Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: March 9, 2013

O. Troan
W. Dec
Cisco Systems
X. Li
C. Bao
CERNET Center/Tsinghua University
S. Matsushima
SoftBank Telecom
T. Murakami
IP Infusion
September 5, 2012

Mapping of Address and Port with Encapsulation (MAP)
draft-ietf-softwire-map-02

Abstract

This document describes a mechanism for transporting IPv4 packets across an IPv6 network, and a generic mechanism for mapping between IPv6 addresses and IPv4 addresses and transport layer ports.

Table of Contents

1. Introduction

Mapping of IPv4 addresses in IPv6 addresses has been described in numerous mechanisms dating back to 1996 [RFC1933]. The Automatic tunneling mechanism described in RFC1933, assigned a globally unique IPv6 address to a host by combining the host's IPv4 address with a well-known IPv6 prefix. Given an IPv6 packet with a destination address with an embedded IPv4 address, a node could automatically tunnel this packet by extracting the IPv4 tunnel end-point address from the IPv6 destination address.

There are numerous variations of this idea, described in 6over4 [RFC2529], 6to4 [RFC3056], ISATAP [RFC5214], and 6rd [RFC5969].

The commonalities of all these IPv6 over IPv4 mechanisms are:

- Automatically provisions an IPv6 address for a host or an IPv6 prefix for a site

- Algorithmic or implicit address resolution of tunnel end point addresses. Given an IPv6 destination address, an IPv4 tunnel endpoint address can be calculated.

- Embedding of an IPv4 address or part thereof into an IPv6 address.

In phases of IPv4 to IPv6 migration, IPv6 only networks will be common, while there will still be a need for residual IPv4 deployment. This document describes a generic mapping of IPv4 to IPv6, and a mechanism for encapsulating IPv4 over IPv6.

Just as the IPv6 over IPv4 mechanisms referred to above, the residual IPv4 over IPv6 mechanism must be capable of:

- Provisioning an IPv4 prefix, an IPv4 address or a shared IPv4 address.

- Algorithmically map between an IPv4 prefix, IPv4 address or a shared IPv4 address and an IPv6 address.

The mapping scheme described here supports encapsulation of IPv4 packets in IPv6 in both mesh and hub and spoke topologies, including address mappings with full independence between IPv6 and IPv4 addresses.

This document describes delivery of IPv4 unicast service across an IPv6 infrastructure. IPv4 multicast is not considered further in this document.

The A+P (Address and Port) architecture of sharing an IPv4 address by distributing the port space is described in [RFC6346]. Specifically section 4 of [RFC6346] covers stateless mapping. The corresponding stateful solution DS-lite is described in [RFC6333]. The motivation for the work is described in [I-D.ietf-softwire-stateless-4v6-motivation].

A companion document defines a DHCPv6 option for provisioning of MAP [I-D.ietf-softwire-map-dhcp]. Other means of provisioning is possible. Deployment considerations are described in [I-D.mdt-softwire-map-deployment].

MAP relies on IPv6 and is designed to deliver production-quality dual-stack service while allowing IPv4 to be phased out within the SP network. The phasing out of IPv4 within the SP network is independent of whether the end user disables IPv4 service or not. Further, "Greenfield"; IPv6-only networks may use MAP in order to deliver IPv4 to sites via the IPv6 network.

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 virtual link. 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 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: The 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 a 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.

4. Architecture

The MAP mechanism uses existing standard building blocks. The existing NAT44 on the CE is used with additional support for restricting transport protocol ports, ICMP identifiers and fragment identifiers to the configured port set. MAP supports the

encapsulation mode specified in [RFC2473]. In addition MAP specifies an algorithm to do "address resolution" from an IPv4 address and port to an IPv6 address. This algorithmic mapping is specified in Section 5.

A full IPv4 address or IPv4 prefix can be used like today, e.g. for identifying an interface or as a DHCP pool. A shared IPv4 address on the other hand, MUST NOT be used to identify an interface. While it is theoretically possible to make host stacks and applications port-aware, that is considered a too drastic change to the IP model [RFC6250].

The MAP architecture described here, restricts the use of the shared IPv4 address to only be used as the global address (outside) of the NAPT [RFC2663] running on the CE. The NAPT MUST in turn be connected to a MAP aware forwarding function, that does encapsulation/decapsulation of IPv4 packets in IPv6.

When MAP is used to provision a full IPv4 address or an IPv4 prefix to the CE, these restrictions do not apply.

For packets outbound from the private IPv4 network, the CE NAPT MUST translate transport identifiers (e.g. TCP and UDP port numbers) so that they fall within the assigned CE's port-range.

The forwarding function uses the Rules table to make forwarding decisions. The table consists of the mapping rules. An entry in the table consists of an IPv4 prefix and PSID.

```
         User N
       Private IPv4
      |  Network
      |
   O--+---------------O
   |  |  MAP CE       |
   | +-----+--------+ |
   | NAPT44|  MAP   | |
   | +-----+      | | |\     ,-------.                     .------.
   |       +--------+ | \ ,-'         `-.                ,-'       `-.
   O------------------O  /              \   O---------O /   Public   \
                         /   IPv6 only   \  |  MAP    |/     IPv4     \
                        (    Network      --+  Border +-   Network     )
                         \  (MAP Domain) /  |  Relay  |\              /
   O------------------O  \              /   O---------O \             /
   |    MAP   CE      |  /".         ,-'                `-.       ,-'
   | +-----+--------+ | /   `----+--'                      ------'
   | NAPT44|  MAP   | |/
   | +-----+        | |
   |   |   +--------+ |
   O---.--------------O
       |
         User M
       Private IPv4
         Network
```

Figure 1: Network Topology

The MAP BR is responsible for connecting external IPv4 networks to the IPv4 nodes in one or more MAP domains.

5. Mapping Algorithm

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. Additional mapping rules are specified to allow for multiple different IPv4 sub-nets 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.

There are three types of mapping rules:

1. Basic Mapping Rule - used for IPv4 prefix, address or port set assignment. There can only be one Basic Mapping Rule per End-user IPv6 prefix. The Basic Mapping Rule is used to configure the MAP IPv6 address or prefix.

   * 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. The Basic Mapping Rule is also a Forwarding Mapping Rule. Each Forwarding Mapping Rule will result in an entry in the Rules table for the Rule IPv4 prefix. The FMR consists of the same parameters as the BMR.

3. Default Mapping Rule - used for destinations outside the MAP domain. A 0.0.0.0/0 entry is installed in the Rules table for this rule.

   * IPv6 address of BR

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 Rules table. 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 assigned to an interface of the MAP node and is used to terminate all MAP traffic being sent or received to the node.

Port-aware IPv4 entries in the Rules table 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. Port mapping algorithm

The port mapping algorithm is used in domains whose rules allow IPv4 address sharing. Different Port-Set Identifiers (PSID) MUST have non-overlapping port-sets. The two extreme cases 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. The port mapping algorithm proposed here is called the Generalized Modulus Algorithm (GMA) and supports both these cases.

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

1. The port number (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.

5.1.1. 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
   +---------------+----------+------+------------------+
   |                     P                               |
   ----------------+----------------+-------------------+
   |        A (j)  |   PSID (K)      |        M  (i)     |
   +---------------+----------+------+------------------+
   |<----a bits--->|<-----k bits---->|<------m bits----->|
```

Figure 2: 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 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.

5.1.2. 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)

5.1.3. Port Algorithm Provisioning Considerations

The number of offset bits (a) and excluded ports are optionally provisioned via the "Rule Port Mapping Parameters" in the Basic Mapping Rule.

The defaults are:

o Excluded ports : 0-4095

o Offset bits (a) : 4

To simplify the GMA port mapping algorithm the defaults are chosen so that the PSID field starts on a nibble boundary and the excluded port range (0-1023) is extended to 0-4095.

5.2. Basic mapping rule (BMR)

```
 |     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 3: IPv6 address format

The Embedded Address bits (EA bits) are unique per end user within a Rule IPv6 prefix. The Rule IPv6 prefix is the part of the End-user IPv6 prefix that is common among all CEs using the same Basic Mapping Rule within the MAP domain. The EA bits encode the CE specific IPv4 address and port information. The EA bits can contain a full or part of an IPv4 prefix or address, and in the shared IPv4 address case contains a Port-Set Identifier (PSID).

The MAP IPv6 address is created by concatenating the End-user IPv6 prefix with the MAP subnet-id and the interface-id as specified in Section 6.

The MAP subnet ID is defined to be the first subnet (all bits set to zero). A MAP node MUST reserve the first IPv6 prefix in an End-user IPv6 prefix for the purpose of MAP.

The MAP IPv6 is created by combining the End-User IPv6 prefix with the all zeros subnet-id and the MAP IPv6 interface identifier.

```
                        Shared IPv4 address:


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

Figure 4: Shared IPv4 address

```
                       Complete IPv4 address:


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

Figure 5: Complete IPv4 address

```
                            IPv4 prefix:


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

Figure 6: 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 (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,
```

5.3. Forwarding mapping rule (FMR)

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

On forwarding an IPv4 packet, a best matching prefix look up is done in the Rules table 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  --->|
```

Figure 7: Deriving of MAP IPv6 address

Example:

```
     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
```

5.4. Default mapping rule (DMR)

The Default Mapping rule is used to reach IPv4 destinations outside of the MAP domain. Traffic using this rule will be sent from a CE to a BR.

The DMR consist of the IPv6 address of the BR.

6. The IPv6 Interface Identifier

The 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 Figure 8.

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

Figure 8

In the case of an IPv4 prefix, the IPv4 address field is right-padded with zeroes 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.

If the End-user IPv6 prefix length is larger than 64, the most significant parts of the interface identifier is overwritten by the prefix.

7. MAP Configuration

For a given MAP domain, the BR and CE MUST be configured with the following MAP elements. The configured values for these elements are identical for all CEs and BRs within a given MAP domain.

- The End-User IPv6 prefix (Part of the normal IPv6 provisioning).
- The Basic Mapping Rule and optionally the Forwarding Mapping Rules, including the Rule IPv6 prefix, Rule IPv4 prefix, and Length of EA bits
- The Default Mapping Rule with the BR IPv6 address
- Hub and spoke mode or Mesh mode. (If all traffic should be sent to the BR, or if direct CE to CE traffic should be supported).

7.1. MAP CE

The MAP elements are set to values that are the same across all CEs within a MAP domain. The values may be configured in a variety of manners, including provisioning methods such as the Broadband Forum's "TR-69" Residential Gateway management interface, an XML-based object retrieved after IPv6 connectivity is established, or manual configuration by an administrator. This document describes how to configure the necessary parameters via a single DHCPv6 option. A CE that allows IPv6 configuration by DHCP SHOULD implement this option. Other configuration and management methods may use the format described by this option for consistency and convenience of implementation on CEs that support multiple configuration methods.

The only remaining provisioning information the CE requires in order to calculate the MAP IPv4 address and enable IPv4 connectivity is the IPv6 prefix for the CE. The End-user IPv6 prefix is configured as part of obtaining IPv6 Internet access.

A single MAP CE MAY be connected to more than one MAP domain, just as any router may have more than one IPv4-enabled service provider facing interface and more than one set of associated addresses assigned by DHCP. Each domain a given CE operates within would require its own set of MAP configuration elements and would generate its own IPv4 address.

The MAP DHCP option is specified in [I-D.ietf-softwire-map-dhcp].

7.2. MAP BR

The MAP BR MUST be configured with the same MAP elements as the MAP CEs operating within the same domain.

For increased reliability and load balancing, the BR IPv6 address MAY be an anycast address shared across a given MAP domain. As MAP is stateless, any BR may be used at any time. If the BR IPv6 address is anycast the relay MUST use this anycast IPv6 address as the source address in packets relayed to CEs.

Since MAP uses provider address space, no specific routes need to be advertised externally for MAP to operate, neither in IPv6 nor IPv4 BGP. However, if anycast is used for the MAP IPv6 relays, the anycast addresses must be advertised in the service provider's IGP.

7.3. Backwards compatibility

A MAP-E CE provisioned with only a Default Mapping Rule, and with no IPv4 address and port range configured by other means, MUST disable

its NAT44 functionality. This characteristic makes a MAP CE compatible with DS-Lite [RFC6333] AFTRs, whose addresses are configured as the MAP BR.

8. Forwarding Considerations

Figure 1 depicts the overall MAP architecture with IPv4 users (N and M) networks connected to a routed IPv6 network.

MAP supports Encapsulation mode as specified in [RFC2473].

A MAP CE forwarding IPv4 packets from the LAN performs NAT44 functions first and create appropriate NAT44 bindings. The resulting IPv4 packets MUST contain the source IPv4 address and source transport number defined by MAP. The resulting IPv4 packet is forwarded to the CE's MAP forwarding function. The IPv6 source and destination addresses MUST then be derived as per Section 5 of this draft.

A MAP CE receiving an IPv6 packet to its MAP IPv6 address sends this packet to the CE's MAP function. All other IPv6 traffic is forwarded as per the CE's IPv6 routing rules. The resulting IPv4 packet is then forwarded to the CE's NAT44 function where the destination port number MUST be checked against the stateful port mapping session table and the destination port number MUST be mapped to its original value.

8.1. Receiving rules

The CE SHOULD check that MAP 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).

8.2. MAP BR

8.2.1. IPv6 to IPv4

A MAP BR receiving IPv6 packets selects a best matching MAP domain 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 BR prefixes or FMR prefix(es). The selected MAP rule allows the BR to determine the EA-bits from the source IPv6 address. The BR MUST perform a validation of the consistency of the source IPv6 address and source port number for the packet using BMR. If the packets source port number is found to be outside the range allowed for this CE 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).

9. ICMP

ICMP message should be supported in MAP domain. Hence, the NAT44 in MAP CE must implement the behavior for ICMP message conforming to the best current practice documented in [RFC5508].

If a MAP CE receives an ICMP message having ICMP identifier field in ICMP header, NAT44 in the MAP CE must rewrite this field to a specific value assigned from the port-set. BR and other CEs must handle this field similar to the port number in the TCP/UDP header upon receiving the ICMP message with ICMP identifier field.

If a MAP node receives an ICMP error message without the ICMP identifier field for errors that is detected inside a IPv6 tunnel, a node should relay the ICMP error message to the original source. This behavior should be implemented conforming to the section 8 of [RFC2473].

10. 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.

10.1. Fragmentation in the MAP domain

Encapsulating an IPv4 packet to carry it across the MAP domain will increase its size (40 bytes). 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 on MAP domain entry is described in section 7.2 of [RFC2473]

The use of an anycast source address could lead to any ICMP error message generated on the path being sent to a different BR. Therefore, using dynamic tunnel MTU Section 6.7 of [RFC2473] is

subject to IPv6 Path MTU black-holes.

Multiple BRs using the same anycast source address could send fragmented packets to the same CE at the same time. If the fragmented packets from different BRs happen to use the same fragment ID, incorrect reassembly might occur.

10.2. 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.

10.3. 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.

11. NAT44 Considerations

The NAT44 implemented in the MAP CE SHOULD conform with the behavior and best current practice documented in [RFC4787], [RFC5508], [RFC5382] and [RFC5383]. In MAP address sharing mode (determined by the MAP domain/rule configuration parameters) the operation of the NAT44 MUST be restricted to the available port numbers derived via the basic mapping rule.

12. IANA Considerations

This specification does not require any IANA actions.

13. 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. Practicalities of these methods are discussed in Section 5.9 of [I-D.dec-stateless-4v6].

[RFC6269] outlines general issues with IPv4 address sharing.

14. Contributors

This document is the result of the IETF Softwire MAP design team effort and numerous previous individual contributions in this area:

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

Gang Chen (China Mobile)
53A,Xibianmennei Ave. Beijing 100053 P.R.China
Email: chengang@chinamobile.com

Yu Zhai
CERNET Center/Tsinghua University
Room 225, Main Building, Tsinghua University
Beijing 100084
CN
Email: jacky.zhai@gmail.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

Rajiv Asati (Cisco Systems)
7025-6 Kit Creek Road Research Triangle Park NC 27709 USA
Email: rajiva@cisco.com

15. Acknowledgements

This document is based on the ideas of many, including Mohamed Boucadair, Gang Chen, Maoke Chen, Wojciech Dec, Xiaohong Deng, Jouni Korhonen, Tomasz Mrugalski, Jacni Qin, Chunfa Sun, Qiong Sun, and Leaf Yeh. The authors want in particular to recognize 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.

16. References

16.1. Normative References

[I-D.ietf-softwire-map-dhcp]
Mrugalski, T., Troan, O., Bao, C., Dec, W., and L. Yeh, "DHCPv6 Options for Mapping of Address and Port", draft-ietf-softwire-map-dhcp-01 (work in progress), August 2012.

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

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

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

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

16.2. 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-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.

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

[RFC1933] Gilligan, R. and E. Nordmark, "Transition Mechanisms for IPv6 Hosts and Routers", RFC 1933, April 1996.

[RFC2529] Carpenter, B. and C. Jung, "Transmission of IPv6 over IPv4 Domains without Explicit Tunnels", RFC 2529, March 1999.

[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.

[RFC3056] Carpenter, B. and K. Moore, "Connection of IPv6 Domains via IPv4 Clouds", RFC 3056, February 2001.

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

[RFC4787] Audet, F. and C. Jennings, "Network Address Translation (NAT) Behavioral Requirements for Unicast UDP", BCP 127, RFC 4787, January 2007.

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

[RFC5214] Templin, F., Gleeson, T., and D. Thaler, "Intra-Site Automatic Tunnel Addressing Protocol (ISATAP)", RFC 5214, March 2008.

[RFC5382] Guha, S., Biswas, K., Ford, B., Sivakumar, S., and P. Srisuresh, "NAT Behavioral Requirements for TCP", BCP 142, RFC 5382, October 2008.

[RFC5383] Gellens, R., "Deployment Considerations for Lemonade-Compliant Mobile Email", BCP 143, RFC 5383, October 2008.

[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.

[RFC5969] Townsley, W. and O. Troan, "IPv6 Rapid Deployment on IPv4 Infrastructures (6rd) -- Protocol Specification", RFC 5969, August 2010.

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

[RFC6250] Thaler, D., "Evolution of the IP Model", RFC 6250, 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.

[RFC6333] Durand, A., Droms, R., Woodyatt, J., and Y. Lee, "Dual-Stack Lite Broadband Deployments Following IPv4 Exhaustion", RFC 6333, August 2011.

Appendix A. Example of MAP

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 CE:  2001:db8:0012:3400:00c0:0002:1200:3400
```

Example 2:

Another example can be made of a hypothetical MAP 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 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 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 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 CE as follows:

```
Default Mapping Rule used by MAP 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 CE:
                     2001:db8:0012:3400:00c0:0002:1200:3400
IPv6 destination address:  2001:db8:ffff:0:0001:0203:0400::
```

Authors' Addresses

Ole Troan
Cisco Systems
Philip Pedersens vei 1
Lysaker 1366
Norway

Email: ot@cisco.com

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

Phone:
Email: wdec@cisco.com

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

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

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

Tetsuya Murakami
IP Infusion
1188 East Arques Avenue
Sunnyvale
USA

Email: tetsuya@ipinfusion.com