

Network Working Group
Internet-Draft
Intended status: Informational
Expires: April 2, 2015

G. Chen
China Mobile
W. Li
China Telecom
T. Tsou
J. Huang
Huawei Technologies
T. Taylor
PT Taylor Consulting
September 29, 2014

Analysis of NAT64 Port Allocation Methods for Shared IPv4 Addresses
draft-chen-sunset4-cgn-port-allocation-05

Abstract

This document enumerates methods of port assignment in Carrier Grade NATs (CGNs), focused particularly on NAT64 environments. A theoretical framework of different NAT port allocation methods is described. The memo is intended to clarify and focus the port allocation discussion and propose an integrated view of the considerations for selection of the port allocation mechanism in a given deployment.

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 April 2, 2015.

Copyright Notice

Copyright (c) 2014 IETF Trust and the persons identified as the document authors. All rights reserved.

China Mobile observed that the port consumption per session on NAT64 is roughly only half that on NAT44. 43 percent of top100 websites have AAAA records, therefore the NAT64 didn't have to assign ports to the traffic going to those websites. The results may be different if more services (e.g. game, web-mail, etc) are considered. But it is apparent that the effects of port saving on NAT64 will be amplified by increasing native IPv6 support.

Apart from the above observation, port allocation can be tuned according to the phase of IPv6 migration. As more content providers and services become available over IPv6, the utilization of NAT64 goes down since fewer destinations require translation progressing. Thus as IPv6 migration proceeds, it will be possible to relax the multiplexing ratio of IPv4 address sharing.

2.2. Classification of Port Allocation Models

This section lists several models to allocate the port information in NAT64 equipment. It also describes example cases for each allocation model.

2.2.1. Stateful vs. Stateless

o Stateful

The stateful NAT can be implemented either by static address translation or dynamic address translation.

In the case of static address assignment, a one-to-one address mapping for hosts between a IPv6 network address and an IPv4 network address is pre-configured on the NAT operation. This case normally occurs when a server is deployed in a IPv6 domain. The static configuration ensures stable inbound connectivity.

Dynamic address assignment would periodically free the binding so that the global address could be recycled for later use. This increases the efficiency of usage of IPv4 addresses.

o Stateless

Stateless NAT is performed in compliance with [RFC6145]. The public IPv4 address is required to be embedded in the IPv6 address. Thus the NAT64 can directly extract the address and has no need to record mapping states.

A promising usage of stateless NAT may appear in the data centre environment where IPv6 server pools receive inbound connections from IPv4 users externally [I-D.anderson-v6ops-siit-dc]. NAT usage in

trend, this likely presents a challenge for network operators who need to optimize connectivity states and avoid port depletion. Those optimizations may even affect the method of port-range allocation, because a subscriber is only allowed to use a pre-configured port resource.

Two optimizations may be considered:

- o Reducing the TIME-WAIT state. The user's behavior normally correlates with system performance. It is rather common that users change video channels often. Investigations have shown that 60% of videos are watched for less than 20% of their duration. The user's access patterns may leave a number of the TIME-WAIT states. Therefore, acceleration of TIME-WAIT state transitions could increase the efficiency of port utilization. [RFC6191] defines a mechanism for reducing TIME-WAIT state by proposing TCP timestamps and sequence numbers.

[I-D.penno-behave-rfc4787-5382-5508-bis] recommended applying [RFC6191] and PAWS (Protect Against Wrapped Sequence numbers, described in [RFC1323]) to NAT. This may also be a way to improve port utilization.

- o Another possibility is to use Address-Dependent Mapping or Address and Port-Dependent Mapping [RFC4787] to increase port utilization. This feature has already been implemented on a vendor-specific basis. However, it should be noted that REQ-7 and REQ-12 in [RFC6888] may reduce the incentive to use anything but the Address-Independent Mapping behaviour recommended by [RFC4787].

2.4.3. Port Randomization

Port randomization is a feature to enhance the defense against hijacking of flows. [RFC6056] specifies that:

"A NAT that does not implement port preservation ([RFC4787], [RFC5382]) should obfuscate selection of the ephemeral port of a packet when it is changed during translation of that packet."

A NAT based on per-session allocation normally follows this recommendation.

See Section 4 for a fuller discussion of port randomization.

- [RFC3022] Srisuresh, P. and K. Egevang, "Traditional IP Network Address Translator (Traditional NAT)", RFC 3022, January 2001.
- [RFC4787] Audet, F. and C. Jennings, "Network Address Translation (NAT) Behavioral Requirements for Unicast UDP", BCP 127, RFC 4787, January 2007.
- [RFC5382] Guha, S., Biswas, K., Ford, B., Sivakumar, S., and P. Srisuresh, "NAT Behavioral Requirements for TCP", BCP 142, RFC 5382, October 2008.
- [RFC5424] Gerhards, R., "The Syslog Protocol", RFC 5424, March 2009.
- [RFC6191] Gont, F., "Reducing the TIME-WAIT State Using TCP Timestamps", BCP 159, RFC 6191, April 2011.
- [RFC6302] Durand, A., Gashinsky, I., Lee, D., and S. Sheppard, "Logging Recommendations for Internet-Facing Servers", BCP 162, RFC 6302, June 2011.
- [RFC6333] Durand, A., Droms, R., Woodyatt, J., and Y. Lee, "Dual-Stack Lite Broadband Deployments Following IPv4 Exhaustion", RFC 6333, August 2011.
- [RFC6346] Bush, R., "The Address plus Port (A+P) Approach to the IPv4 Address Shortage", RFC 6346, August 2011.
- [RFC6877] Mawatari, M., Kawashima, M., and C. Byrne, "464XLAT: Combination of Stateful and Stateless Translation", RFC 6877, April 2013.
- [RFC6887] Wing, D., Cheshire, S., Boucadair, M., Penno, R., and P. Selkirk, "Port Control Protocol (PCP)", RFC 6887, April 2013.
- [RFC7011] Claise, B., Trammell, B., and P. Aitken, "Specification of the IP Flow Information Export (IPFIX) Protocol for the Exchange of Flow Information", STD 77, RFC 7011, September 2013.

Authors' Addresses

Gang Chen
China Mobile
53A, Xibianmennei Ave.,
Xuanwu District,
Beijing 100053
P.R. China

Email: phdgang@gmail.com

Weibo Li
China Telecom
109, Zhongshan Ave. West, Tianhe District
Guangzhou 510630
P.R. China

Email: mweiboli@gmail.com

Tina Tsou
Huawei Technologies
Bantian, Longgang District
Shenzhen 518129
P.R. China

Email: tina.tsou.zouting@huawei.com

James Huang
Huawei Technologies
Bantian, Longgang District
Shenzhen 518129
P.R. China

Email: James.huang@huawei.com

Tom Taylor
PT Taylor Consulting
Ottawa, Ontario
Canada

Email: tom.taylor.stds@gmail.com

IETF work SHOULD function completely on IPv6-only nodes and networks, unless consensus exists that it is unnecessary to use a given feature or protocol on IPv6-only networks.

IETF SHOULD identify and update IPv4-only protocols and applications to support IPv6 unless consensus exists that it is unnecessary for a given feature or protocol.

4. Acknowledgements

Thanks to the following people for their comments: Jari Arkko, Ralph Droms, Scott Brim, Margaret Wasserman, Brian Haberman. Thanks also to Randy Bush, Mark Townsley, and Dan Wing for their discussion in IntArea WG at IETF 81 in Taipei, TW regarding transition technologies, IPv4 life extension, and IPv6 support.

5. IANA Considerations

This memo includes no request to IANA.

6. Security Considerations

This document generates no new security considerations because it is not defining a new protocol. As existing work is analyzed for its ability to operate properly on IPv6-only networks, new security issues may be identified.

7. References

7.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

7.2. Informative References

[RFC6540] George, W., Donley, C., Liljenstolpe, C., and L. Howard, "IPv6 Support Required for All IP-Capable Nodes", BCP 177, RFC 6540, April 2012.

Authors' Addresses

Wesley George
Time Warner Cable
13820 Sunrise Valley Drive
Herndon, VA 20171
US

Phone: +1 703-561-2540
Email: wesley.george@twcable.com

Lee Howard
Time Warner Cable
13820 Sunrise Valley Drive
Herndon, VA 20171
US

Phone: +1-703-345-3513
Email: lee.howard@twcable.com

4. Client Connection Establishment Behavior

PROBLEM 6: Happy Eyeballs [RFC6555] refers to multiple approaches to dual-stack client implementations that try to reduce connection setup delays by trying both IPv4 and IPv6 paths simultaneously. Some implementations introduce delays which provide an advantage to IPv6, while others do not [Huston2012]. The latter will pick the fastest path, no matter whether it is over IPv4 or IPv6, directing more traffic over IPv4 than the other kind of implementations. This can prove problematic in the context of IPv4 sunsetting, especially for Carrier-Grade NAT phasing out because CGN does not add significant latency that would make the IPv6 path more preferable. Traffic will therefore continue using the CGN path unless other network conditions change.

PROBLEM 7: `getaddrinfo()` [RFC3493] sends DNS queries for both A and AAAA records regardless of the state of IPv4 or IPv6 availability. The `AI_ADDRCONFIG` flag can be used to change this behavior, but it relies on programmers using the `getaddrinfo()` function to always pass this flag to the function. The current situation is that in an IPv6-only environment, many useless A queries are made.

5. Disabling IPv4 in Operating System and Applications

It is possible to completely remove IPv4 support from an operating system as has been shown by the work of Bjoern Zeeb on FreeBSD. [Zeeb] Removing IPv4 support in the kernel revealed many IPv4 dependencies in libraries and applications.

PROBLEM 8: Completely disabling IPv4 at runtime often reveals implementation bugs. Hard-coded dependencies on IPv4 abound, such as on the 127.0.0.1 address assigned to the loopback interface, and legacy IPv4-only APIs are widely used by applications. It is hard for the administrators and users to know what applications running on the operating system have implementation problems of IPv4 dependency. It is therefore often operationally impossible to completely disable IPv4 on individual nodes.

PROBLEM 9: In an IPv6-only world, legacy IPv4 code in operating systems and applications incurs a maintenance overhead and can present security risks.

sites [I-D.wing-behave-http-ip-address-literals]. IPv4 address literals may be published on media other than web sites, and may appear in various forms other than URLs. For the operating systems which exhibit the behavior described in [I-D.yourtchenko-ipv6-disable-ipv4-proxyarp], this also means an increase in the broadcast ARP traffic, which may be undesirable.

- [RFC3795] Sofia, R. and P. Nesser, "Survey of IPv4 Addresses in Currently Deployed IETF Application Area Standards Track and Experimental Documents", RFC 3795, June 2004.
- [RFC3796] Nesser, P. and A. Bergstrom, "Survey of IPv4 Addresses in Currently Deployed IETF Operations & Management Area Standards Track and Experimental Documents", RFC 3796, June 2004.
- [RFC3927] Cheshire, S., Aboba, B., and E. Guttman, "Dynamic Configuration of IPv4 Link-Local Addresses", RFC 3927, May 2005.
- [RFC4795] Aboba, B., Thaler, D., and L. Esibov, "Link-local Multicast Name Resolution (LLMNR)", RFC 4795, January 2007.
- [RFC6555] Wing, D. and A. Yourtchenko, "Happy Eyeballs: Success with Dual-Stack Hosts", RFC 6555, April 2012.
- [RFC6762] Cheshire, S. and M. Krochmal, "Multicast DNS", RFC 6762, February 2013.
- [Zeeb] "FreeBSD Snapshots without IPv4 support", <<http://wiki.freebsd.org/IPv6Only>>.

Annex A. Solution Ideas

A.1. Remotely Disabling IPv4

A.1.1. Indicating that IPv4 connectivity is unavailable

One way to address these issues is to send a signal to a dual-stack node that IPv4 connectivity is unavailable. Given that IPv4 shall be off, the message must be delivered through IPv6.

A.1.2. Disabling IPv4 in the LAN

One way to address these issues is to send a signal to a dual-stack node that auto-configuration of IPv4 addresses is undesirable, or that direct IPv4 communication between nodes on the same link should not take place.

A signalling protocol equivalent to the one from [RFC2563] but over IPv6 is necessary, using either Router Advertisements or DHCPv6.

If the ID is created from IPv6 address, e.g. by hashing from an IPv6 address, then naturally it has relationship with the address. If the ID is created regardless of IP address, one way to build association with IPv6 address is to embed the ID into an IPv6 address that is to be configured on the router, e.g. use a /96 IPv6 prefix and append it with a 32-bit long ID. One can also use some record keeping mechanisms, e.g. text file, DNS or other provisioning system like network management system to manage the IDs and mapping relations

with IPv6 addresses, though extra record keeping does introduce additional work.

Authors' Addresses

Will(Shucheng) Liu
Huawei Technologies
Bantian, Longgang District
Shenzhen 518129
China

Email: liushucheng@huawei.com

Weiping Xu
Huawei Technologies
Bantian, Longgang District
Shenzhen 518129
China

Email: xuweiping@huawei.com

Cathy Zhou
Huawei Technologies
Bantian, Longgang District
Shenzhen 518129
China

Email: cathy.zhou@huawei.com

Tina Tsou
Philips Lighting
United States of America

Email: tina.tsou@philips.com

Simon Perreault
Jive Communications
Quebec, QC
Canada

Email: sperreault@jive.com

Peng Fan
Beijing
China

Email: fanp08@gmail.com

Rong Gu
China Mobile
32 Xuanwumen West Ave, Xicheng District
Beijing 100053
China

Email: gurong_cmcc@outlook.com

Chongfeng Xie
China Telecom
China Telecom Beijing Information Science&Technology Innovation Park
Beiqijia Town Changping District, Beijing 102209,
China

Email: xiechf.bri@chinatelecom.cn

Ying Cheng
China Unicom
No.21 Financial Street, XiCheng District
Beijing 100033
China

Email: chengying10@chinaunicom.cn