Facebook Network Appliance

CGNAT-BYPASS SUPPORT

May 2021

FACEBOOK

Copyright & Trademarks

Table of Contents

1. Introduction 4
   - High-level description 4
   - A simple CGNAT-bypass example 5
   - Extending the scheme to support IPv4 and IPv6 dual-stack traffic 5
2. Requirements 6
   - Private IP Space 6
   - Prefix Group Configuration 6
   - Network Connectivity / Routing Configuration 6
3. Examples 7
   - Examples of valid implementations 7
   - Multiple Networks 9
   - Examples of invalid implementations 9
4. How can you check if it's working? 10
5. How to make changes or remove configuration? 10
6. Glossary 11
7. Useful links 12
8. FAQ 13

1. Introduction

There are multiple possible ways networks might implement CGNAT. And there are different ways a CDN, such as Facebook's, can interact with CGNAT as well. This document will explain in detail the requirements and conditions under which you may successfully implement CGNAT bypass with FNA caches in your network.

To properly support CGNAT bypass when subscriber devices (phones, computers, etc) connect with FNAs, Facebook needs to know the private user IP prefixes as well as public CGNAT IP prefixes used by the same set of users. In other words, FNAs need to know how to translate private and public IP addresses in your network, in order to enable traffic targeting.

We support a BGP community-based scheme for ISPs to inform Facebook about such private and public prefix groups via route announcements.

High-level description

- **Prefix Group.** A *Prefix Group* contains private user IP prefixes and public CGNAT IP prefixes used by the same set of users in an ISP network. It is possible to have multiple Prefix Groups configured in a network, subject to requirements listed in the next section.

- **BGP Community Values.** Each Prefix Group from an ISP network is associated with a unique BGP community value within this network. The BGP community value picked by ISP from the range of `32934:10200 - 32934:10299`.

- **Route Announcements.** For each prefix within a Prefix Group, when the route is announced to FNAs, the BGP community value associated with the Prefix Group needs to be tagged. As a result, via BGP, the Prefix Group membership is communicated to Facebook.

A simple CGNAT-bypass example

The figure in the next page shows how a hypothetical CGNAT-bypass traffic flow may look like in an ISP network. In this example, clients are assigned with the `100.64.1.0/24` private IP space, and the CGNAT device is assigned with `1.2.3.0/24` public IP space.

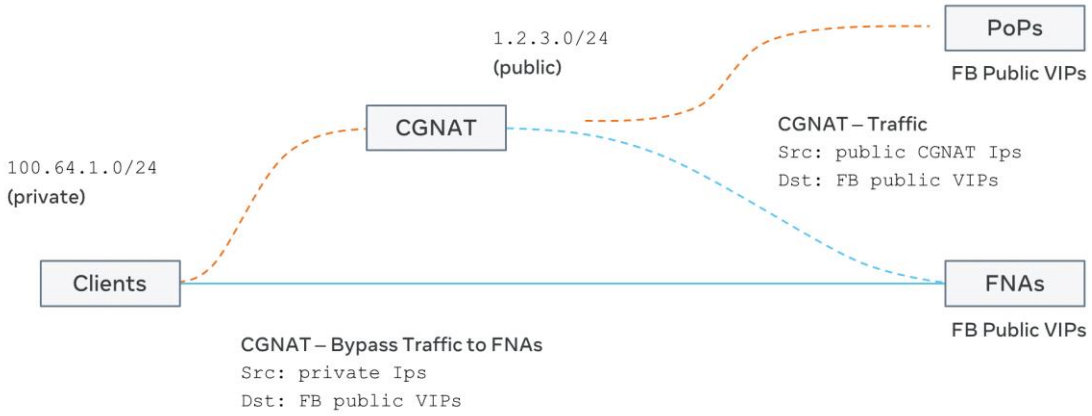

Diagram 1

- The dashed line represents the traffic that goes through CGNAT and the solid line denotes traffic that bypasses CGNAT.
- By applying proper routing changes in your network, it is possible to connect client devices directly to FNAs.
- The client devices still use CGNAT public IPs to connect to Facebook if direct routing is not configured, such as connecting with the Point of Presents (PoPs).

The Prefix Group, as well as the BGP routes observed on FNA, are

```
100.64.1.0/24:[32934:10200]   (private)
1.2.3.0/24:[32934:10200]      (cgnat public ipv4)
```

Extending the scheme to support IPv4 and IPv6 dual-stack traffic

We also extend the scheme for dual-stack traffic management, where a Prefix Group contains public IPv4 and IPv6 prefixes, and these prefixes correspond to the same set of users. The following shows an example Prefix Group configuration:

```
4.5.6.0/24:[32934:10201]        (IPv4 prefix)
2a03:2880:0::/48:[32934:10201] (IPv6 prefix for same users)
```

2. Requirements

Private IP Space

- Private user IP prefixes are required to be within RFC1918 and RFC6598 **exclusively**. If appropriate for your usage, we recommend using the dedicated '`100.64.0.0/10`' CGN space allocated in RFC6598.

Prefix Group Configuration

- Addresses
    - A Prefix Group contains at least one public IP prefix (either IPv4 or IPv6).
    - A Prefix Group contains at least one private user IP prefix.
    - Public IP address overlapping/reuse is not allowed among different Prefix Groups.
    - Private user IP address overlapping/reuse is allowed among different Prefix Groups as long as the overlapping private user IP prefixes are not announced to any common FNAs.
- BGP Community
    - A Prefix Group is required to be assigned with a unique BGP community value within a network, picked from `32934:10200 - 32934:10299`.
    - Route announcements for prefixes within the same Prefix Group should all have the same BGP community value tagged.
    - Non-related prefixes should not be tagged with any BGP community value in the specified range above.

Network Connectivity / Routing Configuration

All public prefixes within a Prefix Group are required to be announced to all FNAs in an FNA Group, where an FNA Group is defined as a set of FNAs serving the same set of users with the same failover policy.

3. Examples

Examples of valid implementations

Below we share multiple examples of valid implementations of CGNAT bypass, with one or multiple FNAs in a network, and with one or multiple prefix groups

- One Prefix Group: only one FNA, private and public prefixes.
- One Prefix Group: same public IP prefixes, same or different private IP prefixes on multiple FNAs.
- Multiple Prefix Groups: different public IP prefixes, same private IP prefixes.

One FNA		
Example	Valid Prefix Group Configurations	Notes
1	FNA1 (ISP1): 100.64.0.0/10:[32934:10200] (private) 1.2.3.0/24:[32934:10200] (cgnat public ipv4)	Minimal configuration
2	FNA1 (ISP1): 10.0.0.0/8:[32934:10200] (private) 100.64.0.0/10:[32934:10200] (private) 1.2.3.0/24:[32934:10200] (cgnat public ipv4) 4.5.6.0/24:[32934:10200] (cgnat public ipv4) 2a03:2880:0::/48:[32934:10200](cgnat public ipv6)	One Prefix Group • Multiple private prefixes • Multiple public prefixes, IPv4 & IPv6
3	FNA1 (ISP1): 10.0.0.0/8:[32934:10200] (private) 1.2.3.0/24:[32934:10200] (cgnat public ipv4) 100.64.0.0/10:[32934:10201] (private) 4.5.6.0/24:[32934:10201] (cgnat public ipv4)	Two Prefix Groups • Note the different BGP community values

Multiple FNAs		
Example	Valid Prefix Group Configurations	Notes
4	FNA1-1 (ISP1): 10.0.0.0/8:[32934:10200] (private) 100.64.0.0/10:[32934:10200] (private) 1.2.3.0/24:[32934:10200] (cgnat public ipv4) FNA1-2 (ISP1): 10.0.0.0/8:[32934:10200] (private) 100.64.0.0/10::[32934:10200] (private) 1.2.3.0/24:[32934:10200] (cgnat public ipv4)	One Prefix Group • Multiple private prefixes • One public prefixes
5	FNA1-1 (ISP1): **100.64.0.0/10**:[32934:10200] (private) 1.2.3.0/24:[32934:10200] (cgnat public ipv4) FNA1-2 (ISP1): **10.0.0.0/8**:[32934:10200] (private) 1.2.3.0/24:[32934:10200] (cgnat public ipv4)	One Prefix Group • Partial private prefixes at different FNAs
6	FNA1-1 (ISP1, dalstack): 10.0.0.0/8:[32934:10200] (private) 100.64.0.0/10:[32934:10200] (private) 1.2.3.0/24:[32934:10200] (cgnat public ipv4) 2a03:2880:0::/48:[32934:10200](cgnat public ipv6) FNA1-3 (ISP1, ipv4-only): 10.0.0.0/8:[32934:10200] (private) 100.64.0.0/10:[32934:10200] (private) 1.2.3.0/24:[32934:10200] (cgnat public ipv4)	One Prefix Group • Partial public prefixes on IPv4-only (or future IPv6-only) FNAs
7	FNA1 (ISP1): 10.0.0.0/8:[32934:**10200**] (private) 1.2.3.0/24:[32934:**10200**] (cgnat public ipv4) FNA2 (ISP1) 10.0.0.0/8:[32934:**10201**] (private) 4.5.6.7/24:[32934:**10201**] (cgnat public ipv4)	One ISP, two Prefix Groups • Private space reuse within the same ISP as long as they are not announced to any common FNA.

Multiple Networks

Organizations that manage multiple networks and have independent CGNAT implementations in each, might be using the same private IP prefix with a different address translation within each network.

This is supported, of course as long as the translation to public prefixes is different in each case:

8	FNA1 (ISP1): 100.64.0.0/10:[32934:10200] (private) 1.2.3.0/24:[32934:10200] (cgnat public ipv4) FNA2 (ISP2): 100.64.0.0/10:[32934:10201] (private) 4.5.6.0/24:[32934:10201] (cgnat public ipv4)	Private IP reuse by different ISPs; two Prefix Groups • Same private prefixes • Different ISP networks

Examples of invalid implementations

- Include example of invalid IP addresses: 192.168.0.0/24
- Include example of CGNAT translations not segmented (they mix regions, users, a mess)
- Include example that works with GGC but not with FNA
    - Example: Use public IPs as private IPs

4. How can you check if it's working?

You can check in the Facebook Partner Portal (http://fb.me/npp), for users that have access.

In the Caching section, the FNA Status table shows the CGNAT Bypass status per cache. If there is no such information, or the column shows a red checkmark, it means that it is not enabled.

From the Caching section and FNA Status table, you may click on the name of a specific cache to look into more details. Within that page for a specific cluster, check the Prefix Announcements table. Here we will show the prefixes received in the BGP sessions as well as the communities associated with them, and you can verify how we are seeing the prefix grouping from the point of view of that specific FNA cache.

5. How to make changes or remove configuration?

Simply update or remove the BGP communities from your prefix advertisement, stop advertising private prefixes to the FNA and enable traffic through your CGNAT. Since Facebook doesn't need to update configuration on the FNA, any changes you make will take place immediately.

6. Glossary

CGNAT: (wikipedia:) **Carrier-grade NAT** (**CGN** or **CGNAT**), also known as **large-scale NAT** (**LSN**), is an approach to IPv4 network design in which end sites, in particular residential networks, are configured with private network addresses that are translated to public IPv4 addresses by middlebox network address translator devices embedded in the network operator's network, permitting the sharing of small pools of public addresses among many end sites. This shifts the NAT function and configuration thereof from the customer premises to the Internet service provider network.

ISP: Internet Service Provider, is an organization that has an IP network and provides services to people or companies (the subscribers) for accessing, using, or participating in the Internet.

POP: also Edge PoP. Edge PoPs are points of presence (PoP) at the edge of Facebook's production network where we interconnect with other networks/carriers/telcos, in order to route traffic and cache content. Edge PoPs extend our network's reach, enable us to route traffic efficiently, and help us decrease latency to our end users.

7. Useful links

- IETF RFC 1918, Address Allocation for Private Internets https://tools.ietf.org/html/rfc1918
- IETF RFC 6598, Reserved IPv4 Prefix for Shared Address Space https://tools.ietf.org/html/rfc6598
- IETF RFC 7021, Assessing the Impact of Carrier-Grade NAT on Network Applications https://tools.ietf.org/html/rfc7021

8. FAQ

Q1. Is Facebook's implementation of CGNAT bypass similar to those of other CDN(s)?

Here's a comparison:

Item	Facebook	Facebook
Private IP Space	RFC1918 or RFC6598.	RFC1918, RFC6598, or usurped globally unique address space.
BGP community values	A range of community values.	Two community values, one for private, one for public.

Prefix Announcements	Private and public prefixes are grouped to Prefix Groups, and each Prefix Group has a unique BGP community value.
	
	Same community value for both private and public prefixes within the same Prefix Group. Example:
	`private:[32934:10200]`
	`public:[32934:10200]`
	
	No private IP address reuse among Prefix Groups if announced to any common FNAs.
		ISPs specify which prefixes are private and which ones are public.
		
		Different community values for private and public prefixes.
		
		Example:
		`private:[asn:bgpvalue1]`
		`Public:[asn:bgpvalue2]`
		
		No private IP address reuse among multiple CGNAT devices if they are announced to the same set of CDN nodes.

Q2. How do I know if FNAs will work with my CGNAT?

Glad you asked! Check section 2 of this document, that includes a detailed list of requirements. Keep in mind that depending on how you have implemented CGNAT in your network, it might be or not be compatible with our solution for CGNAT bypass

Q3. Is it possible to announce the 100.64.0.0/10 and if this is the case how the FNAs will not include the IPs that are not part of the test like 100.101.240.0.0/24 for example?

We suggest not to advertise such a big prefix, and instead advertise the private specific prefixes for the location that the FNA is serving. This will make load balancing easier to achieve.

Q4. Which mechanisms do we use to avoid sending all traffic to the FNAs where private prefixes are advertised?

We target the public cgnat prefixes only and we use our normal targeting system to allocate traffic.

Q5. How will traffic be balanced if ISP enables cgnat bypass in all its network?

Traffic will be balanced using the current mechanism to allocate traffic by our targeting system.

Q6. Now that we have applied CGNAT bypass on FNA, could we also apply bypass on the PNI sessions?

No, you can't, please don't do it.

Q7. May I advertise other public prefixes (IPv4 or IPv6) from other locations and the public cgnat and private prefixes on same FNA?

You can advertise public prefixes of other cities (IPv4/IPv6) in addition to the CGNAT public prefixes and private prefixes. There is no restriction to have full CGNAT or full public advertisements on a FNA.

FACEBOOK