Carrier Grade NAT

Requirements and Challenges in the Real World

Amir Tabdili – Cypress Consulting
amir@cypressconsult.net

Agenda

1. NAT, CG-NAT: Functionality Highlights
2. CPE NAT vs. CG-NAT
3. CGN Requirements for Application Transparency
4. Application Layer Gateways and CG-NAT
5. Flows, Sessions and CG-NAT Deployment
6. Logging Challenges and Port Block Allocation (PBA)
7. Port Forwarding Challenges and PCP
8. Lessons Learned from CG-NAT Deployments

Functionality Highlights

- Network Address Translation (NAT):
  - Involves changing the address and possibly port information of an IP packet
- NAT is commonly used for:
  - Aggregating a set of host addresses on a private network behind a single or a pool of public addresses
  - As a mechanism to conceal the internal network addressing
  - For interconnecting networks with overlapping addressing
  - To fight the depletion of IPv4 address space
    - One of the most common solutions adopted is NAT444 using a CG-NAT

Functionality Highlights (contd.)

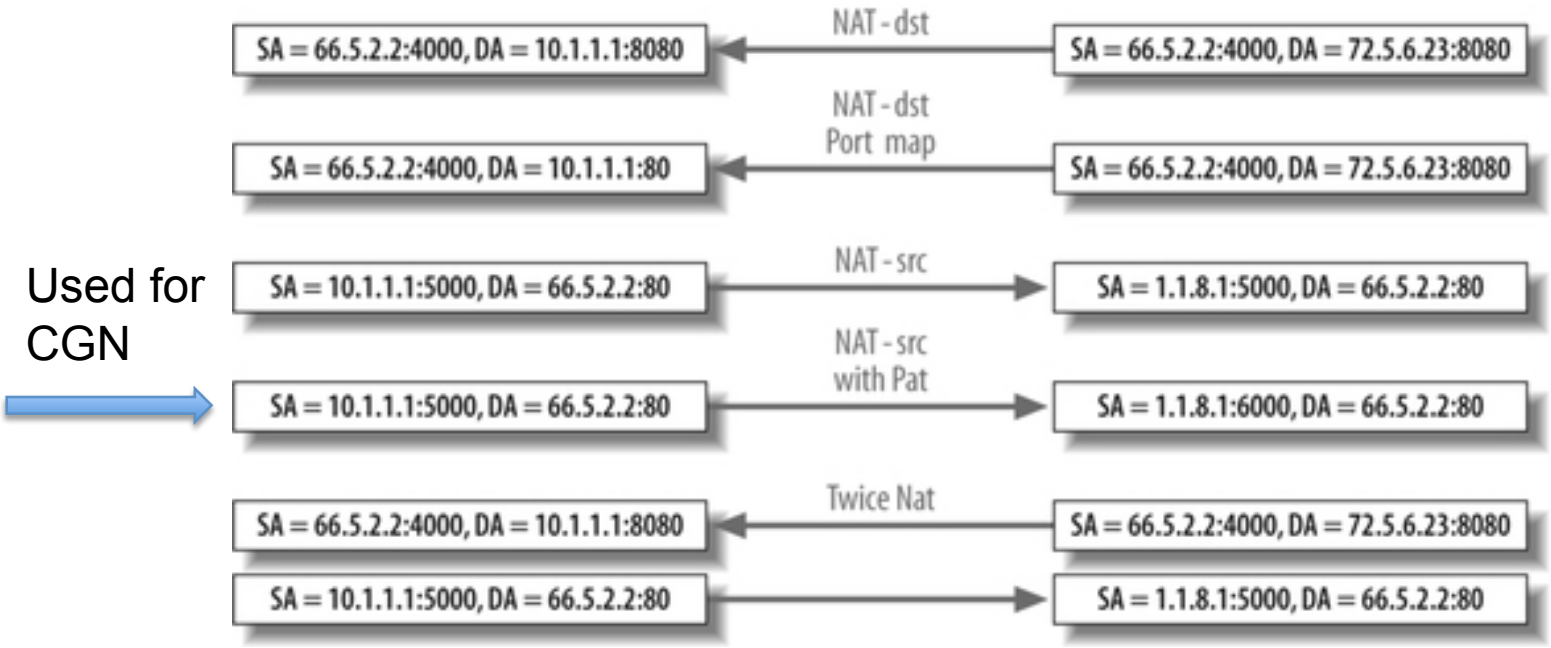
- CG-NAT = Large Scale NAT or LSN
  - A highly scalable NAT placed between the customer premises equipment (CPE) and the core of the network

- Several techniques are available to deal with IPv4 depletion problem. A topic of another presentation

CG-NAT is a cornerstone of at least two of them (NAT444 and DS-Lite).

- This presentation focuses on NAT as it applies to NAT444 applications

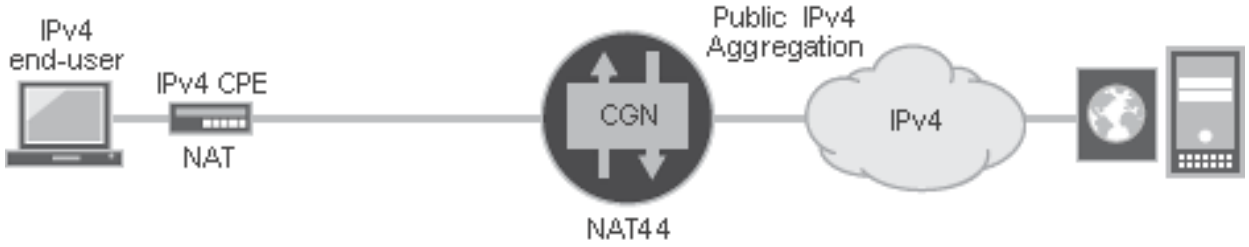
Functionality Highlights (contd.)



Functionality Highlights (contd.)

NAT444 uses three layers of IPv4 addressing:

A private IPv4 block within the user network (behind the CPE NAT)
A different private IPv4 block for the user to provider links (between the CPE NAT and the CG-NAT)
A public IPv4 address on the outside of the CG-NAT



NAT444 has several significant advantages:

There are no changes to end user hosts or the CPEs
NAT44 application-layer gateways (ALGs) can be leveraged.
NAT code is mature

Our Past: CPE NAT (only)

•NAT is prevalent in CPEs (NAT44)
•1:1 mapping of a public IP address to customer local-loop.
- ISP is able to identify without a doubt a customer in a point in time by looking at public IP address logs or RADIUS information.
- Logging and tracking is simple

- CPE NAT scaling needs ranges from a few to thousands of simultaneous devices.
- Easy administration since NAT under customer control.
- Servers on customer (private) network can be accessed from the Internet through port forwarding

Our Future: CG-NAT (also)

Now what happens if all NAT that is performed at customer premises moves (or is also done) at the edge or at the core of the SP?

Subscribers expect same behavior and control as before, but…

- There is no more exclusive 1:1 mapping of customer to public IPv4 address
- Scaling goes from few hundreds of thousands to 10s millions.
- Port management and sessions per second needs to scale accordingly
- Logging becomes increasingly important
  - Needs to deal with increased scalability
  - Needs to carry more information to identify a customer.
- Some customers still need to run servers on their premises easily
- Since subscribers share public IP addresses, session management is needed to enforce fairness.

CGN objectives

- Primary challenge of CG-NAT
  - To *address and balance application transparency vs. public addresses sharing and efficiency*

- That differs from enterprise/CPE NAT which emphasizes on address sharing efficiency
  - Provides much more than just NAT capabilities like firewall functionalities

- "Carrier-grade" has nothing to do with the quality of the NAT device,
  - It is rather a topological/location qualifier
  - The NAT device is placed inside an SP's network and translates the traffic of potentially many subscribers.
  - It however is deployed with NAT options that work in favour of application transparency

CGN Requirements for Application Transparency: APP

APP, Address Pooling ,“Paired” behavior

- A CG-NAT device must use the same external IP address mapping for all sessions associated with the same internal IP address

- It solves the problem of an application opening multiple connections using different source ports.

- Remote servers or peers for the same application reject connections if not all originated by the same IP address.
  - Examples are Instant Messengers

- CGN should adhere to APP even if it means exhausting ports available on a single address
  - Same source can not use ports from two different IP addresses.

CGN Requirements for Application Transparency: EIM

EIM, Endpoint Independent Mapping

• A CG-NAT device must assign the same external address and port for all connections originated from a given internal host if they all use the same internal port

•As a consequence connections originated by same internal IP address, but with different internal port can use a different external IP address

•Enabling EIM allows to have a stable external P address and Port (for a period of time) that external hosts can use to connect.

- Very important for p2p, gaming and the mobile world

•EIM does not decide who from the external realm can connect to the internal host, that is done by EIF instead.

CGN Requirements for Application Transparency: EIF

EIF, Endpoint Independent Filtering

- EIM alone does not influence the inbound filtering behavior.
    - Usually the default filtering behavior is Address and Port dependant (APM)
    - Means that only remote Servers or Peers towards which we opened a connection are allowed to reach the internal host
- EIF filters out only packets not destined to the internal address and port, regardless of the IP address and port of the remote Server or Peer.

APP+EIM+EIF Goal:
To provide as much transparency as possible to the applications.

CGN Requirements for Application Transparency: Other

Preserve Range

RFC4787 defines two port ranges: "Well Known Ports" [0, 1023] and "Registered“/"Dynamic and/or Private" [1024, 65535]
When the source port of the internal host establishing a new connection falls into one of these ranges the CGN tries to allocate an external source port in the same range. If it fails to find a port, connection fails too.

Preserve Parity

CGN tries to allocate a even/odd external source port depending on whether the new connection has an internal even/odd source port

Managing subscriber's sessions

Limiting the maximum number of sessions from same subscriber

- This allows to provide fairness of public IP and port availability to different subscribers
- Prevents DDOS attacks

Port Random-Allocation

- For each IP address in a pool the initial allocated port is assigned randomly and then continuing to allocate ports sequentially from there.
- It lowers the risk of inbound attacks when EIF is enabled

Round Robin address allocation

- For every different internal source address, a different NAT address is allocated in a round robin fashion
- Instead of using all available ports for a specific public IP address, move to the next public IP address whenever there is a new internal host requiring a connection.
- Address Pooling behavior should be unchanged.
    - New sessions from the same internal source address will continue to use the same public IPv4 address

Application Level gateways

Most SPs prefer not to deal with ALGs
But some protocols require them to be enabled

- Iphone uses pptp to connect to a VPN
- ICMP, Traceroute, TFTP, RSH, MS-RPC, PPTP, FTP, H.323

CG-NAT device should allow the administrator to selectively enable ALGs on a per protocol basis

Beware: There a number of NAT-friendly apps that ALGs can interfere with

Flows, Sessions and CGN Deployment

Flow is a unidirectional 5-tuple (source IP, dest IP, protocol, source Port, Dest Port).

A session is two unidirectional flows (C2S, S2C) together

Flow Analysis is a major part of CGN Design

The average number of flows per subscriber, duration, churn, and break down per protocol and application are key data that determine:

- Network deployment strategy
- New hardware design and capacity
- Required storage capacity for logging
- Network planning: oversubscribed or undersubscribed

Cost here is a major concern

PBA: Port Block Allocation Overview

Ports will be allocated to users in blocks

- The block size is fixed per NAT pool and is configurable
- Each user could use multiple blocks
- The administrator can configure a ‘max’ number of blocks per user

The most recently allocated block is the current ‘active’ block.

- New requests for NAT ports will come from this block
- Port will be allocated randomly from the current ‘active’ block
- A block remains active only for a fixed time interval.
  - In practice this is set to infinity: Only timeout the active block when there are no ports assigned from it anymore
- Any non-active block without any ports in use should get freed to NAT pool immediately

A syslog will be generated for each block allocation and release

PBA: Port Block Allocation
The Rational

Service Providers need to be able to trace a NAT session to a subscriber

Generally, two syslog messages (start, stop) are generated for every NAT session.

In large scale deployments the number of sessions/sec is very high

- A large volume of log messages will need to be processed and archived.
- PBA decreases log generation, processing and storage requirements

Also, traffic from the subscribers comes in bursts, meaning that dynamically pre-allocating a block of ports improves performance because it matches traffic patterns.

- Port Block allocation allows for higher ramp rate performances

- Tradeoff is the public IPv4 address efficiency
  - User is assigned a block of fixed size regardless of actual port usage

PORT FORWARDING IN CG-NAT SCENARIO

Today, users can freely configure port forwarding. either:

- Direct connection to the CPE (unmanaged CPE)
- Access to a central configuration server

When moving to a CG-NAT scenario, the requirement is not to break this model.

Using a Central NAT Configuration Server looks like the preferred options to some SPs for now

Future: PCP (Port Control Protocol)

- PCP objectives are to enable applications to receive incoming connections in the presence of an ISP NAT/Firewall.
- Instead of ‘working around’ NATs like other NAT traversal techniques like STUN/TURN/ICE, PCP enables an explicit dialog between applications and the NAT.
- PCP can be seen as a ‘carrier-grade’ evolution of UPnP-IGD and NAT-PMP.

Lessons Learned from CGN Deployments

SPs are deploying CG-NAT regardless of their IPv6 plans. Like it or not! ☺

Redundancy of CG-NAT devices is the most complex part of design

- Inter and intra Chassis fail over
- Most Applications do not behave well with TCP resets
- No Stateful Failover available in Author's experience

L3VPNs is a very common technology to isolate the private domain

Scaling is a concern

- But we do not see those that many subscribers that use more than 100 sessions at peak time.
- Most providers put a cap somewhere between 1200-2000 sessions/ subscriber

Lessons Learned from CGN Deployments

Security is a concern

- Stateful Firewall, Integrated IDP, etc

Port forwarding is done mostly statically

- External Web Page
- PCP in the works

What is widely use and what is not:

- PBA is used to limit the size of log storage. Provides built-in per subscriber port limit
- Other requirements are different from region to region
  - APP is always used
  - EIM/EIF use is mixed
  - Preserve range and Preserve Parity is rarely used
  - ALGs are only used if their presence is required, like FTP
    - Try without ALG first
  - Goal is to not do harm to applications that can traverse NAT

Lessons Learned from CGN Deployments

Conclusions on Flow Analysis

- Number of flows per subscriber or ramp up rate is not an issue.
- Median number of flows is quite low across the board.
- In wireline networks the average number is skewed due to P2P users.
- In a few trial, rarely detected AJAX web sites that open ‘100s’ of sessions concurrently.
- Bandwidth is the more important consideration

Most Operations still not familiar with CG-NAT complexities

There is a lot of trial and error on the part of service providers