

draft-rosenberg-dispatch- cloudsip

Jonathan Rosenberg
Five9

Cullen Jennings
Cisco

Tolga Asveren
Ribbon

Dispatch/GenART IETF 110

Problem Statement

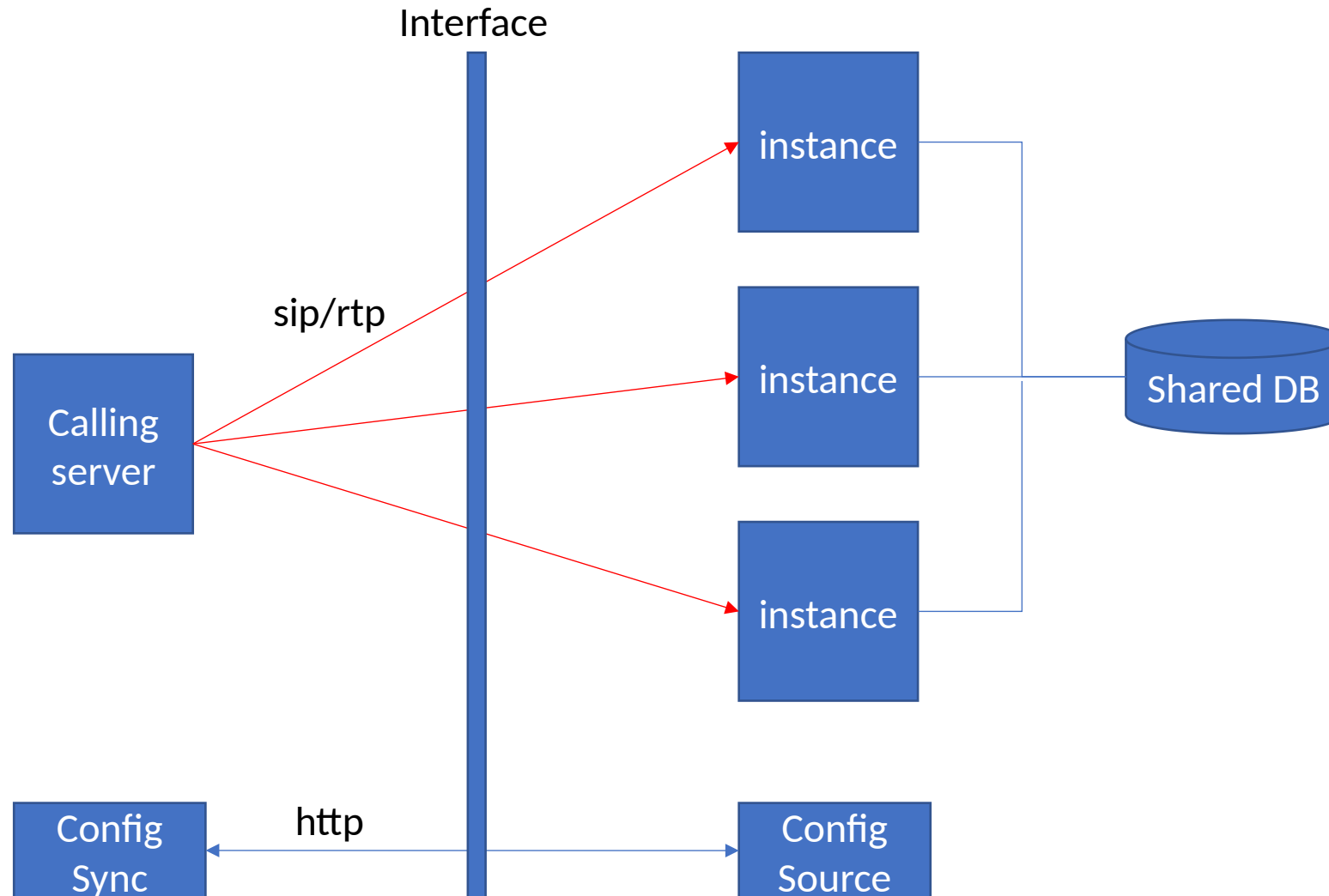
- B2BUA-WM widely deployed – full of state – making HA hard
- Two parts of HA
 - New calls succeed
 - Existing calls continue
- “Existing calls continue” is hard and uncommon
 - IP takeover with significant state replication
 - IP takeover approaches are not amenable to cloud clustering deployment models

Goal – enable HA in clusters of B2BUA-WM

Solution Requirements

- The solution must **enable a call to be recovered in less than 2 seconds**. This time represents the amount of time before which a user would hangup because they cannot hear the other party. A recovered call means that media continues to flow, and future signaling for features or call hangup, can be performed
- The HA technique must not require servers in the cluster to replicate any SIP/SDP/RTP state beyond the dialog identifiers for calls
- The solution should minimize the changes required to the SIP and RTP protocols and their respective implementations
- The solution must support the case where the telco is using traditional SBCs and is not deploying kubernetes or using public cloud
- The solution must enable **fully automated elastic expansion and contraction of cluster**
- The solution must support availability, so that when an instance in a cluster fails, new calls are distributed across the remaining N instances
- The solution must support availability, so that **when an instance of a cluster fails, all of the active calls that were being handled by that instance are spread across the remaining nodes** in the cluster, within 2 seconds
- The solution must support clusters wherein each instance of a cluster has a differing amount of capacity for call handling
- The solution must support the ability for instances of a cluster to **gracefully shut down** without dropping calls

Reference Architecture



Carrier to Enterprise SIP trunks

Intra-domain SIP hops

Carrier to Carrier SIP trunks

Solution

1. Push configuration of cluster members to calling server (facilitates elastic scaling)
2. Rapid failure detection at calling server (SIP OPTIONS or RTP)
3. INVITE w. Replaces sent to a live instance of cluster, rebuilds call and generates a downstream INVITE w. Replaces

Next Steps – Looking for Folks interested to prototype

- Softswitches
- SBCs
- Audio and Video Mixers and SFUs