Behavior Engineering for Hindrance Avoidance
Internet-Draft
Intended status: Informational
Expires: September 4, 2009

D. Miles, Ed.
Alcatel-Lucent
M. Townsley
Cisco Systems
March 4, 2009

Layer2-Aware NAT
draft-miles-behave-l2nat-00

Status of this Memo

Copyright Notice

Abstract

This document describes a "Layer2-Aware" IPv4-to-IPv4 (NAT44) Service Provider NAT function that identifies subscriber traffic based on IP-

independent methods such as a link-layer address, VLAN, PPP session, tunnel, etc. in order to allow one to either avoid "double-NAT" (NAT444) of subscriber IP traffic altogether, the need for additional "Shared Service-Provider" IPv4 address space, or partitioning of RFC 1918 space between subscribers. While the mechanisms described in this document may be applicable to a variety of network architectures, the primary focus is on residential "fixed-line" Internet access.

Table of Contents

1. Introduction

As public IPv4 space diminishes service providers become increasingly concerned about maintaining IPv4 services post exhaustion. The assumption behind this concern is that a significant part of Internet content will remain on IPv4 post exhaustion. IPv4 continuity approaches can be broadly classed as either IPv4-to-IPv6 translation or IPv4 tunneling both combined with a component of IPv4 public address sharing. L2-Aware NAT is a mechanism that permits devices which terminate the subscriber sessions to perform a NAPT function. Examples of a subscriber session may include a Softwire, L2TP session, PPP session, link-layer session, logical/physical interfaces, or future tunneling techniques. If an IPv4 over IPv6 encapsulation is used between the NAT and an individual subscriber, a L2-Aware NAT may be used one side of the tunnel to provide IPv4 access over an IPv6-only network as described in [I-D.durand-dual-stack-lite].

L2-Aware NAT differs from existing IPv4 NATs, or that proposed in[I-D.nishitani-cgn], because it is not reliant on the uniqueness of the NAT inside address to create NAT mappings or to forward downstream traffic towards the subscriber. A Traditional NAT [RFC3022] is deployed between two network segments are commonly referred to as either the inside and outside, or LAN and WAN segments. A L2-Aware NAT has many subscriber sessions (conceptually many inside/LAN segments) which are uniquely identified to allow for each subscriber to have their own NAT mapping table. In L2-Aware NAT, the NAT function supports hosts with the duplicated inside/LAN address provided they exist on different subscriber sessions. This technique can be leveraged post IPv4-exhaustion within the constraints of existing host and CPE implementations to assign the exact same public IPv4 address to every subscriber session and to then perform IPv4-to-IPv4 NAT on the subscriber traffic. For example, multiple PPP subscribers could be assigned the exact same IPv4 address through IPCP and the L2-Aware NAT will translate all packets to an external/WAN public IPv4 address by including subscriber-identifier as an additional delimiter in the NAT mapping table.

L2-Aware NAT assumes that there is some way to identify individual subscriber packets by some method other than the source IPv4 address assigned to the subscriber. While there are a number of ways this can be achieved, it may not always be possible without upgrading RG equipment, altering existing deployment topology, etc. When it can be achieved, subscriber aware NAT can be used to alleviate challenges associated with overlapping RFC 1918 space described in [I-D.shirasaki-isp-shared-addr].

1.1. 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 RFC 2119 [RFC2119].

2. Background

Pre-IPv4 Exhaustion, ISP have usually assigned one unique public IPv4 addresses to each unique subscriber. Subscribers themselves may choose to assign their public IPv4 address directly to hosts, or to traditional NAT devices within their network. In a post-exhaustion world an ISP will find it increasingly difficult to source additional globally unique IPv4 addresses for a growing subscriber population.

IPv6 is the obvious alternative to IPv4 transport in a post-exhaustion world; however, it is appreciated that IPv4 addresses will exhaust prior to significant penetration of IPv6 content, hence IPv4 services will continue to be used and new customers must be able to access any existing IPv4 content. In order to support this model with minimal change to existing IPv4 hosts or customer premise equipment a translation technique akin to NAPT may be employed. Since the operator cannot acquire any more public IPv4 addresses, one public IPv4 address needs to be shared by a number of subscribers.

A problem with IPv4 NAT when deployed within an ISP is that it translates between exactly two network segments. If a Carrier Grade NAT (CGN) is centrally located within the ISP then it is a requirement that hosts within the inside IP segment need a unique IPv4 address.

L2-Aware NAT relies on the identification of the subscriber and their unique logical or virtual network segment using layer 2 mechanisms. L2-Aware NAT allows multiple network segments to share a common public IPv4 address when combined with NAPT. Conceptually L2-Aware NAT is a NAT for multiple non-unique inside networks.

By adopting a L2-Aware NAT, an ISP can avoid change to existing IPv4 host or CPE while maintaining an IPv4-continuity service. Unlike a CGN, L2-Aware NAT allows an ISP to assign a duplicate IPv4 addresses to many subscribers

3. Terminology

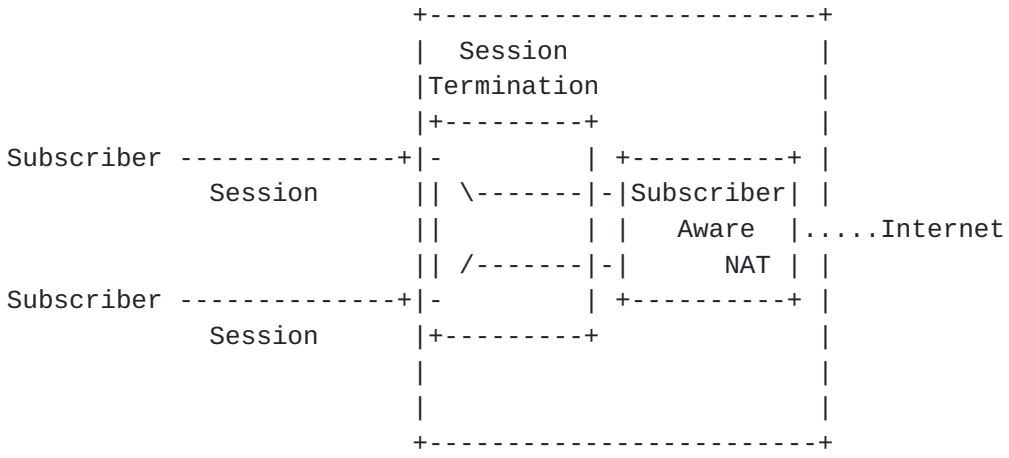
Address	An IP layer identifier for an interface or set of interfaces
BNG	Broadband Network Gateway
Host	A non-routing IPv4 node that sources and receives IP packets
IP	Internet Protocol Version 4 (IPv4)
Node	A device that implements IPv4
Router	A node that forwards packets not directly addressed to itself
Network Segment	A unique layer 3 forwarding topology
NAT	Network Address Translation
CGN	Carrier Grade NAT
Subscriber-ID	A node-specific unique identifier for exactly one subscriber. This may be generalised as a unique link-layer encapsulation identifier

4. Network Topology

The L2-Aware NAT MUST be deployed within the subscriber aware device, examples of which include BRAS, LNS, BNG, CMTS or Softwire Concentrators. Without the loss of generality, when BNG term is used in the following text it equally applies to any type of a device deploying Session-Aware NAT. By deploying L2-Aware NAT inside the same device that terminates the subscriber session, session information including Session-ID can be passed to the L2-Aware NAT function.

L2-Aware NATs terminate multiple network segments into a single NAT function comprised of virtual NAT tables for each subscriber.

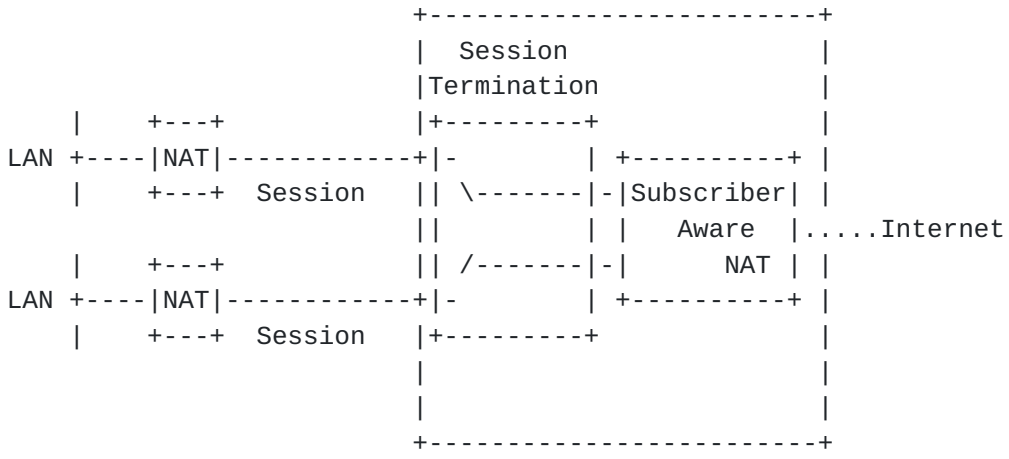
4.1. Link-Layer Subscriber Identification (NAT44)



Subscribers may be identified through the use of a unique per-subscriber link-layer address or the provision of a per-subscriber logical or physical interface. Examples may include PPPoE, L2TP, Softwires, [I-D.durand-dual-stack-lite], per-client VLAN, unique link-layer/MAC address or any other method where by some unique layer 2 information is presented to the L2-Aware NAT. When subscribers are directly attached all subscriber-to-subscriber communication MUST occur through the L2-Aware NAT.

It is expected that all clients will be assigned the same IPv4 address that is in turn translated. This IP address should be an IANA reserved IP address to allow future implementations to know when they are behind a L2-Aware NAT. This IANA-reserved address space could be common with that proposed in [I-D.durand-dual-stack-lite].

4.2. Home Network with NAT (NAT444)



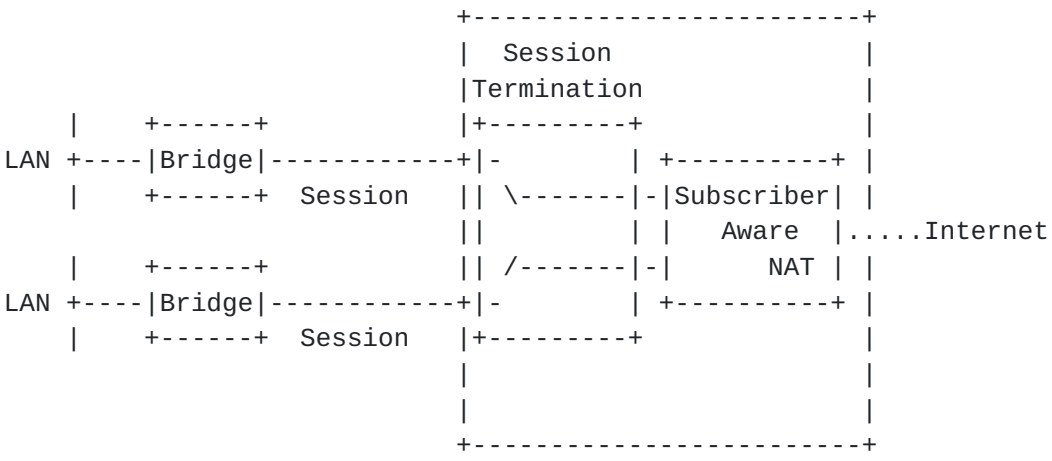
To support fixed broadband networks without change to existing NAT device, the home network NAT may exist between the subscriber LAN and the L2-Aware NAT. In this example the home network LAN would

typically be RFC 1918 address space, and is first NAT by the home network device. The address to which the LAN RFC 1918 address space is translated to should be an IANA reserved IP address and may be common with the IANA-reserved address space proposed in [I-D.durand-dual-stack-lite].

This approach has the advantage of requiring no CPE change though it is subject to a number of limitations associated with historical NAT implementations on consumer CPE while providing a workaround for the shared address space problem described in [I-D.shirasaki-isp-shared-addr].

Note that the presence of a NAT device in front of Session-Aware NAT creates a double-NAT environment and has all the implications thereof.

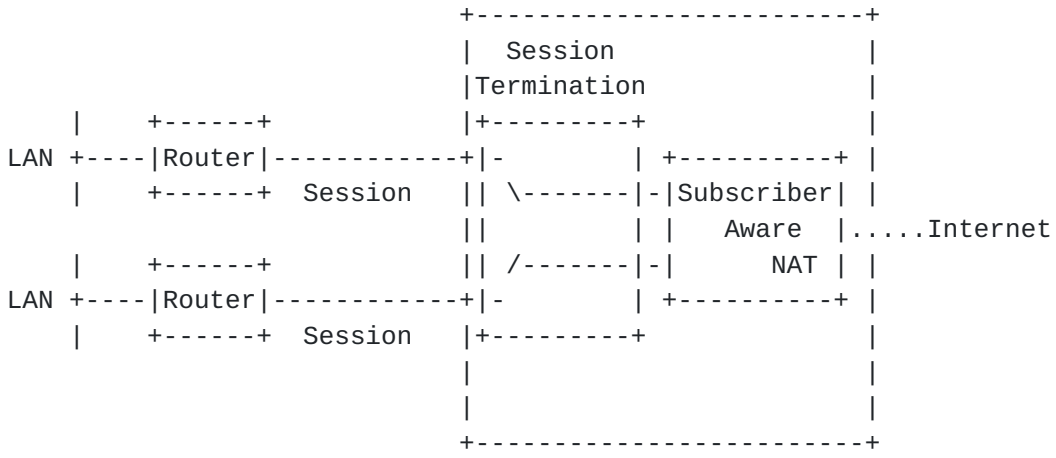
4.3. Bridged Home Network (NAT44)



L2-Aware NAT can support a bridged connection whereby a number of hosts share a single session. In this case each host should be assigned RFC 1918 addresses, and there is a requirement for hosts within a single session to have unique IP addresses. There is no requirement for hosts in different sessions to have unique IP addresses and as a result the same RFC 1918 network may be re-used for different sessions within the L2-Aware NAT.

While a bridge is commonly available, the L2-Aware NAT would be exposed to all link-layer addresses of hosts in the LAN segments and may need to field DHCP/Router Solicitation requests for said hosts. In it not uncommon for subscribers to install their own consumer NAT function in bridge topologies, degenerating this scenario to that described in section 4.2

4.4. Routed Home Network (NAT44)



In the Routed Home Network, subscribers are free to choose their own IP address configuration within the LAN segment. Because the address space in the home network is NATed, and hence not globally routable, there is no requirement for prefix-delegation or management of the address space used within the home network by the L2-Aware NAT. The routing home gateway needs only to install a default route towards the Subscriber Aware NAT - there is no need for routing protocols between routing home gateway and Subscriber Aware NAT.

This approach allows an operator to avoid NAT444 and provide a consistent and common NAT implementation across all subscribers.

A router exists at the boundary of the home network that performs standard IPv4 routing of the LAN private address space towards the L2-Aware NAT device. This device may have an IPv4 address assigned to its WAN interface as part of an IANA reserved network for L2-Aware NAT. The L2-Aware NAT must allow any source IP from the LAN segment and rely on Session-ID for any downstream forwarding.

It is intended that the router at the edge of the home network is of the same design as CPE of today but with the NAT function replaced with an IP Forwarding component. While this is all using well-known capabilities, "consumer-grade" CPEs may not perform this function, though certainly small "enterprise grade" routers likely will.

5. L2-Aware NAT Support

To support L2-Aware NAT, the BNG (or session terminator) MUST allow different sessions to have the same IPv4 address(es). As a result the management of sessions MUST NOT use IP address as a lookup key. Special consideration must be paid to forwarding behaviour between the L2-Aware NAT component and the session itself. Forwarding based

on IP destination will be insufficient between these two functions as IP addresses are no longer unique identifiers.

6. L2-Aware NAT Requirements

6.1. IP Addressing

L2-Aware NAT SHOULD use well-know IP addresses proposed in [I-D.durand-dual-stack-lite].

6.2. Inbound Connections

As the L2-Aware NAT supports the BEHAVE drafts Internet Connectivity Establishment [I-D.ietf-mmusic-ice] methods of TCP and UDP hole-punching are supported. Some IPv4 applications are reliant on accepting unsolicited inbound TCP or UDP sessions (commonly peer-to-peer or server applications) that may require an external port be opened on the NAT public address and mapped to an internal address and port within a single session.

It is important to extend an aspect of control to the users/hosts for inbound mapping behaviour. This control could be exercised through:

- An external management system that provisions permanent or temporary external address/port pair mappings to internal session, address and port tuples. One could envisage this to be via web-based portal or similar.

- Extending the existing UPnP IGD and NAT-PMP protocols to allow hosts direct configuration of the L2-Aware NAT device

- Encouraging IPv6 transport for those services that may require inbound connections

Inbound connections are typically used by peer-to-peer applications or by client-server services where the server is behind the L2-Aware NAT. NAT studies [NAT analysis] suggest the majority of inbound connections are to peer-to-peer applications and these applications are ideal candidates for migration to IPv6 transport, and by using IPv6 the complexities of NAT and client-to-NAT protocol implementation can be avoided.

6.3. UPnP and NAT-PMP

It should be noted that certain session topologies may make the L2-Aware NAT router behave appear as the default router on a subscriber's home network. In these cases it may be attractive to

allow [UPnP_IGD] or [I-D.cheshire-nat-pmp] protocols to operate between clients in the subscriber network and the L2-Aware NAT in the ISP. In routed environments, or scenarios where the customer is performing a NAT function between home network and service provider, a helper function may be needed to relay or proxy UPnP or NAT-PMP protocols to the L2-Aware NAT device.

Note that the [UPnP_IGD] protocol has a significant limitation as it does not have a facility to request a the assignment of a free port - it can only request explicit port numbers and be informed whether the mapping was successful or not successful. [I-D.cheshire-nat-pmp] is not constrained in this way, and through NAT-PMP the L2-Aware NAT could return an available free port. As L2-Aware NAT overloads many subscribers to a single IPv4 address it is inevitable that two UPnP or NAT-PMP clients may request an in-use port. In these cases application developers may require guidance to retry additional ports, and/or introduce a port-randomisation algorithm into their port request.

Some of these issues are not unique to L2-Aware NAT, but are a consequence of the NAT444 model.

7. Forwarding

7.1. Subscriber to Network

A L2-Aware NAT MUST support the IP forwarding techniques of any underlying session or link-layer without change. In links without link-layer addresses (such as PPP and L2TP) a client will send packets towards the L2-Aware NAT based on routing information or other reference.

Where link-layer addresses exist on the session, next-hop link-layer resolution will be performed by the client without any change as a result of the L2-Aware NAT. The L2-Aware NAT MUST present a next-hop IP interface towards the client that responds to any link-layer resolution protocols such as ARP.

From the client perspective the intention is to avoid a change in upstream forwarding behaviour and link-layer resolution. There may be other client impacts associated from the use of public, private or other types of address space, particularly on applications.

7.2. Network to Subscriber

When a packet reaches the L2-Aware NAT it is matched against an existing NAT mapping that returns inside IP, inside port and

session-ID. The L2-Aware NAT translates the packet headers as indicated and forwards the translated packet to the appropriate session. As IP addresses may overlap between sessions, the forwarding function MUST NOT forward based on destination IP.

8. Operational Considerations

8.1. In-band CPE Management

In current ISP networks there may be in-band management of CPE through protocols like SNMP or [TR-69]. Both these protocol examples rely on the management server being able to connect over UDP or TCP directly to the device without solicitation from the CPE. It is expected that most L2-Aware NAT deployments will require overlapping session IP addresses, making the identification of individual CPE by IP address impossible. In instances where IPv6 cannot be used for CPE management the L2-Aware NAT may need to implement SNMP or TR-69 helper or proxy functions whenever management of the CPE over IPv4 is required.

[TR-69] is a HTTP-based management method that operates in two modes. CPE connection initiation, where the CPE connects to the management server (ACS), and; ACS connection initiation, where the management server (ACS) connects directly to the CPE. The TR-69 specification does acknowledge that when an ACS is behind a NAT or firewall ACS connection initiation may not be possible.

Because of the way ACS connection initiations are specified, each CPE generates a unique URL for incoming requests and notifies the ACS of this URL an initialisation time. As the connection is made over HTTP it is feasible for a lightweight HTTP proxy function in the L2-Aware NAT to direct an ACS connection to the relevant CPE. The extensions required to do this are outside the scope of this document.

8.2. Logging and Accounting

As the L2-Aware NAT is deployed within the ISP environment there is an onus on the ISP to maintain accurate records that enable identification of the source of the traffic for purposes of law enforcement. A L2-Aware NAT, like any NAT that overloads external IP addresses, effectively obfuscates the original sender of traffic.

Several approaches are possible within a L2-Aware NAT:

1. Either the L2-Aware NAT MUST create an accurate log, containing Subscriber-ID or unique inside IP, Inside Port, Protocol, Outside IP, Outside Port, Destination IP and timestamp every time a NAT

mapping is created or destroyed, or;

2. the L2-Aware NAT MUST fix a specific port-range mapping and specific external IP address to each unique session for the duration of said session. This port-range must be logged along with information pertaining to the start, and stop of the session.

8.3. External Port Limits

In L2-Aware NAT the external IP addresses are shared among any number of sessions/subscribers. As external ports are a finite resource a mechanism must exist to limit their consumption per session/ subscriber and notify the subscriber when they are approaching this limit.

A L2-Aware NAT MUST implement per-session limits on the number of external ports that can be mapped to enable fair usage of available resources among subscribers. These limits SHOULD be configurable per session/subscriber. When an external port limit is reached for a session, a mapping MUST NOT be created and the L2-Aware NAT MUST follow [I-D.ietf-behave-nat-icmp] REQ-8: with respect to sending an ICMP destination unreachable message, with a code of 13 (Communication administratively prohibited).

In addition to a hard limit of external ports, a L2-Aware NAT SHOULD implement a threshold to allow notifications to the subscriber upon approaching the limit. If thresholds are implemented they SHOULD be configurable per session/subscriber.

8.3.1. HTTP Intercept and Redirect

When either per-session limits or thresholds are reached it is advisable to interactively notify the subscriber. One method to do so is to intercept any new outbound HTTP connections and present the user with a "Warning" page through an HTTP redirect.

In the case of HTTP intercept when an external port warning threshold limit has been reached, the L2-Aware NAT MAY provide a "Warning" page, that is removed upon subscribers acknowledgment. If implemented a mechanism to prevent display of the page on each outbound session initiation MUST be implemented. Details of such mechanism are out of scope for this document.

A L2-Aware NAT MUST NOT destroy any existing NAT mappings when a port limit is reached in an attempt to free ports.

8.3.2. Limit Override

A L2-Aware NAT MAY implement Limit Overrides to allow specific IP destinations, ports/protocols, or a combination of the two to be excluded from the per-session External Port Limit. An ISP may use this capability to allow access to the supporting HTTP server for HTTP intercepts, or to ISP provided services such as mail or account management.

9. IPv6 Co-existance

As L2-Aware NAT is an IPv4-only function, it can co-exist with IPv6 services over the same session subject to the capabilities of the BNG. It should be emphasised that L2-Aware NAT, and any IPv4 NAT, restricts the services available to the ISP subscriber population: the IETF, IANA and RIR community have clearly indicated that a migration to IPv6 must occur as a matter of urgency to alleviate these type of restrictions.

L2-Aware NAT can also function as the NAT-component of Dual-Stack Lite [I-D.durand-dual-stack-lite]. In this case, the tunnel happens to be a Softwire transported over IPv6. For the L2-Aware NAT perspective the underlying tunnel transport is not relevant, only the ability for the L2-Aware NAT device to address packets destined for a particular tunnel/softwire (based on Session ID).

10. Contributors

The authors would like to thank the following for their support: Mark Townsley, Steve Morin, Andrew Dolganow, Mickey Vucic and Philip Matthews.

Comments are solicited and should be addressed to the BEHAVE WG mailing list (behave@ietf.org) and/or the author.

11. IANA Considerations

This document proposes a new IANA managed IPv4 address from the current global address space. The size of this address space should be a /30 and must not be part of RFC 1918, Class D, Class E or otherwise reserved address space. This address space is used by any device behind a L2-Aware NAT where one IP is for hosts, and that IP+1 is reserved for the L2-Aware NAT.

L2-Aware NAT Reserved Address (Client): a.b.c.d

L2-Aware NAT Reserved Address (Default Router): a.b.c.d+1

12. Security Considerations

The operation of L2-Aware NAT is dependant on the unique identification of layer 2 or session parameters. When layer 2 information such as link-layer addresses are used in the creation of a L2-Aware NAT Session-ID mechanisms must exist to ensure link-layer address spoofing cannot occur between host and L2-Aware NAT.

It is also imperative that any NAT mappings are destroyed when a Session is dropped - this will avoid a situation whereby Session-ID might be re-used within a single L2-Aware NAT and earlier mappings may still be active for a new Session.

13. References

13.1. Normative References

[I-D.ietf-behave-nat-icmp]
Guha, S., Sivakumar, S., Ford, B., and P. Srisuresh, "NAT Behavioral Requirements for ICMP protocol", IETF-Draft ietf-behave-nat-icmp-09, September 2008.

[I-D.ietf-behave-tcp]
Guha, S., Biswas, K., Sivakumar, S., Ford, B., and P. Srisuresh, "NAT Behavioral Requirements for TCP", IETF-Draft draft-ietf-behave-tcp-08, September 2008.

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

[RFC3022] Srisuresh, P. and K. Egevang, "Traditional IP Network Address Translator (Traditional NAT)", IETF RFC 4787, January 2001.

[RFC4787] Audet, F. and C. Jennings, "Network Address Translation (NAT) Behavioral Requirements", IETF RFC 4787, January 2008.

13.2. Informative References

[I-D.cheshire-nat-pmp]
Cheshire, S., Krochmal, M., and K. Sekar, "NAT Port Mapping Protocol (NAT-PMP)", IETF Draft draft-cheshire-nat-pmp-03, April 2008.

[I-D.durand-dual-stack-lite]
Durand, A., Droms, R., Haberman, B., and J. Woodyatt, "Dual-stack lite broadband deployments post IPv4 exhaustion", IETF Draft draft-durand-dual-stack-lite-01, November 2008.

[I-D.ietf-mmusic-ice]
Rosenberg, J., "Interactive Connectivity Establishment (ICE): A Protocol for Network Address Translator (NAT) Traversal for Offer/Answer Protocols", IETF Draft draft-ietf-mmusic-ice-19, October 2007.

[I-D.nishitani-cgn]
Nishitani, T., "Carrier Grade Network Address Translator (NAT) Behavioral Requirements for Unicast UDP, TCP and ICMP", IETF Draft draft-nishitani-cgn-00, July 2008.

[I-D.shirasaki-isp-shared-addr]
Shirasaki, Y., "ISP Shared Addresses after IPv4 Exhaustion", IETF Draft draft-shirasaki-isp-shared-addr-00, June 2008.

[TR-69] The Broadband Forum, "CPE WAN Management Protocol", Technical Report TR-69, May 2004.

[UPnP_IGD]
Iyer, P. and U. Warrier, "Internet Gateway Device (IGD) Standardized Device Control Protocol V 1.0 - InternetGatewayDevice:1", UPnP Forum InternetGatewayDevice:1 Standardized DCP, November 2001.

Authors' Addresses

David Miles (editor)
Alcatel-Lucent
L3 / 215 Spring St
Melbourne, Victoria 3000
Australia

Phone: +61 3 9664 3308
Email: david.miles@alcatel-lucent.com

Mark Townsley
Cisco Systems
Paris,
France

Phone:
Email: townsley@cisco.com