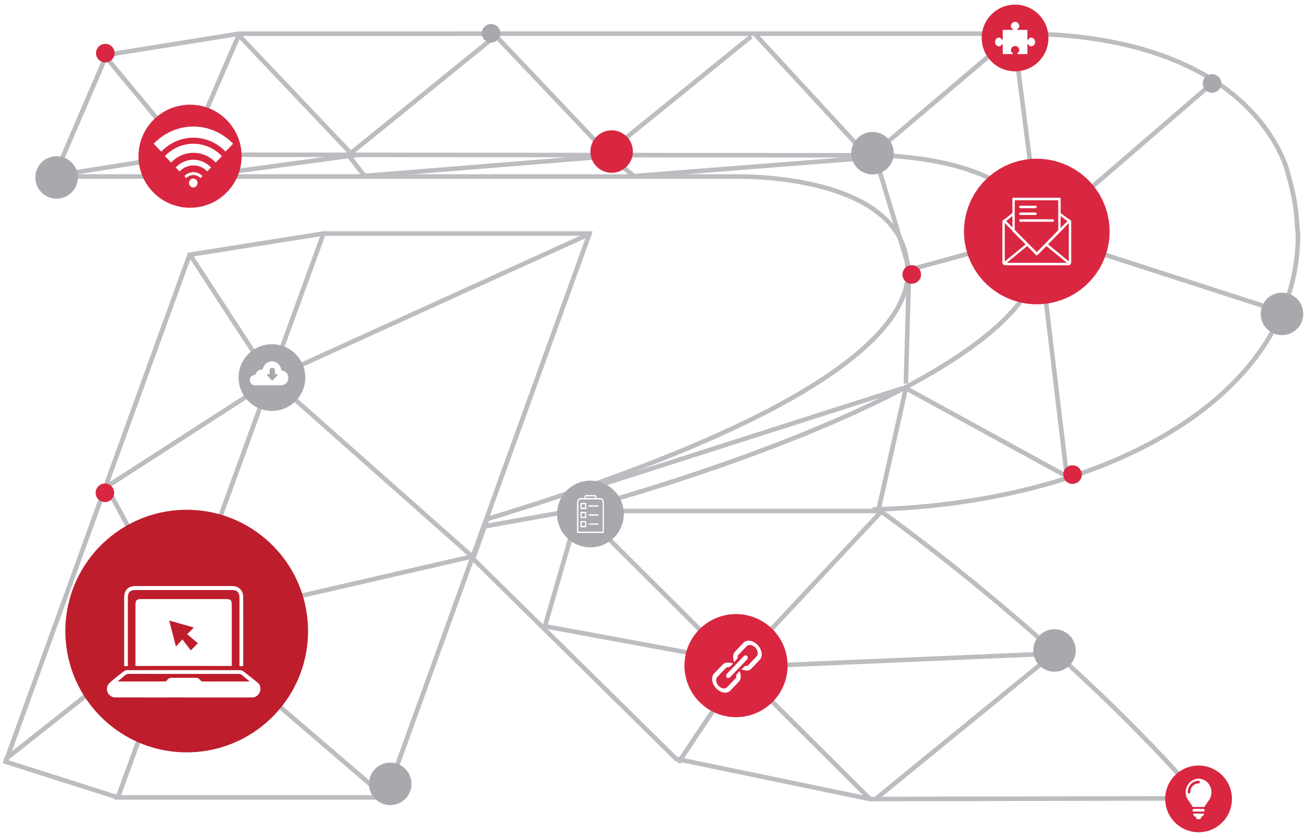
# Ruijie Hot Backup

White Paper



# Contents

This document describes in detail the hot backup redundancy High Availability (HA) technology used in some Ruijie core switches, including technical features and solutions.

# Terminology

| Terminology | Description |
|---|---|
| CPLD | Complex Programmable Logic Device. It is a logic element that is more complex than the Programmable Logic Device (PLD). |
| HA | High Availability |
| NSF | None Stop Forwarding. NSF is an important HA technology, which ensures that data forwarding is uninterrupted even if the control plane of the router encounters a failure (for example, restart due to a failure or route flapping), thereby protecting network traffic from interruption. |
| GR | Graceful Restart. GR ensures that service forwarding is uninterrupted when a routing protocol is restarted, thereby preventing the reset of the control plane. The core of the GR mechanism is described as follows: When a routing protocol of a device is restarted, GR instructs neighbor devices to keep neighbor relationships and routes to the device stable within a certain period of time. After the restart of the routing protocol is complete, neighbor devices assist the device in routing information synchronization, to restore routing information of the device to the state before restart within a short period of time. In the entire protocol restart process, network routes and data forwarding are highly stable, the packet forwarding paths are not changed, and the system can continuously forward IP packets. This process is called GR. |
| STP | Spanning Tree Protocol |
| RIB | Routing Information Base. RIB can store routing updates, which usually includes information about multiple paths to the same destination. |
| GVRP | GARP VLAN Registration Protocol. GVRP is a Generic Attribute Registration Protocol (GARP) application for dynamically configuring and spreading VLAN memberships. |
| VLAN | Virtual Local Area Network. A VLAN is a logical network divided from a physical network. |

# Overview

Nowadays, fast development of the Internet brings about a sharp increase of the network scale and network application fields. Networks have penetrated into all aspects of our life and have become an indispensable part of our social life. With widespread application of networks, enterprises and individuals can hardly make communication and exchange without networks. People show more dependency on networks and apply network more widely, and raise higher requirements for network stability and reliability. In such cases, the transient failure of device functions and temporary interruption of services may lead to severe impact and cause huge losses to every sector. Therefore, the HA of network devices has become an increasingly important demand of people, which requires improvement to the reliability of components and reduction of the error recovery time.

Network applications need to traverse multiple network segments. To ensure the network availability, all network segments must be capable of recovery from failures, and the recovery should be so fast that users and network applications do not perceive the failures.

In general, system downtime can be divided into two categories:

* **Unscheduled downtime (fault)**

Such downtime is uncontrollable and random, and is often related to defects of hardware and software components.

* **Scheduled downtime (maintenance)**

Such downtime can be scheduled within a certain period of time, for system repair, upgrade, or the like. Therefore, its impact on network availability can be minimized.

To prevent network unavailability caused by a Single Point of Failure (SPOF) and implement HA, from the global level, redundant links should be available between two communication nodes, and the topology can be managed automatically (for example, via the layer-2 STP or layer-3 routing protocol) or manually; from a local level, network nodes, especially forwarding devices, should be capable of coping with the failure of a single component. Redundancy is an important policy for addressing SPOFs.

The HA technology requires high reliability of devices. However, the implementation of HA does not rely on the improvement of reliability of components alone. As a further policy, the hardware solution provides the function of replacing components without shutting down devices and backs up some key components. Software support includes hot swapping, fault detection, and fault recovery.

From fault detection to fault recovery, a long recovery time (exception handling + device cold start) is required if a redundant card is cold-started. The status of the device needs to be backed up to implement fast fault recovery. This is the source of hot backup function requirements in RGOS. Through real-time status backup, components and their redundant components can maintain status synchronization, to achieve fast fault recovery. In addition to component reliability, Ruijie Networks provides the hot backup function on some core switches to back up the key part of the system. The purpose is to not only guarantee the normal maintenance of devices but also effectively reduce the probability of device crash and the impact of service traffic interruption. In this way, the system can run continuously, applications are sustained, and the interests of customers are protected.

The hot backup technology is a highly fault-tolerant application solution combining hardware and software. Its biggest advantage lies in use of a backup module to replace the faulty module without interrupting the system running, and lies in the capability of allowing the system to automatically perform synchronization and run after the fault is rectified, thereby effectively eliminating customers' troubles.

Ruijie devices use two Supervisor Engines to implement the hot backup function. The two Supervisor Engines work in master-slave backup mode: The one working in master mode is the master Supervisor Engine and the other working in slave mode is the slave Supervisor Engine. Users cannot directly perform operations and configuration on the slave Supervisor Engine (the master Supervisor Engine supports all CLI commands while the slave Supervisor Engine supports a few commands in special mode of the master Supervisor Engine), but need to perform operations and configuration via the CLI of the master Supervisor Engine. Then, the master Supervisor Engine synchronizes the status, data, and configuration to the slave Supervisor Engine in a timely manner to maintain status and configuration consistency between the master and slave Supervisor Engines. Once the master Supervisor Engine malfunctions or is removed in hot mode (manual master/slave switchover is also supported), the slave Supervisor Engine can automatically and rapidly become the new master Supervisor Engine (this process is called master/slave switchover) and take over the work of the old Supervisor Engine smoothly and rapidly. In this way, the device can run continuously and applications can fully recover in a short period of time.

## • Benefits of Hot Backup

Hot backup can bring the following beneficial effects to network services:

- Improvement of network availability

Data forwarding and user session statuses are maintained during device switchovers.

- Prevention of link flapping detected by neighbors

The data plane is not restarted during the switchover. Therefore, neighbors cannot detect the status change of a link from down to up.

- Prevention of route flapping

The data plane maintains the forwarding communication and the control plane rapidly creates a new forwarding table during the switchover. The process of replacing the old forwarding table with the new one is unperceivable, thereby preventing route flapping. In the switchover process, the original Slave Supervisor Engine completes route convergence and then distributes routes to line cards. Before the convergence is thoroughly complete, line cards use the old forwarding table to forward packets, and the impact on device functions is slight.

- No loss of user sessions

User sessions established before a switchover are not lost because status information is synchronized in real time.

- No loss of services

The forwarding of some packets does not involve Supervisor Engines. Therefore, when the master/slave switchover occurs, the system runs stably and the statuses inside the system including the software and hardware statuses are consistent. Line cards can still forward packets, thereby ensuring no loss of services.

- Fast software upgrade

Hot backup can ensure that services are uninterrupted during software upgrade, thereby achieving high device reliability.

- Hardware replacement

Hot backup allows replacing some hardware without interrupting services.

- No major adverse impact on efficiency

There is no significant difference in the overall operation efficiency between systems that implement hot backup and those that do not.

Undoubtedly, the hot backup function consumes more resources (including software and hardware resources), but in consideration of the reliability and stability brought by the application of this technology, such consumption is worthwhile.

In addition, the hot backup technology has the following limitations:

- The hot backup technology can run properly only when the software and hardware compositions of the two Supervisor Engines are intact and completely consistent (including the program integrity and version consistency). Otherwise, exceptions such as suspension or reset may occur during startup. If software has defects, such as a logic bug, the same fault will occur when the same triggering situations appear after switchovers. In addition, Ruijie hot backup function identifies faults only by detecting hot swapping and heartbeats. For some other faults, such as faults caused by the failure of a single module, the hot backup function cannot perceive or process such faults. In this case, a manual switchover needs to be performed or technical support is required.

- The hot backup solution is applicable only to the forwarding scheme with distributed line cards. The hot backup data channels are deployed on 100M or 1000M switching networks (100M for the M8606-CM and M8610-CM, and 1000M for M8606-CMII and M8610-CMII). The bandwidth of the data channels is sufficient for packet transmission so far.

- In the startup process (both Supervisor Engines are started at the same time or one Supervisor Engine is inserted in hot mode while the other is running), batch synchronization needs to be performed first between the master and slave Supervisor Engines so that the two Supervisor Engines have consistent statuses and data. The period before this process is complete is a window period in which the hot backup function cannot come into full play. If the master Supervisor Engine encounters a software fault or is removed during this period, the system will reset, resulting in data flow interruption.

- Not all forwarding-related functions are synchronized. Based on the degree of support for NSF, switch functions can be classified into the following types:

(1) HA support functions

Status information is synchronized in real time between the master and slave Supervisor Engines, for example, control plane functions directly related to layer-2 forwarding are synchronized in real time.

(2) HA-compatible functions

These functions do not support HA and their status data is not synchronized. When HA is enabled, these functions are still available. After a switchover, these functions start from the initialized state.

(3) HA-incompatible functions

These functions do not support HA and their status data is not synchronized. When HA is enabled, these functions are unavailable. Otherwise, the system may become abnormal. When these functions are enabled, the system degrades from the hot backup state to the boot hot state and the system synchronizes only the real-time configuration file. An example of such functions is GVRP. For details, see the description of interference sources in "Basic Concepts".

## • Technical Indicators

The system availability is generally measured by following indicators:

- **Reliability**

Reliability refers to the probability that a system encounters a fault in a certain period of time. For example, 99.999% reliability means that the probability that a device is free from faults within a year is 0.99999. "Reliability" is non-equivalent to "availability". Reliability refers to the capability that system functions do not fail while availability refers to the capability that a system provides users with service of a certain level.

- **MTBF**

The Mean Time Between Failures (MTBF) is measured by using the average time interval between two failures. An increase in the MTBF indicates that the system reliability is raised.

The fault rate is 1/MTBF.

- **MTTR**

MTTR is short for Mean Time to Repair.

- **Uptime ratio and availability**

Uptime ratio is an indicator that measures the system availability. The annual downtime of the system can be calculated using the following formula:

Downtime per year (minutes) = (1 – Uptime ratio) x 365 x 24 x 60

In general, availability is calculated using the following formula:

$$\text{Availability} = \frac{\text{MTBF}}{\text{MTBF} + \text{MTTR}}$$

The following table lists the average annual downtime for different availability values.

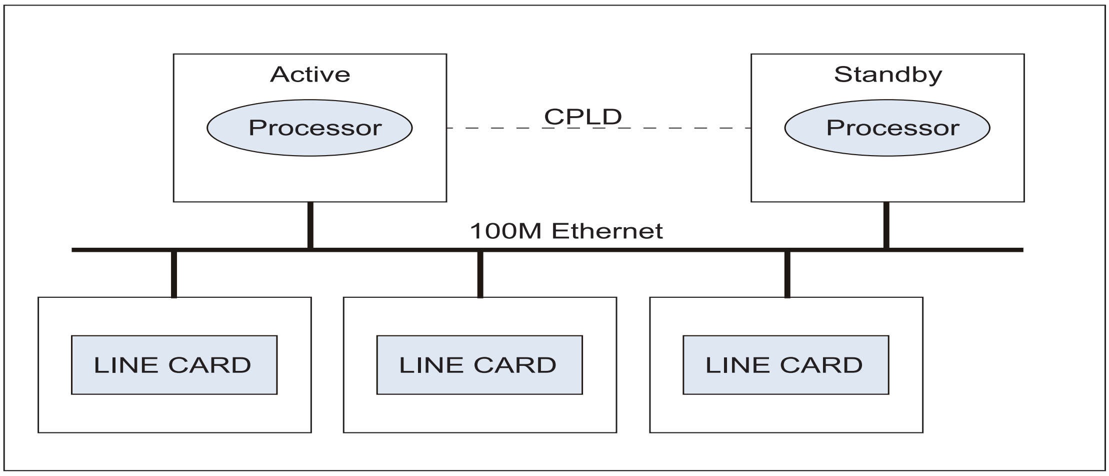
| Availability (%) | Downtime Per Year |
|---|---|
| 99 | 3.65 days |
| 99.9 | 8.76 hours |
| 99.99 | 52.6 minutes |
| 99.999 | 5.26 minutes |
| 99.9999 | 30.00 seconds |

## • Hot Backup Requirements

From the communication support perspective, internal functions of network devices are divided into the data plane, control plane, and management plane. The forwarding service is provided by network devices for users through the data plane. Therefore, as long as the data plane functions properly, the network is available to users. Devices of the chassis-based structure generally consist of Supervisor Engines, line cards, and the chassis. The Supervisor Engines provide the functions of the management plane and control plane, line cards provide the functions of the data plane, and the chassis connects the Supervisor Engines and line cards. Therefore, data can be forwarded provided that line cards work properly. The separation of the management component from the forwarding component is the hardware basis for HA.

To prevent the unavailability of the entire device caused by SPOFs, the Supervisor Engines adopt the redundancy + hot swapping solution. When the master Supervisor Engine malfunctions or is being maintained, the slave Supervisor Engine takes over work of the master Supervisor Engine. The following figure shows the physical model assumed for the hot backup function in RGOS.

**Figure 1 Physical Model of Hot Backup**



The startup of the Supervisor Engine involves the following steps:

(1) The Supervisor Engine starts up from the bootloader to complete the most basic initialization and self-check, and loads the main program. Then, the bootloader guides the main program to run.

(2) The main program initializes the system framework.

(3) The main program obtains line card information.

(4) The main program distributes configuration parameters.

The startup of the Supervisor Engine generally takes several minutes. If hot start is adopted, the Supervisor Engine starts running directly inside the main program, which reduces the startup time. The initialization of line cards always causes communication interruption: Even if line cards are not re-initialized, the protocol supporting topology management sets the management statuses of interfaces to disabled when starting running, and the routing protocol convergence will incur route flapping, resulting in communication interruption. Therefore, to keep the communication uninterrupted, it is necessary to ensure that the physical statuses and management statuses of line cards are not changed during redundancy switchovers.

If the status of the slave Supervisor Engine is always consistent with that of the master Supervisor Engine, the slave Supervisor Engine does not need to modify the statuses of line cards when taking over the work of the master Supervisor Engine. In this way, data forwarding is not affected and such a switchover between the Supervisor Engines is transparent to communication users.

Only the master Supervisor Engine can manage line cards and the slave Supervisor Engine only backs up the statuses. The running environment and operation logic of the two Supervisor Engines are different. It cannot be ensured that they give the same response to the same event. Therefore, the "mirroring" method cannot be adopted.

## • Technical Key Points

The hot backup design focus on handling the following issues:

1. Status synchronization between two Supervisor Engines

This issue concerns how the master Supervisor Engine collects information, what information needs to be collected, and how the slave Supervisor Engine processes the information so as to keep the status synchronous with that of the master Supervisor Engine.

2. Internal status consistency of the slave Supervisor Engine

The statuses of components inside the system are correlated. It cannot be ensured that all components of the slave Supervisor Engine are always in the same state because their synchronization information is sent to the slave Supervisor Engine at different time. Switchovers may occur at any time and there is a time difference from the occurrence of a fault to the detection of the fault. Therefore, for system switchovers, it cannot be assumed that the statuses of the components inside the slave Supervisor Engine are consistent, or that the statuses recorded by the slave Supervisor Engine are consistent with the actual statuses of line cards. Therefore, status consistency check is required for switchovers.

This document does not focus on other technical points, such as the election of the master Supervisor Engine and mechanism of the communication between the two Supervisor Engines. Fault detection is an independent part of the HA solution and is not described in detail in this document.

The construction methods of software components vary greatly due to complex software functions of the RGOS. According to the principle of "separation for management" in software design, hot backup is designed based on the following ideas:

1. The framework of the system is the same regardless of the role of the Supervisor Engine.

2. The system provides a reliable mechanism for the communication between the two Supervisor Engines.

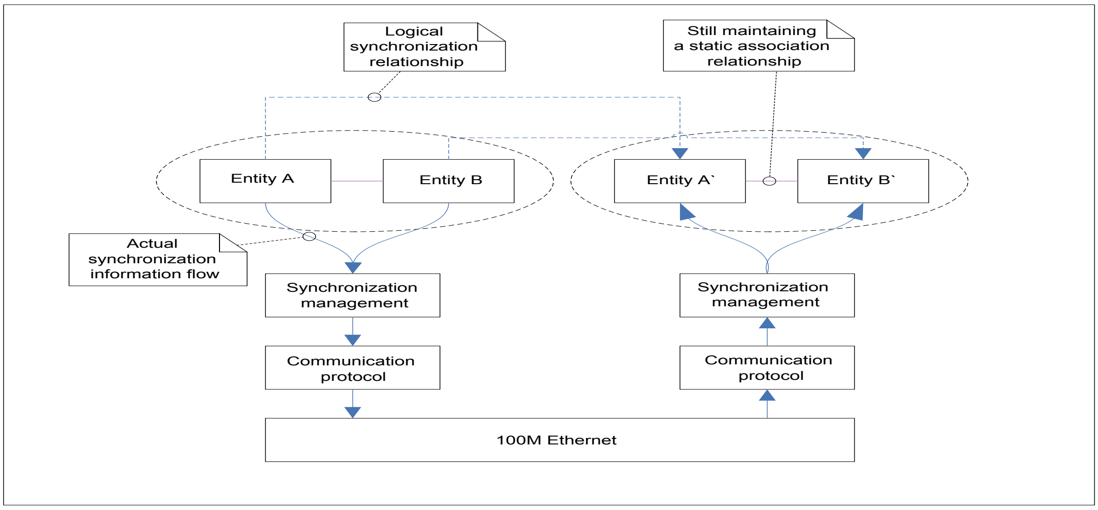
3. Components in the master Supervisor Engine map to those in the salve Supervisor Engine, and status synchronization is also based on this mapping, which is called "peer synchronization". "Peer synchronization" aims to ensure the status synchronization between peer entities, but it cannot guarantee the status consistency between components in the slave Supervisor Engine or between the slave Supervisor Engine and line cards.

4. Because the system status consistency cannot be guaranteed, a status consistency check is required during system switchovers. The status consistency check includes status consistency check within the slave Supervisor Engine and status synchronization between the slave Supervisor Engine and line cards. Separation of the data plane from the control plane is adopted to ensure that communication is uninterrupted during system switchovers.

## Peer Synchronization

Information needs to be synchronized from entities on the master Supervisor Engine to the peers (same entities) on the slave Supervisor Engine. An entity collects only its own information and sends it to the peer. These entities can use various methods to achieve synchronization, for example, directly send statuses or send events received by the entities. A higher-level system mechanism is adopted to embody the association relationship between entities.

**Figure 2 Peer Synchronization**


Logical synchronization relationship
Still maintaining a static association relationship
Entity A
Entity B
Entity A`
Entity B`
Actual synchronization information flow
Synchronization management
Synchronization management
Communication protocol
Communication protocol
100M Ethernet

Peer synchronization needs to ensure that the internal statuses of entities are consistent at any time. Therefore, the intermediate statuses cannot be sent to the peer, but there is no need to ensure synchronization between associated entities on the slave Supervisor Engine. As shown in the figure above, assume that entity A and entity B are associated. The status change of entity A will cause the status of entity B to change. Peer synchronization only requires that A` is synchronous with A. The status consistency between A` and B` is not controlled by A` but by B. The status inconsistency between A` and B` caused by a failure or switchover will be eliminated after the system switchover.

### Separation of the Data Plane from the Control Plane

Hot backup requires that the communication interruption time of the data plane should be within several milliseconds. However, the control plane needs some longer time to recover from a switchover, and deletes old forwarding entries during initialization according to normal operations. As a result, data forwarding will be interrupted after the old entries are deleted and before new entries are generated.

Separation of the data plane from the control plane refers that the old forwarding table is used before the new one is generated. The new forwarding table is used to gradually replace the old forwarding table within a period of time after a system switchover.

In the RGOS hot backup solution, dynamic and static ARP entries in the forwarding table are synchronized, but dynamic entries in the RIB and MAC address table are not synchronized. These entries are cached in the software or hardware in each line card. Therefore, though the Supervisor Engine has no complete RIB, line cards can still forward data after a switchover. In the subsequent convergence process, new entries generated during the running of protocols are stored in the RIB of the Supervisor Engine and gradually replace the old entries in the line cards. The old entries that are not replaced after protocol convergence are cleared by the line cards.

The switchover time often depends on the convergence time of the routing protocol on the control plane. Many protocols support the GR function, and fast re-convergence can be implemented only under the assistance of GR-supported neighbors.

### Limitations of the RGOS Hot Backup Solution

If the data plane cannot reflect network topology changes occurring during a switchover in a timely manner or the information sent by the control plane to other neighbors does not reflect the actual forwarding status of the data plane, the local device or other devices may make incorrect forwarding decisions. The RIB on the Supervisor Engine is not fully backed up during switchovers, and the routing part cannot be carried out if the "route once, switch many" solution is adopted. As a result, new communication requests cannot be processed.

## • Application Model

Hot backup mainly involves the following processes:

(1) Batch synchronization

The master Supervisor Engine has started running normally when the slave Supervisor Engine is started. All status information on the master Supervisor Engine needs to be sent to the slave Supervisor Engine for status consistency. Then, the slave Supervisor Engine can take over the work of the master Supervisor Engine when necessary. A major difficulty in batch backup lies in the large amount of data.

(2) Real-time synchronization

The master and slave Supervisor Engines have consistent statuses after batch backup is complete. Then, status changes on the master Supervisor Engine still need to be backed up to the slave Supervisor Engine. Such backup usually involves only a single entity.

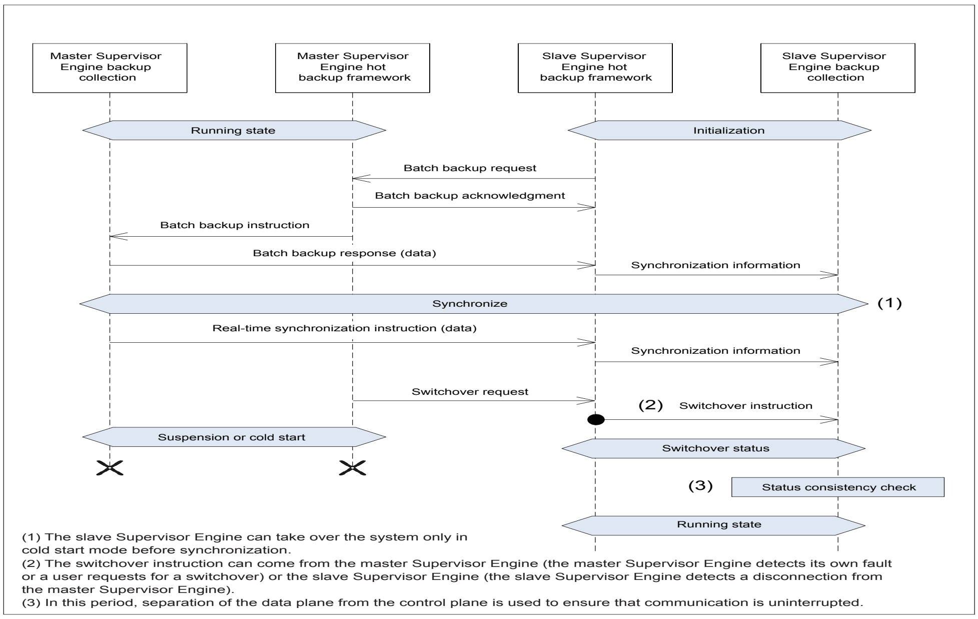
A scheduled backup mechanism can be adopted theoretically, but this mechanism requires information buffering and may cause the loss of a large amount of information upon unexpected failures. Therefore, this mechanism is not adopted in RGOS. To cope with synchronization information burst, entities need to control such burst, for example, an entity can send synchronization information at scheduled time, which is beyond the scope of the hot backup design.

(3) Master/Slave switchover

Master/Slave switchover aims at achieving status consistency inside the slave Supervisor Engine (the new master Supervisor Engine) and between the slave Supervisor Engine and line cards. The separation of the data plane from the control plane is adopted to ensure that communication is uninterrupted during switchovers.

The following figure shows the typical lifecycle model supporting the hot backup system.

**Figure 3 Lifecycle Model**


Master Supervisor Engine backup collection
Master Supervisor Engine hot backup framework
Slave Supervisor Engine hot backup framework
Slave Supervisor Engine backup collection
Running state
Initialization
Batch backup request
Batch backup acknowledgment
Batch backup instruction
Batch backup response (data)
Synchronization information
Synchronize
(1)
Real-time synchronization instruction (data)
Synchronization information
Switchover request
(2)
Switchover instruction
Suspension or cold start
Switchover status
(3)
Status consistency check
Running state

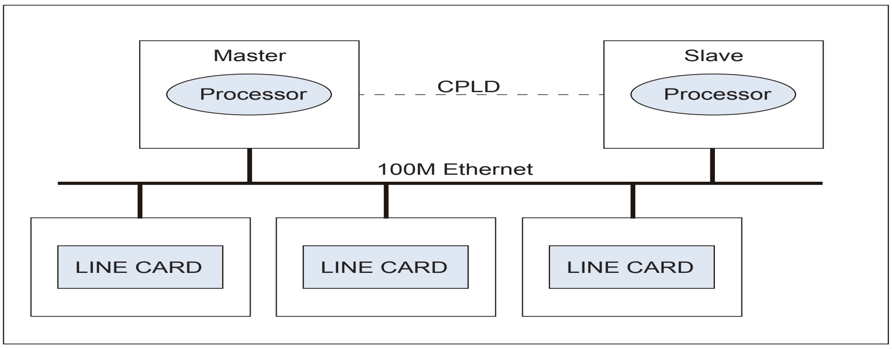
(1) The slave Supervisor Engine can take over the system only in cold start mode before synchronization.
(2) The switchover instruction can come from the master Supervisor Engine (the master Supervisor Engine detects its own fault or a user requests for a switchover) or the slave Supervisor Engine (the slave Supervisor Engine detects a disconnection from the master Supervisor Engine).
(3) In this period, separation of the data plane from the control plane is used to ensure that communication is uninterrupted.

Obviously, apart from fast fault recovery, the hot backup function can be used for software upgrade, hardware replacement, and other purposes.

# Technical Principle

The following figure shows the system physical model used in the hot backup solution (same as Figure 1).

**Figure 4 Outband Communication of a Ruijie Core Switch**



Once the master Supervisor Engine breaks down due to a fault, the slave Supervisor Engine can take over the work of the master Supervisor Engine and become the new master Supervisor Engine to ensure service continuity. At this point, all distributed line cards retain their own routing tables, and therefore they can continue to work.

However, there is still a window period in which the hot backup function cannot work and the hot backup also has limitations. The following briefly describes some specific technologies and advantages and limitations of hot backup.

## • Batch Backup and Real-time Backup

One major function of hot backup is to back up data and statuses. The master Supervisor Engine synchronizes configuration files, running status, and running data to the slave Supervisor Engine in a timely manner so that the slave Supervisor Engine can immediately take over the work of the master Supervisor Engine when the master Supervisor Engine fails. In addition, hot backup ensures configuration consistency during switchovers, so that data flows are uninterrupted and services are always available.

The backup function mainly includes batch backup and real-time backup. Batch backup refers that, after the device is started or a Supervisor Engine is inserted in hot mode, the master Supervisor Engine synchronizes a large amount of configuration and data of the entire system (including all static configuration and some dynamic configuration) to the slave Supervisor Engine at a time when the two Supervisor Engines establishes a hot backup relationship. After batch synchronization is complete, the hot backup relationship between the master and slave Supervisor Engines is successfully set up, and the data and statuses maintained by the master and slave Supervisor Engines are basically consistent. Conditions necessary for batch backup are as follows: The master Supervisor Engine stays in the normal running state and the slave Supervisor Engine has been initialized. Real-time backup refers that the master Supervisor Engine instantly synchronizes the running configuration or data (such as ARP entries and MAC addresses that are just learned, and configuration files saved by users in real time) to the slave Supervisor Engine in an incremental manner. Some data can be backed up regularly.

Real-time backup is the process that the master Supervisor Engine synchronizes real-time information to the slave Supervisor Engine and the slave Supervisor Engine processes such information (the slave Supervisor Engine uses the same method to process information synchronized via batch backup and real-time backup). The following describes the establishment of the hot backup relationship between the master and slave Supervisor Engines and the batch synchronization process.

### Basic Concepts

The following describes related terms:

- Singleton

When the device is just started and no peer is detected or only one Supervisor Engine is inserted, the status of the hot backup system is Singleton, indicating only one Supervisor Engine. The displayed peer status is None.

- Cold

The cold state indicates that batch synchronization is not complete. In this state, the redundancy system is not ready and the status and configuration of the master Supervisor Engine are not fully synchronized to the slave Supervisor Engine.

- Hot

The hot state indicates that batch synchronization is complete. In this state, the redundancy system is ready and the status and configuration of the master Supervisor Engine are fully synchronized to the slave Supervisor Engine. Then real-time synchronization will be performed.

- Boot cold/Boot hot

Boot cold/Boot hot indicates that the slave Supervisor Engine is in the cold backup state with batch synchronization uncompleted/completed.

- Cold/Standby hot

Cold/Standby hot indicates that the slave Supervisor Engine is in the hot backup state with batch synchronization uncompleted/completed.

- Violate

Violate indicates interference sources. Due to implementation limitations, the status may become disordered after a system switchover if certain content is configured in hot backup. Such content is an interference source. For example, if GVRP information is not backed up, the slave Supervisor Engine has only static VLANs but no dynamic VLANs. However, both static VLANs and dynamic VLANs exist on the hardware. The system status is disordered if the system continues running after the switchover. This is equivalent to the high availability incompatible feature on Cisco devices. When an interference source exists in the system, the master and slave Supervisor Engines do not establish a hot backup relationship but a cold backup relationship.

* **Currently, interference sources in the system include the following entities:**

* **GVRP: a GARP application for dynamically configuring and spreading VLAN memberships.**

* **Super VLAN: a VLAN division method. It is also called VLAN aggregation, which is dedicated for optimizing IP address management.**

* **PVLAN: Private VLAN.**

* **MCAST: Multicast. MCAST is the Protocol Independent Multicast (PIM), which includes PIM Sparse Mode (SM) and PIM Dense Mode (DM).**

* **Dot1x: 802.1x, used to control the network access authentication for users and provide the authorization and accounting functions for users.**

* **Protocol VLAN: a VLAN division technology based on the packet protocol type. Packets of the same protocol type and carrying no VLAN ID can be distributed to the same VLAN.**

If hot backup has been established before, after an interference source is set, the master Supervisor Engine resets the slave Supervisor Engine to enter the cold backup mode. If the system works in cold backup mode due to the existence of the interference source, after all interference sources are removed (no software and hardware version inconsistency exists), the master Supervisor Engine resets the slave Supervisor Engine to enter the hot backup mode.

- Hot backup

The master and slave Supervisor Engines are in the hot backup relationship. The master Supervisor Engine synchronizes the status and configuration to the slave Supervisor Engine in a timely manner. Once the master Supervisor Engine malfunctions, the slave Supervisor Engine can immediately take over the work of the master Supervisor Engine to ensure that services are uninterrupted.

- Cold backup

The master and slave Supervisor Engines are in the cold backup relationship. The master Supervisor Engine synchronizes the status and configuration to the slave Supervisor Engine in a timely manner, but some running data cannot be completely synchronized to the slave Supervisor Engine. Once the master Supervisor Engine malfunctions, the slave Supervisor Engine can immediately take over the work of the master Supervisor Engine but it needs certain time to create some data.

- Switchover

When the master Supervisor Engine malfunctions or a forcible switchover is triggered manually, the slave Supervisor Engine will take over the work of the master Supervisor Engine. This process is called switchover. After the switchover, the original slave Supervisor Engine will generally reset the original master Supervisor Engine.

- Hot swapping

The system allows removing or inserting a Supervisor Engine (or line card) during device operation, without adversely affecting the system running. This process is called hot swapping.

- Link flapping

Link flapping refers that a port or link changes between up and down states repeatedly. Such changes can also cause route flapping.

- Route flapping

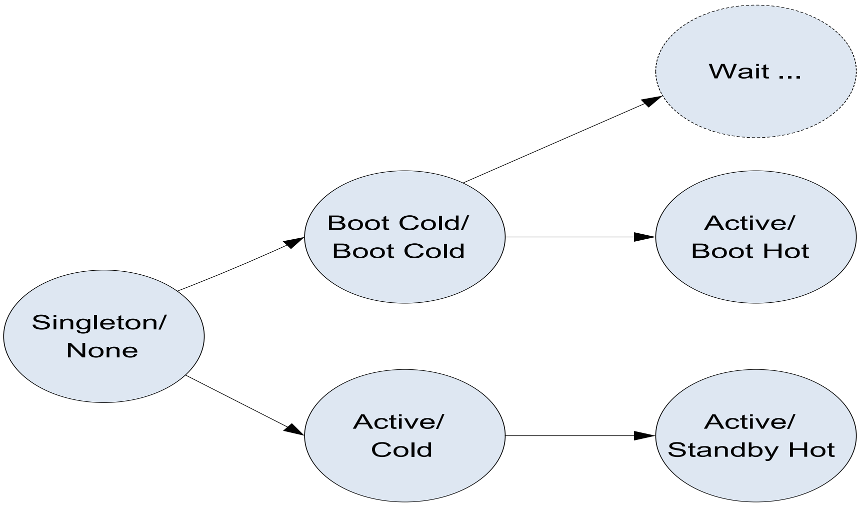
Routes change very frequently in a complex network environment and unstable routes arise easily. Route instability means that a route in the routing table may intermittently disappear and reappear. This is called route flapping. When route flapping occurs, the frequent changes and updates of the routing table will occupy a lot of bandwidth and processing time of the router, affecting the running of the device. In addition, with route reclaiming or broadcast, calculation caused by route flapping will spread inside the autonomous system, resulting in the same problems on internal routers. Currently, the technology used to control route instability is Border Gateway Protocol (BGP) damping. When route flapping occurs inside a network, a network connected to this network can utilize BGP damping to effectively defense against the impact.

In conclusion, route flapping means that the router on a network repeatedly performs route calculation and packet retransmission due to status changes of certain routes (from unavailable to available or from available to unavailable) within a short period of time. This process wastes a large number of bandwidth resources and CPU resources of the device.

## Batch Backup Process

The following figure shows some status changes during the establishment of a hot backup relationship between the master and slave Supervisor Engines (this section describes only states displayed by the show redundancy state command but not the state machine).

**Figure 5 Status Changes in the Batch Backup Process**



Status changes in a normal batch synchronization process are as follows:

1. As shown in the preceding figure, when the device is just started and no peer is detected or only one Supervisor Engine is inserted, the status is Singleton/None, indicating that the current Supervisor Engine does not identify a peer and is in the single Supervisor Engine state.

2. Continued from 1. When the master and slave Supervisor Engines detect the existence of each other, they establish a connection, start a three-way handshake, and compare information (such as the hardware and software version numbers, interference sources, and checkpoint information). If such information is consistent and no interference source exists, the master and slave Supervisor Engines enter the Active/Cold state to start batch synchronization.

3. Continued from 1. When the master and slave Supervisor Engines detect the existence of each other, they establish a connection, start a three-way handshake, and compare information (such as the hardware and software version numbers, interference sources, checkpoint information, and version numbers). If information inconsistency occurs or an interference source exists, the master and slave Supervisor Engines enter the Boot Cold/Boot Cold state to start batch synchronization. However, the information synchronized in batches is not as much as that synchronized in 2, and only some interface information and configuration files are synchronized.

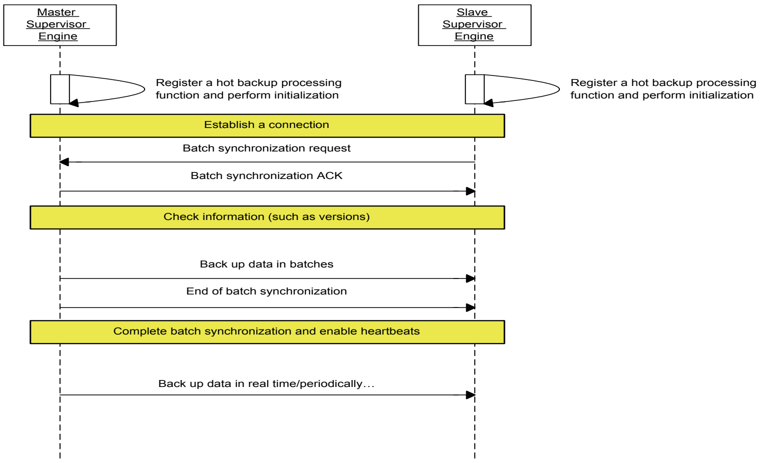
4. Continued from 2. The master and slave Supervisor Engines start batch synchronization and the master Supervisor Engine synchronizes its statuses and data to the slave Supervisor Engine. After the synchronization is complete, the master and slave Supervisor Engines enter the Active/Standby Hot state, and start the heartbeat function. Then, the hot backup relationship between the master and slave Supervisor Engines is completely established.

5. Continued from 3. The master and slave Supervisor Engines start batch synchronization and the master Supervisor Engine synchronizes some statuses and data to the slave Supervisor Engine. After the synchronization is complete, the master and slave Supervisor Engines enter the Active/Boot Hot state, and start the heartbeat function. Then, the cold backup relationship between the master and slave Supervisor Engines is completely established.

6. Continued from 3. When the master and slave Supervisor Engines detect software version inconsistency during the three-way handshake, the hot backup process does not proceed and the Supervisor Engines are in the Boot Cold/Boot Cold state. This state is different from the Singleton/None state because the two Supervisor Engines detect the existence of each other. After the automatic upgrade function upgrades the slave Supervisor Engine to the version same as that of the master Supervisor Engine and resets the slave Supervisor Engine, the two Supervisor Engines begin to establish the hot backup relationship again.

The following figure shows the batch backup process.

**Figure 6 Batch Backup Process**


Master Supervisor Engine
Slave Supervisor Engine
Register a hot backup processing function and perform initialization
Register a hot backup processing function and perform initialization
Establish a connection
Batch synchronization request
Batch synchronization ACK
Check information (such as versions)
Back up data in batches
End of batch synchronization
Complete batch synchronization and enable heartbeats
Back up data in real time/periodically…

The preceding describes the normal batch backup process. If an exception occurs (for example, a connection is broken before synchronization is complete, the CPLD is lost, or the batch backup connection fails to be established), the master and slave Supervisor Engines attempt to recover by means of retry or the reset of the Supervisor Engines or the system. The processing of some exceptions will be described in "Switchover and Hot Swapping".

## • Switchover and Hot Swapping

When the master Supervisor Engine becomes faulty, the slave Supervisor Engine takes over the work of the master Supervisor Engine after it detects the fault or a command is executed manually. This process is called switchover. After a switchover, the original slave Supervisor Engine takes the place of the original master Supervisor Engine to control the system, and the statuses and configuration of the new master Supervisor Engine are consistent with those of the original master Supervisor Engine. This basically ensures that services are uninterrupted. The switchover process includes a series of steps, including status clearing, heartbeat disabling, configuration refresh, and route convergence.

Hot swapping refers that a Supervisor Engine is inserted or removed during device running with no need to shut down the device. All Supervisor Engines and line cards supporting hot backup support hot swapping. The configuration data can be completely recovered after the original card is re-inserted, no configuration data will be lost, the device does not need to be restarted, and services are not affected.

Automatic detection in the RGOS hot backup solution mainly utilizes the underlying hot swapping detection notifications and heartbeat signals (including CPLD signals), to determine whether reset and switchover are required for recovery. In general, there are the following hot swapping cases (the detected hot swapping and heartbeat signal loss are collectively referred to as peer loss):

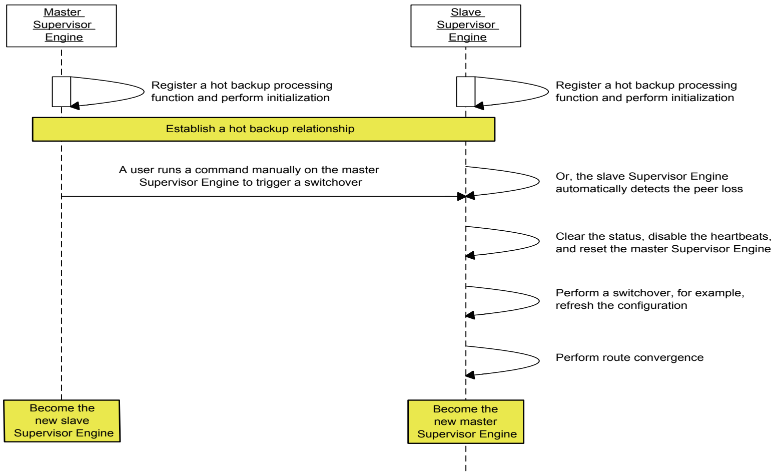
- When detecting the loss of the slave Supervisor Engine, the master Supervisor Engine clears the statuses, disables the heartbeat function, and enters the Singleton state.

- When detecting the loss of the master Supervisor Engine before batch backup is completed, the slave Supervisor Engine clears the statuses, disables the heartbeat function, and resets the system (including all line cards) to ensure configuration consistency.

- When detecting the loss of the master Supervisor Engine after batch backup is complete, the slave Supervisor Engine clears the status, disables the heartbeat function, and performs a switchover to take over the work of the master Supervisor Engine to ensure that data flows are uninterrupted.

The following figure shows the switchover process.

**Figure 7 Switchover Process**


Master Supervisor Engine
Slave Supervisor Engine
Register a hot backup processing function and perform initialization
Register a hot backup processing function and perform initialization
Establish a hot backup relationship
A user runs a command manually on the master Supervisor Engine to trigger a switchover
Or, the slave Supervisor Engine automatically detects the peer loss
Clear the status, disable the heartbeats, and reset the master Supervisor Engine
Perform a switchover, for example, refresh the configuration
Perform route convergence
Become the new slave Supervisor Engine
Become the new master Supervisor Engine

### • Heartbeat Detection

The master and slave Supervisor Engines adopt the heartbeat method to maintain their connection. Heartbeat refers that the master and slave Supervisor Engines send communication signals to each other at intervals to report their current statuses. Once the heartbeat signal indicates that one peer is faulty or one peer fails to receive the heartbeat signal from the other peer within the specified time, the system software considers that a fault occurs and processes the fault as described in "Switchover and Hot Swapping" above. If the heartbeat loss occurs after the master and slave Supervisor Engines have established a hot backup relationship and the slave Supervisor Engine detects the loss of the master Supervisor Engine, the master Supervisor Engine stops working and the slave Supervisor Engine takes over the work of the master Supervisor Engine. In this case, the statuses and data of the master Supervisor Engine have been synchronized to the slave Supervisor Engine. Therefore, the slave Supervisor Engine can become the new master Supervisor Engine smoothly and network services are uninterrupted.

Ruijie heartbeat function involves both hardware and software. When one Supervisor Engine detects that the heartbeat signal sent from the peer times out, it considers that the peer is lost and processes the case as described above.

### • Configuration File Synchronization

A major portion of the status and data synchronization between the master and slave Supervisor Engines is the configuration file synchronization. Configuration files mainly include the startup configuration file (startup-config) and real-time configuration file (running-config), which are collected from entities based on their configuration. The two files share the same syntax and semantics. Users can run the write command to save the real-time configuration file as the startup configuration file. A custom file transfer protocol (ITFTP, akin to TFTP) is adopted to transfer configuration files between the master and slave Supervisor Engines due to a large size of the configuration files. A small amount of other information is synchronized via checkpoint.

The following events will trigger the synchronization of configuration files between the master and slave Supervisor Engines:

- The system switches from the global mode to the privileged mode.

- A user saves the startup configuration file.

- Configuration files are synchronized periodically when users enter commands but do not perform the preceding two operations.

Theoretically, the real-time configuration file can be synchronized in incremental backup mode to reduce the backup overheads. However, there are great difficulties in the design of the incremental backup algorithm, and the full backup method is adopted currently, that is, the complete real-time configuration file is synchronized every time.

## Networking Application of Hot Backup

Ruijie hot backup function is mainly applied to core switches. These devices play very important roles on networks. They can be used in the core of some backbone networks or LANs, and serve thousands of users. If these devices malfunction, services are severely affected, causing huge losses. The hot backup technology in response to such failures can effectively reduce the adverse impact and protect customers' interests.

# Conclusion

Hot backup provides users with higher reliability and stability and can bring real values to users: It can maintain the network stability and reliability to reduce maintenance expenses and costs, reduce the losses caused by failures, help improve users' work efficiency, and help users build a good image and reputation. These are all conducive to users' success.

Ruijie Networks Co.,Ltd

For further information,please visit our website **http://www.ruijienetworks.com**