SIP protocol and NAT problems

Applies to all VoipNow versions!

This article explains the Network Address Translation process and how it impacts the communication process.

About NAT

Traversing NAT (Network Address Translation) is one of the issues hindering SIP communications. With an ideal Internet, all devices would be able to communicate end to end without any intermediaries, except for routers. This implies that each device has a public IP address, that is a public reachable Internet identity.

In reality, today many of the devices connected through the Internet are using a NAT function that occurs in the border router. Not only does this function stop the Internet from initiating connections to the device (which is bad for IP telephony or other forms of peer-to-peer communications), but it also protects the users against malicious attacks. Using NAT, one may also connect multiple devices to the Internet by using only one public IP address. Therefore, NAT comes with advantages and disadvantages at the same time.

Why SIP does not work behind NAT by default

The reason is that many of the communication parameters in SIP are transmitted within the SIP message. Such parameters include the IP and port numbers used for signaling and media. A SIP device behind NAT does not know much about how it will be seen from the Internet, it only knows its own IP address and the ports where the SIP application runs. Once communication with the Internet starts, the NAT device translates the private **IP:port** combination of the SIP device connected on the private NAT interface to a temporary mapping of a public **IP:port** on the interface connected to the Internet.

How to fix NAT problems

The answer depends on the NAT type you have. There are various NAT types (according to the RFC):

- **Full cone**: A full cone NAT is one where all requests from the same internal IP address and port are mapped to the same external IP address and port. Furthermore, any external host can send a packet to the internal host, by sending a packet to the mapped external address.
- **Restricted cone**: A restricted cone NAT is one where all requests from the same internal IP address and port are mapped to the same external IP address and port. Unlike a full cone NAT, an external host (with IP address X) can send a packet to the internal host only if the internal host had previously sent a packet to IP address X.
- **Port restricted cone**: A port restricted cone NAT is like a restricted cone NAT, the only difference is that the restriction includes port numbers. More specifically, an external host can send a packet with source IP address X and source port P to the internal host only if the internal host had previously sent a packet to IP address X and port P.
- **Symmetric**: A symmetric NAT is one where all requests from the same internal IP address and port to a specific destination IP address and port are mapped to the same external IP address and port. If the same host sends a packet with the same source address and port, yet to a different destination, a different mapping will be used. Furthermore, only the external host receiving the packet can send a UDP packet back to the internal host.

For the first 3 NAT types you can use a STUN server. STUN is a client-server protocol. A VoIP phone or software package may include a STUN client, which will send a request to a STUN server. The server then reports back to the STUN client the public IP address of the NAT router and the port that was opened by the NAT to allow incoming traffic back in to the network. The response also allows the STUN client to determine what type of NAT is in use since NAT types handle incoming UDP packets differently.

STUN will not work with Symmetric NAT (also known as bi-directional NAT) which is often found in the networks of large companies. With Symmetric NAT, the IP address of the STUN server is different than that of the endpoint, and therefore the NAT mapping the STUN server sees it different from the mapping that the endpoint would use to send packets through to the client.

Once a client has discovered his external addresses, he can relate them to his peers. If the NATs are full cone, then either side can initiate communication. If they are restricted cone or restricted port cone, both sides must start transmitting together. Protocols such as SIP use UDP packets for the transfer of sound/video/text signaling traffic over the Internet. Unfortunately as both endpoints are often behind NAT, a connection cannot be set up in the traditional way. This is where STUN comes in handy. The STUN server is contacted on UDP port 3478, however the server will hint clients to perform tests on alternate IP and port number too (STUN servers have two IP addresses). The RFC states that this port and IP are arbitrary.

Other options besides NAT

- Eliminate NAT completely from your network setup, by creating a VPN between your phones and your server.
- Another solution would be to force your devices to shorten the time-frame between the REGISTER/OPTIONS packages sent to the server. By doing so, the servers will always know where your device is. Since most devices usually drop NAT connections after 120 seconds, you must set RE-REGISTER TIMEOUT, REGISTER TIMEOUT (the name differs depending on the device manufacturer) to less than 120 seconds. For most of them, 60 seconds or one minute is enough. However, in some cases you must set it to less than 30 seconds.
- You have NAT problems when you can dial out, but you do not seem to be able to receive calls, unless you have just made a call, or just registered the device. In Asterisk, your extension status shown by `sip show peers` is UNKNOWN.

Related articles

- Understanding and blocking ghost calls
- How to monitor VoipNow with Homer
- How to use Homer capture agents with VoipNow
- How SIP forking works in VoipNow

- Understanding codec negotiation