Network Working Group Internet-Draft Intended status: Informational Expires: January 12, 2014

C. Donley C. Grundemann V. Sarawat K. Sundaresan CableLabs 0. Vautrin Juniper Networks July 11, 2013

# Deterministic Address Mapping to Reduce Logging in Carrier Grade NAT Deployments draft-donley-behave-deterministic-cgn-06

## Abstract

In some instances, Service Providers have a legal logging requirement to be able to map a subscriber's inside address with the address used on the public Internet (e.g. for abuse response). Unfortunately, many Carrier Grade NAT logging solutions require active logging of dynamic translations. Carrier Grade NAT port assignments are often per-connection, but could optionally use port ranges. Research indicates that per-connection logging is not scalable in many residential broadband services. This document suggests a way to manage Carrier Grade NAT translations in such a way as to significantly reduce the amount of logging required while providing traceability for abuse response. While the authors acknowledge that IPv6 is a preferred solution, Carrier Grade NAT is a reality in many networks, and is needed in situations where either customer equipment or Internet content only supports IPv4; this approach should in no way slow the deployment of IPv6.

## 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 <u>RFC 2119</u> [<u>RFC2119</u>].

## 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 <a href="http://datatracker.ietf.org/drafts/current/">http://datatracker.ietf.org/drafts/current/</a>.

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 January 12, 2014.

Copyright Notice

Copyright (c) 2013 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to <u>BCP 78</u> and the IETF Trust's Legal Provisions Relating to IETF Documents (<u>http://trustee.ietf.org/license-info</u>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

## Table of Contents

#### **<u>1</u>**. Introduction

It is becoming increasingly difficult to obtain new IPv4 address assignments from Regional/Local Internet Registries due to depleting supplies of unallocated IPv4 address space. To meet the growing demand for Internet connectivity from new subscribers, devices, and service types, some operators will be forced to share a single public IPv4 address among multiple subscribers using techniques such as

Carrier Grade Network Address Translation (CGN) [RFC6264] (e.g., NAT444 [I-D.shirasaki-nat444], DS-Lite [RFC6333], NAT64 [RFC6146] etc.). However, address sharing poses additional challenges to operators when considering how they manage service entitlement, public safety requests, or attack/abuse/fraud reports [RFC6269]. In order to identify a specific user associated with an IP address in response to such a request or for service entitlement, an operator will need to map a subscriber's internal source IP address and source port with the global public IP address and source port provided by the CGN for every connection initiated by the user.

CGN connection logging satisfies the need to identify attackers and respond to abuse/public safety requests, but it imposes significant operational challenges to operators. In lab testing, we have observed CGN log messages to be approximately 150 bytes long for NAT444 [I-D.shirasaki-nat444], and 175 bytes for DS-Lite [RFC6333] (individual log messages vary somewhat in size). Although we are not aware of definitive studies of connection rates per subscriber, reports from several operators in the US sets the average number of connections per household at approximately 33,000 connections per day. If each connection is individually logged, this translates to a data volume of approximately 5 MB per subscriber per day, or about 150 MB per subscriber per month; however, specific data volumes may vary across different operators based on myriad factors. Based on available data, a 1-million subscriber service provider will generate approximately 150 terabytes of log data per month, or 1.8 petabytes per year.

The volume of log data poses a problem for both operators and the public safety community. On the operator side, it requires a significant infrastructure investment by operators implementing CGN. It also requires updated operational practices to maintain the logging infrastructure, and requires approximately 23 Mbps of bandwidth between the CGN devices and the logging infrastructure per 50,000 users. On the public safety side, it increases the time required for an operator to search the logs in response to an abuse report, and could delay investigations. Accordingly, an international group of operators and public safety officials approached the authors to identify a way to reduce this impact while improving abuse response.

The volume of CGN logging can be reduced by assigning port ranges instead of individual ports. Using this method, only the assignment of a new port range is logged. This may massively reduce logging volume. The log reduction may vary depending on the length of the assigned port range, whether the port range is static or dynamic, etc. This has been acknowledged in [<u>RFC6269</u>] and [<u>I-D.sivakumar-behave-nat-logging</u>]. Per [<u>RFC6269</u>]:

"Address sharing solutions may mitigate these issues to some extent by pre-allocating groups of ports. Then only the allocation of the group needs to be recorded, and not the creation of every session binding within that group. There are trade-offs to be made between the sizes of these port groups, the ratio of public addresses to subscribers, whether or not these groups timeout, and the impact on logging requirements and port randomization security (<u>RFC6056</u>) [<u>RFC6056</u>]."

However, the existing solution still poses an impact on operators and public safety officials for logging and searching. Instead, CGNs could be designed and/or configured to deterministically map internal addresses to {external address + port range} in such a way as to be able to algorithmically calculate the mapping. Only inputs and configuration of the algorithm need to be logged. This approach reduces both logging volume and subscriber identification times. In some cases, when full deterministic allocation is used, this approach can eliminate the need for translation logging.

This document describes a method for such CGN address mapping, combined with block port reservations, that significantly reduces the burden on operators while offering the ability to map a subscriber's inside IP address with an outside address and external port number observed on the Internet.

The activation of the proposed port range allocation scheme is compliant with BEHAVE requirements such as the support of APP.

## 2. Deterministic Port Ranges

While a subscriber uses thousands of connections per day, most subscribers use far fewer resources at any given time. When the compression ratio (see Appendix B of RFC6269 [RFC6269]) is low (e.g., the ratio of the number of subscribers to the number of public IPv4 addresses allocated to a CGN is closer to 10:1 than 1000:1), each subscriber could expect to have access to thousands of TCP/UDP ports at any given time. Thus, as an alternative to logging each connection, CGNs could deterministically map customer private addresses (received on the customer-facing interface of the CGN, a.k.a., internal side) to public addresses extended with port ranges (used on the Internet-facing interface of the CGN, a.k.a., external side). This algorithm allows an operator to identify a subscriber internal IP address when provided the public side IP and port number without having to examine the CGN translation logs. This prevents an operator from having to transport and store massive amounts of session data from the CGN and then process it to identify a subscriber.

The algorithmic mapping can be expressed as:

(External IP Address, Port Range) = function 1 (Internal IP Address)

Internal IP Address = function 2 (External IP Address, Port Number)

The CGN SHOULD provide a method for administrators to test both mapping functions (e.g., enter an External IP Address + Port Number and receive the corresponding Internal IP Address).

Deterministic Port Range allocation requires configuration of the following variables:

- o Inside IPv4/IPv6 address range (I);
- o Outside IPv4 address range (0);
- Compression ratio (e.g. inside IP addresses I/outside IP addresses
  O) (C);
- o Dynamic address pool factor (D), to be added to the compression ratio in order to create an overflow address pool;
- o Maximum ports per user (M);
- o Address assignment algorithm (A) (see below); and
- o Reserved TCP/UDP port list (R)

Note: The inside address range (I) will be an IPv4 range in NAT444 operation (NAT444 [<u>I-D.shirasaki-nat444</u>]) and an IPv6 range in DS-Lite operation (DS-Lite [<u>RFC6333</u>]).

A subscriber is identified by an internal IPv4 address (e.g., NAT44) or an IPv6 prefix (e.g., DS-Lite or NAT64).

The algorithm may be generalized to L2-aware NAT [<u>I-D.miles-behave-l2nat</u>] but this requires the configuration of the Internal interface identifiers (e.g., MAC addresses).

The algorithm is not designed to retrieve an internal host among those sharing the same internal IP address (e.g., in a DS-Lite context, only an IPv6 address/prefix can be retrieved using the algorithm while the internal IPv4 address used for the encapsulated IPv4 datagram is lost).

Several address assignment algorithms are possible. Using predefined algorithms, such as those that follow, simplifies the process of

reversing the algorithm when needed. However, the CGN MAY support additional algorithms. Also, the CGN is not required to support all algorithms described below. Subscribers could be restricted to ports from a single IPv4 address, or could be allocated ports across all addresses in a pool, for example. The following algorithms and corresponding values of A are as follow:

0: Sequential (e.g. the first block goes to address 1, the second block to address 2, etc.)

1: Staggered (e.g. for every n between 0 and ((65536-R)/(C+D))-1 , address 1 receives ports n\*C+R, address 2 receives ports (1+n)\*C+R, etc.)

2: Round robin (e.g. the subscriber receives the same port number across a pool of external IP addresses. If the subscriber is to be assigned more ports than there are in the external IP pool, the subscriber receives the next highest port across the IP pool, and so on. Thus, if there are 10 IP addresses in a pool and a subscriber is assigned 1000 ports, the subscriber would receive a range such as ports 2000-2099 across all 10 external IP addresses).

3: Interlaced horizontally (e.g. each address receives every Cth port spread across a pool of external IP addresses).

4: Cryptographically random port assignment (<u>Section 2.2 of</u> <u>RFC6431</u> [<u>RFC6431</u>]). If this algorithm is used, the Service Provider needs to retain the keying material and specific cryptographic function to support reversibility.

5: Vendor-specific. Other vendor-specific algorithms may also be supported.

The assigned range of ports MAY also be used when translating ICMP requests (when re-writing the Identifier field).

The CGN then reserves ports as follows:

 The CGN removes reserved ports (R) from the port candidate list (e.g., 0-1023 for TCP and UDP). At a minimum, the CGN SHOULD remove system ports (<u>RFC6335</u>) [<u>RFC6335</u>] from the port candidate list reserved for deterministic assignment.

Note: Setting D to 0 disables the dynamic pool. This option eliminates the need for per-subscriber logging at the expense of limiting the number of concurrent connections that 'power users' can initiate.

- 3. When a subscriber initiates a connection, the CGN creates a translation mapping between the subscriber's inside local IP address/port and the CGN outside global IP address/port. The CGN MUST use one of the ports allocated in step 2 for the translation as long as such ports are available. The CGN SHOULD allocate ports randomly within the port range assigned by the deterministic algorithm. This is to increase subscriber privacy. The CGN MUST use the preallocated port range from step 2 for Port Control Protocol (PCP, [I-D.ietf-pcp-base]) reservations as long as such ports are available. While the CGN maintains its mapping table, it need not generate a log entry for translation mappings created in this step.
- 4. If D>0, the CGN will have a pool of ports left for dynamic assignment. If a subscriber uses more than the range of ports allocated in step 2 (but fewer than the configured maximum ports M), the CGN assigns a block of ports from the dynamic assignment range for such a connection or for PCP reservations. The CGN MUST log dynamically assigned port blocks to facilitate subscriber-to-address mapping. The CGN SHOULD manage dynamic ports as described in [I-D.tsou-behave-natx4-log-reduction].
- 5. Configuration of reserved ports (e.g., system ports) is left to operator configuration.

Thus, the CGN will maintain translation mapping information for all connections within its internal translation tables; however, it only needs to externally log translations for dynamically-assigned ports.

# 2.1. IPv4 Port Utilization Efficiency

For Service Providers requiring an aggressive address sharing ratio, the use of the algorithmic mapping may impact the efficiency of the address sharing. A dynamic port range allocation assignment is more suitable in those cases.

## 2.2. Planning & Dimensioning

Unlike dynamic approaches, the use of the algorithmic mapping requires more effort from operational teams to tweak the algorithm (e.g., size of the port range, address sharing ratio, etc.). Dedicated alarms SHOULD be configured when some port utilization thresholds are fired so that the configuration can be refined.

The use of algorithmic mapping also affects geolocation. Changes to the inside and outside address ranges (e.g. due to growth, address allocation planning, etc.) would require external geolocation providers to recalibrate their mappings.

### 2.3. Deterministic CGN Example

To illustrate the use of deterministic NAT, let's consider a simple example. The operator configures an inside address range (I) of 100.64.0.0/28 [RFC6598] and outside address (0) of 203.0.113.1. The dynamic address pool factor (D) is set to '2'. Thus, the total compression ratio is 1:(14+2) = 1:16. Only the system ports (e.g. ports < 1024) are reserved (R) . This configuration causes the CGN to preallocate ((65536-1024)/16 =) 4032 TCP and 4032 UDP ports per inside IPv4 address. For the purposes of this example, let's assume that they are allocated sequentially, where 100.64.0.1 maps to 203.0.113.1 ports 1024-5055, 100.64.0.2 maps to 203.0.113.1 ports 5056-9087, etc. The dynamic port range thus contains ports 57472-65535 (port allocation illustrated in the table below). Finally, the maximum ports/subscriber is set to 5040.

	Inside Address / Pool		Outside Address & Port	
 	Reserved		203.0.113.1:0-1023	
İ	100.64.0.1	İ	203.0.113.1:1024-5055	
İ	100.64.0.2	İ	203.0.113.1:5056-9087	
İ	100.64.0.3	İ	203.0.113.1:9088-13119	
I	100.64.0.4		203.0.113.1:13120-17151	
I	100.64.0.5		203.0.113.1:17152-21183	
I	100.64.0.6		203.0.113.1:21184-25215	
I	100.64.0.7		203.0.113.1:25216-29247	
I	100.64.0.8		203.0.113.1:29248-33279	
I	100.64.0.9	Ι	203.0.113.1:33280-37311	
I	100.64.0.10	Ι	203.0.113.1:37312-41343	
I	100.64.0.11		203.0.113.1:41344-45375	
I	100.64.0.12		203.0.113.1:45376-49407	
I	100.64.0.13	Ι	203.0.113.1:49408-53439	
I	100.64.0.14		203.0.113.1:53440-57471	
I	Dynamic		203.0.113.1:57472-65535	

+----+

+----+

When subscriber 1 using 100.64.0.1 initiates a low volume of connections (e.g. < 4032 concurrent connections), the CGN maps the outgoing source address/port to the preallocated range. These translation mappings are not logged.

Subscriber 2 concurrently uses more than the allocated 4032 ports (e.g. for peer-to-peer, mapping, video streaming, or other connection-intensive traffic types), the CGN allocates up to an additional 1008 ports using bulk port reservations. In this example, subscriber 2 uses outside ports 5056-9087, and then 100-port blocks between 58000-58999. Connections using ports 5056-9087 are not logged, while 10 log entries are created for ports 58000-58099, 58100-58199, 58200-58299, ..., 58900-58999.

In order to identify a subscriber behind a CGN (regardless of port allocation method), public safety agencies need to collect source address and port information from content provider log files. Thus, content providers are advised to log source address, source port, and timestamp for all log entries, per [RFC6302]. If a public safety agency collects such information from a content provider and reports abuse from 203.0.113.1, port 2001, the operator can reverse the mapping algorithm to determine that the internal IP address subscriber 1 has been assigned generated the traffic without consulting CGN logs (by correlating the internal IP address with DHCP /PPP lease connection records). If a second abuse report comes in for 203.0.113.1, port 58204, the operator will determine that port 58204 is within the dynamic pool range, consult the log file, correlate with connection records, and determine that subscriber 2 generated the traffic (assuming that the public safety timestamp matches the operator timestamp. As noted in RFC6292 [RFC6292], accurate time-keeping (e.g., use of NTP or Simple NTP) is vital).

In this example, there are no log entries for the majority of subscribers, who only use pre-allocated ports. Only minimal logging would be needed for those few subscribers who exceed their preallocated ports and obtain extra bulk port assignments from the dynamic pool. Logging data for those users will include inside address, outside address, outside port range, and timestamp.

### 3. Additional Logging Considerations

In order to be able to identify a subscriber based on observed external IPv4 address, port, and timestamp, an operator needs to know how the CGN was configured with regards to internal and external IP addresses, dynamic address pool factor, maximum ports per user, and

Internet-Draft

reserved port range at any given time. Therefore, the CGN MUST generate a record any time such variables are changed. The CGN SHOULD generate a log message any time such variables are changed. The CGN MAY keep such a record in the form of a router configuration file. If the CGN does not generate a log message, it would be up to the operator to maintain version control of router config changes. Also, the CGN SHOULD generate such a log message once per day to facilitate quick identification of the relevant configuration in the event of an abuse notification.

Such a log message MUST, at minimum, include the timestamp, inside prefix I, inside mask, outside prefix O, outside mask, D, M, A, and reserved port list R; for example:

[Wed Oct 11 14:32:52 2000]:100.64.0.0:28:203.0.113.0:32:2:5040:0:1-1023,5004,5060.

#### <u>3.1</u>. Failover Considerations

Due to the deterministic nature of algorithmically-assigned translations, no additional logging is required during failover conditions provided that inside address ranges are unique within a given failover domain. Even when directed to a different CGN server, translations within the deterministic port range on either the primary or secondary server can be algorithmically reversed, provided the algorithm is known. Thus, if 100.64.0.1 port 3456 maps to 203.0.113.1 port 1000 on CGN 1 and 198.51.100.1 port 1000 on Failover CGN 2, an operator can identify the subscriber based on outside source address and port information.

Similarly, assignments made from the dynamic overflow pool need to be logged as described above, whether translations are performed on the primary or failover CGN.

#### **<u>4</u>**. Impact on the IPv6 Transition

The solution described in this document is applicable to Carrier Grade NAT transition technologies (e.g. NAT444, DS-Lite, and NAT64). As discussed in [I-D.donley-nat444-impacts], the authors acknowledge that native IPv6 will offer subscribers a better experience than CGN. However, many CPE devices only support IPv4. Likewise, as of July 2012, only approximately 4% of the top 1 million websites were available using IPv6. Accordingly, deterministic CGN should in no way be understood as making CGN a replacement for IPv6 service. The authors encourage device manufacturers to consider [RFC6540] and include IPv6 support. In the interim, however, CGN has already been deployed in some operator networks. Deterministic CGN will provide operators with the ability to quickly respond to public safety

requests without requiring excessive infrastructure, operations, and bandwidth to support per-connection logging.

#### 5. IANA Considerations

This document makes no request of IANA.

## <u>6</u>. Security Considerations

The security considerations applicable to NAT operation for various protocols as documented in, for example, <u>RFC 4787</u> [<u>RFC4787</u>] and <u>RFC 5382</u> [<u>RFC5382</u>] also apply to this document.

Note that with the possible exception of cryptographically-based port allocations, attackers could reverse-engineer algorithmically-derived port allocations to either target a specific subscriber or to spoof traffic to make it appear to have been generated by a specific subscriber. However, this is exactly the same level of security that the subscriber would experience in the absence of CGN. CGN is not intended to provide additional security by obscurity.

#### 7. Acknowledgements

The authors would like to thank the following people for their suggestions and feedback: Bobby Flaim, Lee Howard, Wes George, Jean-Francois Tremblay, Mohammed Boucadair, Alain Durand, David Miles, Andy Anchev, Victor Kuarsingh, Miguel Cros Cecilia, and Reinaldo Penno.

# **8**. References

#### 8.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", <u>BCP 14</u>, <u>RFC 2119</u>, March 1997.
- [RFC4787] Audet, F. and C. Jennings, "Network Address Translation (NAT) Behavioral Requirements for Unicast UDP", <u>BCP 127</u>, <u>RFC 4787</u>, January 2007.
- [RFC5382] Guha, S., Biswas, K., Ford, B., Sivakumar, S., and P. Srisuresh, "NAT Behavioral Requirements for TCP", <u>BCP 142</u>, <u>RFC 5382</u>, October 2008.
- [RFC6264] Jiang, S., Guo, D., and B. Carpenter, "An Incremental Carrier-Grade NAT (CGN) for IPv6 Transition", <u>RFC 6264</u>, June 2011.

[RFC6269] Ford, M., Boucadair, M., Durand, A., Levis, P., and P. Roberts, "Issues with IP Address Sharing", <u>RFC 6269</u>, June 2011.

## 8.2. Informative References

[I-D.donley-nat444-impacts]

Donley, C., Howard, L., Kuarsingh, V., Berg, J., and U. Colorado, "Assessing the Impact of Carrier-Grade NAT on Network Applications", <u>draft-donley-nat444-impacts-04</u> (work in progress), May 2012.

[I-D.ietf-pcp-base]

Wing, D., Cheshire, S., Boucadair, M., Penno, R., and P. Selkirk, "Port Control Protocol (PCP)", <u>draft-ietf-pcp-</u> <u>base-29</u> (work in progress), November 2012.

[I-D.miles-behave-l2nat]

Miles, D. and M. Townsley, "Layer2-Aware NAT", <u>draft-</u> <u>miles-behave-l2nat-00</u> (work in progress), March 2009.

[I-D.shirasaki-nat444]

Yamagata, I., Shirasaki, Y., Nakagawa, A., Yamaguchi, J., and H. Ashida, "NAT444", <u>draft-shirasaki-nat444-06</u> (work in progress), July 2012.

#### [I-D.sivakumar-behave-nat-logging]

Sivakumar, S. and R. Penno, "IPFIX Information Elements for logging NAT Events", <u>draft-sivakumar-behave-nat-</u> <u>logging-05</u> (work in progress), July 2012.

[I-D.tsou-behave-natx4-log-reduction]

ZOU), T., Li, W., and T. Taylor, "Port Management To Reduce Logging In Large-Scale NATs", <u>draft-tsou-behave-</u> <u>natx4-log-reduction-02</u> (work in progress), September 2010.

- [RFC6056] Larsen, M. and F. Gont, "Recommendations for Transport-Protocol Port Randomization", <u>BCP 156</u>, <u>RFC 6056</u>, January 2011.
- [RFC6146] Bagnulo, M., Matthews, P., and I. van Beijnum, "Stateful NAT64: Network Address and Protocol Translation from IPv6 Clients to IPv4 Servers", <u>RFC 6146</u>, April 2011.
- [RFC6292] Hoffman, P., "Requirements for a Working Group Charter Tool", <u>RFC 6292</u>, June 2011.

- [RFC6302] Durand, A., Gashinsky, I., Lee, D., and S. Sheppard, "Logging Recommendations for Internet-Facing Servers", <u>BCP</u> <u>162</u>, <u>RFC 6302</u>, June 2011.
- [RFC6333] Durand, A., Droms, R., Woodyatt, J., and Y. Lee, "Dual-Stack Lite Broadband Deployments Following IPv4 Exhaustion", <u>RFC 6333</u>, August 2011.
- [RFC6335] Cotton, M., Eggert, L., Touch, J., Westerlund, M., and S. Cheshire, "Internet Assigned Numbers Authority (IANA) Procedures for the Management of the Service Name and Transport Protocol Port Number Registry", <u>BCP 165</u>, <u>RFC</u> 6335, August 2011.
- [RFC6431] Boucadair, M., Levis, P., Bajko, G., Savolainen, T., and T. Tsou, "Huawei Port Range Configuration Options for PPP IP Control Protocol (IPCP)", <u>RFC 6431</u>, November 2011.
- [RFC6540] George, W., Donley, C., Liljenstolpe, C., and L. Howard, "IPv6 Support Required for All IP-Capable Nodes", <u>BCP 177</u>, <u>RFC 6540</u>, April 2012.
- [RFC6598] Weil, J., Kuarsingh, V., Donley, C., Liljenstolpe, C., and M. Azinger, "IANA-Reserved IPv4 Prefix for Shared Address Space", <u>BCP 153</u>, <u>RFC 6598</u>, April 2012.

Authors' Addresses

Chris Donley CableLabs 858 Coal Creek Cir Louisville, CO 80027 US

Email: c.donley@cablelabs.com

Chris Grundemann CableLabs 858 Coal Creek Cir Louisville, CO 80027 US

Email: c.grundemann@cablelabs.com

Vikas Sarawat CableLabs 858 Coal Creek Cir Louisville, CO 80027 US

Email: v.sarawat@cablelabs.com

Karthik Sundaresan CableLabs 858 Coal Creek Cir Louisville, CO 80027 US

Email: k.sundaresan@cablelabs.com

Olivier Vautrin Juniper Networks 1194 N Mathilda Avenue Sunnyvale, CA 94089 US

Email: olivier@juniper.net