Genesys Mobile Services Deployment Guide

Multi-site Deployment

4/14/2025

Multi-site Deployment

For a multi-site deployment, you must consider the following questions:

- How are API requests from mobile devices going to be routed?
- How is the mobile/web client going to retry the request in case of network failure?
- How will the client find the addresses for the global load balancer or site load balancer?
- How is the telephony network going to route the call to the number the client is receiving through the data API call?
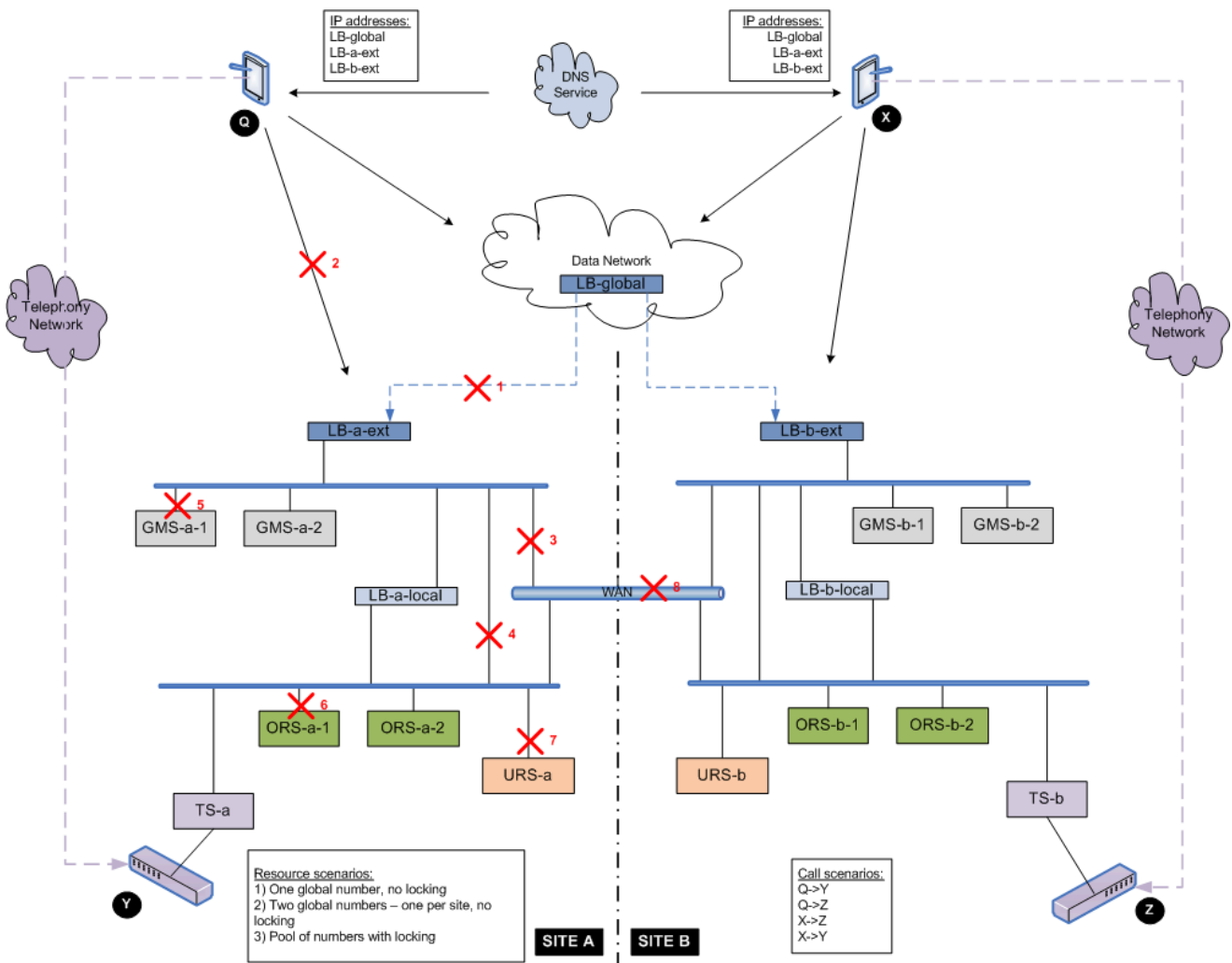- Is the call origination direction user-originated or user-terminated?

The general recommendation for user originated scenarios is to deploy an independent GMS cluster per site, and control call routing by ensuring that no intersecting pools of incoming phone numbers are configured. This way, if the client is requesting an access number through the data API call, it will be routed to Site A by the data network (internet), and then the call initiated by this client will land on the telephony switch/gateway located on (or associated with) Site A.

The matching API call will first be performed against the GMS cluster located on Site A. If Site A fails, then Site B is tried. This provides a good level of site isolation and simplifies maintenance, as well as day-to-day operations. Contact centers will be able to distribute telephony traffic between sites by controlling data API traffic (splitting between sites or directing to one site only) using standard IP load balancers, DNS records, and so on.

Using an approach of separate GMS clusters helps to avoid complex problems of "split brain" scenarios, where typically, some type of manual intervention is required either after the split, or later on when connection between sites is restored and clusters must be reconciled back together.

For user terminated scenarios, where call matching is not required, deploying a single GMS cluster across sites may provide some benefits. Still, it is not clear whether those benefits will outweigh the risks and be more beneficial as compared to the simplicity of a dedicated site cluster approach.

The following diagram shows major components of the solution and connectivity between them through local and wide area network segments. Possible network and/or component failures are numbered and described in more detail in the table below.

Resource scenarios:
1) One global number, no locking
2) Two global numbers – one per site, no locking
3) Pool of numbers with locking

Call scenarios:
Q->Y
Q->Z
X->Z
X->Y

Failure Condition (see diagram)	Resource Locking Scenario		Two Separate GMS Clusters: One on Site A an
Create Service/ Get Access Number	Match ixn Scenario		
	Same Site	Cross Site	
1 or 2	Global number, no locking.	Client has to retry the other site.	Not affected.
	Two global numbers – one per site, no locking.	Client has to retry the other site.	Not affected.
	Pool of numbers with locking.	Client has to retry the other site.	Not affected.

Failure Condition (see diagram)	Resource Locking Scenario	Two Separate GMS Clusters: One on Site A an...	
3	Global number, no locking.	Not affected.	Not affected.
	Two global numbers – one per site, no locking.	Not affected.	Not affected.
	Pool of numbers with locking.	Not affected.	Not affected.
4	Global number, no locking.	Not affected.	Request fails. Retrying the other site is not p...
	Two global numbers – one per site, no locking.	Not affected.	
	Pool of numbers with locking.	Not affected.	
8	Global number, no locking.	Not affected.	Not affected.
	Two global numbers – one per site, no locking.	Not affected.	Not affected.
	Pool of numbers with locking.	Not affected.	Not affected.