Redundancy and Load Balance Mechanism of NAT64

draft-xu-behave-nat64-standby-00

Status of this Memo

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

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt.

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html.

This Internet-Draft will expire on October 29, 2009.

Copyright Notice

Copyright (c) 2009 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust’s Legal Provisions Relating to IETF Documents in effect on the date of publication of this document (http://trustee.ietf.org/license-info). Please review these documents carefully, as they describe your rights and restrictions with respect to this document.
Abstract

NAT64 [NAT64], a simplified NAT-PT [RFC2766] without DNS-ALG, provides a method for IPv6 hosts to initiate communications with IPv4 hosts. This memo defines several mechanisms supporting redundancy and load balance amongst NAT64 boxes.

Conventions used in this document

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].

Table of Contents

1. Introduction ............................................. 3
2. Terminology ............................................. 3
3. Scenario Description ..................................... 3
4. Redundancy Mechanisms ................................... 4
   4.1. Cold Standby Mechanism ............................... 4
   4.2. Hot Standby Mechanism ............................... 5
5. Load Balance Mechanisms ................................. 6
   5.1. Load Balance in a Cold Standby Mechanism .......... 6
   5.2. Load Balance in a Hot Standby Mechanism .......... 6
6. Election Protocol Consideration ........................ 7
7. Security Considerations .................................. 7
8. IANA Considerations ..................................... 8
9. Acknowledgments .......................................... 8
10. References ............................................. 8
    Authors’ Addresses .................................... 9
1. Introduction

NAT-PT [RFC2766] is an IPv4/IPv6 translation solution. However, due to the reasons described in [RFC4966], NAT-PT has been reclassified from proposed standard to historic status. NAT64 [NAT64] is a simplified NAT-PT, which provides a scalable method for IPv6 hosts to initiate communications to IPv4 hosts. In this memo, we specify several mechanisms for redundancy and /or load balance among a set of NAT64 boxes.

2. Terminology

NAT64: is a translator device that translates communications initiated from the IPv6 side to the IPv4 side. See [NAT64] for more information.

Prefix64: is an IPv6 prefix used for synthesizing IPv6 addresses for the IPv4 hosts. See [Prefix64] for more information.

3. Scenario Description

In typical operational scenarios, two NAT64 boxes are usually used for redundancy and /or load balance purpose. Therefore, to benefit the discussion, we describe the mechanisms mainly using the scenarios where there are only two boxes (See Figure 1. Note that the mechanisms proposed in this demo can be easily used in scenarios where there are more than two boxes.

We assume that IPv6 hosts have already obtained the synthesized IPv6 addresses for IPv4 hosts through one of those approaches described...
in [DNS64], [Learn Prefix64] etc. So we will skip the process of obtaining the synthesized IPv6 addresses and directly specify the redundancy and/or load balance mechanisms.

4. Redundancy Mechanisms

This section introduces two standby mechanisms, the cold standby mechanism and the hot standby mechanism. These redundancy mechanisms are able to keep the switchover of the NAT64 boxes transparent to the hosts, especially to the IPv6 hosts, to different extents.

The cold standby mechanism doesn’t need the mapping states to be synchronized among the NAT64 boxes. With this mechanism, the already established connections (e.g., TCP connections) will be interrupted due to the switchover of NAT64 boxes, but later they can be re-established without the notice of applications on the communicating hosts. The hot standby mechanism, in contrast, requires the mapping states to be synchronized in a timely fashion among the involved NAT64 boxes. With this mechanism, the already established connections can be preserved even when the switchover of NAT64 boxes occurs.

4.1. Cold Standby Mechanism

For cold standby, the prefix64s used by the NAT64 boxes in Figure 1 (i.e., NAT64-A and NAT64-B) should be identical. However, there should be no overlap in their IPv4 address pools. By running some election protocol (see section 4.1), one is designated as the Master and the other as the Backup. The Master announces a route to that prefix64 on the IPv6 network side and a route to its own IPv4 address pool on the IPv4 network side. The Backup could either announce the route to that prefix64 at much higher cost or not announce it at all on the IPv6 network side, and it could either announce the route to its IPv4 pool or not on the IPv4 network side. The benefit of advertising those routes by the Backup is to achieve fast failover. However, the cost of the route to that prefix64 advertised by the Backup must be set high enough so as to ensure the route advertised by the Master always to be selected as the best by those routers within IPv6 network despite of topology changes, as long as the Master is still available. Since these NAT64 boxes use different IPv4 address pools, there is no such issue on the IPv4 network side.

Apart from election protocols, we can also achieve the similar effect through manual configuration. For example, we set one box as the Master and the other as the Backup. The Master announces a route...
to that prefix64 on the IPv6 network side and a route to its own IPv4 address pool on the IPv4 network side. The Backup should announce the route to that prefix64 at a high enough cost on the IPv6 network side and a route to its own IPv4 address pool on the IPv4 network side. Once a NAT64 box, no matter the Master or the Backup, loses connections with the IPv4 network or the IPv6 network, it must withdraw the routes announced previously. Once the Master fails, the route to the prefix64 advertised by the Backup, though with a higher cost, will now be looked as the best by those routers within IPv6 network since the route advertised by the Master has either been withdrawn or unavailable.

Since the NAT64 boxes use different IPv4 address pools without any overlap, the already established connections are interrupted when the switchover of the NAT64 boxes occurs. However, the IPv6 hosts can re-establish those connections without the notice of applications on the communicating hosts.

4.2. Hot Standby Mechanism

To achieve hot standby, the two NAT64 boxes shown in Figure 1 should use not only the same prefix64 but also the same IPv4 address pool. By running a certain election protocol, a box is designated as the Master, and the other is designated as the Backup. The Master announces a route to the prefix64 on the IPv6 network side and a route to the IPv4 address pool on the IPv4 network side. The Backup doesn’t need to announce them at all. To reduce the interrupting duration further during the failover, the Backup could also announce those routes but the costs of them must be set high enough so as to guarantee the route advertised by the Master always to be selected as the best by those routers within IPv6 network despite of topology changes, as long as the Master is still available.

Besides the election protocol, we can also achieve the similar effect through manual configuration. For example, we set one box as the Master and the other as the Backup. The Master announces a route to the prefix64 on the IPv6 network side and a route to its IPv4 address pool on the IPv4 network side. The Backup also announces those routes but with much higher costs. Since this mechanism is almost the same as that described in section 3.1.1, we do not go into details.

The packets between the IPv4 network and the IPv6 network travel via the Master in normal cases. Meanwhile, the translation mapping states on the Master are synchronized by a certain mapping state synchronization protocol (e.g., Server Cache Synchronization...
Protocol (SCSP) [RFC2334]) to the Backup in a timely fashion. Once
the Master fails, the Backup becomes the new Master and takes over
the translation and forwarding responsibility to provide
uninterrupted service to the hosts.

Because the Master and the Backup use the same IP address pool and
synchronize the mapping states timely, the established connections
will not be interrupted during the switchover of the NAT boxes.

5. Load Balance Mechanisms

On the basis of the standby mechanisms mentioned above, we can
further realize load balance among a set of NAT64 boxes. The basic
idea is as follows: these NAT64 boxes use two prefix64s(e.g.,
prefix64-A and prefix64-B), and one box is designated as the Master
for prefix64-A and the Backup for prefix64-B and the other as the
Backup for prefix64-A and the Master for prefix64-B, and half of the
IPv6 hosts use prefix64-A and half are using prefix64-B. In this way,
the IPv6 packets towards IPv4 network is balanced among a set of
NAT64 boxes according to their destination addresses with different
prefix64. Note that the both outbound and inbound packets of the
same connection will pass through the same NAT64 box.

This demo does not discuss the issues with achieving best balance by
adjusting the numbers of hosts adopting different prefix64s.
Interested readers are referred to [DNS64] and [Learn Prefix64] for
more details.

5.1. Load Balance in a Cold Standby Mechanism

To achieve load balance in a cold standby mechanism, there are two
options for NAT64s. One option is to use the same IPv4 address pool
corresponding to different prefix64. In this case, a NAT64 box needs
to determine which prefix64 should be used when translating a
received IPv4 packet to a IPv6 packet. So the prefix64 used by each
connection should be recorded in its NAT mapping table. Another
option is to use different IPv4 address pools corresponding to
different prefix64s. In this way, the NAT64 box could easily
determine which prefix64 should be used according to which IPv4
address pool the destination address belongs to.

5.2. Load Balance in a Hot Standby Mechanism

As for load balance in a hot standby mechanism, the NAT64 box should
use different IPv4 address pools corresponding to different
prefix64s. Otherwise, the inbound IPv4 packets may pass through a
different NAT64 box than that the outbound IPv6 packets of the same connection pass through. When translating a received IPv4 packet to an IPv6 packet, the NAT64 box could easily determine which prefix64 should be used according to which IPv4 address pool the destination address belongs to.

6. Election Protocol Consideration

The election protocol is used to dynamically designate one from a set of NAT64 boxes as the Master, which is responsible for IPv4/IPv6 translation and forwarding. Once the election is done, the Master will announce a route to the prefix64 on the IPv6 side and a route to the IPv4 address pool on the IPv4 side, and the Backup will either announce those routes at much higher costs or not announce them at all. If the Master becomes unavailable then the Backup with the highest priority will transit to Master after a short delay. The election protocol will also track the NAT64 box’s connectivity to the IPv4 network and the IPv6 network, once the NAT64 box loses connection to the IPv4 network or the IPv6 network, its priority is set to zero, which means it is not suitable to be a candidate for the Master any more and it will withdraw the route to the prefix64 and the route to the IPv4 address pool advertised previously.

In fact, we can use the VRRP [RFC2338] directly as the automatic election protocol. If two NAT64 boxes are directly connected via an Ethernet network, the VRRP can run directly on the Ethernet interfaces. Otherwise, some extra configurations or protocol changes need to be implemented. One option is to create conditions for the VRRP to run among these boxes. For example, we create a VPLS [RFC4761, RFC4762] instance and enable IP functions and run VRRP on those VLAN interfaces which are bounded to that VPLS instance. If enabling IP function on those interfaces bound to VPLS instances is not supported, we can use the following trick to realize the same goal, but at a cost of consuming two physical interfaces on each NAT64 box. The approach is: create a VPLS instance among a set of NAT boxes, and on each of them one Ethernet interface is bound to that VPLS instance, another IP enabled Ethernet interface is locally connected with that interface. Then the VRRP can run on those IP enabled Ethernet interfaces which are connected through that VPLS instance. Another option is to do some change to the VRRP so that VRRP neighbors can be configured manually and VRRP messages can be exchanged directly between two neighbors in a unicast fashion.

7. Security Considerations

TBD.
8. IANA Considerations

TBD.

9. Acknowledgments

The author would like to thank Dacheng Zhang and Xuwei Wang for their valuable comments and reviews.

10. References


Authors’ Addresses

Xiaohu Xu
Huawei Technologies,
No.3 Xinxi Rd., Shang-Di Information Industry Base,
Hai-Dian District, Beijing 100085, P.R. China
Phone: +86 10 82836073
Email: xuxh@huawei.com