

# Geographical Redundancy

Ericsson Service-Aware Policy Controller

## FACILITY DESCRIPTION

## **Copyright**

© Ericsson España, S.A. 2017. All rights reserved. No part of this document may be reproduced in any form without the written permission of the copyright owner.

## **Disclaimer**

The contents of this document are subject to revision without notice due to continued progress in methodology, design and manufacturing. Ericsson shall have no liability for any error or damage of any kind resulting from the use of this document.

## **Trademark List**

All trademarks mentioned herein are the property of their respective owners. These are shown in the document Trademark Information.

## **Abstract**

This document describes the Geographical Redundancy function provided by the SAPC.



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Revision Information	1
1.2	Concepts	1
<b>2</b>	<b>Function</b>	<b>2</b>
2.1	Overview	2
2.2	Data Mirroring	3
2.3	SAPC Network traffic handling	5
2.4	SAPC Geographical Redundancy States	8
2.5	Geographical Redundancy Supervision and Control Functions	10
<b>3</b>	<b>Traffic Cases</b>	<b>14</b>
3.1	SAPC Active Node Restart	15
3.2	SAPC Standby Node Restart	16
3.3	Replication Channel Unavailable	18
<b>4</b>	<b>Capabilities</b>	<b>20</b>
	<b>Glossary</b>	<b>23</b>
	<b>Reference List</b>	<b>25</b>





# 1 Introduction

This document describes the Geographical Redundancy function provided by the SAPC.

## 1.1 Revision Information

**Rev. A** This is the first release of this document.

**Rev. B** Editorial changes only.

## 1.2 Concepts

**Active SAPC** The SAPC node that is processing traffic and provisioning operations.

### **Asynchronous replication**

The data is committed in the active node and then it is replicated to the standby node. Therefore, the standby node lags behind the active until the data is replicated.

**Mated peer** For a SAPC, the mated peer is the other SAPC that is part of the geographical redundancy function.

**Origin State Id** It is a monotonically increasing value that is advanced whenever a Diameter entity restarts with loss of previous state, for example upon reboot. It is used to allow rapid detection of Diameter terminated sessions.

**Own Origin State Id** The SAPC Origin State Id.

**Replication channel** Connection between the active and standby nodes in a geographical redundant configuration used for data replication and also for geographical redundancy supervision and control functions.

**System Controller** A System controller is a processor in the SAPC cluster providing OAM and provisioning services. There are always two System controllers (SCs) in the SAPC cluster.

**Standby SAPC** The SAPC that is replicating data from the active. This SAPC does not process traffic nor provisioning operations but is ready to take over in case of failure in the active SAPC.



### Virtual IP

It is the regular method to connect a telecommunication node to an external Data Communication Network. Virtual IP (VIP) is the concept for collective addressing. Using VIP, a shared IP address can be used to address distributed functions in a telecommunication node, which is a multi-processing cluster.

## 2 Function

### 2.1 Overview

The SAPC, as a network element, provides high availability as explained in the document *Availability and Scalability*. However, this level of availability does not help in the case of complete power failure, natural disasters, such as fire or earthquakes, or deliberately destructive human behavior, such as bombings or terrorist attacks. Operators may also require the possibility to shut down clusters completely for planned or unplanned maintenance (for example, hardware or software change). The geographical network redundancy function provides this extra level of redundancy at network level. The SAPC with this feature offers a system availability target figure of 99.999%. The SAPC provides a geographical redundancy solution which enhances the In-Service Performance (ISP) for traffic and O&M interfaces. The solution is based on a hot-standby system (1+1 redundancy) composed of two SAPC single nodes and network connection allowing the nodes to communicate.

During normal operation, both nodes work in an active/standby model. The active SAPC processes the incoming traffic and provisioning operations. The standby keeps the state of the active, and it is ready to process the incoming traffic and provisioning when the active node cannot handle it. If the node being active fails down, the control is automatically taken by the other node (see Figure 1); this procedure is known as switchover. This switchover is transparent for the neighboring traffic and provisioning plane nodes, as well as for the nodes sending XML/SOAP requests/notifications.

Both SAPC nodes are interconnected through the network by a connection, called the replication channel link. The replication channel link is used to transfer changes in database information done in the active node to the standby node. The replication channel link is also used for geographical redundancy supervision and control, mainly to monitor the redundancy state of the mated peer and detect when the active SAPC is no longer operational.

The geographical redundancy function is initiated in the SAPC by performing an operational procedure to promote one of the SAPC nodes to active and another to standby.

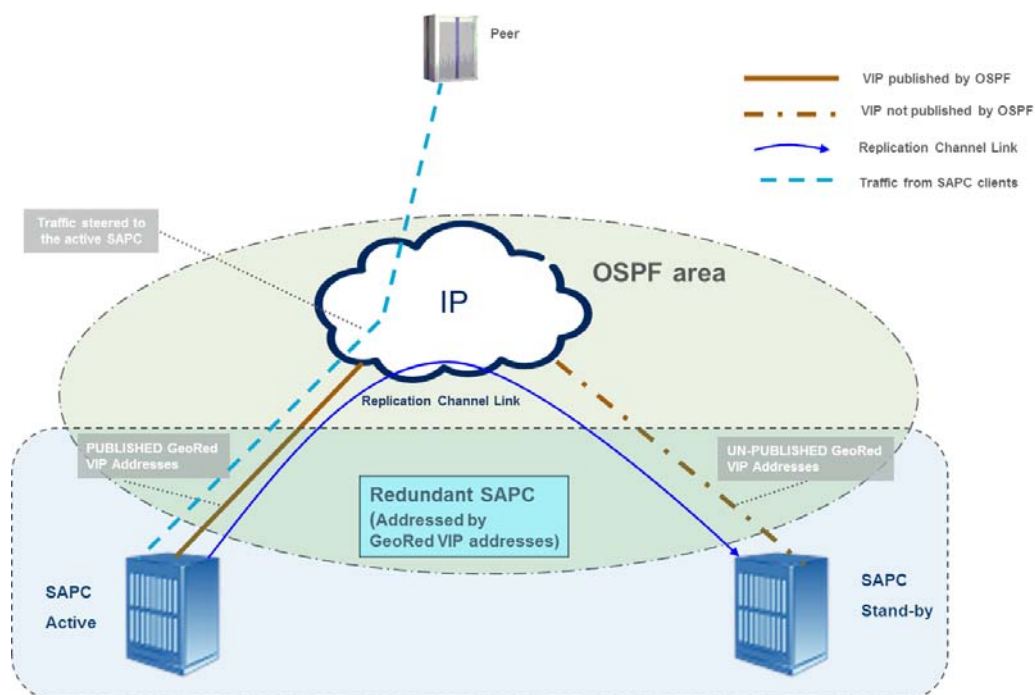


Figure 1 SAPC Geographical Redundancy.

The function consists of the following main parts:

- Data mirroring, which keeps the standby node up-to-date regarding the SAPC data that is replicated from the active node.
- Network traffic handling, which routes requests from the network to the active node.
- Geographical redundancy supervision and control, which maintains the correct redundancy state in each node.

To prevent time discrepancies, for example when active and standby nodes belong to different time zones, the SAPC in geographical redundancy uses the UTC time standard regardless of the configured time zone.

## 2.2 Data Mirroring

The SAPC geographical redundancy solution is based on the replication capability provided by the Database Service (DBS) of the Ericsson Component Based Architecture (CBA) platform. This is an asynchronous replication function, that guarantees that database changes done in one node are mirrored and applied in the mated peer.

Data mirroring (replication) between the nodes keeps data in the standby node synchronized with the active node. The asynchronous replication results in the standby node being a little behind in applying the latest transactions compared



to the active node. The active node forwards data transaction updates to the standby node over the replication channel.

The data mirroring functionality is distributed over all traffic processors in the SAPC node. There is one TCP/IP connection originating from each traffic processor in the active node, towards a traffic processor in the standby node.

**Note:** It is recommended to always have the active and standby SAPC nodes equally sized in terms of processing capacity and memory capacity. However, the function does not depend on an exact match.

### 2.2.1 Redundant Data

To be able to perform a transparent switchover in case of failure between the active and the standby SAPC nodes, the following information is replicated:

- Subscriber data (including Subscriber Fair usage accumulated data).
- Provisioning data (policies, subscriber groups data, services, and profiles).
- Mobile session information, notification data and time trigger data.

The SAPC does not replicate any other data which is not stored in the Database Service (DBS). This comprises the following data:

- The node configuration data that is provided through COM, and through configuration (cfg) files is not replicated. Therefore, the configuration data has to be provided to each SAPC node individually.
- Licensing information is not replicated. Therefore license information has to be managed in each SAPC node individually.
- Alarms, logging events and performance measurements are local to each SAPC node.

### 2.2.2 Backlog

Data mirroring is performed asynchronously for performance reasons. Hence, database changes in the active node are temporarily stored in memory while they are sent to the standby node. Next, the received changes are saved in the standby node and a confirmation is sent back to the active node. The standby node can then apply the received changes in the local database, and the active node can release the transmitted data. These memory buffers can be regarded as a backlog.

This procedure allows the SAPC to handle failures during data mirroring and ensure database consistency by forcing the same order of changes in both the active and standby nodes. Thus, having a backlog is a normal condition. However, there are situations when some transactions in the backlog cannot be processed immediately, for example, due to overload in either the sender or the receiver side, or disturbances in the replication channel. To handle this case in terms of resource utilization, it is possible to configure the maximum amount of memory used by the backlog.





## 2.2.3 Node Synchronization

Data mirroring synchronizes database changes between the active and standby nodes. However, there are a number of cases where the original contents of the databases are different, therefore synchronizing the changes does not make the contents identical either. In the following cases, a complete synchronization of the database is needed.

- After initial startup of a SAPC node.
- When transactions have to be discarded because of memory and network capacity limitations.
- After split-brain situations (loss of network connectivity between active and standby nodes).
- When data mirroring functionality is resumed after being disabled as a result of an operational procedure, for example when one of the nodes is set to Halted state or after a geographical redundancy configuration change.

Synchronization itself transmits a consistent view of the database of the active SAPC node to the standby node, where it is imported. Any changes done on the active node in the meantime are also transferred as normal database changes, which will then be applied in the standby node when the initial database is fully imported. The node synchronization process is automatically triggered by the SAPC when it is required, no manual intervention is needed. The SAPC DBS component sends notifications about the start and the end (successful or unsuccessful) of the synchronization process.

When a SAPC node reloads, it starts synchronization from the still-running peer, that is the node that reboots copies the database from the active (running) node. However, in other situations where a complete synchronization is needed, it is not possible to know which of the SAPC nodes holds the most up-to-date database. In those cases, for example after a split-brain situation, the database from the SAPC node configured as "preferred" is maintained, and the database from the other node is discarded.

## 2.3 SAPC Network traffic handling

### 2.3.1 Hot Standby Concepts

A hot standby solution means that the standby node is always able to take over traffic in the event of a failure in the active node. In addition, SAPC traffic handling is adapted to follow the hot standby principles, by always attempting to route traffic to the correct node. On the other hand, node configuration, operation and maintenance procedures are performed in each SAPC node individually.



### 2.3.1.1 Active Node

The traffic is always routed to the node that is considered as active, which means that it processes all traffic. The active node is automatically switched if a fault is detected, see Section 2.4.1 on page 8. The active node can also be switched manually, for instance, to do planned maintenance of the node.

All provisioning operations (including subscription data, services, profiles and policies) are performed in the active node and replicated in the standby node. This is transparent to the provisioning server. Incoming and outgoing traffic messages, mobile session reauthorizations, AF session events, access to external database, interactions with the online charging system, end user notifications, are also handled by the active node.

### 2.3.1.2 Standby Node

The standby node is not allowed to handle traffic nor provisioning operations, but can adopt the role of the active node, when needed.

### 2.3.1.3 Preferred Node

One of the SAPC nodes must be configured as the preferred node in geographical redundancy, and this is used when resolving some fault situations where is not possible to know which of the SAPC nodes holds the most up-to-date database.

- After recovery of the network connectivity between both SAPC nodes (split-brain situation), the database from the preferred SAPC node is maintained, and the database from the other node is discarded.
- When both SAPC nodes reload and come up nearly at the same time, the preferred SAPC node takes the role of active and provides data to the not preferred SAPC.

## 2.3.2 Network connectivity

There is only one connection point to the redundant SAPC pair seen from the external network independently of the node that is handling traffic and provisioning. The redundant SAPC solution exposes by default one VIP address for traffic and another VIP address for provisioning. These are called redundant virtual IP addresses.

The active SAPC notifies, through Open Shortest Path First (OSPF) mechanism, that it is the one that is processing traffic and provisioning, while the standby SAPC does not. If the standby node has to take the control, it notifies, through OSPF mechanism, that it is going to process the incoming traffic and provisioning. The other SAPC node removes that notification. The advertisement is done with Link-State Advertisements (LSA) messages. For more information, refer to RFC 2328 – OSPF Version 2, Reference [1].

Each SAPC handles the following VIP addresses:



<b>Traffic GeoVIP</b>	This is the VIP address that the SAPC clients use to send diameter traffic to the SAPC. In network deployments with Online Charging System, this is also the VIP address where the SAPC handles the Sy interface. This VIP is handled by both SAPC nodes.
<b>Provisioning GeoVIP</b>	This is the VIP address that the SAPC clients use to send provisioning orders to the SAPC. This VIP is handled by both SAPC nodes.
<b>ExtDB GeoVIP</b>	In network deployments with the CUDB or an external database function, this is the VIP address used to provide access to an external database system and receive SOAP notifications. This VIP is handled by both SAPC nodes.
<b>Replication VIP</b>	This is the VIP address that the SAPC exposes for the replication channel to the mated peer. Each SAPC node has its own replication VIP.
<b>O&amp;M Local VIP</b>	There is an extra local VIP associated to each SAPC node. This is the VIP used to manage the SAPC information model through COM.

The following picture shows all the IP addresses involved in the geographical redundancy scenario with details about which IP address is advertised by each SAPC node.

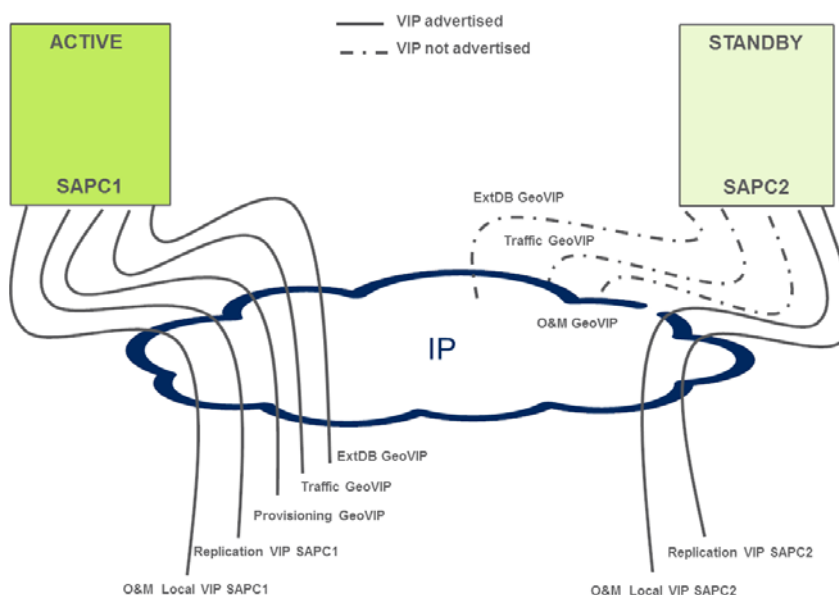


Figure 2 VIP addresses in the SAPC geographical redundancy solution.



## 2.4 SAPC Geographical Redundancy States

The node state reflects the traffic handling ability of each of the nodes. The valid node states are the following:

- Initial: This is the initial state when the SAPC is installed. The node does not handle traffic as it does not publish the GeoVIP addresses for traffic, external database and provisioning. The SAPC can only transition from this state when ordered by operational procedure.
- Active: The SAPC is handling traffic and provisioning while the mated peer is replicating the changes. In this state, the SAPC publishes the GeoVIP addresses for traffic, external database and provisioning.
- Standby: The SAPC is not handling traffic nor provisioning but is replicating the changes from the active SAPC. Therefore, it is ready to take over traffic if the active fails. In this state, the SAPC does not publish the GeoVIP addresses for traffic external database and provisioning.
- Halted: The SAPC is in this state when the geographical redundancy function is stopped as a result of an operational procedure. In this state, the SAPC does not handle traffic nor provisioning. The SAPC does not publish the GeoVIP addresses for traffic, external database and provisioning and does not replicate database changes from the mated peer. However, the O&M Local VIP can be used for SAPC configuration through COM.

### 2.4.1 Transition between States

The SAPC node executes transitions from one state to another depending on the information provided by the supervision functions explained in Section 2.5 on page 10.

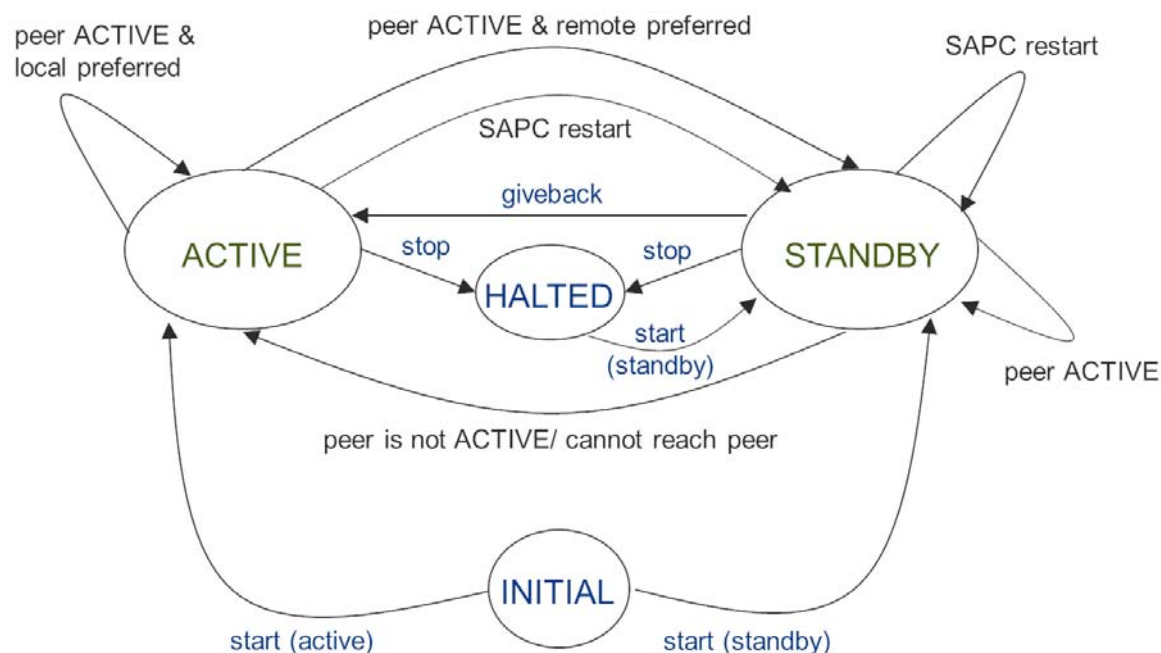


Figure 3 SAPC geographical redundancy state machine

#### 2.4.1.1 Transitions from Initial State

In this state, a SAPC node can make the following transitions when ordered by operational procedure:

- Initial to Active. This is done when the SAPC is ordered to start geographical redundancy with the role of active node.
- Initial to Standby. This is done when the SAPC is ordered to start geographical redundancy with the role of standby node.

#### 2.4.1.2 Transitions from Active State

In this state, a SAPC node can make the following transitions:

- Active to Standby. This is done in the following situations:
  - The SAPC node restarts and determines that the mated peer is available and in active state.
  - The SAPC node detects that the Database Service (DBS) is temporarily unavailable.
  - The SAPC node determines that the mated peer is the preferred active node and is currently in active state. This can happen after a split brain situation as explained in Section 2.5.2.1 on page 12.



- Active to Halted. This is done when ordered by operational procedure. This transition is not allowed if the mated peer is in Initial or Halted state.

The SAPC remains in active state when it is the preferred active node and detects that the mated peer is also in active state.

#### 2.4.1.3 Transitions from Standby State

In this state, a SAPC node can make the following transitions:

- Standby to Active. This is done in the following situations:
  - The SAPC node determines that the mated peer is unavailable, in standby state or in halted state.
  - The SAPC node configured as the preferred node is ordered by operational procedure to return to active state, typically after a takeover.
- Standby to Halted. This is done when ordered by operational procedure. This transition is only allowed if the mated peer is in active state or unavailable.

The SAPC remains in standby state when the SAPC node restarts and determines that the mated peer is available and in active state.

#### 2.4.1.4 Transitions from Halted State

In this state, a SAPC node can make the following transitions:

- Halted to Standby. This happens when the geographical redundancy function is restarted by operational procedure.

## 2.5 Geographical Redundancy Supervision and Control Functions

The Geographical Redundancy control function has the following responsibilities:

- Supervise the own state of the SAPC node to recognize when there is an internal failure in the node, such as if the Database Service (DBS) is temporarily unavailable.
- Monitor the redundancy state of the mated peer, to determine when it is unreachable or unavailable and whether it is in active, standby or halted state.
- Manage the geographical redundancy state machine, and take the appropriate actions in each state or transition. This includes:
  - To publish or not publish the GeoVIP addresses for traffic, access to external database and provisioning.
  - Set the correct value for the node `Origin State Id` upon failure.



- Handle the operational procedures to perform transitions from the Initial state and transitions to the Halted state, and provide configuration, logs and alarms related to the geographical redundancy functionality.

If the geographical redundancy control function detects that the Database Service (DBS) is temporarily unavailable in the active SAPC, a switchover is executed and the standby SAPC node takes the role of active SAPC node.

If the two System Controllers in the SAPC cluster fail in geographical redundancy configuration, the SAPC node restarts. If the failure occurs in the active SAPC node, a switchover is executed and the standby SAPC node takes the role of active SAPC node. This is to ensure that the O&M and provisioning functions are available in the SAPC when the two System Controllers fail.

### 2.5.1

#### Mated Peer Supervision

The SAPC stores its own replication state in a specific Managed Object Class (MOC) that can be managed via the NETCONF interface. This MO also stores the previous state and a time stamp when the transition between states happened.

The SAPC uses a heartbeat mechanism to monitor the availability and redundancy state of the other node through the IP network. This heartbeat is sent to the replication VIP address of the mated peer.

- The active node monitors the availability of the standby node at regular time intervals, by sending a heartbeat signal. This allows the active node to detect a problem in the replication channel or in the mated peer. The heartbeat interval has a default value of 5 seconds that can be adapted to the network operator conditions.
- The standby node answers the heartbeat signal by sending an acknowledgment to the active node.
- If the active node does not receive the acknowledgement to the heartbeat signal after several attempts, the mated peer is declared unreachable and an alarm is raised. The number of reattempts has a default value of 3 retransmissions that can be adapted to the network operator conditions.
- If the standby node does not receive a heartbeat signal during more than a predefined time period, the node transitions to active state. The maximum time period before assuming that the mated peer is unavailable is derived from the heartbeat interval and the number of reattempts as  $\text{Heartbeat\_Timeout} = \text{Heartbeat\_Interval} * (1 + \text{Number\_of\_Reattempts})$
- If the active node transitions to halted state by means of operational procedure, the active node notifies the standby node, and the standby node transitions to active state. This allows the mated peer to react immediately upon the state transition. The SAPC node in halted state also answers the heartbeat signal by sending an acknowledgment to the active node.

- If the standby node transitions to halted state by means of operational procedure, the standby node signals the active node and waits for the answer. This allows the SAPC to determine if the mated peer is in active state, and so permit the state transition.

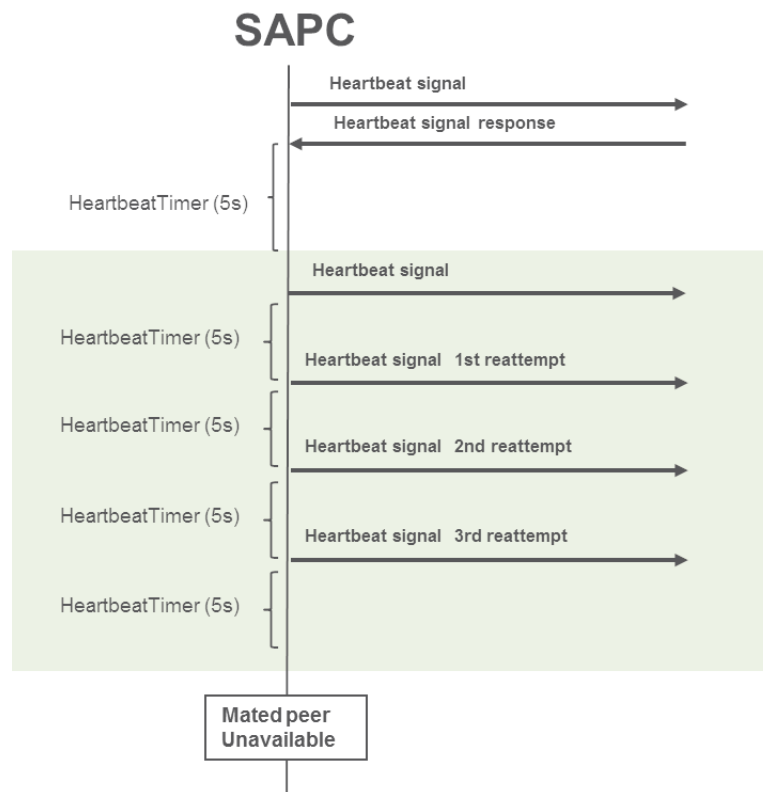


Figure 4 SAPC mated peer supervision

## 2.5.2 Fault Detection and Recovery

The basic principle for the geographical redundancy function is that the active SAPC node processes all incoming traffic and provisioning operations, while the standby node keeps the state of the active, and it is ready to take over when the active node fails. However, there are situations that might cause the standby node not to be synchronized with the active node. This chapter describes those scenarios and how they are handled.

### 2.5.2.1 Split Brain Scenario

This situation happens when the standby SAPC node detects a failure in the mated peer. The standby SAPC switches to active state and starts to announce the GeoVIP addresses to serve the traffic. If the active SAPC node is healthy, that is, only the heartbeat is lost (owing to loss of connectivity between the nodes), the active SAPC continues announcing the GeoVIP address. In this scenario, the end





result consists of two nodes in active state announcing the GeoVIP addresses to serve traffic and provisioning. This situation is known as a split brain scenario.

The split brain has the following consequences:

- The standby node switches to active state, so it starts announcing the GeoVIP addresses and serving traffic.
- Both SAPC nodes may commit traffic and provisioning operations so the databases may become inconsistent because changes cannot be replicated.
- Traffic operations may fail if the messages related to a session are steered to the SAPC node that does not hold the correct session information.

When the connectivity between the SAPC nodes is re-established, the nodes will not communicate database changes until a complete synchronization is performed. Then, the SAPC node that is not configured as the preferred node is automatically restarted. When this node reloads, it recovers the most current database information from the active node and sets the replication state to standby.

### 2.5.2.2 Simultaneous Node Restart

If both SAPC nodes (active and standby) fail, they probably do not reload at the same time. Therefore, the first one that restarts will observe loss of network connectivity to the mated peer and take the necessary actions, as described in Section 2.5.2.1 on page 12.

If both SAPC nodes come up nearly at the same time and both observe that its peer is already running, both of them will try to synchronize data from the peer. Resolving this situation is automatic. The SAPC node that is configured as preferred, provides data to the not preferred node. The procedure requires a complete synchronization of the non-preferred SAPC node.

### 2.5.2.3 Temporary Differences between Databases

The data mirroring functionality makes use of backlogs while transferring database changes from the active to the standby node is described in Section 2.2.2 on page 4. Normally, any specific transaction should be processed soon, and, therefore, be removed from the backlog. If a transaction remains in a queue for more than a minute, an alarm is raised to report that redundancy is compromised. If the database differences turn out to be temporary, the alarm is automatically cleared.

However, it can happen that node overload or network connectivity problems persist, and the backlog reaches the configured memory limit and transactions have to be dropped by either the active or the standby node. In this case, the only way to reach a state where the database contents are the same is to fully synchronize the databases. Then, the SAPC node that is configured as preferred, provides data to the not preferred node by using the procedure described in Section 2.2.3 on page 4.



### 2.5.3 Handling of the Node Origin State Id

In a stand-alone deployment, when the SAPC recovers from a restart, the database information recovered from the backup may not be fully up to date. Hence the SAPC increments its own node `Origin State Id` and includes the new value in every response message alerting the peer diameter nodes about the loss of previous session state.

In a deployment with geographical redundancy, the `Origin State Id` information is replicated between the active and standby SAPC nodes. Upon a node restart, the SAPC does not increment the `Origin State Id`, but obtain the `Origin State Id` value together with the most up-to-date database information during synchronization with the mated peer. This enables the standby SAPC to send the same `Origin State Id` value as the active SAPC, upon a failure in the active node. Transitions between SAPC redundancy states do not increase the value of the node `Origin State Id`, as those are transparent to the peer diameter nodes in the external network.

In a deployment with geographical redundancy, the SAPC only increments the `Origin State Id` if both SAPC nodes restart. This is, when the SAPC recovers from a restart and cannot replicate the `Origin State Id` information (cannot sync with the mated peer, for example due to loss of network connectivity in the replication channel), the SAPC increments its own node `Origin State Id`.

When the active or standby SAPC node detects a split-brain situation, the own node `Origin State Id` is not increased in order to provide service availability, as the session information is available in both SAPC nodes. When the network connectivity is re-established, the SAPC node that is configured as preferred continues providing service, but the `Origin State Id` is not increased.

If a SAPC node restarts in a split-brain situation, the node autonomously increments its own `Origin State Id`, and this information cannot be replicated in the mated peer. As a result each SAPC node may send a different `Origin State Id` value as long as the replication channel is unavailable. When the network connectivity is recovered the `Origin State Id` in the SAPC node that has been configured as preferred node is maintained.

## 3 Traffic Cases

This chapter explains the high level interactions that occur in the most common use cases for the Geographical Redundancy functionality:

- SAPC active node restart.
- SAPC standby node restart.

— Replication channel unavailable.

### 3.1 SAPC Active Node Restart

The following figure shows the high level flow that takes place when the active node restarts, and the main actions taken by the SAPC to perform the Geographical Redundancy functionality.

A failure in the active node makes the standby node transition to active state and start publishing the GeoVIP addresses and processing traffic. When the SAPC recovers from the restart and completes synchronization from the mated peer, it takes the role of the standby node. During the transition, ongoing traffic events may fail and diameter connections need to be reestablished, in a similar way to when a temporary connectivity problem occurs in the diameter link. The duration of the traffic switch depends on several factors, see Section 4 on page 20.

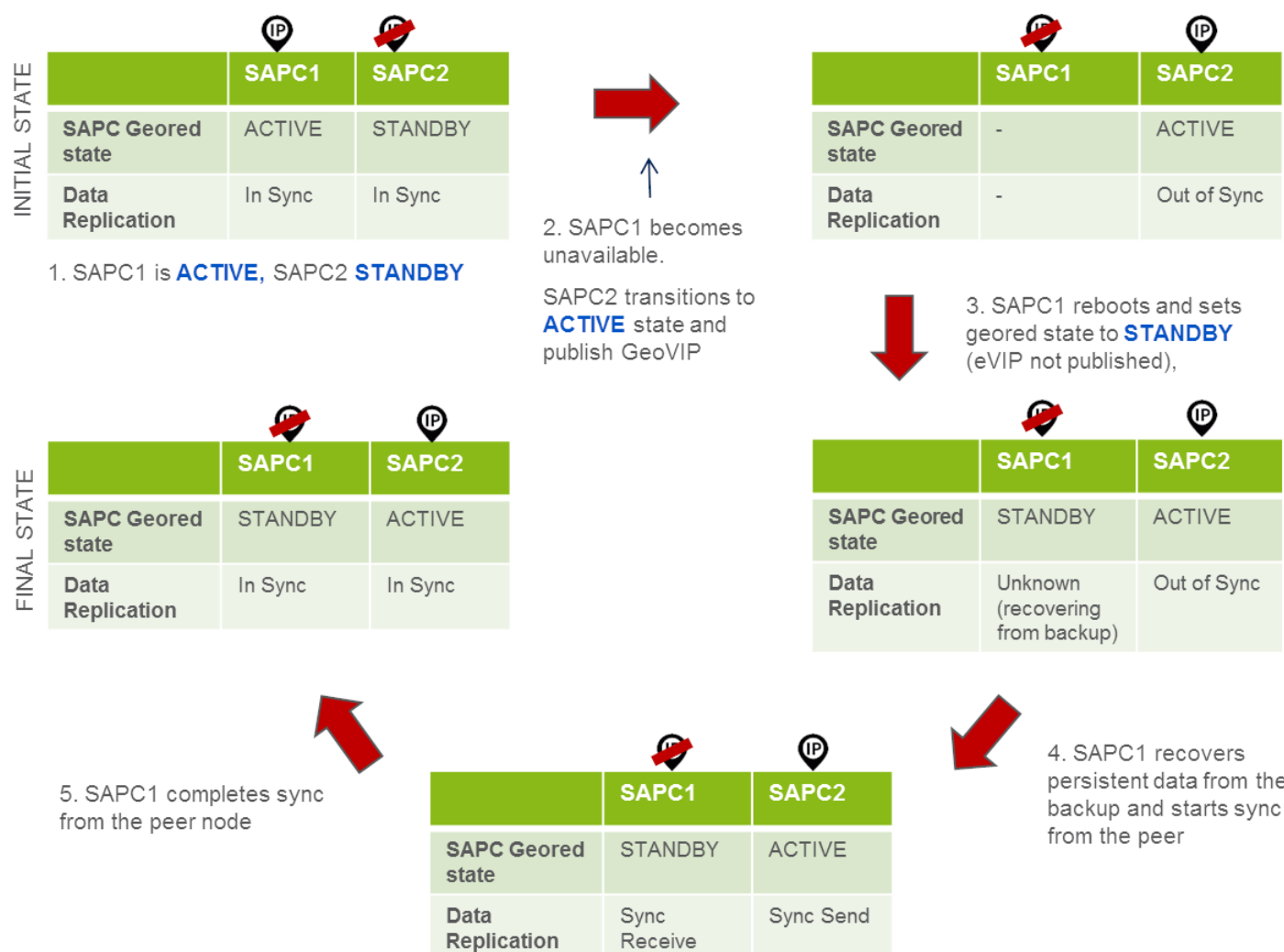


Figure 5 SAPC active node restart



1. This is the initial working condition for geographical redundancy. The SAPC1 is in active state publishing the GeoVIP addresses and processing traffic, whilst the SAPC2 is in standby state, not publishing the GeoVIP addresses, not processing traffic and replicating the changes from the active node. Data mirroring (replication) is fully operational and keeps data in the standby node synchronized with the active node.
2. SAPC1 fails, SAPC2 detects that SAPC1 is not available (due to heartbeat timeout), rises an alarm (Unable to Reach Peer), transitions to active state and publish the GeoVIP addresses. Data mirroring is interrupted and the SAPC DBS component also rises an alarm (Connection Loss). Database transactions that were pending to be replicated, are dropped. As the SAPC2 node starts to handle traffic, database changes are applied but can no longer be replicated in the mated peer. This makes the SAPC DBS component to rise another alarm in the SAPC2 node (Synchronization Needed).
3. SAPC1 completes the software reload, connects to the mated peer (which is in active state) and sets the redundancy state to standby (GeoVIP addresses not published). The SAPC2 clears the corresponding alarms (Connection Loss, Unable to Reach Peer).
4. Simultaneously with the previous step, the SAPC1 recovers all persistent database data from the latest backup and detects that the local database is out of sync. Then the SAPC DBS component rises an alarm (Initial Synchronization Needed) and starts synchronization from the active node. The synchronization process transmits a snapshot of the database of the active SAPC node to the standby node, where it is imported. Any changes done on the active node in the meantime are also transferred as normal database changes, which will then be applied in the standby node when the base view is fully imported. The SAPC DBS component in the SAPC2 node also clears the corresponding alarm (Synchronization Needed).
5. SAPC1 completes successfully synchronization from the active SAPC2 node and clears the corresponding alarm (Initial Synchronization Needed). In the final state, the SAPC2 is in active state publishing the GeoVIP addresses and processing traffic, whilst the SAPC1 is in standby state replicating the changes from the active node.

## 3.2 SAPC Standby Node Restart

The following figure shows the high level flow that takes place when the standby node restarts, and the main actions taken by the SAPC to perform the Geographical Redundancy functionality.

A failure in the standby node does not affect the SAPC traffic handling capability and is not detected by the external network nodes. However, there is a time period during which the standby node is not ready to take over with up-to-date database information, in the event of a failure of the active node.

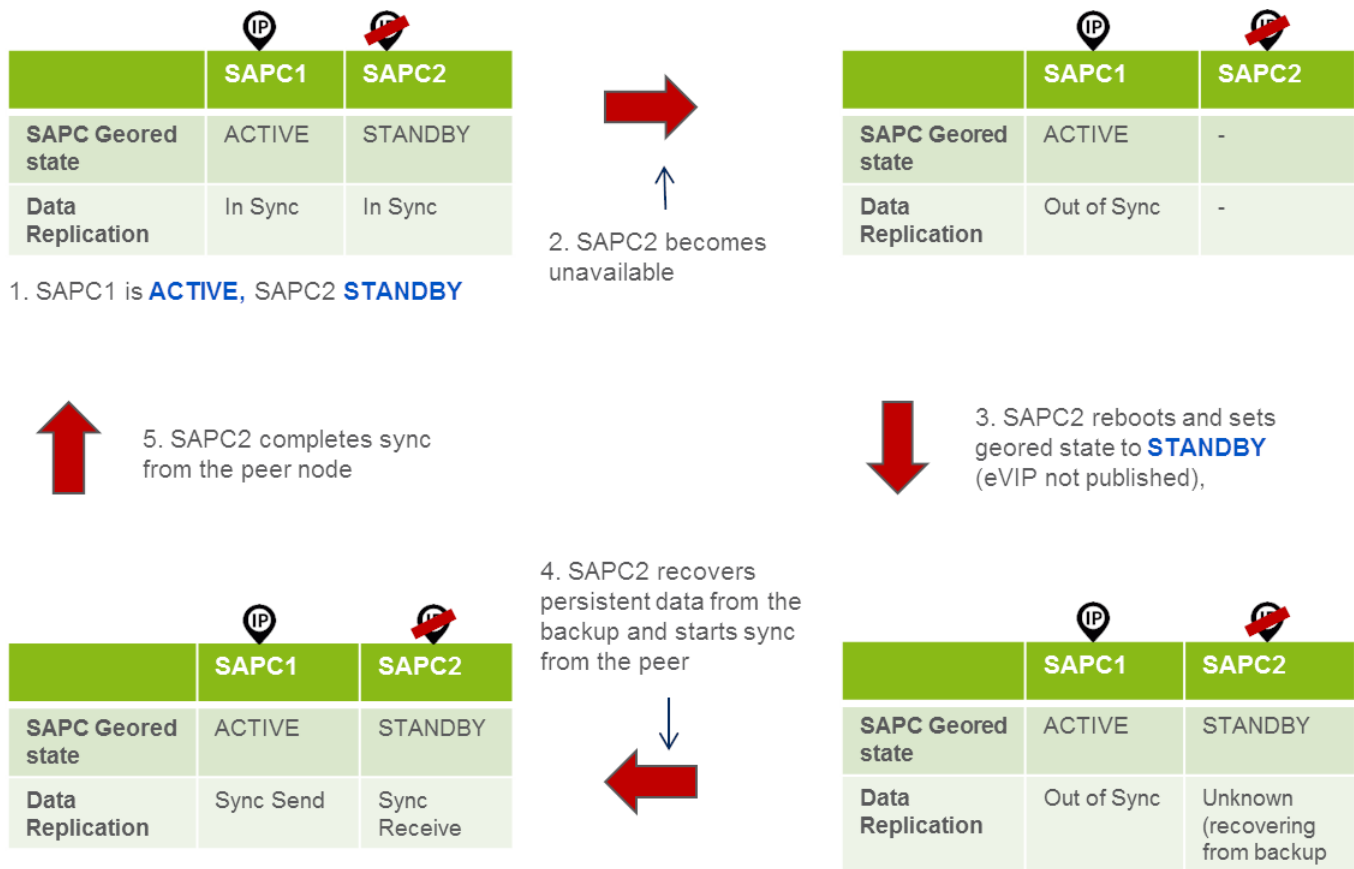


Figure 6 SAPC standby node restart

1. This is the initial working condition for geographical redundancy. The SAPC1 is in active state publishing the GeoVIP addresses and processing traffic, whilst the SAPC2 is in standby state, not publishing the GeoVIP addresses, not processing traffic and replicating the changes from the active node. Data mirroring (replication) is fully operational and keeps data in the standby node synchronized with the active node.
2. SAPC2 fails, SAPC1 detects that SAPC2 is not available (due to heartbeat timeout) and rises an alarm (Unable to Reach Peer). Data mirroring is interrupted and the SAPC DBS component also rises an alarm (Connection Loss). Database transactions in the active node that are pending to be sent to standby node and those that arrived from the active node but are still not applied in the standby node, are dropped. As the SAPC1 node continues to handle traffic, database changes are applied but can no longer be replicated in the standby node. This makes the SAPC DBS component to rise another alarm in the SAPC1 node (Synchronization Needed).
3. SAPC2 completes the software reload, connects to the mated peer (which is in active state) and sets the redundancy state to standby (GeoVIP addresses not published). The SAPC1 clears the corresponding alarms (Connection Loss, Unable to Reach Peer).



4. Simultaneously with the previous step, the SAPC2 recovers all persistent database data from the latest backup and detects that the local database is out of sync. Then the SAPC DBS component rises an alarm (*Initial Synchronization Needed*) and starts synchronization from the active node. The synchronization process transmits a snapshot of the database of the active SAPC node to the standby node, where it is imported. Any changes done on the active node in the meantime are also transferred as normal database changes, which will then be applied in the standby node when the base view is fully imported. The SAPC DBS component also clears the corresponding alarm in the SAPC1 node (*Synchronization Needed*).
5. SAPC2 completes successfully synchronization from the active SAPC1 node and clears the corresponding alarm (*Initial Synchronization Needed*). The system goes back to the initial state, where SAPC1 is active and SAPC2 standby.

### 3.3 Replication Channel Unavailable

The following figure shows the high level flow that takes place when the replication channel becomes temporarily unavailable, and the main actions taken by the SAPC to perform the Geographical Redundancy functionality.

A failure in the replication channel results in both SAPC nodes taking the role of active node, publish the GeoVIP addresses and start handling traffic. In this situation, the databases become inconsistent. When the connectivity is re-established, the database from the SAPC node configured as preferred is maintained, and the database from the other node is discarded.

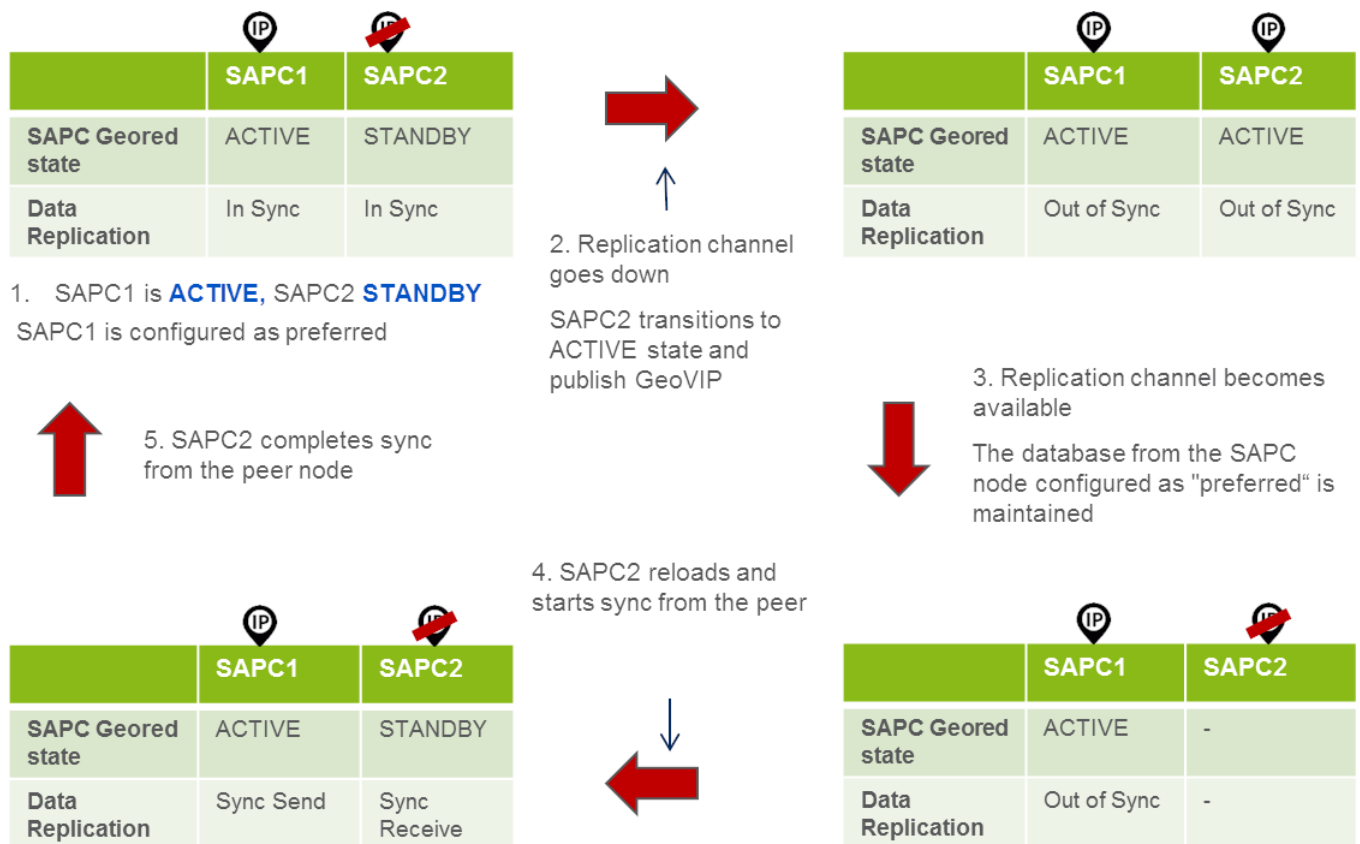


Figure 7 Replication channel temporarily unavailable

1. This is the initial working condition for geographical redundancy. The SAPC1 is in active state publishing the GeoVIP addresses and processing traffic, whilst the SAPC2 is in standby state, not publishing the GeoVIP addresses, not processing traffic and replicating the changes from the active node. Data mirroring (replication) is fully operational and keeps data in the standby node synchronized with the active node. The SAPC1 is configured as the preferred node for geographical redundancy.
2. The replication channel fails, SAPC1 detects that SAPC2 is not available and rises an alarm (Unable to Reach Peer). SAPC2 detects that SAPC1 is not available, also rises an alarm, transitions to active state and publish the GeoVIP addresses. Data mirroring is interrupted and the SAPC DBS component also rises an alarm (Connection Loss) in both nodes. This is a split brain scenario. Both SAPC nodes start handling traffic, database changes are applied locally but can no longer be replicated, so they become inconsistent.
3. The replication channel becomes available and both nodes clear the corresponding alarms (Unable to Reach Peer, Connection Loss). Data mirroring restarts and detects that a complete synchronization is needed. Then, the SAPC node that is not configured as the preferred node is automatically restarted.



4. SAPC2 completes the software reload, connects to the mated peer (which is in active state) and sets the redundancy state to standby (GeoVIP addresses not published). Then the SAPC2 recovers all persistent database data from the latest backup and starts synchronization from the active node. In the meantime the SAPC1 node continues to handle traffic.
5. SAPC2 completes successfully synchronization from the active SAPC1 node and the system goes back to the initial state.

## 4 Capabilities

For Geographical Redundancy the following capabilities must be considered:

- Enough bandwidth. The replication channel must be dimensioned to be able to handle the required bandwidth according to the traffic scenario and hardware configuration.
- Link quality. The characteristics of the link (latency and error rate) set a limit in the maximum throughput that can be achieved in the replication channel, and this throughput must fulfil the bandwidth requirements. To avoid the effects of poor performance in the replication channel link, the maximum One-Way Delay (OWD) of the replication channel, must be no more than 20 ms and the packet loss rate must be no more than 0,0001.
- System dimensioning. The system limit is imposed by the maximum sustained load at which the standby SAPC can replicate from the active SAPC without lagging behind.
- Backlog size. The maximum size of the backlog must be correctly dimensioned in order to cope with temporary overload in either the sender or the receiver side, or disturbances in the replication channel.

Regarding response times, the following capabilities must be considered for SAPC transitions:

- The time to detect loss of connectivity to the mated peer depends on the configured values for the heartbeat interval and number of re-attempts, according to the reliability of the operator transport network. Typical values to perform a transition from standby to active, may vary from 5 seconds to 20 seconds.
- Time to stop advertising the GeoVIP addresses at transition from active to other state: a Link State Update is sent after one second (controlled by the transmit\_delay OSPF parameter) without the VIP address. The LSA is (the same OSPF parameter) flooded on all the interfaces from the SAPC router, and throughout the network after one second. The next step is that the SAPC





routers and the other routers in the network recalculate their routing tables. The exact time depends on the time needed to recalculate the routes. The time required to run the algorithm depends on a combination of the size of the area and the number of routes in the database. It can take up to 10 seconds. After that, the routers stop sending packets to the former active SAPC node.

- Time to start advertising the GeoVIP at transition from standby to active state: a Link State Update packet is sent after one second (controlled by the `transmit_delay` OSPF parameter provided by eVIP component) with the VIP address. The LSA is (the same OSPF parameter) flooded on all the interfaces from the SAPC router and throughout the network after one second. The next step is that the SAPC routers and the other routers in the network recalculate their routing tables. The exact time depends on the time needed to recalculate the routes. The time required to run the algorithm depends on a combination of the size of the area and the number of routes in the database. It can take up to 10 seconds. After that, the routers start sending packets to the active node.





# Glossary

**LSA**

Link-State Advertisements (LSA)

**O&M**

Operation and Maintenance

**OSPF**

Open Shortest Path First

**SAPC**

Ericsson Service-Aware Policy Controller

**SC**

System Controller

**TP**

Traffic Payload

**VIP**

Virtual IP





## Reference List

### **Ericsson Documents**

- [1] Availability and Scalability

### **Standards**

- [2] RFC 2328 – OSPF Version 2