Individual Submission
INTERNET-DRAFT
Expired: November 2003
Filename:
draft-park-scalable-multi-natpt-00.txt

S. Daniel Park
SAMSUNG Electronics
Senthil Sivakumar
Cisco Systems, Inc
Pyda Srisuresh
Caymas Systems, Inc
May 2003

Scalable mNAT-PT Solution

Status of This Memo

Abstract

This document provides scalability extension to NAT-PT. The extension is based on the use of DNS-ALG and exchange of load metrics amongst a cluster of NAT-PT devices. We refer such a NAT-PT device as mNAT-PT. mNAT-PT is valuable in connecting large V6 domains to legacy V4 domain.

Table of Contents

1. Introduction

In order to widely deploy IPv6 network, V4/V6 transition mechanisms are essential. NAT-PT and TRT transition solutions are proposed for enable connectivity between IPv6-only and IPv4 networks. However, both these solutions have limitations for large size V6 networks. This draft focuses on scaling extensions for traditional NAT-PT. Specifically, a method to permit outbound sessions for IPv6 hosts in a large IPv6-only domain to the legacy IPv4 domain using multiple mNAT-PT translators.

2. Scaling Considerations

The NAT-PT solution defined in [NATPT] does not address scalability issue. Although TRT [TRT] considered scaling, it mandates reconfiguration of existing systems such as host and DNS-server by the network administrator. This may not be feasible or desirable in many circumstances, especially when the V6 domain is constituted of mobile nodes. In this document, we propose mNAT-PT as an efficient scalable solution to address such environments. Unlike

TRT, mNAT-PT will not mandate changes to existing end-nodes.

3. Terminology

The following terminology is used throughout the document.

- NAT-PT device — A device that implements traditional NAT-PT function as described in [NATPT].
- mNAT-PT function — Stands for "Multiple NAT-PT". mNAT-PT function makes use of NAT-PT, DNS-ALG and real-time load monitoring functions to scale NAT-PT function to multiple NAT-PT devices.
- mNAT-PT device — A device that implements mNAT-PT function.
- Address Pool — IPv4 addresses to translate between IPv6 and IPv4 realms.
- Mapping Table — Mapping between IPv4 and IPv6 addresses in the NAT-PT (or) mNAT-PT device.
- Load threshold — This is an upper ceiling of NAT BINDings configured for a given mNAT-PT device. When a mNAT-PT device reaches the threshold, the mNAT-PT device attempts to assign a different mNAT-PT device for new sessions crossing the realms.
- ALG — Application layer gateway.

o DNS-ALG DNS Application layer gateway, as described in [NATPT].

4. V4/V6 topology with a single NAT-PT device

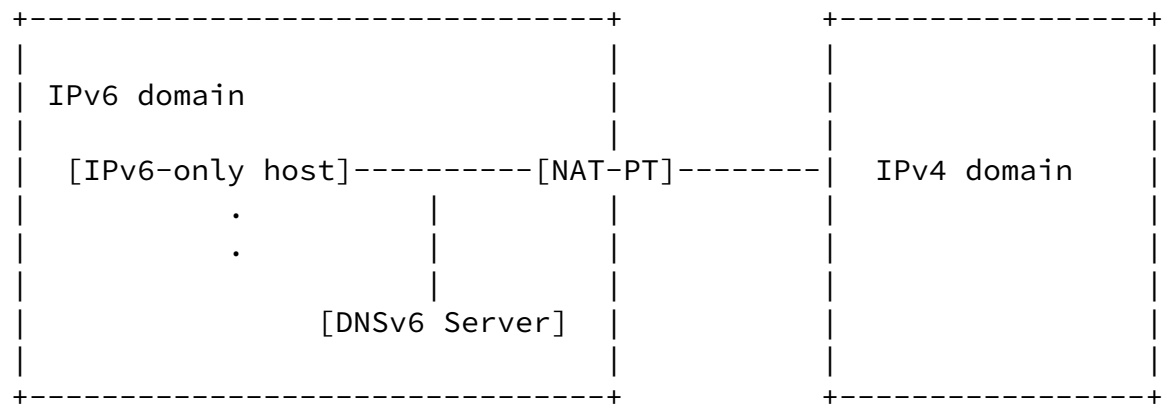


Figure 1 : single NAT-PT topology

As described in figure 1 above, IPv6-only hosts in the IPv6 domain are translated into IPv4 address by the NAT-PT device. NAT-PT is listed as the default router in IPv6 network. Also, there may be a DNSv6 Server within the IPv6 domain. The above solution cannot scale to support large no. of V6 hosts.

5. V4/v6 topology with multiple mNAT-PT devices

With growing deployment of IPv6-only networks, a single NAT-PT will not be able to scale. Support for large number of mobile users is a requirement in a 3GPP mobile network. We propose deploying multiple mNAT-PT devices on the border of the IPv6 domain as described below in figure 2. Each mNAT-PT device will perform NAT-PT function, host one or more unique IPv6 prefixes and advertise the prefixes within the IPv6 domain.

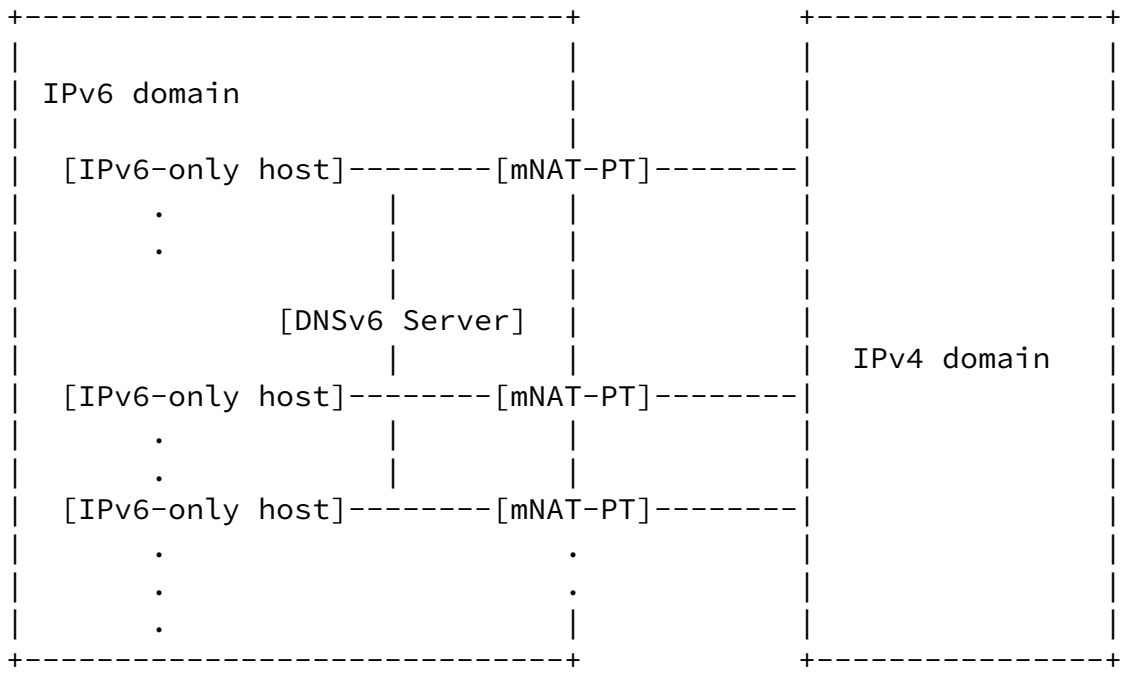
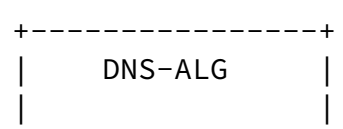


Figure 2 : mNAT-PT topology

6. Method of operation for mNAT-PT devices

The following layout describes how mNAT-PT function may be accomplished with the aid of DNS-ALG and Cluster Load-Monitor as an extension to NAT-PT function.



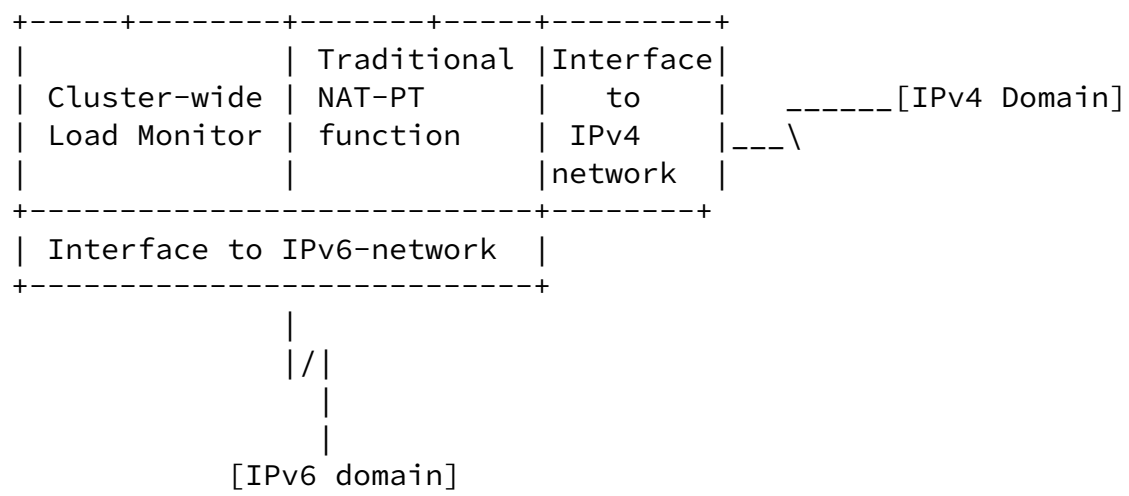


Figure 3 : mNAT-PT functional framework

A cluster of mNAT-PT devices may be configured to provide a scalable mNAT-PT solution to large V6 networks. All member nodes within a mNAT-PT cluster carry the same Cluster Identifier (ClusterId). The Load-Monitor entity within each mNAT-PT is responsible for relaying the real-time load on the hosting mNAT-PT periodically and in turn, collecting the load from member nodes within the cluster. The Load-monitor supplies real-time up-uo-date load data of member nodes to the DNS-ALG.

The NAT-PT entity is required to invoke DNS-ALG for all DNS queries and responses traversing the domains. While processing the DNS response, the DNS-ALG will use load on member nodes as the basis to assign a IPv6 prefix in the AAAA response to the V6-node that originated the DNS query. The IPv6 prefix assignment will be based on the load on the mNAT-PT node associated with the prefix.

6.1. Load-monitor and Communication amongst mNAT-PT cluster members

The following assumptions are made throughout the document. A few of these assumptions may be changed to suit specific deployment scenarios.

* Communication amongst the cluster members may take place within the IPv6 domain.

* It is assumed that each of the mNAT-PT member nodes have non-conflicting address maps and host a IP-v6 prefix that is also non-conflicting.

* This document uses NAT-BINDing load on member nodes as the criteria for determining which of the mNAT-PT devices is assigned to carry out a NAT-PT translation. However, load need not be the criteria or the only criteria to make the device selection. There may be other criteria or policies that decide the right mNAT-PT device.

* The document assumes that all mNAT-PT devices equal access to all V4 routes in the V4 domain. This need not be the case. In such a case, the mNAT-PT devices must either setup V4-over-V6 VPNs between themselves so this is ensured (or) exchange the V4 route reachability between themselves so the right prefix is assigned based on route reachability. The former is the preffered and recommended choice.

* The document assumes that the Cluster-ID, Cluster-controller and member nodes within a cluster are preconfigured on the mNAT-PT device. Dynamic discovery of Cluster members is out of the scope of this document. Election of initial or subsequent cluster controller is also outside the scope of this document. Cluster controller is assumed to be the first node of the cluster to come up. A cluster controller is required to be alive throughout the duration of the cluster.

* The document assumes there exists a V6 path for any of the V6-only hosts to reach any of the mNAT-PT devices.

* Lastly, [NIQ] does not seem like the right choice for cluster communication. Cluster communication requires reliable point-to-point data communication (say, TCP based sessions to a well-knwon port) and reliable point-to-multipoint communication. The document does not address the transport or message formats used for cluster communication at this time.

* The communication amongst mNAT-PT cluster memebers devices should be properly authenticated to avoid any malicious devices trying to add a IPv6 prefix in the active list and thereby causing the traffic to be directed to the malicious device.

An mNAT-PT node joins the cluster by sending a message with the following mNAT-PT configuration data to Cluster controller. The cluster controller, in turn, accepts the node into cluster by sending the existing cluster membership info, Load-data polling duration and mNAT-PT configuration for each of the member nodes. The periodic load-data update will also be used to validate the liveness of a memeber node. Alternately, member nodes may terminate their memebership by explicitly sendign a termination message to the cluster controller.

- The Cluster ID
- Hosted IP-V6 Prefix
- NAT-PT address map configuration
- BINDings threshold limit

Further to joining the cluster, the mNAT-PT nodes report their load-data to the cluster controller periodically. The load-data is essentially the count of active NAT-PT BINDings at the time of reporting.

6.2. Redirection using DNS-ALG

Each mNAT-PT must be configured with a threshold of NAT BINDings and one or more IP-v6 prefixes (as desccribed in the previous section) in advance. The DNS-ALG is supplied with this information from the Load-Monitor.

Typically, all the communications between a IPv4 device and IPv6 device starts with a DNS request. DNS-ALG receives the DNS request and converts the A query to a AAAA query and vice versa. When the DNS-ALG receives the reply then it will decide which IPv6 prefix to use to translate the IPv4 address. If the load on the hosting mNAT-PT is within threshold configured for the node, then the prefix assigned to the host mNAT-PT is included in the DNS response. Otherwise, DNS-ALG will select a member node with the least percentage of load utilization vis-a-vis the threshold setting.

7. Security Considerations

Cluster communication between cooperatign mNAT-PT devices must be authenticated so that sessions are not hijacked by a rogue node that pretends to be a member mNAT-PT device.

8. Intellectual Property

The following notice is copied from RFC 2026 [Bradner, 1996], Section 10.4, and describes the position of the IETF concerning intellectual property claims made against this document.

The IETF takes no position regarding the validity or scope of any intellectual property or other rights that might be claimed to pertain to the implementation or use other technology described in this document or the extent to which any license under such rights might or might not be available; neither does it represent that it has made any effort to identify any such rights. Information on the IETF's procedures with respect to rights in standards-track and standards-related documentation can be found in BCP-11. Copies of claims of rights made available for publication and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF Secretariat.

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights which may cover technology that may be required to practice this standard. Please address the information to the IETF Executive Director.

9. Copyright

The following copyright notice is copied from RFC 2026 [Bradner, 1996], Section 10.4, and describes the applicable copyright for this

document.

10. Authors' Addresses

Soohong Daniel Park
Mobile Platform Laboratory
SAMSUNG Electronics, KOREA
Email:soohong.park@samsung.com

Senthil Sivakumar
Cisco Systems, Inc.
170 West Tasman Dr.
San Jose CA 95134, US
Email: ssenthil@cisco.com

Pyda Srisuresh
Caymas Systems, Inc
1179-A North McDowell Blvd.
petaluma, CA 94954
USA
Email: srisuresh@yahoo.com

11. References

[Trans] Gilligan, R. and E. Nordmark, "Transition Mechanisms for IPv6 Hosts and Routers", RFC 1933, April 1996.

[NATPT] Tsirtsis G. and P. Srisuresh "Network Address Translation-Protocol Translation(NAT-PT)", RFC 2766, February 2000.

[DSTM] Bound, J, et al. "Dual Stack Transition Mechanism (DSTM)" draft-ietf-ngtrans-dstm-08.txt.

[TRT] J.Hagino, K.Yamamoto, "An IPv6-to-IPv4 Transport Relay Translator", RFC 3142, June 2001.

[DNSALG] Srisuresh, P., Tsirtsis, G., Akkiraju, P. and A. Heffernan, "DNS extensions to Network Address Translators(DNS_ALG)", RFC 2694, September 1999.

[NDP] Narten, T, Nordmark, E, Simpson, W "Neighbor Discovery for IP Version 6 (IPv6)", RFC 2461, December 1998.

[ISSUE] Durand, A., "Issues with NAT-PT DNS ALG in RFC2766" Internet-Draft, January 2003, work in progress.

[NIQ] Crawford, M, "IPv6 Node Information Queries", Internet-Draft, May 2002, work in progress.

[IPV6] Deering, S, Hinden, R, "Internet Protocol, Version 6 (IPv6) Specification", RFC 2460, December, 1998.

[LINK] Hinden, R, et. al. "An IPv6 Aggregatable Global Unicast Address Format", RFC 2374, July 1998.