Independent Submission
Request for Comments: 7422
Category: Informational
ISSN: 2070-1721

C. Donley
CableLabs
C. Grundemann
Internet Society
V. Sarawat
K. Sundaresan
CableLabs
O. Vautrin
Juniper Networks

December 2014

Deterministic Address Mapping to Reduce Logging in Carrier-Grade NAT Deployments

Abstract

In some instances, Service Providers (SPs) 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 logging solutions for Carrier-Grade NATs (CGNs) require active logging of dynamic translations. CGN port assignments are often per connection, but they 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 CGN translations in such a way as to significantly reduce the amount of logging required while providing traceability for abuse response. IPv6 is, of course, the preferred solution. While deployment is in progress, SPs are forced by business imperatives to maintain support for IPv4. This note addresses the IPv4 part of the network when a CGN solution is in use.

Status of This Memo

Table of Contents

1. 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 NAT (CGN) [RFC6264] (e.g., NAT444 [NAT444], Dual-Stack Lite (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 [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 SP will generate approximately 150 terabytes of log data per month, or 1.8 petabytes per year. Note that many SPs compress log data after collection; compression factors of 2:1 or 3:1 are common.

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 it 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 [RFC6269], which recommends the logging of source ports at the server and/or destination logging at the CGN, and [NAT-LOGGING], which describes information to be logged at a NAT.

However, the existing solutions still pose 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 Application-specific functions (APP).

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. 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 RFC 6269 [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:

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

Note: The inside address range (I) will be an IPv4 range in NAT444 operation (NAT444 [NAT444]) and an IPv6 range in DS-Lite operation (DS-Lite [RFC6333]).

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 [L2NAT], but this requires the configuration of the Internal interface identifiers (e.g., Media Access Control (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 of the 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 follows:

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 (Section 2.2 of RFC6431 [RFC6431]). If this algorithm is used, the SP 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 rewriting the Identifier field).

The CGN then reserves ports as follows:

1. 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 [RFC6335] from the port candidate list reserved for deterministic assignment.

2. The CGN calculates the total compression ratio (C+D), and allocates 1/(C+D) of the available ports to each internal IP address. Specific port allocation is determined by the algorithm (A) configured on the CGN. Any remaining ports are allocated to the dynamic pool.

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 pre-allocated port range from step 2 for Port Control Protocol (PCP, [RFC6887]) 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 [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 SPs 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 and 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 198.51.100.0/28 [RFC6598] and outside address (O) of 192.0.2.1. 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 pre-allocate ((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 198.51.100.1 maps to 192.0.2.1 ports 1024-5055, 198.51.100.2 maps to 192.0.2.1 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	192.0.2.1:0-1023
198.51.100.1	192.0.2.1:1024-5055
198.51.100.2	192.0.2.1:5056-9087
198.51.100.3	192.0.2.1:9088-13119
198.51.100.4	192.0.2.1:13120-17151
198.51.100.5	192.0.2.1:17152-21183
198.51.100.6	192.0.2.1:21184-25215
198.51.100.7	192.0.2.1:25216-29247
198.51.100.8	192.0.2.1:29248-33279
198.51.100.9	192.0.2.1:33280-37311
198.51.100.10	192.0.2.1:37312-41343
198.51.100.11	192.0.2.1:41344-45375
198.51.100.12	192.0.2.1:45376-49407
198.51.100.13	192.0.2.1:49408-53439
198.51.100.14	192.0.2.1:53440-57471
Dynamic	192.0.2.1:57472-65535

When subscriber 1 using 198.51.100.1 initiates a low volume of connections (e.g., < 4032 concurrent connections), the CGN maps the outgoing source address/port to the pre-allocated 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 192.0.2.1, 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 192.0.2.1, 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 RFC 6292 [RFC6292], accurate timekeeping (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 pre-allocated 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.

Note that in a production environment, operators are encouraged to consider [RFC6598] for assigning inside addresses.

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 regard to internal and external IP addresses, dynamic address pool factor, maximum ports per user, and 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]:198.51.100.0:28:192.0.2.0:32:2:5040:0:1-1023,5004,5060.

3.1. 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 198.51.100.1 port 3456 maps to 192.0.2.1 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.

4. Impact on the IPv6 Transition

The solution described in this document is applicable to CGN transition technologies (e.g., NAT444, DS-Lite, and NAT64). As discussed in [RFC7021], the authors acknowledge that native IPv6 will offer subscribers a better experience than CGN. However, many Customer Premises Equipment (CPE) devices only support IPv4. Likewise, as of October 2014, only approximately 5.2% 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; however, until such time as IPv6 content and devices are widely available, 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. Privacy Considerations

The algorithm described above makes it easier for SPs and public safety officials to identify the IP address of a subscriber through a CGN system. This is the equivalent level of privacy users could expect when they are assigned a public IP address and their traffic is not translated. However, this algorithm could be used by other actors on the Internet to map multiple transactions to a single subscriber, particularly if ports are distributed sequentially. While still preserving traceability, subscriber privacy can be increased by using one of the other values of the Address Assignment Algorithm (A), which would require interested parties to know more about the Service Provider's CGN configuration to be able to tie multiple connections to a particular subscriber.

6. Security Considerations

The security considerations applicable to NAT operation for various protocols as documented in, for example, RFC 4787 [RFC4787] and RFC 5382 [RFC5382] 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. References

7.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997, <http://www.rfc-editor.org/info/rfc2119>.

[RFC4787] Audet, F. and C. Jennings, "Network Address Translation (NAT) Behavioral Requirements for Unicast UDP", BCP 127, RFC 4787, January 2007, <http://www.rfc-editor.org/info/rfc4787>.

[RFC5382] Guha, S., Biswas, K., Ford, B., Sivakumar, S., and P. Srisuresh, "NAT Behavioral Requirements for TCP", BCP 142, RFC 5382, October 2008, <http://www.rfc-editor.org/info/rfc5382>.

[RFC6264] Jiang, S., Guo, D., and B. Carpenter, "An Incremental Carrier-Grade NAT (CGN) for IPv6 Transition", RFC 6264, June 2011, <http://www.rfc-editor.org/info/rfc6264>.

[RFC6269] Ford, M., Boucadair, M., Durand, A., Levis, P., and P. Roberts, "Issues with IP Address Sharing", RFC 6269, June 2011, <http://www.rfc-editor.org/info/rfc6269>.

7.2. Informative References

[L2NAT] Miles, D. and M. Townsley, "Layer2-Aware NAT", Work in Progress, draft-miles-behave-l2nat-00, March 2009.

[LOG-REDUCTION] Tsou, T., Li, W., Taylor, T., and J. Huang, "Port Management To Reduce Logging In Large-Scale NATs", Work in Progress, draft-tsou-behave-natx4-log-reduction-04, July 2013.

[NAT-LOGGING] Sivakumar, S. and R. Penno, "IPFIX Information Elements for logging NAT Events", Work in Progress, draft-ietf-behave-ipfix-nat-logging-04, July 2014.

[NAT444] Yamagata, I., Shirasaki, Y., Nakagawa, A., Yamaguchi, J., and H. Ashida, "NAT444", Work in Progress, draft-shirasaki-nat444-06, July 2012.

[RFC6146] Bagnulo, M., Matthews, P., and I. van Beijnum, "Stateful NAT64: Network Address and Protocol Translation from IPv6 Clients to IPv4 Servers", RFC 6146, April 2011, <http://www.rfc-editor.org/info/rfc6146>.

[RFC6292] Hoffman, P., "Requirements for a Working Group Charter Tool", RFC 6292, June 2011, <http://www.rfc-editor.org/info/rfc6292>.

[RFC6302] Durand, A., Gashinsky, I., Lee, D., and S. Sheppard, "Logging Recommendations for Internet-Facing Servers", BCP 162, RFC 6302, June 2011, <http://www.rfc-editor.org/info/rfc6302>.

[RFC6333] Durand, A., Droms, R., Woodyatt, J., and Y. Lee, "Dual-Stack Lite Broadband Deployments Following IPv4 Exhaustion", RFC 6333, August 2011, <http://www.rfc-editor.org/info/rfc6333>.

[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", BCP 165, RFC 6335, August 2011, <http://www.rfc-editor.org/info/rfc6335>.

[RFC6431] Boucadair, M., Levis, P., Bajko, G., Savolainen, T., and T. Tsou, "Huawei Port Range Configuration Options for PPP IP Control Protocol (IPCP)", RFC 6431, November 2011, <http://www.rfc-editor.org/info/rfc6431>.

[RFC6598] Weil, J., Kuarsingh, V., Donley, C., Liljenstolpe, C., and M. Azinger, "IANA-Reserved IPv4 Prefix for Shared Address Space", BCP 153, RFC 6598, April 2012, <http://www.rfc-editor.org/info/rfc6598>.

[RFC6887] Wing, D., Cheshire, S., Boucadair, M., Penno, R., and P. Selkirk, "Port Control Protocol (PCP)", RFC 6887, April 2013, <http://www.rfc-editor.org/info/rfc6887>.

[RFC7021] Donley, C., Howard, L., Kuarsingh, V., Berg, J., and J. Doshi, "Assessing the Impact of Carrier-Grade NAT on Network Applications", RFC 7021, September 2013, <http://www.rfc-editor.org/info/rfc7021>.

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, Fred Baker, Brian Carpenter, and Reinaldo Penno.

Authors' Addresses

Chris Donley
CableLabs
858 Coal Creek Cir
Louisville, CO 80027
United States

EMail: c.donley@cablelabs.com

Chris Grundemann
Internet Society
Denver, CO
United States

EMail: cgrundemann@gmail.com

Vikas Sarawat
CableLabs
858 Coal Creek Cir
Louisville, CO 80027
United States

EMail: v.sarawat@cablelabs.com

Karthik Sundaresan
CableLabs
858 Coal Creek Cir
Louisville, CO 80027
United States

EMail: k.sundaresan@cablelabs.com

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

EMail: olivier@juniper.net