Network Working Group O. Troan, Ed.
Internet-Draft cisco
Intended status: Standards Track November 25, 2011
Expires: May 28, 2012

Mapping of Address and Port (MAP)
draft-mdt-softwire-mapping-address-and-port-02

Abstract

This document describes a generic mechanism for mapping between an IPv4 prefix, address or parts thereof, and transport layer ports and an IPv6 prefix or address.

Table of Contents

1. Introduction

The mechanism of mapping IPv4 addresses in IPv6 address 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 an 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 differences are the use of well known IPv6 prefixes, or Service Provider assigned IPv6 prefixes, and the exact position of the IPv4 bits embedded in the IPv6 address. Teredo [RFC4380] added a twist to this to achieve NAT traversal by also encoding transport layer ports into the IPv6 address. 6rd, to achieve more efficient encoding, allowed for only the suffix of an IPv4 address to be embedded, with the IPv4 prefix being deducted from other provisioning mechanisms.

NAT-PT [RFC2766](deprecated) combined with a DNS ALG used address mapping to put NAT state, namely the IPv6 to IPv4 binding encoded in an IPv6 address. This characteristic has been inherited by NAT64 [RFC6146] and DNS64 [RFC6147] which rely on an address format defined in [RFC6052]. [RFC6052] specifies the algorithmic translation of an IPv6 address to IPv4 address. In particular, [RFC6052] specifies the address format to build IPv4-converted and IPv4-translatable IPv6 addresses. RFC6052 discusses the transport of the port set information in an IPv4-embedded IPv6 address but the conclusion was the following (excerpt from [RFC6052]):

"There have been proposals to complement stateless translation with a port range feature. Instead of mapping an IPv4 address to exactly one IPv6 prefix, the options would allow several IPv6 nodes to share an IPv4 address, with each node managing a different set of ports.

If a port set extension is needed, could be defined later, using bits currently reserved as null in the suffix."

The commonalities of all these mechanisms are:

- o Provisions an IPv6 address for a host or an IPv6 prefix for a site

- o Algorithmic or implicit address resolution for tunneling or encapsulation. Given an IPv6 destination address, an IPv4 tunnel endpoint address can be calculated. Likewise for translation, an IPv4 address can be calculated from an IPv6 destination address and vice versa.

- o Embedding of an IPv4 address or part thereof and optionally transport layer ports 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 more generic mapping of IPv4 to IPv6 that can be used both for encapsulation (IPv4 over IPv6) and for translation between the two protocols.

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

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

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

The unified mapping scheme described here supports translation mode, encapsulation mode, in both mesh and hub and spoke topologies.

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

In particular the architecture of a shared IPv4 address by distributing the port space is described in [RFC6346]. The corresponding stateful solution DS-lite is described in [RFC6333]. The motivation for work is described in

[I-D.ietf-softwire-stateless-4v6-motivation].

A companion document defines a DHCPv6 option for provisioning of MAP [I-D.mdt-softwire-map-dhcp-option].

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: A set of 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 MAP node in the domain has the same set of rules.

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.

Shared IPv4 address: An IPv4 address that is shared among multiple CEs. Each node has a separate part of the transport layer port space; denoted as a port set. Only ports that belong to the assigned port set can be used for communication.

End-user IPv6 prefix: The IPv6 prefix assigned to an End-user CE by other means than MAP itself.

MAP IPv6 address: The IPv6 address used to reach the MAP function of a CE from other CE's and from BR's.

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

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.

IPv4 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. 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 itself with an IPv4 address, IPv4 prefix or shared IPv4 address from an End-user IPv6 prefix. 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 can be specified to allow for e.g. multiple different IPv4 subnets to exist within the domain. Additional mapping rules are recognized by having a Rule IPv6 prefix different from the base End-user IPv6 prefix.

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

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 a route in a conceptual routing table for the Rule IPv4 prefix.

   * 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. A 0.0.0.0/0 route is installed in the conceptual routing table for this rule.

   * Rule IPv6 prefix (including prefix length)

   * Rule BR IPv4 address

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 (or prefix) is formed from the BMR Rule IPv6 prefix. This address MUST be assigned to an interface of the MAP node and is used as to terminate all MAP traffic being sent or received to the node.

Routes in the conceptual routing table are installed for all the Forwarding Mapping Rules and an IPv4 default route for the Default Mapping Rule.

In the hub and spoke mode, all traffic should be forwarded using the Default Mapping Rule. Hub and spoke mode is achieved with a BMR IPv4 rule prefix length of 32 and no further Forwarding Mapping Rules.

4.1. Port mapping algorithm

Different Port Set Identificators (PSID) MUST have non-overlapping port sets. The two extreme cases are: (1) the port number is not contiguous for each PSID, but uniformly distributed across the whole port range (0-65535); (2) the port number is 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 well-known port numbers (0 - 4096) 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.

4.1.1. 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----->|
                |k-c |<--c bits-->|<------m bits----->|
```

Figure 1: 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, A MUST be larger than 0. This ensures that the algorithm excludes the well known ports. For a = 0, A MAY be 0 to allow for the provisioning of the well known ports.

When m = 0, GMA becomes a modulo operation. When a = 0, GMA becomes division operation.

4.1.2. GMA examples

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

```
           Port set-1                 Port set-2
PSID=0   | 4096, 4097, 4098, 4099, | 8192, 8193, 8194, 8195, | 12288
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=127 | 4604, 4605, 4606, 4607, | 8700, 8701, 8702, 8703, | ....
```

Figure 2: Example

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]
```

Figure 3: Example with offset = 0 (a = 0)

4.1.3. GMA Provisioning Considerations

The sharing ratio (R), the PSID (K) and the PSID length are derived from existing information.

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

4.1.4. Features of the Algorithm

The GMA algorithm has the following features:

1. There is no waste of the port numbers, except the well-known ports.

2. The algorithm is flexible, the control parameters are PSID offset (a) and PSID length (c) / Sharing ratio.

3. The algorithm is simple to perform effectively.

4. It allows Service Providers to define their own address sharing ratio, the theoretical value is from 1:1 to 1:65536 and a more practical value is from 1:1 to 1:4096.

5. It supports differentiated port ranges per mapping rule.

6. It support exclusion of the well-known ports.

7. It supports assigning the well-known ports to a CE.

8. It supports legacy RTP/RTCP compatibility.

4.2. Basic mapping rule (BMR)

```
|     n bits         |  o bits    | m bits  |   128-n-o-m bits      |
+--------------------+-----------+---------+------------+----------+
| Domain IPv6 prefix |  EA bits  |subnet ID|     interface ID      |
+--------------------+-----------+---------+-----------------------+
|<---  End-user IPv6 prefix  --->|
```

Figure 4: IPv6 address format

The Embedded Address bits (EA bits) are unique per end user within a Domain IPv6 prefix. The Domain 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. There MUST be a Basic Mapping Rule with a Rule IPv6 prefix equal to the Domain IPv6 prefix. 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 ~0 and the interface-id as specified in the Interface-id section.

Shared IPv4 address:

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

Figure 5

Complete IPv4 address:

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

Figure 6

IPv4 prefix:

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

Figure 7

If only a part of the IPv4 address/prefix is encoded in the EA bits, the Domain 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 from the EA bits, are concatenated with the Domain 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 in the Rule EA-bits length parameter.

If o + r < 32, then an IPv4 prefix is assigned. The IPv4 prefix length is equal to r bits + Rule EA-bits length.

If o + r is equal to 32, then a full IPv4 address is to be assigned. The address is created by concatenating the Domain 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 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.

Example:

```
  Given:
   End-user IPv6 prefix: 2001:db8:0012:34::/56
   Domain IPv6 prefix:   2001:db8:00::/40
   IPv4 prefix:          192.0.2.0/24
   Basic Mapping Rule:   {2001:db8:00::/40, 192.0.2.0/24, 256, 6}
```

We get IPv4 address and port set:

```
      EA bits offset:       40
      IPv4 suffix bits (p): 32 - 24 = 8
      IPv4 address:         192.0.2.18

      PSID start:           40 + p = 40 + 8 = 48
      PSID length:          o - p = log2(256) - 8 = 8.
      PSID:                 0x34.
```

4.3. Forwarding mapping rule (FMR)

On adding a FMR rule an IPv4 route is installed the conceptual routing table for the Rule IPv4 prefix.

On forwarding an IPv4 packet a lookup is done in the routing table and the correct FMR is used.


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

Figure 8

The subnet ID for MAP is defined to be ~0. I.e. the last subnet in an End-user IPv6 prefix allocation is used for MAP. A MAP node MUST reserve the topmost IPv6 prefix in a End-user IPv6 prefix for the purpose of MAP. This prefix MUST NOT be used for native IPv6 traffic.

Example:

```
  Given:
   IPv4 destination address: 192.0.2.18
   IPv4 destination port:    1232
   Forwarding Mapping Rule:  {2001:db8:00::/40, 192.0.2.0/24,
                              Sharing ratio: 256, PSID offset: 4}

  We get IPv6 address:
   IPv4 suffix bits (p): 32 - 24 = 8 (18)
   PSID length:          8 (sharing ratio)
   PSID:                 0x17 (1232)
   EA bits:              0x1217
   IPv6 address:         2001:db8:0012:17FF:<interface-identifier>
```

4.4. 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 Rule IPv4 prefix in the DMR is: 0.0.0.0/0. The Rule IPv6 prefix is the IPv6 address or prefix of the BR. Which is used is dependent on the mode used. For example translation requires that the IPv4 destination address is encoded in the BR IPv6 address, so only a prefix is used in the DMR to allow for a generated interface identifier. For the encapsulation mode the Rule IPv6 prefix can be the full IPv6 address of the BR.

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

An example of a DMR is:

Default Mapping Rule: {2001:db8:0001:0000:<interface-id>:/128, 0.0.0.0/0, BR IPv4 address: 192.0.2.1, }

In most implementations of a routing table, the next-hop address must be of the same address family as the prefix. To satisfy this requirement a BR IPv4 address is included in the rule. Giving a default route in the routing table:

```
0.0.0.0 -> 192.0.2.1, MAP-Interface0
```

5. The IPv6 Interface Identifier

In an encapsulation solution, an IPv4 address and port is mapped to an IPv6 address. This is the address of the tunnel end point of the receiving MAP CE. For traffic outside the MAP domain, the IPv6 tunnel end point address is the IPv6 address of the BR. The interface-id used for all MAP nodes in the domain MUST be deterministic.

When translating, the destination IPv4 address is translated into a corresponding IPv6 address. In the case of traffic outside of the MAP domain, it is translated to the BR's IPv6 prefix. For the BR to be able to reverse the translation, the full destination IPv4 address must be encoded in the IPv6 address. The same thing applies if an IPv4 prefix is encoded in the IPv6 address, then the reverse translator needs to know the full destination IPv4 address, which has to be encoded in the interface-id.

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


```
<-8-><-------- L>=32 ------><48-L><8->
+---+----------------+------+-----+---+
| u |  IPv4 address  | PSID |  0  | L |
+---+----------------+------+-----+---+
```

Figure 9

The L field denotes the length of the IPv4 address, IPv4 prefix or shared IPv4 address. In the case of an full IPv4 address L = 32, in case of an IPv4 prefix L < 32, in the case of an shared IPv4 address 32 < L <= 48.

If the End-user IPv6 prefix length is larger than 64, the most significant parts of the interface identifier is overwritten by the prefix. For translation mode the End-user IPv6 prefix MUST be 64 or shorter.

6. IANA Considerations

This specification does not require any IANA actions.

7. Security Considerations

There are no new security considerations pertaining to this document.

8. Contributors

The members of the MAP design team are:

Congxiao Bao, Mohamed Boucadair, Gang Chen, Maoke Chen, Wojciech Dec, Xiaohong Deng, Jouni Korhonen, Xing Li, Satoru Matsushima, Tomasz Mrugalski, Tetsuya Murakami, Jacni Qin, Necj Scoberne, Qiong Sun, Tina Tsou, Dan Wing, Leaf Yeh and Jan Zorz.

9. Acknowledgements

The authors would like to thank Guillaume Gottard.

10. References

10.1. Normative References

[I-D.mdt-softwire-map-dhcp-option]
Mrugalski, T., Boucadair, M., Troan, O., Deng, X., and C. Bao, "DHCPv6 Options for Mapping of Address and Port", draft-mdt-softwire-map-dhcp-option-01 (work in progress), October 2011.

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

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

10.2. Informative References

[I-D.ietf-softwire-stateless-4v6-motivation]
Boucadair, M., Matsushima, S., Lee, Y., Bonness, O., Borges, I., and G. Chen, "Motivations for Stateless IPv4 over IPv6 Migration Solutions", draft-ietf-softwire-stateless-4v6-motivation-00 (work in progress), September 2011.

[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.

[RFC2766] Tsirtsis, G. and P. Srisuresh, "Network Address Translation - Protocol Translation (NAT-PT)", RFC 2766, February 2000.

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

[RFC4380] Huitema, C., "Teredo: Tunneling IPv6 over UDP through Network Address Translations (NATs)", RFC 4380, February 2006.

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

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

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

[RFC6146] Bagnulo, M., Matthews, P., and I. van Beijnum, "Stateful NAT64: Network Address and Protocol Translation from IPv6 Clients to IPv4 Servers", RFC 6146, April 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.

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

Author's Address

Ole Troan (editor)
cisco
Oslo
Norway

Email: ot@cisco.com