SNAT with NAT Gateway

Index

SNAT

- allows traffic from a private virtual network to connect to the internet while remaining fully private.
- SNAT rewrites the source IP and port of the originating packet to a public IP and port combination.
- Ports are used as unique identifiers to distinguish different connections from one another.
- SNAT also allows multiple private instances within a virtual network to use the same single public IP address or set of IP addresses (prefix) to connect to the internet.
- ! When NAT gateway makes multiple connections to the same destination endpoint, each new connection uses a different SNAT port so that connections can be distinguished from one another.
- SNAT #port_exhaustion occurs when a source endpoint has run out of available SNAT ports to differentiate between new connections.
- When #port_exhaustion occurs, connections fail.

Scale SNAT for NAT Gateway

- Scaling NAT gateway is primarily a function of managing the shared, available #SNAT_port_invenntory
- #SNAT_port_invenntory is provided by public IP addresses, public IP prefixes or both attached to NAT gateway.
- SNAT port inventory is made available on-demand to all instances within a subnet attached to NAT gateway.
- When multiple subnets within a virtual network are attached to the same NAT gateway resource, the SNAT port inventory provided by NAT gateway is shared across all subnets.
- source

NAT gateway dynamically allocates SNAT ports

- NAT gateway dynamically allocates SNAT ports across a subnet's private resources, such as virtual machines.
- Pre-allocation of SNAT ports to each virtual machine is required for other SNAT methods.
- This pre-allocation of SNAT ports can cause SNAT port exhaustion.
- With NAT gateway, pre-allocation of SNAT ports isn't required, which means SNAT ports aren't left unused by virtual machines not actively needing them.
- After a SNAT port is released, it's available for use by any virtual machine within subnets configured with NAT gateway.

NAT gateway SNAT port selection and reuse

#SNAT-ports/reuse

- NAT gateway selects a SNAT port at random out of the available inventory of ports to make new outbound connections.
- If NAT gateway doesn't find any available SNAT ports, then it reuses a SNAT port.
- The same SNAT port can be used to connect to multiple different destinations at the same time.

- A SNAT port can be reused to connect to the same destination endpoint. Before the port is reused.
- NAT gateway places a SNAT port reuse timer for cool down on the port after the connection closes.
- The SNAT port reuse down timer helps prevent ports from being selected too quickly for connecting to the same destination.
- This process is helpful when destination endpoints have firewalls or other services configured that place a cool down timer on source ports.
