Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: May 3, 2012

O. Troan, Ed.
cisco

October 31, 2011

Mapping of Address and Port (MAP)
draft-mdt-softwire-mapping-address-and-port-01

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.

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

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

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

This Internet-Draft will expire on May 3, 2012.

Table of Contents

1. Introduction

The mechanism of mapping IPv4 addresses in IPv6 address has been described in numerous mechanisms dating back to [RFC1933] from 1996. The Automatic tunneling mechanism described in RFC1933, assigned a globally unique IPv6 address to a host by combining the hosts 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], 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 an IPv4 address suffix 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 suffix to be embedded, with the deducted from other provisioning mechanisms. DNS ALG used address IPv4 binding encoded in it a corresponding IPv4 address, and vice versa. 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:

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

- Embedding of an IPv4 address or part thereof and optionally transport layer ports into an IPv6 address.

In the later 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:

- 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 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.

Other work that has motivated the work on a unified mapping mechanism for translation and encapsulation are:
[I-D.sun-softwire-stateless-4over6]
[I-D.murakami-softwire-4v6-translation]
[I-D.despres-softwire-4rd-addmapping]
[I-D.chen-softwire-4v6-add-format] [I-D.bcx-address-fmt-extension]
[I-D.mrugalski-dhc-dhcpv6-4rd]
[I-D.boucadair-dhcpv6-shared-address-option]
[I-D.despres-softwire-sam] [I-D.chen-softwire-4v6-pd]
[I-D.boucadair-softwire-stateless-requirements]
[I-D.dec-stateless-4v6] [I-D.boucadair-behave-ipv6-portrange]
[I-D.bsd-softwire-stateless-port-index-analysis]
[I-D.despres-softwire-stateless-analysis-tool]
[I-D.xli-behave-divi-pd] [I-D.murakami-softwire-4rd].

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]

Outstanding issues, Requirements and deployment considerations are temporarily kept in Appendix A to D. The appendixes are in no way to be considered normative.

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. In a residential broadband deployment, this type of device is sometimes referred to as a "Residential Gateway" (RG) or "Customer Premises Equipment" (CPE). A typical MAP CE adopting MAP rules will serve a residential site with one WAN side interface, 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 map from an End-user IPv6 prefix to an IPv4 prefix, address 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 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.

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.

   * 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 RIB 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 RIB 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.

Routes in the conceptual RIB 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.

4.1. Port mapping algorithm

Several port mapping algorithms have been proposed with their own set of advantages and disadvantages. Since different 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 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 = (1024 / M) / R to ((65536 / M) / R) - 1, if the well-known port numbers (0 - 1024) 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:

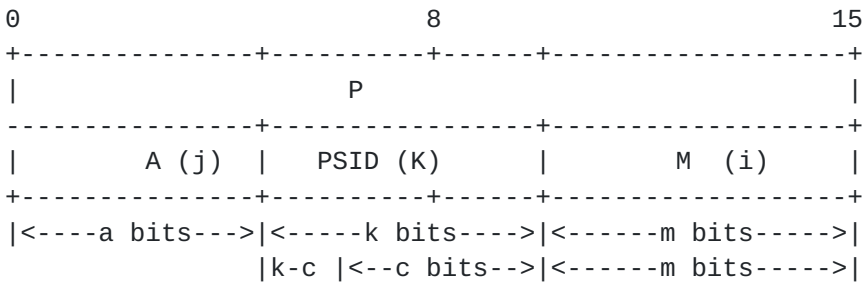


Figure 1: Bit representation

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.

Note that in above figure there is a PSID prefix length (c). Based on this definition, PSID can also be represented in "CIDR style" and more ports can be assigned to a single CE when PSID prefix length (c < k).

When m = 0, GMA becomes a modulo operation. When a = 0, GMA becomes division operation. The port mapping algorithm in [I-D.despres-softwire-4rd-addmapping] can be represented by the algorithm usng a=4 and each PSID may have different prefix length c).

4.1.2. GMA examples

For example, for R=128, M=4,

```
          Port set-1                Port set-2
PSID=0   | 1024, 1025, 1026, 1027, | 1536, 1537, 1538, 1539, | 2048
PSID=1   | 1028, 1029, 1030, 1031, | 1540, 1541, 1542, 1543, | ....
PSID=2   | 1032, 1033, 1034, 1035, | 1544, 1545, 1546, 1547, | ....
PSID=3   | 1036, 1037, 1038, 1039, | 1548, 1549, 1550, 1551, | ....
...
PSID=127 | 1532, 1533, 1534, 1535, | 2044, 2045, 2046, 2047, | ....
```

Figure 2: Example

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:

- Excluded ports : 0-1023
- Offset bits (A) : 6

The defaults of Offset bits (A), which determines excluded ports, remains to be chosen. At least if MAP and native-IPv6 prefixes are the same, two values are considered: 6 and 4. With offset=6, there are 1024 excluded ports, but the maximum sharing ratio is less than the requirement of R-4 (1024). With offset=4, compliance with R-4 is ensured, but there are 4096 excluded ports, which reduces by 4.8% the number of non-well-known ports that can be unused 4096-1024)/(65536-1024). Comparative merits of R-4 compliance and full optimization of port-set sizes remain to be evaluated. If MAP and native-IPv6 prefixes are different, having a different default, e.g. offset=0 has also been proposed.

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 sharing ratio (R), the continue port range (M) and PSID prefix length (c).

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 deployments using differentiated port ranges.

6. It could support differentiated port ranges within a single shared IPv4 address, depending on the IPv6 format chosen (see Appendix A).

7. It support excluding the well known ports 0-1023.

8. It supports assigning well known ports to a CE.

9. It supports legacy RTP/RTCP compatibility.

4.2. Basic mapping rule (BMR)

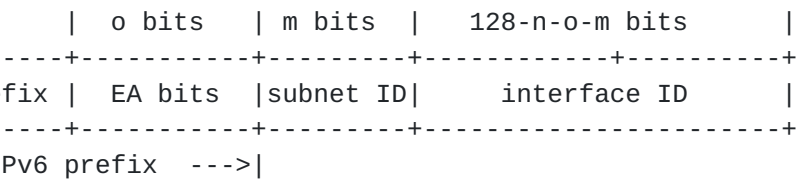


Figure 3: IPv6 address format

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).

Shared IPv4 address:

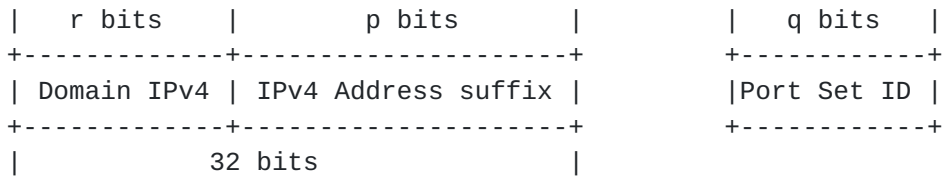


Figure 4

Complete IPv4 address:

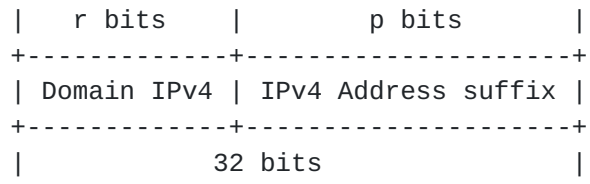


Figure 5

IPv4 prefix:

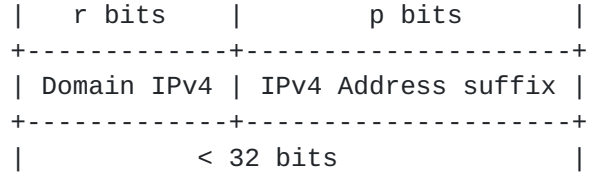


Figure 6

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.

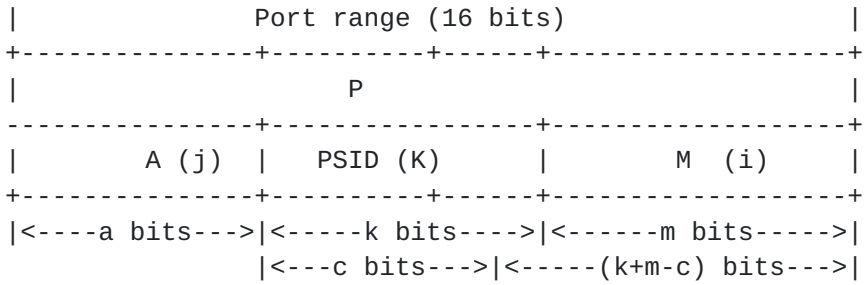


Figure 7

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 RIB (conceptual) for the Rule IPv4 prefix.

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

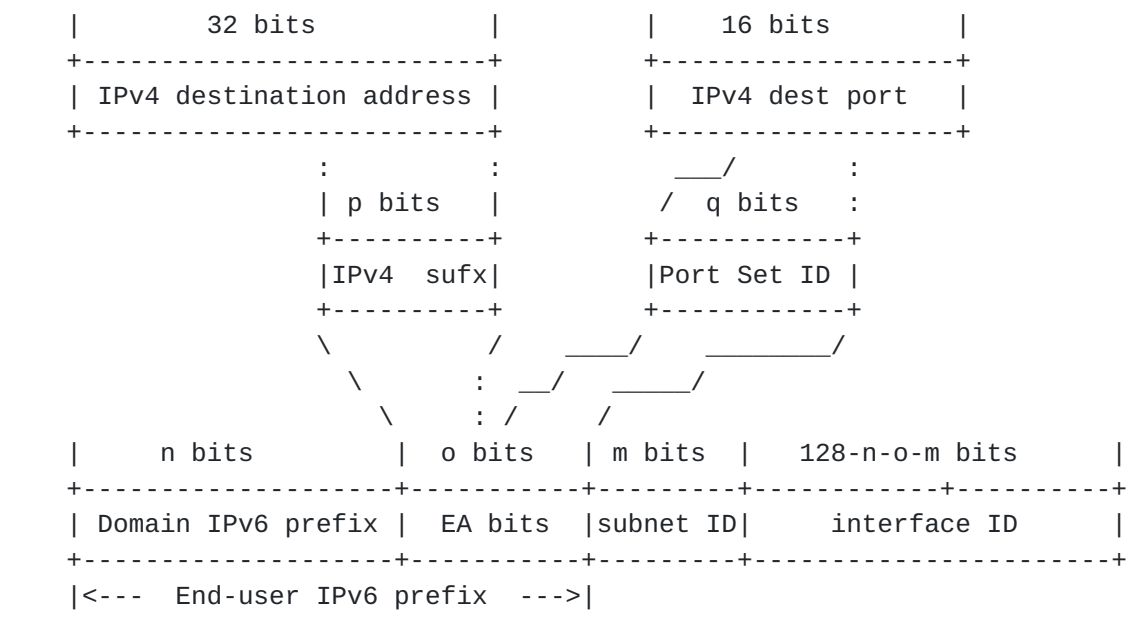


Figure 8

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: 6}

We get IPv6 address:
IPv4 suffix bits (p): 32 - 24 = 8 (18)
PSID length: 8 (sharing ratio)
PSID: 0x34 (1232)
EA bits: 0x1234
IPv6 address: 2001:db8:0012:34FF:<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.

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 RIB, 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 RIB:

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

5. Use of 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. As long as the interface-id is well known or provisioned and the same for all MAP nodes, it can be any interface identifier. E.g. ::1.

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.

There are multiple proposals for how to encode the IPv4 address, and if also the destinatin port or PSID should also be included. A couple of the proposals are shown in the figure below.

Note: 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. The format finally agreed upon here, will apply for both encapsulation and translation.

Existing IANA assigned format [RFC5342]:

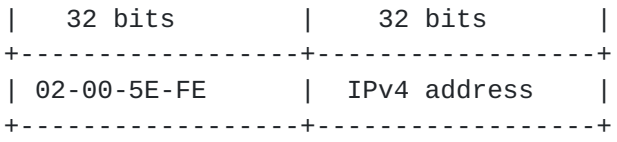


Figure 9

Parsable format including the extended IPv4 prefix length (L) and PSID:

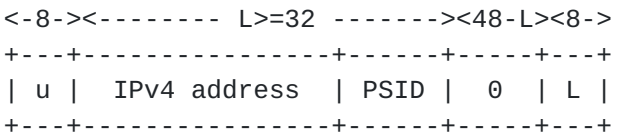


Figure 10

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

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, Remi Despres, Jouni Korhonen, Xing Li, Satoru Matsushima, Tomasz Mrugalski, Tetsuya Murakami, Jacni Qin, Qiong Sun, Tina Tsou, Dan Wing, Leaf Yeh and Jan Zorz.

9. Acknowledgements

10. References

10.1. Normative References

[I-D.mdt-softwire-map-dhcp-option]
Mrugalski, T., Boucadair, M., and O. Troan, "DHCPv6 Options for Mapping of Address and Port", draft-mdt-softwire-map-dhcp-option-00 (work in progress), October 2011.

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

[RFC5342] Eastlake, D., "IANA Considerations and IETF Protocol Usage for IEEE 802 Parameters", BCP 141, RFC 5342, September 2008.

[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.bcx-address-fmt-extension]
Bao, C. and X. Li, "Extended IPv6 Addressing for Encoding Port Range", draft-bcx-address-fmt-extension-02 (work in progress), October 2011.

[I-D.boucadair-behave-ipv6-portrange]
Boucadair, M., Levis, P., Grimault, J., Villefranque, A., Kassi-Lahlou, M., Bajko, G., Lee, Y., Melia, T., and O. Vautrin, "Flexible IPv6 Migration Scenarios in the Context of IPv4 Address Shortage", draft-boucadair-behave-ipv6-portrange-04 (work in progress), October 2009.

[I-D.boucadair-dhcpv6-shared-address-option]
Boucadair, M., Levis, P., Grimault, J., Savolainen, T., and G. Bajko, "Dynamic Host Configuration Protocol (DHCPv6) Options for Shared IP Addresses Solutions", draft-boucadair-dhcpv6-shared-address-option-01 (work in progress), December 2009.

[I-D.boucadair-softwire-stateless-requirements]
Boucadair, M., Bao, C., Skoberne, N., and X. Li, "Requirements for Extending IPv6 Addressing with Port Sets", draft-boucadair-softwire-stateless-requirements-00 (work in progress), September 2011.

[I-D.bsd-softwire-stateless-port-index-analysis]
Boucadair, M., Skoberne, N., and W. Dec, "Analysis of Port Indexing Algorithms", draft-bsd-softwire-stateless-port-index-analysis-00 (work in progress), September 2011.

[I-D.chen-softwire-4v6-add-format]
Chen, G. and Z. Cao, "Design Principles of a Unified Address Format for 4v6", draft-chen-softwire-4v6-add-format-00 (work in progress), October 2011.

[I-D.chen-softwire-4v6-pd]
Chen, G., Sun, T., and H. Deng, "Prefix Delegation in 4V6", draft-chen-softwire-4v6-pd-00 (work in progress), August 2011.

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

[I-D.despres-softwire-4rd-addmapping]
Despres, R., Qin, J., Perreault, S., and X. Deng, "Stateless Address Mapping for IPv4 Residual Deployment (4rd)", draft-despres-softwire-4rd-addmapping-01 (work in progress), September 2011.

[I-D.despres-softwire-4rd-u]
Despres, R., "Unifying Double Translation and Encapsulation for 4rd (4rd-U)", draft-despres-softwire-4rd-u-01 (work in progress), October 2011.

[I-D.despres-softwire-sam]
Despres, R., "Stateless Address Mapping (SAM) - a Simplified Mesh-Softwire Model", draft-despres-softwire-sam-01 (work in progress), July 2010.

[I-D.despres-softwire-stateless-analysis-tool]
Despres, R., "Analysis of Stateless Solutions for IPv4 Service across IPv6 Networks - A synthetic Analysis Tool", draft-despres-softwire-stateless-analysis-tool-00 (work in progress), September 2011.

[I-D.mrugalski-dhc-dhcpv6-4rd]

Mrugalski, T., "DHCPv6 Options for IPv4 Residual Deployment (4rd)", draft-mrugalski-dhc-dhcpv6-4rd-00 (work in progress), July 2011.

[I-D.murakami-softwire-4rd]
Murakami, T., Troan, O., and S. Matsushima, "IPv4 Residual Deployment on IPv6 infrastructure - protocol specification", draft-murakami-softwire-4rd-01 (work in progress), September 2011.

[I-D.murakami-softwire-4v6-translation]
Murakami, T., Chen, G., Deng, H., Dec, W., and S. Matsushima, "4via6 Stateless Translation", draft-murakami-softwire-4v6-translation-00 (work in progress), July 2011.

[I-D.sun-softwire-stateless-4over6]
Sun, Q., Xie, C., Cui, Y., Wu, J., Wu, P., Zhou, C., and Y. Lee, "Stateless 4over6 in access network", draft-sun-softwire-stateless-4over6-00 (work in progress), September 2011.

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

[I-D.xli-behave-divi-pd]
Li, X., Bao, C., Dec, W., Asati, R., Xie, C., and Q. Sun, "dIVI-pd: Dual-Stateless IPv4/IPv6 Translation with Prefix Delegation", draft-xli-behave-divi-pd-01 (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.

[RFC3194] Durand, A. and C. Huitema, "The H-Density Ratio for Address Assignment Efficiency An Update on the H ratio", RFC 3194, November 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.

Appendix A. Open issues / New features

A.1. Max PSID

It has been proposed to keep independence of IPv6 routing plans from IPv4 considerations and yet to be able to support variable sized port sets per shared IPv4 address. A mechanism proposed for this is called "MAX PSID". The idea is that a source, transmitting a packet to a CE doesn't need to know the length of the PSID field of that CE. All port bits after offset bits are copied in the encoded IPv6 address. This implies that a MAP CE be capable of receiving MAP traffic for multiple addresses within its delegated prefix, e.g. using the same mechanism as used for double translation when CEs are allocated IPv4 prefixes shorter than /32.

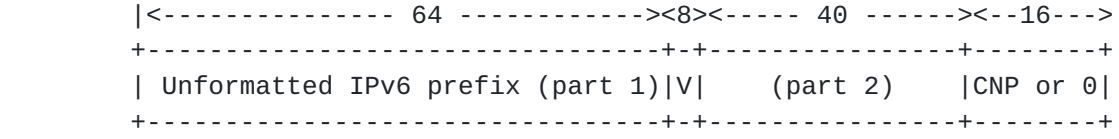
A.2. Interface identifier - V octet and Checksum neutrality

There are multiple issues related to the Interface-identifier encoding.

- The V octet is required to distinguish between MAP and native IPv6 traffic if the same End-user IPv6 prefix is used. If a separate End-user IPv6 prefix is used for MAP traffic, requiring a special flag in the interface-identifier is not required.

- The Checksum-neutrality preserver (CNP). It is for MAP packets to be acceptable by IPv6 functions that check UDP/TCP checksums, without needing for this to consider transport-layer fields. Checksum neutrality is useful for double translation and, possibly more important, it permits to envisage a unified solution which has significant advantages of both encapsulation and double translation [I-D.despres-softwire-4rd-u]. With encapsulation, the field can be set to 0.

With both mechanisms, IPv6 addresses have the following format:



The V octet deterministically differentiates MAP addresses from other IPv6 addresses by having its bits 6 and 7 set to 1 and 1 (they are 1 and 0 in modified EUI-64 Interface-ID format, and bit 6 is 0 in the privacy extension of [RFC 3041]. V is proposed to be 0x03 (which leaves 2^6 values of bits 0 to 5 for other Interface ID formats that

could be useful in the future).

The Unformatted IPv6 prefix starts with bits derived from the IPv4 address being mapped (e.g. Rule IPv6 prefix, IPv4 suffix, and PSID, or Max PSID if applicable). The remainder to reach 104 bits is filled with 0s.

The CNP field is, in one's complement arithmetic, the sum of the two halves of the IPv4 address, minus the sum of the seven 16-bit fields that precede the CNP in the IPv6 address.

A.3. Optional BR per Rule within a domain

With BR IPv6 address/prefix as optional parameters in mapping rules, it has been proposed to support ISP networks that have IPv4 prefixes coming from several providers necessitating geographically dispersed BRs. In such configurations, each provider exercises ingress filtering so that a CE MUST sent its traffic going to the Internet via the right BR (that whose locally routed IPv4 prefixes include one that matches the IPv4 address or prefix of the CE) [I-D.despres-softwire-4rd-addmapping].

Appendix B. Requirements

This list of requirements for a stateless mapping of address and ports solution may not be complete. The requirements are listed in no particular order, and they may be conflicting.

R-1: To allow for a single user delegated IPv6 prefix to be used for native IPv6 service and for MAP, the representation of an IPv4 prefix, address or shared IPv4 address and PSID must be efficient. As an example it must be possible to represent a shared IPv4 address and PSID in 24 bits or less. (Given a typical prefix assignment of /56 to the end-user and a MAP IPv6 prefix of /32.)

R-2: The IPv6 address format and mapping must be flexible, and support any placement of the embedded bits from IPv4 prefix/ address and port set in the IPv6 address.

R-3: Algorithm complexity. The mapping from an IPv4 address and port to an IPv6 address is done in the forwarding plane on MAP nodes. It is important that the algorithm is bounded and as efficient as possible.

R-4: MAP must allow service providers to define their own address sharing ratio. MAP MUST NOT in particular restrict by design the possible address sharing ratio; ideally 1:1 and 1:65536 should be supported. The mapping must at least support a sharing ratio of 64, 1024 ports per end-user.

R-5: The mapping may support deployments using differentiated port-sets. That is, end-users are assigned different sized port-sets and direct communication between MAP CEs are permitted.

R-6: The mapping should support differentiated port sets within a single shared IPv4 address. (i.e., be able to assign port sets of different sizes to customers without requiring any per customer state to be instantiated in network elements involved in data transfer).

R-7: The MAP solution should support excluding the well known ports 0-1023.

R-8: It MUST be possible to assign well known ports to a CE.

R-9: There must not be any dependency between IPv6 addressing and IPv4 addressing. With the exception where full IPv4 addresses or prefixes are encoded. Then IPv6 prefix assignment must be done so that martian IPv4 addresses are not assigned.

R-10: The mapping must not require IPv4 routing to be imported in IPv6 routing.

R-11: The mapping should support legacy RTP/RTCP compatibility. (Allocating two consecutive ports).

R-12: The mapping may be UPnP 1.0 friendly. A UPnP client will keep asking for the next port (as in current port + 1) a scattered port allocation will be more UPnP friendly.

R-13: For out of domain traffic the mapping must support embedding a full IPv4 address in the IPv6 interface identifier. This is required in the translation case. It also simplifies pretty printing and other operational tools.

R-14: For Service Providers requiring to apply specific policies on per Address-Family (e.g., IPv4, IPv6), some provisioning tools (e.g., DHCPv6 option) may be required to derive in a deterministic way the IPv6 address to be used for the IPv4 traffic based on the IPv6 prefix delegated to the home network.

R-15: It should/must/may be possible to use the same IPv6 prefix (/64) for native IPv6 traffic and MAPed traffic.

R-16: When only one single IPv6 prefix is assigned for both native IPv6 communications and the transport of IPv4 packets, the IPv4-translatable IPv6 prefix MUST have a length less than /64. When distinct prefixes are used, this requirement is relaxed.

R-17: The same mapping must support both translation and encapsulation solutions.

Appendix C. Deployment considerations

C.1. Flexible Assigment of Port Sets

Different classes of customers require port sets of different size. In the context of shared IPv4 addresses, some customers would be satisfied with an shared IPv4 addresses, while others may need to be assigned a single IPv4 address or delegated an IPv4 subnet.

C.2. Traffic Classification

Usually, ISPs adopt traffic classification to ensure service quality for different classes of customers. This feature is also helpful for customer behavior monitoring and security protection. For example, DIA (Dedicated Internet Access) has been provided by operators for corporations to cater for their Internet communications needs. Service is made by means of the edge router features and key systems, like ACL (Access List Control) to classify different traffic. Five tuples would be identified from IP header and UDP/TCP header. Currently, it is very well supported in IPv4. Vendors are delivering or committed to support that feature for IPv6. In order to facilitating IPv6 deployment, MAP solution should support this feature on IPv6 plane.

C.3. Prefix Delegation Deployment

Prefix delegation is an important feature both for broadband and mobile network. In mobile network, prefix delegation is introduced in 3GPP network in Release 10. The deployment of PD would be supported in 4v6 case. Variable length of IPv6 prefix is assigned to CPE for deriving IPv4 information.

C.4. Coexisting Deployment

4v6 solutions(i.e. encapsulation and translation) would not only coexist with each other, but also can harmonize with other deployment cases. Here lists some coexisting cases. (Note: more coexisting cases are expected to be investigated in future.)

o Case 1: Coexisting between 4v6 encapsulation and 4v6 translation

o Case 2: Coexisting between 4v6 translation and NAT64 (Single Translation)

o Case 3: Coexisting between 4v6 solutions and SLAAC

C.5. Friendly to Network Provisioning

Network management plane normally has an ability to to identify different users and the compatible with the address assignment techniques in the domain. 4v6 would conform to current practices on management plane. In 3GPP network, for example, only the IPv6 prefix is assigned to the devices, used to identify different users. And management plane for one family address is better than two, namely the operating platform does not need to manage both IPv4 and IPv6. Since only IPv6 prefix is assigned, 4v6 on the management plane is naturally conducted only via IPv6.

C.6. Enable privacy addresses

User privacy should be taken into account when 4v6 solution is deployed. Some security concern associated with non-changing IPv6 interface identifiers has been expressed in RFC4941[RFC4941]. Ability to change the interface identifier over time makes it more difficult for eavesdroppers and other information collectors to identify when different addresses used in different sessions actually correspond to the same node.

C.7. Facilitating 4v6 Service

Some ISPs may need to offer services in a 4v6 domain with a shared address, e.g. 4v6 node hosts FTP server. The service provisioning may require well-know port range(i.e. port range belong to 0-1023). MAP would provide operators with possibilities to generate a port range including the 0-1023. Afterwards, operators could decide to assign it to any requesting user.

C.8. Independency with IPv6 Routing Planning

The IPv6 routing is easier to plan if it's not impacted by the encoded IPv4 or port ID information. MAP would prohibit IPv4 routing imported in IPv6.

C.9. Optimized Routing Path

MAP could achieve optimized routing path both for hub case and mesh case. Traffic in hub and spoke case could follow asymmetric routing, in which incoming routes would not be binded to a given border point but others geographically closed to traffic initiators. In mesh cases, traffic between CPEs could directly communicate without going through remote border point.

Appendix D. Guidelines for Operators

This appendix is purposed to (1) clarify the difference and relevance of MAP address mapping format and what has published in standard track; (2) provide a referential guideline to operators, illustrating a common use-case of MAP deployment.

D.1. Additional terms

The following terms are listed, mainly used in this appendix only, as an add-on to the terminology of the main text.

4pfx	the index for an IPv4 prefix, either generated with coding or as same as the IPv4 prefix itslef.
ug-octet	the octet consisting of 64-71 bits in the IPv6 address, containing the bits u and g defined by EUI-64 standard.
Common prefix	an aggregate decided by a domain for the MAP deployment. It is a subset of the operator's aggregates by its RIR or provider.
IPv4 suffix	the part of IPv4 address bits used for identifying CEs.
Host suffix	the IPv6 suffix assigning to an end system. NOTE: it doesn't mean this should be really configured on a certain interface of a host.
MAP-format	the address mapping format defined by this document.
RFC6052-format	the address mapping format defined by [RFC6052] and its succeeding extensions (or updates) for port-space sharing, for example, [I-D.xli-behave-divi].

D.2. Understanding address formats: their difference and relevance

MAP introduces an address format of embedding IPv4 information to IPv6 address. On the other hand, we also have [RFC6052] defines an address format with the similar property. With extending port-set id, it can also support address sharing among different CEs [I-D.xli-behave-divi]. What are their difference and relevance?

We present a common abstract format for them both, as is depicted in

Figure 11. For the easy expression, we exclude the ug-octet, which is not concerned in this appendix.

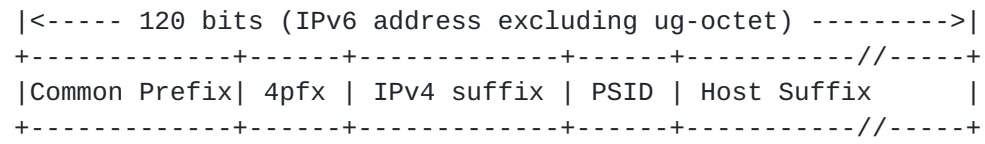


Figure 11: Abstract view of MAP- and RFC6052-formats

Only two parts in Figure 11 are different for MAP- and RFC6052-formats. We compare them in Figure 12 and following paragraphs.

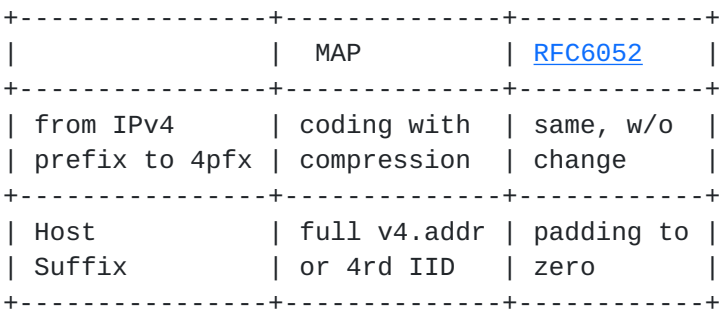


Figure 12: Difference between MAP- and RFC6052-formats

The comparison clarifies that the major role of full IPv4 address embedded in the RFC6052 format is replaced by the MAP's coded IPv4 prefix index and the uncoded IPv4 suffix. The Figure 13 illustrates this relevance.

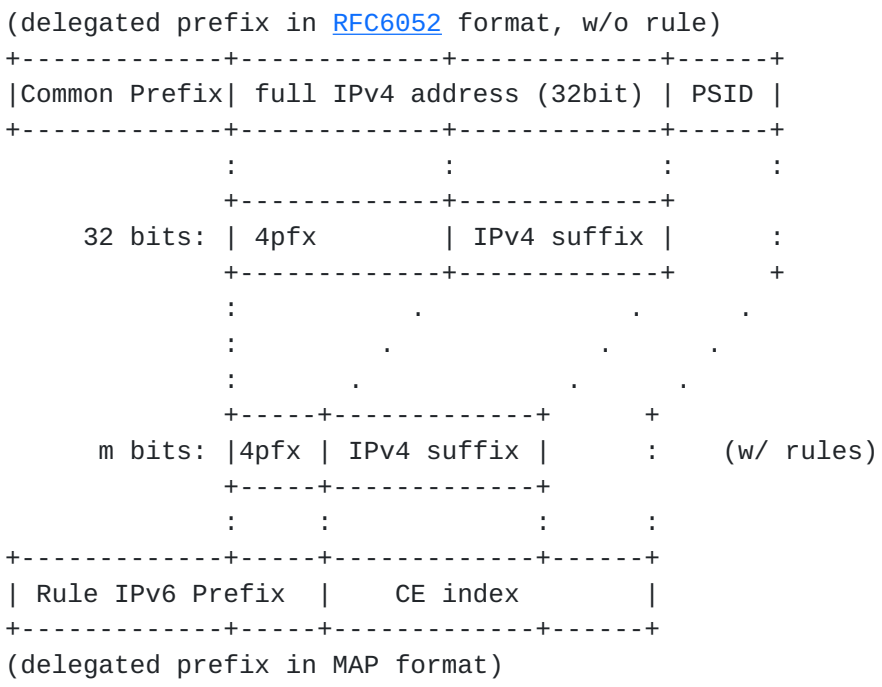


Figure 13: Relevance between MAP- and RFC6052-formats

o Why is it needed to compress the IPv4 prefix?

Precisely speaking, it is not "to compress the IPv4 prefix" but "to establish correspondence between IPv6 delegated prefixes and the residual IPv4 prefixes."

It is important for an operator to understand what the MAP is designed for and where it could be applied. A keyword for MAP is "residual deployment", referring to the deployment of an IPv6 network with utilizing the residual IPv4 address spaces for the subnets/host having IPv4 communication, without introducing per-session states.

Therefore, the delegated CE prefixes are determined prior to finding a proper IPv4 address block in hand to be mapped to the CE index and the IPv4 prefix index (4pfx) as well as the Rule IPv6 prefix.

IPv6 delegation planning, independent of the IPv4 addressing, also implies to follow the common convention of assigning a /64 prefix to any IPv6 local network. It is highly impossible to directly match some IPv4 prefixes to the already-determined IPv6 prefixes,

and therefore the prefixes have to be coded and typically it is a compression.

If we have a short-enough Common Prefix, it is also possible to deploy a direct matching where 4pfx is equal to IPv4 prefix. Only in this case, the MAP-format is equivalent to the RFC6052-format and the rule set could be simplified to a unique rule for 0.0.0.0/0.

Once the unique rule for 0.0.0.0/0 is defined, the special rule for the out-of-domain traffic towards the BR is not needed any more. The route with the common prefix itself can play the role of less-specific routes for the whole IPv4 space. This is a feature of the RFC6052-format.

- Why does MAP copy IPv4 address in the suffix?

  The full IPv4 address is copied in the Host suffix of MAP with two reasons.

  First of all, for the traffic going out of the domain, compress coding makes full IPv4 address information not directly appear in the IPv6 prefix for BR at all. To enable the double translation, it is had to embed this information in the Host Suffix of MAP-format for the peer IPv4 address outside of the domain.

  Further, it is not necessary to separate the processing for the in-domain addresses and that for the out-domain addresses. Making a symmetric format is perferred.

  Another concern is the simplicity. Even though the delegated prefix is theorectally sufficient to extract the corresponding IPv4 address for the CE, it relies on retrieving rules for every datagram. Embedding the full IPv4 address in the suffix simplifies the processing at IPv6-to-IPv4 translator when utilizing MAP for double translation. It also helps in setting filters at middle boxes, with exposing the IPv4 full addresses dispatched to the CEs.

MAP is designed for the residual deployment, including the case of recalling deployed IPv4 addresses and reallocating them for the deployment in IPv6 networks. To this extent, MAP can understood as "4rd-MAP".

In practice, MAP-format can be also used for the objective of providing stateless encapsulation or double translation for the already deployed IPv4 networks, without renumbering, whose provider backbone is upgraded to IPv6. Unlike the residual deployment, this use-case unavoidably introduces IPv4 routing entropy into the IPv6 routing infrastrucutre. On the other hand, for the old IPv4 network or IPv6 network upgraded from IPv4, it is not necessarily having 64 bits for their host identifiers. Therefore longer-than-/64 prefix is not a strict constrain. Therefore, RFC6052-format is recommended in this case of non-residual deployment. RFC6052-format is motivated with keeping temporal uniqueness of end-to-end identifiers throughout the period of transition and providing the rule-free simplicity.

D.3. Residual deployment with MAP

This section illustrate how we can use MAP in the operation of residual deployment.

NOTE: Applying MAP for a use-case other than residual deployment should follow different logic of address planning and therefore, because of the reason mentioned above, not included in this Appendix.

Residual deployment starts from IPv6 address planning. A simple example is taken inline for easy understanding.

(A) IPv6 considerations

(A1) Determine the maximum number N of CEs to be supported, and, for generality, suppose N = 2^n.

For example, we suppose n = 20. It means there will be up to about one million CEs.

(A2) Choose the length x of IPv6 prefixes to be assigned to ordinary customers.

Considering we have a /32 IPv6 block, it is not a problem for the IPv6 deployment with the given number of CEs. Let x = 60, allowing subnets inside in each CE delegated networks.

(A3) Multiply N by a margin coefficient K, a power of two (K = 2 ^ k), to take into account that:

- Some privileged customers may be assigned IPv6 prefixes of length x', shorter than x, to have larger addressing spaces than ordinary customers, both in IPv6 and IPv4;

- Due to the hierarchy of routable prefixes, many theoretically delegatable prefixes may not be actually delegatable (ref: host density ratio of [RFC3194]).

In our example, let's take k = 0 for simplicity.

(B) IPv4 considerations

(B1) List all (non overlapping, not yet assigned to any in-running networks) IPv4 prefixes Hi that are available for IPv4 residual deployment.

Suppose that we hold two blocks and not yet assigned to any fixed network: 192.32../16 and 63.245../16.

(B2) Take enough of them, among the shortest ones, to get a total whose size M is a power of two (M = 2 ^ m), and includes a good proportion of the available IPv4 space.

If the M < N, addresses should be shared by N CEs and thus each is shared by N/M = 2^(n - m) CEs with PSID length of (n - m).

If we use both blocks, M = 2^16 + 2^16, and therefore m = 17. Then PSID length could be 3 bits, the corresponding sharing ratio is also determined so that each CE can have 8192 ports to use under the shared global IPv4 address.

(B3) For each IPv4 prefix Hi of length hi, choose a "rule index", i.e., the 4pfx in Fig.C-1 and Fig.C-3, say Ri of length ri = m - (32 - hi).

All these indexes must be non overlapping prefixes (e.g. 0, 10, 110, 111 for one /10, one /11, and two /12).

Then we have:

H1 = 192.32../16, h1 = 16, r1 = 1 => R1 = bin(0);
H2 = 63.245../16, h2 = 16, r2 = 1 => R2 = bin(1);

(C) After (A) and (B), derive the rule(s)

(C1) Derive the length c of the "Common prefix" C that will appear at the beginning of all delegated prefixes (c = x - (n + k)).

(C2) Take any prefix for this C of length c that starts with a RIR-allocated IPv6 prefix.

(C3) For each IPv4 prefix Hi, make a rule, in which the key is Hi, and the value is the Common prefix C followed by the Rule index Ri. Then this i-th rule's Rule IPv6 Prefix will have the length of (c + ri).

Then we can do that:

c = 40 => C = 2001:0db8:ff00::/40
Rule 1: Rule IPv6 Prefix = 2001:0db8:ff00::/41
Rule 2: Rule IPv6 Prefix = 2001:0db8:ff80::/41

As a result, for a certain CE delegating 2001:0db8:ff98:
7650::/60, its parameters are:

```
Rule IPv6 Prefix = 2001:0db8:ff80::/41 => Rule 2
IPv4 Suffix = bin(001 1000 0111 0110 0)
                              PSID = bin(101) = 0x5
Rule IPv4 Prefix = 63.245../16
CE IPv4 Address = 63.245.48.236
```

If different sharing ratio is expected, we may partition CEs into groups and do (A) and (B) for each group, determining the PSID length for them separately. However, this might cause a fairly complicated work in the address planning.

Author's Address

Ole Troan (editor)
cisco
Oslo
Norway

Email: ot@cisco.com