Room 225, Main Building, Tsinghua University Beijing 100084 CN

Email: jacky.zhai@gmail.com

13. Acknowledgements

This document is based on the ideas of many. In particular Remi Despres, who has tirelessly worked on generalized mechanisms for stateless address mapping.

The authors would like to thank Guillaume Gottard, Dan Wing, Jan Zorz, Necj Scoberne, Tina Tsou for their thorough review and comments.

14. References

14.1. Normative References

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

[RFC6052] Bao, C., Huitema, C., Bagnulo, M., Boucadair, M., and X. Li, "IPv6 Addressing of IPv4/IPv6 Translators", RFC 6052, October 2010.

[RFC6145] Li, X., Bao, C., and F. Baker, "IP/ICMP Translation Algorithm", RFC 6145, April 2011.

[RFC6346] Bush, R., "The Address plus Port (A+P) Approach to the IPv4 Address Shortage", RFC 6346, August 2011.

14.2. Informative References

[I-D.dec-stateless-4v6]
Dec, W., Asati, R., and H. Deng, "Stateless 4Via6 Address Sharing", draft-dec-stateless-4v6-04 (work in progress), October 2011.

[I-D.ietf-softwire-map]
Troan, O., Dec, W., Li, X., Bao, C., Matsushima, S., and T. Murakami, "Mapping of Address and Port with Encapsulation (MAP)", draft-ietf-softwire-map-02 (work in progress), September 2012.

[I-D.ietf-softwire-stateless-4v6-motivation]
Boucadair, M., Matsushima, S., Lee, Y., Bonness, O., Borges, I., and G. Chen, "Motivations for Carrier-side Stateless IPv4 over IPv6 Migration Solutions", draft-ietf-softwire-stateless-4v6-motivation-04 (work in progress), August 2012.

[I-D.ietf-tsvwg-iana-ports]
Cotton, M., Eggert, L., Touch, J., Westerlund, M., and S. Cheshire, "Internet Assigned Numbers Authority (IANA) Procedures for the Management of the Service Name and Transport Protocol Port Number Registry", draft-ietf-tsvwg-iana-ports-10 (work in progress), February 2011.

[I-D.mdt-softwire-map-dhcp-option]
Mrugalski, T., Troan, O., Bao, C., and W. Dec, "DHCPv6 Options for Mapping of Address and Port", draft-mdt-softwire-map-dhcp-option-03 (work in progress), July 2012.

[I-D.xli-behave-divi]
Shang, W., Li, X., Zhai, Y., and C. Bao, "dIVI: Dual-Stateless IPv4/IPv6 Translation", draft-xli-behave-divi-04 (work in progress), October 2011.

[RFC0792] Postel, J., "Internet Control Message Protocol", STD 5, RFC 792, September 1981.

[RFC0897] Postel, J., "Domain name system implementation schedule", RFC 897, February 1984.

[RFC2473] Conta, A. and S. Deering, "Generic Packet Tunneling in IPv6 Specification", RFC 2473, December 1998.

[RFC2663] Srisuresh, P. and M. Holdrege, "IP Network Address Translator (NAT) Terminology and Considerations", RFC 2663, August 1999.

[RFC2827] Ferguson, P. and D. Senie, "Network Ingress Filtering: Defeating Denial of Service Attacks which employ IP Source Address Spoofing", BCP 38, RFC 2827, May 2000.

[RFC3633] Troan, O. and R. Droms, "IPv6 Prefix Options for Dynamic Host Configuration Protocol (DHCP) version 6", RFC 3633, December 2003.

[RFC4443] Conta, A., Deering, S., and M. Gupta, "Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6) Specification", RFC 4443, March 2006.

[RFC4953] Touch, J., "Defending TCP Against Spoofing Attacks", RFC 4953, July 2007.

[RFC5508] Srisuresh, P., Ford, B., Sivakumar, S., and S. Guha, "NAT Behavioral Requirements for ICMP", BCP 148, RFC 5508, April 2009.

[RFC5961] Ramaiah, A., Stewart, R., and M. Dalal, "Improving TCP's Robustness to Blind In-Window Attacks", RFC 5961, August 2010.

[RFC6056] Larsen, M. and F. Gont, "Recommendations for Transport-Protocol Port Randomization", BCP 156, RFC 6056, January 2011.

[RFC6147] Bagnulo, M., Sullivan, A., Matthews, P., and I. van Beijnum, "DNS64: DNS Extensions for Network Address Translation from IPv6 Clients to IPv4 Servers", RFC 6147, April 2011.

[RFC6219] Li, X., Bao, C., Chen, M., Zhang, H., and J. Wu, "The China Education and Research Network (CERNET) IVI Translation Design and Deployment for the IPv4/IPv6 Coexistence and Transition", RFC 6219, May 2011.

[RFC6269] Ford, M., Boucadair, M., Durand, A., Levis, P., and P. Roberts, "Issues with IP Address Sharing", RFC 6269, June 2011.

[RFC6324] Nakibly, G. and F. Templin, "Routing Loop Attack Using IPv6 Automatic Tunnels: Problem Statement and Proposed Mitigations", RFC 6324, August 2011.

Appendix A. Example of MAP-T translation

Example 1:

Given the MAP domain information and an IPv6 address of an endpoint:

```
IPv6 prefix assigned to the end user:  2001:db8:0012:3400::/56
Basic Mapping Rule:  {2001:db8:0000::/40 (Rule IPv6 prefix),
   192.0.2.0/24 (Rule IPv4 prefix), 16 (Rule EA-bits length)}
Sharing ratio:  256 (16 - (32 - 24) = 8. 2^8 = 256)
PSID offset:  4
```

A MAP node (CE or BR) can via the BMR determine the IPv4 address and port-set as shown below:

```
EA bits offset:  40
IPv4 suffix bits (p)  Length of IPv4 address (32) - IPv4 prefix
   length (24) = 8
IPv4 address  192.0.2.18 (0xc0000212)
PSID start:  40 + p = 40 + 8 = 48
PSID length:  o - p = 16 (56 - 40) - 8 = 8
PSID:  0x34

Port-set-1:  4928, 4929, 4930, 4931, 4932, 4933, 4934, 4935, 4936,
   4937, 4938, 4939, 4940, 4941, 4942, 4943
Port-set-2:  9024, 9025, 9026, 9027, 9028, 9029, 9030, 9031, 9032,
   9033, 9034, 9035, 9036, 9037, 9038, 9039
...  ...
Port-set-15  62272, 62273, 62274, 62275, 62276, 62277, 62278,
   62279, 62280, 62281, 62282, 62283, 62284, 62285, 62286, 62287
```

The BMR information allows a MAP CE also to determine (complete) its IPv6 address within the indicated IPv6 prefix.

```
IPv6 address of MAP-T CE:  2001:db8:0012:3400:00c0:0002:1200:3400
```

Example 2:

Another example can be made of a hypothetical MAP-T BR, configured with the following FMR when receiving a packet with the following characteristics:

IPv4 source address: 1.2.3.4 (0x01020304)
IPv4 source port: 80
IPv4 destination address: 192.0.2.18 (0xc0000212)
IPv4 destination port: 9030

Configured Forwarding Mapping Rule: {2001:db8:0000::/40 (Rule IPv6 prefix), 192.0.2.0/24 (Rule IPv4 prefix), 16 (Rule EA-bits length)}

MAP-T BR Prefix 2001:db8:ffff::/64

The above information allows the BR to derive as follows the mapped destination IPv6 address for the corresponding MAP-T CE, and also the mapped source IPv6 address for the IPv4 source.

IPv4 suffix bits (p) 32 - 24 = 8 (18 (0x12))
PSID length: 8
PSID: 0x34 (9030 (0x2346))

The resulting IPv6 packet will have the following key fields:

IPv6 source address 2001:db8:ffff:0:0001:0203:0400::
IPv6 destination address: 2001:db8:0012:3400:00c0:0002:1200:3400
IPv6 source Port: 80
IPv6 destination Port: 9030

Example 3:

An IPv4 host behind the MAP-T CE (addressed as per the previous examples) corresponding with IPv4 host 1.2.3.4 will have its packets converted into IPv6 using the DMR configured on the MAP-T CE as follows:

```
Default Mapping Rule used by MAP-T CE:  {2001:db8:ffff::/64
(Rule IPv6 prefix), 0.0.0.0/0 (Rule IPv4 prefix), null (BR IPv4
   address)}

IPv4 source address (post NAT44 if present)  192.0.2.18
IPv4 destination address:  1.2.3.4
IPv4 source port (post NAT44 if present):  9030
IPv4 destination port:  80
IPv6 source address of MAP-T CE:
                     2001:db8:0012:3400:00c0:0002:1200:3400
IPv6 destination address:  2001:db8:ffff:0:0001:0203:0400::
```

Authors' Addresses

Xing Li
CERNET Center/Tsinghua University
Room 225, Main Building, Tsinghua University
Beijing 100084
CN

Email: xing@cernet.edu.cn

Congxiao Bao
CERNET Center/Tsinghua University
Room 225, Main Building, Tsinghua University
Beijing 100084
CN

Email: congxiao@cernet.edu.cn

Wojciech Dec
Cisco Systems
Haarlerbergpark Haarlerbergweg 13-19
Amsterdam, NOORD-HOLLAND 1101 CH
Netherlands

Phone:
Email: wdec@cisco.com

Ole Troan
Cisco Systems
Oslo
Norway

Email: ot@cisco.com

Satoru Matsushima
SoftBank Telecom
1-9-1 Higashi-Shinbashi, Munato-ku
Tokyo
Japan

Email: satoru.matsushima@tm.softbank.co.jp

Tetsuya Murakami
IP Infusion
1188 East Arques Avenue
Sunnyvale
USA

Email: tetsuya@ipinfusion.com