Application-Initiated Flow High Availability through PCP

draft-vinapamula-flow-ha-02

S. Vinapamula (sureshk@juniper.net)

S. Sivakumar(ssenthil@cisco.com)

M. Boucadair(mohamed.boucadair@orange.com)

Overall Context

- Best Effort is the norm, but serviceability is key
  - Internet service continuity is critical in ISPs environments.
  - This is part of quality of experience KPIs.
  - May be an important differentiation factor when it comes to select a network provide: This is even important for Enterprise Market
- Operators have their own network dependability strategies
  - Goal: is to ensure a service that is up and running almost all the time (4 nines or even 5 nines for some services)
  - Regulatory bodies in some countries may benchmark the availability of networks.
- The current practice is to deploy redundant nodes
  - Depending on the network function, a state may be created for each or a set of connections.
  - This state has to be check-pointed on backup systems.
  - ***It is expensive to checkpoint each and every connections' state.***
  - ***All business critical connections' state have to be check-pointed.***

Limitations of Current Approaches

- Hard for a network to identify/guess which connection is (business) critical
  - This is subscriber-specific: a flow can be sensitive for a User#1 while it is not for another User#2.
  - Can vary in time: A flow can be sensitive in Hour X, while it is not later.
- Heuristics are not deterministic
- DPI based identification is not only expensive, but also fails to identify if the payload is encrypted.
- Not every long lived connection is business critical
  - Think about free subscribers of a streaming service.
- Not every short-lived connection is not business critical
  - Consider the example of Shorter phone conversations such as emergency calls

Rationale

- Applications/Users are the best judge to figure if any of its connections is critical or not
- An application/user can signal to the network such indication
- The network can decide to honor such request or not
- The application is aware about the behavior to be expected from the network
- Network planning operations and HA strategies are triggered with real needs
  - HA resources will be reserved accordingly
  - Cost due to over-dimensioning can be optimized

PCP Solution

- A PCP client can signal its expected HA behavior associated with a flow
  - Achieved with a CHECKPOINT-REQUIRED option
- A PCP server can honor or discard such request
- A PCP client can update its HA behavior with a refresh PCP request
  - clear or set CHECKPOINT-REQUIRED behavior
- A PCP client can include PREFER_FAILURE if check-pointing is mandatory for a given flow
  - The server may discard such request as per any PCP request!
- A PCP server can set quota per subscriber to limit the amount of entries that can be elected to be HA

Advantages

- Only business critical flows are check-pointed
- Effective utilization of resources
- The mechanism is not specific to NAT, but applies to every flow-aware function in the network.
- An upstream device can notify other devices about the HA behavior received from an application

Next steps

- The proposed approach is straightforward and aligned with the current networking trend
  - Indeed, Operators are currently investigating open network APIs to interact with applications/services
  - PCP is an interesting tool to signal flow information to the network
  - Policy-decision making process at the network side will be enriched with application-initiated information
- The proposed approach is simple and deterministic
  - Superior to heuristic-based approached in current deployments
- The proposed approach allows for per-subscriber policies
- Any interest to continue this work in PCP?