

SRDF/Asynchronous: A Technical Description

DRAFT

Abstract

This paper provides a technical overview of SRDF/Asynchronous (SRDF/A), along with key features and options such as SRDF/Star, and how it works with Symmetrix DMX systems. This paper is not intended to provide a general overview of SRDF/A.

Published 01/04

Copyright © 2004 EMC Corporation. All rights reserved.

EMC believes the information in this publication is accurate as of its publication date. The information is subject to change without notice.

THE INFORMATION IN THIS PUBLICATION IS PROVIDED “AS IS.” EMC CORPORATION MAKES NO REPRESENTATIONS OR WARRANTIES OF ANY KIND WITH RESPECT TO THE INFORMATION IN THIS PUBLICATION, AND SPECIFICALLY DISCLAIMS IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Use, copying, and distribution of any EMC software described in this publication requires an applicable software license.

EMC², EMC, Symmetrix, Celerra, CLARiiON, CLARAlert, DG, E-Infostructure, HighRoad, Navisphere, PowerPath, ResourcePak, SnapView/IP, SRDF, VisualSAN, The EMC Effect, and where information lives are registered trademarks and EMC Automated Networked Storage, EMC ControlCenter, EMC Developers Program, EMC Enterprise Storage, EMC Enterprise Storage Network, EMC OnCourse, EMC Proven, EMC Snap, Access Logix, AutoAdvice, Automated Resource Manager, AutoSwap, AVALONidm, C-Clip, CacheStorm, Celerra Replicator, Centera, CentraStar, CLARevent, Connectrix, CopyCross, CopyPoint, CrosStor, Direct Matrix, Direct Matrix Architecture, EDM, E-Lab, Engenuity, FarPoint, FLARE, InfoMover, MirrorView, NetWin, OnAlert, OpenScale, Powerlink, PowerVolume, RepliCare, SafeLine, SAN Architect, SAN Copy, SAN Manager, SDMS, SnapSure, SnapView, StorageScope, SupportMate, SymmAPI, SymmEnabler, Symmetrix DMX, TimeFinder, Universal Data Tone, and VisualSRM are trademarks of EMC Corporation. All other trademarks used herein are the property of their respective owners.

Part Number C1058.2

Table of Contents

SRDF Family of Remote Replication Solutions	<u>4</u>	Deleted: 2
SRDF/Asynchronous.....	<u>4</u>	Deleted: 2
Purpose and Scope.....	<u>6</u>	Deleted: 2
SRDF/A Product Description.....	<u>6</u>	Deleted: 2
Ensuring Data Consistency – The Dependent Write Principle.....	<u>6</u>	Deleted: 2
How Does SRDF/A Work?	<u>7</u>	Deleted: 2
Symmetrix Dependent Write Processing	<u>7</u>	Deleted: 2
Dependent Write Consistency	<u>8</u>	Deleted: 2
SRDF/A Data Flow Description.....	<u>8</u>	Deleted: 2
Ensuring Data Consistency—TheDelta Set Switch Process	<u>10</u>	Deleted: 2
SRDF/Consistency Groups’ Multi-Session Controls: Operational Overview	<u>10</u>	Deleted: 2
Entering SRDF/CG for SRDF/A Multi-Session Mode	<u>11</u>	Deleted: 2
SRDF/A States.....	<u>11</u>	Deleted: 2
SRDF/A Mode Transitions and Recovery	<u>12</u>	Deleted: 2
NR State (System Startup).....	<u>12</u>	Deleted: 2
Inactive State	<u>12</u>	Deleted: 2
Switching to SRDF/Asynchronous Mode	<u>12</u>	Deleted: 2
Transition from Adaptive Copy WP Mode to SRDF/A.....	<u>12</u>	Deleted: 2
Transition from Adaptive Copy Disk Mode to SRDF/A.....	<u>13</u>	Deleted: 2
Active State: System Working in Asynchronous Mode, and R2 is Consistent.....	<u>13</u>	Deleted: 2
Dropping Out of Active State	<u>13</u>	Deleted: 2
Deactivating SRDF/A.....	<u>13</u>	Deleted: 2
SRDF/A Recovery Scenarios.....	<u>13</u>	Deleted: 2
Temporary Link Loss.....	<u>19</u>	Deleted: 2
Permanent Link Loss	<u>19</u>	Deleted: 2
Primary Symmetrix Global Memory Full Condition	<u>20</u>	Deleted: 2
Failback from R2	<u>20</u>	Deleted: 2
Summary	<u>20</u>	Deleted: 2

SRDF Family of Remote Replication Solutions

Providing continuous business operations can be a daunting task. You may be trying to gain a certain level of certification—dictated by the Securities and Exchange Commission—or specific service level agreements (SLAs) that you offer your customers. It may be that you simply want to limit your exposure to planned and unplanned downtime. Or perhaps you simply need to provide your organization with flexible and efficient data replication. No matter what the challenge is, there is one underlying theme: data protection and faster business restart in the event of a disaster or unplanned outage is critical across the organization.

The SRDF® family of remote replication solutions can help. SRDF-based solutions will help you manage and conquer your current business challenges while enabling new processes and procedures that will help you gain a significant competitive advantage—something all businesses strive for in today's competitive marketplace.

You will notice that there are several key areas that must be understood with each approach to remote replication, be it synchronous, asynchronous, or some sort of adaptive copy approach:

The response time impact to your production application. For example, with any synchronous solution, your application must wait for the acknowledgement from the remote system before proceeding with the next dependant write. You also have speed-of-light issues which impact the maximum distance that your locations can be from each other.

The infrastructure—What additional equipment is required to support the remote replication process?

The communication links—How big and how expensive will the communication links need to be to support the process?

And most importantly, **what is the recovery point at the target?** That is, how much data exposure will you experience as part of the operation: none, seconds, minutes, hours?

The SRDF family of remote replication solutions can help you balance these “pain points”. The SRDF family consists of three base solutions – SRDF/Synchronous, SRDF/Asynchronous, and SRDF/Data Mobility. There are a number of additional options and features that can be added to the base solutions to solve specific service level requirements. These options include: SRDF/Consistency Groups for data consistency, SRDF/Cluster Enabler for integration with host-based clustering products such as MSCS, VCS, SRDF/Star for advanced multi-site failover with continuous protection, a Mode Change option to dynamically switch between SRDF/A and SRDF/S and advanced SRDF/Automated Replication solutions to meet very specific remote replication service level requirements.

SRDF/Asynchronous

SRDF/Asynchronous (SRDF/A) is a remote mirroring solution for the Symmetrix® DMX series. Its innovative Delta Set architecture delivers a remote replication solution that has no impact on production applications and no distance limitations. This architecture also enables significant operational savings through reduced bandwidth requirements; SRDF/A only needs enough bandwidth to support the average production workload vs. peak workloads.

SRDF/A is a single solution supporting both mainframe and open systems attaches. It also complements SRDF solutions to meet mixed service level requirements; in fact, it can also share the same communication links as SRDF/S.

SRDF/Asynchronous has some every significant advantages over other, traditional asynchronous ordered write approaches:

Performance—With SRDF/A there is no host application impact regardless of the distance between sites. Also, there are no side file puncture issues with EMC's Delta Set architecture.

Availability—SRDF/A provides a consistent, restartable image at the target side at all times and enables fast and efficient incremental failback, unlike other vendors who have to perform a full offline re-sync to return home.

Functionality—A unique capability of SRDF/A is its ability to share the exact same Symmetrix DMX™, extension equipment, and communication links as SRDF/S and/or SRDF/DM. This includes the EMC GigE Director with on-board compression. SRDF/A also works with the complete TimeFinder™ family of local replication solutions.

Economics—With its unique Delta Set architecture, SRDF/A can significantly lower the required bandwidth necessary to perform the mirroring. IT also offers the best TCO available by supporting the native GigE Director vs. having to purchase external extension equipment. And best of all, it offers a single-vendor solution.

SRDF/A Options and Features

Multiple SRDF/A Groups

Enginuity 5x71 expands SRDF/A to allow multiple (up to 64 depending on configuration) SRDF/A groups per Symmetrix array. Implicit in this feature is the ability to define a source for one SRDF/A group and the target to another SRDF/A group in the same Symmetrix DMX. By doing this, you now have the ability to use SRDF/A with bi-directional operations.

Note that you cannot have bi-directional operation within a single SRDF/A group, but you can have it between the same systems in a SRDF/A pair by using multiple SRDF/A groups. All source-to-target data flow operations within a group still remain unidirectional.

Concurrent SRDF/A and SRDF/S from a single source volume

This feature provides the ability to replicate a group of devices in Synchronous mode to one target site and in Asynchronous mode to a secondary extended distance site. The same source device can be replicated synchronously using SRDF/Synchronous mode down one link and asynchronously using SRDF/Asynchronous mode down another link.

This capability provides a limited multi-site protection capability. In the event of a primary site failure, there would be no data loss since it is synchronously replicated to a local/regional recovery site. In the event of a regional disruption, data has been simultaneously replicated to a recovery site much further away.

The limitation to this solution is that in the event of a loss of the primary site, establishing remote replication between the remaining sites would mean configuration changes and a full synchronization process. To avoid this limitation, EMC has developed SRDF/Star (see below).

SRDF/Star

SRDF/Star provides advanced multi-site business continuity protection. It enables concurrent SRDF/S and SRDF/A operations from the same source volumes with the ability to incrementally establish an SRDF/A session between the two remote sites in the event of a primary site outage—a capability only available through SRDF/Star software.

This capability takes the promise of concurrent synchronous and asynchronous operations (from the same source device) to its logical conclusion. SRDF/Star allows you to quickly re-establish protection between the two remote sites in the event of a primary site failure, and then just as quickly restore the primary site when conditions permit.

With SRDF/Star, enterprises can quickly resynchronize the SRDF/S and SRDF/A copies by replicating only the differences between the sessions—allowing for much faster resumption of protected services after a source site failure.

A complete description of SRDF/Star can be found below.

SRDF/A to SRDF/S Mode Change Feature

This feature allows the capability to switch to and from SRDF/S to SRDF/A while ensuring dependent write consistency at the R2 side throughout the process. This capability offers those currently using SRDF/S the option of switching into SRDF/A during high workload periods to minimize performance impacts to applications. It also allows switching in reverse from SRDF/A to SRDF/S (for example, to catch up tracks owed) before then reverting to SRDF/A again.

Purpose and Scope

This document provides an in-depth explanation of how SRDF/A and SRDF/Star operate. It is not intended as a general overview of either SRDF/S or SRDF/A.

For more general information on the SRDF family of products and more in-depth discussions of SRDF/S and SRDF/DM, please visit our website at www.EMC.com.

SRDF/A Product Description

Beginning with Solutions Enabler version 5.3 and Enginuity version 5670, Symmetrix DMX systems support the new asynchronous replication product named SRDF/Asynchronous (SRDF/A). Asynchronous mode provides a point-in-time image on the target (R2) device that is only slightly behind the source (R1) device. SRDF/A session data is transferred to the remote Symmetrix system in Delta Sets, eliminating the redundancy of same-track changes being transferred over the link, reducing the required bandwidth.

SRDF/A provides a long-distance replication solution with no impact on performance. This level of protection is intended for customers who require no host application impact while maintaining a consistent, restartable image of their data on the R2 side at all times. In the event of a disaster at the R1 site, or if RDF links are lost during data transfer, a partial Delta Set of data can be discarded, preserving consistency on the R2 with a data loss of no more than two SRDF/A cycles.

Ensuring Data Consistency – The Dependent Write Principle

Any asynchronous remote replication solution, host-based or storage-controller-based, must ensure that data at the remote site is always consistent in order to enable restartability and maintain data integrity.

To achieve this objective, asynchronous replication solutions must honor the logical dependency of write I/Os that are embedded in the I/O stream by the logic of an application, operating system, or database management system. This ensures *dependent write consistency* and is the basic principle behind all of EMC's consistency technology solutions. Understanding this concept is key to understanding how SRDF/A is able to provide consistent data at the remote site at all times.

The definition of a dependent write is a good place to begin understanding how SRDF/A is able to ensure data consistency. A dependent write is a write I/O that will not be issued by an application until some prior related write I/O has completed.

A real-world illustration of this concept is the relationship that a Database Management System (DBMS) creates between the I/O that updates a database file and the I/O that writes to the log of the DBMS. No DBMS will ever write

changes to the database itself until after it has written those updates to its log. In this case, the database write is the dependent write and the log write is the predecessor write. Without this dependent write logic, no DBMS could restart with data integrity in the event of a power failure of the CPU on which it was running.

As an example, assume two write I/Os, A and B, have a relationship in the application such that I/O B will not be issued until I/O A has been successfully completed. Ensuring dependent write consistency means that no failure in the remote replication infrastructure can create the condition where I/O B has been replicated to the remote site, but I/O A has not been replicated. Such a missing sequence condition would destroy the integrity of the data at the remote site.

There are several ways to ensure that this does not occur and that consistency of the remote site is preserved. These include:

- Ordering write I/Os based on timestamps
- Ordering write I/Os based on sequence numbers
- Honoring dependent writes

SRDF/A approaches preserving dependent write consistency by honoring the dependent write relationships imbedded in the I/O stream by the application. This makes it an implicitly enterprise-wide, heterogeneous host solution.

The relationship between dependent I/Os is one of logic, not timing. By respecting the logical relationships between I/Os, there is no need for any type of temporal references (timestamps or sequence numbers) to create the consistency. The Symmetrix system has no way of knowing which of the specific I/Os it receives have dependent relationships, so it must assume that any I/O B that arrives after another I/O A has been acknowledged as complete to the host, must in fact be dependent on I/O A. Therefore, the system must ensure that A and B are either in the same Delta Set or that B is in a Delta Set that is processed to the remote side after the Delta Set that contains A is transmitted.

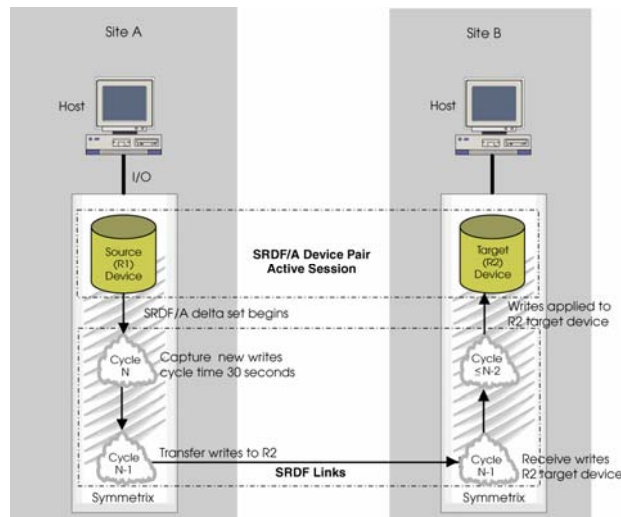
This means that it's actually the Delta Sets, rather than the individual I/Os themselves, that are being ordered by SRDF/A. The key to preserving the consistency of data within SRDF/A is to ensure that dependent write relationships are honored during the Delta Set switch process. The frequency of the switch process determines the data lag experienced by the remote site, for example, the amount of data that would be lost in the event of a primary site disaster.

Note that in the event of a planned failover or shutdown, no data is lost. Data loss exposure only comes into play when the source site suddenly becomes unavailable to the target and cannot be recovered. The amount of this exposure is dependent upon several factors, and is not unique to SRDF/A; all asynchronous replication solutions have this exposure and it is one of the decision criteria in choosing an asynchronous solution over a synchronous solution.

How Does SRDF/A Work?

Symmetrix Dependent Write Processing

Different from ordered write approaches, Symmetrix systems implement asynchronous host writes from the source to the target using predetermined timed Delta Sets. Each Delta Set contains groups of I/Os for processing, which are managed for consistency by the Enginuity operating environment. Using Delta Sets, SRDF/A transfers sets of data, one Delta Set at a time, between the R1 and the R2. If the same data block is written to more than one time within a Capture Delta Set on the R1 side, SRDF/A sends the update over the link only once. This approach lowers the link bandwidth compared with other ordered write processing approaches that transfer each and every



write separately. Refer to Figure 1 for a depiction of the SRDF/A Delta Sets.

Figure 1. SRDF/Asynchronous

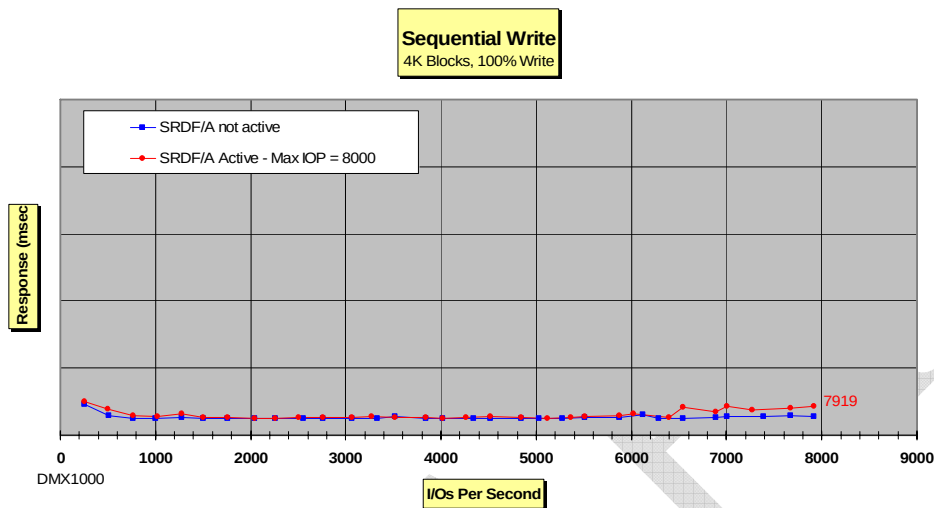


Figure 2. Performance with and without SRDF/A

Dependent Write Consistency

Dependent write consistency is achieved through the processing of four ordered SRDF/A Delta Sets between the source (R1) and the target (R2). Refer to Figure 1, which depicts the four SRDF/A Delta Sets. Dependent write consistency ensures that all writes to the R2 are processed in sequential Delta Sets to maintain a consistent copy of data between the R1 and the R2.

When the first SRDF/A Capture Set is active, it collects any new writes on the R1, overwriting any previously written data blocks intended for data transfer over the link. The Delta Set is active for a predetermined minimum amount of time, which, in future releases can be configured on the Symmetrix system. The default time is 30 seconds. After the minimum time has been reached, the Capture Delta Set data becomes the Transmit Delta Set and begins transferring the data over the link into the R2. A new Capture Set then begins collecting new writes in global memory.

On the R2 side, the Receive Delta Set collects all the data from the Transmit Delta Set. Once the Delta Set has been received in its entirety, it is promoted to an Apply Delta Set and committed to disk (R2). This ensures a consistent, restartable R2 copy with each application of an Apply Delta Set.

One Delta Set is dependent upon the other for achieving write consistency. No Delta Set can begin destaging to disk until the prior one has completed. All data is transferred at the block level.

SRDF/A Data Flow Description

At their most basic levels, continuous asynchronous copy solutions are data flow control problems. Data flows into the system at a certain rate (the channel write I/O bandwidth), where it is buffered (creating an additional global memory requirement) and sent out the inter-controller links (outgoing bandwidth) to the remote controller where it is buffered again (additional global memory) before being written to the remote disk. Keeping these resources in balance is the key to a successful implementation.

A key difference between SRDF/A and other approaches is that less data is transferred and the data is handled fewer times. By exploiting the locality of reference phenomenon, through a process called write folding, within an application, data that is updated multiple times in the same Capture Delta Set is sent across the SRDF/A links only once, and does not have to be duplicated within the global memory of the sending control unit to preserve data consistency. This phenomenon is called *Write Folding* and is illustrated in Figure 3.

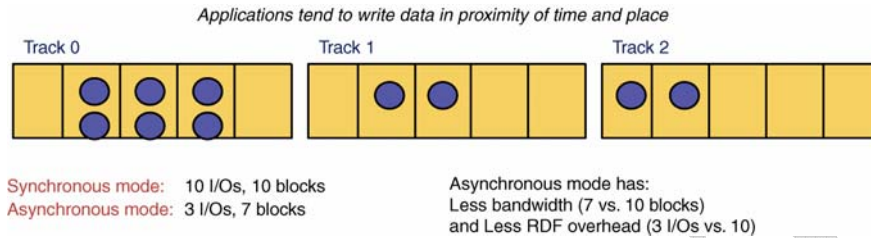


Figure 3. SRDF/A Write Folding Example

The SRDF/A flow control model, at any single point in time, has up to four Delta Sets in motion (note that the Transmit and Receive Delta Sets are the same — with the exception that one is only on the R1 side and the other on the R2), working to achieve the movement of data to the remote site. SRDF/A also handles the data much less than alternative solutions because there is no manipulation of the data or metadata for each update, as is done in ordered-write solutions.

At the R1 site, the Capture Delta Set is collecting all new writes and holding them in cache until the Delta Set switch process promotes them to the Transmit Delta Set. The Transmit Delta Set, which is not receiving any new data, is pushing the data it had collected when it was the Capture Delta Set, to the R2 side. These two Delta Sets switch roles from Capture to Transmit during the Delta Set switch process, which is driven by the Symmetrix Remote Adapters (RAs) and is initiated by the R1 side. These Delta Sets are physically implemented as a collection of pointers to global memory slots.

At the remote site, there is also an inactive Receive Delta Set, which is receiving data from its partner Transmit Delta Set in the R1 site and placing it in global memory. The Apply Delta Set at the R2 site is draining the data from global memory from a previous Delta Set that had been received in its entirety, and therefore represents a consistent image of the data. It does this by simply marking all tracks as write pending to the R2s, and then lets the normal disk adapter (DA) algorithms de-stage the data to the disk for maximum performance.

The switching between Delta Sets on the R2 side is driven by the R1 side issuing a *commit* message to the RAs on the R2 side. This happens once the minimal Delta Set time has elapsed and it has finished sending all the data in its Transmit Delta Set.

Once the Delta Set switch is complete on the R2 side, a message is sent back to the R1 side completing the loop and enabling the R1 side to perform its Delta Set switch if all conditions for this switch are met.

The frequency of the Delta Set switches can be set to between 5 and 30 seconds.

Also, it is important to note that the unit of transfer for each update is at the block level, just as in SRDF/S synchronous mode. It is not at the track level, except for resynchronization scenarios where SRDF/A was suspended and tracks have been marked invalid.

Ensuring Data Consistency—The Delta Set Switch Process

SRDF/A ensures data consistency through the Delta Set switch process. In a single Symmetrix SRDF/A implementation, this consistency must be assured across all volumes in the SRDF group. Since SRDF/A is ensuring the consistency of the data, there is no need for the Consistency Group for z/OS product or PowerPath® for open systems to be installed for SRDF/A devices. Up to 64 SRDF/A groups per system can be defined.

The Symmetrix host adapter obtains the Capture Delta Set number from a single location in global memory and assigns it to each I/O at the beginning of the I/O. The I/O retains that Delta Set number even if a Delta Set switch occurs during its life. This ensures dependent write consistency within SRDF/A group of devices.

This results in the Delta Set switch process being *atomic for dependent-write sequences*, even though it is not physically an atomic event across a range of volumes. As a result, two I/Os with a dependent relationship between them will either be in the same Delta Set, or the dependent I/O will be in a subsequent Delta Set.

The Delta Set switch algorithm will **not** allow a dependent I/O *B* to be placed in transmit Delta Set when its predecessor I/O *A* is placed in the new Capture Set. This is because the predecessor I/O *A* will retain its assigned Delta Set number, even if the Capture Delta Set number becomes the Transmit Delta Set during *A*'s execution. This Delta Set number retention will necessarily result in *B* being assigned to the new Capture Set. (This is because *B* is only issued by the application after *A*'s execution is completed. If the Delta Set switched during *A*'s execution, then *B* must be assigned to the new Delta Set.)

The new Capture Set will then be sent to the R2 site only after the previous Transmit Delta Set has been successfully transmitted and a Delta Set switch has occurred, thus preserving dependent write consistency. In no case will *B* ever be transmitted to the remote side ahead of *A*.

Beginning with Enginuity 5670 56.53, SRDF/A is supported in configurations where there are multiple Symmetrix systems on the R1 side. Initially this support is available only for z/OS environments and requires an equal number of Symmetrix systems, each containing a single SRDF/A group, on each side of the configuration.

Achieving data consistency across multiple SRDF/A groups simply requires that the Delta Set switch process described earlier be coordinated among the participating Symmetrix systems, and that the switch occur during a very brief time period when no host writes are being serviced by the subsystems.

Achieving this requires a single coordination point to drive the cycle switch process in all participating Symmetrix systems. This function is provided by a z/OS host running the Symmetrix Control Facility Version 5.4 through an SCF service called SRDF/A Multi-Session Control (MSC) that is managed and controlled by the SRDF Host Component V5.2.1 (or later) on the mainframe, and through Solutions Enabler in Open Systems environments (new in 5x71).. This capability is available in the SRDF/Consistency Groups option.

It should be noted that this design uses system cache to stage and transfer data. Making efficient use of cache instead of first writing the data to a journal file, and then reading it off again is far more efficient on system resources.

SRDF/Consistency Groups' Multi-Session Controls: Operational Overview

From a single Symmetrix perspective, I/O is processed exactly the same way in SRDF/CG multi-session mode for SRDF/A, as in single-session mode. The host writes to the Capture Delta Set on R1 side and SRDF/A transfers the data in the Transmit Delta Set from R1 to R2. The R2 Receive Delta Set collects the data and the Apply Delta Set commits it to disk.

During this process the status and location of the Delta Sets are communicated between the R1 and R2 Symmetrix systems. For example, when the R1 finishes sending the Transmit Delta Set, it sends a special indication to R2 to let it know that it has completed the Delta Set transfer. Similarly, when R2 finishes restoring the Apply Delta Set, it sends a special indication to let the R1 system know that its Apply Delta Set is empty.

At this point in the process, that is, when the Transmit Delta Set on the R1 side is empty and the Apply Delta on the R2 side is empty, SRDF/A is ready for a cycle switch. In single-session mode, SRDF/A itself would perform a cycle switch once all the conditions are satisfied (the default cycle time for multi-session SRDF/A remains 30 seconds).

However, when using SRDF/CG multi-session controls in a mainframe or open systems environment, SRDF/CG indicates its “switch readiness” to the host (mainframe or open systems), which is polling for this condition, and the switch command is issued by the host to all Symmetrix systems once they are in this state.

Entering SRDF/CG for SRDF/A Multi-Session Mode

In order for the host to control the cycle switch process, the Symmetrix systems must be aware that they are running in SRDF/CG multi-session mode. The prerequisite for entering multi-session mode is that SRDF/A must be active on the Symmetrix.

Note that simply activating SRDF/A does not place it in SRDF/CG multi-session mode, and conversely exiting SRDF/CG multi-session mode does not drop or deactivate SRDF/A, it merely places it in normal or “single-session” mode. However, if SRDF/A is dropped or deactivated, then SRDF/CG multi-session mode is necessarily terminated and would need to be re-entered once SRDF/A was made active again.

SRDF/A enters multi-session mode upon receipt of a host command. As part of the command to enter multi-session mode, and with each switch command issued thereafter, the host provides a tag to each active Delta Set that is retained throughout that Delta Set’s life. This tag is a value that is common across all participating Symmetrix systems and eliminates the need to synchronize the Delta Set numbers across them. The tag is the mechanism by which consistency is assured and represents all the Capture Delta Sets in the host-coordinated cycle switch.

SRDF/A States

SRDF/A can be in one of three logical states:

- **Active:** System is running in asynchronous mode.
- **Inactive:** Devices run in their basic modes—sync/semi-sync/adaptive copy write pending or adaptive copy disk mode.
- **NR:** Devices are NR (Not Ready) on the link, no data is moving to the R2 side.

When running SRDF/A in asynchronous mode (active state), the R2 side can be either consistent or inconsistent. SRDF/A will attempt to make the R2 devices consistent as fast as possible. It will declare the R2 consistent once all the updated blocks for remote mirrors of any R1 devices have been transferred to the R2, and the last Delta Set containing any resynchronization data is fully copied to global memory and in the Apply Delta Set on the R2 side. Figure 4 shows the allowed state transitions for SRDF/A.

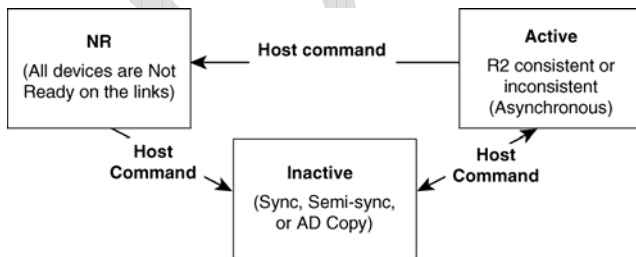


Figure 4. SRDF/A Allowed State Transitions

SRDF/A Mode Transitions and Recovery

This section provides a general overview of how SRDF/A moves between the states discussed in the section, *SRDF/A States*.

For specific open systems commands, refer to the *Symmetrix SRDF CLI Product Guide*. For mainframe commands, refer to the *SRDF Host Component for z/OS Product Guide*.

NR State (System Startup)

When an SRDF/A environment is configured, and the SRDF/A links come up, all SRDF/A volumes are in *not ready* (NR) state by default. This means that all the remote devices on the R1 side are not ready on the communication link. The user can make all devices ready on the link by issuing a command from the host software. By doing so, the user is switching to the inactive SRDF/A state.

Inactive State

In inactive state, devices are ready on the link, SRDF/A is *inactive*, and all devices work in their basic modes (synchronous, semi-synchronous, or Adaptive Copy WP/Disk mode). The user may choose to put SRDF/A in an inactive state using a DEACTIVATE command, but doing so will compromise the consistency of the data on the R2 side unless a point in time copy of the volumes is first established using TimeFinder/Mirror BCVs. The use of BCVs is strongly encouraged to retain a consistent restartable image of the data volumes on the R2 side during periods of resynchronization. BCVs can be established and split in concert with the SRDF/A Delta Set switch process; SRDF/A does not need to first be suspended to perform TimeFinder/Mirror operations.

Users are strongly encouraged to inquire about the ability to maintain consistent restartable images during resynchronization of local and remote volumes for any asynchronous replication solution they may have deployed – especially while the replication sessions are active.

Switching to SRDF/Asynchronous Mode

The user can ask the R1 Symmetrix to switch to SRDF/A mode.

SRDF verifies that all the devices are ready, and then moves the system to active state (on both Symmetrix systems). As a result, Delta Sets are established on the R1 and R2 sides, and the SRDF/A mechanism is enabled.

If the devices were in sync and running SRDF/A base mode prior to activating SRDF/A, the R2 side is by definition consistent.

If there were invalid tracks to be copied, there is no consistency at the R2 side until the last invalid track has been sent to the remote side by way of the Delta Set switch mechanism and is in the Apply Delta Set. In fact, devices operating in an SRDF/A base mode such as sync, semi-synchronous, or Adaptive Copy mode will automatically switch to asynchronous mode when SRDF/A is made active. Any outstanding write-pending tracks or invalid tracks, marked as a result of Adaptive Copy mode skew values, will be merged into SRDF/A Delta Sets over time (similar to how invalid tracks are normally sent to the R2 side).

Once the Delta Set containing the last invalid track is fully on the R2 side and in the Apply Delta Set, SRDF/A will indicate that the R2 is consistent. The user can determine the consistency state of the R2 side at any time using a host software query command.

Transition from Adaptive Copy WP Mode to SRDF/A

When devices are in SRDF/A's adaptive copy write-pending base mode, it is helpful to allow them to transition into asynchronous mode without moving first to a synchronous base mode. When SRDF/A is activated, all adaptive copy write pending devices are moved into the adaptive copy pending off mode. With SRDF/A active, when the DA scans a device in the pending off mode, rather than creating a separate SRDF queue record, it adds the slot to the Capture Delta Set. When there are no slots left write pending to the mirror that are not in an SRDF/A Delta Set, the device can

transition out of the pending off mode. Once out of pending off mode, two Delta Set switches are required for the R2 side to be consistent.

Transition from Adaptive Copy Disk Mode to SRDF/A

Transition into SRDF/A mode from Adaptive Copy Disk mode is immediate. Invalid tracks “owed” to the R2 device as a result of Adaptive Copy Disk skew are scheduled as re-sync operations. SRDF/A will produce a consistent image of data at the R2 after all re-sync operations are complete and two Delta Set switches have occurred.

A direct transition to Adaptive Copy write pending or disk mode, while SRDF/A is active, is not allowed. See the section,

Deactivating SRDF/A

Formatted: Font: *Italic*

Deleted: *Deactivating SRDF/A*

Active State: System Working in Asynchronous Mode, and R2 is Consistent

The active state is the normal running state of SRDF/A. The R2 always represents a dependent write consistent image of the data.

Dropping Out of Active State

SRDF/A supports several methods of dropping out of active state into not ready state. One option, called DROP, puts the devices in NR state immediately. This results in tracks being converted to invalids on both the R1 and R2 sides of the relationship. Resuming SRDF/A would then require resolving the invalids in the normal way, via SRDF Host Component for z/OS *REFRESH* and *RFR-RSUM* commands or Solutions Enabler *symrdf establish* commands. (Remember that dropping out of SRDF/A asynchronous mode does not compromise the consistency of the data on the R2. SRDF/A always provides a consistent image of data at the remote site at all times. It is only during a resynchronization activity that data consistency will be compromised, but this can be avoided by using TimeFinder/Mirror to preserve a copy of the R2 prior to the resynchronization.)

Another option, called PEND-DROP, puts the devices in NR state only at the end of the current in-process Delta Set. Write pending blocks in the Transmit Delta Set will be converted to invalids on the R1 side only. By dropping on a Delta Set boundary, there will be no need to first resolve R2 invalids upon resuming SRDF/A.

Deactivating SRDF/A

SRDF/A also offers the option of moving out of active state, but leaving the devices ready on the communication link, and by default, in their last primary mode of operation (synchronous or semi-synchronous). Since the devices are left ready on the link, data will continue to flow and the consistency of the data at the R2 side will be compromised.

The actual command syntax to control SRDF/A will be different between the SRDF Host Component for z/OS and the Solutions Enabler *symcli* commands for open systems (as is true today for normal SRDF control), but the behavior and functionality offered will be the same. For specific open systems commands, refer to the *Symmetrix SRDF CLI Product Guide*. For mainframe commands, refer to the *SRDF Host Component for z/OS Product Guide*.

Concurrent SRDF/A and SRDF/S from a single source volume

Concurrent operations from a single source device, one leg running SRDF/S to a local/regional site and the other running SRDF/A to a far remote site, is new with the Engenuity 5x71 release. Each R2 at the remote site requires a symmetrix mirror position but does not otherwise impose system or configuration limitations on the primary Symmetrix DMX array over and above what would normally be required for SRDF/S and SRDF/A operations. Performance will be generally comparable to a conventional SRDF/S configuration; the addition of the SRDF/A leg does not impose additional performance restrictions.

Disruption of either remote location will not effect operations of the remaining replication operations, and full SRDF Family capabilities are supported for each leg.

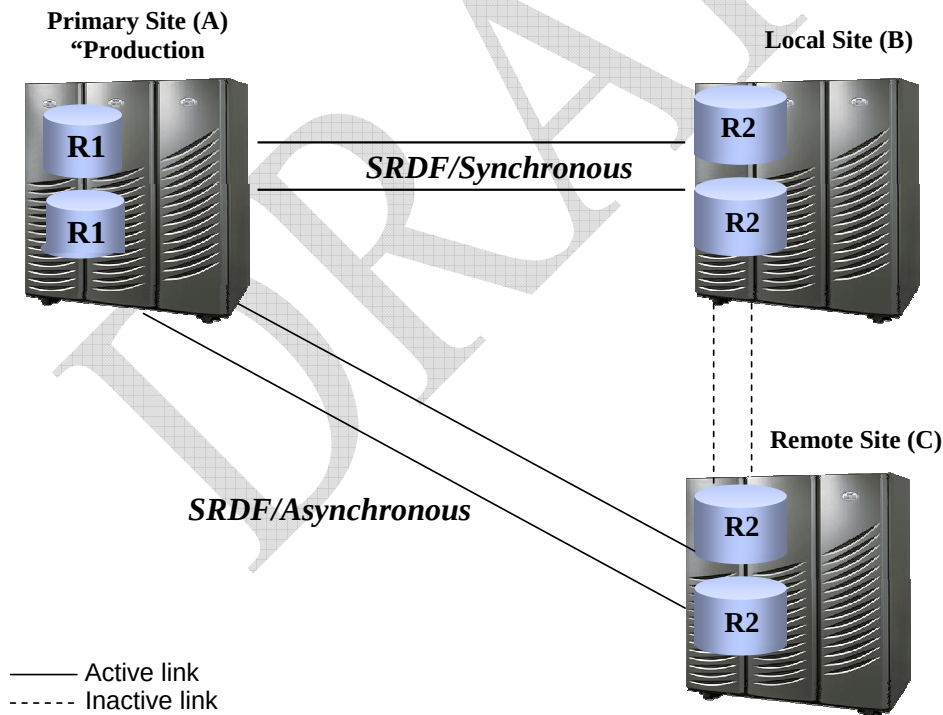
No support is provided to maintain a third replication link between the two remote arrays unless SRDF/Star is deployed. In the event of a primary site failure, an SRDF/A link would have to be manually configured between the two remote locations and a full synchronization would be required.

SRDF/Star Operations

SRDF/Star Provides advanced multi-site business continuity protection. It enables concurrent SRDF/S and SRDF/A operations from the same source volumes with the ability to incrementally establish an SRDF/A session between the two remote sites in the event of a primary site outage—a capability only available through SRDF/Star software.

SRDF/Star is a combination of host software (mainframe only in the initial release, with open systems support scheduled at a later time) and Enginuity functionality that operates in a concurrent SRDF configuration ($A \rightarrow B$ and $A \rightarrow C$) where one remote mirror operates in SRDF/S mode ($A \rightarrow B$) and the other remote mirror operates in SRDF/A mode ($A \rightarrow C$).

SRDF/Star provides a very rapid re-establishment of cross-site protection in the event of Primary Site (A) failure. Rather than a full re-synchronization between sites B and C, SRDF/Star provides a differential B-C synchronization, dramatically reducing the time it takes to remotely protect the new production site. SRDF/Star also provides a mechanism for the user to determine which site (B or C) has the most current data in the event of a rolling disaster that affects site A. In all cases the choice of which site to use in the event of a failure is left to the discretion of the customer.



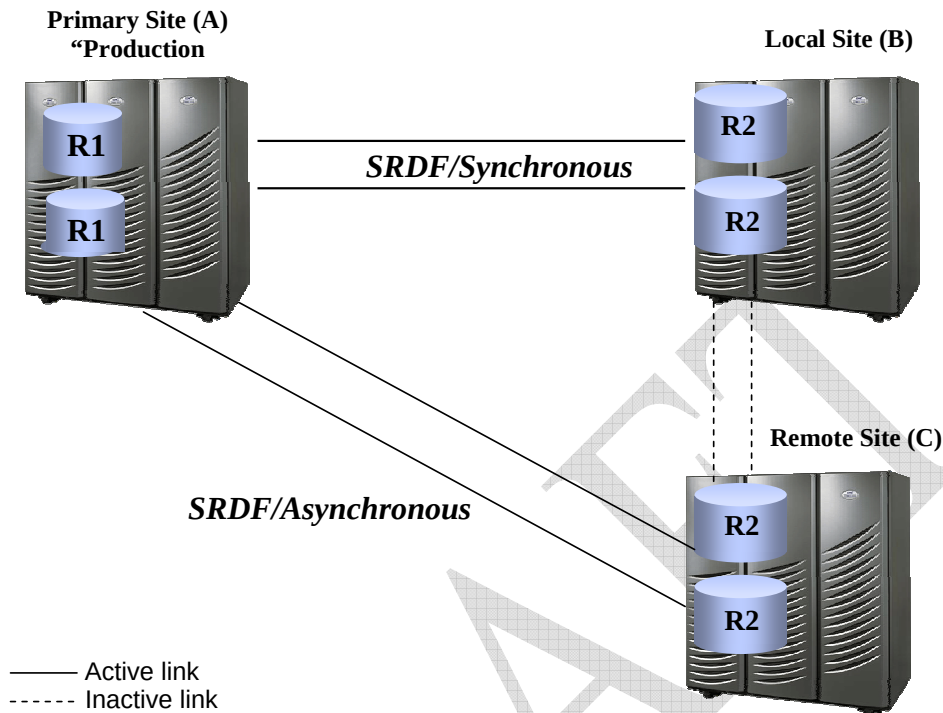


Figure 2 Concurrent SRDF/A and SRDF/S Configuration configured for SRDF/Star support

Note that a mainframe host can control and coordinate an SRDF/Star configuration in an Open Systems only or mixed environment

Problem Statement and Solution Overview

The concurrent configuration option of SRDF/A (figure 1) offers customers the ability to restart their environments at long distances with minimal data loss, while simultaneously providing a zero data loss restart capability at a local site. Such a configuration provides protection for both a site disaster and a regional disaster while minimizing performance impact and loss of data.

Customers who invest in a three data center concurrent replication solution wish to ensure that their regional disaster protection is always enabled, even if the primary site (which is the source of all remote replication) fails. In a concurrent SRDF/A configuration without the SRDF/Star functionality, the loss of the primary 'A' site would normally mean that the long distance replication would stop and data would no longer propagate to the 'C' site.

Data at 'C' would continue to age as production was resumed at site 'B'. Resuming SRDF/A between sites B and C would require a full resynchronization to re-enable disaster recovery protection. This is both a time and resource consuming process. SRDF/Star avoids this full resynchronization by allowing a B→C (or C→B) resynchronization to be done differentially using host software commands and procedures.

In addition, SRDF/Star allows the SRDF personalities of sites A and B to be swapped and the SRDF/A relationship to be transferred to sites B and C during planned outages with no data movement at all. (A 'NOCOPY' option has been added to dynamic SRDF host software commands to allow SRDF relationships

to be created between sites B and C, and declared fully synchronized by the user so SRDF/A can be resumed without a synchronization process of any kind.)

It is important to note that SRDF/Star provides the capability to perform differential resynchronizations between sites B and C, to 'close the triangle' created by a concurrent SRDF/S and SRDF/A configuration. SRDF/Star by itself does not do these resynchronizations automatically. If 'automatic' resynchronization is desired in the event of a site A failure, this must be driven by host based automation software provided by the user. EMC will provide and support (if unaltered by the user) automation to perform the necessary functions at the remote sites to: 1) achieve a differential resynchronization, 2) resume SRDF/A protection and 3) present a restartable image of data to the user.

SRDF/Star Operational Description

Normal Operation

SRDF/Star makes use of the Symmetrix Differential Data Facility (SDDF) of Enginuity to achieve the differential resynchronization between sites B and C. These sessions are created and managed in the B and C site Symmetrix systems by the SRDF/A Multi Session Control software running on a server/host at site A. When configured for SRDF/Star, the MSC software is required even when running a single SRDF/A session. In addition Consistency Group software is required at the primary site regardless of the number of Symmetrix systems installed.

Note: SRDF/Star is an enterprise solution, meaning that an MSC task can support a SRDF/Star configuration containing both FBA and CKD devices (FICON or ESCON connectivity to each Symmetrix is required). Initially only the MSC task running on z/OS will support SRDF/Star, but support by the open systems MSC task is planned.

Each R2 device in the B site Symmetrix has two SDDF session bitmaps associated with it. (see figure 2 below)

The first session is actively tracking which tracks were changed by the R1 side host. (This is possible because the devices are in a synchronous SRDF relationship). This session is called the 'active' session.

The second session is called the 'inactive' session and is used to retain the knowledge of which tracks contain data that needs to be transmitted to the site C Symmetrix via SRDF/A. Periodically, the inactive session is cleared, and the two sessions reverse roles. This clearing of the inactive session will only occur after at least 2 cycle switches have been performed on the SRDF/A leg. At that point, the data tracked by the inactive session will have been received in its entirety and been promoted to the apply cycle at site C. The inactive session can become the 'new' active session and continue to track updated tracks.

The inactive and active SDDF sessions act as SRDF/Star's 'memory' of what tracks are changed at B but not yet in C, and the periodic rotation of these sessions serves to minimize the number of tracks that are moved when a resynchronization between sites B and C is performed.

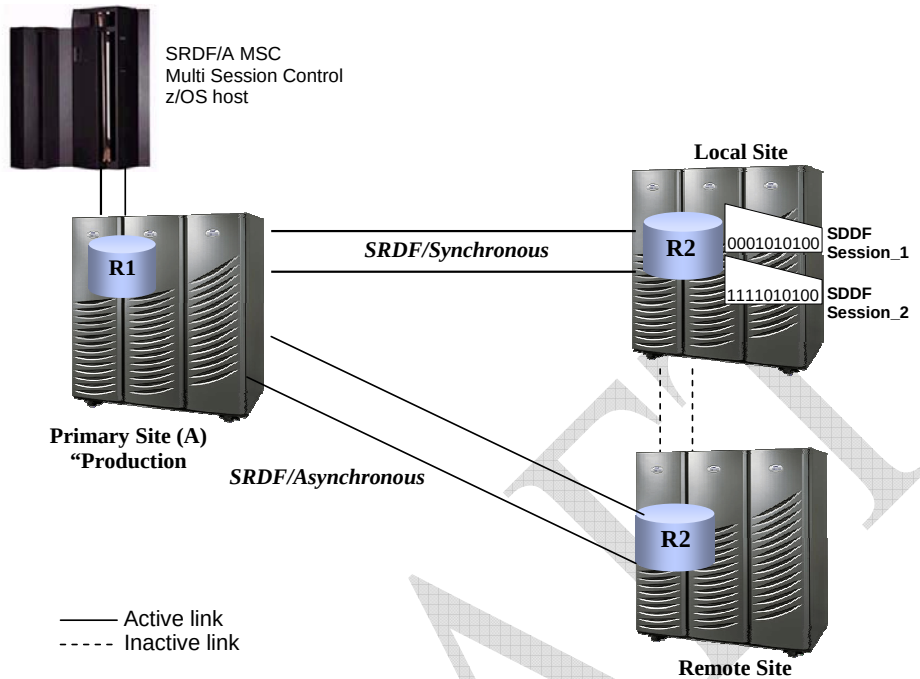


Figure 3 SRDF/Star Normal Operation

Primary Site Failure Operation

When there is a disaster at site A, MSC must be restarted in the hosts at site B or C. A new SRDF relationship must be established using dynamic SRDF between the Symmetrix at site B and the Symmetrix at site C that were both remote partners of the failed R1 Symmetrix. This new relationship is created dynamically without requiring bin file changes.

Once the pairs are created the two SDDF sessions are inclusively OR'd with each other. The result is then OR'd with an SDDF session at C (which may have been activated in certain failure scenarios, such as the loss of the synchronous A→B link). The resulting bitmap represents the data that needs to move during a differential resynchronization, and as such is used to mark R2 tracks invalid in the 'new' R1's track tables in site B. SRDF/A can now be resumed as normal after a drop (a BCV should be split on the site B or C Symmetrix to protect against resynchronization failures) and once the invalid tracks have been cleared and two cycle switches have occurred, consistency is achieved at site C. (Normal indicators for consistency are used to reflect the consistent state.)

This is depicted in figure 3 below:

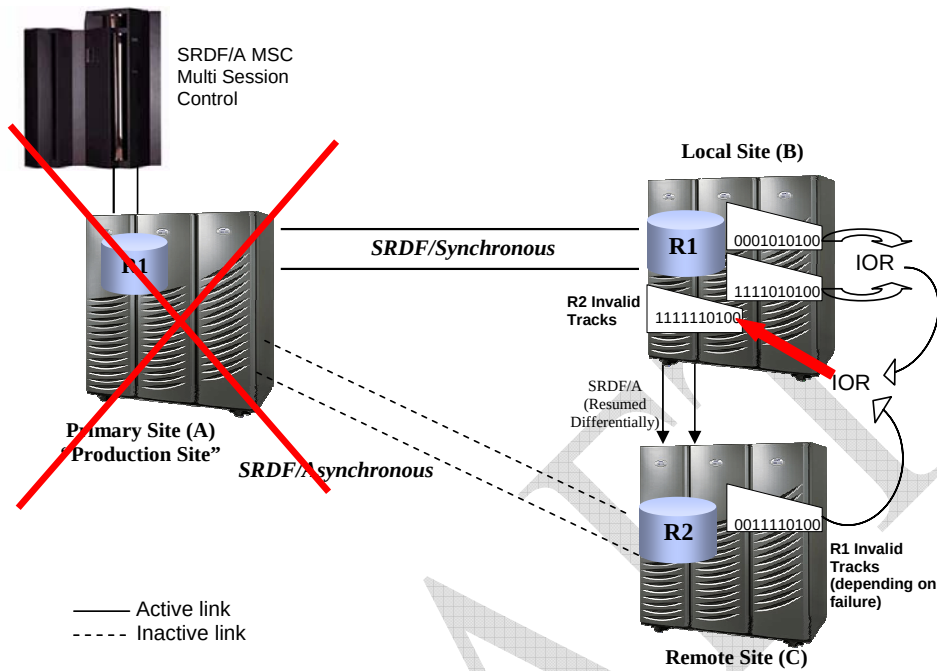


Figure 4 SRDF/Star SRDF/A session failover

Note that the R1 device need not always be at site B. A differential resynchronization is possible in either direction, and SRDF/A could be resumed with the R1 device at site C.

Determining the location of the most current data

SRDF/Star also provides a mechanism for the user to determine which site (B or C) has the most current data in the event of a rolling disaster that affects site A.

The need to do this may not be obvious as it is easy to assume that the synchronous site (site B) will always be more current than the asynchronous site (site C). This is not necessarily the case. Imagine a rolling disaster that first disabled the synchronous link between site A and B. Processing at site A continues for many minutes before failing and SRDF/A performs two or more cycle switches before dropping due to the disaster.

In such a scenario, the data at site C is more current than the data at site B and the customer may choose to restart operations at site C (or resync site C to site B and run at B).

SRDF/Star provides this information to the user by incrementing a token value in the Symmetrix at site C (see figure 4). This token value comes into being upon MSC software being informed of a ConGroup trip on the synchronous R2 devices. Once in existence it is incremented at every cycle switch on the R2 side.

The value of the token counter can be displayed on the C site via a host query command and the user can determine if site C contains more current data than site B. A business decision can then be made as to which site's data will be used for recovery.

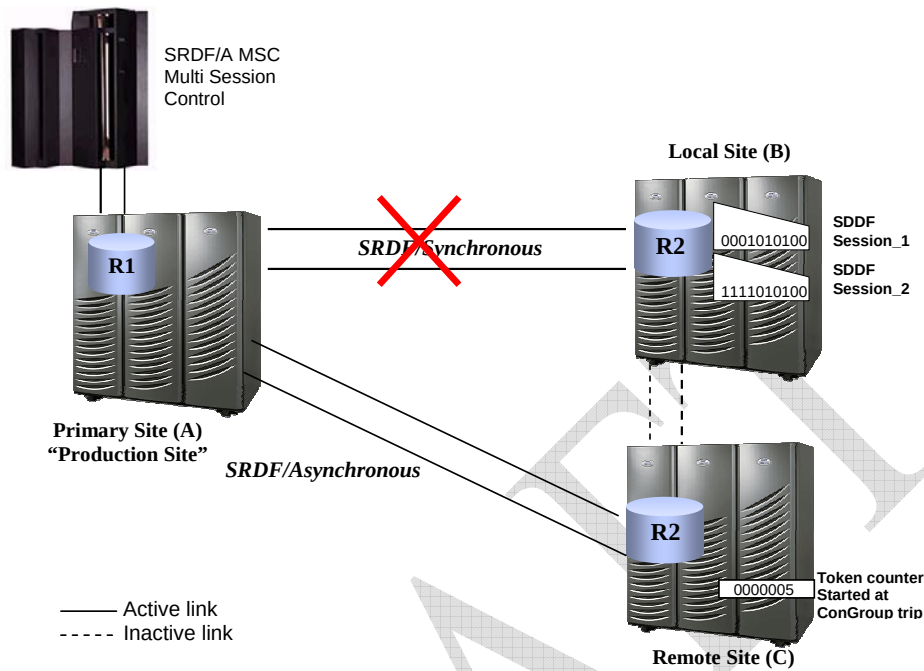


Figure 5 SRDF/Star - most current data determination

SRDF/A Recovery Scenarios

Temporary Link Loss

If SRDF/A suffers a temporary link loss (<10 seconds) of all links, the SRDF/A state will remain active and data will continue to accumulate in global memory. This may result in an elongated Delta Set, but the remote consistency will not be compromised, and the R1-R2 relationships will not be suspended. The amount of time SRDF/A waits until it declares a link loss permanent is configurable (between 0 and 10 seconds).

Permanent Link Loss

If SRDF/A experiences a permanent link loss, it will drop all the devices on the link to not ready (NR) state. This will result in all data in the Capture and Transmit R1 side Delta Sets being changed from write pending for the remote mirror to invalid on the remote mirror. In addition, any new work that enters the Symmetrix will result in tracks being marked invalid. At the R2 side, the Receive Delta Set tracks are marked invalid on the remote mirror, and the Apply Delta Set data completes its commit to local devices.

When the links are restored, normal SRDF/A recovery procedures are followed. The track tables are OR'd using normal host recovery procedures (REFRESH and RFR-RSUM on the SRDF Host Component for z/OS, or *symrdf establish* in Solutions Enabler for open systems). The data is then resynchronized by sending over the invalid tracks as part of the SRDF/A Delta Sets.

The data on the R2 volumes is always consistent in SRDF/A, even when the links have failed. However, the act of starting a resynchronization activity between the R1 and the R2 after a link failure will temporarily compromise the consistency of the R2 data until the resynchronization is fully completed. For this reason, it is recommended that the user employ a method for preserving the data on the R2 volumes prior to commencing resynchronization. This may be done, for example, using TimeFinder/Mirror business continuance volumes.

Primary Symmetrix Global Memory Full Condition

It is possible that an imbalance may occur within SRDF/A between the incoming write I/O workload and the outgoing SRDF/A bandwidth, such that the global memory in the primary Symmetrix becomes full. (The Capture and Transmit Delta Sets consume all the available write memory in the Symmetrix.)

In this situation, the user has a choice as to how SRDF/A will behave:

1. The Symmetrix system can throttle the host at the speed of the links and keep SRDF/A running. In this case, the performance will be equivalent to SRDF/S in synchronous mode.
2. The Symmetrix system throttles the host for a user-specified period of time, and if the condition has not resolved itself at the expiration of this time, then SRDF/A is dropped.
3. SRDF/A will be dropped immediately. The default behavior is to drop SRDF/A immediately.

To avoid a memory full condition, the SRDF/A environment must be properly designed and configured. Factors and variables that may cause an imbalance in the SRDF/A environment may include bandwidth, global memory, and workload allocation for the specific implementations.

Failback from R2

In the event that a disaster occurs on the primary site, the data on the R2 devices represents a dependent write consistent image of data that can be used to restart an environment with minimal data loss. Once the primary site has been repaired, the process for returning to the R1 side uses exactly the same methods as are used for SRDF/S synchronous failback.

Once the workload had been transferred back to the primary site hosts, SRDF/A can be activated, and normal asynchronous mode protection can be resumed.

Summary

Implementing cost-effective extended-distance replication solutions while simultaneously ensuring remote data restartability and local application performance is a key competitive differentiator for your business. The SRDF family of remote replication solutions can help you achieve these goals.

SRDF/A is an innovative approach to asynchronous replication — for both open systems and mainframe environments — that delivers a consistent and restartable remote copy of your production data at all times, over any distance, with no host application impact. Built with the underpinnings of the proven SRDF/S software, SRDF/A is the highest-performance extended-distance solution available today.

SRDF/A can lower your remote replication TCO by reducing the amount of overall bandwidth required compared to traditional ordered write asynchronous approaches. SRDF/A can also utilize native GigE connectivity, further reducing infrastructure costs and overall TCO. SRDF/A is the optimal extended-distance replication solution when your service-level requirements dictate that economics and application performance are more critical than zero data exposure.