Network Working Group   Q. Sun
Internet-Draft   C. Xie
Intended status: Standards Track   China Telecom
Expires: March 28, 2012   Y. Cui
                                                              J. Wu
                                                              P. Wu
                                                Tsinghua University
                                                            C. Zhou
                                                Huawei Technologies
                                                             Y. Lee
                                                            Comcast
                                                 September 25, 2011

Stateless 4over6 in access network
draft-sun-softwire-stateless-4over6-00

Abstract

   This document specifies a stateless 4over6 mechanism which moves the
   translation function from tunnel concentrator (AFTR) to initiators
   (B4s). It is used for service providers offering residual deployment
   of the IPv4 service across IPv6-only domains. The concentrator will
   record the full set of prefix mapping rules independent of the number
   of subscribers, while the initiator will only keep a single rule of
   its own sub-domain. The BNG devices can learn the rule set further,
   to achieve traffic optimization between initiators. In this way, it
   not only achieve scalability and simplicity feature benefits from
   stateless address mapping, but also keep CPE functions simple.
   Besides, flexible addressing and scattered IPv4 address blocks can be
   supported as well. Traffic optimization can be deployed
   incrementally when necessary.

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 March 28, 2012.

Table of Contents

1. **Introduction**

   Global IPv4 address exhaustion is becoming reality. The dramatic
   growth of the Internet is accelerating the consumption of available
   IPv4 addresses together with the scattered IPv4 address blocks, which
   makes the address shortage problem even worse. It is widely accepted
   that IPv6 is the only answer to solve the address shortage problem
   and sustain the long-term growth of the Internet. However, IPv6
   deployment is a huge systematic project and a lot of challenges will
   arise, especially in large commercial operational network.

   In the course of IPv6 transition, CPE is always a big issue for the
   whole industry. Since CPE is a customer-side device with restricted
   resource, it is cost sensitive. Moreover, there is lots of legacy
   CPEs which are owned by customers, in this case, ISP can not upgrade
   them . Besides, due to the huge number of CPEs, it is of vital
   importance to keep it simple for easy management and trouble
   shooting.

   As discussed in [I-D.ietf-softwire-stateless-4v6-motivation],
   stateless approaches would have good features of high scalability,
   reliability and robustness, etc. It is easy to achieve multi-vendor
   redundancy. However, since stateless solutions have essentially
   infered the mapping relationship between IPv4 address and IPv6
   address, it would have some kind of impact on network planning, e.g.
   address planning, routing, CPE, etc. From service provider
   perspective, there are several operational requirements on stateless
   approaches.

   Firstly, flexible addressing is needed especially when the remaining
   IPv4 address blocks are rather scattered. Network providers might do
   address reallocation in their network in the future, e.g. reallocate
   global IPv4 address for address sharing. And this kind of addressing
   adjustment need to have little impact on the whole network
   infrastructure, e.g. informing thoudsands of CPEs within the same
   domain.

   Secondly, we need to keep CPE as simple as possible. This will not
   only reduce the great burden on CPE management and trouble shooting,
   but also reduce cost and implemenation complexity.

   Thirdly, stateless solutions should be capable of covering a large
   area with centralized deployment. Since stateless approaches can
   achieve high scalability and no performance bottleneck, centralized
   deployment is good for network management and multi-vendor industry.
   Given the fact that large scale service providers would normally have
   few thousand BNGs with different vendors and different versions, it
   is a huge burden to update all existing BNGs with the new service

card.

Finally, CPE-CPE traffic optimization should be deployed
incrementally. Currently, network architecture for some service
providers is already rather flat, and there is no direct link between
CPEs. The traffic within the same BNG occupies a small percentage,
so CPE-CPE traffic optimization is NOT a MUST in general. It can be
only deployed incrementally when necessary.

In this document, we propose a stateless 4over6 approach for service
providers offering residual deployment of the IPv4 service across
IPv6-only domains. In stateless 4over6, the concentrator will record
the full set of prefix mapping rules including IPv4 prefix, IPv6
prefix and multiplexing ratio, while the initiator will only keep a
single rule of its own sub-domain. Thus, all traffic will be firstly
tunneled to concentrator by default. The BNG devices can learn the
rule set further, to achieve traffic optimization between initiators.

This procedure keeps the advantages of stateless 4over6 mechanisms,
and prevents the customer devices from complex configuration/
delegation. It "moves" the mapping rules from customer devices up to
concentrator and BRAS devices, in which the amount of rules is
acceptable. It can still keep the flexible feature of centralized
addressing. This is also an extended case which covers stateless
addressing sharing for [I-D.cui-softwire-host-4over6]. However, this
extension will decrease IPv4 address utilization. Moreover, since
port randomization algorithm must use ports within this port-range,
it may cause the port randomization algorithm more predictable.

2. Requirements Language

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 [RFC2119].

3. Terminology

   Stateless 4over6: stateless 4over6 is an IPv4-over-IPv6 hub and spoke
   mechanism proposed in this document, which supports address sharing
   to alleviate the problem of IPv4 address exhaustion, and places the
   IPv4 translation on the initiator side.

   Stateless 4over6 initiator (or "initiator" for short): the tunnel
   client module of stateless 4over6 mechanism. The stateless 4over6
   initiator could be a host directly connected to IPv6, or an dual-
   stack CPE in front of an IPv4 local network. It is responsible for
   IPv4 public-private translation besides tunnel encapsulation and
   decapsulation.

   Stateless 4over6 concentrator (or "concentrator" for short): the
   tunnel servicing module in stateless 4over6 mechanism. The stateless
   4over6 concentrator connects the ISP IPv6 network and IPv4 Internet.
   It provides tunnel encapsulation and decapsulation but no IPv4
   private-public translation.

   Stateless 4over6 domain(or "SL Domain" for short): The IPv6 network a
   concentrator covers. Its coverage depends on the placement of the
   concentrator. For example, if the concentrator is deployed at the
   Core of MAN, the Domain will cover the whole MAN.

   Stateless 4over6 sub-domain(or "sub-domain" for short): A IPv6 sub-
   network where different initiators can share a same IPv4/IPv6 prefix
   for stateless mapping. It usually depends on the range of IPv4
   address block and multiplexing ratio.

   Stateless 4over6 subPre6(or "subPre6" for short): A common IPv6
   prefix in the sub-domain for stateless address mapping. subPre6 is
   only a part of IPv6 perfix/address delegated to customers, which will
   be followed by "user-id" , etc.(refer to section 6 for detail).

   Stateless 4over6 subPre4(or "subPre4" for short): A common IPv4
   prefix in the sub-domain for stateless address mapping. Different
   subscribers in the common sub-domain will have a different identifier
   (indicated by "Free bits"). For example, if a given subPre4 is
   "202.112.1/24" and the shared public address for the specific
   subscriber is "202.112.1.5", then the "Free bits" is "5".

   Stateless sub-domain Rule(or "sub-domain rule" for short): A mapping
   rule inculding subPre6, subPre4 and Multiplexing ratio for a specific
   sub-domain.

4. **Stateless 4over6 Overview**

   Figure 1 provides an overview of the stateless 4over6 mechanism. A
   stateless 4over6 initiator which can be either an IPv6 hosts or CPE,
   and the stateless 4over6 concentrator are seperated by the IPv6
   network in the middle, so they use IPv4-in-IPv6 tunnel to build IPv4
   connectivity. One concentrator will cover a SL Domain and there will
   be one or more sub-domains in it. Each sub-domain will only have one
   sub-domain rule indicating the common subPre6, subPre4 and
   multiplexing ratio. These sub-domains can be divided in a rather
   flexible way. One BNG can have one or more sub-domains, and multiple
   BNGs can also belong to one sub-domain. The only rule for sub-domain
   to follow is to have a unique sub-domain rule with a unified
   multiplexing ratio.

```
                   +------------------------------+
                   |     Stateless 4over6 Domain  |
              +---+---------------+--------------|
 +--------+   +      Sub-domain   |              |
 |        |   | +---------+     +--+--+          |
 |IPv4 LAN|--|SL 4over6|     |     |          |
 |        |   | |Initiator|=====| BNG | =====+---------+   +-----------+
 +--------+   +---------+     +--|--+         |SL 4over6|   |   IPv4    |
 +--------+       +--------------+             |Concen-  |---| Internet  |
 |        |   +---|-----+     +--+--+          |trator   |   |           |
 |IPv4 LAN|--|SL 4over6|=====|     | =====+---------+   +-----------+
 |        |   | |Initiator|     | BNG |          |
 +--------+   +---------+     +--+--+          |
              +      Sub-domain   |              |
              +------------------+--------------+
```
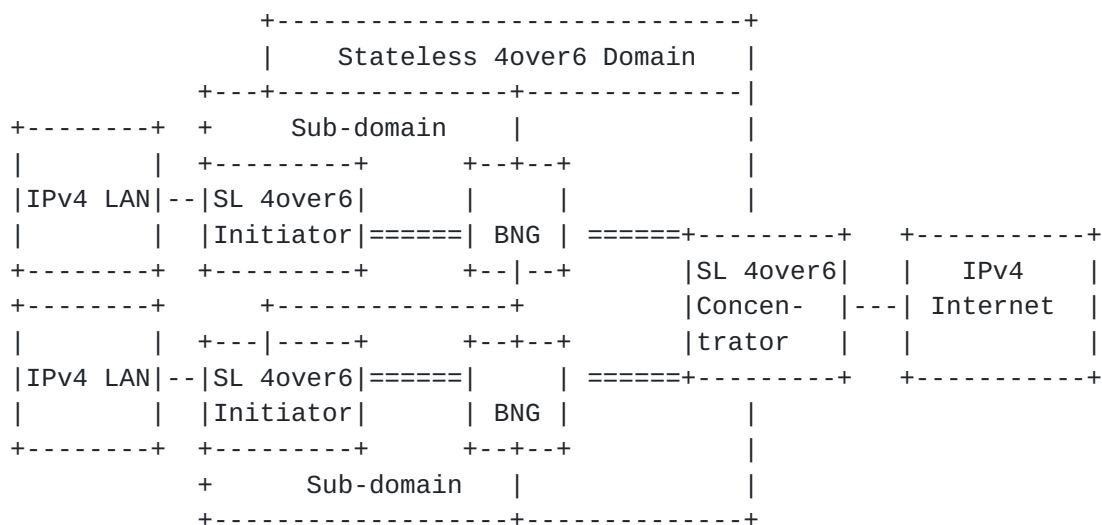
   Figure 1 Stateless 4over6 overview

   An initiator will get its sub-domain rule containing subPre6, subPre4
   and multiplexing ratio. And it will also get a user prefix which
   includes a user-id via PPPoE, etc. By combining the user-id and
   multiplexing ratio, we can deduce the vaule of the "Free bits" and
   "Port-set id". SubPre4 and "Free bits" will then construct a public
   address for the customer, while "Port-set id" can be mapped to a
   certain port set by port mapping algorithm.

   In this way, an initiator has known its shared IPv4 address and valid
   port range. For a IPv4 outbound packet arriving at initiator, it
   will do port-restricted NAT44, then encapuslate with IPv6 header and
   tunneled to the concentrator by default. The source address and
   destination address format in IPv6 header will be discussed in the
   next section. The concentrator will only do packet de-capsulation
   for outbound packet. When the inbound traffic arriving at

concentrator, it will lookup sub-domain rule table based on
destination IPv4 address and find out its corresponding subPre6.
Then it will construct an IPv6 header algorithmically. Since the
number of sub-domain rules is independent of subscriber numbers, it
is still a algorithm-based stateless method. Then the inbound
traffic will be sent back to customer and perform NAT44 again.

5. **Port Restricted IPv4 Address Allocation**

   As described above, a stateless 4over6 initiator needs the sub-domain
   rule containing subPre6, subPre4 and multiplexing ratio. Since these
   parameters are relatively stable, it can be configured in a variety
   of provisioning methods, including
   DHCPv6[I-D.cui-softwire-dhcp-over-tunnel], "TR-069", etc. For
   customer's IPv6 prefix, it will be delegated just in traditional way,
   e.g. DHCPv6, Prefix Delegation, etc. And there is no impact on
   current IPv6 addressing model. The concentrator address can be
   passed through the same option of ds-lite AFTR address.

6. Address/Prefix Mapping

6.1. Address/Prefix Mapping format

   The subscriber source address mapping format is in consistent with
   [I-D.xli-behave-divi-pd] expect for the last 64 bits. Since
   stateless 4over6 is a tunnel-based solution, we just set the last 64
   bits to zero (depicted in Figure2).

```
+--+---------+-------+------+-----+----+----------+----------+----+
|PL| 0-------32------48-----56----64---72----------104---------120  |
+--+---------+-------+------+-----+----+----------+----------+----+
|64|         prefix            | u  |          0                   |
+--+---------+-------+------+-----+----+----------+----------+----+
|e |  Prefix-sub |user(CPE)-id| 0 | u  |          0           | 0  |
|x |-------------+------------+---+----+----------+----------+----|
|t |(d) bits    |(s+k) bits  |(m)|(8) |          (48)        |(8) |
+--+---------+-------+------+-----+----+----------+----------+----+
```
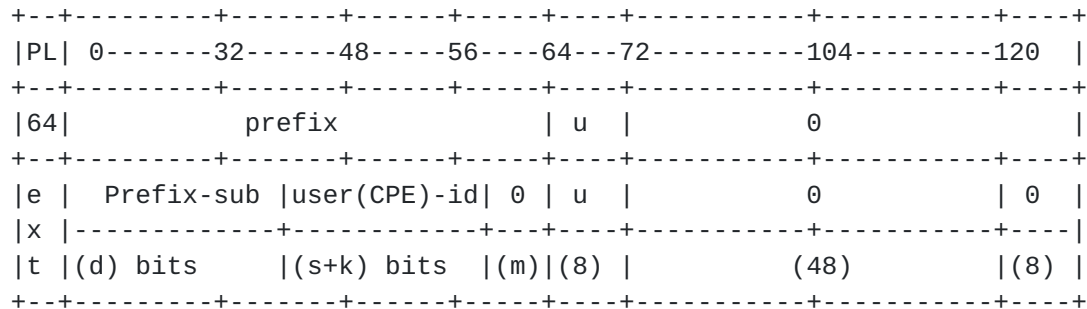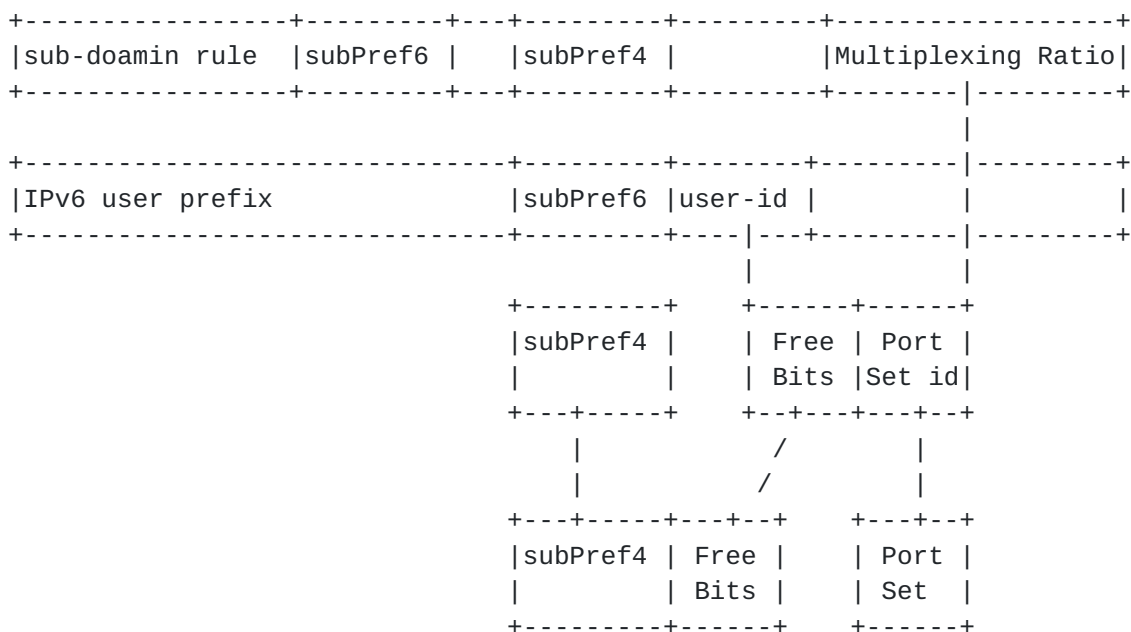
   Figure 2 Subscriber Source address mapping format

   The user-id is consisted of the "Free Bits" and "Port-set id".

   The destination address is the concentrator address in CPE. However,
   when BRAS is processing traffic optimization, the destination address
   needs to be performed in a similar way with subscriber source
   address.

6.2. Address/Prefix Mapping Procedure

```
+------------------+---------+---+---------+---------+-----------------+
|sub-doamin rule  |subPref6 |   |subPref4 |         |Multiplexing Ratio|
+------------------+---------+---+---------+---------+--------|---------+
                                                             |
+------------------------------+---------+--------+---------|---------+
|IPv6 user prefix             |subPref6 |user-id |         |         |
+------------------------------+---------+----|---+---------|---------+
                                             |   |         |
                              +---------+    +------+------+
                              |subPref4 |    | Free | Port |
                              |         |    | Bits |Set id|
                              +---+-----+    +--+---+---+--+
                                  |             /        |
                                  |            /         |
                              +---+-----+---+--+    +---+--+
                              |subPref4 | Free |    | Port |
                              |         | Bits |    | Set  |
                              +---------+------+    +------+
```

                             Shared IPv4 address

      Figure 2 Address/Prefix Mapping Procedure

      The above figure depicts address mapping procedure in stateless
      4over6. An initiator will get its sub-domain rule containing
      subPre6, subPre4 and multiplexing ratio. And it will also get a user
      prefix incluing subPref6 and user-id. By combining the user-id and
      multiplexing ratio, we can deduce the vaule of the "Free bits" and
      "Port-set id". SubPre4 and "Free bits" will then construct a public
      address for the customer, while "Port-set id" can be mapped to a
      certain "Port Set" by port mapping algorithm.

      Detailed descriptions will be added soon...

6.3. Port mapping algorithm

      A unified Port mapping algorithms should be defined to determine the
      mapping rule between Port-set id and Port-set. You can refer to
      [I-D.xli-behave-divi-pd] for example. And we will complete this
      section soon ...

7. **Stateless 4over6 Initiator Behavior**

   When requiring IPv4 access, the initiator needs to get its IPv6
   address/prefix via normal IPv6 address allocation precedure first,
   then it will calculate its port-restricted IPv4 address and valid
   port range.

   The data plane functions of the initiator includes translation and
   encapsulation/decapsulation. For CPE initiator case, the initiator
   runs a standard local NAT44 with the address pool consist of the
   calculated port restricted address. When sending out an IPv4 packet
   with private source address, it performs NAT44 function and
   translates the source address into public. Then it encapsulates the
   packet with concentrator's IPv6 address as destination IPv6 address,
   and forwards it to the concentrator. The source address of the IPv6
   packet will be generated as described above. When receiving an IPv4-
   in-IPv6 packet from the concentrator, the initiator decapsulates the
   IPv6 packet to get the IPv4 packet with public destination IPv4
   address. Then it performs NAT44 function and translates the
   destination address into private one.

   For host initiator case, it is suggested that the host runs a local
   NAT to map randomly generated ports into the restricted, valid port
   range. Private-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, vailid range to applications. Either
   way makes sure that every port number in the packets sent out by
   itself falls into the allocated port range.

8. **Stateless 4over6 Concentrator Behavior**

   The stateless 4over6 concentrator should record the full set of sub-
   domain rules. They can also be configured in a variety of methods,
   and they can be refreshed or deleted once there is an adjustment on
   address planning (no Initiator needs to be informed with this
   adjustment).

   The data plane functions of the concentrator is purely encapsulation
   and de-capsulation. When receiving an IPv4-in-IPv6 packet from an
   initiator, the concentrator de-capsulates it and forwards it to IPv4
   Internet. When receiving an IPv4 packet from the Internet, it uses
   the destination address and port from this packet to lookup the sub-
   domain rules and calculate the specific "user-id" with port mapping
   algorithm. Then the concentrator encapsulates this packet and
   forwards it to the correct initiator based on native IPv6 routing.
   The source address of IPv6 packet is just the concentrator address.

9. **CPE-CPE optimization**

   In order to support CPE-CPE optimization, it will delegate the full
   sub-domain rules to BRAS device in the 4over6 domain. When the IPv4
   destination of an IPv4-in-IPv6 packet is in the 4over6 domain, BRAS
   performs IPv6 destination address translation, according to sub-
   domain rules from the set, and the IPv4 destination address + port in
   the packet. After the translation, the IPv6 packet can be forwarded
   to the destination CPE directly without going through the
   concentrator. However, this functionality is not a must in cast the
   network is flat and there is no direct link between different CPEs.

10. Mechanism Analysis

   Stateless 4over6 does not need to inform multiple prefix mapping
   rules to CPEs directly. Thus, there is no need to introduce dynamic
   protocol for distributing and maintaining 4rd rules. This will turn
   CPE to be a much simpler/dummier client, which will also
   significantly reduce the burden of CPE management and trouble
   shooting.

   Besides, stateless 4over6 would be easier for address planning. SPs
   only need to pre-determine the mapping relationship between IPv4
   prefix and IPv6 prefix. Only concentrators(and BRAS in CPE-CPE
   optimization case) need to be informed when changing a mapping rule.
   And it fits well for scattered IPv4 address blocks. Moreover,
   incremental CPE-CPE optimization can be achieved when network
   providers would like to adjust their traffic when necessary.

   The costs of stateless 4over6 for achieving the above benefits are
   less optimized traffic by default. All traffic will be tunneled to
   the concentrator when there is no prefix mapping BNGs. This is still
   acceptable in flat network and few direct physical links between
   different BNGs. However, if the network architecture is not flat and
   concentrator is in the upper level of the network, it is recommended
   that CPE-CPE optimization should be introduced.

11. Security Considerations

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

12. 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.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.

[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: Lightweight address family
          transition for IPv6", draft-sun-v6ops-laft6-01 (work in
          progress), March 2011.

[I-D.tsou-pcp-natcoord]
          Zhou, C., ZOU), T., Deng, X., Boucadair, M., and Q. Sun,
          "Using PCP To Coordinate Between the CGN and Home Gateway
          Via Port Allocation", draft-tsou-pcp-natcoord-03 (work in
          progress), July 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.

   [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.

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

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

Authors' Addresses

   Qiong Sun
   China Telecom
   Room 708, No.118, Xizhimennei Street
   Beijing 100035
   P.R.China

   Phone: +86-10-58552936>
   Email: sunqiong@ctbri.com.cn

   Chongfeng Xie
   China Telecom
   Room 708, No.118, Xizhimennei Street
   Beijing 100035
   P.R.China

   Phone: +86-10-58552116>
   Email: xiechf@ctbri.com.cn

   Yong Cui
   Tsinghua University
   Department of Computer Science, Tsinghua University
   Beijing 100084
   P.R.China

   Phone: +86-10-62603059
   Email: yong@csnet1.cs.tsinghua.edu.cn

   Jianping Wu
   Tsinghua University
   Department of Computer Science, Tsinghua University
   Beijing 100084
   P.R.China

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

   Peng Wu
   Tsinghua University
   Department of Computer Science, Tsinghua University
   Beijing 100084
   P.R.China

      Phone: +86-10-62785822
      Email: weapon@csnet1.cs.tsinghua.edu.cn

      Cathy Zhou
      Huawei Technologies
      Section B, Huawei Industrial Base, Bantian Longgang
      Shenzhen 518129
      P.R.China

      Phone: +86-10-58552116>
      Email: cathyzhou@huawei.com

      Yiu L. Lee
      Comcast
      One Comcast Center
      Philadelphia, PA 19103
      USA

      Email: yiu_lee@cable.comcast.com