Network Working Group Xiaohong Deng
Internet Draft M. Boucadair
Intended status: Standards Track France Telecom
Expires: May 2012 C. Zhou
Huawei Technologies
October 31, 2011

NAT offload extension to Dual-Stack lite
draft-zhou-softwire-b4-nat-04

Abstract

Dual-Stack Lite, combining IPv4-in-IPv6 tunnel and Carrier Grade NAT technologies, provides an approach that offers IPv4 service via IPv6 network by sharing IPv4 addresses among customers during IPv6 transition period. Dual-stack lite, however, requires CGN to maintain active NAT sessions, which means processing performance, memory size and log abilities for NAT sessions should scale with number of sessions of subscribers; Hence increasing in CAPEX for operators would be resulted in when traffic increase.

This document propose the NAT offload extensions to DS-Lite, which allows offloading NAT translation function from centralized network side (AFTR) to distributed customer equipments (B4), thereby offering a trade-off between CAPEX (e.g. less performance requirements on AFTR device) and OPEX (e.g., easy and fast deployment of Dual-Stack Lite) for operators. The ability of easily co-deploying with basic Dual-Stack Lite is essential to NAT offload extension to DS-Lite.

This Internet-Draft will expire on April 26, 2012.

Copyright Notice

Table of Contents

1. Background

The basic idea of NAT offload extension to DS-lite, is to reuse the basic DS-Lite infrastructure, including tunneling transport and provisioning method, and ICMP and fragmentation processing as well.

The NAT offload extension makes the AFTR table scales with customer number other than traffic sessions. Based on this NAT offload extension, log entries for per subscriber instead of per session is achievable. IPv4 address utilization efficiency depends on port allocation strategies, e.g., per port on demand, or a buck of ports pre-allocation, which would be elaborated in Section 5.

Besides, this method allows unique IPv6 address for delivery both IPv4 over IPv6 traffic and native IPv6 traffic without introduce any IPv4 addressing/rouging into IPv6 address/routing, as it reuses Dual Stack Lite tunneling transport infrastructure, unlike stateless solutions with port set allocation such as aplusp and 4rd, that either requires two IPv6 addresses separately for either IPv4 traffic over IPv6 or native IPv6 traffic, or require carefully design to avoid introduce IPv4 routing to IPv6 routing when using unique IPv6 address to transport both IPv4 over IPv6 traffic and native IPv6 traffic.

2. NAT offload extended DS-Lite Overview and terminologies

Figure 1 provides an overview of the NAT offload extended DS-Lite.

```
                      +--------------------------+
                      |     IPv6 ISP Network     |
                      |                          |
                      |                          |
                      +--------------------------+
                        |
                        |                   +------------+   +-----------+
                        |                   |NAT offload|   |   IPv4    |
      +---------+       |      IPv4-in-IPv6 |AFTR       |---| Internet |
      |         |  +----------+             |            |   |           |
      |IPv4 LAN|--|NATed     |===============+------------+   +-----------+
      |         |  |B4       |CPE/HOST             |
      +---------+  +----------+                    |
                      |                            |
                      |                            |
                      +--------------------------+
```

Figure 1 : NAT offload extended DS-Lite Overview

NATed B4: A NAT offload extended B4 which is called NATed B4 in this document can be either an IPv6 hosts or a CPE. NATed B4 performs IP address and port translation function, besides establishment of IPv4 in IPv6 tunnel with AFTR.

NAT offload AFTR: A NAT offload extended AFTR which is called NAT offload AFTR is responsible for establishing IPv4 in IPv6 tunneling with NATed B4 to transport IPv4 over IPv6 while the NAT translation function is offloaded to NATed B4.

A NATed B4 uses IPv4 address with a restricted port set for this IPv4 connectivity, which may be provisioned via either DHCPv4 with the AFTR, or via PCP with the PCP server. The AFTR keeps the mapping between B4's IPv6 address, allocated IPv4 address, and a restricted port set ID on a per customer basis.

For host NATed B4 case, the host gets public address directly. It is also suggested that the host run a local NAT to map randomly generated ports into the restricted port set. Private to public address translation would not be needed in this NAT. Another

solution is to have the IP stack to only assign ports within the restricted port set to applications. Either way the host guarantees that every port number in the packets sent out by itself falls into the allocated port set.

3. NATed B4 Behavior

The NATed B4 is responsible for performing NAT and/ALG functions, basic B4 functions, as well as supporting NAT Traversal mechanisms (e.g., UPnP or NAT-PMP).

The tunneling provisioning of the B4 element should reuse what has defined in [I-D.ietf-softwire-dual-stack-lite].

3.1. Plain IPv4 Address

A NATed B4 MAY be assigned with a plain IPv4 address.

When a plain, IPv4 address is assigned, the NAT operations are enforced as per current legacy CPEs. The NAT in the AFTR is disabled for that user.IPv4 datagrams are encapsulated in IPv6 as specified in [I-D.ietf-softwire-dual-stack-lite].

3.2. Restricted IPv4 Address and port set provisioning

3.2.1. Restricted port allocation strategies and requirements

Restricted port allocation strategies for this approach could either be allocating per port on demand, or be pre-allocating a port set (no matter a continuous port range, or multiple non-continuous sub port sets),which leads to trade-off between provisioning efficiency and IPv4 utilization efficiency.

Note that efficiency on log is reported by operators as a practical requirement for AFTR, hence port set decoding should take this requirement into account, no matter which port allocation strategy is adopt.

Unlike stateless 4over6 solutions such as [I-D.murakami-softwire-4rd], the restricted port sets allocation for NAT offload extended DS-Lite has no requires on careful planning of the IPv6 and IPv4 addressing together. It therefore offers more flexibility for ISPs, when it comes to managing the IPv6 access network, and introduces no impact on IPv6 routing.

3.2.2. Restricted IPv4 Address and port set provisioning method

Either DHCP for example, [I-D.bajko-pripaddrassign] or PCP would be candidate for delivery Restricted IPv4 and port set.

With PCP, The basic PCP protocol allows per port on demand allocation, while an extension to PCP [I-D.tsou-pcp-natcoord] supports pre-allocate bulk of ports.

3.3. Outgoing Packets Processing

Upon receiving an IPv4 packet, the B4 performs NAT using the public IPv4 address and port set assigned to it. Then B4 encapsulates the resulting IPv4 packet into an IPv6 packet, and delivers it through IPv6 connectivity to AFTR which will then decapsulate the encapsulated packet and forward it through IPv4. The destination IPv6 address used for encapsulation should be the AFTR's address.

3.4. Incoming Packets Processing

Upon receipt of IPv4-in-IPv6 packet from AFTR, B4 will decapsulate the packet and translate the public IPv4 address to the private IPv4 address. Finally, it delivers the packet to the host using the translated IPv4 address. The source IPv6 address used for encapsulation at AFTR is the AFTR's address, and the destination address is set to the external address of B4.

3.4.1. Incoming Ports considerations on a given restricted IPv4 address

As described in [I-D.ietf-intarea-shared-addressing-issues], a bulk of incoming ports can be reserved as a centralized resource shared by all subscribers using a given restricted IPv4 address. In order to distribute incoming ports as fair as possible among subscribers sharing a given restricted IPv4 address, other than allocating a continuous range of ports to each, a solution to distribute bulks of non-continuous ports among subscribers, which also takes port randomization into account, is elaborated in Section 3.1.

4. NAT offload AFTR Behaviour

The NAT offload AFTR may be co-located with IP and /or restricted port set allocation server (e.g., a DHCP server, or a PCP server).

The AFTR only maintains a static mapping entry per customer consist of IPv6 address, IPv4 address and port set ID, other than maintains NAT entries per session.

4.1. Outgoing Packets Processing

For outgoing packets, the NAT offload AFTR simply decapsulates it and forwards it to IPv4 Internet.

4.2. Incoming Packets Processing

For inbound traffic, NAT offload AFTR would use the IPv4 destination address and port as the index to retrieve mapping table in order to find a destination IPv6 address, and then encapsulates it into IPv6, so that native IPv6 routing could be used to forward the IPv4 in IPv6 traffic.

5. Fragmentation and Reassembly and DNS

No change to Section 5.3 of [I-D.ietf-softwire-dual-stack-lite. The DNS behavior is the same as described in [I-D.ietf-softwire-dual-stack-lite].

6. Security Considerations

As port randomization is one protection among others against blind attacks, a simple non-contiguous port sets distribution mechanism is therefore proposed to distribute bulks of non-continuous ports among subscribers, and to enable subscribers operating port randomized NAT.

In this section, a non-continuous restricted port set encoding/decoding and an algorithm of random ephemeral port selection within the allocated restricted port set example proves that port randomization is applicable this approach.

On every external IPv4 address, according to port set size N, log2(N) bits are randomly choosing by NAT offload AFTR as subscribers identification bits(s bit) among 1st and 16th bits. Take a sharing ration 1:32 for example, Figure 1 shows an example of 5 random selected bits of s bits.

```
|1st |2nd |3rd |4th |5th |6th |7th | 8th|
+----+----+----+----+----+----+----+----+
| 0  |  s | 0  | 0  | s  | 0  | s  |  0 |
+----+----+----+----+----+----+----+----+
|9th |10th|11th|12th|13th|14th|15th|16th|
+----+----+----+----+----+----+----+----+
| s  | 0  |  s | 0  | 0  | 0  | 0  | 0  |
+----+----+----+----+----+----+----+----+
```

Figure 2 : A s bit selection example (on a sharing ration 1:32 address).

Subscriber ID pattern is formed by setting all the s bits to 1 and other trivial bits to 0. Figure 2 illustrates an example of subscriber ID pattern on a sharing ration 1:32 address. Note that the subscriber ID pattern will be different, guaranteed by the random s bit selection, on every restricted IP address no matter whether the sharing ratio varies.The NAT offload AFTR can use subscriber ID pattern as port set ID on a per restricted IPv4 address basis, which allows log entries scale on a subscriber basis, hence meets the log efficiency requirements described in Section 3.1.2.

```
|1st |2nd |3rd |4th |5th |6th |7th | 8th|
+----+----+----+----+----+----+----+----+
| 0  | 1  | 0  | 0  | 1  | 0  | 1  |  0 |
+----+----+----+----+----+----+----+----+
|9th |10th|11th|12th|13th|14th|15th|16th|
+----+----+----+----+----+----+----+----+
| 1  | 0  | 1  | 0  | 0  | 0  | 0  | 0  |
+----+----+----+----+----+----+----+----+
```

Figure 3 : A subscriber ID pattern example (on a sharing ration 1:32 address).

Subscribers ID value is then assigned by setting subscriber ID pattern bits (s bits shown in the following example) according to a customer value and setting other trivial bits to 1.

```
|1st |2nd |3rd |4th |5th |6th |7th | 8th|
+----+----+----+----+----+----+----+----+
| 1  | s  | 1  | 1  | s  | 1  | s  | 1  |
+----+----+----+----+----+----+----+----+
|9th |10th|11th|12th|13th|14th|15th|16th|
+----+----+----+----+----+----+----+----+
| s  | 1  |  s | 1  | 1  | 1  | 1  | 1  |
+----+----+----+----+----+----+----+----+
```

Figure 4 : A subscriber ID value example (0# subscriber on this restricted address).

Subscriber ID pattern and subscriber ID value together uniquely defines a non-overlapping port set on a restricted IP address.

Pseudo-code shown in the Figure 4 describe how to use subscriber ID pattern and subscriber ID value to implement a random ephemeral port selection function in a restricted port set.

```
         do{
             restricted_next_ephemeral = (random()| customer_ID_pattern)
                                         & customer_ID_value;
             if(five-tuple is unique)
             return restricted_next_ephemeral;
         }
```

Figure 5   : Random ephemeral port selection of restricted port set algorithm.

7. IANA Considerations

TBD.

8. References

8.1. Normative References

[RFC2119]

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

8.2. Informative References

[I-D.bajko-pripaddrassign]

Bajko, G., Savolainen, T., Boucadair, M., and P. Levis, "Port Restricted IP Address Assignment", draft-bajko-pripaddrassign-03 (work in progress), September 2010.

[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.cui-softwire-dhcp-over-tunnel]
Cui, Y., Wu, P., and J. Wu, "DHCPv4 Behavior over IP-IP tunnel", draft-cui-softwire-dhcp-over-tunnel-01 (work in progress), July 2011.

[I-D.cui-softwire-host-4over6]
Cui, Y., Wu, J., Wu, P., Metz, C., Vautrin, O., and Y. Lee, "Public IPv4 over Access IPv6 Network", draft-cui-softwire-host-4over6-06 (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.sun-v6ops-laft6]

Sun, Q. and C. Xie, "LAFT6: NAT offload address family transition for IPv6", draft-sun-v6ops-laft6-01 (work in progress), March 2011.

9. Acknowledgments

Thank Alain Durand, Ole Troan and Ralph Dorm for their valuable feedback and discussion to this appraoch, and thanks to Qiong Sun for a discussion from operators needs' perspective.

Appendix A. Variants of this approach

A.1. Introduction

This section defines variants of deployment for this NAT offload DS-Lite approach. A.2 describes its combination with stateless encapsulation.

A.2 Stateless Encapsulation

B4 may implement the stateless encapsulation specified in Section 4.4 of [I-D.ymbk-aplusp].

Authors' Addresses

Xiaohong Deng
France Telecom
Email: xiaohong.deng@orange.com

Mohamed Boucadair
France Telecom
Rennes,   35000
France

Email: mohamed.boucadair@orange.com

Cathy Zhou
Huawei Technologies
Bantian, Longgang District
Shenzhen 518129
P.R. China

Phone:
Email: cathyzhou@huawei.com

Tina Tsou
Huawei Technologies (USA)
2330 Central Expressway
Santa Clara, CA 95050
USA

Phone: +1 408 330 4424
Email: tena@huawei.com

Gabor Bajko
Nokia

Email: gabor.bajko@nokia.com