Redundancy and Load-Balancing Framework for Stateful NAT

draft-xu-behave-stateful-nat-standby-03

Xiaohu Xu  
(xuxh@huawei.com)

Mohamed Boucadair  
(mohamed.boucadair@orange-ftgroup.com)

Yiu Lee  
(yiu_lee@cable.comcast.com)

Gang Chen  
(chengang@chinamobile.com)

IETF 77, Anaheim
Changes from -01

- Some editorial changes;
- Two new co-authors added;
Motivations

- Due to the forthcoming IPv4 address exhaustion, CGN will be deployed more widely to share IPv4 addresses among a large number of customers. Hence,
  - Single point of failure should be avoided so as to avoid impacting the service delivery to a large number of customers.
  - To scale the CGN deployment better, traffic loads should be shared among a group of CGNs.
- This draft describe a framework for stateful NAT redundancy and load-balancing requirements.
  - e.g., the address pool configuration and route announcement requirements.
Scenarios and Terminology

- Internal realm is the network where the session initiator is located, while the external realm is the network where the session responder is located.
  - E.g., in the case of NAT64, the internal realm is IPv6 network/Internet while the external realm is IPv4 network/Internet.

- The mechanism is suitable for all stateful CGNs, including NAT64, NAT44 (e.g., DS-Lite) and NAT46.
  - The following will take NAT64 as an example.
Two Standby Modes

- **Cold Standby**
  - To keep the NAT failover transparent to the internal hosts. In other words, when the Primary NAT fails, all the existing established sessions will be flushed out. The internal hosts are required to re-establish sessions to the external hosts;

- **Hot Standby**
  - To keep the established sessions intact during NAT failover. In other words, when the Primary NAT fails, the Backup NAT will take over all the existing established sessions. The internal hosts are not required to re-establish sessions to the external hosts.
Address Pool Configuration in Cold Standby Mode

- The addresses for external hosts should remain as is despite NAT failover.
  - NAT64 devices belonging to a redundancy group should be configured with an identical prefix64.

- To avoid the newly elected Primary from occasionally assigning an address/port pair, which had even been assigned to internal host A by the failed Primary, to internal host B.
  - Each NAT64 device should be configured with different external address pools without any overlapping.
Route Announcement in Cold Standby Mode

- Through manual configuration or election mechanism (e.g., VRRP), one NAT router is designated as the Primary, and the other as the Backup.
- In the internal realm, the Primary NAT announces a route to the Prefix64.
  - In election mode, the Backup could do nothing.
  - In manual configuration mode, the Backup should also announce into the internal realm a route to the prefix64 with a higher enough cost or in a larger granularity for potential takeover.
  - In either mode, once the connectivity to the external realm is lost, the Primary should withdraw the route to the Prefix64 advertised before.
- In the external realm, each NAT device should announce a route to its own external address pool.
Address Pool Configuration in Hot Standby Mode

- The addresses for external hosts should remain as is despite NAT failover.
  - NAT64 devices belonging to a redundancy group should be configured with an identical Prefix64.

- The external addresses for internal hosts should also remain as is in the event of failover.
  - NAT64 devices of a redundancy group should be configured with an identical external address pool and synchronize their NAT states among them.
Route Announcement in Hot Standby Mode

- Through manual configuration or election mechanism (e.g., VRRP), one is designated as the Primary, the other as the Backup.
- The Primary announces into the internal realm a route to the prefix64, and announces into the external realm a route to the external address pool.
  - In election mode, the Backup could do nothing.
  - In manual configuration mode, the Backup should also announce the same routes as those advertised by the Primary, but with a higher enough cost or in a larger granularity for potential takeover.
  - In either mode, once the connectivity to the external or internal realm is lost, the Primary should attempt to withdraw the routes it had advertised before.
Load-balancing

- Associate different prefix64s to different redundancy groups.
  - Each group is associated with a distinct Prefix64.
  - Each group is configured with a different address pool without any overlapping.
- Alternative: anycast?
Next-Step

- Adopt it as a new charter item?