Understand NAT on Nexus 9300

Contents

Introduction

Prerequisites

- Requirements
- Components Used

IntroduceNATSupport on N9K

- Terminology
- NAT TCAM Resource
  - NAT region
  - TCP Aware region
  - NAT Rewrite Table

Configuration and Verification

- Topology
- N9K-NAT Configuration
- Verification

Frequently Asked Questions

- What Happens when NAT TCAM exhausted?
- What Happens when Max-entries is Reached?
- Why are Some NAT Packets Punted to the CPU?
- Why NAT Works without proxy-arp on Nexus 9000?
- How add-route Argument Works on N9K and Why It is Mandatory?
- Why does NAT Support a Maximum of 100 ICMP entries

Related Information

Introduction

This document describes NAT feature on Nexus 9000 switches equipped with a Cisco Cloud-Scale ASIC that runs NX-OS software.

Prerequisites

Requirements

Cisco recommends that you have a familiarity with the Cisco Nexus Operating System (NX-OS) and basic Nexus architecture before you proceed with the information that is described in this document.

Components Used

The information in this document is based on these software and hardware versions:

- N9K-C93180YC-FX3
- nxos64-cs.10.4.3.F

The information in this document was created from the devices in a specific lab environment. All of the devices used in this document started with a cleared (default) configuration. If your network is live, ensure that you understand the potential impact of any command.

Introduce NAT Support on N9K

Terminology

- NAT - NAT is a technique used in networking to modify the source or destination IP address of IP packets.
- PAT - Port Address Translation, also known as "Overloading NAT", multiple internal IP addresses share a single external IP address, differentiated by unique port numbers.
- TCP Aware NAT - TCP-aware NAT support enables NAT flow entries to match the state of TCP sessions and get created and deleted accordingly.

NAT TCAM Resource

By default no TCAM entries are allocated for the NAT feature on Nexus 9000. You must allocate the TCAM size for the NAT feature by reducing the TCAM size of other features.

There are three types of TCAM involved in NAT operations:

- **NAT region**

NAT utilizes the TCAM NAT region for packet matching based on IP address or port.

Each NAT/PAT entry for inside or outside source addresses requires two NAT TCAM entries.

By default, ACL atomic update mode enabled, 60% of Non-Atomic scale number is supported.

- **TCP Aware region**

For each NAT inside policy with "x" aces, "x" number of entries is required.

For each configured NAT pool, one entry is required.

TCP-NAT TCAM size must be doubled when atomic update mode is enabled.

- **NAT Rewrite Table**

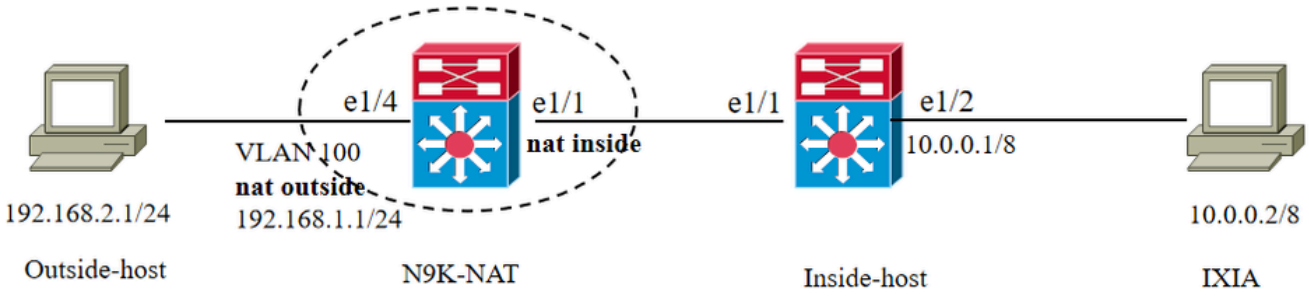
NAT rewrites and translations are stored in the "NAT Rewrite Table," which exists outside of the NAT TCAM region. The 'NAT Rewrite Table' has a fixed size of 2048 entries for Nexus 9300-EX/FX/FX2/9300C and 4096 entries for Nexus 9300-FX3/GX/GX2A/GX2B/H2R/H1. This table is exclusively used for NAT translations.

Each Static NAT/PAT entry for inside or outside source addresses requires one "NAT Rewrite Table" entry.

For more detail about TCAM on Nexus 9000, you can reference Classification TCAM with Cisco CloudScale ASICs for Nexus 9000 Series Switches White Paper.

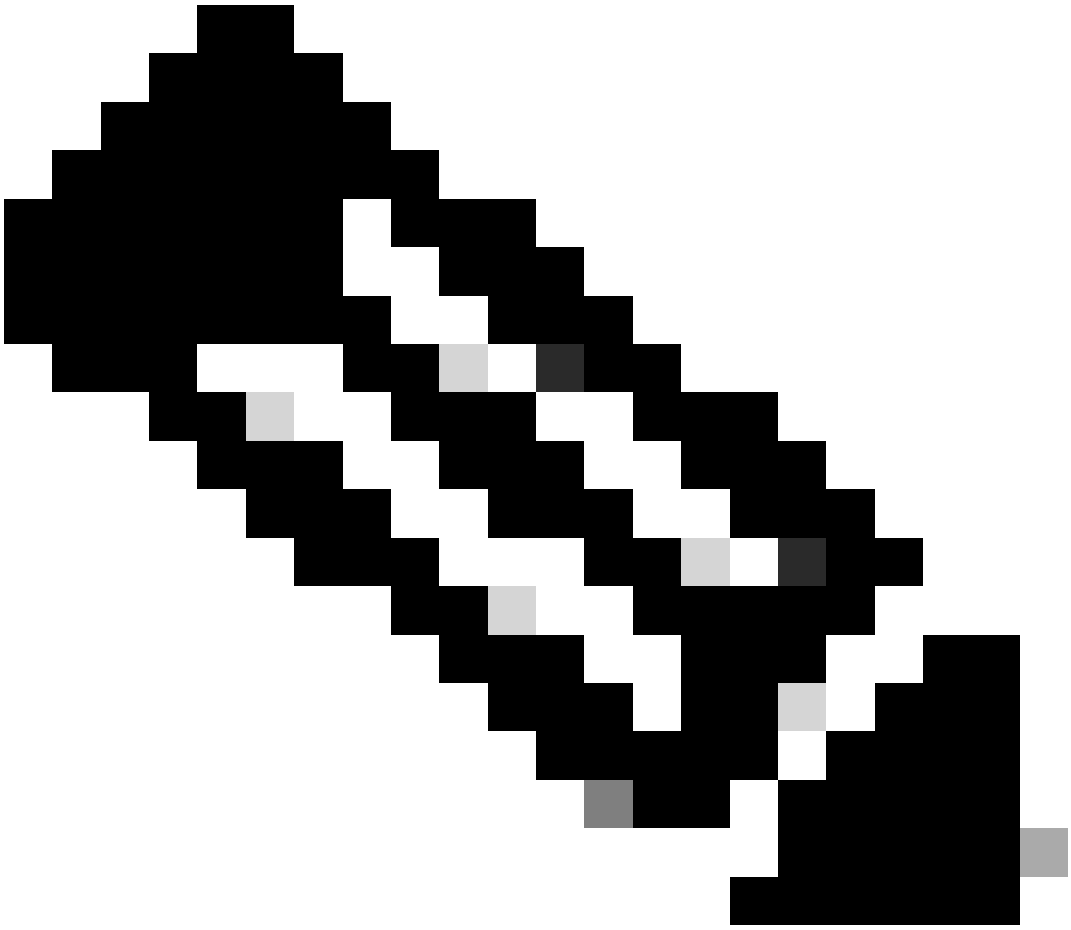
Configuration and Verification

Topology



N9K-NAT Configuration

hardware access-list tcam region nat 1024 hardware access-list tcam region tcp-nat 100 ip nat translation max-entries 80

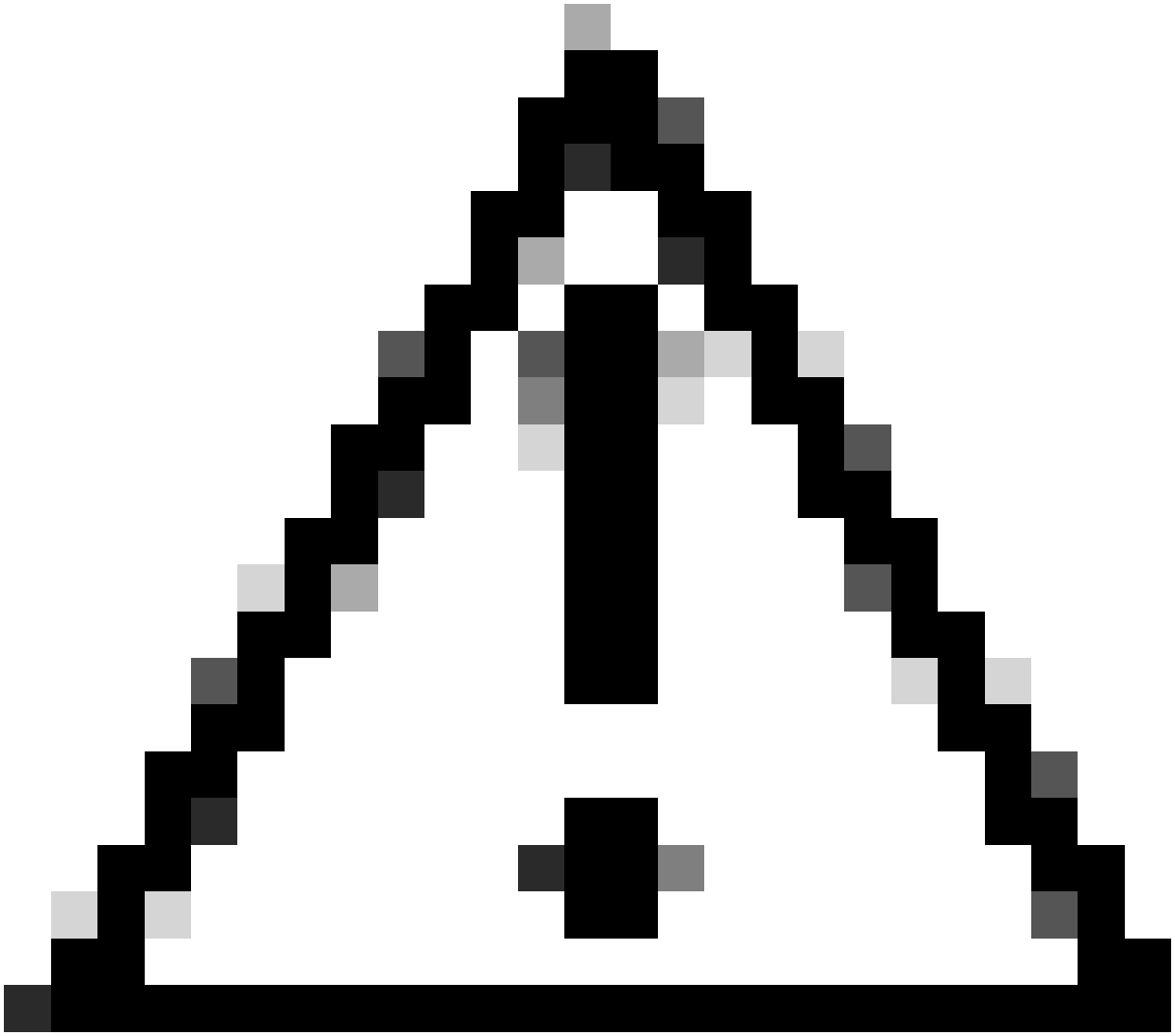


Note: By default, the dynamic nat translation max-entries is 80.

---

ip access-list TEST-NAT 10 permit ip 10.0.0.1/8 192.168.2.1/24 ip nat pool TEST 192.168.1.10 192.168.1.10 netmask 255.255.255.0 ip nat inside source list TEST-NAT pool TEST overload

---



Caution: The interface overload option for inside policies option is not supported on the on the Cisco Nexus 9200, 9300-EX, 9300-FX 9300-FX2, 9300-FX3, 9300-FXP, and 9300-GX platform switches for both outside and inside policies

---

interface Vlan100 no shutdown ip address 192.168.1.1/24 ip nat outside
interface Vlan100 no shutdown ip address 192.168.1.1/24 ip nat outside

Verification

Inside-host Ping

Source IP of the data packet: 10.0.0.1 Converted to IP: 192.168.1.10

Destination IP: 192.168.2.1

Inside-host# ping 192.168.2.1 source 10.0.0.1 PING 192.168.2.1 (192.168.2.1): 56 data bytes 64 bytes from 192.168.2.1: icmp_seq=0 ttl=63

time=0.784 ms 64 bytes from 192.168.2.1: icmp_seq=1 ttl=63 time=0.595 m

NAT Translation Table Check

N9K-NAT# show ip nat translations icmp 192.168.1.10:60538 10.0.0.1:48940 192.168.2.1:0 192.168.2.1:0 icmp 192.168.1.10:60539 10.0.0.1:0 192.168.2.1:0 192.168.2.1:0

NAT Statistics

N9K-NAT# show ip nat statistics IP NAT Statistics == Stats Collected since: Tue Sep 3 14:33:01 2024 ---------- Total active translations: 82 / Number of translations active in the system. This number is incremented each time a translation is created and is decremented each time a translation is cleared or times out. No.Static: 0 / Total number of static translations present in the system. No.Dyn: 82 / Total number of dynamic translations present in the system. No.Dyn-ICMP: 2 ---------- Total expired Translations: 2 SYN timer expired: 0 FIN-RST timer expired: 0 Inactive timer expired: 2 ---------- Total Hits: 10475 / Total number of times the software does a translations table lookup and finds an entry. Total Misses: 184884 / Total number of packet the software dropped Packet. In-Out Hits: 10474 In-Out Misses: 184884 Out-In Hits: 1 Out-In Misses: 0 ---------- ---------- Total SW Translated Packets: 10559 / Total number of packets software does the translation. In-Out SW Translated: 10558 Out-In SW Translated: 1 ---------- Total SW Dropped Packets: 184800 / Total number of packet the software dropped Packet. In-Out SW Dropped: 184800 Out-In SW Dropped: 0 Address alloc. failure drop: 0 Port alloc. failure drop: 0 Dyn. Translation max limit drop: 184800 / Total number of packets dropped due to configured maximum number of dynamic translation entry limit reached. (ip nat translation max-entries <1-1023>) ICMP max limit drop: 0 Allhost max limit drop: 0 -------- ---------- Total TCP session established: 0 Total TCP session closed: 0 ---------- NAT Inside Interfaces: 1 Ethernet1/1 NAT Outside Interfaces: 1 Vlan100 ---------- Inside source list: ++++++++++++++++++++ Access list: TEST-NAT RefCount: 82 / Number of current references to this access list. Pool: TEST Overload Total addresses: 1 / Number of addresses in the pool available for translation. Allocated: 1 percentage: 100% Missed: 0

Frequently Asked Questions

What Happens when NAT TCAM exhausted?

If TCAM resources are exhausted, the error log reported.

2024 Aug 28 13:26:56 N9K-NAT %ACLQOS-SLOT1-2-ACLQOS_OOTR: Tcam resource exhausted: Feature NAT outside [nat-outside] 2024 Aug 28 13:26:56 N9K-NAT %NAT-2-HW_PROG_FAILED: Hardware programming for NAT failed:Sufficient free entries are not available in TCAM bank(3)

What Happens when Max-entries is Reached?

By default, the NAT translation max-entries is 80. Once the dynamic NAT translation entries exceed the maximum limit, the traffic be punted to the CPU, resulting in an error log and drop.

Ping test failure: Inside-host# ping 192.168.2.1 source 10.0.0.1 count unlimited interval 1 PING 192.168.2.1 (192.168.2.1): 56 data bytes Request 0 timed out N9K-NAT Error log: 2024 Sep 5 15:31:33 N9K-NAT %NETSTACK-2-NAT_MAX_LIMIT: netstack [15386] NAT: Can't create dynamic translations, max limit reached - src:10.0.0.1 dst:192.168.2.1 sport:110 dport:110 Capture file from CPU: N9K-NAT# ethanalyzer local interface inband limit-captured-frames 0 Capturing on 'ps-inb' 15 2024-09-05 15:32:44.899885527 10.0.0.1 → 192.168.2.1 UDP 60 110 → 110 Len=18

Why are Some NAT Packets Punted to the CPU?

Normally, there are two scenarios in which traffic be routed to the CPU.

The first occurs when NAT entries have not yet been programmed to the hardware, at this time the traffic

need to be processed by the CPU.

Frequent hardware programming puts strain on the CPU. To reduce the frequency of programming NAT entries in the hardware, NAT programs translations in one-second batches. The command**ip nat translation creation-delay** delays session establishment.

The second scenario involves packets that are sent to the CPU for processing during the initial phase of establishing a TCP session and during the termination interactions of that.

Why NAT Works without proxy-arp on Nexus 9000?

There is a feature called **nat-alias** added from version 9.2.X . This feature is enabled by default and resolves NAT ARP issues. Unless you disable it manually, you do not need to enable ip proxy-arp or ip local-proxy-arp.

NAT devices own Inside Global (IG) and Outside Local (OL) addresses and are responsible for responding to any ARP requests directed to these addresses. When the IG/OL address subnet matches the local interface subnet, NAT installs an IP alias and an ARP entry. In this case, the device uses local-proxy-arp to respond to ARP requests.

The no-alias feature responds to ARP requests for all the translated IPs from a given NAT pool address range if the address range is in the same subnet as the outside interface.

How add-route Argument Works on N9K and Why It is Mandatory?

On Cisco Nexus 9200 and 9300-EX, -FX, -FX2, -FX3, -FXP, -GX platform switches, the add-route option is required for both inside and outside policies due to the ASIC hardware limitation. With this argument, the N9K adds a host route. TCP NAT traffic from outside to inside gets punted to the CPU and can drop without this argument.

Before:

192.168.1.0/24, ubest/mbest: 1/0, attached *via 192.168.1.1, Vlan100, [0/0], 10:23:08, direct 192.168.1.0/32, ubest/mbest: 1/0, attached *via 192.168.1.0, Null0, [0/0], 10:23:08, broadcast 192.168.1.1/32, ubest/mbest: 1/0, attached *via 192.168.1.1, Vlan100, [0/0],10:23:08, local

After:

192.168.1.2/32, ubest/mbest: 1/0 *via 10.0.0.2, [1/0], 00:02:48, nat >>route created by NAT feature 10.0.0.2/32, ubest/mbest: 1/0 *via 192.168.100.2, [200/0], 06:06:58, bgp-64700, internal, tag 64710 192.168.1.0/24, ubest/mbest: 1/0, attached *via 192.168.1.1, Vlan100, [0/0], 20:43:08, direct

Why does NAT Support a Maximum of 100 ICMP entries

Normally, ICMP NAT flows time out after the expiration of the configured sampling-timeout and translation-timeout. However, when ICMP NAT flows present in the switch become idle, they time out immediately after the expiration of the sampling-timeout configured.

Beginning with Cisco NX-OS Release 7.0(3)I5(2), hardware programming is introduced for ICMP on Cisco Nexus 9300 platform switches. Therefore, the ICMP entries consume the TCAM resources in the hardware. Because ICMP is in the hardware, the maximum limit for NAT translation in Cisco Nexus platform Series switches is changed to 1024. Maximum of 100 ICMP entries are allowed to make the best usage of the resources. It is fixed, and there is no option to adjust the maximum ICMP entries.

Related Information

Cisco Nexus 9000 Series NX-OS Interfaces Configuration Guide, Release 10.4(x)

Classification TCAM with Cisco CloudScale ASICs for Nexus 9000 Series Switches White Paper

Cisco Nexus 9000 Series NX-OS Verified Scalability Guide