Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: March 26, 2012

X. Li
C. Bao
CERNET Center/Tsinghua University
W. Dec
R. Asati
Cisco Systems
C. Xie
Q. Sun
China Telecom
September 23, 2011

dIVI-pd: Dual-Stateless IPv4/IPv6 Translation with Prefix Delegation
draft-xli-behave-divi-pd-01

Abstract

This document presents the address specifications and deployment considerations of address-sharing dual stateless IPv4/IPv6 translation with prefix delegation (dIVI-pd). The dIVI-pd keeps the features of stateless, end-to-end address transparency and bidirectional-initiated communications of the original stateless IPv4/IPv6 translation, while it can utilize the IPv4 addresses more effectively. In addition, it does not require the DNS64 and ALG, and can be used with prefix delegation.

Status of this Memo

Copyright Notice

Table of Contents

1. Introduction

The experiences for the IPv6 deployment in the past 10 years strongly indicate that for a successful transition, the communication between IPv4 and IPv6 address families should be supported.

Recently, the stateless and stateful IPv4/IPv6 translation methods are developed and became the IETF standards. The original stateless IPv4/IPv6 translation (stateless 1:1 IVI) is scalable, maintains the end-to-end address transparency and support both IPv6 initiated and IPv4 initiated communications [RFC6052], [RFC6144], [RFC6145], [RFC6147], [RFC6219]. But it can not use the IPv4 addresses effectively. The stateful IPv4/IPv6 translation can share the IPv4 addresses among IPv6 hosts, but it only supports IPv6 initiated communication [RFC6052], [RFC6144], [RFC6145], [RFC6146], [RFC6147]. In addition, both stateless and stateful IPv4/IPv6 translation technologies require the application layer gateway (ALG) for the applications which embed IP address literals. Furthermore, in ADSL and 3G environment, it requires the prefix delegation (assigning an IPv6 /64 or shorter) to the customer router/L3-device rather than assigning a single IPv4-translatable address to the customer device defined in [RFC6052].

In this document, we present address specifications and deployment considerations for address-sharing dual stateless IPv4/IPv6 translation with prefix delegation (dIVI-pd), which is based on basic dIVI model [I-D.xli-behave-divi] with the support of prefix delegation. The dIVI-pd can solve the IPv4 address sharing, the ALG and prefix delegation problems mentioned above, though still keeps the stateless, end-to-end address transparency and supporting of both IPv6 initiated and IPv4 initiated communications.

The dIVI-pd is in the family of stateless IPv4/IPv6 translation, alone with IVI and dIVI. These techniques are used for the following scenarios defined in [RFC6144].

- o Scenario 1: An IPv6 network to the IPv4 Internet.
- o Scenario 2: The IPv4 Internet to an IPv6 network.
- o Scenario 5: An IPv6 network to an IPv4 network.
- o Scenario 6: An IPv4 network to an IPv6 network.

2. Applicability

The address-sharing dual stateless IPv4/IPv6 translation with prefix delegation (dIVI-pd) can be used in ADSL or 3G environment when prefix delegation is required. An ADSL example is shown in the following figure.

```
        ----             -----
      //    \\         //     \\         -----
     /        \       /         \      //      \|------[CPE.1]--{H.1}
    |         +-----+            +----+         |
    |         |     |   Metro    |BRAS|         |------[CPE.2]--{H.2}
    | Backbone|Core |   Area     |/SR | Access  |
    | Network |Route|  Network   +----+ Network |------[CPE.k]--{H.k}
    |         |     |            |AAA |         |
    |         +-----+            +----+         |------[CPE.n]--{H.n}
     \        /       \          /      \\     /|
      \\    //         \\      //         ----
        ----            ----|--                         |
                            |                           |
    IPv4/IPv6           [XLAT1]       IPv6-only     [XLAT2.x] IPv4/IPv6
    Internet                |         Network           |      Hosts
                            |                           |
    Data path:              |                           |
    IPv6 --------------------------------IPv6-----------|------IPv6----
    IPv4 -------------------|------------IPv6-----------|------IPv4----
                            |                           |
    Address assignment:     |                           |
    IPv6                    |  [BRAS]->(DHCPv6 PD)->[CPE]->SLAAC->[Host]
    IPv4                    |                           | \-DHCP-/
```

Figure 1: BRAS

Where the ISP's backbone network is dual stack, as well as part of the metro-area network. The core IPv4/IPv6 translator (XLAT1) is performing the IPv4 address-sharing stateless IPv4/IPv6 translation and connects the dual-stack part and the IPv6-only part of the metro-area networks. The access network is IPv6-only and multiple IPv4/IPv6 translators (XLAT2.x) are connected to the access network and provide dual-stack access to the customer devices. Each dual-stack customer get a whole IPv6 /64 (or shorter) and a fractional public IPv4 address.

The data path of this user case are: The IPv6 packets from customer devices and the IPv6 Internet are not translated, while the IPv4 packets from customer devices and the IPv4 Internet are translated twice via stateless IPv4/IPv6 translation technology. Due to the

stateless nature, the dual stateless IPv4/IPv6 translation is almost equivalent to tunneling with header compression.

There are two address assignment processes: (1) From BRAS to CPE is via IPv6CP and DHCPv6 prefix delegation; (2) From CPE to customer device, the IPv6 is via SLAAC and the IPv4 is via DHCP. Note that if more than one customer device requires IPv4 addresses, a built-in NAT44 in each CPE can be used to translate a fractional IPv4 address to several [RFC1918] defined IPv4 addresses.

3. Terminologies

This document uses the terminologies defined in [RFC6144].

The key words MUST, MUST NOT, REQUIRED, SHALL, SHALL NOT, SHOULD, SHOULD NOT, RECOMMENDED, MAY, and OPTIONAL, when they appear in this document, are to be interpreted as described in [RFC2119].

4. Address Specifications

The address-sharing dual stateless IPv4/IPv6 translation with prefix delegation (dIVI-pd) requires the header translation specifications defined in [RFC6145].

In addition, it uses the address format of PL=64 defined in Figure 1 of [RFC6052] with two kind of extensions (CPE index and suffix), as shown in the following figure.

```
+--+---------+-------+------+-----+----+-----------+-----------+----+
|PL| 0-------32------48-----56----64---72----------104---------120  |
+--+---------+-------+------+-----+----+-----------+-----------+----+
|64|           prefix             | u  |  v4(32)   |    suffix      |
+--+---------+-------+------+-----+----+-----------+-----------+----+
|e |domain prefix |CPE index  | 0 | u  |  v4(32)   |suffix     | 0  |
|x |--------------+-----------+---+----+-----------+-----------+----|
|t |(d) bits      |(s+k) bits |(m)|(8) |(32) bits  |(16) bits  |(8) |
+--+---------+-------+------+-----+----+-----------+-----------+----+
```

Figure 2: Extended IPv4-translatable address format

4.1. Attributes

A dIVI-pd system has the following attributes used in the algorithmatic mapping process:

1. Domain prefix

This parameter is assigned by the network operator per dIVI-pd domain and used by the XLAT1 and CPEs for constructing the delegated prefixes for each CPE. The length of the domain prefix is (d) bits.

2. Total number of CPEs in this domain

The total number of CPEs served in a specific domain determines the length of the CPE index (s+k) bits.

3. IPv4 Address sharing ratio

Each dIVI-pd domain is assigned a target IPv4 address-sharing ratio N, as a power of 2, for addresses assigned to that domain.

4. Prefix length assigned to a specific CPE

For a customer with IPv6 prefix delegation, the longest prefix length is /64. However, other prefix lengths are also permitted, for example /63, /62, ..., /56. The length of prefix delegation determines the zero padding length (m).

5. CPE index

The CPE index is used to uniquely identify different CPEs in the specific domain. The length of the CPE index is (s+k) bits.

6. Suffix

The suffix is used for several IPv6 nodes sharing an IPv4 address, with each node managing a different range of ports. The suffix contains the IPv4 address sharing ratio N of 4 bits and the Port-set ID of 12 bits, so the length of suffix is 16 bits, fixed. Note that the suffix is the same as in [I-D.xli-behave-divi]. This makes the CPE working in both dIVI and dIVI-pd environment.

4.2. CPE index Coding

The CPE index is used to uniquely identify different CPEs in the specific domain. For the reason of simplicity, the CPE index consists of (k) bits identifying the IPv4 devices using different port-sets of same IPv4 address and (s) bits for the IPv4 subnet to

serve for this specific domain.

4.3. Suffix Coding

The suffix is used for several IPv6 nodes sharing an IPv4 address, with each node managing a different range of ports [I-D.xli-behave-divi]. The most significant 4 bits define the multiplexing ratio and the least significant 12 bits define the IPv6 node index. The multiplexing ratio, the suffix range and the number of corresponding concurrent ports are as shown in the following figure.

ratio	suffix range	# of Ports
1	0000 - 0000	65,536
2	1000 - 1001	32,768
4	2000 - 2003	16,384
8	3000 - 3007	8,192
16	4000 - 400f	4,096
32	5000 - 501f	2,048
64	6000 - 603f	1,024
128	7000 - 707f	512
256	8000 - 80ff	256
512	9000 - 91ff	128
1,024	a000 - a3ff	64
2,048	b000 - b7ff	32
4,096	c000 - cfff	16

Figure 3: Suffix for Port Range Encoding

4.4. Port-set algorithm

The algorithm used to derive available port-set for a specific CPE, or by XLAT1 to construct, per domain, the CPE index based on an IPv4 address and TCP/UDP port.

For a domain's multiplexing ratio N, the port-set numbers of a CPE with port-set-id K is composed of P=j*N + K + 1024, for all the values of j=0, 1, ..., (65536-N)/N.

For a destination port number (P), the port-set-id of a given CPE with port-set-id K is determined by the modulo operation: K=((P-1024)%N) (% is the Modulus Operator).

4.5. Domain prefix selection

The domain prefix is inside the ISP prefix. The domain prefix length (d) can be determined by

(d)=64-(s)-(k)-(m)

Where (s) is the number of bits of the IPv4 subnet, (k) is the bits representing different port-set ID sharing the same IPv4 address and (m) is the number of bits representing the size of the PD compared with standard PD (/64). When the domain prefix length is determined, the domain prefix can be selected by network administrator.

5. Deployment Considerations

5.1. IPv4-converted address

The IPv4-converted address is for presenting the IPv4 hosts in the IPv4 Internet. For avoiding the collision between IPv4-converted address and SLAAC address, a common IPv4-converted address block (/64) is selected for all CPEs in the specific domain. This IPv4-converted address block is a subset of the domain prefix and it can use the IPv4 subnet identifier or broadcast address, therefore, the IPv6-converted address block is different from all CPE prefix without wasting the IPv4 addresses and IPv6 addresses.

5.2. IPv4-translatable address

In order to minimize the operational overhead, it is desirable to delegate a single prefix (IPv6 /64 or shorter) to each customer for SLAAC and this prefix also contains the IPv4-translatable address for communicating with the IPv4 Internet. It can be shown that the probability of collision between SLAAC and Ipv4-translatable address is almost zero.

5.3. Hairpin

Since the IPv4-converted addresses and IPv4-translatable addresses are using different prefixes in this case, the hairpin function is required, which can be done in XLAT1 or in BRAS/SR using IPv6-to-IPv6 Network Prefix Translation discussed in [RFC6296].

5.4. Dual translation

The advantage of dual IVI is that the DNS64/DNS46 and ALG are not required.

Special MTU and fragmentation actions must be taken in the case of dual translation.

6. Experimental Evaluation

The basic stateless IPv4/IPv6 translation (IVI) has been deployed since 2007. It connects [CERNET] and [CNGI-CERNET2].

The dual stateless translation with IPv4 address sharing (dIVI) has been deployed in [CERNET] and [CNGI-CERNET2] since 2009. The design and implementation results are presented in [I-D.xli-behave-divi].

The dIVI has also been tested in China Telecom. The [I-D.sunq-v6ops-ivi-sp] summarizes the testing results.

The dIVI-pd presented in this document has been running in [CERNET] and [CNGI-CERNET2] since Jan. 2011. The experimental results indicate that the CPE index coding, the suffix coding and port-set ID mapping algorithm work for existing applications without any problem.

7. Security Considerations

See security considerations presented in [RFC6052] and [RFC6145].

8. IANA Considerations

This memo adds no new IANA considerations.

Note to RFC Editor: This section will have served its purpose if it correctly tells IANA that no new assignments or registries are required, or if those assignments or registries are created during the RFC publication process. From the author's perspective, it may therefore be removed upon publication as an RFC at the RFC Editor's discretion.

9. Acknowledgments

The authors would like to acknowledge the following contributors in the different phases of the address-sharing IVI and dIVI development: Hong Zhang, Yu Zhai, Wentao Shang, Weifeng Jiang, Bizhen Fu, Guoliang Han and Weicai Wang.

The authors would like to acknowledge the following contributors who provided helpful inputs: Heyu Wang, Lu Yan, Dan Wing, Fred Baker,

Dave Thaler, Randy Bush, Kevin Yin and Bobby Li.

10. References

10.1. Normative References

[RFC1918] Rekhter, Y., Moskowitz, R., Karrenberg, D., Groot, G., and E. Lear, "Address Allocation for Private Internets", BCP 5, RFC 1918, February 1996.

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

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

[RFC6144] Baker, F., Li, X., Bao, C., and K. Yin, "Framework for IPv4/IPv6 Translation", RFC 6144, April 2011.

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

[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.

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

[RFC6296] Wasserman, M. and F. Baker, "IPv6-to-IPv6 Network Prefix Translation", RFC 6296, June 2011.

10.2. Informative References

[CERNET] "CERNET Homepage: http://www.edu.cn/english_1369/index.shtml".

[CNGI-CERNET2]
"CNGI-CERNET2 Homepage:

http://www.cernet2.edu.cn/index_en.htm".

[I-D.sunq-v6ops-ivi-sp]
Sun, Q., Xie, C., Li, X., Bao, C., and M. Feng, "Considerations for Stateless Translation (IVI/dIVI) in Large SP Network", draft-sunq-v6ops-ivi-sp-02 (work in progress), March 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-03 (work in progress), July 2011.

Appendix A. Example of dual-stateless IPv4/IPv6 translation with prefix delegation (dIVI-pd)

Assume:

1. ISP prefix is 2001:db8::/32.

2. Total number of CPEs in this domain is 30

3. Address sharing ratio N=16. This means that k=4 bits.

4. Prefix length assigned to a specific CPE is /63. This means that m=1 bit.

5. The length of IPv4 subnet s=2 bits. This is obtained by 30/16 and note the fact that an IPv4 /30 should be used for 2 IPv4 hosts.

6. Suffix has 16 bits, fixed.

7. Domain prefix length d=64-s-k-m=64-2-4-1=57 bits. For operational convenience, we can make it in the 8 bit boundary of the IPv6 address, this results in d=56 bits. Then we choose 2001:db8:a4a6:4600::/56 as the domain prefix.

The serial number of CPE, the IPv4 addresses and the constructed IPv4-translatable addresses are shown in the following figure, where v and h represent the IPv4 subnet (hex) and host index (hex), respectively.

	IPv4	v	h	IPv4-translatable address
1	192.168.1.1	1	0	2001:db8:a4a6:4640:c0:a801:140:0
2	192.168.1.1	1	1	2001:db8:a4a6:4644:c0:a801:140:100
3	192.168.1.1	1	2	2001:db8:a4a6:4648:c0:a801:140:200
4	192.168.1.1	1	3	2001:db8:a4a6:464c:c0:a801:140:300
5	192.168.1.1	1	4	2001:db8:a4a6:4650:c0:a801:140:400
6	192.168.1.1	1	5	2001:db8:a4a6:4654:c0:a801:140:500
7	192.168.1.1	1	6	2001:db8:a4a6:4658:c0:a801:140:600
8	192.168.1.1	1	7	2001:db8:a4a6:465c:c0:a801:140:700
9	192.168.1.1	1	8	2001:db8:a4a6:4660:c0:a801:140:800
10	192.168.1.1	1	9	2001:db8:a4a6:4664:c0:a801:140:900
11	192.168.1.1	1	a	2001:db8:a4a6:4668:c0:a801:140:a00
12	192.168.1.1	1	b	2001:db8:a4a6:466c:c0:a801:140:b00
13	192.168.1.1	1	c	2001:db8:a4a6:4670:c0:a801:140:c00
14	192.168.1.1	1	d	2001:db8:a4a6:4674:c0:a801:140:d00
15	192.168.1.1	1	e	2001:db8:a4a6:4678:c0:a801:140:e00
16	192.168.1.1	1	f	2001:db8:a4a6:467c:c0:a801:140:f00
17	192.168.1.2	2	0	2001:db8:a4a6:4680:c0:a801:240:0
18	192.168.1.2	2	1	2001:db8:a4a6:4684:c0:a801:240:100
19	192.168.1.2	2	2	2001:db8:a4a6:4688:c0:a801:240:200
20	192.168.1.2	2	3	2001:db8:a4a6:468c:c0:a801:240:300
21	192.168.1.2	2	4	2001:db8:a4a6:4690:c0:a801:240:400
22	192.168.1.2	2	5	2001:db8:a4a6:4694:c0:a801:240:500
23	192.168.1.2	2	6	2001:db8:a4a6:4698:c0:a801:240:600
24	192.168.1.2	2	7	2001:db8:a4a6:469c:c0:a801:240:700
25	192.168.1.2	2	8	2001:db8:a4a6:46a0:c0:a801:240:800
26	192.168.1.2	2	9	2001:db8:a4a6:46a4:c0:a801:240:900
27	192.168.1.2	2	a	2001:db8:a4a6:46a8:c0:a801:240:a00
28	192.168.1.2	2	b	2001:db8:a4a6:46ac:c0:a801:240:b00
29	192.168.1.2	2	c	2001:db8:a4a6:46b0:c0:a801:240:c00
30	192.168.1.2	2	d	2001:db8:a4a6:46b4:c0:a801:240:d00
31	192.168.1.2	2	e	2001:db8:a4a6:46b8:c0:a801:240:e00
32	192.168.1.2	2	f	2001:db8:a4a6:46bc:c0:a801:240:f00

Figure 4: Address example

Note that the CPE prefixes can be obtained from corresponding IPv4-translatable addresses, for example 2001:db8:a4a6:4610:c0:a801:140:0 results in 2001:db8:a4a6:4610::/63.

A common IPv4-converted address block can be selected as 2001:db8:a4a6:4690::/64.

Authors' Addresses

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

Phone: +86 10-62785983
Email: xing@cernet.edu.cn

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

Phone: +86 10-62785983
Email: congxiao@cernet.edu.cn

Wojciech Dec
Cisco Systems
Haarlerberdweg 13-19
Amsterdam 1101 CH
NL

Email: wdec@cisco.com

Rajiv Asati
Cisco Systems
7025-6 Kit Creek Road
Research Triangle Park NC 27709
USA

Email: rajiva@cisco.com

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