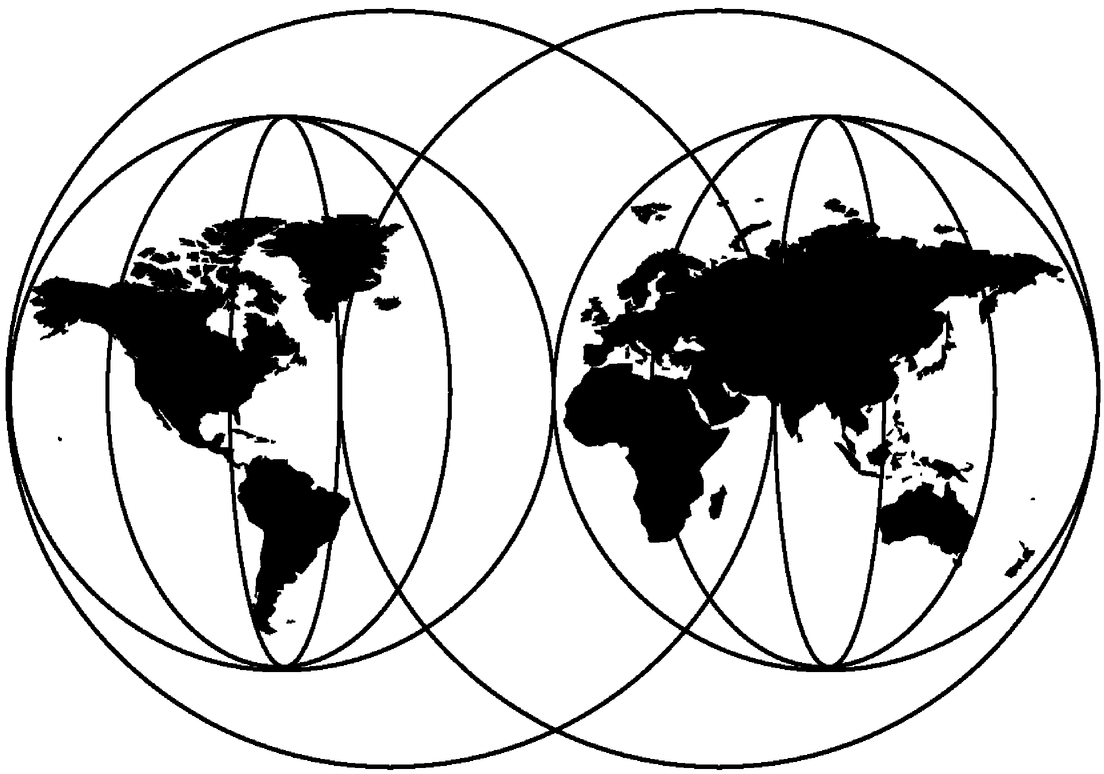
OS/390 Parallel Sysplex Configuration Volume 3: Connectivity

Frank Kyne, Luiz Fadel, Michael Ferguson, Rafael Garcia, Keith George, Masayuki Kojima, Alan Murphy, Henrik Thorsen



International Technical Support Organization

www.redbooks.ibm.com

SG24-5639-00

IBM

International Technical Support Organization

SG24-5639-00

OS/390 Parallel Sysplex Configuration Volume 3: Connectivity

August 2000

Take Note!

Before using this information and the product it supports, be sure to read the general information in Appendix B, "Special Notices" on page 167.

First Edition (August 2000)

This edition applies to

Program Name, Program Number	Version, Release Number
CICS/ESA, 5655-018	4.1
CICS TS for OS/390, 5655-147	1.3
DB2 for OS/390, 5655-DB2	5.1
DB2 UDB for OS/390, 5645-DB2	6.1
DB2 UDB for OS/390, 5675-DB2	7.1
DFSMS/MVS, 5695-DF1	1.5
IMS/ESA, 5695-176	5.1
IMS/ESA, 5655-158	6.1
IMS/ESA, 5655-B01	7.1
System Automation for OS/390, 5645-045	1.3

for use with:

Program Name	Version, Release Number
OS/390, 5645-001	1.3
OS/390, 5647-A01	2.4
OS/390, 5647-A01	2.5
OS/390, 5647-A01	2.6
OS/390, 5647-A01	2.7
OS/390, 5647-A01	2.8
OS/390, 5647-A01	2.9
OS/390, 5647-A01	2.10

Comments may be addressed to:
IBM Corporation, International Technical Support Organization
Dept. HYJ Mail Station P099
2455 South Road
Poughkeepsie, New York 12601-5400

Contents

OS/390 Parallel Sysplex Configuration Volume 2: Cookbook

OS/390 Parallel Sysplex Configuration Volume 3: Connectivity

Figures

OS/390 Parallel Sysplex Configuration Volume 1: Overview

OS/390 Parallel Sysplex Configuration Volume 2: Cookbook

OS/390 Parallel Sysplex Configuration Volume 3: Connectivity

Tables

OS/390 Parallel Sysplex Configuration Volume 1: Overview

OS/390 Parallel Sysplex Configuration Volume 2: Cookbook

OS/390 Parallel Sysplex Configuration Volume 3: Connectivity

Preface

The Parallel Sysplex Configuration redbooks consist of three volumes that help you design and configure a Parallel Sysplex. These redbooks assist you in making the key hardware and software choices necessary to meet your performance, capacity, and availability requirements.

In addition to discussing many sysplex-related topics not documented elsewhere, these books contain, for the IBM large systems customer, practical information that has been collected from many sources and brought together into a single, comprehensive reference.

The redbooks provide rules of thumb for configuring a Parallel Sysplex. They discuss central processing complex sizing, CF sizing, the Sysplex Timer, connectivity guidelines, the implications of Parallel Sysplex for IBM software such as OS/390 V2R8 and related subsystem configurations, and useful tools and services. Network, systems management and availability considerations are also discussed.

These redbooks are a starting point for those involved in designing and configuring a Parallel Sysplex. They also may be used by those already having a Parallel Sysplex when further exploitation is being considered. The books refer to other relevant publications that cover specific areas in more depth.

These books are an update to an existing set of books:

- OS/390 MVS Parallel Sysplex Configuration Volume 1: Overview, SG24-2075
- OS/390 MVS Parallel Sysplex Configuration Volume 2: Cookbook, SG24-2076
- OS/390 MVS Parallel Sysplex Configuration Volume 3: Connectivity, SG24-2077

New information since the last books were published (January 1998) has been added, and information pertaining to back levels of hardware and software has been removed. The old *books* continue to be orderable for customers using those hardware or software levels.

The Team That Wrote These Redbooks

These projects were designed, managed and partly documented by:

Henrik Thorsen	International Technical Support Organization, Poughkeepsie
Frank Kyne	International Technical Support Organization, Poughkeepsie

The authors of these documents are:

Erik Bach	IBM Denmark	**Kuno Kern**	IBM Germany
Chris Barwise	IBM UK	**Masayuki Kojima**	IBM Japan
Frank Byrne	IBM UK	**Frank Kyne**	IBM Ireland
Dave Clitherow	IBM UK	**Geoff Leong**	IBM Canada
Richard Cwiakala	IBM USA	**Jock Marcks**	IBM Australia
Noshir Dhondy	IBM USA	**Dennis Moore**	IBM USA
Luiz Fadel	IBM Brazil	**Alan Murphy**	IBM Ireland
Linda Ferdinand	IBM USA	**Rachel Pickering**	IBM UK
Michael Ferguson	IBM Australia	**Alvaro Salla**	IBM Brazil
Per Fremstad	IBM Norway	**Alain Servais**	IBM France
Rafael Garcia	IBM USA	**Keld Teglgaard**	IBM Denmark
Keith George	IBM UK	**Henrik Thorsen**	IBM Denmark
Dick Jorna	IBM Netherlands	**Ian Waite**	IBM UK
Kazumasa Kawaguchi	IBM Japan		

These publications are the result of a series of residencies conducted at the International Technical Support Organization, Poughkeepsie Center. However, much valuable information and assistance was received from various organizations across IBM. The production of these redbooks is a team effort. The organizations that participated include:

- **Washington System Center, US M&S**
- **S/390 Parallel Center, S/390 Division**
- **S/390 Development, S/390 Division**

Also, thanks to the following people for reviewing these publications, providing material and offering invaluable advice and guidance. Their contributions have made these books eminently more accurate and more readable than they otherwise might have been.

Bill Abbot	Computer Associates
Robert M. Abrams	IBM USA
Riaz Ahmad	IBM USA
Michelle Ahearn	Computer Associates
Jay Aiken	IBM USA
Maria Sueli Almeida	ITSO Almaden
Steve Anania	IBM USA
Jaime Anaya	IBM USA
Stuart Arnell	IBM UK
John Arwe	IBM USA
Alan Ashcroft	IBM UK
Judi Bank	IBM USA
Paola Bari	IBM USA
Harley Beier	IBM USA
Ed Berkel	IBM USA
Evanne Bernardo	IBM USA
Connie Beuselinck	IBM USA
Page Borchetta	IBM USA
Mark Brooks	IBM USA
Jane Brockbank	IBM UK
Paolo Bruni	ITSO Almaden
Fred Bucci	IBM USA
Charlie Burger	IBM USA
Chuck Calio	IBM USA
John Campbell	IBM USA
Walt Caprice	IBM USA
Bernice Casey	IBM USA
Jose Castano	IBM USA
Alfred Christensen	IBM USA
Rich Conway	IBM USA
Roy Cornford	IBM UK
Curt Cotner	IBM USA
Mike Cox	IBM USA
Mary Crisman	IBM USA
Vijay Dalal	IBM USA
Jerry Dearing	IBM USA
Mac Devine	IBM USA
George Dillard	IBM USA
Kristin Donceel	IBM Belgium
Terry Draper	IBM UK
James Doyle	IBM USA
Joann Duguid	IBM USA
Gregory Dunlap	IBM USA
Mary Duval	IBM USA
Greg Dyck	IBM USA
Scott Fagen	IBM USA
Bob Fantom	IBM USA
Erika Fenn	Software AG
Martin Ferrier	IBM UK
John Fitch	IBM USA
Jonathan Follows	IBM ITSO Raleigh
Jim Fyffe	IBM USA
David Fox	IBM USA
Peter Garvin	IBM USA
Rob Geiner	IBM USA
Tony Giaccone	IBM USA
Rusty Godwin	IBM USA
Dionne Graff	IBM USA
Steve Grabartis	IBM USA
Jennifer Green	IBM USA
Susan Greenlee	IBM USA
Bob Haimowitz	IBM USA
Jim Hall	IBM USA
Dick Hannan	IBM USA
Ricardo Haragutchi	IBM Brazil
Johnathan Harter	IBM USA
Evan Haruta	IBM USA
Steve Heinbuch	IBM Canada
Audrey Helffrich	IBM USA
Gary Henricks	IBM USA
Steffen Herzog	IBM Germany
Ken Hewitt	IBM Australia
Juergen Holtz	IBM Germany
Janice Hui	IBM Australia
Patricia Jakubik	IBM USA
Dave Jeffries	IBM France
Curt Jews	IBM USA
Kim Johnson	IBM USA
Paul Johnson	IBM UK
Michael Jordan	IBM USA
Jeff Josten	IBM USA
Ken Kapulka	IBM USA
Gary King	IBM USA
Carl Klitscher	IBM New Zealand
Raimo Korhonen	CompMeas, Finland
Jeff Kubala	IBM USA
Ravi Kumar	IBM USA
Paul Lekkas	IBM Germany
Dave Levish	IBM USA
Rich Z. Lewis	IBM USA
Patrick LiVecchi	IBM USA
Alan Little	IBM USA
Sherman Lockard	IBM USA
Nick Matsakis	IBM USA
John Matush	IBM USA
Geoff Meacock	IBM UK
Kershaw Mehta	IBM USA
Geoff Miller	IBM USA
Roy Moebus	IBM USA
Thomas Murphy	IBM USA
Stephen Nichols	IBM USA
Arthur Neil	IBM UK
Geoff Nicholls	IBM Australia
Jeff Nick	IBM USA
Madeline Nick	IBM USA
Ron Northrup	IBM USA
Marianne Orman	IBM UK
Ron Parrish	IBM USA
Betty Patterson	IBM USA
Anthony Pearson	IBM USA
Kris Perry	IBM USA
James Perlik	IBM USA
David B. Petersen	IBM USA
Silvio Podcameni	IBM Brazil
Paul Powell	IBM USA
Richard Prewitt	IBM USA
David Raften	IBM USA
Tom Rankin	IBM USA
Patrick Rausch	IBM USA
Frank Ricchio	IBM USA
Paul Rogers	IBM USA
Tom Russell	IBM Canada
Eugene Sale	IBM USA
Hany Salem	IBM USA
Chris Saul	IBM USA
Jonathan Scott	IBM UK
Joachim Schmalzried	IBM Germany
Martin Schulman	IBM USA
Tom Shaw	IBM USA
Joseph Shelak	IBM USA
Sim Schindel	IBM France
Carol Shanesy	IBM USA
Beth Somers	IBM USA
Orrin Stevens	Computer Associates
Nat Stevenson	IBM USA
Bill Stillwell	IBM USA
Jeff Stokes	IBM USA
David Surman	IBM USA
John Swift	IBM UK
Lisa Taylor	IBM USA
Clarisse Taaffe-Hedglin	IBM USA
Michael Teuffel	IBM Germany
Douglas Trimboli	IBM USA
Ken Trowell	IBM USA
John Tuccio	IBM USA
Søren Understrup	IBM Denmark
Robert Vaupel	IBM Germany
Ambrose Verdibello	IBM USA
Phil Wakelin	ITSO Almaden
Kathy Walsh	IBM USA
Sharon Wang	IBM USA
James Warnes	IBM USA
Vern Watts	IBM USA
Klaus Wegele	IBM Germany
Maryela Weihrauch	IBM USA
Suzie Wendler	IBM USA
Gail Whistance	IBM USA
David Whitney	IBM USA
Joseph Winkelbauer	IBM USA
Lynn Winkelbauer	IBM USA
David Yackel	IBM USA
Douglas Zobre	IBM USA

Further thanks to the staff and the editorial team of the International Technical Support Organization, Poughkeepsie.

To everyone else who contributed, our sincere thanks.

Comments welcome

Your comments are important to us!

We want our redbooks to be as helpful as possible. Please send us your comments about this or other redbooks in one of the following ways:

- Fax the evaluation form found in "IBM Redbooks evaluation" on page 211 to the fax number shown on the form.
- Use the online evaluation form found at `http://www.redbooks.ibm.com/`
- Send your comments in an Internet note to `redbook@us.ibm.com`

Chapter 1. I/O Connectivity in Parallel Sysplex

One of the objectives of an "ideal" Parallel Sysplex is to be able to run any critical application on more than one system in the sysplex. A prerequisite of this capability is the required connectivity—connectivity from the application to its programs and data, and connectivity from the user to the application. In this chapter, we discuss the considerations for providing the appropriate connectivity between the application and its programs and data. "Appropriate" in this context means that sufficient capacity and availability are provided. In Chapter 2, "Network Connectivity for Parallel Sysplex" on page 19, we discuss the considerations for connecting users to the application.

In a Parallel Sysplex, the trend is for "any-to-any" connectivity, and nondisruptive changes. The "standard" environment would consist primarily of ESCON channels and ESCON control units, connected via ESCON directors.

However, due to increases in system capacity and the volume of online data, some customers are experiencing limitations in their S/390 Channel environment. These may be overcome by using the new channel architecture, the FIbre CONnection (FICON) channel.

Use of

"FICON" in This Book

As previously mentioned, at the time of writing, FICON channels are only available in FICON Bridge (FCV) mode, whereby the FICON channel is connected to a FICON Bridge card in a 9032-5, which in turn is connected to existing ESCON control units.

In the future, as native FICON control units and FICON directors become available, FICON channels will operate in native FICON (FC) mode. However, all references in this document to FICON channels actually refers to FICON channels operating in FCV mode.

The primary reasons for using ESCON and FICON (FCV) is to ensure cross-connectivity to all devices in the Parallel Sysplex, and to benefit from the support these channels provide for nondisruptive changes.

Recommended Sources of Further Information

The following sources provide support for the information in this chapter:

- *FICON Implementation*, SG24-5169
- *FICON Native Planning Guide*, SG24-5974 (currently in development)
- *FICON Physical Layer*, SA24-7172
- *FICON Planning*, SG24-5443
- *IBM Fiber Saver (2029) Implementation Guide*, SG24-5608
- *IBM RAMAC Scalable Array Storage Introduction*, GC26-7212
- *IBM RAMAC Virtual Array Storage Introduction*, GC26-7168
- *IBM RAMAC Virtual Array Storage Operation and Recovery*, GC26-7171
- *IBM RAMAC Virtual Array Storage Planning, Implementation and Usage Guide*, GC26-7170
- *RAMAC Virtual Array: Implementing Peer-to-Peer Remote Copy*, SG24-5338
- *IBM RAMAC Virtual Array Storage Reference*, GC26-7172
- *IBM 3990/9390 Storage Control Introduction*, GA32-0253
- *IBM 3990/9390 Storage Control Planning, Installation and Storage Administration Guide*, GA32-0100
- *IBM 3990/9390 Storage Control Reference*, GA32-0274
- *IBM 9397 RAMAC Electronic Array Storage Introduction*, GC26-7205
- *IBM 9729 Optical Wavelength Division Multiplexer*, 9729 Package
- *Introduction to S/390 FICON*, SG24-5176
- *IOCP User's Guide and ESCON Channel-to-Channel Reference*, GC38-0401
- *Network Products Reference*, GX28-8002
- *Planning for Fiber Optic Channel Links*, GA23-0367
- *System Overview - G6*, GA22-1030
- *S/390 I/O Connectivity Handbook*, SG24-5444
- *S/390 PR/SM Planning Guide*, GA22-7236

Use the roadmap in Table 1 on page 3 to guide you through this chapter efficiently.

Table 1. Connectivity in the Parallel Sysplex Roadmap			
You want to configure:	**If you are especially interested in:**	**Subtopic of interest:**	**Then refer to:**
A Parallel Sysplex			
	Adequate connectivity of I/O devices		Chapter 1, "I/O Connectivity in Parallel Sysplex" on page 1
		FICON and EMIF benefits in a Parallel Sysplex	1.1, "Fibre Connection (FICON) in FCV Mode"
		ESCON and EMIF benefits in a Parallel Sysplex	1.2, "Enterprise Systems Connection (ESCON)" on page 6
		Logical path requirements, and ESCON directors	1.3.1, "ESCON Logical Paths" on page 14
		DASD control unit logical path requirements	1.3.2, "DASD Control Unit Considerations" on page 15
		CTC considerations	1.3.3, "ESCON Channel-to-Channel (CTC) Considerations" on page 16
		Tape considerations	1.3.4, "Tape Control Unit Considerations" on page 16
		OEM control unit considerations	1.3.5, "OEM Hardware Connectivity" on page 17

1.1 Fibre Connection (FICON) in FCV Mode

The FICON channel is a new channel architecture. FICON provides improved connectivity and other advantages over ESCON. At the time of writing, the only type of FICON channel that is available is one operating in FICON Bridge (FCV) mode—that is, a FICON channel connected to an ESCON control unit via a FICON bridge card in a 9032-5 ESCON director.

What Is a FICON Bridge Card?

A FICON Bridge Card is a special adaptor for use within an IBM 9032-5 that permits the attachment of a FICON channel (operating in FCV mode), allowing communication to ESCON control units using ESCON channel protocols.

Each bridge card supports just 1 port, compared to 8 ports supported on an ESCON card. However, each FICON bridge card supports up to 8 concurrent I/O operations, each to a different ESCON port in the same 9032-5. Each of the ESCON links will go to a different control unit.

You can have a maximum of 16 bridge cards per 9032-5.

Enhancements to your S/390 ESCON channel environment may be provided by S/390 FICON channels for system environments that are experiencing:

- S/390 256 ESCON channel constraint

- 9672 ESCON channel 1024 device address limitation per channel
- Data rate performance droop at extended distances (over 9 km)
- High fiber cabling costs between local and remote sites
- Distance limitations for dark fiber

FICON (FCV) provides the functions of ESCON except that the CTC function available on ESCON TYPE=CTC channels is not currently available on FICON channels. However, FICON (FCV) channels can participate as the "other end" in an ESCON CTC connection in the same manner as an ESCON CNC channel. FICON (FCV) provides the following added advantages over ESCON:

- Fewer connections than ESCON as between 5 and 8 ESCON channels (depending on workload on channels) can be consolidated on 1 FICON (FCV) channel. This results in fewer channels, cables, and patch panel ports.
- FICON (FCV) channels operate in full duplex mode.
- FICON (FCV) provides a 100 MB/sec link rate.
- Up to 8 concurrent I/Os in any combination of reads and/or writes are supported.
- 16 K unit addresses per channel compared to 1024 devices on an ESCON channel.
- Negligible data droop rate up to 100 km (allowable device connectivity distances must still be considered on the ESCON links).
- FICON (FCV) supports distances of up to 10 km (up to 20 km with RPQ 8P1984) between the CPC and the ESCON director when using single mode cables with a longwave Fiber Optic SubAssembly (FOSA).

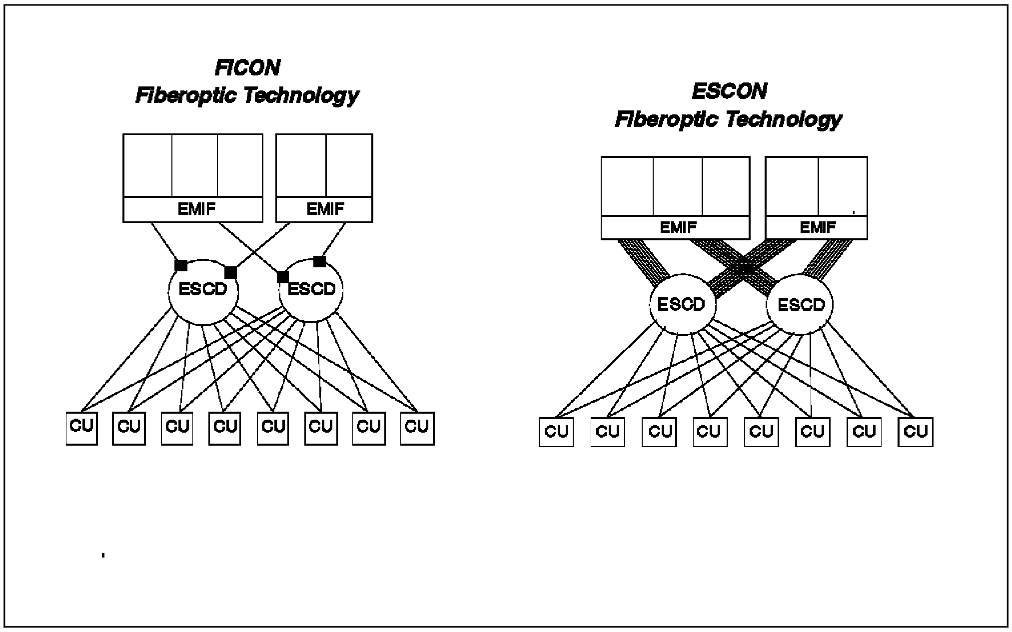


Figure 1. FICON vs. ESCON Cabling

For other benefits of FICON, refer to 1.2, "Enterprise Systems Connection (ESCON)" on page 6.

1.1.1 FICON Exploitation of Existing ESCON Infrastructure

FICON is able to use your existing investment in hardware by continuing to use your current 9032-5s ESCON directors and single or multimode fibre.

FICON bridge cards are required to connect FICON channels to your existing ESCON directors. The FICON bridge card has 1 port per card. The FICON bridge card can be used in place of 1 ESCON port card in the ESCON director. An ESCON port card has 8 ports per card, so the net result of adding a FICON bridge card is the reduction of ESCON director ports by 8. However, as 1 FICON channel can replace up to 8 ESCON channels, this should not cause a problem.

FICON long wavelength FOSA channels (the only type available at the time of writing) use single mode 9 micron fibre cables. The FICON channel and the FICON bridge card in the 9032-5 use SC single mode, duplex receptacles. However, if your existing ESCON fibre infrastructure is based on multimode fibre cables, and you wish to reuse these for the FICON channels, this requires the use of a mode conditioner patch (MCP) cable. In addition to allowing the use of multimode cables for FICON, these are required to adapt the fibre channel duplex connector to the ESCON multimode receptacles.

Although FICON can use existing multimode fibre, optic technology limits the distance to 550 meters. This may be a problem in some installations and needs to be considered when planning for FICON. A sample configuration showing the use of mode conditioning patch cables with FICON channels is shown in Figure 2.

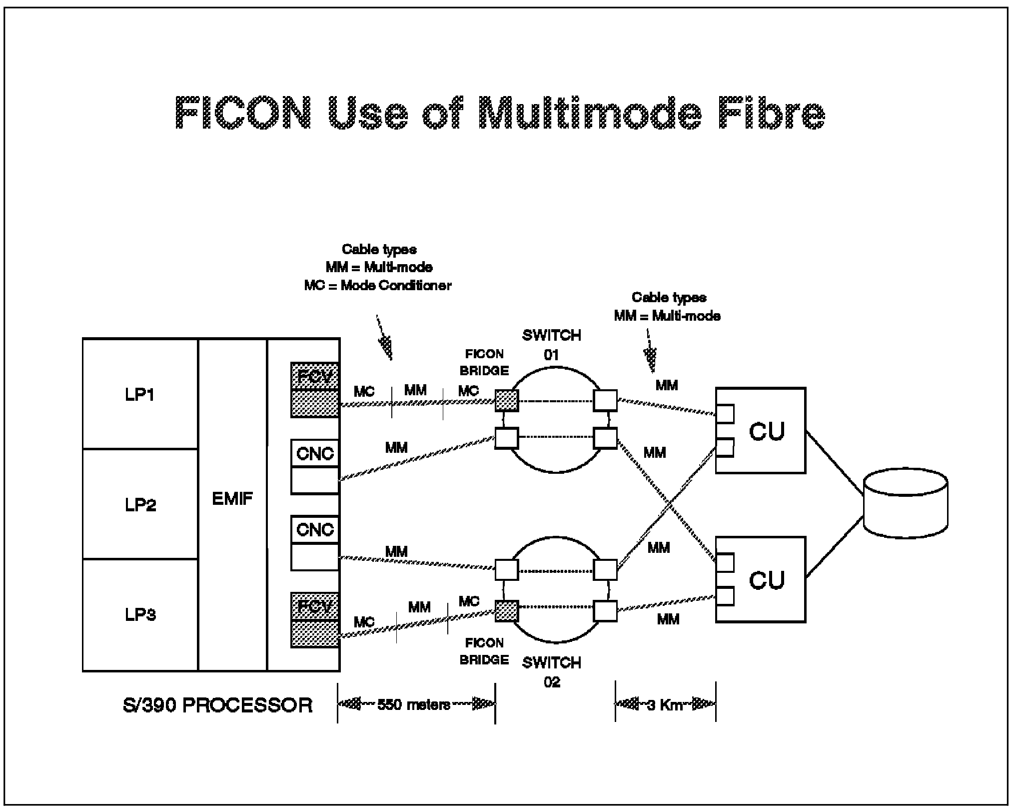


Figure 2. FICON Use of Fiber Cables

1.1.2 System Automation for OS/390

Recommendation to Use System Automation for OS/390

The I/O Operations component of System Automation for OS/390 (previously available as ESCON Manager) is now an integrated function of System Automation for OS/390 (5645-005). Its use is highly recommended in the Parallel Sysplex environment, in order to provide the ability to manage the safe switching of your ESCON directors on a sysplex-wide basis.

The I/O Operations component is a systems management tool designed to enhance user control and manageability in the ESCON and/or FICON environment when changing ESCON director switch matrix configurations.

I/O Operations can perform simultaneous configuration changes on multiple ESCON directors, an action that is not feasible with local control of each ESCON director. Therefore, we recommend that you install System Automation for OS/390 on each CPC with ESCON and FICON connections to installed ESCON directors.

The I/O Operations component can also help you integrate your ESCON director definitions into HCD, thus providing a single point of definition and control for the whole configuration. Without I/O Operations, all ESCON director definitions had to be entered through the local ESCON director console, thus breaking the configuration definition into two areas that have to be managed and controlled separately. There is, however, a drawback to managing your director switch matrix configurations with HCD: every configuration change that you make must be activated (using the dynamic I/O reconfiguration support) on every system that is attached to the director. (In contrast, if the switch matrix is only managed on the director, the change must only be activated once, at the director itself.) You must decide whether centralized management of all your configuration information is valuable enough to you to offset the added overhead of having to do multiple ACTIVATEs.

More information about the System Automation for OS/390 product can be found in A.4.3, "System Automation for OS/390" on page 142.

1.2 Enterprise Systems Connection (ESCON)

Recommendation to Use ESCON and FICON Channels

Note: There are references to FICON (FCV) in this ESCON section as the benefits here also apply to FICON (FCV).

ESCON and FICON (FCV) channels should always be configured as an alternative to parallel channels. Even in smaller configurations, where the connectivity capabilities of ESCON or FICON (FCV) may not be essential, parallel channels should always be viewed as non-strategic.

ESCON and FICON (FCV) channels enable the use of Enhanced Multiple Image Facility (EMIF) on the CPC. This allows for improved channel utilization by consolidating channels for improved channel resource management, improved I/O flexibility, and enhanced connectivity.

ESCON and FICON (FCV) channels and ESCON directors improve connectivity and I/O switching. Also, System Automation for OS/390 helps to manage ESCON, FICON (FCV), and non-ESCON products in the configuration and provides a single point of control for managing I/O resources.

ESCON and FICON (FCV) architectures provide significant benefits beyond parallel (OEMI) channels, including:

- Increased channel-to-control unit distances.
- Smaller and lighter cables.
- Built-in channel-to-channel adapter (CTC) capability. (It is not currently possible to define a FICON (FCV) channel as TYPE=CTC. However, a FICON (FCV) channel can connect to a ESCON channel operating in CTC mode).
- Enhanced switching through the ESCON director and System Automation for OS/390.
- Enhanced availability through "hot plugging" of cables.
- Ability to reduce the number of control unit channel adapters and processor channel ports in some environments.
- Ability to share ESCON and FICON (FCV) channels across LPs.
- Most performance enhancements in the area of channel to control unit connectivity have been to serial interfaces (ESCON) and not parallel.

1.2.1 Enhanced Multiple Image Facility (EMIF)

EMIF was formerly known as ESCON Multiple Image facility but the name has been changed to Enhanced Multiple Image Facility with the introduction of FICON.

IBM 9672 CPCs support up to 15 PR/SM LPs. This is useful in consolidating workloads onto fewer CPCs, and in providing additional hot standby or test partitions.

EMIF allows sharing of ESCON or FICON (FCV)-attached control units across the PR/SM partitions within a CPC. This can reduce the number of channels, control unit channel adapters, and ESCON director ports otherwise required to share control units across LPs.

The Enhanced Multiple Image Facility (EMIF) supports sharing of:

- FICON channels
- ESCON channels
- CF links at the sender end
- OSA channels

1.2.2 ESCON Extended Distance Feature (XDF)

ESCON XDF allows extended distances using laser-driven fiber. Although originally announced as available on some IBM CPCs, this feature was withdrawn due to the availability of external channel extenders. However, it is still available on ESCON directors.

The distances available through the use of XDF are device-dependent. For example, an IBM 3174 can be up to 43 km from the CPC; an IBM 3490 can be up to 23 km from the CPC; and the maximum distance for an Enterprise Storage Server is 43 km.

1.2.3 ESCON Director (ESCD) Configuration Guidelines

The 9032-5 is the only ESCON director currently available from IBM. The 9032-5 is used to provide connectivity between ESCON and FICON (FCV) channels and ESCON control units, another ESCON director, or other ESCON channels.

The 9032 Model 5 is floor-standing and is available in sizes ranging from 24 ESCON ports to 248 ESCON ports, in increments of 8 ports. It also supports up to 16 FICON bridge ports at the expense of 16 ESCON port cards. One of the provided features of the 9032 model 5 is a migration tool to provide simple port definition migration from a 9032 model 3 (now withdrawn).

The 9032-5 also has:

- Optional hardware component redundancy
- Concurrent LIC upgrade capability
- Dynamic addition of ports (hot pluggable)

For information on the use of trunk cables with an ESCON director, see 1.2.7, "FTS Direct Attach" on page 13.

It is possible to attach the 9032 ESCON director to the same Token-Ring as the Hardware Management Console of any S/390 CMOS CPC. The console function for the ESCON directors can be provided from the Hardware Management Console instead of a stand-alone console. This capability makes the purchase of a separate console unnecessary and helps consolidate the operational requirements to a single point of control.

Recommendation about ESCON Director Models

It is good practice to plan for a larger potential ESCON and FICON bridge port capacity than initially needed. This simplifies the future addition of CPCs and I/O devices in a Parallel Sysplex.

While the 9032-5 supports the non-disruptive installation of additional port cards, configuring the director with a small number of spare ports gives you the flexibility to make minor changes to your configuration. Should you make a significant change to your configuration, such as ordering an additional CPC, that order should include additional port cards for all the directors to accommodate the additional channels.

The ESCON director, with its dynamic switching capability, is the "hub" of the ESCON topology, as shown in Figure 3 on page 9.

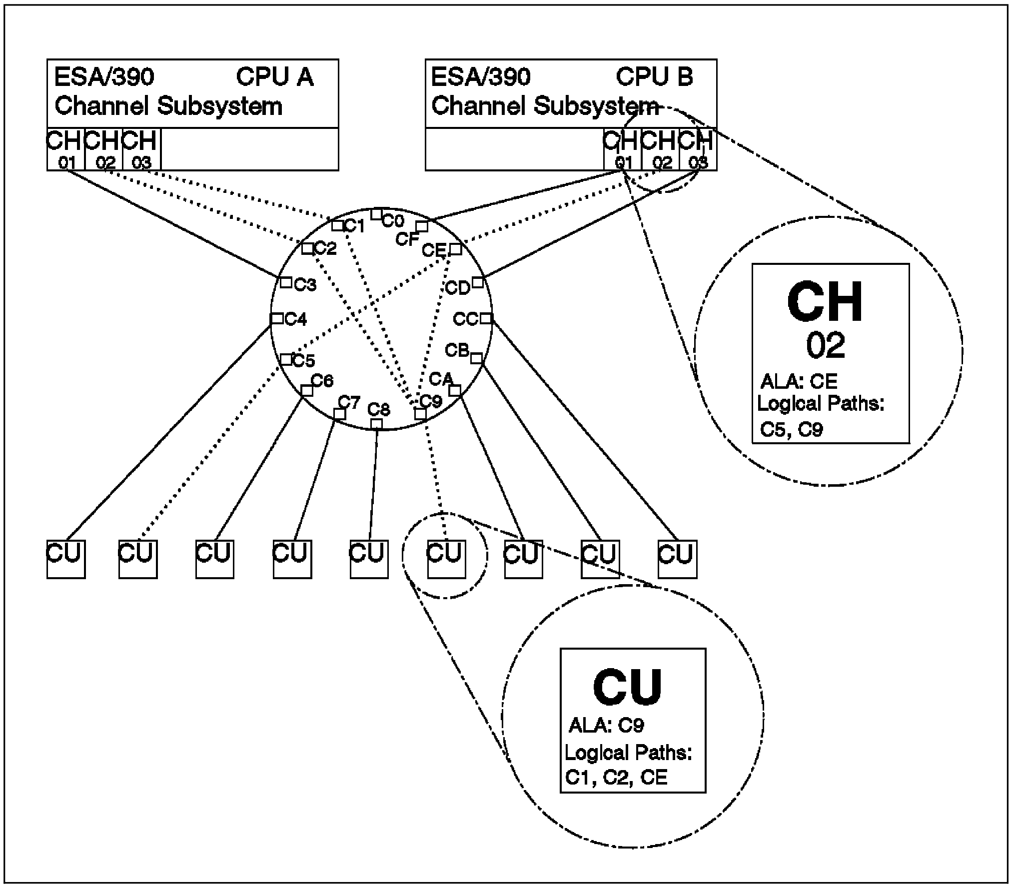


Figure 3. Relationship between Logical Paths, ESCDs, and Links

The ESCON director routes serial frame transmissions from one port to any other port in the same ESCON director and provides concurrent data and command exchange for multiple channels and control units.

ESCON directors are non-blocking, in the sense that there are enough internal paths for up to 124 pairs of ports to communicate simultaneously. Signals travel from one port to another, converting from optical signals to electrical, and back to optical. This allows for some "elasticity" in the bit stream (source and destination ports do not have to be locked at exactly the same time) and for re-powering the optical signal, balancing for any losses at interfaces or on the fiber cables themselves.

The ESCON director also provides a control unit function that can control port connectivity. A port's connection state can be dedicated, prohibited from communicating to other ports, or blocked from communicating.

The control unit function is invoked either by the ESCON director console or by the control unit port, which is a reserved port internal to the ESCON director. System Automation for OS/390 can communicate with the control unit port and provide the user with a high-level, easy-to-use interface.

Configuring for Availability

In order to optimize the availability of the installed hardware, there are a number of things to bear in mind when designing your ESCON director configuration:

- You should have *at least* two ESCON directors
- The paths to a given control unit should be divided evenly over the installed directors.
- The channels from a given CPC should be divided over as many directors as possible.
- The directors should be configured with dual power supplies, and each director should have at least two consoles. From a centralized management point of view, it may be attractive to use the HMCs as the consoles. However, care must be taken that the status of both the directors and the CPCs can be monitored concurrently. Also, not having a console close to the director may make maintenance activities more awkward.
- You should define the Control Unit Port (CUP) on each director in HCD, even if you do not have System Automation for OS/390 installed. If the CUP is not defined, the director has no way of reporting internal errors back to the attached systems.

1.2.4 IBM 9036 Remote Channel Extender Configuration Guidelines

The IBM 9036 Model 3 for Sysplex Timer (RPQ 8K1919) is used to extend the distance between:

- Sysplex Timer model 2 and the CPC up to 26 km
- Two model 2 Sysplex Timers in a high availability configuration up to 26 km

A pair of these is required for each function. For more information about this RPQ, see 3.3, "9037 RPQs" on page 91.

The IBM 9036 ESCON Remote Channel Extender model 1 and 2 have now been replaced by:

- Soneplex ESCON Channel Extender (ECX) - System ECX Model 201. This replaces the 9036 model 1. This model supported multimode to single-mode fiber conversion and signal repeat.
- Soneplex ESCON Channel Extender (ECX) - System ECX Model 101. This replaces the 9036 model 2. This model supported single-mode to single-mode fiber signal repeat.

At the time of writing there was little information available on the Soneplex ESCON Channel Extenders. The following boxed information is based on the IBM 9036 model 1 and 2.

Remote Channel Extender

You can have a maximum of three Remote Channel Extenders and two ESCON directors within a channel-to-control unit, or CPC, path. Each Remote Channel Extender adds 200 nanoseconds of latency (equivalent to approximately 40 meters of fiber) to the link. Link timeouts and data rates are unaffected, other than by the latency added by the Remote Channel Extender in the link.

There is also an IBM 9036 Model 4 ESCON Remote Channel Extender (RPQ 8P1841). This model is a chassis that supports up to 14 cards, each of which may be single-mode to single-mode (RPQ 8P1842), or multimode to single-mode (RPQ 8P1843).

1.2.5 IBM 9729 Optical Wavelength Division Multiplexer

9729 Withdrawn from Marketing

At the time of writing, the IBM 9729 has been withdrawn from marketing. However, as this unit is still in use at many customers sites, we have left the following section in the book. The 9729 has been replaced by the IBM 2029 Fiber Saver, described in 1.2.6, "IBM 2029 Dense Wavelength Division Multiplexer" on page 13.

The IBM 9729 Optical Wavelength Division Multiplexer enables multiple bit streams, each using a different communications protocol, bit rate, and frame format, to be multiplexed onto a single optical fiber for transmission between separate locations. The IBM 9729 Multiplexer can multiplex ten full duplex bit streams, each at up to 1 Gbps over a single single-mode optical fiber. The distance between the two locations can be up to 50 km for a 200 Mbps rate per channel. This allows ten ESCON channels to use one fiber, but the location of ESCON devices, of course, is limited by the supported distances.

Additionally, the IBM 9729 supports the Sysplex Timer (IBM 9037-2) links. For the Sysplex Timer, the links between two IBM 9037-2 models in an IBM 9037-2 Expanded Availability configuration and links from the IBM 9037-2 to the CPC are supported to distances of up to 26 km.

The 9729 Model 1 supports HiPerLink single-mode CF links at distances of up to 20 km.

The IBM 9729 supports two fibers between the A-box and the B-box, where one fiber is the primary and used for active transmission, and the other fiber is the backup. The time for the IBM 9729 to detect a "loss of light" and automatically switch from the primary fiber to the backup is less than 2.5 seconds.

The fact that the IBM 9729 has a backup fiber between the A-box and B-box does not guarantee a high availability configuration. For example, assume that *all* the cross-site links in a multi-site Parallel Sysplex go through a pair of MuxMasters (A-box and B-box) with dual fibers (each taking a different route) and the primary fiber is severed. The following are the results:

- **FICON (FCV) links:** The channel subsystem detects the loss of light and attempts to redrive the I/O through other non-functional, paths resulting in the I/O being unsuccessful.

- **ESCON links:** The channel subsystem detects the loss of light and attempts to redrive the I/O through other non-functional paths, resulting in the I/O being unsuccessful.
- **PPRC links:** The primary DASD CU detects the loss of light within 1 second (ESCON architected channel timeout) and redrives the I/O through other paths within tens of milliseconds, resulting in the PPRC volume pairs being suspended.

  The primary DASD CU detects when the links are repaired and automatically reestablishes the paths. The installation must issue CESTPAIR to reestablish the duplexed PPRC volume pairs.
- **CF links:** Depending upon the error, XES support either redrives the request for up to one second and validates pathing if the operation was unsuccessful, or immediately validates pathing. Path validation tries to identify and activate the message path for up to 2.5 seconds. If the identify and activate processing succeeds, the path is considered good and the request is redriven; otherwise, the CF connectivity is lost. The IBM 9729 will switch to the backup fiber prior to XES performing path validation and redriving the request.
- **Sysplex Timer:** The secondary clock goes offline. This can take from a couple of milliseconds up to one million microseconds (1.048576 seconds), depending upon when the fault occurs. The systems in the same data center as the secondary clock do not receive timer signals (from either the local links or the cross-site links) and therefore enter disabled wait states.

1.2.5.1 Configuring for High Availability

To achieve a high availability configuration, using two pairs of IBM 9729 A- and B-boxes guarantee connectivity if the first pair primary and second pair backup use one route and the first pair backup and second pair primary use another route.

- **FICON (FCV) links:** The channel subsystem detects the loss of light and redrives the I/O through other functional paths, resulting in the I/O being successful.
- **ESCON links:** The channel subsystem detects the loss of light and redrives the I/O through other functional paths, resulting in the I/O being successful.
- **PPRC links:** The primary DASD CU detects the loss of light within 1 second (ESCON architected channel timeout) and redrives the I/O through other functional paths within 10s of milliseconds.
- **CF links:** Depending upon the error, XES support either redrives the request for up to one second and validates pathing if the operation was unsuccessful, or immediately validates pathing. Path validation will try to identify and activate the message path for up to 2.5 seconds. If identify and activate processing succeeds, the path is considered good and the request is redriven; otherwise the link and CF connectivity are lost. The IBM 9729 switches to the backup fiber prior to XES performing path validation and redriving the request; if path validation is unsuccessful, XES loses the link and redrives the request over the other link (assuming it goes to the same CF).
- **Sysplex Timer:** The secondary clock does not go offline (there are two fibers, each going through a different pair of IBM 9729 models) and the systems do not enter disabled wait states (the CPC may switch to the other clock if the active clock time signal is lost).

The links between the IBM 9037 and the CPC become operational automatically once the fiber problem is corrected.

1.2.6 IBM 2029 Dense Wavelength Division Multiplexer

The IBM 2029 Dense Wavelength Division Multiplexer enables multiple bit streams, each using a different communications protocol, bit rate, and frame format, to be multiplexed onto a single optical fiber for transmission between separate locations. The IBM 2029 Multiplexer can multiplex 32 full duplex channels, each at up to 1.25 Gbps over two single-mode optical fibers. When fully configured, the 2029 can provide 64 full duplex channels over 2 fibre pairs. It is possible to configure the 2029 with 2 fibres and support either:

- 32 full duplex channels with backup, or
- 64 full duplex channels without backup

The 2029 also has redundancy built into it which allows the use of only one 2029 while ensuring no single point of failure. This redundancy includes such things as dual power supplies and dual cooling. The distance between the two locations can be up to 50 km (70 km with RPQ). These distances are dependent on the support of the attached devices. Device distances are as follows:

- FICON (FCV) - 50 km (70 km with RPQ)
- ESCON - 43 km
- Sysplex timer - 40 km
- CF links - 40 km

The 2029 provides support for a Geographically Dispersed Parallel Sysplex (GDPS) up to 40 km (limited by the Sysplex Timer support).

1.2.7 FTS Direct Attach

The Fiber Transport Services (FTS) Direct Attach feature permits quick connect and disconnect of fiber trunk cables at both the CF and the sender CPC ends. If the CF is moved, the trunk cables are quickly disconnected from the direct attach harnesses, which remain installed and plugged into the CF ports. Once the CPC and trunk cables are relocated, the trunk cables are easily reconnected to the "Direct Attach" harnesses. One trunk connector is equivalent to plugging 6 or 72 duplex jumper cables. *Fiber Transport Services (FTS) Direct Attach Physical and Configuration Planning*, explains how to plan an FTS direct attach installation for the following:

- IBM 9672
- IBM 9032 ESCON models 005

Installation Instructions for Fiber Transport Services (FTS) Direct Attach Trunking System in 9672 CPCs CFs provides instructions and a pictorial description of how to install the direct attach harnesses and trunks in all IBM 9672 CPCs.

Also, refer to the announcement letter *IBM Enhances Fiber Transport Services*, (696-025).

1.3 Channel Configuration

Whenever new channels are acquired or hardware is changed, the opportunity should be taken to move toward ESCON and FICON. See Chapter 1, "I/O Connectivity in Parallel Sysplex" on page 1 for the reasons behind this strategy.

Use EMIF to share physical channels between LPs on the same CPC. This can reduce the number of channels required. EMIF does not support parallel channels, and does not allow channels to be shared between CPCs.

1.3.1 ESCON Logical Paths

Besides understanding the physical configuration requirements of the I/O subsystem, it is also necessary to have an awareness of the ESCON Logical Path configuration, which can impact both performance and availability. Under some circumstances, providing a physical path from a CPC to a device does not guarantee a working (logical) path.

Each ESCON-capable control unit supports a specific number of ESCON Logical Paths, regardless of the number of physical channels connected and the number of IOCP paths defined. In addition to being a connectivity problem, this can also become an availability issue if the mapping and manipulation of the Logical Path are not understood.

For example, Figure 4 on page 15 shows that each of the ten systems in the Parallel Sysplex has multiple paths defined and configured to the 3990-6 control unit. The 3990-6 supports up to 16 logical paths per host adaptor, so this configuration should not cause a problem.

However, if you inadvertently set up HCD so that multiple channels from the one CPC used the same destination port on the ESCON director (rather than each channel using a different destination port as you would expect for this configuration) you could exceed this limit of 16 paths.

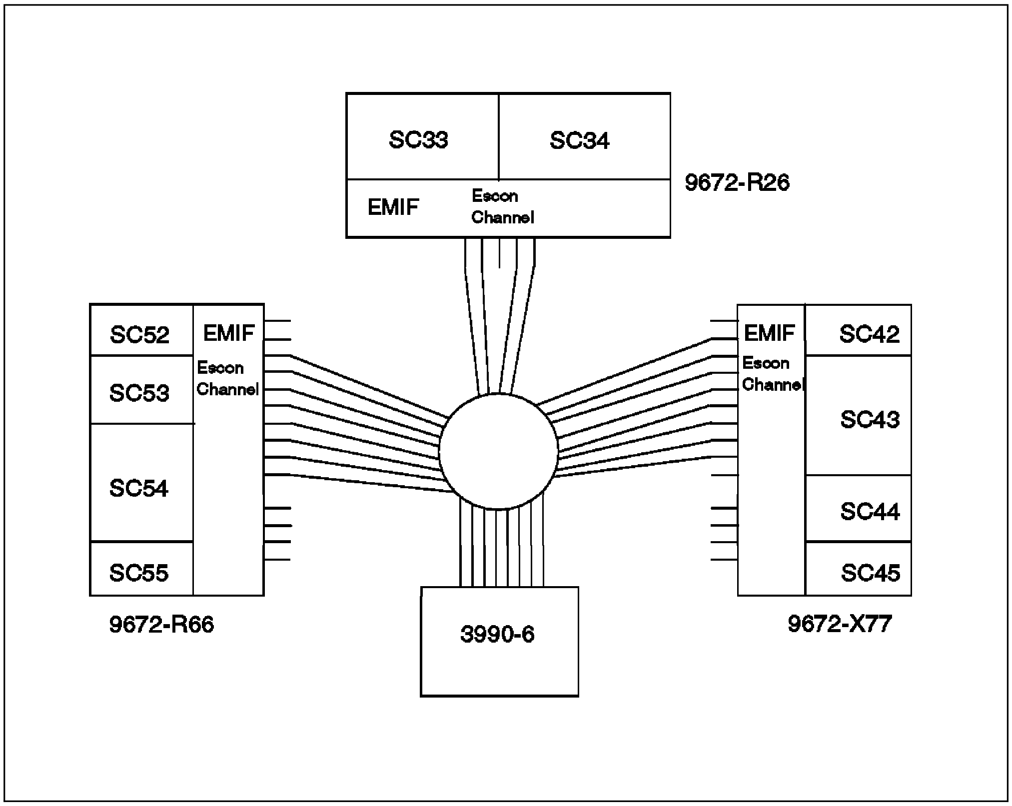


Figure 4. ESCON Logical Paths Configuration

The ESCON Logical Path considerations for each specific control unit type are discussed in detail in the following sections.

Logical paths are established at LPAR activation for any control unit that is on a channel that contains that LPAR in its access list. Requests to establish logical paths are based on information contained in the IOCDS. To help you verify your understanding of how many logical paths are in use, you can use the ANALYZE command of the ICKDSF program. The output from that command will tell you which paths are allocated, which host adaptor each is associated with, and the CPU serial number and partition number of the host that is using each logical path.

Check That Logical Path Support Is Not Exceeded

Remember to check that the control unit supports the number of logical paths that you have defined; it would be easy to exceed the number of logical paths on older control units using ESCON and FICON.

1.3.2 DASD Control Unit Considerations

IBM offers multiple DASD solutions. Remember that regardless of the DASD solution chosen, the connectivity issues (especially logical paths) remain.

DASD Logical Paths

Ensure that DASD devices are connected through control units that provide enough logical paths.

The Enterprise Storage Server supports a total of 2048 logical paths, 128 per logical control unit, with a maximum of 64 per ESCON port.

The IBM 3990-6 has up to 128 logical paths, 16 per ESCON port. Similar numbers apply to the 9390 control unit.

The IBM RAMAC Virtual Array Model 2 subsystem supports up to 128 logical paths.

The IBM RAMAC Scalable Array (RSA) and RAMAC Electronic Array (REA) have up to 512 logical paths, 32 logical paths per ESCON port.

1.3.3 ESCON Channel-to-Channel (CTC) Considerations

With the advent of HiperLink coupling links, it is unlikely that installations will be defining new CTCs, in addition to those already in use. For this reason, we have removed the section of this book that talks about a CTC connection methodology. If you *do* have a need for additional CTC connections, we recommend that you refer to the corresponding chapter in the previous version of this book, *OS/390 MVS Parallel Sysplex Configuration Volume 3: Connectivity*, SG24-2077.

1.3.4 Tape Control Unit Considerations

Unlike disk devices which, by their design, are intended to be concurrently used by more than one user, tape drives would normally only ever be in use by one system at a time. This design is enforced by the ASSIGN command which ensures that a drive can only be assigned to one system at a time.

However, to allow tape drives to be shared between systems, support is provided to allow a tape device to be online to more than one system at a time. This capability is intended to be used in conjunction with software to control the sharing. For example, JES3 makes use of this capability to have the drives online to many systems, and JES3 then controls which jobs can use a particular drive at a given time.

Tape Drive Sharing in a Parallel Sysplex

OS/390 provides automatic tape drive sharing within a Parallel Sysplex. JES2 installations can share tape drives (3480, 3490, 3490E, and 3590) between systems in a Parallel Sysplex. Tape drive switching across Parallel Sysplexes is not provided. Tape drives are either dedicated or *auto-switched*.

If all your tape drives are accessed from systems in one Parallel Sysplex, then you may no longer need special tape drive switching software.

For more information on the tape unit sharing CF structure, refer to 2.9, “Tape Allocation Structures” on page 122 in Volume 2: Cookbook, SG24-5638.

Logical Paths in Tape Subsystems

- The 3590-A50 tape subsystem can have up to two ESCON adapters, each with 64 logical paths. The 3590-A60 tape subsystem can have up to eight ESCON adapters, each with 64 logical paths, providing a total of up to 512 logical paths.
- Existing IBM 3490/3490E tape drives may be upgraded with ESCON channel interfaces (if not already installed) to provide 16 logical paths per ESCON port.

Older tape subsystems, such as the IBM 3803, IBM 3420, IBM 3422, and IBM 3480 and the IBM 3490/3490E that are equipped with parallel adapters, can be accommodated through use of the 9034 ESCON Converter, or through parallel channel interfaces.

ESCON-Attached 3494 Virtual Tape Server (VTS)

The 3494 VTS connects to the 3494 tape library data server. From an ESCON perspective it looks like 2 or 4 3490E control units. Data is written to DASD contained in the VTS which emulate 3490 cartridges. The OS/390 system believes it is writing to a physical cartridge.

This virtual cartridge is offloaded from the DASD in the VTS to a physical volume at some later time by the VTS. Because the VTS looks like 3490E control units to ESCON, the number of logical paths are the standard 3490E discussed previously.

IBM 3494 Connectivity Considerations

There are no unique or additional considerations, relating to connectivity, for the IBM 3494. The control units within the IBM 3494 are the standard IBM 3490E and IBM 3590 control units discussed previously.

1.3.5 OEM Hardware Connectivity

Connectivity is the primary area that should be looked at. Considerations are limited to ensure there are enough physical and logical paths to support the Parallel Sysplex environment. It is normal practice to ensure that all control units (OEM and IBM) are at an appropriate microcode level if a new CPC is installed as part of a Parallel Sysplex migration.

Chapter 2. Network Connectivity for Parallel Sysplex

This chapter describes the options for connecting your network to a Parallel Sysplex from both a VTAM and TCP/IP point of view. The section on VTAM looks at the general networking aspects related to the configuration of the VTAM systems inside and outside the Parallel Sysplex. In particular, it discusses the requirement for APPN and High Performance Routing (HPR) and shows how these requirements are integrated with existing non-APPN networks. The section on TCP/IP discusses the options available with IBM's TCP/IP for exploiting Parallel Sysplex.

This chapter also discuss hardware aspects of connecting the network to the Parallel Sysplex. The choice of the gateway between the network and the Parallel Sysplex can affect availability in some configurations.

Recommended Sources of Further Information

The following sources provide support for the information in this chapter:

- *IBM Router Interoperability and Migration Examples*, SG24-5865
- *IBM 3746 Nways Controller Models 900 and 950: APPN Implementation Guide*, SG24-2536
- *Inside APPN: The Essential Guide to the Next-Generation SNA*, SG24-3669
- *NCP Tuning with NTune*, GG24-2520
- *NCP V7 Resource Definition Guide*, SC31-6223
- *Network Products Reference*, GX28-8002
- *Nways Multiprotocol Access Services Configuration Reference Vol 1*, SC30-3884
- *Nways Multiprotocol Access Services Configuration Reference Vol 2*, SC30-3885
- *Open Systems Adapter 2 Implementation Guide*, SG24-4770
- *Open Systems Adapter (OSA) Presentation Pack*, MDOSA package
- *OSA-Express Customer's Guide and Reference*, SA22-7403
- *OS/390 eNetwork Communications Server V2R7 TCP/IP Implementation Guide Volume 1: Configuration and Routing*, SG24-5227
- *OS/390 eNetwork Communications Server V2R7 TCP/IP Implementation Guide Volume 2: UNIX Applications*, SG24-5228
- *OS/390 eNetwork Communications Server V2R7 TCP/IP Implementation Guide Volume 3: MVS Applications*, SG24-5229
- *OS/390 MVS Parallel Sysplex Test Report*, GC28-1963
- *Performance for OS/390 Open Systems Adapter*, OSAPERF package
- *Planning for S/390 Open Systems Adapter Feature*, GC23-3870
- *S/390 OSA-Express Gigabit Ethernet Implementation Guide* SG24-5443
- *S/390 OSA-Express Gigabit Ethernet Performance Report*, OSAEPERF package
- *S/390 OSA-Express Implementation Guide*, SG24-5948
- *Security in OS/390-based TCP/IP Networks*, SG24-5383
- *SNA and TCP/IP Integration*, SG24-5291
- *SNA in a Parallel Sysplex Environment*, SG24-2113
- *Subarea Network to APPN Network Migration Experiences*, SG24-4656
- *Subarea to APPN Migration: HPR and DLUR Implementation*, SG24-5204
- *TCP/IP in a Sysplex*, SG24-5235
- *3745/3746 Overview*, GA33-0180
- *3745 Communications Controller Models A, 3746 Nways Multiprotocol Controller Models 900 and 950*, GA33-0457

Use the roadmap in Table 2 on page 21 to guide you through this chapter efficiently.

Table 2. Network Connectivity for the Parallel Sysplex Roadmap			
You want to configure:	**If you are especially interested in:**	**Subtopic of interest:**	**Then refer to:**
Network connections to a Parallel Sysplex			
	General VTAM networking issues. When do I need to use APPN? When do I need to use HPR? What about HPDT?		2.1, "Network Access to Parallel Sysplex Using VTAM"
		What difference do session managers make?	2.1.5, "Session Manager Implications in Parallel Sysplex" on page 31
		How do I use a CMC?	2.1.6, "Communications Management Configuration" on page 33
		Can I use SNI?	2.2.2, "SNI Considerations" on page 39
		Where does Enterprise Extender fit in?	2.2.3, "Enterprise Extender" on page 39
	Using TCP/IP in a Parallel Sysplex environment.		2.3, "Network Access to Parallel Sysplex Using TCP/IP" on page 42
		TCP/IP availability features	2.3.1, "Availability Enhancements" on page 42
		TCP/IP load balancing features	2.3.2, "Workload Balancing Enhancements" on page 46
	Network Gateways to a Parallel Sysplex. What difference does the choice of gateway make?		2.4, "Connecting to the Network" on page 57
		IBM 3745/6 considerations.	2.4.1, "IBM 3745/6 Communications Controller Family" on page 57
		Open Systems Adapter considerations.	2.4.2, "Open Systems Adapter" on page 61
		Channel-attached routers.	2.4.3, "Channel-Attached Routers" on page 68

2.1 Network Access to Parallel Sysplex Using VTAM

There are some networking issues that relate to functions discussed in other parts of this book. This section addresses some of the more general issues.

For details on availability, refer to 3.3, "Network Considerations for Availability" on page 106 in Volume 1: Overview, SG24-5637. For details on workload balancing, refer to 4.5, "Network Workload Balancing Capabilities" on page 221 in Volume 1: Overview, SG24-5637.

2.1.1 APPN Overview

The Advanced Peer-to-Peer Networking (APPN) architecture is an extension to SNA that enables systems to communicate and operate in a peer-oriented manner. VTAM V4.1 and higher provide APPN for the S/390 server platform.

The APPN architecture defines two node types:

- Network node
- End node

The APPN *network node* is responsible for network services and performs functions such as topology and directory services on behalf of the end nodes.

The APPN *end node* is typically where applications are run, and it has a subset of the networking services of a network node.

The end node uses one particular network node for its network services, and this network node is known as the *network node server* for that end node.

Network Nodes and Generic Resource Requirements

The network node server must be in the same Parallel Sysplex as the end nodes it serves if the generic resources function is implemented. The network node server does not need to be in the Parallel Sysplex for Multi-Node Persistent Sessions (MNPS).

The VTAM generic resources feature is discussed in 3.3.1, "VTAM Generic Resources Function" on page 106 in Volume 1: Overview, SG24-5637, and MNPS is discussed in 3.3.3, "Multi-Node Persistent Sessions" on page 107 in Volume 1: Overview, SG24-5637.

An end node has a control session with its network node server, known as its *control point session*. This control point session is used for many functions, including dynamic application registration and session setup requests. If the selected network node server fails, its end nodes must connect to an alternative network node in the Parallel Sysplex and establish control point sessions. If there is no alternative network node for an end node to select as its network node server, then the end users in the network will not be able to establish new sessions with generic resource applications on that end node.

VTAM's implementation of APPN is different from other APPN nodes, because VTAM can operate as a subarea SNA system while also acting as an APPN node. Consequently, there are various types of configurations that VTAM can use:

Network node
: An APPN network node with no subarea appearance.

Interchange node
: An APPN network node with a subarea appearance.

End node
: An APPN end node with no subarea appearance.

Branch Network node
: A node that appears to be a network node to a branch location, but appears as an end node to the mainframe. This is the only APPN node type now supported by Cisco.

Low Entry Networking (LEN) node
: An SNA Node Type 2.1 which doesn't understand APPN.

Migration data host
: An APPN end node with a subarea appearance.

Subarea node
: A subarea node with no APPN appearance
All VTAMs prior to VTAM V4.1 are defined this way. All VTAMs from V4.1 onwards can optionally be configured as subarea nodes, which is the default option.

APPN is enabled in VTAM by specifying the NODETYPE start option. In the simplest configurations, this is *all* that must be done. Refer to the *OS/390 Parallel Sysplex Test Report* for a description of the VTAM definitions used in a sample Parallel Sysplex.

The APPN architecture supports all LU types, including LU0, LU1, LU2, LU3, LU6, LU6.1 and LU6.2.

BSC3270 connections are not supported over APPN links.

BSC3270 and APPN

For BSC3270 connections, a BSC device can logon to a Generic Resource, as long as the following are all true:

- There is a subarea path from the VTAM that owns the BSC device to the Resource Selector Node.
- There is a subarea path from the VTAM that owns the BSC device to the real instance of GR.
- The necessary ADJSSCP tables are in place to ensure subarea routing can be accomplished successfully.

Note: You cannot use any APPN searching, for any of these conditions, even if the APPN route is VR-TG.

2.1.1.1 Virtual-Route-Based Transmission Group

In most configurations, especially where the network is large and complex, there are some other VTAM parameters that must be considered in addition to NODETYPE. One very important consideration when planning a migration is the integration of subarea and APPN routing.

VTAM V4.2 introduced a routing function called the *virtual-route-based transmission group* (VR-TG) to assist in subarea-to-APPN migrations. This function allows VTAMs that operate in subarea mode using SSCP sessions to also bring up APPN control point sessions over subarea routes. This enables APPN connectivity over a subarea network without having to define separate APPN connections. All of the subarea routes between the two VTAMs can now be described as a single APPN transmission group, providing optimum routing over the subarea network in a mixed APPN/subarea environment.

Using VR-TGs is very simple. They are defined by specifying VRTG=YES on the cross-domain resource manager (CDRM) statements used to define other VTAMs. You do not have to code VRTG=YES on CDRM definitions if you want VR-TGs with every other adjacent CDRM in your network. Instead, you can simply code VRTG=YES as a start option.

For more information on VR-TGs and mixed APPN/subarea environments, refer to *VTAM for MVS/ESA V4.4 Network Implementation Guide*.

2.1.1.2 Dependent LU Requester/Server

Dependent LUs, such as 3270 terminals and printers, continue to be supported by all releases of VTAM. This is true even if VTAM is configured as an APPN node. However, these types of LUs are dependent on VTAM for their session establishment, and they are supported by a function known as the *boundary function*. The boundary function is provided by VTAM or the NCP, and dependent LUs must be logically adjacent to the VTAM or NCP to get this function.

If you are migrating from an existing subarea network to APPN, but you do not wish to make any changes to the network that supports dependent LUs, then no additional functions are required. However, if the migration to APPN is happening to enable additional APPN routing nodes to be used in the network, then additional function may be required. You add additional network nodes, for example, when 3745s are replaced with 3746-950 network nodes. If the APPN network causes additional APPN routing nodes to be placed in the network, then the dependent LUs may no longer be adjacent to the boundary function.

The feature known as *Dependent LU Requester/Server* (DLUR) allows remote dependent LUs to be placed in an APPN network, thus removing the requirement to be adjacent to the boundary function in VTAM or NCP, and allowing more flexible APPN network design. It also brings APPN routing benefits to the dependent LUs because routes through the network can now be optimized.

The function is provided in two parts. The dependent LU server component was first introduced in VTAM V4.2. The dependent LU requester is provided by the following platforms (minimum levels shown):

- Communication Manager/2 V1.11 for OS/2
- Communications Server for OS/2 WARP V4
- Communications Server for AIX V4.2
- Communications Server for NT V5.0
- Personal Communications 3270 V4.1 for Windows 95
- 2216 Model 400
- 3174 Configuration Support Level C6
- 3746 Model 900/950
- Non-IBM products

2.1.2 Use of APPN in Parallel Sysplex

It is possible to implement a Parallel Sysplex without using APPN. However, the following functions will require the use of APPN protocols inside the Parallel Sysplex:

- Generic Resources (GR)
- Multi-Node Persistent Sessions (MNPS)

There is another reason for using APPN, which is to get the benefits of the OS/390 cloning enhancements that were introduced by VTAM V4.3. This is discussed in 3.3.5, "VTAM Systems Management" on page 109 in Volume 1: Overview, SG24-5637. Although APPN is not required to support cloning, it is much easier to exploit it if the VTAM systems are defined as APPN nodes rather than subarea SNA nodes.

If APPN is being used for any of these reasons, then some of the VTAM systems will be defined as APPN end nodes and some will be defined as APPN network nodes. It is possible to have just one APPN network node and configure the rest of the systems as APPN end nodes. However, in a Parallel Sysplex, the recommendation is to have a minimum of two VTAM network nodes for availability reasons. The rest of the VTAM systems can be configured as end nodes.

The network nodes can be pure network nodes or interchange nodes. The end nodes can be pure end nodes or migration hosts. If it is possible to add any new VTAM systems to the Parallel Sysplex as pure APPN nodes with no subarea appearance, then this will provide the simplest solution to configure. There are very few additional definitions to be made in APPN, compared to what was done for subarea systems.

How Do I Configure My VTAM Systems?

1. Define existing subarea nodes as interchange nodes or migration data hosts, and make use of VR-TGs. This will enable APPN connectivity, without having to change any of your subarea routing.
2. Use APPN for new images. APPN allows the new systems to be added with very few definitions, without having to change any of your subarea routing.

If you define VTAM systems as APPN nodes, it is still possible to access applications on the systems from workstations that are not using the APPN protocols. This requires the use of an interchange node somewhere in the network. The interchange node can be inside or outside the Parallel Sysplex.

It is possible to run user applications on the VTAM network nodes in a Parallel Sysplex. Both end nodes and network nodes can support subsystems such as CICS, IMS, DB2, and TSO.

Generic Resources and Network Nodes

If the VTAM generic resource function is being used, the actual application instances can run on VTAM end nodes or VTAM network nodes. However, for availability reasons, it may be desirable to use the VTAM network nodes purely for networking services, and run applications only on the VTAM end nodes.

Multi-Node Persistent Session Applications

While instances of GR applications can reside on *either* sysplex end nodes or network nodes, OS/390 V2R6 or higher is required if you wish to run MNPS applications on network nodes. All VTAM nodes providing Multi-Node Persistent Session must be defined as Rapid Transit Protocol (RTP) level nodes (HPR=RTP start option).

2.1.2.1 APPN Planning Considerations

It is not possible to make a recommendation about how to design the APPN configurations for the VTAM systems in a Parallel Sysplex. This choice is dependent on the particular environment and on the way the network is configured and used. However, there are a number of general points that should be considered and these are listed in this section. There are also some more specific considerations for generic resources listed in 4.5.1.4, "Generic Resource Planning Considerations" on page 227 in Volume 1: Overview, SG24-5637.

- Ensure that every VTAM end node has logical links to all potential VTAM network node servers to carry the control point sessions. In a sysplex, the best option is to use XCF for APPN CP-to-CP sessions. You can also use a direct channel path between the VTAM end node and the VTAM network node, or it could be a path from the VTAM end node to an NCP that is owned by the VTAM network node.
- The generic resources function requires the use of APPN on the VTAM systems that are supporting the application instances. These VTAM systems can be configured as "pure" APPN nodes (end nodes and network nodes), or as APPN nodes with subarea support (migration data hosts and interchange nodes).
- It is possible to configure VTAM systems so that there is a mixture of node types in the Parallel Sysplex. As long as all the VTAMs that are supporting generic resources are configured as APPN nodes, the remaining systems can still be subarea only.
- If a VTAM system owns NCPs (running in 37xx controllers), then the VTAM must be configured as an Interchange Node.
- If you are migrating from an environment where VTAM systems are already set up as subarea nodes, then redefine the nodes to be interchange nodes and migration data hosts. Use VR-TGs to carry the control point sessions. This allows APPN communications, but with no changes to existing routing, and enables the simplest migration.
- If you are planning to add a number of new VTAM systems to your Parallel Sysplex, then try to make these "pure" end nodes. This allows you to make use of the APPN dynamic routing to add the new systems with the minimum

amount of definitions. It also enables you to make optimum use of the OS/390 cloning facilities.

- If there are any BSC3270 terminals in the network that require access to applications on VTAM systems in a Parallel Sysplex, the VTAMs must be defined as subarea nodes or APPN nodes with a subarea appearance. In addition, there must be a subarea path from the network to each VTAM system that the BSC3270 terminals need to talk to.

For more details on all of these points, and general advice on migration, refer to the *VTAM for MVS/ESA V4.4 Network Implementation Guide*, and the ITSO redbooks *Subarea Network to APPN Network Migration Experiences*, SG24-4656 and *SNA in a Parallel Sysplex Environment*, SG24-2113.

2.1.3 High Performance Routing (HPR)

High Performance Routing is an extension to APPN that provides a number of benefits:

- Nondisruptive path switching
- Improved performance
- Improved throughput in intermediate routing nodes

HPR support is provided in two parts:

HPR routing function A node that provides HPR routing can act as an intermediate node for an HPR connection. Any node that provides HPR support will provide this level of function.

In the HPR architecture, HPR routing is known as *Automatic Networking Routing* (ANR).

HPR endpoint function The HPR endpoint function is required at each end of an HPR connection. It is provided by some, but not all, HPR implementations. All nodes that support the HPR endpoint function must also support the HPR routing function.

In the HPR architecture, the HPR endpoint function is known as *Rapid Transport Protocol* (RTP).

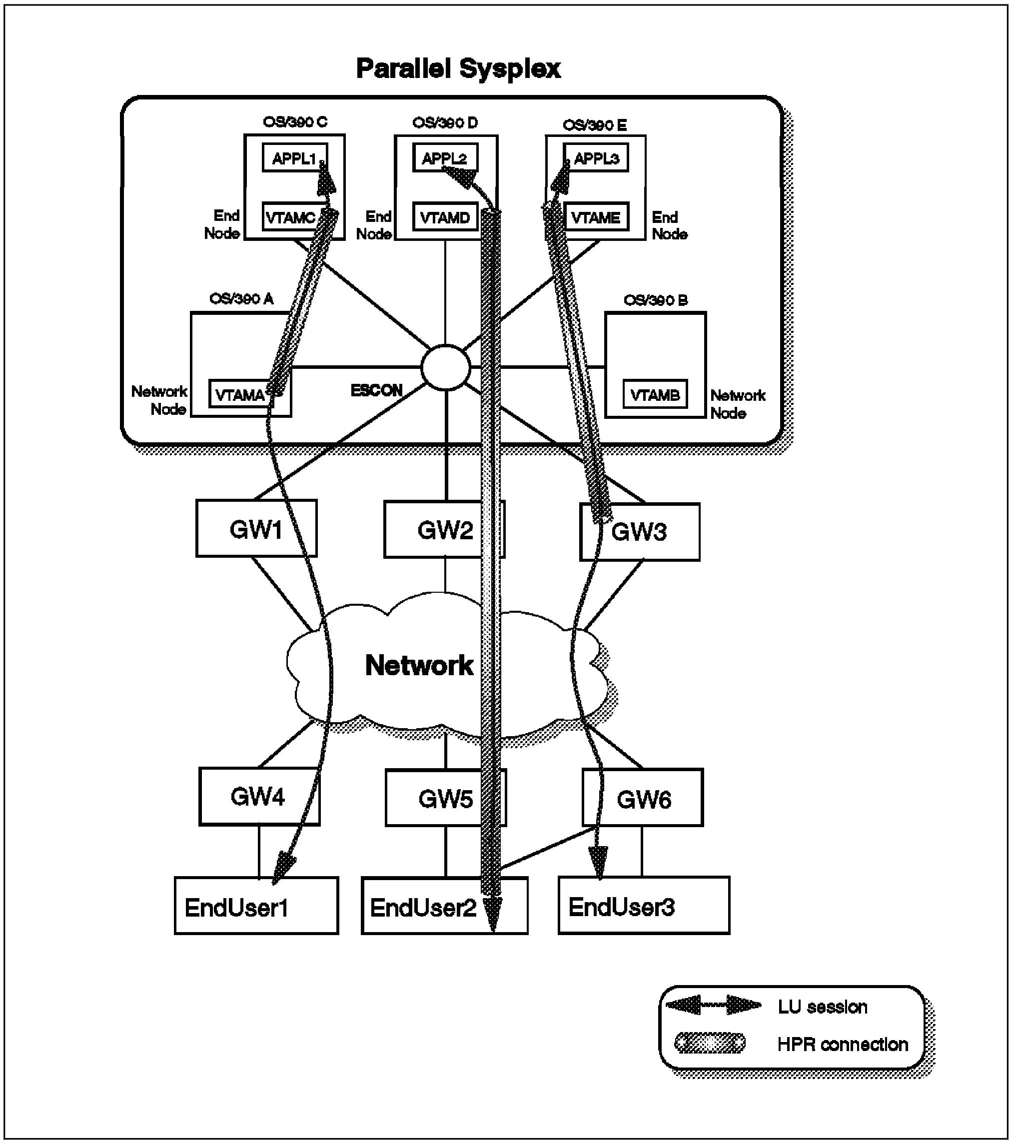


Figure 5. HPR Connections in a Parallel Sysplex

Figure 5 shows a number of different HPR configurations. You can see that HPR connections can carry the whole of an end-user session, or just a part of it.

- **Enduser1 Considerations**

  Enduser1's session cannot be supported over an end-to-end HPR connection. This could be because the workstation itself does not support the HPR endpoint function, or because the intermediate nodes GW1 and GW4 do not support the HPR routing function.

  Without HPR, this session may be established using a direct channel path from GW1 to VTAMC (not shown in Figure 5). The figure shows what will happen if we want to use Multi-Node Persistent Sessions. Multi-Node Persistent Session support requires an HPR connection from the VTAM end node. The assumption is that the selected route will use GW1 and GW4 and that this route does not support HPR in the network. In this case, to use a Multi-Node Persistent Session, VTAM itself must provide the HPR connection inside the Parallel Sysplex. So, the HPR connection will be made between VTAMC and one of the VTAM network nodes, VTAMA or VTAMB.

This connectivity supports Multi-Node Persistent Sessions.

- **Enduser2 Considerations**

  Enduser2 is using a workstation that supports the HPR endpoint function. When the session was established with APPL2 on VTAMD, it was determined that all the nodes along the path supported the HPR routing function. So the whole session is carried over an HPR connection. This offers the best possible availability, because if any intermediate node or line fails, HPR will dynamically move the HPR connection to a different path without disrupting the end user session. Note that the recovery can take place on the existing path, if the failing component (link or node) recovers in time.

  This connectivity supports Multi-Node Persistent Sessions. This support allows the HPR endpoint itself to be moved to a different end node, if OS390D (or VTAMD or the hardware on the image) fails.

- **Enduser3 Considerations**

  Enduser3's session cannot be supported over an end-to-end HPR connection. This could be because the workstation itself does not support the HPR endpoint function, or because the intermediate node GW6 does not support the HPR routing function. However, the front end node GW3 does provide the HPR endpoint function, and so an HPR connection can be set up between VTAME and GW3 to carry that part of the end user's session. This allows nondisruptive path switching around failures inside the Parallel Sysplex (for example, if a channel fails). However, Enduser3 will lose the session if GW3 or GW6 fails.

  This connectivity supports Multi-Node Persistent Sessions.

If the configuration in Figure 5 on page 28 really existed, with GW2 and GW3 supporting HPR and GW1 having no HPR support, then you should try to ensure that as many HPR-capable links as possible are chosen for Multi-Node Persistent Sessions. You can affect how much of the session is running over HPR by connecting users who need to access Multi-Node Persistent Session applications to the GW2 and GW3.

In a migration phase, of course, this may not always be possible, and so there will be users connected via gateways like GW1, which does not yet support HPR. VTAM always does its best to ensure that an HPR-capable link is chosen for at least the first hop of a Multi-Node Persistent Session. So the solution, in migration cases, is the route via a VTAM network node.

2.1.3.1 HPR Implementations

The HPR function is usually delivered as a software update and is available on most APPN platforms. The IBM products listed in Table 3 comprise the initial release that supports HPR:

Table 3 (Page 1 of 2). HPR Support in IBM Products

Product	HPR Routing Function (ANR)	HPR Endpoint Function (RTP)
Communication Server for OS/390 SNA Services	Yes	Yes
NCP V7.8	Yes	No
Communications Server for AIX V4.2	Yes	Yes

Table 3 (Page 2 of 2). HPR Support in IBM Products		
Product	**HPR Routing Function (ANR)**	**HPR Endpoint Function (RTP)**
Communications Server for OS/2 WARP V4	Yes	Yes
Communications Server for NT V5.0	Yes	Yes
Personal Communications 3270 V4.3 for Windows 95/98	Yes (1)	Yes
2216 M400	Yes	Yes
OSA-2	Yes(2)	No
OSA Express	Yes(2)	No
3174 Configuration Support Level C6	Yes (RPQ 8Q1594)	No
3746 Model 900/950	Yes	Yes
AS/400 (OS/400) R4	Yes	Yes

Notes:

(1) PCOMM for Windows can only run on an APPN end node. So it can do HPR routing, but it cannot act as an intermediate routing node.

(2) OSA itself does not support HPR, the support is provided by VTAM. In order to support an HPR connection from a VTAM node over an XCA connection such as provided by OSA-2, 3172 or some 2216 configurations, you will require at least VTAM V4R4 plus the additional support provided by APARs OW25950 and OW26732 and the minimum microcode levels for 3172/OSA-2/2216 as specified in this APAR. It's likely that anyone reasonably current in maintenance on these products will satisfy these prerequisites today.

2.1.3.2 High Performance Routing (HPR) for External Communications Adapter (XCA)

HPR is the technique used by APPN to deliver improved availability through non-disruptive network connections. External Communications Adapter (XCA) is one of the methods used by VTAM to connect to LAN-attached devices in SNA networks.

In the past, it was not possible to have HPR routes established over LAN connections defined using XCA major nodes. IBM has provided an enhancement to VTAM V4R4 which provides support for ANR routing over LAN-attached devices that are defined using XCA major nodes. As a result, both throughput and performance for LAN devices are improved. OSA-2 and OSA-Express both take advantage of HPR for XCA for both DIAL=NO and DIAL=YES connections. HPR is also supported over Ethernet connections. This capability is valid for both OSA-2 and OSA-Express features that support the SNA mode.

2.1.4 High Performance Data Transfer (HPDT)

The Communications Server for OS/390 provides High Performance Data Transfer (HPDT) facilities offering improved throughput for APPC applications.

2.1.4.1 High Performance Data Transfer (HPDT) for APPC Applications

To better capitalize on high-speed networking, the communications server introduces HPDT services and an HPDT interface to optimize performance for communications server APPC applications, particularly those that transfer larger blocks of data.

HPDT services are available to applications written to the communications server's APPCCMD interface and whose sessions connect two intraCPC applications or traverse one of the following high-bandwidth network attachments:

- XCF links between CPCs in a sysplex
- S/390 OSA-2 connected to a native ATM network
- S/390 OSA-Express in HPDT ATM Native mode
- APPN CPC-to-CPC channel connections
- IBM 2216 Multiaccess Connector Model 400
- Multiaccess Enclosure of IBM 3746 NWays Multiprotocol Controller Models 900 or 950

This interface allows applications requiring efficient bulk data transfer to gain performance improvements and entirely eliminate the data copy, because data is transferred between the APPCCMD and VTAM. A communication storage manager allows VTAM and applications to exchange ownership of commonly addressable storage, so there is no need to copy data at the APPCCMD application programming interface.

No application change is required to realize performance benefits.

For more information, refer to *VTAM for MVS/ESA V4.4 Release Guide*.

2.1.5 Session Manager Implications in Parallel Sysplex

Session managers are used to provide a customized front-end to SNA applications, and they often also allow multiple active sessions. They are run as VTAM programs, and can act in two different ways:

- Pass mode
- Relay mode

In *pass mode*, the end-user session is connected directly between the end user and the application. The session manager is only involved when the session is established, and then again when the session is taken down. There are no special implications for session managers using pass mode in a Parallel Sysplex environment.

In *relay mode*, there are actually two sessions for each end user. There is a session from the end user to the session manager, and a second session from the session manager to the application. Most session managers operate this way, as this is the method used when they implement multiple sessions. In a Parallel Sysplex environment, there can be problems with session managers that use relay mode, if you are also using generic resources.

The placing of the session manager is of concern. In an environment with a large number of VTAM systems, it is usual to run one or two copies of the session manager. You would normally place a session manager on a VTAM network node, although it is technically possible to put it on an end node.

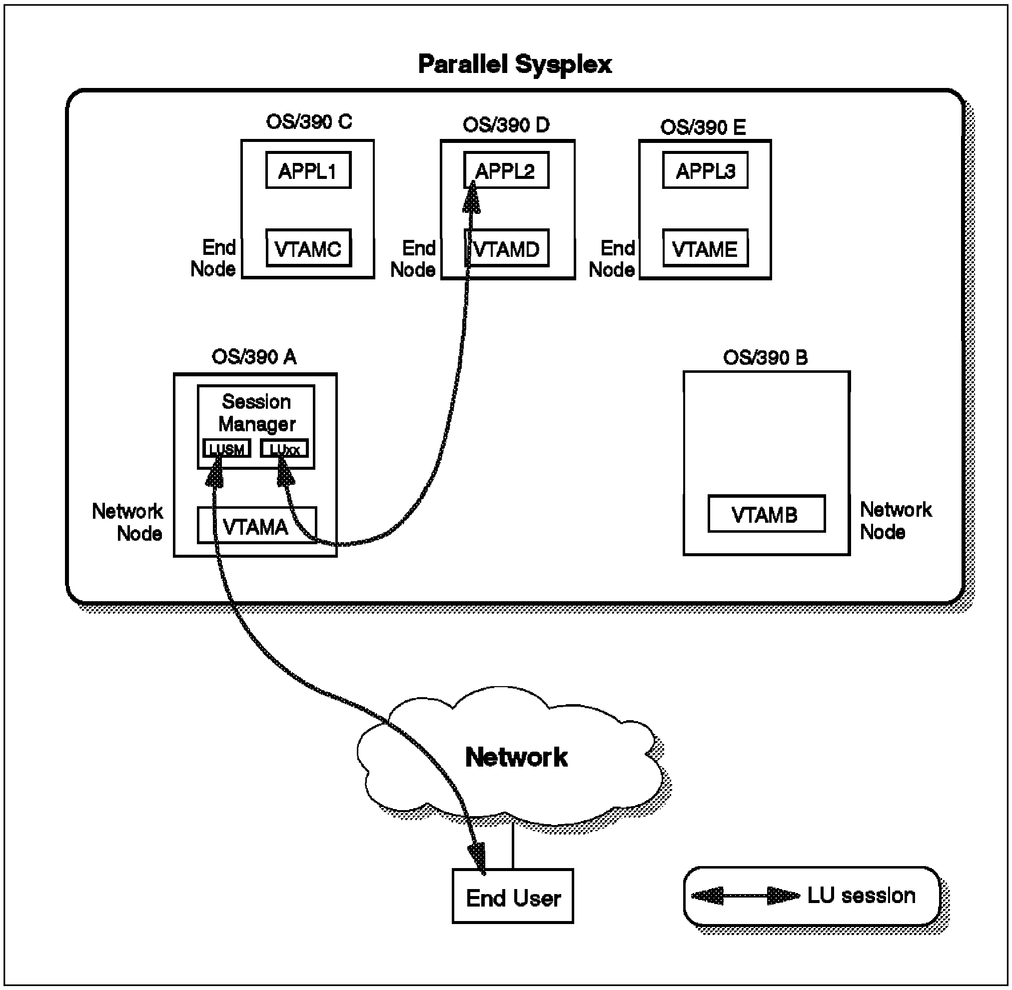


Figure 6. Session Managers in a Parallel Sysplex. Implications of Using Relay Mode

Figure 6 shows a configuration where the session manager in relay mode is running on OS/390 A. The end users have a session (or multiple sessions) with the session manager LUSM. Then, for each remote end user LU, the session manager will create a local LU on VTAMA, which then has a session with the application. In the diagram, these local LUs are shown as LUxx.

In this configuration, it is important that no actual application instances of generic resources are located on VTAMA. This is because all the end-user LUs (LUxx) appear to VTAMA as local LUs. If you define a generic resource that does use an actual application instance on VTAMA, then VTAMA would always use this local application instance. Consequently, the other application instances (APPL1, APPL2, APPL3 and so on) will never be chosen, and session balancing will not work. Refer to "Local Access to Generic Resources" on page 225 in Volume 1: Overview, SG24-5637 for more details.

In this environment, the session manager node is always a single point of failure. If the session manager fails, or if the VTAM network node fails, then all the end sessions are lost. In addition, if there are two session managers, for availability, the end user will have to know the name of the alternate one to use after a failure.

The solution to this single point of failure is for the session manager to support generic resources itself. In this case, when sessions are lost after a failure, the user can log on again to the same generic name, and the sessions will be with an alternate session manager in the same generic resource.

Just before this book was published, support for Netview Access Services (NVAS) to both access generic resources and for NVAs itself to be defined as a VTAM GR. More information about this support is contained in Information APAR II12403.

In addition, Computer Associates has a session manager product called "Teleview" that allows logons on applications defined as VTAM generic resources, as well as being capable of being defined as a generic resource itself. More information about Teleview may be found at:
`http://www.cai.com/products/announcements/teleview_mvs.htm`

2.1.6 Communications Management Configuration

The *communications management configuration* (CMC) is a description for a configuration that is often used in networks with large numbers of NCPs running in 37xx controllers. The CMC concept is based on the following principles:

CMC One VTAM system is responsible for the ownership of all NCPs in the network, both local and remote.

Backup CMC It is usual to have another VTAM system, in a different part of the network, that can take over ownership of the NCPs in case the primary CMC fails.

Data host The rest of the VTAM systems will be defined as data hosts. This means they can run applications, but they do not have responsibility for ownership of NCPs.

This configuration can increase availability for the end users. If a system where applications are running goes down, the rest of the network will be unaffected. Only sessions to applications on that system will be lost. This configuration has been used for many years, especially for large networks.

The CMC configuration can still be used if a Parallel Sysplex is implemented. The CMC itself can be placed inside or outside the Parallel Sysplex.

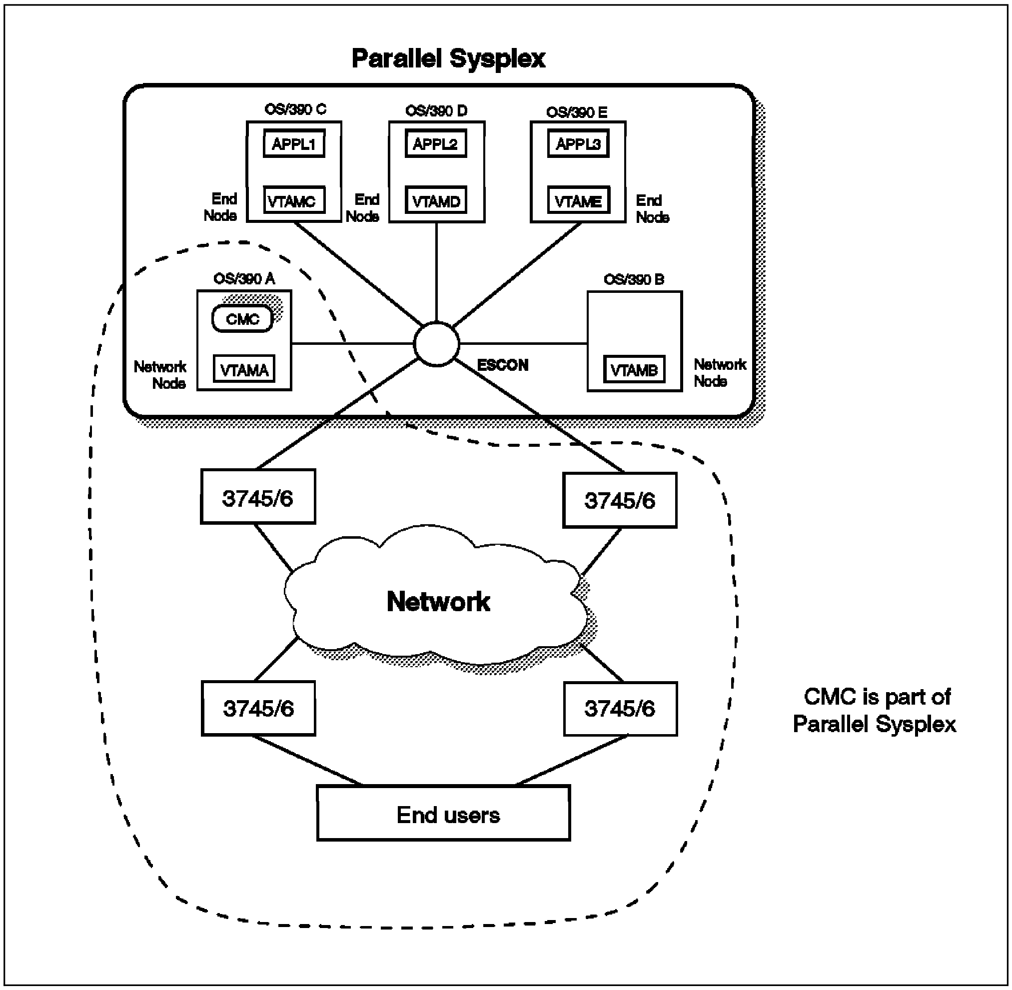


Figure 7. CMC Is Part of Parallel Sysplex

Figure 7 shows a configuration where the CMC is one of the VTAM network nodes in the Parallel Sysplex. This configuration will be applicable if there are a small number of VTAM systems, or it is not acceptable to add extra images purely for the network node function. The CMC is reconfigured to become a network node, and acts as network node server for the served VTAM end nodes. This means that the CMC will operate as an interchange node, because it is an APPN network node, and it owns NCPs.

2.2 NCPs in a CMC Environment

The main planning activity in this configuration is to decide how the local NCPs should be configured to attach to the channel-attached VTAM nodes. This is discussed in detail in 2.2.1.1, "APPN Connectivity for NCP-Owning VTAMs" on page 36.

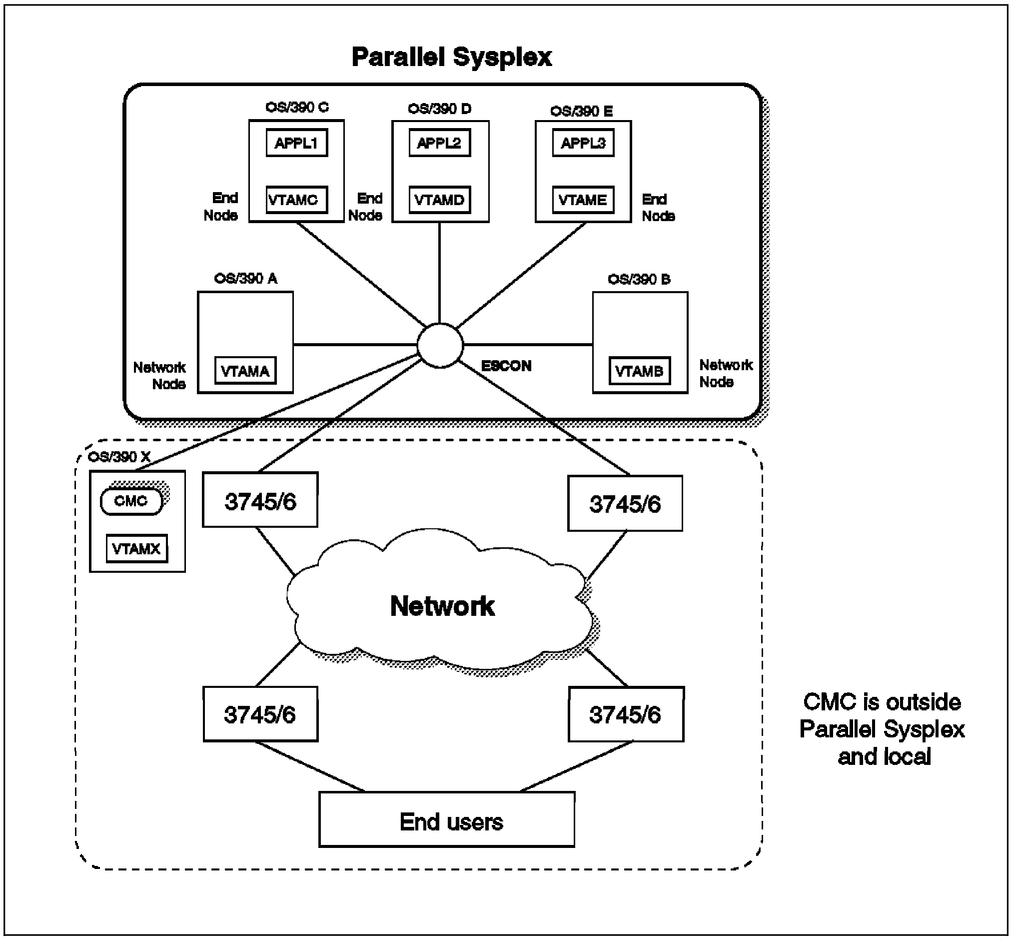


Figure 8. CMC Is Outside Parallel Sysplex with Local Connectivity

Figure 8 shows a configuration where the CMC is outside the Parallel Sysplex, but has channel connectivity to the Parallel Sysplex systems.

This is a valid configuration, but there may be some connectivity implications if the CMC VTAMx is non-APPN and the Parallel Sysplex systems are using APPN. If VTAMx is APPN, then you must decide how the local NCPs should be configured to attach to the channel-attached VTAM nodes. Refer to 2.2.1.1, “APPN Connectivity for NCP-Owning VTAMs” on page 36 for more details.

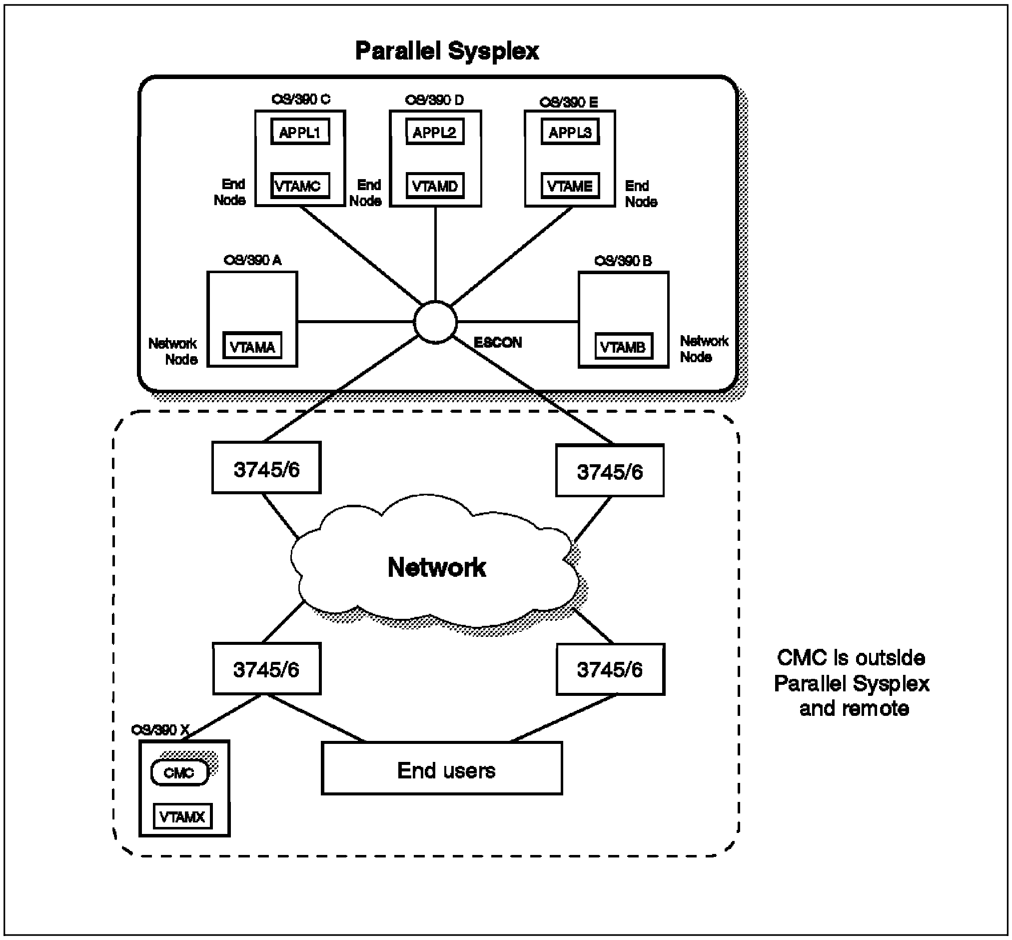


Figure 9. CMC Is Outside Parallel Sysplex and Remote

Figure 9 shows a configuration where the CMC is outside the Parallel Sysplex, and remote from it. Therefore, there is no channel connectivity to the Parallel Sysplex systems.

This is also a valid configuration, but there may be some connectivity implications if the CMC VTAMx is non-APPN and the Parallel Sysplex systems are using APPN. If VTAMx is APPN, then you must decide how the local NCPs should be configured to attach to the channel-attached VTAM nodes. Refer to 2.2.1.1, "APPN Connectivity for NCP-Owning VTAMs" for more details.

2.2.1.1 APPN Connectivity for NCP-Owning VTAMs

This discussion applies to the CMC configuration where a single VTAM system (possibly with a backup system) is responsible for all NCP ownership. However, the discussion also applies to any VTAM system that has NCP ownership and so is relevant for installations who share NCP ownership among a number of VTAM systems. For convenience, the term CMC is used to mean any VTAM system which owns NCPs.

CMC Is Inside Parallel Sysplex: When the CMC is also one of the Parallel Sysplex network nodes, it must be configured as an interchange node. It will have control point sessions to the VTAM end nodes that can run over either VR-TGs or APPN links. These control point sessions will run over channel connections.

The local channel-attached NCPs do not necessarily have to be APPN-capable. The channel links from the local NCPs can be configured to support APPN connections to the VTAM end nodes, or they can remain as subarea connections. This is discussed in "Connectivity for Channel-Attached NCPs" on page 38.

CMC Is Outside Parallel Sysplex: There are a number of configuration possibilities:

1. The CMC is configured as an interchange node, and it has control point sessions with the VTAM network nodes inside the Parallel Sysplex over APPN connections.

   The local channel-attached NCPs must be APPN capable, and are configured to support APPN T2.1 connections over the channel. The implications of this are discussed in "Connectivity for Channel-Attached NCPs" on page 38.

   This configuration provides the simplest solution, because APPN dynamics can be used to replace the definitions used in the subarea configuration. However, the growth in NCP storage may be unacceptable.

2. The CMC is configured as an interchange node. It has control point sessions with the VTAM network nodes inside the Parallel Sysplex over subarea connections, using VR-TG.

   The local channel-attached NCPs do not have to be APPN-capable. They can be down-level on hardware or software, and the channel connections to the VTAM end nodes use subarea protocols. This is discussed in "Connectivity for Channel-Attached NCPs" on page 38.

   This configuration provides a good solution where there is a large existing network. It is not always possible or desirable to change existing devices, and the VR-TG allows the use of APPN without changing the subarea network.

3. The CMC is configured as a subarea node.

   It has no APPN appearance, and therefore can only have subarea connection to other VTAMs.

   This configuration may result in less than optimum routing for certain sessions. This can happen when there are both APPN and subarea routes to the VTAM end nodes (which are actually configured as migration data hosts). In the ideal case, the end-user session will be routed across the network, through the local 3745/6 and then across the ESCON channel to the appropriate VTAM system. However, in some cases, the session will be routed across the network, through the local 3745/6, through the VTAM network node, and then to the appropriate VTAM system.

   This has implications for availability, because each session passes through an additional node. It is possible that in this configuration, some sessions use the optimum route across the channel, while others route through the VTAM network node. The reason for this difference has to do with whether the search for the application is done using APPN or subarea logic.

All of these CMC configurations are valid. However, the third configuration can cause less than optimum routes to be selected, so it is always recommended to upgrade the CMC to be APPN-capable.

Recommendation for Planning Purposes

> If you are using a CMC configuration, always upgrade the CMC itself to be an APPN interchange node.

Connectivity for Channel-Attached NCPs: If the CMC or NCP-owning system has been upgraded to be APPN-capable, then you have the option of migrating the channel connections from the local NCPs to the VTAM end nodes to be APPN links, or leaving them using subarea protocols. If the CMC is not APPN-capable, then the channel connections from the local NCPs must continue to use subarea protocols.

These options are discussed below:

- Local channel connections use APPN

  The NCPs that the CMC owns may be configured to support APPN. In this case, NCP V6.2 and upwards is required, which implies the use of a 3745 communications controller. There will also be implications for 3745 resources, such as storage and engine utilizations. This is also discussed in 2.4.1.3, "3745 Software" on page 59.

  The definitions for the channel connections between the NCPs and the VTAM end nodes are now for APPN T2.1 links, rather than subarea links. Consequently, they are much simpler to define, since APPN links have very few parameters and do not require predefined routing. In the future, it will be easier to add new nodes into the Parallel Sysplex without having to do a subarea routing exercise. This configuration also provides the opportunity to do HPR routing to support Multi-Node Persistent Sessions.

  VTAM V4.4 supports HPR over VR-TGs. Therefore, using VR-TGs to the sysplex end nodes supports MNPS sessions.

  One restriction in this configuration is that BSC3270 connections are not supported over APPN links. So, if you still use BSC3270, you should not migrate your channel connections to be APPN links.

- Local channel connections use subarea protocols

  The NCPs that the CMC owns may continue to use subarea protocols. The NCP subarea definitions for the connections to the VTAM systems are unchanged. The CMC will use VR-TGs to communicate with the VTAM systems. In this case, there is no requirement for a specific version of NCP, and no growth in 37xx resources.

  However, this does mean that new subarea routes have to be defined as new images are added into the Parallel Sysplex, and this has implications for all VTAMs and NCPs in the network.

Recommendation for Planning Purposes

> To avoid increased storage and CCU usage in NCP, it is recommended to use VR-TGs between the CMC and the Parallel Sysplex VTAM systems.

2.2.2 SNI Considerations

SNA Network Interconnection (SNI) is widely used in subarea networking to connect networks with different network identifiers (netids). SNI gateways are provided by a combination of functions in VTAM and NCP. At a minimum, one NCP is required for an SNI connection.

There are a number of configurations to be considered in a Parallel Sysplex environment:

- SNI Gateway Outside Parallel Sysplex

  If an SNI gateway is outside the Parallel Sysplex, it is possible for remote end users from a different network to establish sessions with applications inside the Parallel Sysplex. The GWSSCP function will be provided by a VTAM outside the Parallel Sysplex, for example by an external CMC.

- SNI Gateway at Parallel Sysplex Boundary

  It is also possible for an SNI gateway to be provided at the Parallel Sysplex boundary. For example, the VTAM network nodes and end nodes can be part of one netid, with the rest of the network outside the Parallel Sysplex in a different network. The GWSSCP function will be provided by one of the Parallel Sysplex network nodes, which will be configured as an interchange node.

- Different Networks Inside the Parallel Sysplex

  It is also possible to configure different netids within the Parallel Sysplex. The VTAM end nodes can have different netids from the VTAM network nodes. This is a standard APPN feature, and does not require SNI at all. However, this configuration *cannot* be used for generic resources.

  Generic Resources and MNPS Requirements for Netids

  The generic resources function requires that all the VTAM systems in the sysplex are part of the same network (use the same netid).

  The MNPS function requires that all the end nodes in the sysplex have the same netid. However, contrary to the generic resources case, MNPS allows the sysplex ENs to have a different netid from the network nodes, if desired. This configuration does not support generic resources.

  To avoid any potential conflicts, it is recommended that you configure all the VTAM systems inside a Parallel Sysplex with the same netid.

- An alternate approach is to use an APPN border node for cross-network communication, but this requires that both networks implement APPN.

2.2.3 Enterprise Extender

One of the predominant trends in the industry is the move towards IP backbone networks. However, most S/390 applications are still SNA-based. In most installations therefore, it is likely that there is, or will be, a requirement to access SNA applications from networks that are running the IP protocol.

Initially, this requirement was addressed by AnyNet, but that product has effectively been replaced by Enterprise Extender for most situations. Enterprise Extender allows you to route SNA traffic over an IP network. Enterprise Extender is used for APPN resources, and is positioned as a replacement for some of the earlier solutions, where those solutions are used with APPN. Non-APPN traffic must continue to use one of the previous solutions. Similarly, if you wish to route IP traffic over an SNA backbone, then you can use Anynet—this function is not provided by Enterprise Extender.

One of the common techniques for carrying SNA applications over an IP backbone network is Data Link Switching (DLSw), used by many vendors. However, DLSw has disadvantages in the Parallel Sysplex environment for the following reasons:

- DLSw does not support HPR and so availability is compromised if using MNPS or APPN/HPR without MNPS. Switching of MNPS sessions can only be done inside the Parallel Sysplex, thus incurring an additional VTAM hop on the session path. End-to-end high availability can never be achieved.
- DLSw uses TCP which is connection-oriented and so the routers at the edge of the IP network (where the DLSw conversion is performed) are single points of failure. If either of these routers fails, then the end-user session is always disrupted.
- DLSw cannot handle the SNA Class of Service and so SNA interactive traffic can suffer degradation of performance in a congested IP backbone network. To get around this, it is possible to prioritize *all* SNA traffic over IP traffic. This results in SNA batch traffic being treated the same as SNA interactive traffic, but it provides a workable solution in most cases.

Because of these restrictions, the APPN architecture has been enhanced with a function called Enterprise Extender which is an alternative to DLSw when carrying SNA applications over an IP backbone network. The Enterprise Extender function is placed in an edge router (as is DLSw), but it uses a very different technique which removes the three restrictions of DLSw explained above.

Instead of using TCP to carry the SNA traffic, Enterprise Extender uses the UDP protocol. The SNA Class of Service is mapped into a UDP port number and this port number is carried in the packet header. All routers today can prioritize the UDP port number and so it is possible to prioritize SNA even inside the IP backbone network. This means that no new hardware is required in the backbone network to support Enterprise Extender.

One key benefit of Enterprise Extender is that is uses APPN/HPR. This is crucial for the Parallel Sysplex environment where high availability is required. Using APPN/HPR end-to-end provides the support for non-disruptively rerouting SNA sessions around failures in the network.

The Enterprise Extender function is normally provided by Communications Server for OS/390 (requires OS/390 Version 2.6 with enabling PTF) at the central site and on the remote gateway/router at the branch. This means that APPN/HPR is used between the Communications Server/390 and the branch. However, in the high availability environment, MNPS uses APPN/HPR as a minimum inside the Parallel Sysplex. So, the combination of MNPS and Enterprise Extender means that APPN/HPR is extended from the mainframe out to the remote branch, even with an IP network. All network devices between the two Enterprise Extender nodes only have to be capable of supporting IP routing. From an APPN/HPR perspective, the IP network is a single APPN hop, regardless of how many IP router nodes are actually traversed, and each of these routers only sees the data as a UDP datagram. So, if Enterprise Extender is running at one end on OS/390, the network gateway out of the Parallel Sysplex only needs to suppport IP routing. Cisco provides support Enterprise Extender in conjunction with their BrNN implementation in the levels of their code current at the time of writing. This makes Enterprise Extender an attractive solution for installations with Cisco router networks.

If APPN/HPR is moved further out into the branch workstations, then full end-to-end high availability can be achieved. Many workstation platforms today support the APPN/HPR, both on APPN network nodes and APPN end nodes.

Finally, the use of APPN/HPR means that the gateways/routers implementing the Enterprise Extender are not single points of failure. If the nodes behind the Enterprise Extender nodes support APPN/HPR (for example, VTAM in the Parallel Sysplex or a Comms Server platform on the workstation), and if alternate Enterprise Extender nodes are available, then APPN/HPR will non-disruptively switch traffic around failed gateway/routers. Thus there are no single points of failure.

For these reasons, the Enterprise Extender function is strongly recommended when high availability is required for transporting SNA applications over an IP backbone network in the Parallel Sysplex environment.

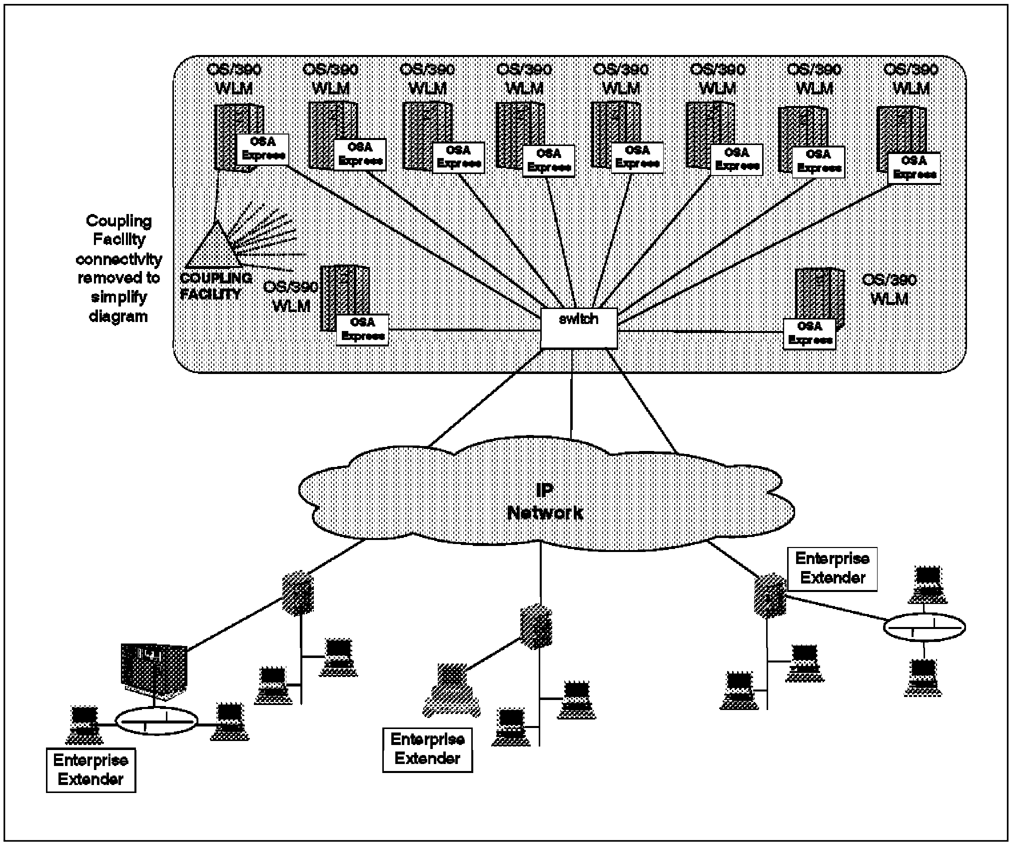


Figure 10. Use of Enterprise Extender in a Mixed SNA/IP Environment

Figure 10 shows two examples of SNA sessions using the Enterprise Extender technology:

- The session on the left shows what can be achieved if all the products involved in the session path provide the full APPN/HPR support. There is an HPR connection from the S/390 server in the Parallel Sysplex all the way down to the SNA workstation, even though there is an IP backbone network.

This configuration provides an excellent solution for SNA sessions using MNPS for high availability. Assuming that backup nodes and paths exist, it is possible to bypass any hardware failures in the network or the Parallel Sysplex and non-disruptively move the SNA session to a new path. There are a number of assumptions here:

1. The S/390 servers in the Parallel Sysplex support both HPR RTP and the MNPS function.
2. The remote workstation supports the HPR RTP endpoint function.
3. The Enterprise Extender function is required in the remote router and on the S/390 Communications Server/390 (Ver 2.6 with enabling PTF).
4. Routers in the IP backbone network must provide UDP port prioritization.

- The session on the right shows the minimum configuration for Enterprise Extender. This configuration provides the ability to support the transport of SNA applications over the IP backbone network, without providing full end-to-end HPR support for high availability. Part of the LU session is carried over an HPR connection. The rest of the session is routed using subarea SNA or APPN routing.

2.3 Network Access to Parallel Sysplex Using TCP/IP

Two aspects to be considered when deciding how to connect your TCP/IP users to your S/390 applications are availability and workload balancing.

The task of workload balancing in a IP environment is not nearly as straightforward as in an SNA environment, due to the wide variety of IP workload distribution choices. To choose the best method for your network, it is important to understand your current and future network requirements as well as your network configuration.

Similarly, from an availability point of view, there are different things to consider. TCP does not have a facility like SNA's MultiNode Persistent Sessions, however, there are enhancements that can help minimize the impact of an application or system outage. Also remember that the ability to do dynamic workload balancing impacts availability—if you have a number of instances of an application running, dynamic workload balancing can help hide the unavailability of one of those instances from the users.

In this section, we review the enhancements to TCP/IP over the last few releases of OS/390, and discuss the considerations for each one.

2.3.1 Availability Enhancements

There are two aspects to the availability of an application from a network point of view:

1. You want to have dynamic workload balancing so that new users logging on will be shielded from the fact that one instance of the application has failed. This support is provided by the various workload balancing features of TCP/IP and OS/390.
2. You also want to be able to restart failed applications as quickly as possible, in as automated a manner as possible. If the system the application was running on is still available, the application should be restarted there. If not, you want to be able to either move the application (and its associated IP

address) to another system, or else to just move the IP address to another system and have it associated with a clone of the failed application.

This latter requirement is addressed by the Virtual IP Addressing capability of TCP/IP and by subsequent enhancements based on that capability. These are discussed in the following sections.

2.3.1.1 Virtual IP Addressing

An IP network provides nondisruptive rerouting of traffic in the event of a failure, but only within the routing network itself, not at the endpoint hosts. For most client hosts (PCs or workstations), failure of the host or the network adapter or the link will just isolate the client application from the network, if it does not take down the client application altogether.

For servers, on the other hand, particularly large-capacity and highly scalable servers such as OS/390, it is extremely common to have more than one link into an OS/390 image and its associated IP stack. While connections may be distributed among the various links and adapters, failure of one such will mean loss of all TCP connections associated with the failing device or link, because the TCP connection is in part defined by the IP address of the failed adapter.

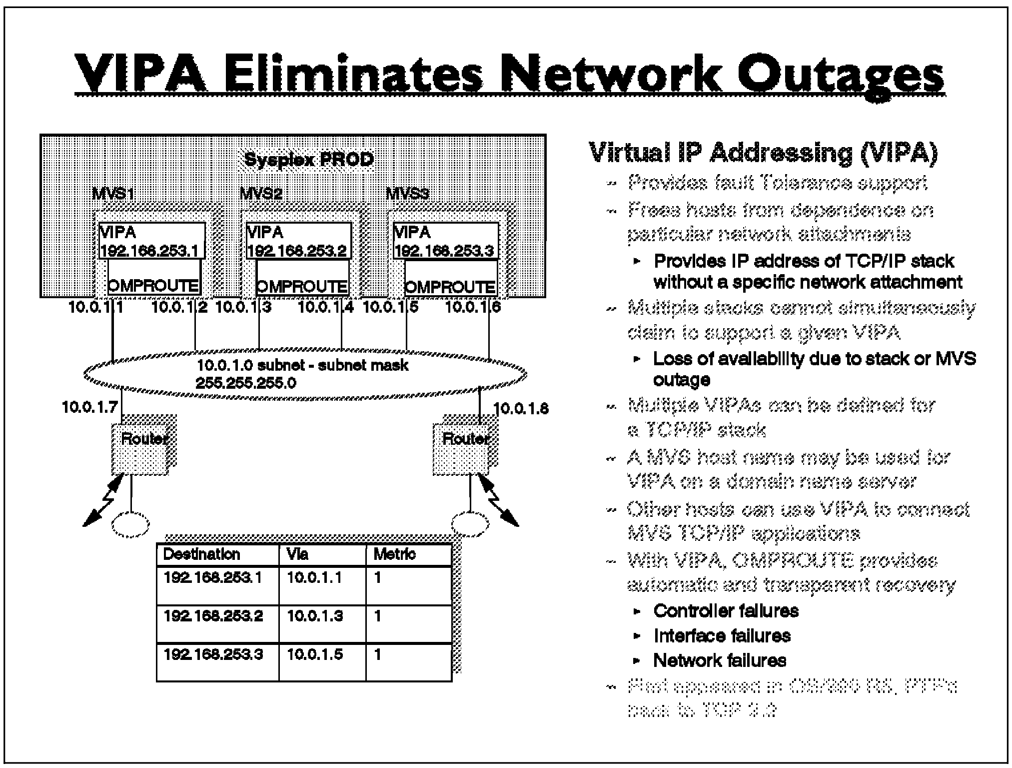


Figure 11. VIPA Configuration

CS for OS/390 addresses the requirement for nondisruptive rerouting around a failing network adapter by allowing the customer to define a virtual adapter with an associated virtual IP address (VIPA). A virtual adapter (interface) has no real existence, and a VIPA is really associated with the stack as a whole. To the routers attached to the stack via physical adapters, a VIPA appears to be on a subnet on the other side of the OS/390 IP stack, and the TCP stack looks like another router.

On the OS/390 IP stack, on the other hand, the VIPA acts somewhat like a loopback address: incoming packets addressed to the VIPA are routed up the stack for handling by TCP or UDP as with any other home IP interface. Dynamic routing protocols can now provide transparent rerouting around the failure of an adapter on the endpoint stack, in that the VIPA still appears reachable to the routing network via one of the other adapters.

In the example in Figure 11 on page 43, the host MVS1 has a VIPA address of 192.168.253.1. The client knows that address, and sends the address in the request to the router. The router uses 10.0.1.1 to reach that address. However, if 10.0.1.1 is unavailable for some reason, the router will retry on 10.0.1.2 instead. If the client had used the real IP address instead of the VIPA, the router would not have been able to redirect the requests following the failure of 10.0.1.1.

VIPA was introduced by TCP/IP V3.2.

2.3.1.2 Dynamic VIPA Takeover

While VIPA removes a single hardware interface and the associated transmission medium as a single point of failure for a large number of connections, the connectivity of the server can still be lost through a failure of a single stack or an MVS image. Of course, we can move a VIPA manually to the other stack, but customers require automatic recovery wherever possible, especially in a sysplex. Therefore, CS for OS/390 Release 8 provides improvements to the VIPA takeover function. VIPA takeover builds on the VIPA concept, but automates the movement of the VIPA to an appropriate surviving stack. There are two forms of VIPA takeover:

- Automatic VIPA takeover allows a VIPA address to move automatically to a stack where an existing suitable application instance already resides, allowing that instance to serve clients formerly going to the failed stack/node.
- Dynamic VIPA for an application server allows VIPAs to be defined and activated by individual applications (with or without modifying the applications), so that the VIPA moves when the application is moved.

This means that application instances on a failing cluster node may be distributed among many (or all) of the surviving nodes, reducing the amount of spare processing capacity that must be provided for each node that may participate in takeover. To support these new enhancements, CS for OS/390 Release 8 now provides three ways to define a VIPA:

- Traditional static VIPA definition, as in Release 7
- Dynamic VIPA definition for automatic VIPA takeover
- Dynamic VIPA activation by application action

The last two ways are new in Release 8 and are configured via the new VIPADYnamic block in the TCP/IP profile.

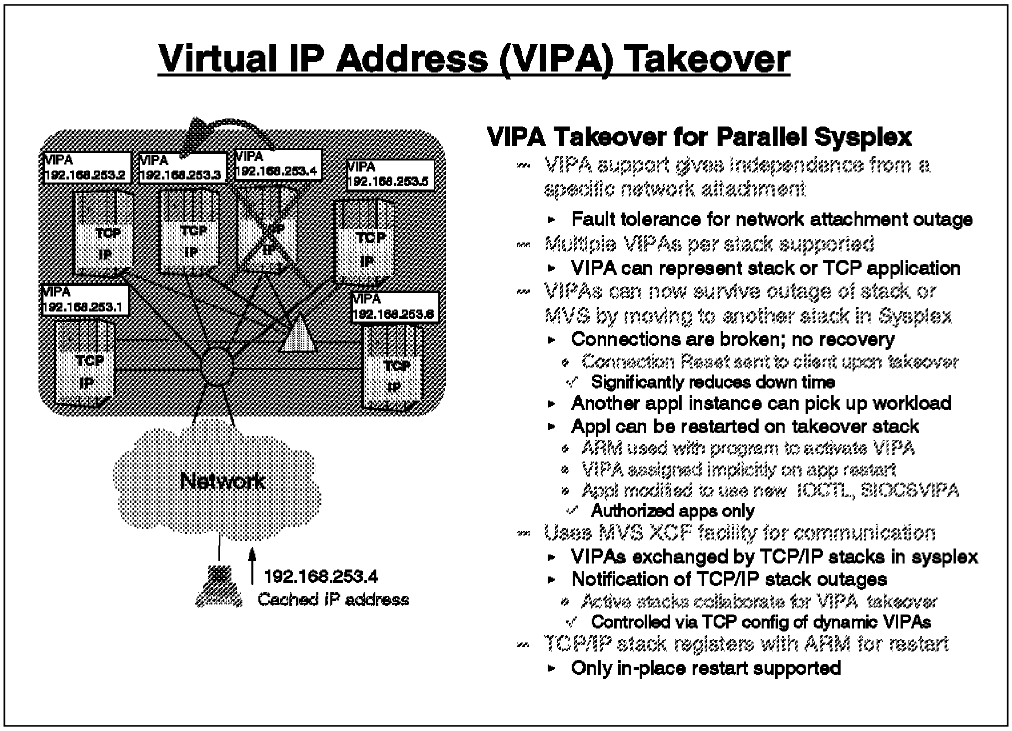


Figure 12. Dynamic VIPA Takeover Support

Figure 12 shows an example of how an application on a failed image is restarted on another image, and automatically reacquires its VIPA, following which all client requests for that IP address will be routed to the new application instance on the alternate system.

More information on the use of Dynamic VIPA may be found in the redbook *TCP/IP in a Sysplex*.

2.3.1.3 Non-Disruptive VIPA Takeover

One of the restrictions of VIPA is that only one application may use a Dynamic VIPA at a time. This is not a problem in the case of a failure—the failed application will be restarted somewhere else, and there will still only be one user of the VIPA. However, when the original instance of the application is restarted, it cannot use the VIPA as it is already in use elsewhere. The only way to move the VIPA back to its original location was for the alternate instance to close the VIPA (and thereby disrupt all connections) before the restarted instance tries to open it. This results in a period of time when the application is not available to users.

This restriction has been addressed by a further enhancement to VIPA OS/390 V2R10. It is now possible to set up a VIPA so that more than one application can use the Dynamic VIPA at the one time. As a result, all new connections will be routed to one instance of the application, while existing connections can continue to use a different instance. This new capability is referred to as Non-Disruptive VIPA Takeover.

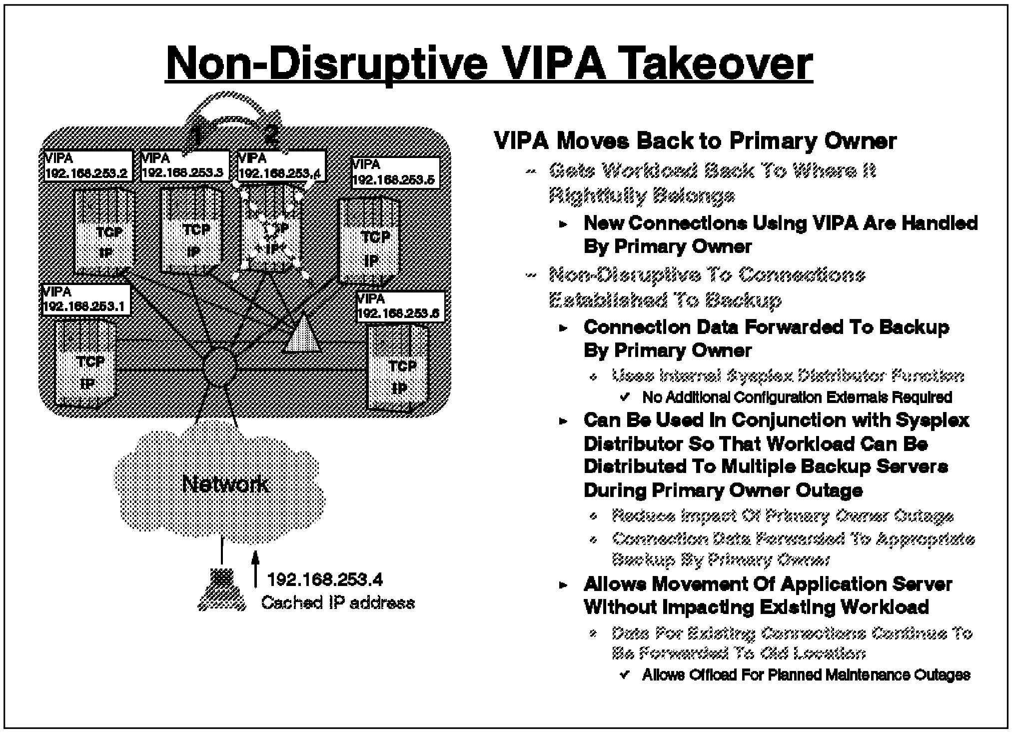


Figure 13. Non-Disruptive VIPA Takeover Support

The picture in Figure 13 shows how the VIPA is moved back to the restarted application, but existing connections continue to be serviced by the backup application instance. All new connection requests will be routed to the restarted primary instance of the application as soon as that instance opens the VIPA.

2.3.2 Workload Balancing Enhancements

By making the Parallel Sysplex look like a single system to the client, the burden of knowing the physical makeup of the data center and choosing the appropriate application instance is shifted from the client to the Parallel Sysplex itself. The Parallel Sysplex provides a single system image by having either a single name which represents the replicated application instances (DNS method), or by a single IP address which represents the replicated application instances (cluster method). By allowing applications to be moved or added to new locations within the sysplex and still be considered part of the same single system image application group, the data center is allowed to grow without impact to the other sysplex members, the WAN or the clients.

In this section, we will briefly describe the workload balancing enhancements in each of the recent OS/390 releases that support both the DNS method and cluster methods. However, for the latest information, and detailed implementation considerations, you should refer to the redbook *TCP/IP in a Sysplex*, and the CS for OS/390 library for your release of OS/390.

2.3.2.1 TCP/IP Utilization of XCF

The TCP/IP stack in OS/390 V2R4 and later, supports XCF connections between OS/390 systems. APAR PQ04890 allows TCP/IP to route IP across VTAM MPC and XCF connections. This is part of the move towards a common Communications Server for OS/390 data link controls (DLCs).

2.3.2.2 Dynamic XCF Support

XCF Dynamics can be used to generate dynamic definitions for TCP/IP stacks that reside on another OS/390 MVS host in a sysplex and for additional TCP/IP stacks that reside on the same OS/390 MVS host. At initialization, each TCP/IP stack configured for XCF (via activation of the VTAM XCF major node, ISTLSXCF) joins a well-known XCF group. When other stacks in the group discover the new stack, the definitions are created automatically, the links are activated, and the remote IP address for each link is added to the routing table. After the remote IP address has been added, IP traffic proceeds as usual.

Dynamic XCF support, introduced in OS/390 V2R7, provides greater flexibility and reduced administration in a sysplex environment, and provides a base that will be used in future releases of OS/390.

Further information about Dynamic XCF support, and how to set it up, is available in *OS/390 Communications Server IP Configuration Guide*.

2.3.2.3 TCP/IP Connection Optimization

Connection optimization provides intelligent sysplex distribution of requests through cooperation between WLM and the Domain Name System (DNS) name server.

For Those Unfamiliar with the World of TCP/IP

The Domain Name System provides a way of mapping IP addresses to domain names, which can be more easily remembered by users. When a client connects to a server, it can use a name instead of having to know the IP address. To do this, the TCP/IP software in the client sends a request to a name server for the IP address of a given name. The name server will return the IP address that maps to that name, and the client can then connect to that address.

If you elect to place a name server in an OS/390 sysplex, DNS will invoke WLM sysplex routing services to determine the “best” system to service a given client request. This is functionally similar to the VTAM Generic Resources support for SNA networks in the Parallel Sysplex environment.

With connection optimization, the sysplex name is a subdomain that you add to your DNS name space. Groups of server applications or hosts can now be given a name (group name) and the name server will return a TCP/IP address based on round-robin logic and load-based ordering. Clients must use a name in the form groupname.sysplex_domain_name or sysplex_domain_name when connecting to servers in order to use connection optimization. To achieve maximum availability, both TCP/IP stacks and server applications should register themselves to WLM. Currently, the TCP/IP stack, TN3270 server, the FTP server and DB2 for OS/390 V5 and later, provide WLM registration. At the time of writing, a future release of CICS is also expected to provide support for WLM registration. Keywords are provided (SYSPLEXROUTING and WLMCLUSTERNAME) for the TCP/IP stack and TN3270 server.

The DNS/WLM option requires that name servers that are located outside the sysplex cannot be configured as primary or secondary name servers for the sysplex domain. Also, name servers outside the sysplex should not be allowed cache information about sysplex domain resources, because this defeats the purpose of connection optimization. To achieve this, the name server returns a time-to-live (TTL) value of zero by default for sysplex domain resources. In a large network with many name servers this may result in increased network traffic. It is possible to change the default TTL for sysplex domain resources, but this must be balanced against the needs of real time load-balancing and availability.

For a full explanation of connection optimization together with examples of its use, see *OS/390 Communications Server IP Configuration Guide*.

2.3.2.4 TN3270 Server

The TN3270 function provides 3270 terminal emulation from a workstation that runs only TCP/IP. Telnet works by using two components:

- The TN3270 client function operates on the end-user's workstation. It provides a 3270 session appearance to the end user, but carries the traffic over a TCP connection.
- The TN3270 server function operates somewhere between the TN3270 client and the 3270 application on the S/390 server. It provides an LU for each TCP connection from a remote TN3270 client.

The TN3270 server provides an interesting example of using both TCP/IP and VTAM to provide load balancing and high availability in a Parallel Sysplex environment, as shown in Figure 14 on page 49. Connection optimization is used in this scenario to load balance across TN3270 servers on OS/390 A and OS/390 B. The placement of the TN3270 server function has implications on the workload balancing in a Parallel Sysplex.

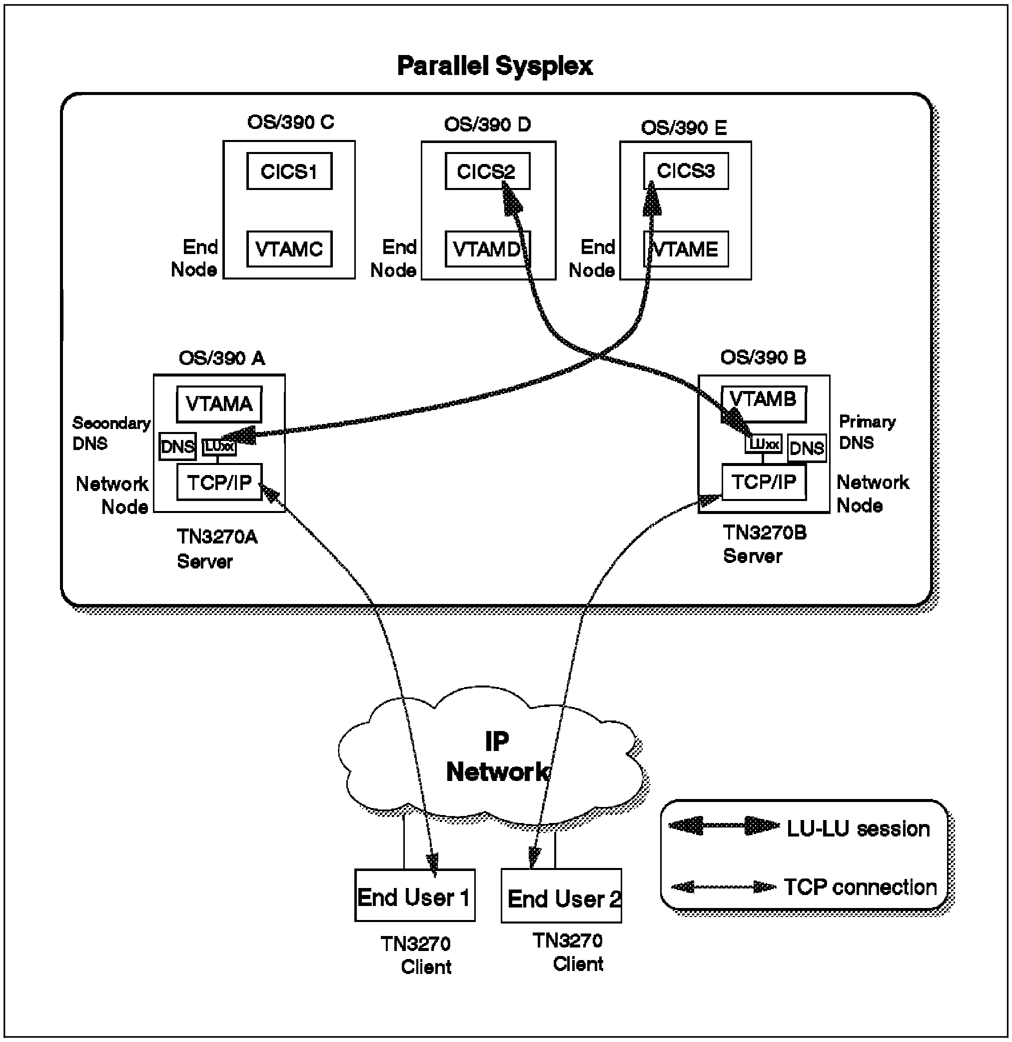


Figure 14. Using Telnet to Access 3270 Applications in a Parallel Sysplex

Figure 14 shows an environment where the TN3270 server function is performed by TCP/IP on OS/390 A and OS/390 B. In this example both TCP/IP stacks and TN3270 servers have registered themselves to WLM. The TN3270 servers use a groupname of tnsplex and the sysplex name is wtsplex. The DNS is enabled for WLM. There is a VTAM Generic Resource defined called CICSG of which CICS1, CICS2 and CICS3 are members.

In this scenario End User 1 requests a TN3270 session with tnsplex.wtsplex. The primary name server returns the TCP/IP address of an interface on OS/390 A. The name server decides which interface address to return based on information from WLM about available interfaces and load. End User 1 then requests the generic resource CICSG and is logged onto CICS3 on OS/390 E. End User 2 requests a TN3270 session with tnsplex.wtsplex and the primary name server returns the TCP/IP address of an interface on OS/390 B. End User 2 then requests the generic resource CICSG and is logged onto CICS2 on OS/390 D.

It is important to note that no actual application instances of generic resources are located on VTAMA or VTAMB. If you defined a generic resource that *does* use an actual application instance on VTAMB, then it appears to VTAMB as though all the LUs are local on OS/390 B. Therefore, the local application instance will always be chosen, and session balancing will not work across the other actual application instances in the Parallel Sysplex. Refer to "Local Access to Generic Resources" on page 225 in Volume 1: Overview, SG24-5637 for more details.

Typically, you would use a small number of TN3270 servers for network access, rather than configuring one on every S/390 server system. Therefore, it is recommended that the TN3270 servers be configured on the same OS/390 systems as the VTAM network nodes, and the generic resource application instances should be run only on the VTAM end nodes. In this way, workload balancing can be achieved for TN3270 clients.

Where Do You Put the TN3270 Server?

Define it on the VTAM network node system and do not run any generic resources there.

An alternative to using the TN3270 server function in TCP/IP for OS/390 is to use a separate gateway, such as 2216, Multiaccess Enclosure of 3746, an RS/6000, or a Cisco CIP/CPA.

Putting the TN3270 server in a gateway outside the Parallel Sysplex will ensure that the workload balancing done by the VTAM generic resources function is not affected in any way. It will not, however, allow you to use connection optimization. Placing the server outside the Parallel Sysplex also creates a single point of failure—if either it or its network connectivity should fail, then all sessions will be lost. Placing the TN3270 server in OS/390 is still a single point of failure if that system fails, but then the chances are that in this case the target SNA application will also be lost, so there's no real extra loss. And, by using VIPA, network failures can be bypassed without loss of active sessions.

2.3.2.5 Support for OS/390 System Symbols

TCP/IP introduced support for OS/390 system symbols in OS/390 V2R7. System symbols can be very useful for TCP/IP in a Parallel Sysplex, provided good naming standards are implemented. They can also make it easier to implement improved naming standards, by reducing the amount of work in changing definitions for lots of similar stacks and servers.

They work in much the same way as symbols in JCL procedures. The value given to any one symbol depends on the value defined in the IEASYMxx PARMLIB member used by the OS/390 system under which TCP/IP is running. The symbols are set in OS/390 at IPL, and used in TCP/IP at TCP/IP Initialization. Their main value in relation to TCP/IP is to allow you to use the same set of TCP/IP definitions for systems which have the same or nearly the same TCP/IP set up.

While the support for system symbols does not in itself provide a workload balancing capability, it does facilitate workload balancing by making it easier to maintain multiple instances of the applications across different images.

2.3.2.6 Interactive Network Dispatcher

Network Dispatcher is a feature that optimizes the performance of servers by forwarding TCP/IP connection requests and datagrams to different servers within a group. Thus, it attempts to balance the traffic across all the servers according to the load on them. The forwarding is transparent to the users and to applications. Network Dispatcher may be used for server applications such as HTTP, FTP, and Telnet; it can be used also for load balancing UDP traffic across a server group.

The Network Dispatcher function does not use a DNS name server for load balancing. It balances traffic among servers through a unique combination of load distribution and management software. Network Dispatcher also can detect a failed server and forward traffic only to the remaining available servers.

With Network Dispatcher, it is possible to combine many individual servers into what appears to be a single generic server. The site thus appears as a single IP address to the world. Network Dispatcher functions independently of a domain name server; all requests are sent to the IP address of the Network Dispatcher machine.

Network Dispatcher can load balance requests to different servers based on the type of request, an analysis of the load on servers, or a configurable set of assigned weights. Network Dispatcher includes a WLM agent that is used to collect utilization information about the various servers in the sysplex—that information can then be used to decide which server to route the request to.

Network Dispatcher runs with an IBM 2216, or on AIX on an RS/6000. Network Dispatcher uses the cluster method for dynamic workload balancing.

Following the IBM-Cisco alliance, IBM has announced that the 2216 will be withdrawn from marketing. As a result, Network Dispatcher will be effectively replaced by Cisco's MultiNode Load Balancing (described in the following section) or the Sysplex Distributor (announced as part of OS/390 V2R10 and described in 2.3.2.8, “Sysplex Distributor” on page 52) or a combination of the two.

2.3.2.7 Cisco MultiNode Load Balancing

Shortly before publication of this document, IBM and Cisco Systems announced that they would be working together to develop world-class e-business solutions, maximizing the high availability aspects of OS/390 and Cisco's IP technology. One of the projects being worked on is the use of OS/390's Workload Manager (WLM) by Cisco routers to identify the most appropriate host to route an incoming request to.

The two main features include the OS/390 Workload Manager and Cisco MultiNode Load Balancing, called MNLB. The Workload Manager dynamically manages assignment of resources based on customer-defined business priorities. The MNLB provides intelligent load balancing based on input received from WLM.

MNLB has three components:

- The Services manager residing in the Cisco LocalDirector.
- The forwarding agent which resides in IOS for routers and switches.
- The workload agent which resides in the S/390 to communicate with WLM.

Initial requests are routed to the LocalDirector. The Cisco MultiNode Load Balancing (MNLB) feature of the Cisco LocalDirector product works with WLM to select the appropriate S/390 server for the particular end-user connection. This ensures that the workload is distributed across the Parallel Sysplex servers. If one or more servers are experiencing high utilization, new connections are routed to other servers. Once the server is selected, traffic is routed directly from the router or switch via the MNLB forwarding agent to the server, without further involving the MNLB services agent in the LocalDirector. In fact, subsequent network traffic for the session does not even pass through the Local Director box itself in any way.

To increase availability, MNLB will detect LPAR outages and route traffic to alternate LPAR servers. Failed routers and switches are bypassed using IP routing. The LocalDirector can be backed up with a mirror image to provide immediate fail-over if the primary LocalDirector fails.

For more information about the Cisco LocalDirector support of WLM, refer to: `http://www.cisco.com/univercd/cc/td/doc/product/software/wla/wla_ug/`

2.3.2.8 Sysplex Distributor

The Sysplex Distributor function is actually shared among the TCP/IP stacks in the Parallel Sysplex by utilizing OS/390 V2R7's XCF dynamics support for inter-sysplex communication and OS/390 V2R8's dynamic VIPA support for configuration and recovery.

The role of each stack is established by configuring a dynamic VIPA which has been defined with a distribution server list for a particular port or ports. When the ALL keyword is specified in the distribution server list, any TCP/IP stack on an existing or new sysplex image automatically becomes a candidate for workload distribution. This can reduce the administrative burden significantly in a rapidly changing environment (e-Commerce, for example) by allowing the complete flexibility to move application servers or add new application server instances to new locations within the sysplex and still be considered part of the same single system image application group.

Once a dynamic VIPA becomes a sysplex-wide VIPA, workload can be distributed to multiple server instances without requiring changes to clients or networking hardware and without delays in connection setup, thus the data center and the customers business are allowed to grow non-disruptively.

Configuration: The stack defined as the primary owner of a dynamic VIPA (that is, via a VIPADEFINE statement in the TCP/IP profile) receives all IP datagrams destined for that particular VIPA. If this dynamic VIPA definition includes a distribution server list for a particular port or ports, then the Sysplex Distributor code running in the primary owning stack is activated so that it can distribute connection setup requests destined for that particular VIPA and port(s).

The stacks identified as candidates for workload distribution for particular port(s) require no configuration since they are notified of their role via inter-sysplex communication with the primary owner. In order to avoid the primary owning stack being a single point of failure, it is recommended that one or more backup stacks are defined (that is, via a VIPABACKUP statement in the TCP/IP profile).

The backup stack(s) can inherit the distribution server list via inter-sysplex communication with the primary owner, or can specify an entirely different distribution server list to be used if the primary owner experiences an outage. It is also possible to specify a distribution server list at the backup stack so that distribution only occurs during an outage of the primary owner. This allows the additional workload originally destined to the primary owner to be spread across multiple servers, thus lessening the overall impact to the data center during the outage.

For maximum flexibility with both IP and SNA applications, the primary and backup distribution servers should be placed on the systems that are active as APPN Network Nodes. The picture in Figure 15 shows a sample configuration with a primary and a backup distribution server, and a number of “hidden” servers that the load can be distributed across. You will notice that all servers are using the same VIPA.

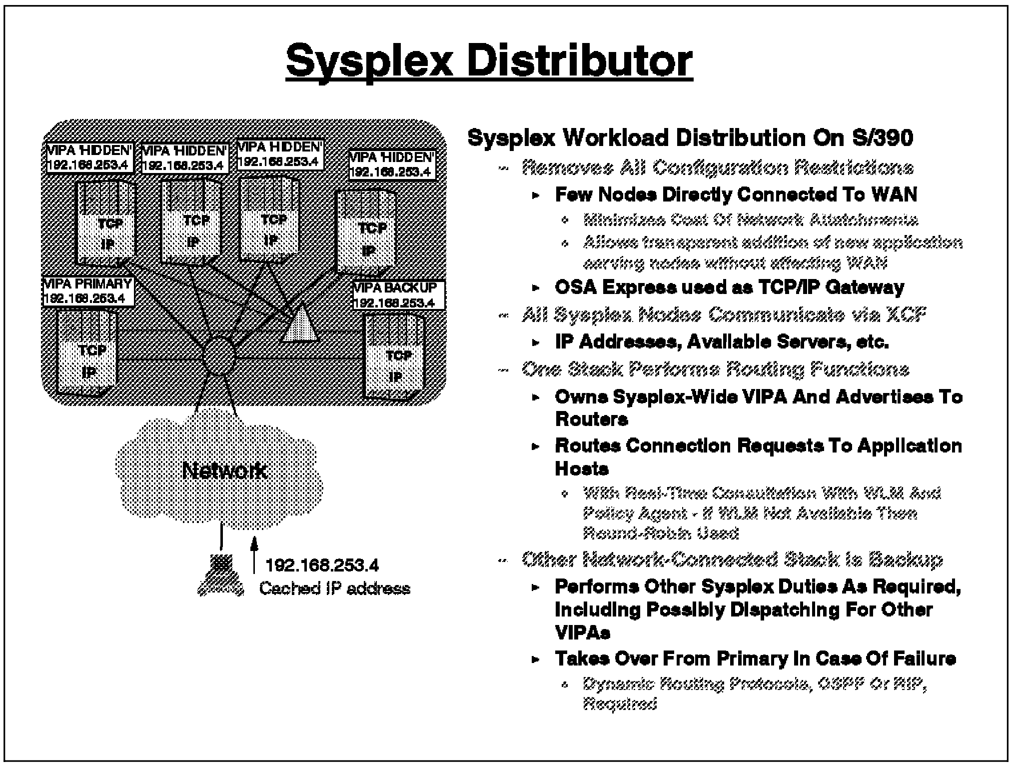


Figure 15. Sysplex Distributor Configuration

Functionality: Because the Sysplex Distributor resides in the Parallel Sysplex itself, it has the ability to factor “real-time” information regarding policy, quality of service (QoS) and application status into the distribution decision. By combining these real-time factors with CPU utilization information, the Sysplex Distributor has the unique ability to ensure that the best destination server instance is chosen for every particular client connection while ensuring that client/server-specific service level agreements are maintained.

Unlike other workload distribution methods, the Sysplex Distributor uses a variety of sources to obtain its distribution decision criteria. In addition to using information obtained from the S/390 Workload Manager (WLM), it also uses information from the Communications Server for OS/390s Service Policy Agent and information directly obtained from the target stacks themselves.

While it is very desirable to factor in the CPU utilization supplied by WLM to understand the workload on a particular system, it's not enough since it does not consider the network performance (that is, Quality of Service) in its workload balancing algorithm. Network performance is often the bottleneck in the overloaded Internet/ISP network, and is a critical factor in the end-to-end response time. Also, enterprise networks often have alternate paths to address network availability/reliability, and yet they are not taken into consideration in the optimization of end-to-end response time and availability/reliability.

This makes it difficult for the service provider to adhere to service level agreements. For example, it may be desirable to route more incoming connection requests to a more loaded server (higher CPU utilization) with better network performance, than to a less loaded server with much worse network performance.

Therefore the Service Policy Agent will inform the Sysplex Distributor whenever a particular server instance is not adhering to the QoS specified in its service level agreement.

The Sysplex Distributor has also been updated to include policy concerning the clients characteristics into the workload distribution decision. This policy can include the IP characteristics of the client (IP address and port, IP Subnet, etc.), time of day, day of week, as well as any other policy rule supported by the Service Policy Agent. An example of the usefulness of this function is in application hosting or in the Internet service provider marketplace, where clients accessing the same application can be assigned to different servers having different capability and/or connectivity.

The target stacks also assist the Sysplex Distributor in making the best distribution decision possible by supplying immediate server status information via inter-sysplex communication. Whenever an application server binds and listens to a port on a target stack being serviced by the Sysplex Distributor, the target stack sends a notification via inter-sysplex communication to the primary owner indicating that an application server exists and is ready to accept connections. When the application terminates or closes the listening socket, a corresponding notification is sent to the primary owner so that no additional connection requests will be distributed to this stack. The Sysplex Distributor has up-to-date information on available servers on target stacks so there is no need for application-specific advisors to issue periodic null application requests to determine existence of functioning servers, as is the case with many other workload distribution methods.

In addition to providing workload distribution, the Sysplex Distributor also enhances the availability provided by the dynamic VIPA support. The dynamic VIPA support in OS/390 V2R8 allowed a backup stack to takeover the VIPA in cases where the primary stack experienced a system outage. It did not, however, allow non-disruptive movement of the VIPA during normal operations. There was no way to preserve existing connections while relocating an application server which was using the VIPA via BIND or IOCTL DVIPA or while recovering the VIPA ownership at the primary stack upon its recovery from the outage.

The non-disruptive VIPA takeover function, also delivered in OS/390 V2R10, allows for this freedom of movement by maintaining the active connections established with the backup stack and allowing the VIPA to move immediately back to the primary owning stack. The primary owner will then be allowed to accept all new connections requests and internally use the Sysplex Distributor function to forward the IP datagrams to the backup stack for connections which were active at the time of the non-disruptive takeover. This ensures minimal impact for planned or unplanned system outages, since workload can be redirected without affecting existing connections.

Therefore, it can be seen that Sysplex Distributor combines the best features of both methods—up-to-date information for intelligent routing of *all* connection requests, and independence from factors such as where in the network the router is located, or whether the clients are “well-behaved.” It is envisioned that future enhancements to the workload balancing and availability characteristics of TCP/IP on OS/390 will be built upon the foundation provided by Sysplex Distributor.

2.3.3 Summary of Alternatives

All of the listed workload balancing functions fall into one of two categories—clustering based on a single name (using DNS), or clustering based on a single IP address (cluster method). In this section we will discuss the advantages and disadvantages of each.

2.3.3.1 DNS Method

A DNS method for TCP/IP application workload distribution uses a single name to represent the replicated application instances within the Parallel Sysplex. There are two types of DNS methods available in the marketplace today: DNS weighted round-robin, available on a variety of platforms, and DNS/WLM support provided by S/390.

The DNS/WLM support available on OS/390 V2R5 and above uses input from WLM to factor in the current health of the application instance in terms of CPU consumption and availability (that is, is it currently up and running on a particular stack). This provides an advantage over the round-robin approach, which does not factor in these variables.

Because of the time needed to perform a DNS resolution for every connect request from the client in the DNS/WLM model, it is generally recommended only for long duration connections like Telnet and FTP. Another important factor to consider is whether clients and/or the primary DNS (i.e. a DNS which is more local to the client than the S/390 DNS) cache the IP address even though the S/390 DNS supplied a time-left value of zero on the initial DNS resolution. If the “time-left” value of zero is not honored, the perceived availability of the application is severely compromised since the client may attempt to reconnect using the IP address of a “down” application server.

The VIPA takeover function in OS/390 V2R8 significantly improves the availability in this scenario, since the cached IP address of the server can be a dynamic VIPA address which can automatically be recovered on a backup stack. However, the WLM information is not factored into the reconnect decision.

2.3.3.2 Cluster Method

A cluster method for TCP/IP application workload distribution uses a single IP address which represents the replicated application instances within the Parallel Sysplex. There are a variety of implementations in the marketplace, including Network Dispatcher, the Cisco's MNLB feature described above, and the new Sysplex Distributor.

The Network Dispatcher and Cisco MNLB implementations use an IP address advertised as being owned by a particular router, and therefore all inbound connection requests for that IP address are directed to that particular router.

Upon arrival at the distributing router, the connection is forwarded to the destination server chosen, based on MVS WLM information obtained via agents running on the S/390 providing feedback to the router. The destination server processes the inbound connection request since the S/390 TCP/IP stacks have also been defined with a loop back alias address which corresponds to the cluster address owned by the router.

Cisco's MNLB has the advantage of being able to physically reside anywhere in the IP network, whereas the Network Dispatcher must be on the same IP subnet as the S/390 servers in the Parallel Sysplex. This Cisco flexibility requires that there are only Cisco routers upstream to the datacenter—this is because the Cisco routers communicate via their Dynamic Feedback Protocol (DFP) to enable forwarding of the data packets to the "real" destination server instead of all inbound data packets being sent to the router owning the cluster address.

The advantage of this method is that there is no dependence on clients not caching the IP address of the target node. Also, there is no need to resolve the name requests, resulting in faster turnaround for short (Web-like) transactions. On the other hand, because the WLM information is not obtained for every connection request, there is the possibility that every request will not be routed to the most appropriate server.

The Sysplex Distributor implementation is different. Rather than being located outboard of OS/390, the Sysplex Distributor function resides within the sysplex. However, it also is independent of whether clients cache the IP address or not. And, as a result of residing within OS/390, current WLM information is available to be used for every connection request, rather than having to use WLM information that may be outdated.

Table 4 contains a list of considerations and the most appropriate mechanism for each connection type.

Table 4. TCP/IP Workload Balancing Options

Requirement	**DNS/WLM**	**Cluster**	**Sysplex Distributor**
Need to distribute connection requests across multiple servers.	√	√	√
Need to ensure that the connection is always routed to an available server.			√
Every connection should be routed to a server capable of provided the required level of service.	√		√
Cannot guarantee that clients are not caching server IP address.		√	√
Connections are short in nature - Web transactions, for example.		√	√
Connections are longer in nature - TN3270 sessions, for example.	√		√
The application does not support WLM and/or DNS registration.		√	√
The applications are spread over multiple sysplexes, or even heterogeneous platforms.		√	

In addition to these, there are also some other things to consider:

- Will S/390 DNS be used?
- Is there enough capacity on the system that will own the cluster IP address?
- What is the network attachment to the data center?
- What is the desired location of the distribution point in the network?
- Do all clients have access to the network resources needed for the distribution model?
- Do different groups of clients require different QoS or Policy?

Each of these factors will have an impact on the method that is the most appropriate for your environment.

2.4 Connecting to the Network

Although the connectivity requirements of most installations are increasing at a dramatic rate, both in terms of bandwidth and diversity, the capabilities of the hardware used to provide that connectivity is also increasing rapidly. As a result, for a S/390 environment, there are three alternatives that should meet most of your requirements:

1. For an SNA-based WAN, which are still common in most large organizations, a 3745 can be used to provide the required connectivity, performance, and availability.
2. For an IP network, one option is an OSA adapter.
3. The other option, for an IP network, is to use a channel-attached router.

Whichever option(s) you select, direct channel connection from each OS/390 image in the Parallel Sysplex to the network gateway is recommended in order to provide the most efficient path for end-user sessions.

This section does not provide a comparison of all the feeds and speeds of the network gateways—they all change too frequently for such a list to have any value. However, it does discuss the APPN and HPR support of each of the options, because that may affect the availability of functions such as Multi-Node Persistent Sessions.

2.4.1 IBM 3745/6 Communications Controller Family

The 3745/6 communications controller family of products provides the largest and most flexible channel-attached gateway from the network to the Parallel Sysplex. Traditionally it acts as a frontend processor for SNA traffic. It also provides support for TCP/IP, both as a network gateway for access to TCP on the S/390 server, and as an independent IP router.

2.4.1.1 3745/6 Hardware

The 3745/6 combination provides the following connectivity options:

- Parallel or ESCON channels
- Token-Ring, Ethernet, Fast Ethernet or FDDI LAN connections (the Fast Ethernet and FDDI connections will no longer be available after the MAe is withdrawn from marketing)
- Low/medium speed line connections over V.24/V.35/X.21

- Medium speed connections (up to 2 Mbps) over V.35/X.21
- High speed serial line connections (HSSI) up to 52 Mbps
- ATM 155 Mbps (requires the MAe)
- ISDN Primary Rate (T1/J1/E1) (requires the MAe)
- Supported data link controls include SDLC, frame relay, X.25, BSC3270 and ISDN
- Supported protocols include SNA, APPN, HPR and IP
- APPN network node (with 3746 Model 900/950 NNP, or as part of a VTAM composite network node with NCP, or as part of Multiaccess Enclosure)
- IP routing support (using RIP, OSPF, or BGP)

The 3745 will often be configured with one or more connected 3746 expansion frames. In most cases, a 3746 Model 900 with or without Multiaccess Enclosure will be configured, as this adds increased connectivity and processing capacity to the 3745. In particular, the 3746 Model 900 and the Multiaccess Enclosure both offer ESCON channel connectivity.

The 3746 M950 Nways Controller can also be used as a network gateway. The 3746 Model 950 provides all the functions of the 3746 Model 900, but it acts as a stand-alone routing node, without being attached to a 3745 and without the requirement for the NCP.

The adapters in a 3746 Model 900 can be controlled by the NCP running in the 3745, or by a Network Node Processor (NNP), which runs in the Model 900 itself. The adapters in a 3746 M950 are always controlled by the NNP. The NNP provides an APPN network node for the 3746.

A 3745 with 3746 Model 900 (with or without Multiaccess Enclosure), or a 3746 M950 (with or without Multiaccess Enclosure), is a prerequisite to using ESCON channels. Also, if you require the channel connections to be defined as APPN links, then a 3745, 3745/6, 3746 M950, or 2216 M400 is required. However, if neither ESCON nor APPN links are required, then older communications controllers such as the 3725 or 3720 can be used.

The Multiaccess Enclosure adds all the connectivity options of the 2216 to the 3746 M900 or stand-alone 3746 M950. In addition, it provides access to functions such as the Network Dispatcher to all the adapters of the 3746.

2.4.1.2 3745/6 S/390 Channel Considerations

ESCON adapters are physically installed in a 3746 M900 or a stand-alone 3746 M950. Each ESCON adapter is connected to an ESCON channel processor on the 3746. If you are using a 3745 and you wish to configure ESCON adapters, you need to add a 3746 Model 900 to the existing 3745. At the same time, the 3745 has to be upgraded to a Model 17A, 21A, 31A, 41A, or 61A. The ESCON adapters can be controlled by the NCP in the 3745, or the NNP in the Model 900/950.

The theoretical limit of a maximum of 16 ESCON adapters is imposed because the 3746 Model 900/950 has slots for 16 processors, which can be any of the following:

- ESCON Channel Processor
- Communications Line Processor (for serial lines)
- Token-Ring Processor (for Token-Ring or Ethernet LAN connections)

So, when planning the installation of ESCON adapters, you need to be aware of the other processors that may be installed on the 3746. The maximum number of 16 processors per 3746 Model 900 applies to the 3745 M17A, 21A, and 31A. If you are using a 3745 M41A or 61A, the maximum number of processors supported in the attached 3746 Model 900 is 15.

The base 3745 may optionally have an additional 3746 Model A11 or A12, and additional IBM 3746 Model L13, L14, or L15 expansion frames attached to it. The 3746 Model 900 must be bolted to the IBM 3745 or, if present, to the 3746 Model A11 or A12.

Each ESCON adapter can connect to a maximum of 16 S/390 servers. These connections may be attached to physically separate S/390 servers, to different channels on the same S/390 server, to shared channels, or to different VTAMs on the same channel. The ESCON adapters support the EMIF capability of ESCON S/390 servers.

If ESCON adapters are not used, parallel channel adapters are configured on the 3745. The maximum number supported is 16, depending on the model. It is possible to install both parallel channel adapters and ESCON channel adapters on a 3745/6, without restrictions.

EP/PEP Guidelines: A number of configurations use the Emulation Program (EP) in a 37xx controller to support BSC and asynchronous connections to BTAM applications. If you do have an EP or PEP in your 37xx controller, then it is still possible to access these applications in a Parallel Sysplex. Parallel channel adapters must be used, as the EP does not support ESCON.

EP/PEP Still Has a Requirement for Parallel Channels

If you use EP/PEP in your Parallel Sysplex, reserve parallel channels for it.

For availability, configure two ESCON adapters and two paths to each S/390 server image. If more than one SSCP uses a path, the configuration should be checked for performance using the CF3745 tool.

ESCON Byte-Multiplexer (CBY) Channel Support: Support for the CBY channel path allows ESCON channels attached to an IBM 9034 ESCON Converter Model 1 to connect to control units that require byte-multiplexer channels. CBY channel paths operate as byte-multiplexer channel paths. However, you *cannot* define CBY channel paths for LPs running in S/370 mode. CBY channel paths support the following control units:

- IBM 3745
- IBM 3725
- IBM 3720

2.4.1.3 3745 Software

The software used in the 3745 is the *Network Control Program* (NCP). If you have decided to keep the connections to the Parallel Sysplex as subarea links, and use VR-TGs between the VTAMs to support APPN control point sessions, then there is no prerequisite NCP level. For more information on VR-TGs, refer to 2.1.1.1, “Virtual-Route-Based Transmission Group” on page 23.

However, if you want to add new nodes into the Parallel Sysplex as APPN end nodes or network nodes, then you will need to define APPN transmission groups between the VTAM APPN nodes and the NCP in the 3745/6. This means that you need to provide APPN support in the NCP. NCP V6.2 and higher provide the support for APPN.

Storage Requirement for APPN

If you migrate existing subarea links to become APPN links, additional storage may be required on the 3745/6 communications controller. You should use CF3745 to check the NCP storage requirements.

2.4.1.4 3746 Software

The 3746 M950 does not use the NCP, so there are no NCP software prerequisites. If the adapters of the 3746 Model 900 are controlled by the NNP, there are also no NCP prerequisites. There are a number of levels of microcode in the 3746 that provide different levels of function. You should use CF3745 to check these EC levels.

2.4.1.5 3745/6 Networking Considerations

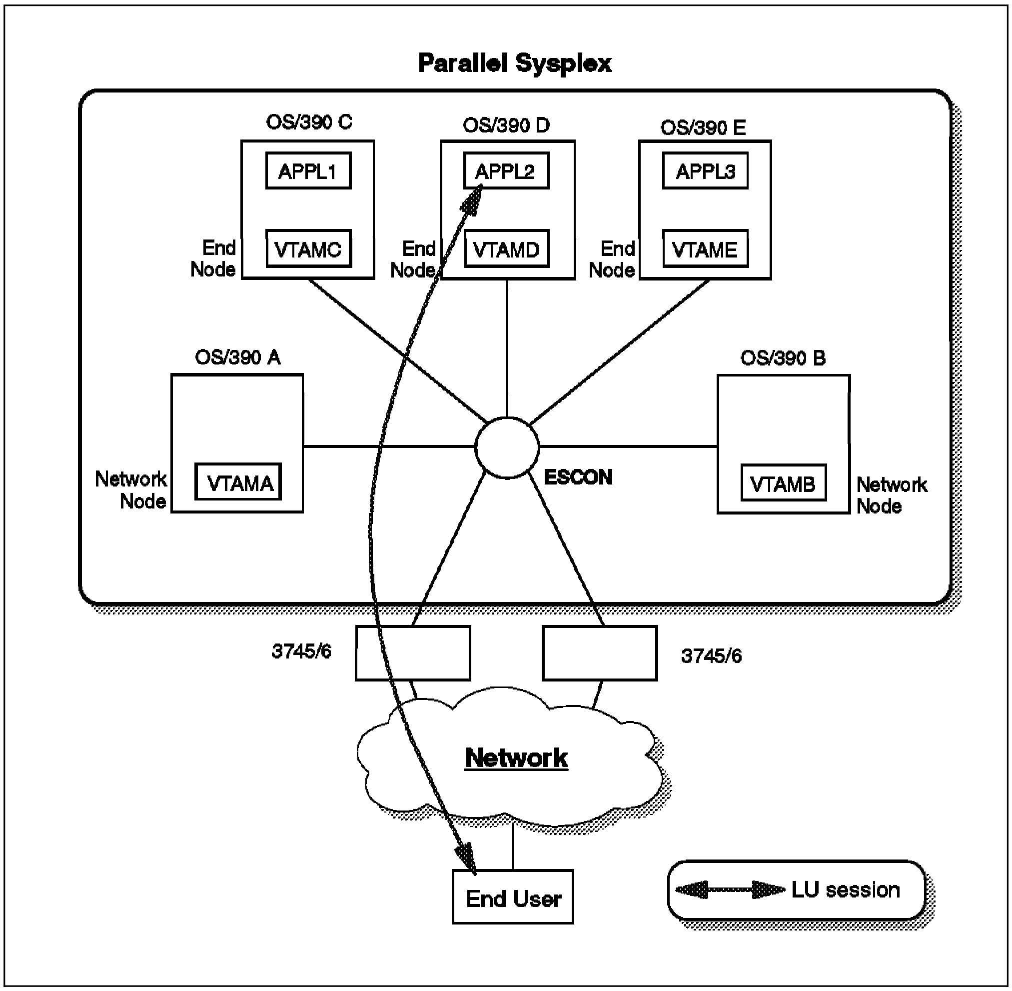


Figure 16. Using 3745 or 3746 As a Network Gateway to a Parallel Sysplex

Figure 16 shows a typical environment where a network-attached end user has a session with an application on Parallel Sysplex system VTAMD. The 3745/6 provides the SNA *boundary function* to support sessions from dependent LUs. This function allows sessions from end users to be routed through the NCP and directly up a channel path to the VTAMD system. The boundary function is an SNA component used for all dependent LUs, and is provided here by the NCP.

If the ESCON channel adapter is controlled by the Network Node Processor (NNP) of the 3746 Model 900 or 950, or by the network node of the Multiaccess Enclosure, then the end-user session will still take the path shown in Figure 16 on page 60. In this case, the boundary function will be provided by the DLUR function, which is provided by the NNP. Refer to 2.1.1.2, "Dependent LU Requester/Server" on page 24, for an explanation of the Dependent LU Requester/Server.

If you want to use Multi-Node Persistent Sessions, the 3745/6 is positioned to support HPR. Both the NCP and the 3746 Model 900/950 support HPR routing. The 3746 Model 900/950 provides HPR endpoint function, which is useful if the network has not been migrated to APPN/HPR.

Using 3745/6 for Multi-Node Persistent Sessions provides optimum availability benefits, because the end-user session can be routed directly through the 3745/6 without having to route through an intermediate VTAM network node. In Figure 5 on page 28, the 3745 and 3746 provide a gateway with the same functions as GW2. They can also both act as remote intermediate nodes, such as GW5 in Figure 5 on page 28.

2.4.2 Open Systems Adapter

The S/390 Open Systems Adapter (OSA) feature of S/390 offers direct LAN connections using industry-standard network interfaces which are packaged under the covers of the S/390 server. For network access to S/390 severs, the interface of choice may vary by many factors, such as the S/390 server machine type and model, the network infrastructure and protocols deployed, and the throughput requirements.

OSA has delivered three generations of product: OSA-1 and OSA-2 were delivered in 1995, and OSA-Express was delivered in 1999. OSA-1 has been withdrawn from marketing, so we will not discuss it further in this chapter. OSA-2 is a standard feature of G3 and G4 9672s and Multiprise 2000. OSA-2 and OSA-Express are standard features of G5 and G6 9672s. Both OSA-2 and OSA-Express provide for the transport of TCP/IP, SNA (APPN, subarea, and HPR), and IPX traffic to S/390 servers.

2.4.2.1 OSA-2 Adapter

The OSA-2 is the second generation adapter to provided direct LAN attachment to IBM CPCs. The OSA adapter physically plugs into the CPC I/O cage in place of a normal channel card. The OSA-2 card attaches to the Channel Request Handler bus, which is also used for ESCON channels and operates at 17 MB/sec. A CPC can support up to 12 OSA-2 cards. OSA-2 is available on the following models:

- IBM 9672 R2/R3
- IBM 9672 G3
- IBM 9672 G4

- IBM 9672 G5
- IBM 9672 G6
- IBM Multiprise 2000
- IBM Application StarterPak

The Multiprise 3000 supports connection to Token Ring and Ethernet, but does not use the OSA for this support.

The OSA-2 provides connection to the following LAN types:

- Token Ring
- Ethernet
- Fast Ethernet (10/100)
- ATM 155 (OC3)
- FDDI

All OSA-2 cards can be shared between multiple LPARs by using EMIF. If the system is running in LPAR mode, communications between the logical partitions (LP-to-LP) depends on the type of OSA-2 and in what mode of operation it is running. There are five different feature codes that can be used to order an OSA-2 card:

- 5202 - FDDI
- 5208 - Fast Ethernet (only on G3 and later)
- 5201 - Ethernet and Token Ring
- 5204 - ATM Multimode
- 5206 - ATM Single Mode

Token Ring: OSA-2 can provide LPAR to LPAR communication, but in half duplex mode only.

Ethernet: Communications between the logical partitions (LPAR-to-LPAR) to which the OSA is defined are not allowed as the OSA has an identical source and destination MAC address.

ATM: The ATM OSA-2 can provide communications between the logical partitions (LPAR-to-LPAR) to which the OSA is defined, but the network traffic flow between LPARS will be via the ATM switch or router to which the OSA is connected.

FDDI: The FDDI OSA-2 also supports communications between the logical partitions (LPAR-to-LPAR) to which the OSA is defined. The communication goes across the LAN rather than being handled directly on the card.

IP Network Data Routing: The OSA-2 has an OSA Address Table (OAT) which is created and maintained by the OSA Support Facility. In this table you map an IP address to a device address and an LPAR. So, for example, device 0300,0301 (TCP/IP has a read and write address—the read address is the lower address number) are assigned an IP address. You can also set up one LPAR in the OAT table as the PRImary LPAR and another LPAR as the SECondary.

All IP addresses that are sent to the OSA are read by the OSA card. If there is a match in the OAT table, the TCP/IP network data is sent to the corresponding LPAR. If there is no match for the TCP/IP address in the OAT table, the data is sent to the PRImary LPAR. If this LPAR is inactive, the data is sent to the SECondary LPAR.

2.4.2.2 OSA-Express

The OSA-Express is the third generation OSA to provide direct LAN attachment from IBM CPCs.

The OSA-Express card plugs into an I/O slot in the CPC and is connected directly to the Self-Timed Interconnect (STI). This is a full duplex 333 MB/sec bus. This makes the OSA-Express a significantly faster attachment than the OSA-2. The OSA-Express also provides additional functions, including offloading the mainframe IP stack, and allowing more complex configurations involving VIPA. OSA-Express is available on the following CPCs:

- IBM 9672 G5
- IBM 9672 G6

Three different types of OSA-Express adapter are currently available. All of them have a single LAN port which can be shared among LPARs by using EMIF:

- Gigabit Ethernet (Gbe)
- Fast Ethernet (FENET)
- ATM 155

Multiple cards can be installed, up to a maximum of 12.

The Gbe Adapter operates at 1.0 Gb/sec (gigabits per second) and includes Jumbo 9,000 byte frame support. This adapter supports the TCP/IP protocol. SNA traffic is also supported when Enterprise Extender is enabled.

The Fast Ethernet Adapter can auto-sense both 10 Mb/sec and 100 Mb/sec while providing protocol support for:

- TCP/IP
- SNA

The 155 ATM Adapter operates at 155 MB/sec while providing protocol support for:

- TCPIP (classical IP or RFC 1577)
- SNA
- LAN emulation for Token Ring and Ethernet

All three types of OSA-Express adapters support LPAR-to-LPAR communication. In all cases, data is sent directly from one LPAR to the other without traversing the LAN.

A significant enhancement to OSA-Express is the ability to configure the card in Queued Direct Input/Output mode (QDIO). QDIO is supported by the TCP/IP protocol and requires OS/390 V2R7 (or later) for the Gigabit Ethernet (Gbe) Adapter and OS/390 V2R8 (or later) for the Fast Ethernet and 155 ATM adapters. QDIO transfers data from the adapter's memory directly to Comm Server for OS/390. A QDIO-defined OSA adapter provides continuous data exchanges using memory queuing with minimal I/O interrupts and continuously active communication. This has reduced the TCP/IP path length by 25%. You also have the ability to set up priority queuing rather than the best effort priority normally assigned to all TCP/IP traffic.

OSA-Express channels that use QDIO are defined as TYPE=OSD in HCD and therefore require OS/390 V2R7 or later. Those that do not use QDIO are defined as TYPE=OSE, and therefore require OS/390 R3 or later. OSA-Express channels defined as OSE provide the same protocol support as the OSA-2 adapters but at higher speed.

For OSA-Express cards in QDIO mode, the OAT is dynamic. OAT data is sent directly to the card from Comms Server for OS/390 every time the device is started. After that, the decision about which LPAR to route the incoming TCP/IP data to is handled in the same way as described in 2.4.2.1, “OSA-2 Adapter” on page 61.

2.4.2.3 OSA-2 and OSA-Express Considerations for SNA

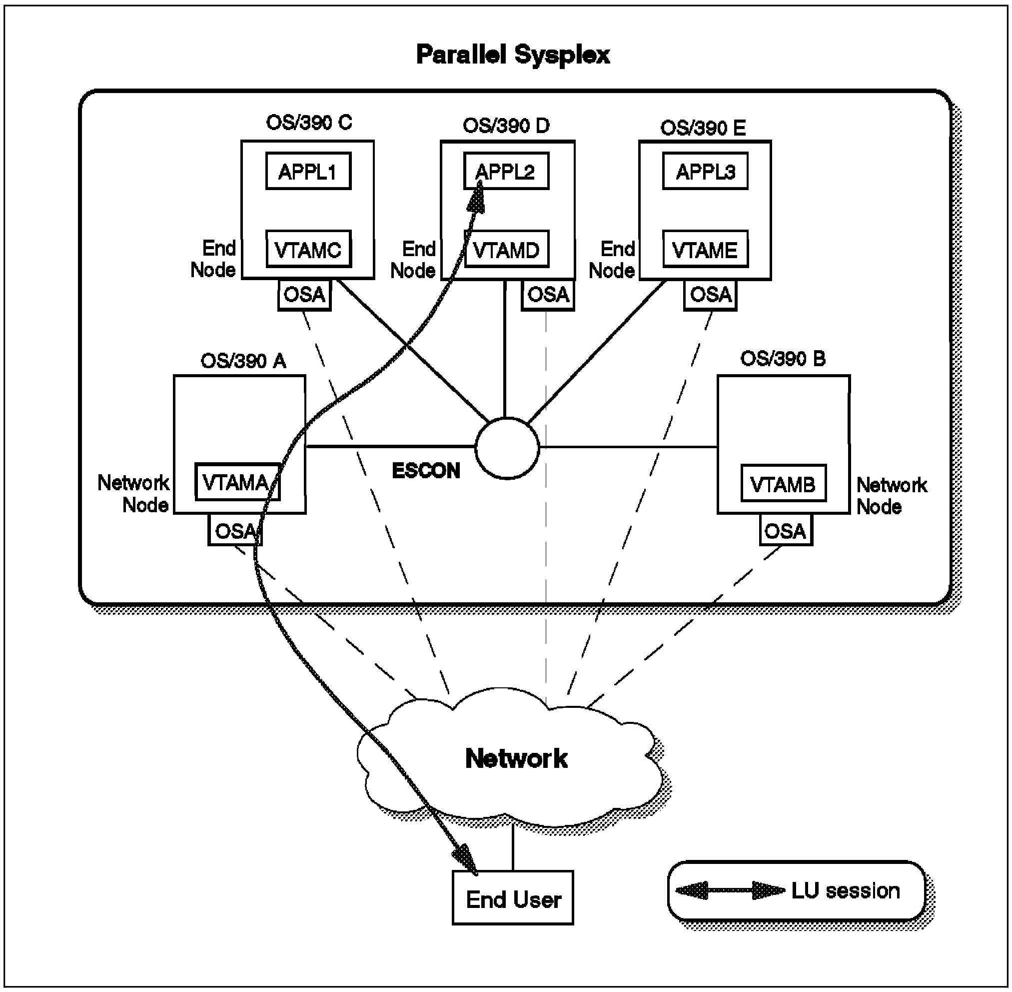


Figure 17. Using the OSA As Network Gateway to a Parallel Sysplex

Figure 17 shows an environment where a network-attached end user has a session with an application on Parallel Sysplex system VTAMD. OSA does not provide the SNA boundary function itself, but makes use of the boundary function in VTAM. This means that sessions from dependent LUs have to be routed through the VTAM that provides the boundary function. This is the normal way OSA operates; it is not a special case for the Parallel Sysplex.

In this environment, it is most likely that the boundary function will be provided by one of the VTAM network nodes. The sessions will pass from the end user across the network and into the VTAM network node via the OSA, and will typically then be routed from the VTAM network node to the selected VTAM end node using another connection, such as an ESCON CTC.

One factor which changes the routing of the end-user session is the use of APPN in the network. If APPN is used in the network and on the end-user's workstation, then it is possible to use a facility of APPN called *connection network* support. This supports media such as LANs, where all the nodes are logically adjacent. In this case, the end user-session can be routed directly across the network and into the OSA on the VTAM end node, without having to route through an intermediate VTAM network node.

A recent enhancement to VTAM has added the ability for VTAM to support the APPN/HPR ANR routing function when using the OSA. (In this configuration, VTAM does not see the OSA as a separate routing node; the OSA appears as if it is part of the VTAM node itself.) So, in a high availability configuration, using either MNPS or just APPN/HPR, the OSA can now act as a gateway without compromising the overall availability and performance.

The OSA does not provide an outboard APPN or IP router, and so it uses the routing functions of Comms Server for OS/390. Therefore, it can connect both SNA and IP networks to the S/390. However, the lack of the outboard router means that additional S/390 cycles are used compared to the solutions based on 3745/6.

For SNA use of OSA and Token Ring, it is possible to set up the OSA so that if one of the SNA hosts fails, traffic destined for that host can be handled by an alternate host instead. This is achieved by defining two OSAs with the same MAC address. On one OSA you set a delay timer so that one OSA always responds to the SNA request before the other. The second (delayed) OSA will only get used when the first OSA fails to respond. This procedure is described in detail in the section entitled "SNA Session Availability Options" in *OSA Planning Guide*.

Users wanting the same type of resilience for Ethernet networks should consider using APPN/HPR.

2.4.2.4 OSA-2 or OSA-Express Considerations for TCP/IP

The OSA does not provide its own workload balancing capability. There is no equivalent of IND for OSA. However, OSA *does* support the use of VIPA addresses, and the ability to move a VIPA from one LPAR to another.

Workloads can be moved between servers, on a planned basis, without disruption to established connections. System administrators can perform on-line transfers of applications without effecting end-users. In addition, a primary application server can take back work (VIPA take-back requires OS/390 V2R10), after it has recovered, without disrupting the connections established on the backup server. Used in conjunction with the Sysplex Distributor, this provides distribution of workload to multiple backup servers during a primary server outage.

Further information about the use of VIPA with OSA is available in the INFO APAR II10581.

As well as providing support for application outages with VIPA, OSA also provides mechanisms to provide continued availability across the failure of an OSA card. There are several ways to ensure network availability should failure occur at either the logical partition or the CHPID/network connection level. Port sharing, redundant paths, and the use of primary and secondary ports all provide some measure of recovery. A combination of these can guarantee network availability regardless of the failing component.

When TCP/IP is started in QDIO mode, it downloads all the home IP addresses in the stack and stores them in the OSA-Express feature. This is a service of QDIO architecture. The OSA-Express then responds to ARP requests for its own IP address, as well as for virtual IP addresses (VIPAs). If an OSA-Express feature fails while there is a backup OSA-Express available on the same network or subnetwork, TCP/IP informs the backup OSA-Express which IP addresses (real and VIPA) to take over, and the network connection is maintained.

Figure 18 shows how IP address control is automatically taken over from a failing OSA-Express

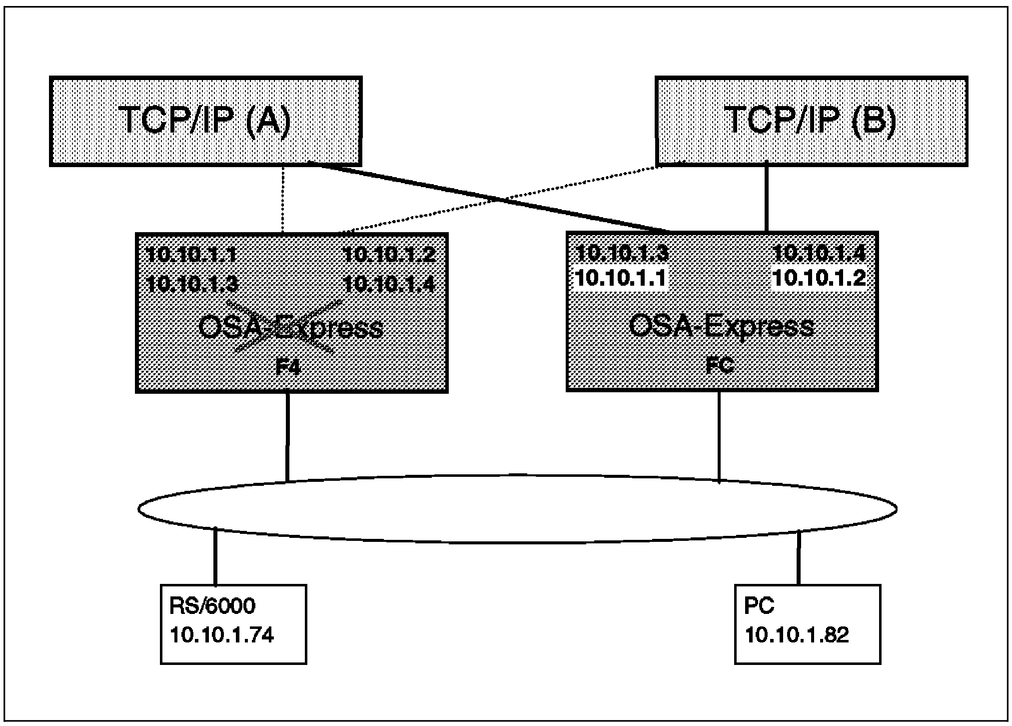


Figure 18. OSA Failure Scenario 1

This figure shows two OSA-Express features, each storing the same four IP addresses downloaded by TCP/IP. Under normal conditions, OSA-Express CHPID F4 controls the network connection to IP addresses 10.10.1.1 and 10.10.1.2, while CHPID FC controls the connection to 10.10.1.3 and 10.10.1.4. If CHPID F4 fails, OSA-Express CHPID FC is automatically reconfigured to take over control of IP addresses 10.10.1.1 and 10.10.1.2, while maintaining control of 10.10.1.3 and 10.10.1.4.

When multiple TCP/IP instances are using the same OSA-Express feature, you can designate a TCP/IP instance as the primary default or secondary default for handling unknown IP address destinations. (CS for OS/390 uses the terms *primary router* and *secondary router*) For the OSA-Express feature, you must designate a primary or secondary instance of TCP/IP in order to access an IP address on another LAN. Only one TCP/IP instance can be registered as the primary default on any OSA-Express. Likewise, only one TCP/IP instance can be registered as the secondary default. (See the description of the PRIROUTER and SECROUTER parameters on the DEVICE statement for MPCIPA in *OS/390 Communications Server: IP Configuration* for details on how to designate a primary or secondary router.)

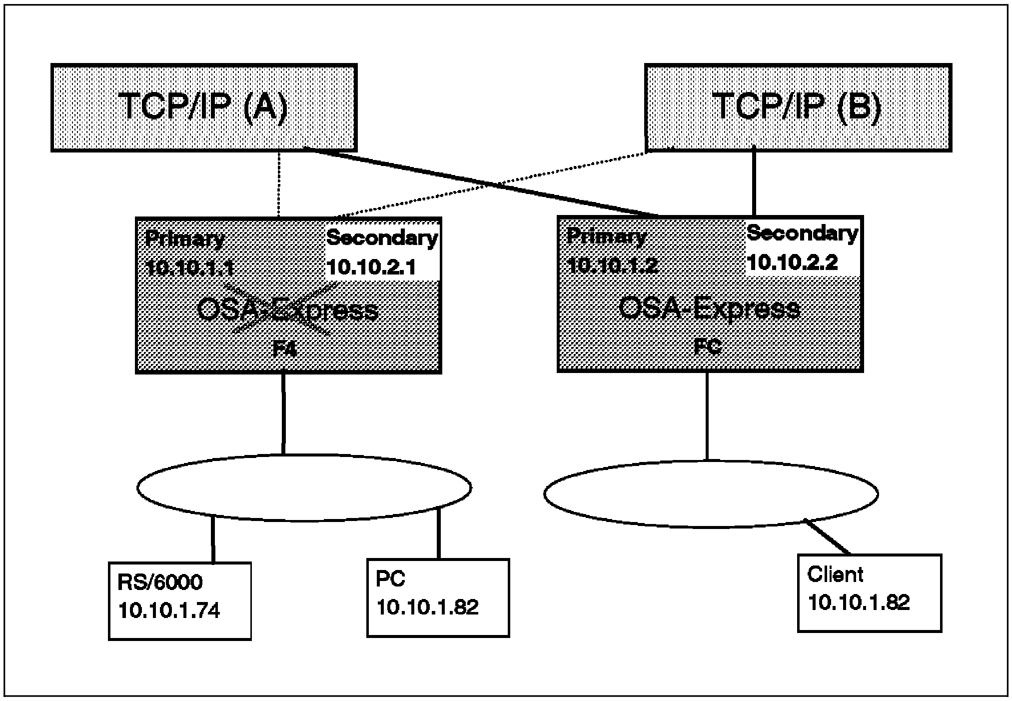


Figure 19. OSA Failure Scenario 2

Figure 19 shows two TCP/IP instances, one of which is designated Primary TCP/IP Router and the other, Secondary TCP/IP Router. Two OSA-Express features provide redundant paths to both TCP/IP instances, ensuring that both the RS/6000 and the workstation can access the client at IP address 10.30.8.5 on another LAN. When TCP/IP instance A failed, thus removing the primary path between the 10.10 LAN and the 10.30 LAN, TCP/IP instance B provided a secondary path for recovery.

2.4.2.5 OSA Support for Enterprise Extender

The Enterprise Extender function was added to OS/390 V2R6 (via PTF). With the Enterprise Extender function on the S/390, the SNA Parallel Sysplex functions, such as Generic Resources and MultiNode Persistent Sessions, will be enabled even when the transport network is an IP network and access to the S/390 is provided using an OSA (to provide the IP connectivity).

With Enterprise Extender, the SNA transport can be replaced by an IP transport, while client access can be provided by either TN3270 clients, or SNA clients. By using Enterprise Extender, the decision to implement an IP transport end-to-end does not require the replacement of SNA applications with TCP/IP-based applications, enabling a continuance of the parallel sysplex high-availability functions that were implemented for SNA applications with IP transport network.

2.4.3 Channel-Attached Routers

Following from the announcement of the IBM-Cisco Alliance, IBM will no longer market the IBM 2216 Router, or any other channel-attached router (except the IP Router capability of the 3745/3746). As a result, if you need information about the Parallel Sysplex considerations for these devices you should contact the vendor for the routers you have or intend to purchase.

Chapter 3. Sysplex Timer Considerations

This chapter addresses the various factors that have to be considered when attaching an IBM 9037 Sysplex Timer to CPCs in a Parallel Sysplex.

Mandatory HW Requirement for Multi-CPC Parallel Sysplexes

The 9037 Sysplex Timer is a mandatory hardware requirement for a Parallel Sysplex consisting of more than one CPC. In this case, XCF requires that the CPCs be connected to the same External Time Reference (ETR) network.

The 9037 provides the synchronization for the Time-of-Day (TOD) clocks of multiple CPCs, and thereby allows events started by different CPCs to be properly sequenced in time. When multiple CPCs update the same database and database reconstruction is necessary, all updates are required to be time-stamped in proper sequence.

In a sysplex environment, the allowable differences between TOD clocks in different CPCs are limited by inter-CPC signalling time, which is very small (and is expected to become even smaller in the future). Some environments require that TOD clocks be accurately set to an international time standard. The Sysplex Timer and the Sysplex Timer attachment feature enable these requirements to be met by providing an accurate clock-setting process, a common clock-stepping signal, and an optional capability for attaching to an external time source.

A primary goal when designing the Sysplex Timer was to provide a time facility whose availability exceeded the availability of any of the individual sysplex elements. It was also essential that the integrity of the timing information be ensured. This is accomplished by extensive error detection and correction and by high-priority interruptions for situations where there is a loss (or possible loss) of Sysplex Timer synchronization. These interruptions provide an alert to the system control programs in all participating systems, notifying them that they must initiate either immediate recovery or an orderly shutdown to maintain data integrity.

Recommended Sources of Further Information

The following sources provide support for the information in this chapter:

- *IBM 9037 Sysplex Timer and S/390 Time Management*, GG66-3264
- *IBM Hardware Timer Settings for Year 2000*, `http://www.s390.ibm.com/year2000/`
- *Logical Partition (LPAR) Sysplex Test Datesource Enhancement*, `http://www.s390.ibm.com/pso/`
- *Migration Planning for the 9037 Model 2 Sysplex Timer*, 9037MIGR package
- *Planning for the 9037 Model 2*, SA22-7233
- *S/390 Time Management and IBM 9037 Sysplex Timer*, SG24-2070
- *Sysplex Timer Planning*, GA23-0365
- *Using the 9037 Model 2 Sysplex Timer*, SA22-7230

The following roadmap guides you through the chapter, providing a quick reference to help you find the appropriate section.

Table 5. Sysplex Timer Considerations Roadmap

You want to configure:	**If you are especially interested in:**	**Subtopic of interest:**	**Then refer to:**
Sysplex Timer			
	Overview		3.1, "Overview of Time Management" on page 71
	Configuration Considerations		3.2, "Configuration Considerations for an ETR Network" on page 75
		How many Sysplex Timers do I need?	3.2.1, "Planning for Availability of Sysplex Timer" on page 75
		Which attachment feature do I use for S/390 CPCs?	3.2.3, "S/390 CPC Sysplex Timer Attachment Features" on page 81
		9037 Port granularity	3.2.4, "9037 Port Configurations" on page 84
		What are the functions of the 9037-2 console?	3.2.5, "9037-2 Console Selection and Considerations" on page 85
		How do I improve time accuracy?	3.2.6, "Planning for Time Accuracy" on page 87
		What are the distance limitations between components of an ETR network?	3.2.7, "9037 Distance Limitations" on page 88
		Who supplies the cables?	3.2.8, "Planning for Cabling Requirements" on page 90
		Available 9037 RPQs	3.3, "9037 RPQs" on page 91
		Supported migration procedures	3.4.3, "Supported Sysplex Timer Migration Procedures" on page 93
		How to configure for continuous availability	3.5, "Recommendations for Continuous Sysplex Timer Availability" on page 94
		Options for nondisruptive maintenance and future upgrade	3.6, "Concurrent Maintenance and Upgrade Considerations" on page 97

3.1 Overview of Time Management

There is a long-standing requirement for accurate time and date information in data processing. As single systems have been replaced by multiple, coupled systems, this need has evolved into a requirement for both accurate and consistent clocks among the systems. Clocks are said to be consistent when the difference or offset between them is sufficiently small. An accurate clock is consistent with a standard time source.

Refer to *IBM 9037 Sysplex Timer and S/390 Time Management*, or *S/390 Time Management and IBM 9037 Sysplex Timer*, for a more detailed overview.

3.1.1 TOD Clock

The TOD clock was introduced as part of the S/370 architecture to provide a high-resolution measure of real time, suitable for the indication of the date and time of day.

3.1.1.1 64 Bit Clock

The TOD clock is a 64-bit unsigned binary counter with a period of approximately 143 years. The value of the TOD clock is directly available to application programs by use of the STCK instruction, which stores the value of the clock into a storage location specified by the instruction. The TOD clock is incremented, so that 1 is added into bit position 51 every microsecond. The architecture requires that the TOD clock resolution be sufficient to ensure that every value stored by a STCK instruction is unique, and that consecutive STCK instructions always produce increasing values.

The Sysplex Timer steps the CPC TOD clock at a rate of about 16 MHz. In S/370 architecture, when more than one TOD clock exists within a shared-storage MP, the stepping rates are synchronized so that all TOD clocks are incremented at exactly the same rate, and the architectural requirement for unique and increasing TOD clock values still applies. In the case where simultaneous STCK instructions are issued on different CPs, uniqueness may be ensured by inserting CP-specific values in bit positions to the right of the incremented position.

A carry out of bit 32 of the TOD clock occurs every 2^{20} microseconds (1.048576 seconds). This interval is sometimes called a "mega-microsecond" (Mμs). This carry signal is used to start one clock in synchronization with another, as part of the process of setting the clocks. The carry signals from two or more clocks may be checked to ensure that all clocks agree to within a specified tolerance.

The use of a binary counter for the time of day, such as the TOD clock, requires the specification of a time origin, or *epoch*; that is, the time at which the TOD clock value would have been all zeros. The S/370 architecture established the epoch for the TOD clock as January 1, 1900, 0 a.m. Greenwich Mean Time (GMT). The 64-bit clock uses 52 bits to hold the number of microseconds since the epoch.

3.1.1.2 128 Bit Clock

On the G5 9672 CPC, IBM introduced a new 128-bit TOD clock. The reasons for this change are as follows:

- Provide a date beyond the year 2041.

  The existing 64-bit TOD clock will wrap some time in the year 2042. Additional bits are required on the "left" side of the clock to provide integrity beyond that date.

- Sysplex-wide unique clock value continues to be possible

  The 390 architecture guarantees that no two users will be provided with the same TOD value when running on the same CPC. However, it is possible that two users running on two CPCs in a sysplex could issue a Store Clock instruction at exactly the same time and both receive exactly the same TOD value. Both users would have to be running on the same CP on their respective CPCs in order to get duplicate values, so this is an unlikely, but not impossible event.

  In order to provide a TOD value that is unique within the sysplex, some of the new bits are used to store the XCF slot number for this system. The slot number for each system is unique within that sysplex, thus ensuring that even if every system issued a Store Clock on the same CP number at exactly the same time, they would all receive unique values in return.

- More granular timestamps.

  As processor speeds increase, the granularity of the existing 64-bit clock will become insufficient. Remember that the 390 architecture ensures that every requestor will be provided with a unique TOD value. When the speed of a CP exceeds the granularity of the 64-bit clock, and a user issues two Store Clock instructions in quick succession, the CP would have to spin, waiting for the clock to be incremented before it could return the value for the second request. As processors get faster, the amount of time spent spinning on such requests could become significant. The addition of these bits to the TOD clock ensures that CPs can get faster by many orders of magnitude before this becomes a concern again.

Support for the new 128-bit TOD is provided back to OS/390 R3 by APAR OW38015. APAR PQ27386 adds High Level Assembler support for the related new commands—STCKE and SCKPF.

OS/390 provides emulation for the 128-bit clock on CPCs that do not have the actual hardware installed. This ensures that a sysplex-unique TOD value will be returned even if running on an older CPC.

Applications can continue using the existing 64-bit TOD instructions and macros, and will be returned the 64-bit portion of the new 128-bit TOD. It is not necessary to convert to the new format instructions until your application requires a sysplex-unique value, or until the year 2042 is approaching!

3.1.2 ESA/390 Time-Coordination Requirements

The requirement for consistency between TOD clocks in coupled CPCs is illustrated by the following scenario:

1. CPC A executes an STCK instruction (time stamp *x*), which places the clock contents in storage.
2. CPC A then signals CPC B.

3. On receipt of the signal, CPC B executes STCK (time stamp *y*).

For time stamps *x* and *y* to reflect the fact that *y* is later than *x*, the two TOD clocks must agree within the time required to send the signal. The consistency required is limited by the time required for signalling between the coupled CPCs and the time required by the STCK instruction itself.

Consider a transaction-processing system in which the recovery process reconstructs the transaction data from log files. If time stamps are used for transaction data logging, and the time stamps of two related transactions are transposed from the actual sequence, the reconstruction of the transaction database might not match the state that existed before the recovery process.

For additional time-coordination requirements refer to *IBM 9037 Sysplex Timer and S/390 Time Management*, or *S/390 Time Management and IBM 9037 Sysplex Timer*.

3.1.3 External Time Reference (ETR) Time

In defining an architecture to meet ESA/390 time-coordination requirements, it was necessary to introduce a new kind of time, sometimes called ETR time, that reflects the evolution of international time standards, yet remains consistent with the original TOD definition. Until the advent of the ETR architecture, the TOD clock value had been entered manually, and the occurrence of leap seconds had been essentially ignored. Introduction of the ETR architecture provides a means whereby TOD clocks can be set and stepped very accurately, based on an external Coordinated Universal Time (UTC) time source, so the existence of leap seconds cannot be ignored.

ETR time can be computed from UTC by adding the number of accumulated leap seconds between 1972 and the time that is to be converted:

ETR time = UTC + Leap Seconds

Leap seconds, and whether they should be specified or not, are discussed in detail in *S/390 Time Management and IBM 9037 Sysplex Timer*, and *Planning for the 9037 Model 2*.

A very few sites run their own atomic clock. These sites use the TAI (French for International Atomic Time) time standard. Sites using a TAI time standard must enter the correct leap second values into the 9037 console. The 9037 will then adjust for leap seconds and will adjust the TOD clocks to the correct value.

3.1.4 ETR Network

An ETR network consists of the following three types of elements configured in a network with star topology:

1. ETR sending unit
2. ETR link
3. ETR receiving unit

The sending unit is the centralized external time reference, which transmits ETR signals over dedicated ETR links. The IBM 9037 Sysplex Timer is an implementation of the ETR sending unit. The receiving unit in each CPC includes the means by which the TOD clocks are set, and by which consistency with ETR time is maintained. The receiving unit implementation is model-dependent, and is also called the Sysplex Timer attachment feature. The ETR network may comprise one or more sysplexes and CPCs not belonging to a sysplex.

It is likely that a large installation may have more than one ETR network, in which case it is important that all CPCs within a sysplex be part of the same ETR network. The ETR data includes network ID information, which is verified by the system control programs running in the attached CPCs, to ensure true consistency of all TOD clocks within a sysplex.

3.1.5 OS/390 Support for the Sysplex Timer

Starting with MVS/ESA SP V4.1, the Sysplex Timer provides a source for time information that allows the operating system to set and maintain synchronization of the TOD clock in each CP of a CPC and across multiple CPCs. OS/390 also includes services to applications, subsystems, and operators for determining whether the clocks are synchronized.

Two modes of time synchronization are available:

- Local synchronization mode, where OS/390 uses the services of the CP's TOD clock
- ETR synchronization mode, where OS/390 uses the services of the Sysplex Timer

The default is to synchronize to a Sysplex Timer.

The Sysplex Timer is not "defined" to OS/390 like a device; however, there are parameters in the SYS1.PARMLIB data set that control its use. The parameters in the CLOCKxx member of SYS1.PARMLIB that are important for time-coordination are:

```
TIMEZONE d.hh.mm.ss
ETRMODE YES|NO
ETRZONE YES|NO
OPERATOR PROMPT|NOPROMPT
SIMETRID nn
ETRDELTA nn
```

These parameters are discussed in detail in *S/390 Time Management and IBM 9037 Sysplex Timer*, and *IBM 9037 Sysplex Timer and S/390 Time Management*.

3.1.6 PR/SM LPAR and Sysplex Timer

The PR/SM, or LPAR LIC, supports the Sysplex Timer and manages changes to the CPC TOD based on Sysplex Timer signals. The LPAR LIC also presents notification of the Sysplex Timer events to those operating systems that support the Sysplex Timer.

At the completion of a CPC power-on reset in LPAR mode, the PR/SM LPAR LIC sets the TOD clocks to be in synchronization with the 9037. Each LP has its own logical TOD clock through the use of a TOD-clock-offset control block for each LP. This allows the logical TOD clock in a LP and the physical TOD clock to have different values. Each LP's logical TOD Clock is initially set to the same value as the physical TOD value at LPAR activation.

When the Sysplex Timer is attached to a CPC operating in LPAR mode, the LPAR LIC accepts the ETR interrupt. LPAR synchronizes the physical TOD to the Sysplex Timer TOD and adjusts the offset of each LP's TOD from the physical

TOD to compensate for the change in the physical TOD. This offset is referred to as the logical TOD offset. The adjustment is transparent to operating systems running in any of the LPs.

When an LP is deactivated and activated, the CPC TOD and the LP's TOD will be synchronized. If a LP is not deactivated, any previously existing offset between the CPC TOD and the LP's TOD is maintained across IPLs. An example of a previously existing offset may be the offset generated when an operator has specified a new GMT in response to the TOD prompt during an IPL. After LPAR has made the necessary adjustments in response to an ETR interrupt, the interrupt is passed to the operating system in each LP. The response of the operating system depends on its capabilities and setup. For a more detailed discussion, refer to the ITSO redbook *S/390 Time Management and IBM 9037 Sysplex Timer*.

3.1.7 Multiple Clocks and Their Uses

Three types of clocks are available on the 9672 CPCs:

- HMC Battery-Operated Clock (BOC)
- Support Element (SE) BOC
- Time of Day Clock (TOD)

In addition to these real hardware clocks, PR/SM allows a CPC to be logically divided into multiple LPs. Each LP has its own logical TOD clock.

3.2 Configuration Considerations for an ETR Network

There are several factors that must be considered when configuring an ETR network. Availability, time accuracy, cabling, console requirements, and attachment feature options are some examples.

3.2.1 Planning for Availability of Sysplex Timer

The 9037 Sysplex Timer is available in three configurations:

1. Expanded availability configuration
2. Expanded basic configuration
3. Basic configuration

Parallel Sysplex Availability Recommendation

The recommended configuration in a Parallel Sysplex environment is the 9037 *expanded availability* configuration. This configuration is fault-tolerant to single points of failure, and minimizes the possibility that a failure can cause a loss of time synchronization information to the attached CPCs.

APAR OW19728 documents how Sysplex Timer failures affect images in a sysplex. It also explains the effect on IPL, if ETR signals are not present. In a multisystem sysplex, any OS/390 image that does not have the PTF for APAR OW44231 applied, and loses ETR signals, is placed into a non-restartable wait state.

However, if the PTF for that APAR *is* applied, the system is placed in a spinloop and the operator is presented with a synchronous WTOR asking if the system should be quiesced or if the 9037 will be restarted in an expanded basic configuration. The effect of this APAR is discussed in more detail in 3.7, "Planning for System Availability" on page 98.

3.2.1.1 Sysplex Timer Expanded Availability Configuration

The *expanded availability configuration* consists of two 9037s. The 9037-2 expanded availability configuration is shown in Figure 20 on page 77, and the 9037-1 expanded availability configuration is shown in Figure 21 on page 77. Differences between the 9037-1 and 9037-2 Sysplex Timers are described in the ITSO redbook *S/390 Time Management and IBM 9037 Sysplex Timer*.

In an expanded availability configuration, the TOD clocks in the two 9037s are synchronized using the Control Link Oscillator (CLO) card and the CLO links between the 9037s. Both 9037s are simultaneously transmitting the same time synchronization information to all attached CPCs. The connections between 9037 units are duplicated to provide redundancy, and critical information is exchanged between the two 9037s every 1.048576 second (Mµs), so that if one of the 9037 units fails, the other 9037 unit will continue transmitting to the attached CPCs.

Redundant fiber optic cables are used to connect each 9037 to the same Sysplex Timer attachment feature for each CPC. Each Sysplex Timer attachment feature consists of two ports: the active or stepping port, and the alternate port. If the CPC hardware detects the stepping port to be non-operational, it forces an automatic port switchover; the TOD then steps to signals received from the alternate port. This switchover takes place without disrupting CPC operations. Note that the 9037s do not switch over, and are unaware of the port change at the CPC end.

Sysplex Timer Availability Recommendation

For an effective fault-tolerant expanded availability configuration, Port 0 and Port 1 of the Sysplex Timer attachment feature in each CPC must be connected to different 9037 units.

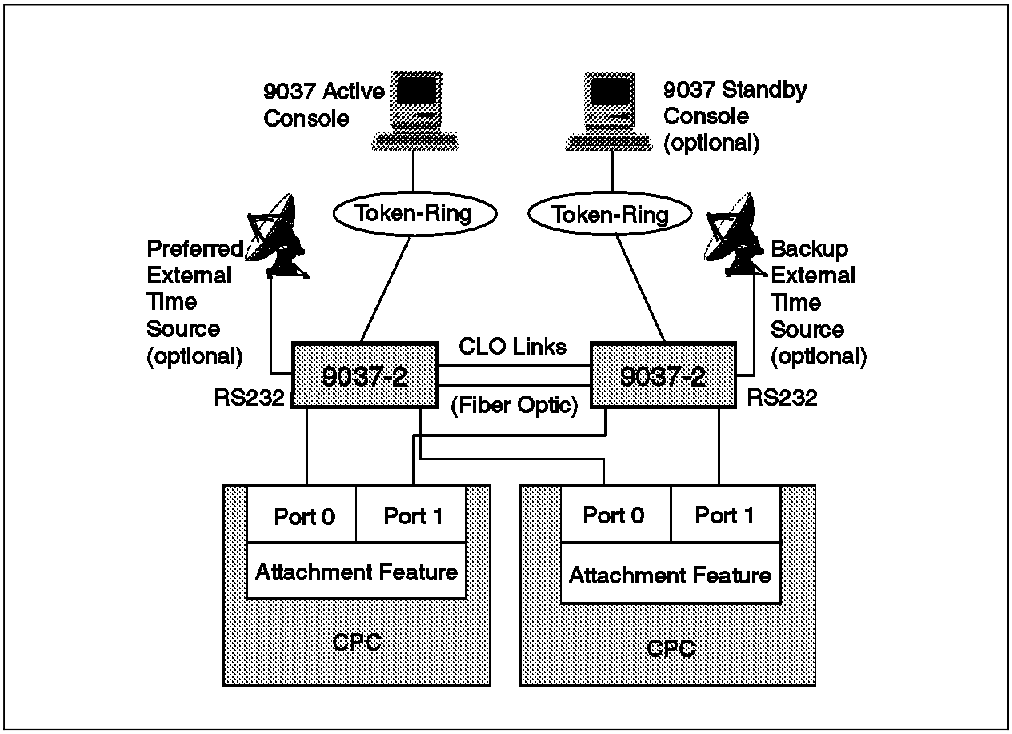


Figure 20. 9037-2 Expanded Availability Configuration

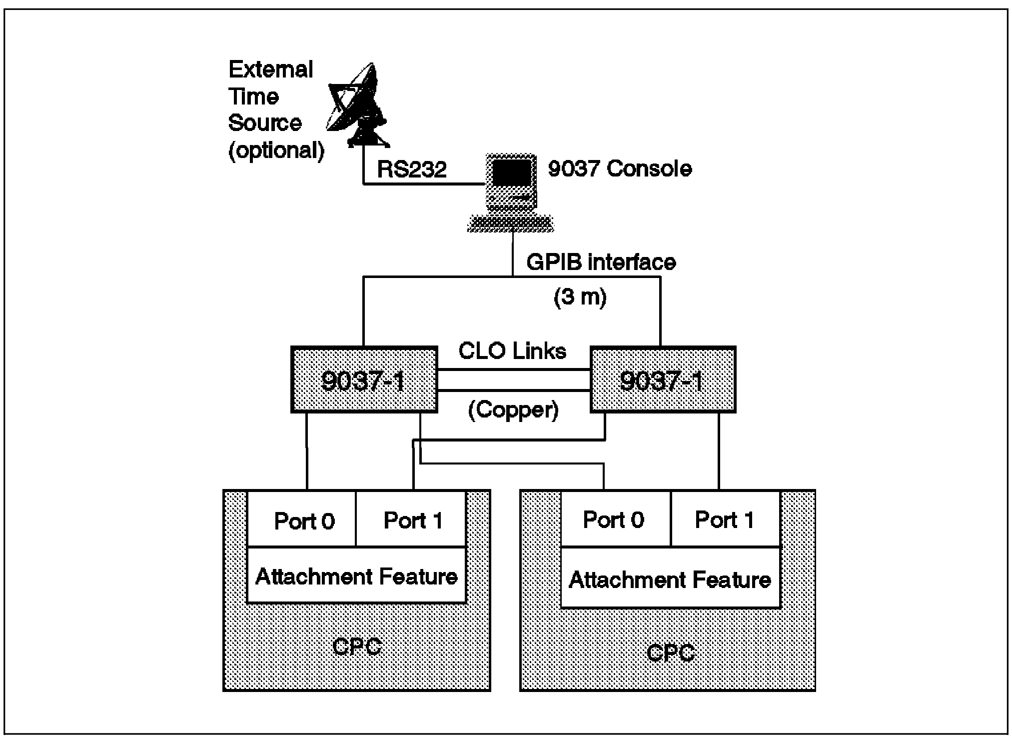


Figure 21. 9037-1 Expanded Availability Configuration

3.2.1.2 Sysplex Timer Basic Configuration

The 9037-2 and 9037-1 *basic configurations* are shown in Figure 22 and Figure 23 on page 79, respectively. The basic configuration is not recommended in a Parallel Sysplex environment because of the limited fault tolerance it provides. It can be used for applications when setting and maintaining the time of multiple CPCs to the same external time reference is desirable. When the 9037 provides this source, the user does not need to set the CPC TOD clock at IPL.

This configuration consists of one 9037, and can provide synchronization to all attached CPCs. The impact of a fiber optic cable or port failure can be minimized by connecting each port of the Sysplex Timer attachment feature to a 9037 port. For a 9037-2, there is an added availability benefit if the two ports of the attachment feature are connected to 9037-2 ports on separate cards, since the 9037-2 port cards are hot-pluggable.

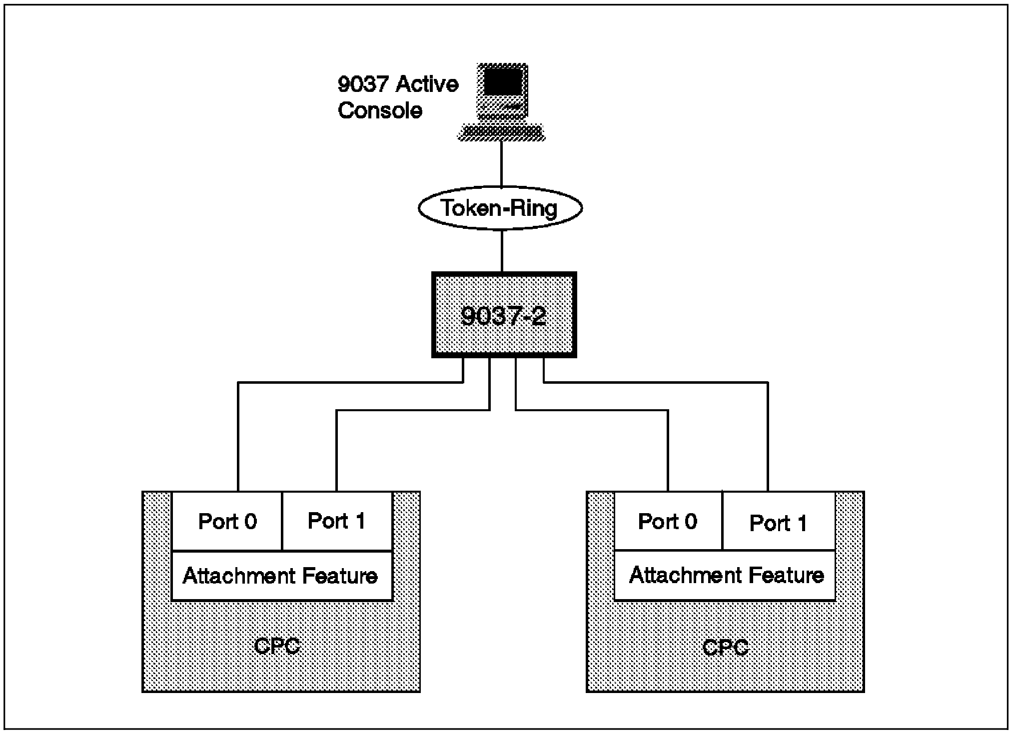


Figure 22. 9037-2 Basic (One 9037) Configuration

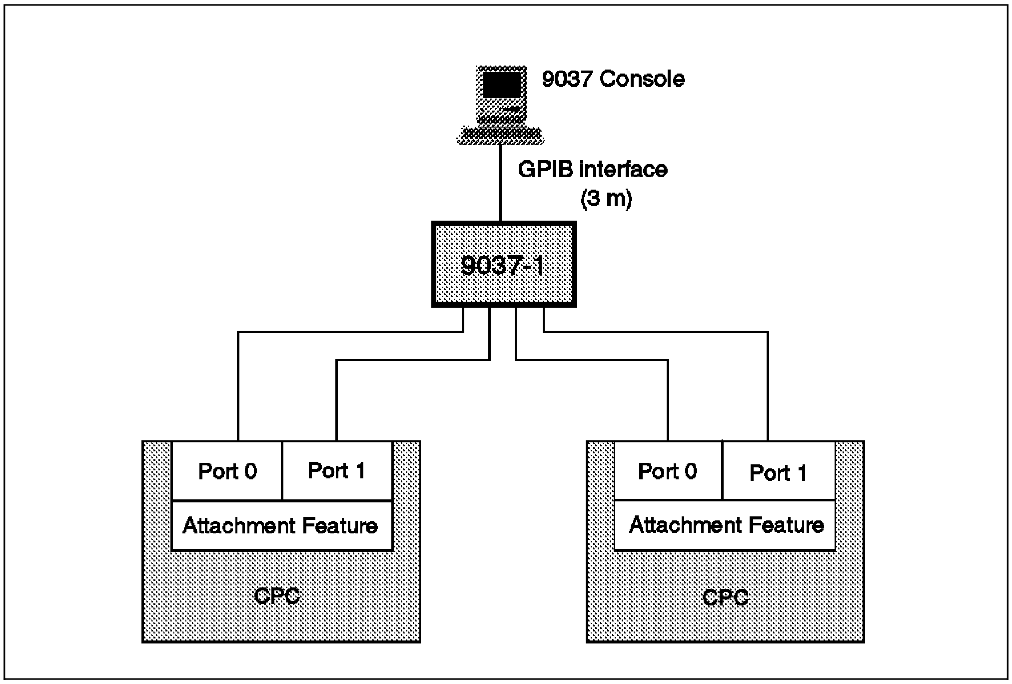


Figure 23. 9037-1 Basic (One 9037) Configuration

3.2.1.3 Sysplex Timer Expanded Basic Configuration

The *expanded basic configuration* is similar to the basic configuration in that only one 9037 is connected to the CPCs. The 9037-2 expanded basic configuration is shown in Figure 24 on page 80, and the 9037-1 expanded basic configuration is shown in Figure 25 on page 80. In the expanded basic configuration, a Control Link Oscillator card is installed in the Sysplex Timer. With this card installed, the 9037 can "steer" to an external time source. This configuration can also be enhanced to the fully redundant expanded availability configuration, without any outage of your operations.

The expanded basic configuration configuration is not recommended in a Parallel Sysplex environment because of the limited fault tolerance it provides. It can be used for applications when setting and maintaining the time of multiple CPCs to the same external time reference is desirable. When the 9037 provides this source, the user does not need to set the CPC TOD clock

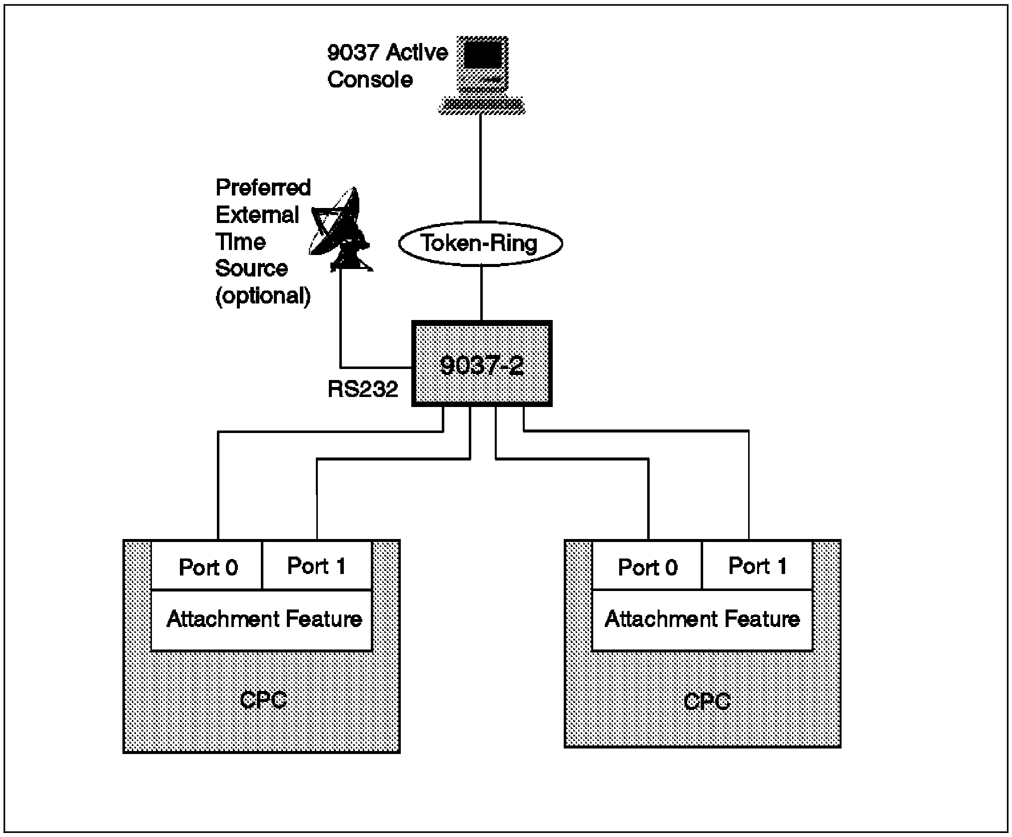


Figure 24. 9037-2 Expanded Basic (One 9037) Configuration

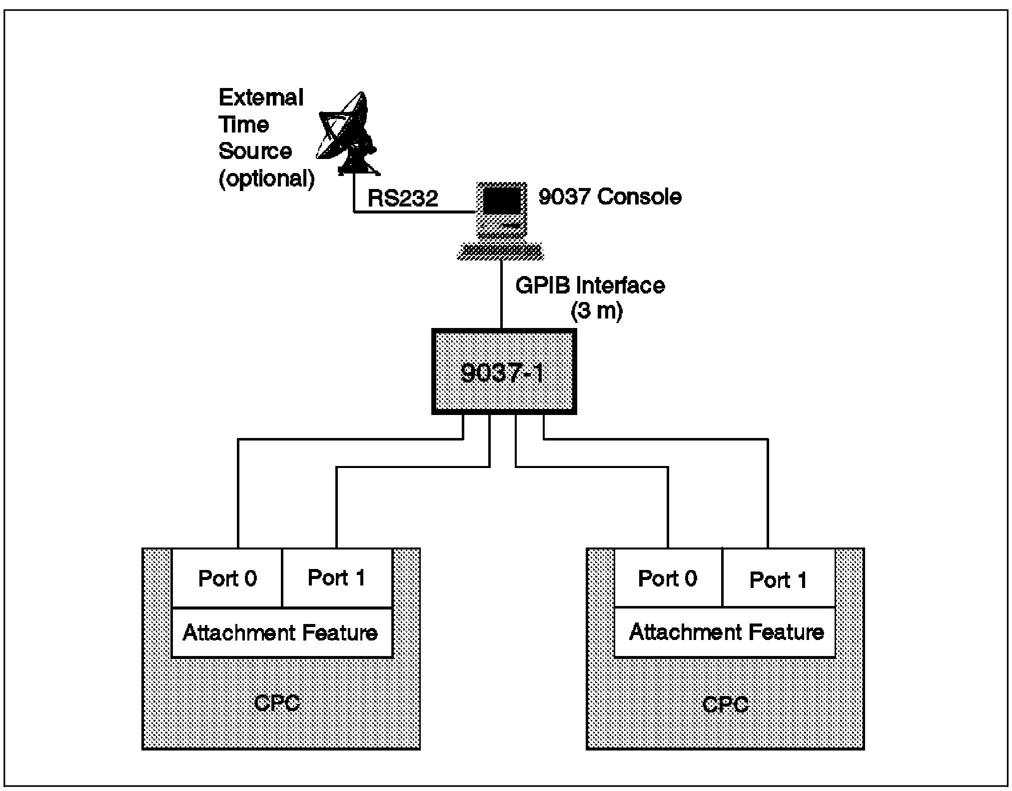


Figure 25. 9037-1 Expanded Basic (One 9037) Configuration

3.2.2 CPC Support for Sysplex Timer

This section lists the hardware and software that support the 9037-1 and the 9037-2. For more detailed information, IBM employees may refer to the sales manual on HONE.

3.2.2.1 CPC Support: Hardware

The 9037-2 can attach to the following CPCs:

- All IBM 9672 and Multiprise Models
- 9021 520- and 711-based and 9121 511-based CPCs

The 9037-1 can attach to the following CPCs:

- IBM 9672 and Multiprise CPCs
- 9021 340-based, 520-based and 711-based CPCs
- 9121 320-based and 511-based CPCs
- 9221, all models
- ES/3090 models 180J, and above

Note: For a discussion of S/390 CPC Sysplex Timer attachment features, refer to 3.2.3, "S/390 CPC Sysplex Timer Attachment Features." Contact your IBM representative for the appropriate feature codes.

3.2.2.2 CPC Support: Software

The following operating systems are supported by the 9037-2 and the 9037-1:

- OS/390 R1 or higher
- MVS/ESA SP V4.3 or higher[1]
- Transaction Processing Facility (TPF) V4.1

In addition to OS/390, both models of the 9037 are supported by System Automation for OS/390 R2. It provides exception and status monitoring for the 9037 and its associated fiber optic cable connections.

3.2.3 S/390 CPC Sysplex Timer Attachment Features

This section describes the attachment feature considerations required to connect a CPC to a Sysplex Timer (either 9037-1 or 9037-2).

3.2.3.1 Attachment Features for the 9672 CPCs

Two different types of Sysplex Timer attachment features can be used for the 9672 CPCs. They are:

- Dual port card
- Master card (not available on 9672 R1 models)

Dual Port Card: This card, shown in Figure 26 on page 82, is recommended for most configurations, since it does not have the same limitations as the master card. The card has two fiber optic ports, each of which is attached to a Sysplex Timer using a fiber optic cable. In an expanded availability configuration, each fiber optic port should be attached to a different Sysplex Timer in the same ETR network. This ensures that redundant Sysplex Timer paths are made available to all attached CPCs. This card does not have any output ports to redistribute the 9037 signals.

1 9037-1 is supported by MVS/ESA V4.1 or above

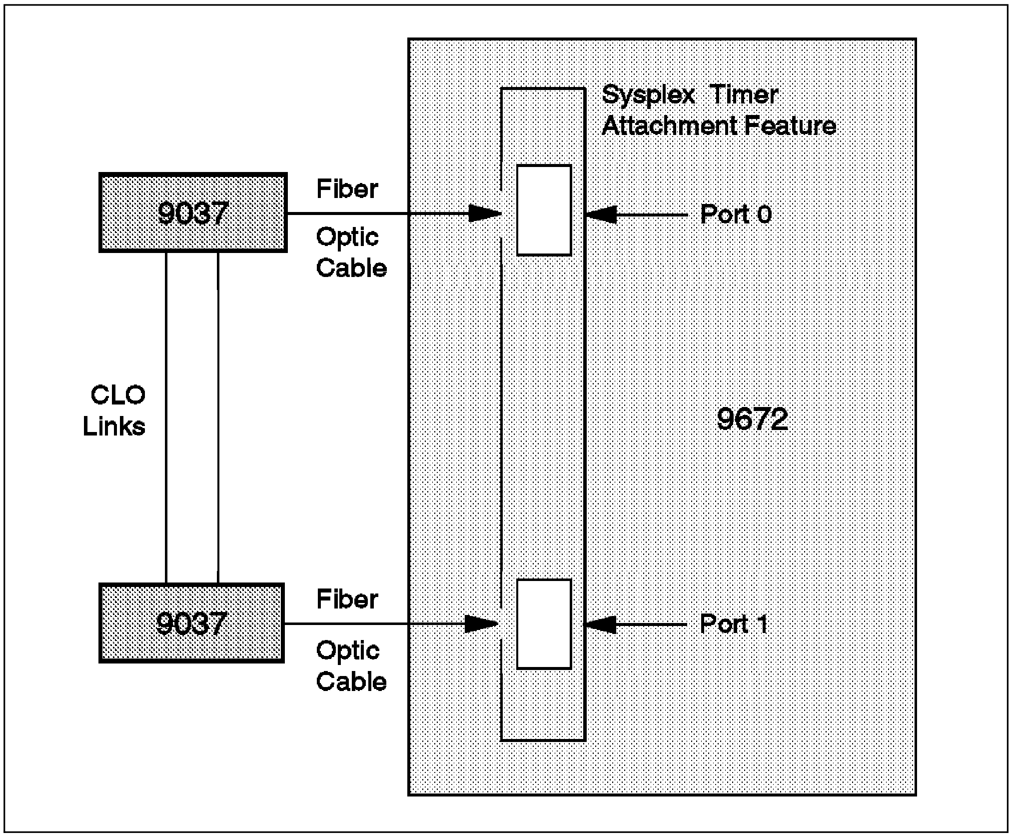


Figure 26. Dual Port Attachment to 9037 Sysplex Timers

Master Card: *The master card should be used only in configurations that require a Sysplex Timer port capability in excess of 24 ports.* The master card allows you to extend the physical 24-port limitation of the 9037-2 or the 16-port limitation of the 9037-1, by allowing you to attach a 9037-2 to a maximum of 48 CPC Sysplex Timer attachment feature ports or a 9037-1 to a maximum of 32 CPC Sysplex Timer attachment feature ports.

Each master card has two input ports and one output port. The *master input port* is a fiber optic port that is attached to a Sysplex Timer using a fiber optic cable. The *slave input port* is an electrical port, which receives redriven timer signals from the other master card's output port. The *master output port* distributes timer signals to the *other master card* through a 25-foot external cable. Refer to Figure 27 on page 83. In an expanded availability configuration, each master card's master input port should be attached to a different Sysplex Timer unit in the same ETR network. This ensures that redundant Sysplex Timer paths are made available to a pair of CPCs.

The following limitations of the master card must be carefully considered before selecting it for use:

- The pair of CPCs must be physically adjacent to each other, since the external cable is only 25 feet long.
- Whenever one of the CPCs is powered off, the remaining CPC has only one connection to a Sysplex Timer.

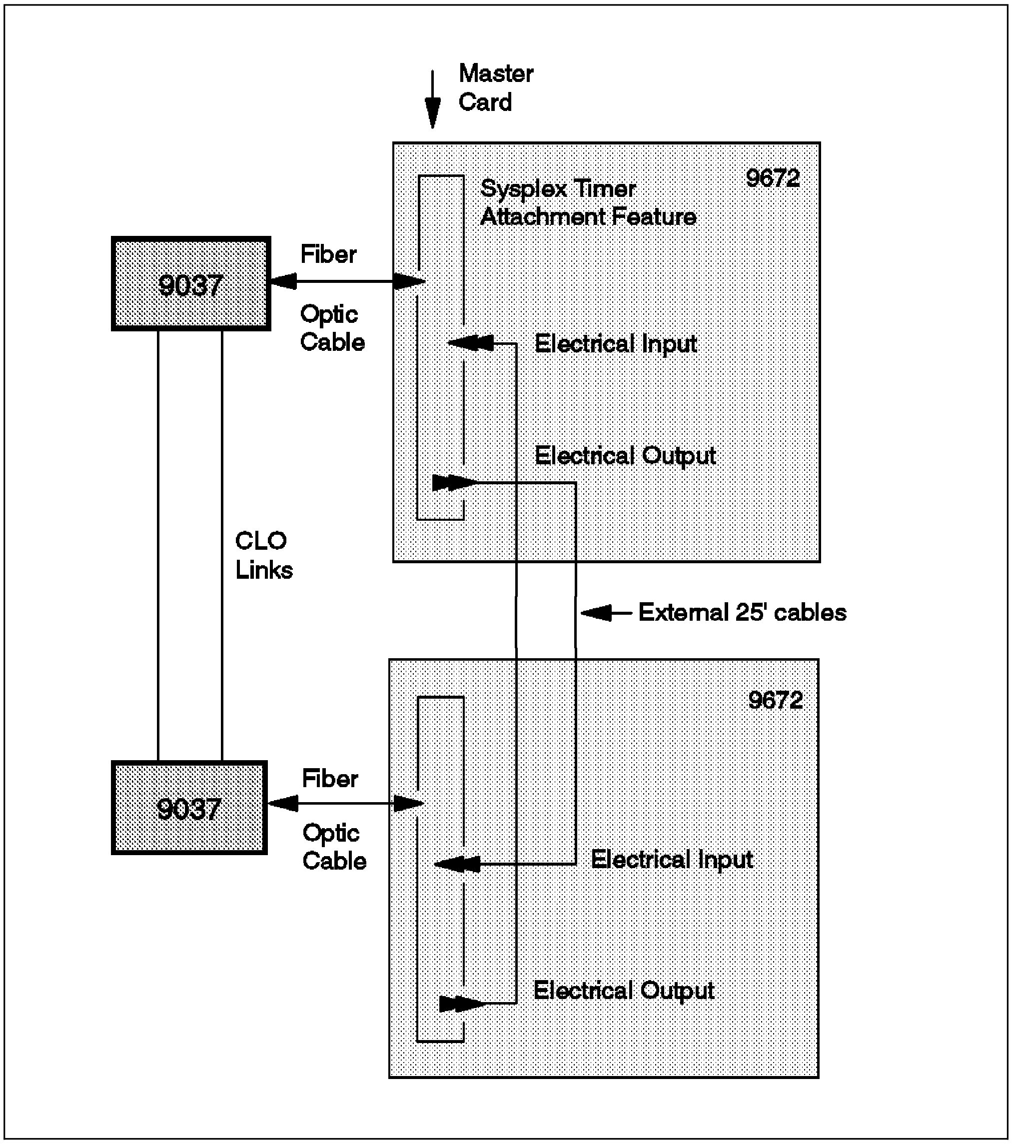


Figure 27. Example of Two 9672 CPCs with Master Cards and External Cables

Figure 28 on page 84 shows how, with only two fiber optic connections from each 9037, four 9672 CPCs can continue to receive ETR signals in an expanded availability configuration.

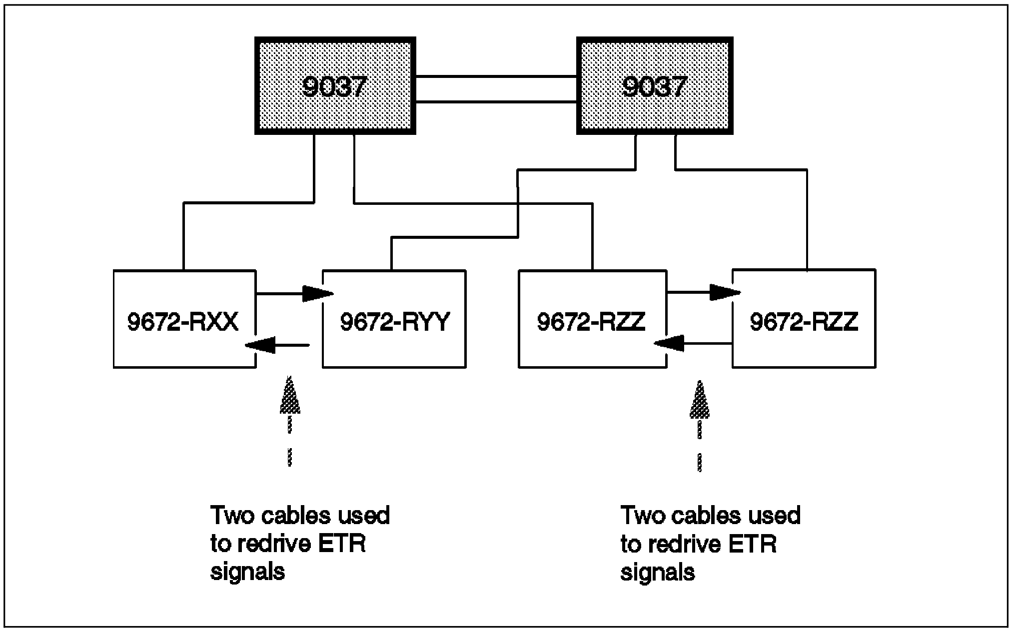


Figure 28. Example of Four 9672 CPCs with Master Cards and External Cables

Note: The 9674 CF *does not* require a connection to the Sysplex Timer; therefore, a Sysplex Timer attachment feature is not required.

3.2.4 9037 Port Configurations

The 9037-2 can have a minimum of four ports and a maximum of 24 ports. Ports can be added with a granularity of four ports per additional 9037-2 port card. 9037-2 port cards are hot-pluggable, which means that they can be plugged in without requiring the 9037-2 unit power to be turned off. 9037-2 port cards can therefore be concurrently maintained, added, or removed.

The 9037-1 can have a minimum of four ports and a maximum of 16 ports. Ports can be added with a granularity of four ports per additional 9037-1 port card. The 9037-1 port cards are *not* hot-pluggable.

The number of ports needed is equal to the total number of CPC connections, up to 24 ports. More than 24 CPC connections are possible in a configuration with S/390 CPCs. Refer to “Master Card” on page 82. Although there is no requirement to install spare ports, having them enables a quick transfer of connections, since the 9037-1 or 9037-2 units do not have to be powered off to connect a fiber optic link to a port.

If a basic configuration or expanded basic configuration is used, it is recommended that a redundant port connection is provided to each CPC. For a 9037-2, these connections should be made from different port cards. This reduces the probability of a single point of failure, as well as allowing a defective port to be replaced without disrupting transmission of timer signals.

Notes:

1. The Sysplex Timer is not channel-attached and, therefore, is not defined to the Input/Output Configuration Program (IOCP), MVS Configuration Program (MVSCP), or Hardware Configuration Definition (HCD).
2. The Sysplex Timer links cannot be routed through an ESCON director, because the two architectures are different.

3. Sysplex Timer links can be routed through a pair of 9036 Model 3 repeaters (RPQ 8K1919), a pair of 9729 MuxMasters, or a pair of 2029 Fiber Savers when extended distances beyond 3 km are required. See 3.2.7, "9037 Distance Limitations" on page 88 for more information.

3.2.5 9037-2 Console Selection and Considerations

Note: This section describes the console application and selection considerations for a 9037-2 console. If you have a 9037-1 unit, refer to *Sysplex Timer Planning*.

A console, dedicated to each 9037-2 configuration, is used to:

- Enter and display initialization data, such as time and offsets.
- Adjust the time.
- Enable and disable ports.
- Select ETS type and its operational parameters.
- Manage and monitor the Sysplex Timer network.
- Collect error information from the 9037s and generate an error log.
- Maintain a change log for data entered by the user.
- Define or modify connection definitions from the 9037-2 console to the 9037-2 unit.

The console connects to the 9037-2 over a 4/16 Mbps Token-Ring Local Area Network (LAN). The LAN used to connect a 9037-2 console to a 9037-2 unit can be public or private. IBM recommends that the 9037-2 units and consoles should be attached to a private LAN. Other users of the private LAN may include:

- 9672 and 9674 SEs
- HMCs
- ESCON directors
- ESCON director consoles
- Hardware Configuration Manager
- OSA/Support Facility

Private LAN Recommendation

A private LAN is recommended, since the Sysplex Timer is a critical component of the Parallel Sysplex and access on a public LAN could jeopardize LAN security. Also, a LAN with heavy traffic should be avoided, otherwise the 9037-2's capability to report errors to the console may be impaired.

A 9037-2 may have sessions with up to two consoles at a time. One of these can be configured as the *active console*, and the other as an *optional standby console*.

A Sysplex Timer network can be controlled by only one active console at a time. Operators at either console may view configuration and status information, but only the operator at the active console can make changes to the operating state of the 9037-2 units. All event reports are sent to both consoles if they are present. When changes are made at either console, the other console is notified of the changes and its configuration and status information is updated. If the active console fails, the standby console can be reconfigured as the active console.

When a 9037-2 is placed in "maintenance mode," a subset of functions is available on the standby console for changing configuration information while maintenance actions are being performed. For a detailed listing of which functions are available in "maintenance mode" on the standby console, refer to *Using the 9037 Model 2 Sysplex Timer*.

In an expanded availability configuration, only one 9037-2 is attached to an active console over the Token-Ring. The console communicates with the second 9037-2 over the fiber optic links (CLO links) between the 9037-2 units. A standby console can attach to the second 9037-2 through the same Token-Ring or a separate dedicated LAN, if distances exceed the limits of a Token-Ring LAN. Figure 29 shows an expanded availability configuration with active and standby consoles in a bridged LAN environment. For other examples of possible console configurations, refer to *S/390 Time Management and IBM 9037 Sysplex Timer*, and *Planning for the 9037 Model 2*.

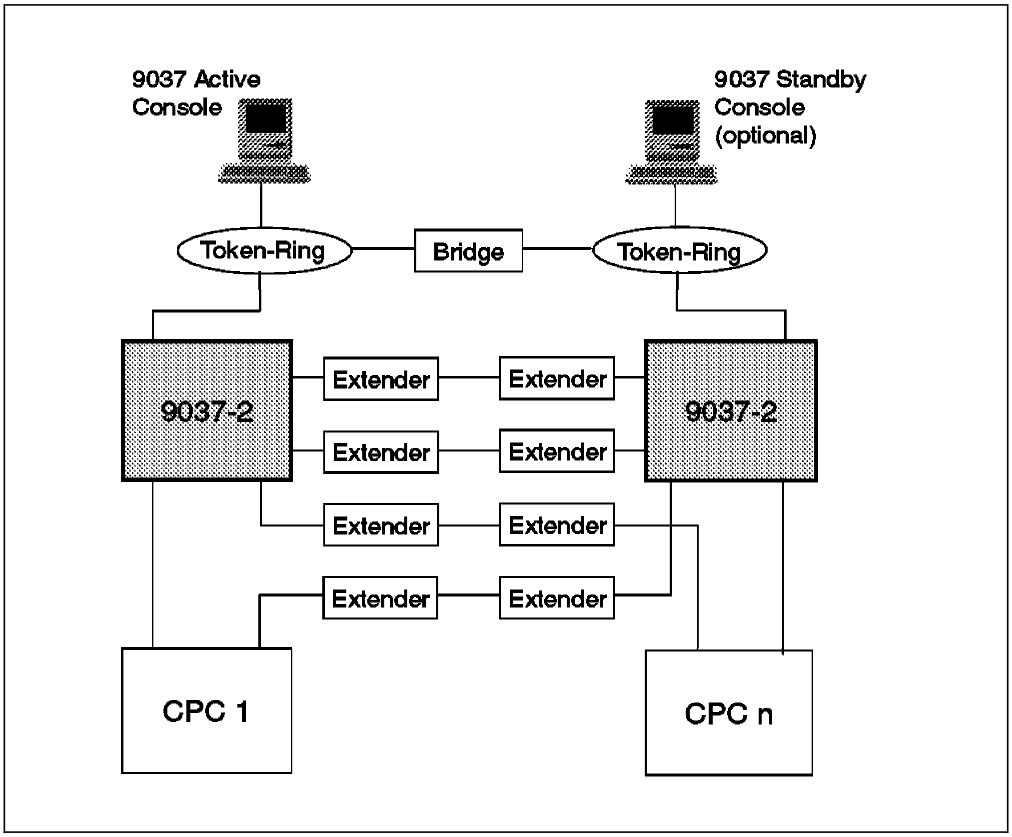


Figure 29. Sysplex Timer Network in a Bridged LAN Environment. Note the extender can be either a pair of 9036-3s (RPQ 8K1919), a pair of 9729s or a pair of 2029s (both the 9729 and 2029 require RPQ 8P1955).

3.2.5.1 9037-2 Console Platforms

The WIN-OS/2-based console application can exist on three different platforms:

1. The *recommended platform* is a dedicated IBM PC with appropriate hardware and software. The console is not automatically included as part of a 9037-2 order. You can order the console with the 9037-2 from IBM, or you can order it separately, either from IBM or from a non-IBM source. If you order it separately, make sure you meet the minimum hardware and software requirements specified in *Planning for the 9037 Model 2*.

2. The IBM PC console used with a 9032 Model 3, 9032 Model 5, or 9033 Model 4 ESCON director.
3. The S/390 CPC's Hardware Management Console(HMC).

3.2.5.2 Console Availability

The 9037-2 console should *always* be attached to the 9037-2 units. If a 9037-2 console failure occurs, the console should be repaired as soon as possible. The 9037-2 continues to transmit data to all attached CPCs while the console is being repaired. However, console functions listed in 3.2.5, "9037-2 Console Selection and Considerations" on page 85 cannot be performed. Also, information necessary to service the 9037-2 could be lost, if the console is not available.

3.2.5.3 Console Remote Operations

Using Distributed Console Access Facility (DCAF) software, it is possible to initiate a remote session to either the active or standby 9037-2 console. The remote session has all the command privileges of the target console.

Note: It is *not possible* to operate the 9037-1 console from a remote location. The distance between the 9037-1 console and the 9037-1 unit should not exceed 2.2 m.

3.2.6 Planning for Time Accuracy

If you have specific requirements to provide accurate time relative to some external time standard for your data processing applications, you need to consider attaching an ETS to the Sysplex Timer and using the leap second offset. The leap second offset is discussed in detail in *S/390 Time Management and IBM 9037 Sysplex Timer*.

Accuracy without an External Time Source

When either the 9037-1 or 9037-2 is *not attached* to an external time source, the accuracy of the 9037 is typically ± 32 seconds per year from the initial time setting. Actual accuracy could vary with the effects of component aging and machine environment.

The 9037 ETS function allows the time accuracy of the 9037-2 and the 9037-1 to be enhanced. Instead of setting the 9037 time manually, the ETS function allows the 9037 time to be set from an ETS device in all three 9037 configurations. The basic configuration only allows the 9037 time to be set from an ETS device, but does not provide the capability to continuously track the 9037 time to the ETS device. The expanded availability configuration and expanded basic configuration configurations provide this capability. This tracking is also referred to as *steering*.

In the case of a 9037-2, an ETS device can attach directly to a 9037-2 unit using an RS-232-D serial port connection; refer to Figure 20 on page 77, and Figure 24 on page 80. In the case of a 9037-1, an ETS device attaches to the 9037-1 console using an RS-232-D serial port connection; refer to Figure 21 on page 77, and Figure 25 on page 80.

Accuracy with an External Time Source

For a 9037-2:

The external time source function can be used to maintain accuracy to within ± 1 ms of a stable external time source. Using the one pulse-per-second (PPS) high precision option of the external time source function, the accuracy can be increased to within ± 100 µs of a stable external time source.

For a 9037-1:

The external time source function can be used to maintain accuracy to within ± 5 ms of a stable external time source.

A variety of external time sources can be used. The three supported configurations for a 9037-2 are listed below. For a list of supported configurations for a 9037-1, refer to *Sysplex Timer Planning*.

- You can use an external modem that attaches to the ETS port of the 9037-2. This allows the 9037-2 to dial-up an available time service, such as the Automated Computer Time Service (ACTS) or international equivalent.
- You can use a time-code receiver that attaches to the ETS port of the 9037-2 and to an antenna. An optional 1 PPS output signal from the receiver can be added to improve the accuracy of the 9037.
- You can use a time-code generator or translator that attaches to the ETS port of the 9037-2 and to an external time code input. An optional 1 PPS output signal from the time-code generator or translator can be added to improve the accuracy of the 9037.

Note: For a complete listing of available time services, supported RS-232C protocols, and considerations to be used in selecting an external time source, refer to *Planning for the 9037 Model 2*.

3.2.6.1 Preferred External Time Source

The 9037-2 allows you to attach an ETS to both units in an expanded availability configuration. The ETS source attached to one of the 9037-2s has to be designated as the *preferred* ETS at the console. The ETS source attached to the other 9037-2 automatically gets assigned as the backup. The preferred ETS is the time source toward which the Sysplex Timer network will steer. If the preferred ETS fails, the backup ETS automatically becomes the active ETS to which the Sysplex Timer network is steered. When the preferred ETS is restored, it becomes the active ETS automatically.

Note: The two ETSes can be of different types. For example, the preferred ETS could receive signals from a radio receiver, and the backup could be a dial-up service.

3.2.7 9037 Distance Limitations

There are distance limitations on:

- How far the 9037s can be located from the attached CPCs
- How far two 9037s in an expanded availability configuration can be separated from each other
- How far the console can be located from the 9037s

3.2.7.1 9037-2- or 9037-1-to-CPC Distance Limitations

Note: Both 9037-2 and 9037-1 have the same Timer-to-CPC distance limitation.

The total cable distance between the 9037 and the CPC attachment port, which includes jumper cables, trunk cables, and any distribution panels required, cannot exceed 3 km for 62.5/125-micrometer fiber and 2 km for 50/125-micrometer fiber. When installing these cables, an additional 2 m of slack must be provided to allow for the positioning of the 9037 and the cable routing within the 9037.

The distance of 3 km can be extended to 26 km and the distance of 2 km can be extended to 24 km by routing the links through the 9036 Model 3 repeater (RPQ 8K1919). See 3.3, “9037 RPQs” on page 91 for more details of RPQ 8K1919.

The total cable distance between the 9037 and the CPC attachment port can further be extended to 40 km by routing the links through a pair of 9729s or a pair of 2029s (both the 9729 and 2029 require RPQ 8P1955).

See 3.3, “9037 RPQs” on page 91 for more details of RPQ 8P1955.

3.2.7.2 9037-2-to-9037-2 Distance Limitations

The total cable distance between 9037-2 units, which includes jumper cables, trunk cables, and any distribution panels required, cannot exceed 3 km for 62.5/125-micrometer fiber and 2 km for 50/125-micrometer fiber. When installing these cables, an additional 2 m of slack must be provided to allow for the positioning of the 9037 and the cable routing within the 9037.

The distance of 3 km can be extended to 26 km and the distance of 2 km can be extended to 24 km by routing the 9037 links through a 9036 Model 3 repeater (RPQ 8K1919). The total cable distance between the 9037 and the CPC attachment port can be extended to 40 km by routing the links through a pair of 9729s or a pair of 2029s (both the 9729 and 2029 require RPQ 8P1955).

See 3.3, “9037 RPQs” on page 91 for more details of RPQ 8P1955.

See 3.3, “9037 RPQs” on page 91 for more details of RPQ 8K1919.

3.2.7.3 9037-1-to-9037-1 Distance Limitations

The distance between the two 9037-1 units in an expanded availability configuration should not exceed 2.2 m. This includes any allowance for internal routing and slack.

3.2.7.4 9037-2-to-9037-2 Console Distance Limitations

The console can be located anywhere up to the limit of the installed Token-Ring. It can also connect to the local LAN containing the 9037-2 through a bridged LAN. Figure 29 on page 86 shows a bridge between the two Token-Rings. The 9037-2 console *cannot connect* to a 9037-2 through a TCP/IP router.

3.2.7.5 9037-1-to-9037-1 Console Distance Limitations

The total distance between the 9037-1 console and the 9037-1 should not exceed 2.2 m. This includes any allowance for internal routing and slack.

3.2.8 Planning for Cabling Requirements

It is important to note which cables are supplied with the 9037 and which cables you, the customer, are expected to supply.

3.2.8.1 IBM-Supplied Cables

The following cables are supplied with each 9037-2:

- A 6.1 m cable to connect the 9037-2 to the Token-Ring
- An AC power cord which is site-dependent and varies in length and outlet plug type

The console adapter cables used to attach the console to the 9037-1s, and the control link cables used to attach the two 9037-1s to each other (expanded availability configuration), are provided with the 9037-1.

3.2.8.2 Customer-Supplied Cables

You, the customer, must supply:

- The fiber optic cabling used to attach the 9037s to the sysplex attachment feature of each CPC
- Two fiber optic cables used to attach the 9037-2 units to each other
- Any cables required to attach the external time source, if the external time source function is used.

  Note: A null-modem cable or adapter is required if the ETS device is wired as Data Terminal Equipment (DTE).

A fiber optic cabling environment could use jumper cables, trunk cables, distribution panels, and various connector types. Either 62.5/125-micrometer or 50/125-micrometer multimode fiber can be used. The distance limitations are outlined in 3.2.7, “9037 Distance Limitations” on page 88. Refer to 3.3, “9037 RPQs” on page 91 if single-mode fiber support or extended distance support is required.

Each connection from a 9037 port to a CPC attachment port requires two individual fibers, one that carries signals from the 9037 to the CPC and another that carries signals from the CPC back to the 9037. Each jumper cable is comprised of these two individual fibers. A trunk cable has many individual strands of fiber cables. The 9037 automatically compensates for propagation delay through the fiber optic cables, thus allowing each connection from the 9037 to the CPC to be of different length.

It is important to note that for the effects of different fiber cable lengths to be nearly transparent to CPC TOD clock synchronization, the difference in length between the pair of individual fibers that make up a connection *must be less than 10 m*. Only then can the 9037 perform an effective propagation delay compensation. This specification for the allowable difference in length applies to both the connection between the 9037 and the CPCs and, in the case of the 9037-2, to the connection between 9037-2 units in an expanded availability configuration configuration. Maintaining a difference of less than 10 m is not a problem with duplex-to-duplex jumper cables, where the fibers are the same length. However, this is an important consideration when laying trunk cables and using distribution panels.

Make sure you have reviewed important considerations for routing cables in an expanded availability configuration, outlined in 3.5, “Recommendations for Continuous Sysplex Timer Availability” on page 94.

3.3 9037 RPQs

The following 9037 RPQs are available to you when required:

- 8K1903 (9037 RPQ Pointer to RPQ 8K1919)

  RPQ 8K1903 is a 9037 RPQ, which refers the reader to RPQ 8K1919 for the special purpose 9036-3 Extender.

- 8K1919 (single-mode fiber and extended distance support)

  RPQ 8K1919, which is a 9036-3, was designed specifically to handle timer signals. Its primary function is to convert ETR signals from a multimode fiber interface to a single-mode fiber interface. The ETR signals are not modified as they are routed through the 9036-3. Two 9036-3 units (two RPQs) are required for each link, since both the 9037 and the Sysplex Timer attachment feature support only multimode fiber.

  RPQ 8K1919 allows the following two Sysplex Timer configuration requirements to be met:

  1. It allows the distance between the 9037 and the CPC, or the distance between 9037-2 units, to be extended to a maximum cabling distance of 26 km.
  2. When the Sysplex Timers are located in two separate sites and only single-mode fiber is available from a telecommunications carrier to make the cross-site connections either from the 9037 to the CPC or between 9037-2 units, the 9036-3 provides the necessary conversion, even if the cabling distance is less than 3 km.

  Note: ESCON directors (9032 or 9033), or 9036-1 or 9036-2 converters *cannot* be used instead of the 9036-3, since they do not support the ETR architecture.

- 8K1787 (Basic configuration ETS Tracking)

  RPQ 8K1787 has been withdrawn. This RPQ provided an ordering vehicle for the Feature Code required to perform the ETS tracking for the 9037-1 basic configuration machines. This Feature Code may now be ordered with a 9037-1 basic configuration; thus, RPQ 8K1787 is no longer available.

- 8P1955 (Extended distance support RPQ)

  This RPQ documents that S/390 Development has approved the requested deviation from the allowed link distance, link budget or other physical layer specification for Sysplex Timer and Coupling Facility Links when used in conjunction with the IBM 9729 Wavelength Division Multiplexer and the 2029 Fiber Saver. For example, these distances may be required when implementing Geographically Dispersed Parallel Sysplex.

  This RPQ is valid for the 9729 Model 001 and the 9729 Model 041. This RPQ also supports Fiber Saver, machine type 2029, models: RS0, RS1, RS2, RS3, RS4, 001, 002, 003 and 004.

  No hardware or software is provided with this RPQ. However, this RPQ must be ordered to document and ensure that the machine records reflect S/390 development's approval for the deviation from the allowed link distance, link budget or other physical layer specification for Sysplex Timer and Coupling Facility Links.

3.4 Migration from a 9037-1 to a 9037-2 Sysplex Timer Configuration

If you have a Parallel Sysplex and are using the 9037-1, you may want to install a 9037-2 because of the features it provides. This section discusses the supported and unsupported migration procedures and the requirements that need to be met during this migration.

3.4.1 Requirements during Sysplex Timer Migration

When migrating from a 9037-1 to a 9037-2, the fundamental requirement is to *maintain data integrity*. Also, since one of the key attributes of the Parallel Sysplex is continuous availability, the expectation is to perform the migration without any outage, or at worst a limited outage (not all systems), of the Parallel Sysplex.

3.4.2 Sysplex Timer Migration Procedures Not Supported

It is important to understand that any migration considerations involve the complete replacement of an existing 9037-1 with a 9037-2.

3.4.2.1 9037-1/9037-2 Coexistence Not Supported

You *cannot connect* a 9037-1 to a 9037-2 in an expanded availability configuration, as shown in Figure 30.

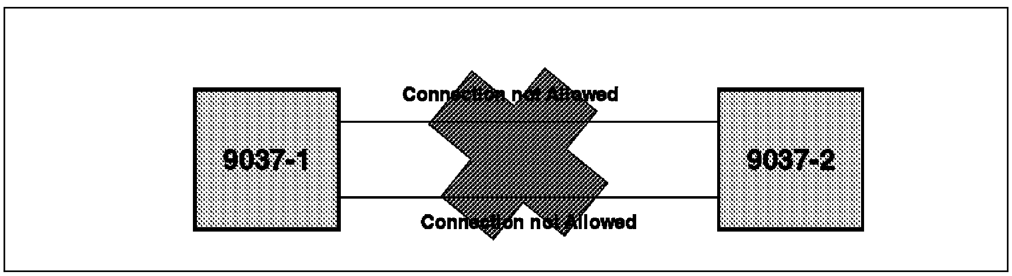


Figure 30. Non-Configurable Expanded Availability Configuration

3.4.2.2 MES Upgrade of 9037-1 Not Supported

The 9037-2 is a completely redesigned box with a different chassis, power supply, card form factors, interfaces to the console, and so on. Therefore, you *cannot* upgrade (MES) from a 9037-1 to a 9037-2 by replacing cards or upgrading code.

3.4.2.3 Migration with Continuous Availability of Parallel Sysplex Not Supported

Since a 9037-1 and a 9037-2 cannot be synchronized, there is no procedure that will allow *all* the CPCs in a Parallel Sysplex to continue participating in data sharing during the migration window. A potential risk of compromising data integrity exists, if such a migration is attempted. However, a migration with limited outage of Parallel Sysplex is supported, as outlined in 3.4.3, "Supported Sysplex Timer Migration Procedures" on page 93.

To illustrate the potential data integrity exposure, consider the following:

Assume that in Figure 31 on page 93, all the CPCs are participating in data sharing. At one stage of the migration procedure, the Time-of-Day clocks of all the CPCs step to signals received from the 9037-1 (Port 0 of each CPC in this example), and the alternate ports of all CPCs are connected to enabled ports from the 9037-2.

If a failure were to occur at this stage, in the connection between the 9037-1 and CPC 1, such that Port 0 of CPC 1 is non-operational, the hardware in CPC 1 automatically switches ports and starts stepping to signals received on Port 1 from the 9037-2. A *data integrity exposure* now exists, since time stamps for images in CPC 1 are obtained from the 9037-2, while time stamps for images in CPCs 2 and 3 are obtained from a non-synchronized 9037-1.

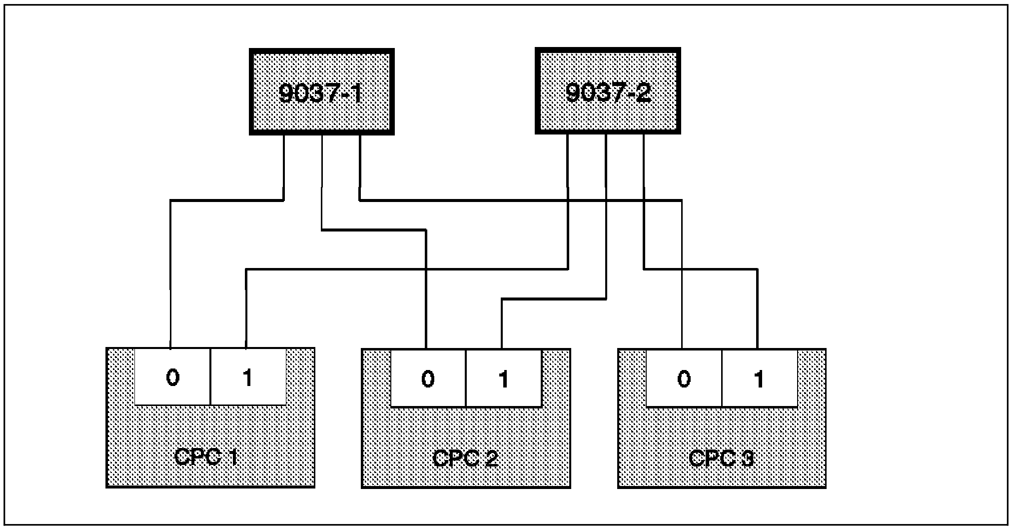


Figure 31. Non-Supported Migration: All CPCs Continuously Available

Observe the following in Figure 31:

- All CPCs are participating in data sharing during the migration.
- 9037-1 and 9037-2 are not synchronized.

3.4.3 Supported Sysplex Timer Migration Procedures

Two procedures have been developed that meet the fundamental requirement of maintaining data integrity during the migration, even though they result in either a limited outage or a planned outage of the Parallel Sysplex. They are:

1. Migration with limited outage of sysplex

   This procedure meets the fundamental requirement of maintaining data integrity during the migration. It allows one of the CPCs in a multisystem sysplex to *continue processing workloads* during the migration. All the OS/390 images on this CPC can continue processing during the migration. On each of the remaining CPCs in the Parallel Sysplex, subsystems such as CICS, IMS, and DB2 that participate in data sharing are shut down. By shutting down these subsystems, new requests for work cannot be processed by these CPCs during the migration window. OS/390 does not need to be shut down on any of the CPCs during the migration. By not having to re-IPL OS/390 on any of the CPCs, the time required for the outage can be minimized.

2. Migration with planned outage of sysplex

As the name implies, you must plan an outage and shut down all OS/390 images running with ETRMODE YES.

9037MIGR package on MKTTOOLS (available via your IBM service representative) provides the planning information required to determine which of these supported procedures is best suited for your configuration. The migration planning also minimizes the time required for a limited or planned outage.

3.5 Recommendations for Continuous Sysplex Timer Availability

Even though the 9037-2 expanded availability configuration is fault-tolerant to single points of failure, careful planning is required to maximize the probability that at least one 9037-2 will continue to transmit signals to the attached CPCs in the event of failures. The recommendations discussed here are to maximize availability of all components of a 9037-2 expanded availability configuration. This includes the 9037-2 units, consoles, external time sources, and any repeaters that may be needed for extended distance connectivity. Some of these recommendations may not apply to your configuration.

3.5.1 Sysplex Timer Recovery Overview

Before outlining the recommendations, it is important to understand the recovery design when two Sysplex Timers are configured in an expanded availability configuration. In an expanded availability configuration, the TOD clocks in the two Sysplex Timers are synchronized using the hardware on the CLO card and the CLO links between the Sysplex Timers.

In Figure 32 on page 95, both Sysplex Timers are simultaneously transmitting the same time-synchronized data to the CPCs, and no CPC information is transmitted back to the Sysplex Timers. The Time-of-Day (TOD) clocks in each CPC can step to signals received from either CPC port (0 or 1). For example, the TODs in CPCs 1 and 3 could be stepping to signals received from 9037-2 .A. and the TOD in CPC 2 could be stepping to signals received from 9037-2 .B..

When the CPC hardware detects the stepping port to be non-operational, it forces an automatic port switchover and now the TOD steps to signals received from the "other" 9037-2. The Sysplex Timers do not switch over, and are unaware of the port change at the CPC end.

Because the ETR system design allows TODs in each CPC to step to ETR signals received at either port 0 or port 1, the following Sysplex Timer design rule *must* be implemented in order to ensure data integrity among the different OS/390 images participating in a Parallel Sysplex:

If, at any time, two Sysplex Timers in an expanded availability configuration lose the capability to synchronize with each other, at least one of the Sysplex Timers must discontinue transmission to the CPCs.

To implement this rule, the Sysplex Timers in an expanded availability configuration arbitrate at initialization time to determine which Sysplex Timer is primary and which is secondary. The primary and secondary designation is not transmitted to the CPCs.

The significance of the primary Sysplex Timer is twofold. First, this is the Sysplex Timer that will continue to transmit (as long as it is operational) to the attached CPCs, in the event of loss of synchronism between Sysplex Timers.

Also, if there is an ETS device in the Sysplex Timer network, both Sysplex Timers will track to the ETS device that is attached to the primary Sysplex Timer.

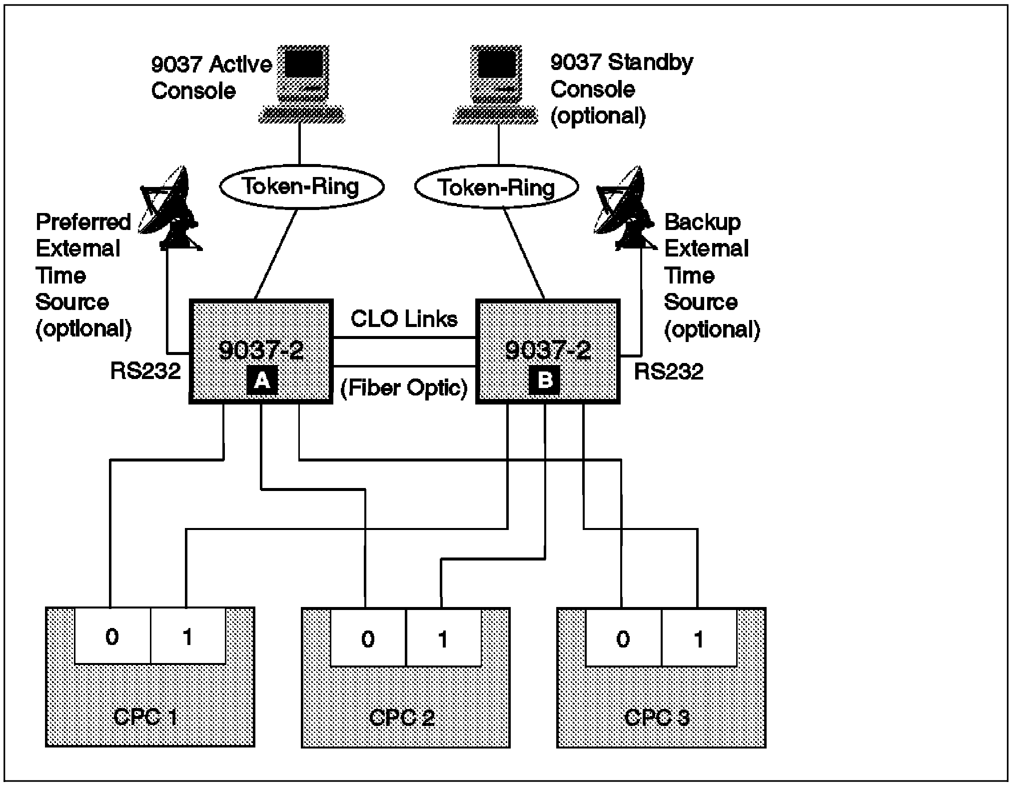


Figure 32. Example of Multisystem Sysplex with 3 CPCs

When a 9037-2 detects a failure (hardware component, software, power outage, and so on), it transmits a special sequence of symbols called the Off Line Sequence (OLS) on the CLO links to signal the "other" timer that it is going offline (that is, discontinuing transmission to the attached CPCs).

The following are recovery rules pertaining to OLS:

- Whether a primary 9037-2 receives or does not receive OLS from the secondary 9037-2, it maintains its designation as the primary and continues to transmit to the CPCs.
- If a secondary 9037-2 receives OLS from the primary, the secondary 9037-2 becomes the primary and continues to transmit to the CPCs.
- If a secondary 9037-2 detects a loss of signalling from the primary 9037-2, but does not receive OLS from the primary, it discontinues transmission to the attached CPCs. As the secondary 9037-2 does not know whether the primary 9037-2 is still functioning, it must discontinue transmission in order to ensure that the attached CPCs do not receive two non-synchronized signals from the two timers.
- If a failure is detected on only one of the control links (CLO links) between the 9037-2s, OLS is not transmitted by either 9037-2, and both timers continue to transmit to the CPCs.

These rules ensure that at least one 9037-2 continues to transmit for the following types of failures:

- Hardware or software failures detectable in each timer

- Power outage of either timer
- Failure of one or both CLO links
- Failures of extenders used between timers
- A disaster at the data center with the secondary timer

Note: This is not a comprehensive list of failures.

3.5.2 Primary Timer Assignment of a 9037-2

The following recommendations allow you to indirectly enforce which 9037-2 in an expanded availability configuration is assigned as the primary at initialization time. It should be noted that certain failures after initialization may necessitate a temporary switchover of the primary designation, until the failure is repaired.

To ensure that a 9037-2 located in a data center is assigned as primary during the arbitration performed during initialization, you must implement the following:

- If your configuration is using ETS, the preferred ETS must be attached to the 9037-2 that you want assigned as the primary 9037-2. See 3.2.6.1, "Preferred External Time Source" on page 88 for more details. Also, when defining the Sysplex Timer Network to the active console, use the IP address of the 9037-2 that you want assigned as the primary 9037-2.
- If your configuration is not using ETS, use the IP address of the 9037-2 that you want assigned as the primary 9037-2, when you are defining the Sysplex Timer Network to the active console.

Recommendation for 9037-2 Availability

The two timers should be physically separated. The greater the separation, the lower the probability that an environmental failure will affect both units.

For continuous ETS availability, an ETS device should be attached to each 9037-2. See 3.2.6.1, "Preferred External Time Source" on page 88 for more details.

For continuous availability of consoles, an active and a standby console should be configured. See 3.2.5, "9037-2 Console Selection and Considerations" on page 85 for more details.

3.5.3 Power Sources for Timers and Extenders

The following recommendations apply to the Sysplex Timers and the extenders between the 9037-2s:

- Each 9037-2 unit should be connected to a separate AC power source. At a minimum, each 9037-2 should be connected to a separate power circuit.

  If a 9037-2 detects a power outage, it transmits an OLS on the CLO links between the 9037-2s before completely shutting down. If the primary 9037-2 detects a power outage, receipt of the OLS signal by the secondary 9037-2 is critical to allow the secondary 9037-2 to become the primary and continue transmission to the CPCs. If the OLS signal is not received by the secondary 9037-2, it also discontinues transmission of signals to the attached CPCs. This places all OS/390 images in a non-restartable wait state.

- If you have extenders between timers, you should connect at least one of the pairs of extenders (same CLO link) to a separate power source from the one used by the primary 9037-2. By doing so, you greatly improve the probability that the OLS signal from the primary timer is received by the secondary timer during a power outage. If the primary 9037-2 and the extenders on each CLO link experience a power outage at exactly the same time, the OLS signal may not be received by the secondary timer, resulting in all OS/390 images to be placed in a non-restartable wait state.

- Protect the power connections by adding a dedicated Uninterruptible Power Supply (UPS) for each 9037-2 and extender.

3.5.4 Fiber Optic Cable Routing

The following recommendations apply to the fiber optic links between 9037-2s (CLO links) and the links from each 9037-2 to the attached CPCs:

- Each CLO link should be routed across a separate path. The greater the separation, the lower the probability that an environmental failure (accidental or otherwise) will affect both links. In addition, the CLO links should be routed across separate paths from the fiber optic cables that attach the 9037-2 to the CPCs. If the CLO links are accidentally severed, the 9037-2 in the primary data center is still able to transmit to the CPCs located in the secondary or remote data center.
- Separately route the redundant fiber optic cables from each 9037-2 to the CPC across a separate path.

3.6 Concurrent Maintenance and Upgrade Considerations

The considerations of concurrent maintenance and upgrade are different for a basic and expanded basic configuration, compared to an expanded availability configuration. The expanded availability configuration offers the best capability for both concurrent maintenance and nondisruptive upgrade.

3.6.1 Concurrent Maintenance of Sysplex Timers

Concurrent maintenance of all 9037 Field Replaceable Units (FRUs) is possible in a correctly configured 9037 expanded availability configuration. The 9037 requiring repair can be powered off to replace the FRU, since the other 9037 continues to transmit time information to the attached CPC.

Note: Verify that the 9037 ports are configured as described in 3.2.1, "Planning for Availability of Sysplex Timer" on page 75, before powering off either 9037.

The 9037-2 expanded availability configuration also allows concurrent download of Licensed Internal Code (LIC).

In a 9037 basic configuration or expanded basic configuration, limited concurrent maintenance can be performed. If the failure is associated with a specific fiber optic port or associated link and there are spare ports available, the repair can be done concurrently by moving the transmission of ETR signals to the spare port.

The 9037-2's port cards are hot-pluggable. This means that they can be removed or replaced concurrently even in a basic configuration or expanded basic configuration.

If the 9037 console or console cables need to be repaired, concurrent maintenance can also be performed in the basic or expanded basic configuration, since the 9037 console can be powered off without impacting the transmission of ETR signals.

3.6.2 Upgrade Options for Sysplex Timers

The following upgrade options are available:

- Upgrade from a basic to an expanded basic or expanded availability configuration
- Upgrade from an expanded basic to an expanded availability configuration
- Increase the number of port attachments (in groups of four)

Upgrading from a basic to an expanded basic or expanded availability configuration is always a disruptive process. The single 9037 running in the basic configuration has to be powered off to add the card required to synchronize the two 9037s.

Upgrading from an expanded basic to an expanded availability configuration can be done concurrently and is therefore recommended if any future planning includes adding a second 9037-2.

The 9037-2's port cards are hot pluggable. This means that increasing the number of port cards can be done concurrently in all three 9037-2 configurations.

In the case of the 9037-1, adding port cards to a basic configuration requires that the 9037-1 be powered off. In a 9037-1 expanded availability configuration, it is possible to add port cards without the attached CPCs losing time information. This is essentially accomplished by removing and upgrading one 9037-1 at a time while the other continues to transmit to the attached CPCs.

Note: Verify that the 9037-1 ports are configured as described in 3.2.1, "Planning for Availability of Sysplex Timer" on page 75, before powering off either 9037-1.

3.7 Planning for System Availability

APAR OW44231 makes changes to the way OS/390 reacts to the loss of all timer signals. Prior to this APAR, if all timer signals are lost, OS/390 will place itself in a disabled wait.

This APAR changes this so that OS/390 will now enter a spin loop if all timer signals are lost. A synchronous WTOR (IEA015A) is presented to the operator asking if he wishes to RETRY to see if the signal has been restored or if he wishes to CONTINUE. If RETRY is specified, the system will resume operation if the timer signal has been restored. If not, message IEA015A will be reissued and the system will once again enter a spin loop.

If CONTINUE is specified, the system will enter a disabled wait state if the system is a member of a multisystem sysplex. If the system is not a member of a multisystem sysplex, it will resume operation in local TOD clock mode.

The application of this APAR is highly recommended, especially in a GDPS environment, where it is possible that the systems in one center continue to operate, but the timer link to the alternate site is severed. In this case, the timer in the alternate location will discontinue transmission because it has lost contact with the primary timer and has not received an OLS signal. In this case, you should attempt to repair the connection, and reply RETRY to restart the systems.

However, if the connection cannot be repaired, and then you should reply CONTINUE to stop all systems that are in the same sysplex as the systems in the primary site.

More information about the impact of this APAR is available in WSC FLASH 20006, *Surviving a 9037-002 Sysplex Timer Failure - OS/390 APAR Enhancement*.

3.7.1 SFM Considerations

There are two scenarios that can affect SFM. The first is when only some of the systems in the sysplex lose their signal, and subsequently issue the IEA015A message.

In this case, SFM on the surviving systems will detect a STATUS UPDATE MISSING condition after their failure-detection-interval expires, assuming that the affected systems have not been successfully restarted during that time. Once the STATUS UPDATE MISSING is detected, the affected systems will be partitioned out of the sysplex.

You *may* want to reconsider your failure detection interval to allow for this situation; however, you need to weigh the likelihood of losing timer signals to a subset of the systems against the likelihood of one system having a problem and holding on to resources longer than necessary, while waiting for the failure detection interval to expire.

The other scenario is when all systems in the sysplex are affected. In this case, as soon as the first system is successfully restarted, it will immediately recognize STATUS UPDATE MISSING for the other members of the sysplex. If you have not specified PROMPT, SFM will start partitioning the other systems out of the sysplex as soon as ISOLATETIME expires, which may be before you get a chance to reply RETRY on those other systems.

You can work around this by increasing the value of ISOLATETIME, but once again you have to balance the likelihood of getting a complete loss of timer signals against the likelihood of one system hanging and holding on to vital resources for longer than necessary.

Another option is to specify the PROMPT keyword in your SFM policy, so that SFM will not automatically partition the other systems out of the sysplex. However, this option is not recommended as it would potentially cause unacceptable delays in the partitioning of systems that should be partitioned.

Chapter 4. Consoles and Parallel Sysplex

This chapter discusses the configuration aspects of consoles in a Parallel Sysplex. The Hardware Management Console (HMC) associated with the 9672 family of CPCs is also discussed.

Recommended Sources of Further Information

The following sources provide support for the information in this chapter:

- *Controlling S/390 CMOS Processors Using the HMC*, SG24-4832.
- *DCAF V1.3 Installation and Configuration Guide*, SH19-4068
- *DCAF V1.3 User's Guide*, SH19-4069
- *Getting the Most Out of a Parallel Sysplex*, SG24-2073
- *Hardware Management Console Guide*, GC38-0453
- *Hardware Management Console Operations Guide*, GC38-0470
- *IBM Token-Ring Network Introduction and Planning Guide*, GA27-3677
- *Managing Your Processors*, GC38-0452
- *MVS/ESA SP JES3 V5 Implementation Guide*, SG24-4582
- *OS/390 MVS Initialization and Tuning Guide*, SC28-1751
- *OS/390 MVS Multisystem Consoles Implementing Sysplex Operations*, SG24-4626
- *OS/390 MVS Planning: Operations*, GC28-1760
- *Programming Interfaces*, SC28-8141
- *System Overview - G6*, GA22-1030
- *TCP/IP Tutorial and Technical Overview*, SG24-3376

The following roadmap guides you through the chapter, by providing a quick reference to help you find the appropriate section.

You want to configure:	If you are especially interested in:	Subtopic of interest:	Then refer to:
Consoles			
	Multisystem consoles		4.1, “Multisystem Consoles in Parallel Sysplex” on page 102
		OS/390 console considerations	4.1.1, “OS/390 Consoles in Parallel Sysplex” on page 103
		How many consoles do I need?	4.1.2, “How Many Consoles Do You Need” on page 108
		Dealing with automation	4.1.4, “Automation Implications” on page 110
	Hardware Management Console		4.2, “Hardware Management Consoles in a Parallel Sysplex” on page 111
		What is the HMC?	4.2.1, “Hardware Management Console (HMC)” on page 111
		What can I do with the HMC LAN?	4.2.3, “Configuring the LAN” on page 113
		Can I have a remote HMC?	4.2.4, “Operating the HMC Remotely” on page 113
		Can I automate HMC operation?	4.2.5, “Automating Operation of the IBM 9672” on page 114
		Can I use HMC as an OS/390 console?	4.2.6, “Can I Use Integrated Consoles Instead of OS/390 Consoles” on page 116
		How many HMCs should I order?	4.2.7, “How Many HMCs Should I Order” on page 117
		What security should I consider?	4.2.11, “Security Considerations” on page 118

Table 6. Consoles and Parallel Sysplex Roadmap

4.1 Multisystem Consoles in Parallel Sysplex

The introduction of Parallel Sysplex into the OS/390 environment provides a simpler and more flexible way to operate consoles in a multisystem environment. Many changes were introduced into multiple console support (MCS) to support the Parallel Sysplex environment. These changes began with MVS/ESA SP V4 and have continued with each new OS/390 release.

Since the introduction of sysplex, a relationship exists between each console on separate OS/390 systems in the same sysplex. Any console in the sysplex can operate any OS/390 system in the sysplex. They can all be considered as part of the same console ring. It is important to understand this difference and configure your consoles correctly due to the impact this may have on availability.

The console address spaces form an XCF group, which makes communication possible between any of the images and any of the active consoles in the

Parallel Sysplex.

Therefore, any console can:

- Be attached to any system.
- Receive messages from any system in the Parallel Sysplex.
- Route commands to any system in the Parallel Sysplex.

This also means that the limit of 99 consoles per image is now a limit across the Parallel Sysplex, and because all consoles are known to each image, they must be uniquely identified.

Therefore, new factors need to be considered when defining MCS consoles in this environment, such as:

- There can only be one master console active in the sysplex; however, multiple consoles can have master command authority.
- The 99-console limit for the Parallel Sysplex can be extended with the use of extended MCS consoles.
- MCS consoles, including subsystem consoles, should be named.
- A Parallel Sysplex, which can be up to 32 systems, can be operated from a single console.
- There is no requirement that each system have consoles attached.

For detailed information on all aspects of Multisystem Consoles in a Parallel Sysplex, refer to the ITSO redbook, *OS/390 MVS Multisystem Consoles Implementing MVS Sysplex Operations*.

4.1.1 OS/390 Consoles in Parallel Sysplex

Cross-coupling services (XCF) enables MCS messages and commands to be transported between systems in the sysplex. This means that both MCS and extended MCS consoles can issue commands to, and receive replies from, any system in the sysplex. Because all consoles on all systems are known across an entire sysplex, you need to understand the new roles played by master and alternate consoles before planning the console configuration.

Additional documentation is available in *OS/390 MVS Planning: Operations* and *OS/390 MVS Initialization and Tuning Reference*.

In the OS/390 world, there are the following different types of consoles, as shown in Figure 33 on page 104:

- MCS consoles
- Extended MCS consoles
- Integrated (System) consoles
- Subsystem consoles

It is mainly MCS and extended MCS consoles that are affected by changes in a Parallel Sysplex configuration and require some planning consideration.

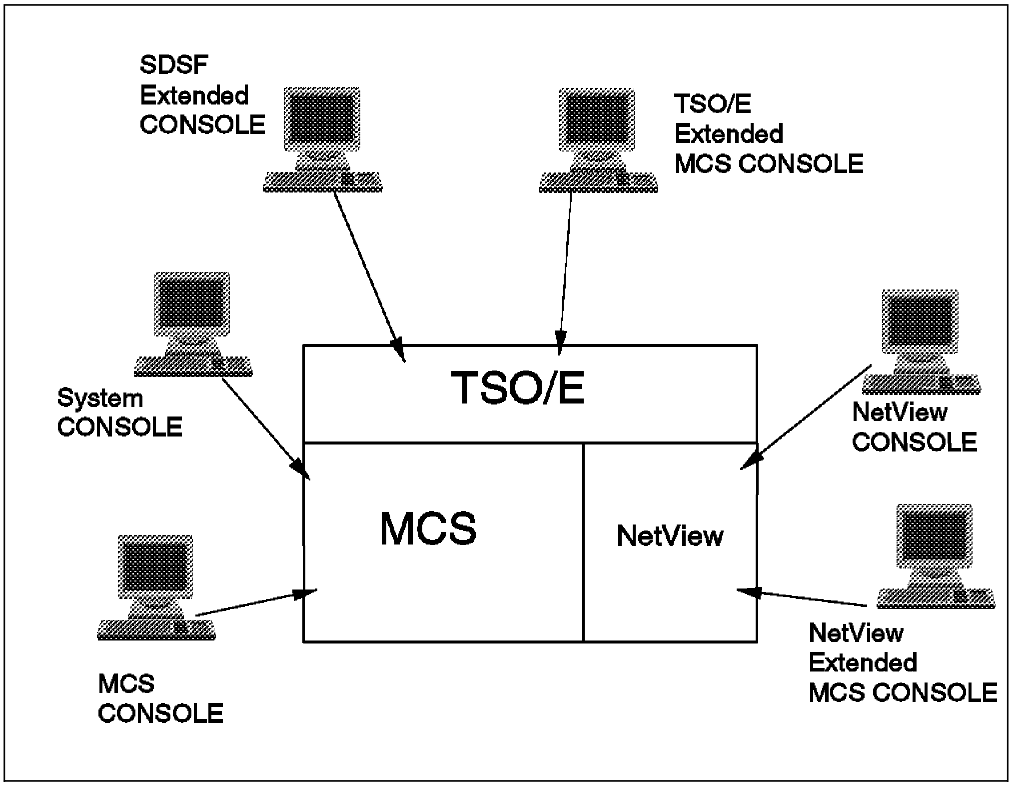


Figure 33. Console Environment

4.1.1.1 MCS Consoles

MCS consoles are display devices that are attached to an OS/390 image and provide the basic communication between operators and OS/390. MCS consoles must be locally channel-attached to non-SNA 3x74 control units; there is no MCS console support for any SNA-attached devices.

Note: An ESCON-attached 3174 can be used; however, it must be configured in non-SNA mode. This also means that it will only support a single CPC because multi-CPC ESCON support requires SNA mode on the 3174.

You can define MCS consoles in a console configuration according to different functions. For example, one MCS console can function as a master console for the sysplex, while another MCS console can be used as a monitor to display messages to an operator working in a functional area, such as a tape pool library, or to display printer-related messages.

4.1.1.2 Extended MCS Consoles

An extended MCS console is actually a program that acts as a console. It is used to issue OS/390 commands and to receive command responses, unsolicited message traffic, and the hardcopy message set. There are two ways to use extended MCS consoles:

- Interactively, through IBM products such as TSO/E, SDSF, and NetView
- Through user-written application programs

Generally speaking, an extended MCS console is used for almost any of the functions that are performed from an MCS console. It can also be controlled in a manner that is similar to an MCS console. For example, you can control command authorization and message routing, provide alternate consoles, and so on, as from an MCS console. In addition, because extended MCS consoles are nothing more than the exploitation of a programming interface, there is no restriction that the consoles must be locally channel-attached to a non-SNA 3x74 control unit. TSO/E's and SDSF's console implementation both allow you to establish an extended MCS console session using a terminal that is SNA-attached.

Note: The major drawback to using a TSO/E console session in place of an MCS console is that it is not a full-screen application. When you enter a command, you must continue to press Enter until the command response is displayed. There is an IBM-supplied sample application that makes a TSO/E console session more usable in that it basically simulates full-screen mode. The documentation on the application is found in member IEATOPSD in SYS1.SAMPLIB.

SDSF has a similar problem to TSO/E. It has an 8-line display area under the command line where command responses are received. Occasionally command responses are not received within sufficient time for SDSF to display, or the response received is greater than 8 lines. However, full responses are always received and logged by SDSF and can be viewed using the ULOG option of SDSF.

4.1.1.3 Integrated System Console

This term refers to the interface provided by the Hardware Management Console (HMC) on an IBM 9672. It is referred to as SYSCONS and does not have a device number. There are three system functions which might use this interface:

- Nucleus Initialization Program (NIP)
- Disabled Console Communication Facility (DCCF)
- Multiple Console Support (MCS)

SYSCONS cannot be specified as a console to NIP. However, if none of the consoles specified in NIPCONS are available, then NIP automatically uses the interface.

DCCF is invoked if OS/390 is unable to continue processing without intervention from the operator. If no consoles are available to the image, then SYSCONS is used and no definition is required.

The definition of SYSCONS to MCS has several unique features. In the CONSOLxx member of SYS1.PARMLIB, it does not have to be given a name and it does not have a device address; the parameter is specified as DEVNUM(SYSCONS). This definition can be used by every system in the Parallel Sysplex because the console is automatically assigned a name equal to the OS/390 system name. This console is assigned routing codes, but in normal operation no messages are sent to it.

Because this interface is rather slow, it is not recommended that this console be used except where there is a very low message volume and it is an emergency situation.

If this console is used to IPL a system, it will stop receiving messages once the system has joined the Parallel Sysplex, and messages from this system will be routed to other consoles in the Parallel Sysplex.

In summary, this console should be regarded as a "last resort." For a more detailed discussion refer to *OS/390 MVS Multisystem Consoles Implementing MVS Sysplex Operations*.

4.1.1.4 Subsystem Console

Subsystem allocatable consoles are console definitions that are available to authorized applications, such as JES3 or NetView, for example. These consoles are allocated and used by the individual applications through the subsystem interface (SSI).

4.1.1.5 Master Console

When not running as a member of a Parallel Sysplex, every system has a master console. When running in a Parallel Sysplex, however, there is only *one* master console for the entire Parallel Sysplex, regardless of the number of systems. There are other consoles, though, that can be defined to have the same authority as the master console.

Initially, the master console will be determined by the system that initializes the Parallel Sysplex and will be the first available MCS console that has an authority of master in its definition in CONSOLxx.[2]

Subsequent consoles defined with an authority of master are simply consoles with master authority. So what is so unique about the master console? Actually, there is little distinction between the master console and a console with master authority. Here are the differences:

- The master console is the last resort for a console switch. If a console cannot be switched anywhere else, it will be switched to the master console.
- The master console receives undeliverable (UD) messages when there is no other console that is eligible. These are messages for which there is no console available that is eligible to receive the message. For example, if a write-to-operator (WTO) was issued with a route code of 27 and there was no console online that was eligible to receive that route code, the message would be considered a UD message.
- The master console receives messages issued to CONSID=0. These are such things as command responses to IEACMD00/COMMNDxx,[3] issued commands, and a variety of system-initiated messages.

Notice about Master Consoles in Parallel Sysplex

There is only one active master console in the entire Parallel Sysplex. Therefore, when planning the console configuration, you must ensure that no matter which console is the master, there is always an alternate console available somewhere in the Parallel Sysplex. This is necessary so that the console can be *switched* should the master console fail or the system to which it is attached be taken out of the sysplex. This is true for both scheduled and unscheduled outages.

2 A member in SYS1.PARMLIB

3 More members of SYS1.PARMLIB

4.1.1.6 Other Consoles in a Sysplex

You can define a maximum of 99 MCS consoles, *including* any subsystem allocatable consoles for an OS/390 image. In a Parallel Sysplex, the limit is also 99 consoles for the entire Parallel Sysplex, which means that you may have to consider this in your configuration planning. One possible way to alleviate this restriction is through the use of extended MCS consoles.

Note: Consoles that contribute to the maximum of 99 are those that are defined in CONSOLxx members through CONSOLE statements, MCS consoles, and subsystem consoles. It does not include extended MCS consoles.

> **Use NetView/390 and Upwards in Parallel Sysplex**
>
> NetView/390 supports both subsystem and extended MCS consoles.
>
> Because of the limit of 99 consoles in a sysplex, IBM recommends that NetView be implemented to use extended MCS consoles, if possible.

In non-sysplex implementations, it has long been a recommendation to have a pair of dedicated 3x74 control units on each system to avoid a single point of failure.

In a Parallel Sysplex implementation, there is no requirement that you have an MCS console on every image in the Parallel Sysplex. Using command and message routing capabilities, from one MCS console or extended MCS console, it is possible to control multiple systems in the Parallel Sysplex. Although MCS consoles are not required on all systems, you should plan the configuration carefully to ensure that there is an adequate number to handle the message traffic and to provide a valid configuration, regardless of the number of systems in the Parallel Sysplex at a time. For example, in a two CPC shop you should have two 3x74s. However, if you add a third CPC, the two existing 3x74s may be sufficient.

Note: When operating a system which does not have a physically attached MCS console, commands and responses are routed from other consoles in the sysplex using XCF services. In the case where all XCF signalling paths to this system are lost, the system will no longer be accessible using the other MCS consoles. Although some extended MCS consoles may be used, if the subsystems providing these extended MCS consoles are affected by the system problem, the integrated system console will be needed to enter commands to recover the signalling paths on this system.

The first system to be IPLed in a Parallel Sysplex should have an attached MCS console. Alternate consoles must also be considered across the entire sysplex, especially for the sysplex master console. You should plan the console configuration so that there is always an alternate to the sysplex master console available at all times. If you do not do so, unnecessary operator action will be required, and messages may be lost or sent to the hardcopy log. Many of the console-related parameters are sysplex in scope, which means:

- They are defined once and are in effect for the life of the sysplex; some may be changed through operator commands.
- They are known to all systems in the sysplex.

For example, the first system in the sysplex will establish the following:

Master console: The first console available with master authority

CNGRP: Which member is to be used, if any

RMAX: Maximum number for write-to-operator-with-reply (WTORs)

RLIM: Maximum number of outstanding WTORs allowed

AMRF: Whether the action message retention facility will be active

Note: For information on these parameters, refer to *OS/390 MVS Initialization and Tuning Reference*.

Consoles can and should be named, especially subsystem consoles. Subsystem consoles are treated as part of a sysplex-wide pool of subsystem-allocatable consoles available to any system in the sysplex. Because there is no system affinity to a subsystem console definition, even the subsystem consoles that were defined in a system's CONSOLxx member do not get deleted when that system leaves the Parallel Sysplex. When this system IPLs and rejoins the Parallel Sysplex, unless you have named the subsystem consoles, OS/390 has no way of knowing that the subsystem consoles had been defined previously and it will add them again.

As you can see, because subsystem consoles count against the maximum of 99 consoles, after a few IPLs you will have exceeded the limit. APAR OW05419 was created to address this problem. To be effective, the APAR requires that consoles are named.

4.1.2 How Many Consoles Do You Need

When planning a Parallel Sysplex, the issue of console configuration should be thoroughly reviewed and understood if operational problems are to be avoided.

Although it is not possible to eliminate all of the non-SNA 3x74s or MCS consoles, it may be possible to reduce the number from what was required when running the same number of individual single systems outside a Parallel Sysplex. Because of the multisystem console support in Parallel Sysplex, there is no longer a requirement that every system have MCS consoles physically attached. There is no longer a restriction that an MCS console have an alternate on the same system.

The situations that you must plan for are those involving the complete loss of the Parallel Sysplex, either through planned or unplanned outage.

A complete planned outage of the total Parallel Sysplex will be a rare occurrence. For better control, during the shutdown and startup procedure, you should ensure that the last system leaving the sysplex and the first system coming up have a physically attached console. This will make it easier to control the flow of the systems leaving or joining the sysplex, and facilitate a fast recovery in case of problems.

In the event of an unplanned outage, the requirement will be to get prime services online as quickly as possible. If there are specific systems (such as a network owner) that have to be available first, then these are good candidates for the placement of consoles.

It is also recommended that to manage the Parallel Sysplex, the configuration should include *at least* two screens capable of acting as OS/390 consoles. These screens should be attached through dedicated non-SNA control units, on different channels, on different CPCs. Figure 34 on page 109 shows an example configuration.

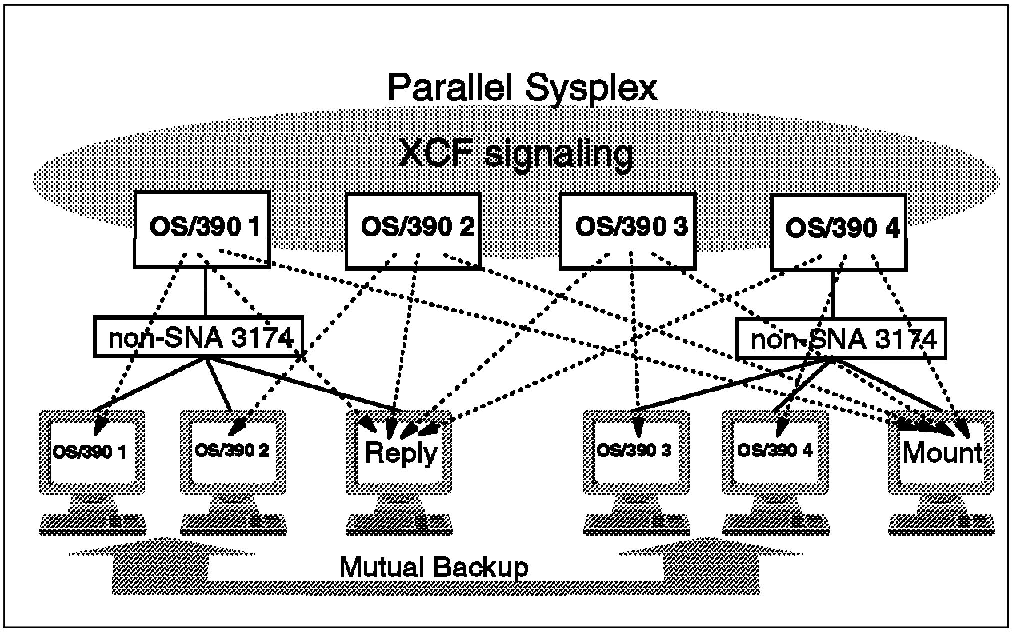


Figure 34. Example Console Configuration

4.1.3 Availability and Performance Considerations

As mentioned earlier in this chapter, it is necessary to ensure you have enough consoles in your configuration to handle the console traffic generated from the systems in the Parallel Sysplex. Incorrectly designed console configurations can cause performance problems in your Parallel Sysplex. Far worse than performance problems, is that in some cases this may lead to reduced availability.

Recommendations for Console Configuration

The following are very important considerations when designing your console configuration.

- Refer to WSC flash W99005 for important information regarding known console problems, recommendations and related APAR's. Flashes can be found at

  `http://www.ibm.com/support/techdocs/atsmastr.nsf`

- Ensure that the consoles you have are attached to the fastest CPCs in the sysplex. This will ensure that a small CPC is not trying to process huge amounts of console traffic generated by much larger OS/390 images.
- Keep console maintenance at the highest level available.
- Do regular searches in IBMLINK using the following search argument:

  `5752SC1CK HIPER`

  This will ensure the most critical console maintenance is applied to your systems.
- OEM products using console functions have often brought a console problem to the surface. Upgrading these products could produce errors not previously seen.
- Keep the number of consoles that have MSCOPE=*ALL to a minimum. This helps to minimize console traffic in the Parallel Sysplex. Configurations with a large number of consoles defined as MSCOPE=*ALL in a Parallel Sysplex have been known to have availability and performance problems.
- Use console groups and alternate groups instead of alternate consoles.
- Do not place the SYSZMCS resource in the GRS RNL list.
- If you experience any contention problems, use the following RMF III reports:
  - ENQR - Resource Enqueue Delays Report

    This shows enqueue delays on a system basis.
  - SYSENQ - Sysplex Enqueue Delays Report

    This shows enqueue delays on a sysplex wide basis

  Refer to B.1.6, "Parallel Sysplex RMF Monitor III Reports" on page 203 in Volume 2: Cookbook, SG24-5638 for more information.

4.1.4 Automation Implications

The greatest impact of this change to command and message processing is to the automation code residing in the message processing facility (MPF) exits and listening on the subsystem interface. Automation processing must take these changes into consideration:

- Automation can now use OS/390-provided services to have a command issued on a CPC other than the one on which it is operating. Furthermore, the command responses will be returned to the issuing system. Automation no longer has to provide its own vehicle for issuing "off-CPC" commands and retrieving the responses.

- Messages that automation gathers from the subsystem interface (SSI) can no longer be considered as always originating on the system where it resides. The automation product must determine which system issued the message, and act accordingly.
- Automation products can determine if the message originated "off-CPC" by bit WQERISS being set in the write-to-operator queue element (WQE). The originating system can be determined by examining the WMJMSNM field of the WQE.
- Automation code residing in an MPF message exit will see only messages originating on the system on which it resides.
- Solicited messages that automation sees through the SSI or MPF exits may not be destined for display on the system where the automation code resides.

4.2 Hardware Management Consoles in a Parallel Sysplex

This section covers the Hardware Management Console, or HMC for 9672. See 4.1.1, "OS/390 Consoles in Parallel Sysplex" on page 103 for information on OS/390 consoles. Much of the information in this section also applies to a non-sysplex environment.

4.2.1 Hardware Management Console (HMC)

Prior to the introduction of the IBM 9672, each CPC needed its own console to perform power-on resets (PORs), problem analysis, and other tasks necessary for system operation.

The Hardware Management Console (HMC) and the Service Element were introduced with the IBM 9672 range of CPCs. The SE is a PC, running OS/2, and is part of the IBM 9672; it provides an interface to the CPC. The HMC is a specialized IBM PC running an OS/2 application, and uses a Token-Ring LAN connection to communicate with the SE. When tasks are performed at the HMC, the commands are sent to one or more SEs through the LAN. The SEs then issue commands to their associated CPCs. CPCs can be grouped at the HMC, so that a single command can be passed along to all of the CPCs in the LAN, if desired. This can help to reduce the number of hardware consoles and to consolidate operations. For example, if you have a configuration with four 9672s, then there will be four CPCs and SEs that can be operated across the LAN by one HMC.

Each HMC can be connected to up to 32 SEs, and each SE can be connected to up to 16 HMCs.

The 9672 G6 CPCs introduced an alternate SE. This is optional on 9672 G5 CPCs. The alternate SE provides primary SE functions should the primary fail. To keep the two SEs synchronized, a daily mirroring function occurs between 10:00 am and 11:00 am daily. This mirroring can also be invoked manually and is recommended after changing:

- IOCDS
- Profiles
- Time
- Network or domain data

The alternate SE is not defined separately to TCP/IP or the HMC. If the alternate SE takes over from the primary SE, it uses the TCP/IP and network information of the primary SE.

Figure 35 shows an HMC and SE configuration.

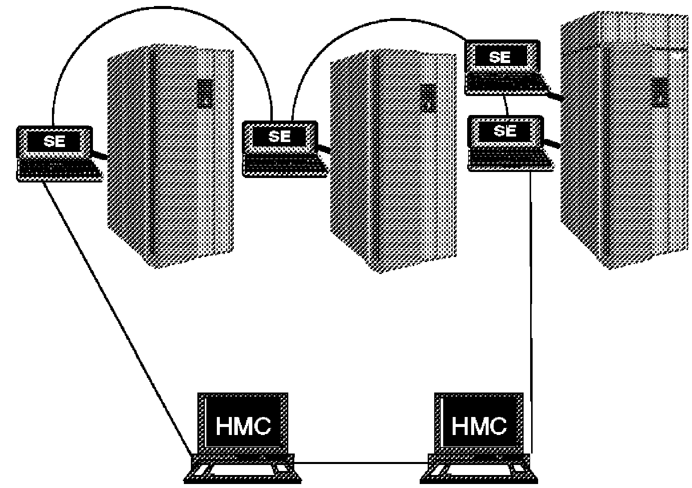


Figure 35. The HMC and the Support Elements

The HMC runs OS/2, Communications Manager/2, the distributed console access facility (DCAF), and the Hardware Management Console application, which is Licensed Internal Code (LIC).

The HMC is a composite of the IBM PC and the LIC running on it. You cannot use an alternative PC, as the LIC is only licensed on the supplied machine. You can run other OS/2 or WIN/OS2 applications on the HMC, although this is generally not recommended.

The HMC user interface is designed to provide the functions you need through an object-oriented design. Through this design, you can directly manipulate the objects that make up your IBM 9672 cluster, and be aware of any changes to any IBM 9672 status as they are detected. It is recommended to order the HMCs with the larger screens for better usability.

A good overview of the HMC and its setup is contained in the ITSO redbook *Controlling S/390 CMOS Processors Using the HMC*.

4.2.2 HMC Configuration Considerations

The HMC is ideally suited as a single point of control from a hardware point of view. As such, the HMC can serve as the console for the ESCON director by installing the ESCON director application onto the HMC. See WSC Flash W9712C for more information about this.

The Sysplex Timer Model 2 also supports the HMC as its console; this is discussed in more detail in Chapter 3, "Sysplex Timer Considerations" on page 69. Also, see WSC Flash W9706B for more information about using the HMC as the 9037 console.

Note: Currently, if you perform an EC upgrade that changes all the HMC code, then any other additional applications besides the ones mentioned will also need to be reinstalled.

An HMC does not have to map to a Parallel Sysplex. It can manage systems that are stand-alone or in multiple sysplexes, provided they are all in the same HMC management domain. Up to 16 HMCs can participate in the same management domain, which is set on the HMC Settings panel.

The configuration choices are simple if all of the CPCs in the installation are capable of being connected to the HMC and are physically close to the operations center. The following areas, which are discussed in more detail in the following sections, could have a bearing on the final design:

- System Automation for OS/390 (SA OS/390)
- Security
- Local consoles required for maintenance activities

For a detailed discussion of the possible scenarios, refer to *Controlling S/390 CMOS Processors Using the HMC*.

4.2.3 Configuring the LAN

It is recommended that you have a *dedicated* 16 Mbps Token-Ring LAN for operation of the IBM 9672 systems. The reason for this is that it is undesirable to have other LAN problems interfering with operation of the CPCs. See also 4.2.11, "Security Considerations" on page 118.

Using 8230s when connecting other hardware to the SE will improve system availability by preventing the other hardware from taking down the LAN for problems such as beaconing.

A 75-foot cable is supplied with each HMC ordered. Distances up to the limit supported by Token-Ring LANs are supported if you supply cables. If you are unfamiliar with Token-Ring LANs, or you want to know more about Token-Ring limit and cable information, refer to *IBM Token-Ring Network Introduction and Planning Guide*.

It is possible to use a Wide Area Network (WAN) between the HMC and the SE, provided certain criteria are met. See 4.2.4, "Operating the HMC Remotely" for the alternative solutions.

A gateway can be configured into the LAN (and in fact may be required), if you want to operate the HMC through NetView.

4.2.4 Operating the HMC Remotely

If you want to operate the HMC from a remote location, the recommended solution is a fully configured HMC in the remote site. This remote HMC must be kept at the same level of LIC, or higher, than the SEs it controls. The LAN bridge used must allow for either SNA LU 6.2 or TCP/IP, in support of communication between the HMC and the Service Element(s). While the response is network-dependent, it has proven to be adequate across 9600 SDLC and over ISDN links. Currently the LAN bridge used must allow for SNA LU 6.2, TCP/IP, and NetBIOS in support of communication between the HMC and the Service Elements.

A second option allows for access to the HMC through a Web browser. The HMC has a Web Server service which allows multiple users to access a subset of tasks at the one HMC at the same time. The subset excludes Single Object Operations but you will be able to perform many other functions.

The Web Server service includes security to ensure that only the appropriate people can access the HMC through the Web. This form of remote operations would be useful for systems programmers working from home and requiring access to the HMC.

For further information regarding the use of the HMC, refer to *HMC Operations Guide*.

There is a third option which can be used. The HMC uses a product called Distributed Control Access Facility (DCAF) to communicate to the SE. This product is no longer marketed by IBM. HMC is continuing to use this product and will support its use to the HMC. If you do wish to operate the HMC using DCAF, the remote user only needs to install a PC, OS/2, CM/2, and DCAF. DCAF provides support for a variety of connection types including SDLC switched connections, SNA Token-Ring connections, and TCP/IP connections. This allows full graphical, mouse-driven operation of the HMC from a remote location. The target workstation user may regain control by using a hot-key combination.

For more information on remote consoles, see *Hardware Management Console Guide*. The DCAF requirements can be found in the *DCAF V1.3 Installation and Configuration Guide*.

Note that performance of the HMC using DCAF will depend on the network performance between the HMC and the remote PC. Dropout problems have been experienced on large networks using TCP/IP. It could be very slow if the network performance is not good. Using a dial-up facility, any line speed less than 14,400 Mbps is unusable. Even 14,400 Mbps is probably only acceptable for emergency use. For regular use, using DCAF across a low to medium loaded 16 Mbps Token-Ring provides an acceptable response.

4.2.5 Automating Operation of the IBM 9672

There are at least four different methods of automating the operation of 9672 systems:

- Using the Operations Management Facility through NetView/390 in an SNA environment.
- Using an SNMP-based systems management product, such as NetView/6000, in a TCP/IP environment.
- Using customer-developed applications (written in C, C++, or REXX) running on the HMC or on any system with TCP/IP connectivity to the HMC.
- Using systems management products from other vendors.

4.2.5.1 Using the Operations Management Facility through NetView in an SNA Environment

A central site processor may be designated as the focal point or S/390 CPC from which other systems will be operated. Other networked systems managed from the focal point are often referred to as the target systems, entry points, or remote nodes. Since the CPC and networked systems are connected through channels or telecommunications lines, they could be located across the room from each other or across the country. The communication link that connects the central site to the target processor is known as the *operational path*. This high-speed line supports the day-to-day operations and processing of the enterprise's line-of-business applications. The operational path is also used for NetView-to-NetView communications when the NetView licensed program is used for software and SNA-resource network management support.

For CPCs, there is an additional communication link used for hardware network management. This link connects the central site to the support elements using a Token-Ring LAN. It is called the *initialization/recovery path.* Using this path, the CPC carries out administrative tasks for its distributed nodes. Administrative tasks include such activities as activating and deactivating systems, and reacting to system problems. The main advantage of using the initialization/recovery path to transport network management requests is that the operational path may not always be available. Unavailability of the operational path may be due simply to the fact that the system has not yet been activated. A more serious reason may be that the system had a hardware or software failure. In any event, the initialization/recovery path provides a connection that makes communication possible for system activation or problem reporting.

Clearly, with the HMC being icon-driven by the user, it would not be easy to automate without another interface.

The operations management facilities offer a unique way to carry out routine tasks on an unattended CPC. Examples of such tasks are:

- Powering on distributed CPCs in a network
- Performing power-on reset (IML) of these CPCs
- Performing IPL of systems on these CPCs
- Powering off the CPCs

It eliminates the need for local operators and optimizes the use of network operators at the central site. With the introduction of the distributed control access facility (DCAF), a network operator at the central site could do these tasks from a remote console. For more information on remote consoles, see *Hardware Management Console Guide*.

Operations management commands may be issued through NetView command lists, which automate, and often eliminate, certain operator tasks. They ensure consistency in the handling and processing of tasks, thus decreasing the chance for errors. See *S/390 Managing Your PrGcessors*, GC38-0452, for details of the commands and responses that may be issued, and other information on automation and HMC operation.

NetView provides facilities that enable automation support for SNA networks. You can exploit NetView's automation capabilities by using automation products, such as SA OS/390, that provide automation routines that can manage the IBM 9672, or you can write your own automation routines using NetView command lists (CLISTs) or REXX. Problem reports are generated and sent to NetView using SNA generic alerts.

SA OS/390 supports the coupled systems environment of the IBM 9672 with the integrated CFs (ICMF or CFCC code) as target systems.

SA OS/390 interfaces directly with the operations command facility (OCF) implemented in the support element, using the same LAN that interconnects the support element and the HMCs. The SA OS/390 pass-through function, which enables system or operator console frames of screen-oriented CPCs to be displayed on the NetView console at the focal-point system, is supported on CPCs only for operator consoles attached to an SA OS/390 PC. Function equivalent to the CPC console is provided by the HMCs use of the Distributed Console Access Facility (DCAF).

To use the CPC hardware facilities, the following supporting software is needed at your central site CPC:

- NetView R2.3 (or higher)
- VTAM Release 3.4 (or higher)

4.2.5.2 Using an SNMP-Based Systems Management Product

The HMC contains an SNMP agent that can report status changes, hardware messages, and console integration messages through SNMP traps. It also allows full remote operation through SNMP attributes. These are published in the IBM Management Information Base (MIB). This allows any SNMP-based systems management product, such as IBM's NetView/6000, to manage the CPCs attached to a HMC. This is a very important capability when 9672 systems are installed in an environment that is primarily TCP/IP-based and managed by SNMP-based systems management products. Additional information on SNMP can be found in the ITSO redbook *TCP/IP Tutorial and Technical Overview*.

4.2.5.3 Using the HMC APIs to Automate Operations

The HMC provides two sets of application program interfaces (APIs) that in turn use SNMP to communicate with the HMC SNMP agent. The first API provides a library of functions which can be called from OS/2-based C or C++ programs. The second API provides a set of functions that can be called from OS/2-based REXX programs. These programs can run on an HMC, or on any OS/2 system with TCP/IP connectivity to an HMC. These APIs provide complete monitoring and operation of all CPCs attached to the HMC. The APIs are shipped with the HMC. Sample programs written in C and REXX are also shipped with the HMC. For more information, refer to *Programming Interfaces*.

4.2.5.4 Using Non-IBM Systems Management Products to Automate Operations

Several companies have developed systems management products that are based on the HMC APIs described in the previous section. These programs provide remote and automated operation of all of the CPCs attached to the HMC. This allows 9672 systems to be managed consistently with other systems in environments where one of these system management programs is being used.

4.2.6 Can I Use Integrated Consoles Instead of OS/390 Consoles

The practical answer is no. This support was never intended to be the sole means of operating a sysplex. The 9672 supports console integration so the HMC can be an OS/390 NIP console in a no-consoles situation. However, having the HMC as the only OS/390 console in a sysplex, even though possible, is not recommended. In a sysplex, you should have *at least* 2 channel-attached MCS consoles. This allows one for the sysplex master console, and a backup for the master to switch to in the event of a failure in the sysplex master. You may operate a number of LPs in the sysplex with no channel-attached consoles, provided that the messages are routed to a channel-attached MCS console attached to another system in the sysplex.

4.2.7 How Many HMCs Should I Order

It is very easy to configure one HMC with each IBM 9672 ordered, and to end up with too many HMCs.

Each HMC can manage up to 32 CPCs. Each SE can be managed by up to 16 Hardware Management Consoles. To achieve maximum availability, *IBM recommends that you have at least two Hardware Management Consoles.* Since HMCs operate as peers, this provides full redundancy. It is also important to remember that any failure in an SE or HMC will not bring down the operating system or systems running on the CPC or CPCs.

Many installations will have two (or more) computer sites, and they operate these sites remotely. In this situation, one HMC needs to be in each computer site for maintenance service. This local HMC can also have ESCON director and Sysplex Timer console support installed for the local timer and directors. In addition, each remote operations area needs to have an HMC for awareness and management of the remote 9672s. In most cases, all 9672s will be defined on all HMCs.

Operations can use DCAF to access the timer and director console on the remote HMC for management of this equipment. Service personnel can use the local HMC console for 9032 and 9037 at the computer site when necessary. More details on this method of operation are given in the ITSO redbook *S/390 Time Management and IBM 9037 Sysplex Timer*.

Another way to increase the availability of operation functions is to put a 3174L-type control unit on the S/390 Token-Ring LAN to allow NetView/390 to interact with the support elements. You can use NetView/390 with or without SA OS/390. To use it without SA OS/390, you need to write your own automation routines using NetView/390 CLISTs or REXX.

4.2.7.1 Ordering Considerations for the HMCs

There are several items you need to consider when ordering the HMCs. Since the physical setup is rather complicated, it is worth involving both you and an IBM specialist early to ensure everything ordered is right first time. The following list outlines a few items to consider:

- How many should I order?

  See 4.2.7, "How Many HMCs Should I Order" for details.

- Cables for the LAN.

  One 75-foot cable is shipped with each HMC. If this is not what you want, you will need to order (and pay for) additional cables.

- Optical diskettes.

  You require a number of 3.5 inch 128 MB formatted optical rewritable disks. They are used to back up the LIC and customer data on the HMC/SE.

- Console printer.

  The HMC console application has no explicit printer support, but an OS/2-supported printer can be used with the "print screen" function. The HMC PC system units tht IBM has been shipping have an available parallel port and an available serial port. Over time, these system units could change; you should use the parallel port if possible.

4.2.8 Installing a New CPC on an Existing HMC LAN

An HMC can control any CPC with an SE with LIC at or below the level of the HMC LIC. Before an HMC can control a new CPC with a *higher* level SE LIC, the HMC LIC must be upgraded. Ensure that action is taken to confirm the required HMC LIC EC level and upgrade it if necessary.

Installing a New CPC

Take care if ordering a new CPC without a console. In many cases existing HMCs will need to be replaced if they are going to control a new CPC, and this requires an MES upgrade to the CPC to which the existing HMC is attached.

4.2.9 Battery Backup of the HMC

The Exide battery backup module cannot provide power for the HMCs through the rack. There are many inexpensive PC UPSes on the market that would work for HMCs.

4.2.10 Updating LIC on the HMC/SE

Licensed Internal Code (LIC) is contained on the fixed disks of the HMCs and SEs. The HMC manages all the LIC for an HMC and each support element. In a distributed network, the central site focal point can supervise the LIC change process. Changes to LIC are provided in sets sometimes referred to as maintenance change levels (MCLs). Each LIC change may have defined prerequisites that must be in place before it can be used. The prerequisites are automatically checked at the time the LIC change is prepared for system activation.

IBM controls the level of LIC for CPCs and releases changes as necessary. Such changes are all mandatory and must be applied sequentially by level, as defined by IBM. This sequencing is handled automatically when you apply the changes. The internal code changes are provided to you by the IBM Support System or optical disk and are applied to the HMC and each SE.

4.2.11 Security Considerations

Security is covered in more detail in *System Overview G6*.

To ensure total security, connect all HMCs, all CPCs, and all attaching control units to a private LAN. Using a private LAN for your configuration offers several security, availability, and performance advantages, as follows:

- Direct access to the LAN is limited to those HMCs and CPCs attached to it. Outsiders cannot connect to it.
- Traffic disruption due to temporary outages on the LAN is reduced, including minor disruptions caused by plugging in and powering on new devices on the LAN, and such catastrophic disruptions as LAN adapters being run at the wrong speed.
- LAN traffic is minimized, reducing the possibility of delays at the HMC user interface. Additionally, HMC and SE activity, if included on a non-private LAN, could potentially disrupt other LAN activity.

- If using a private LAN is not practical, isolate the LAN used by the HMCs, SEs, and control units by providing a LAN bridge between the isolated LAN and the backbone LAN to provide an intermediate level of security.
- Assign a unique domain name that includes all the CPCs controlled from one or more HMCs.
- Install one or more HMCs that have defined to them all the CPCs you want to control. Place at least one of these HMCs in the machine room near the CPCs that form its domain.
- Create user IDs and passwords for the HMC, or change the passwords of the default user IDs.
- Control DCAF access to the HMC by setting the appropriate enablement option on the HMC.
- Change the DCAF passwords on the HMC as appropriate.
- Do not publicize the SNA names or TCP/IP addresses of the HMCs or the SEs.
- Do not modify the HMC communications configuration to make it a Network Node.
- If using the Web Server, ensure that only the appropriate people are given access and that all others have none.

Normal PC security issues should also be satisfied, such as ensuring the PC is in a secure area, and that the keyboards are locked when not in use. *It is clearly possible to cause significant service disruption from the HMC console if unauthorized access is possible.*

Appendix A. Systems Management Products for Parallel Sysplex

This appendix describes systems management aspects and modifications that an installation needs to be aware of in order to configure systems management products in a Parallel Sysplex.

Recommended Sources of Further Information

The following sources provide support for the information in this appendix:

- *http://www.s390.ibm.com/rmf*, RMF home page
- *http://www.s390.ibm.com/racf*, RACF home page
- *http://www.s390.ibm.com/hcm*, HCM, HCD home page
- *http://www.s390.ibm.com/wlm*, Workload Manager home page
- *http://www.s390.ibm.com/os390/support/os390tst/*, OS/390 Integration Test home page
- *http://www.tivoli.com*, Tivoli home page
- *Automating CICS/ESA Operations with CICSPlex SM and NetView*, SG24-4424
- *Automation for S/390 Parallel Sysplex*, SG24-4549
- *Enhanced Catalog Sharing and Management*, SG24-5594
- *Getting the Most Out of a Parallel Sysplex*, SG24-2073
- *Hardware Configuration Manager User's Guide*, SC33-6469
- *Hints and Tips Checklist for Parallel Sysplex*, CHKLIST on MKTTOOLS
- *MVS/ESA SP V5 Sysplex Migration Guide*, SG24-4581
- *MVS/ESA SP V5 WLM Performance Studies*, SG24-4532
- *OS/390 Critical Dataset Placement Considerations*, WSC Flash 10017
- *OS/390 MVS An Introduction to OS/390*, GC28-1725
- *OS/390 MVS Hardware Configuration Definition: User's Guide*, SC28-1848
- *OS/390 MVS Setting Up a Sysplex*, GC28-1779
- *OS/390 MVS Parallel Sysplex Test Report*, GC28-1963
- *OS/390 MVS Planning: Global Resource Serialization*, GC28-1759
- *OS/390 MVS Planning: Workload Management*, GC28-1761
- *OS/390 MVS Sysplex Application Migration*, GC28-1863
- *OS/390 MVS System Commands*, GC28-1781
- *OS/390 R4 Implementation*, SG24-2089
- *OS/390 R5 Implementation*, SG24-5151
- *OS/390 RMF User's Guide*, GC28-1949
- *OS/390 RMF Performance Management Guide*, SC28-1951
- *OS/390 Software Management Cookbook*, SG24-4775
- *OS/390 Workload Manager Implementation and Exploitation*, SG24-5326
- *Parallel Sysplex Automation Guidelines*, SG24-5441
- *Parallel Sysplex Automation: Using System Automation for OS/390*, SG24-5442
- *Parallel Sysplex: Managing Software for Availability*, SG24-5451
- *RACF V2.2 Installation and Implementation Guide*, SG24-4580
- *RACF V2.2 Technical Presentation Guide*, GG24-2539
- *S/390 MVS Parallel Sysplex Performance*, SG24-4356
- *System Automation for OS/390 General Information*, GC28-1541
- *TME 10 OPC V2 Installation Guide*, SH19-4379
- *XCF Service Recommendations to Maximize Sysplex Availability*, WSC Flash 98038

The following roadmap guides you through the appendix, by providing a quick reference to help you find the appropriate section.

Table 7. Systems Management Products in a Parallel Sysplex Roadmap

You want to configure:	**If you are especially interested in:**	**Then refer to:**
Systems Management Products in a Parallel Sysplex		
	Having a quick view of the major IBM systems management products	A.1, "Systems Management Software in Parallel Sysplex" on page 124
	Investigating security aspects	A.2, "OS/390 Security Server (RACF) Sysplex Exploitation" on page 124
	Making your Parallel Sysplex perform well	A.3, "Performance Management Products and Parallel Sysplex" on page 128
	Operating a Parallel Sysplex	A.4, "Operations Management Products and Parallel Sysplex" on page 139
	Managing your Parallel Sysplex	A.5, "Change Control and Configuration Management" on page 148

A.1 Systems Management Software in Parallel Sysplex

The following table provides an overview of IBM products covering systems management disciplines.

Table 8. IBM Systems Management Software Supporting the Parallel Sysplex

Name	Systems Management Discipline	Remark
OS/390 Resource Measurement Facility (RMF), 5647-A01	Performance Management	OS/390 RMF to support the Parallel Sysplex.
TME 10 NetView for OS/390, 5697-B82	Systems Management	TME 10 NetView for OS/390 recommended for Parallel Sysplex.
OS/390 Security Server (RACF), 5647-A01	Security Management	OS/390 Security Server (RACF) to support the Parallel Sysplex.
Performance Reporter for OS/390, formerly SystemView Enterprise Performance Data Manager/MVS (EPDM), 5695-101	Performance Management and Capacity Planning	EPDM V1.1 enhanced to support the Parallel Sysplex.
System Automation for OS/390, 5645-005	Operations Management	An integrated set of operations management products including AOC/MVS, ESCON Manager, TSCF and the Parallel Sysplex Hardware Management Console.
OS/390 Systems Modification Program Extended (SMP/E), 5647-A01	Change Management	OS/390 SMP/E for Parallel Sysplex.
OS/390 System Display and Search Facility (SDSF), 5647-A01	Operations Management	OS/390 SDSF to support the Parallel Sysplex.
Hardware Configuration Dialogue (HCD), part of OS/390	Operations and Configurations Management	MVS/ESA SP V5.1 HCD and upwards support the Parallel Sysplex.
Data Facility System Storage Manager (DFSMS/MVS), 5695-DF1	Operations Management	DFSMS/MVS V1.4 and upwards recommended for the Parallel Sysplex.

A.2 OS/390 Security Server (RACF) Sysplex Exploitation

In OS/390, RACF forms part of the OS/390 Security Server. Throughout this section, all references to RACF equally apply to the RACF part of the OS/390 Security Server.

A.2.1 RACF Data Sharing Modes

RACF can utilize sysplex communications to allow it to operate in three different modes in a Parallel Sysplex. The mode that RACF is operating in can be controlled via the RVARY command or by an install option. These three modes are *non-data-sharing mode, sysplex data sharing mode*, and *read-only mode*. Figure 36 on page 125 shows the three modes.

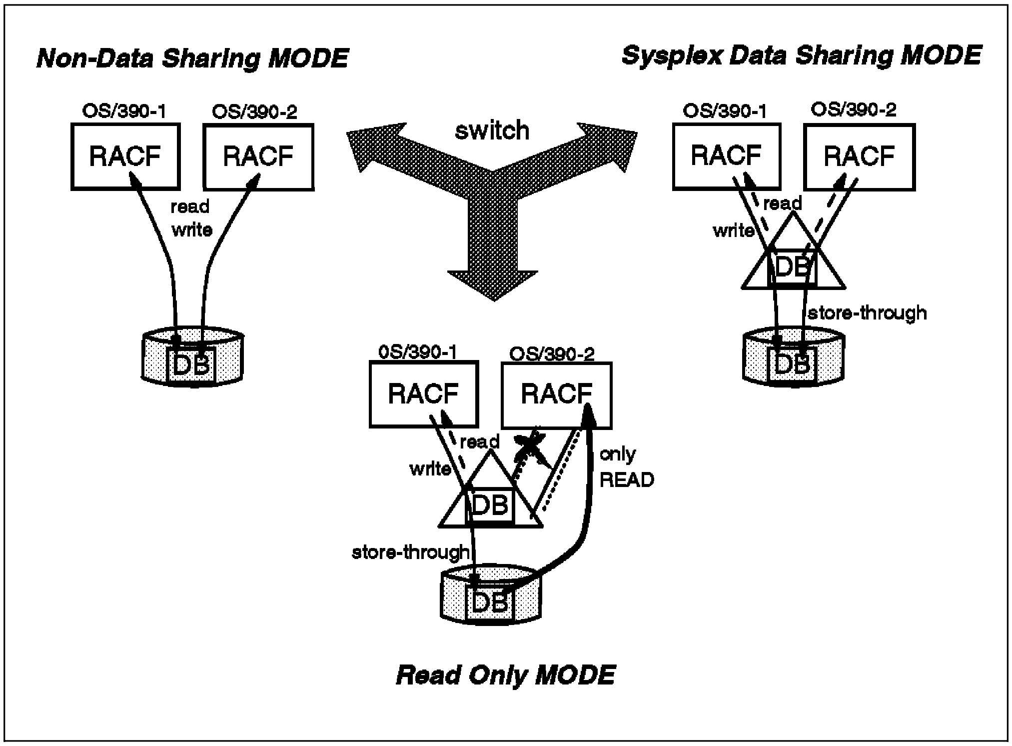


Figure 36. RACF Modes

A.2.1.1 Non-Data-Sharing Mode

Non-data-sharing mode provides RACF command propagation through XCF among the RACF members without using the CF. In terms of commands, this mode is going to offer a single image view because any RACF command can be entered in any system and propagated to all systems in the sysplex. RACF databases are stored on DASD shared among all the systems, with the reserve/release protocol being used to maintain database integrity. For the non-data-sharing option, the database must be on a device configured as shared, and each system must be:

- Running OS/390 and OS/390 Security Server
- In a sysplex and joined to a RACF XCF data sharing group
- Not in XCF-local mode

A.2.1.2 Sysplex Data Sharing Mode

Sysplex data sharing mode provides performance, systems management, and availability enhancements designed for RACF installations that run OS/390 systems in a Parallel Sysplex. Sysplex data sharing mode uses CF cache structures as a store-through cache for the RACF primary and backup databases. The structures act like a buffer for the RACF read activity. Any time you make changes to RACF, the DASD version is always updated before the structure. If you lose the RACF structure, you go to "read-only mode" for RACF, where no changes are allowed. One cache structure is created for each data set that makes up the primary and backup RACF databases. These cache structures are shared by systems in the Parallel Sysplex, and thus allow buffering of a much larger part of the RACF database than can be accomplished with single system in-storage buffers.

RACF Can Gain Performance through Use of the CF

One major reason for sysplex data sharing is to reduce the I/O rate to the RACF databases.

End users of RACF and programs using RACF interfaces are unaware that RACF sysplex data sharing is taking place within the Parallel Sysplex. The databases must be on a device configured as shared, and each system must be:

- Running with OS/390 and OS/390 Security Server
- In the same sysplex and part of the RACF XCF group
- Accessing the same CF
- Not in XCF-local mode

A RACF data sharing address space, RACFDS, is used to support CF data sharing. The address space is started automatically when RACF enters sysplex data sharing. If the RACFDS address space cannot be started, RACF will only enter read-only mode. The sysplex data sharing option also allows command propagation through XCF and the use of GRS for ENQ/DEQ serialization.

Important Note about SYSZRACF

You should check the values currently set in the GRS parmlib member, (GRSRNLxx), to verify that SYSZRACF is *not* in the SYSTEMS exclusion list.

If SYSZRACF is in the SYSTEMS exclusion list, command propagation will not work and if you enter sysplex data sharing mode, your database will be corrupted.

To enter sysplex data sharing mode requires RACF to be enabled for sysplex communication. Although sysplex data sharing mode can be entered with a RACF RVARY command, enablement for sysplex communication requires reassembly of the RACF dataset name table (ICHRDSNT) followed by an IPL of all systems.

Once in RACF sysplex data sharing mode, the database cannot be shared with other systems that do not participate in the data sharing group. Systems not participating in the RACF data sharing group can, of course, access another RACF database. For information about sizing the RACF structures in the CF, refer to 2.5, “RACF Structures” on page 91 in Volume 2: Cookbook, SG24-5638.

For information about setting up RACF data sharing, as well as information about using the Parallel Sysplex Configuration Assistant to set up the RACF CF structures, refer to the redbook *Parallel Sysplex Resource Sharing*, SG24-5666.

For information about having several RACF systems in the same Parallel Sysplex, refer to 2.3.3, “What Different 'Plexes Are There?” on page 41 in Volume 1: Overview, SG24-5637.

Consider the following when defining multiple RACFplexes in a Parallel Sysplex:

- Because the name of the RACF CF structure is fixed, there can only one RACFplex per Parallel Sysplex enabled for sysplex communications and data sharing. The sysplex-enabled RACFplex:
  - Can operate in sysplex data sharing or non-data-sharing mode

  - Cannot include any systems outside the sysplex
  - Must have all systems in the RACFplex enabled for sysplex communication

- One or more RACFplexes that are not enabled for sysplex communication may exist within a sysplex, although this configuration is not recommended. These RACFplexes:
  - Cannot be enabled for sysplex communications
  - Do not support RACF command propagation
  - Cannot operate in data sharing mode
  - Can coexist with each other and with a sysplex-enabled RACFplex
  - Can include systems inside and outside of the sysplex
  - Use reserve/release for serialization
  - Must not share the same RACF database as the sysplex-enabled RACFplex, otherwise, the RACF database will be damaged

Single RACFplex Recommended

Although it is possible to have multiple RACFplexes (not enabled for sysplex communication) in a sysplex, the recommended configuration is one RACFplex, enabled for sysplex communication, and in sysplex data sharing mode.

A.2.1.3 Read-Only Mode

The system enters read-only mode, which is an *emergency* mode, when use of the CF is specified, but an error has occurred, making the CF either inaccessible to, or unusable by, RACF. This mode is also entered if the RACFDS address space will not start or fails.

In this mode, no updates to the databases are allowed (except statistics updates during logon and job initiation, and other statistics updates made with ICHEINTY ALTER request). TSO logons are permitted.

Immediately after a *structure failure* has been detected, RACF starts the rebuilding of the RACF structure in another available CF. This function can be done automatically, or it may be installation-controlled.

After the successful rebuild, the mode is converted from read-only to sysplex data sharing mode again. Following a CF *connection failure*, RACF switches to read-only mode and stays there until someone initiates a rebuild for the structure. In RACF read-only mode, you can either:

- Issue an RVARY command, which:
  - Disconnects from the structure.
  - Creates a new one based on the preference list and connects to that structure.
- Issue an XCF rebuild command. When the rebuild completes successfully an RVARY command will be issued internally by RACF, causing RACF to go back into full sysplex data sharing mode.

RACF supports the XCF rebuild option of LOC=OTHER and REBUILDPERCENT for connectivity loss.

Note: The RVARY DATASHARE command should be password protected. This ensures that the operator is prompted for a password. The password defaults to

YES. This may complicate operating procedures, especially in recovery scenarios, where a read-only environment needs to be established using a RVARY NODATASHARE command.

A.2.2 RACF Remote Sharing Facility

RACF remote sharing facility (RRSF) is a RACF facility that allows you to:

- Manage remote RACF databases. For example, a security administrator can maintain the remote RACF database of the disaster recovery center, without the need to log on to that remote system. Each RACF command can be qualified by the keyword AT (also called command direction).
- Keep RACF databases synchronized. This is helpful for recovery and workload balancing purposes. RACF TSO commands that update the RACF database are automatically directed to execute on a remote node, after being executed on the local node.
- Keep passwords synchronized. For example, a user who owns more than one user ID on the same system, or different user IDs on different systems, can keep passwords synchronized between all his user IDs.
- Enhancements to RRSF, available with the OS/390 Security Server for OS/390 R3, provide the capability for application updates to be automatically directed, in addition to command updates and password changes.

  When activated, automatic direction of application updates will propagate RACROUTE REQUEST=DEFINE and RACROUTE REQUEST=EXTRACT,TYPE=REPLACE requests. If an application invokes ICHEINTY, then ADD, ALTER, DELETE, DELETEA, and RENAME requests will be propagated.
- Issue RACF TSO commands from the OS/390 operator console when the subsystem RACF address space is active.

The main consideration for RRSF in a Parallel Sysplex is that all RACF database updates that are propagated through RRSF to one member of the Parallel Sysplex are available for all other members of the Parallel Sysplex, regardless of which data sharing mode the sysplex is running.

This implies that if the members in a sysplex share the RACF database, you should configure the RRSF network in such a way that database updates are only propagated to a single member of the sysplex. Refer to the ITSO redbook *RACF V2.2 Installation and Implementation Guide*, SG24-4580, to investigate how this can be achieved in your environment.

A.3 Performance Management Products and Parallel Sysplex

Performance management is the activity in an installation that monitors and allocates data processing resources to applications according to a *service level agreement* (SLA) or equivalent.

- There are three ways to solve performance problems:
  - Buy more resources.
  - Create the illusion you bought more resources. This is known as *tuning*. The capability to tune successfully implies you have been previously wasting resources.

  - Steal it from a less important application. Again there is a cost. Here the cost is reduced service to the application from which the resources were stolen.
- There is usually a conflict between *throughput* and *response* objectives. Throughput one aims to heavily use the system (high utilization), while response aims to serve better the key transactions (high availability of resources).
- Performance management is an heuristic task, implying an ongoing set of measurements and related modifications (based on the measured values).

The most important performance element in Parallel Sysplex is the Workload Manager (WLM) MVS/ESA SP V5, or OS/390 component. The next topic discusses WLM together with products for doing performance measurements in Parallel Sysplex.

A.3.1 Workload Manager (WLM) in the Parallel Sysplex Environment

Important Notice - End of Support for Compatibility Mode Announced

The announcement letter for OS/390 V2R10 contained the following Statement of Direction:
"The operating system release scheduled for availability in the second half of 2001 will be the last release to support WLM compatibility mode. Goal mode will be the only supported mode starting with the operating system release scheduled for the first half of 2002."

To assist customers that have not yet moved to goal mode, IBM is making a tool available to help with the migration. The tool is known as the Goal Mode Migration Aid, and is available from the WLM Home Page.

In addition, OS/390 V2R10 contains a number of changes that have been requested by customers to make the migration easier. Both of these enhancements are described in more detail below.

WLM is a component of the basic control program (BCP), introduced in MVS/ESA SP V5.1. WLM is active whether systems are running in *goal mode* or *compatibility mode*. We highly recommend that you run systems in a Parallel Sysplex in goal mode (and, from the Statement of Direction above, goal mode will soon be the *only* supported WLM mode). There are a number of reasons for this:

- It is easier to define performance objectives in goal mode, as the objectives are defined in similar terms to those used in your Service Level Agreements (response times, throughput, and so forth). The WLM ISPF dialog is used to define goal mode objectives, while compatibility mode continues to use the older ICS/IPS SYS1.PARMLIB members.
- Many new OS/390 functions, such as WLM-managed initiators, are only available on systems in goal mode.
- Once a system is in goal mode, it is easier to add a new workload with specific performance objectives.

WLM has the following objectives:

- To provide more effective use of the available resources
- To improve productivity by:
  - Reduced complexity
  - Providing hands-off systems management
  - Providing effective service management
  - Providing effective performance management
- More effective use of Parallel Sysplex

You tell WLM about its *goals* through a service policy. In this policy, the workload is associated with *service classes* that are WLM constructs describing goals. WLM attempts to maintain the defined goals by dynamically adjusting the dispatching priority of the task and/or allocating more processor storage.

WLM also provides other functions to manage work and report on it. These functions include the following:

- For both JES2 and JES3 environments, WLM can optionally control the number of initiators available for specified job classes. WLM will adjust the number of started initiators in response to the Performance Index of the jobs waiting to execute, and the relative priority of that work compared to other work running on the system. (The Performance Index is an indicator of whether a service class is meeting, exceeding, or missing it goal.) These initiators are known as WLM-managed initiators. They were introduced for JES2 in OS/390 V2R4, and subsequently added to JES3 in OS/390 V2R8.

- WLM provides a resource affinity scheduling environment. The scheduling environment is defined to WLM and is made up of a number of defined resources. The resources are defined to WLM and consist of a user-defined name representing an OS/390 resource such as devices, database managers or systems, and a state of on, off, or reset. In order for a scheduling environment to be available, the resources defined to this environment must be in a required state. JES2 and JES3 check whether the scheduling environment is available before starting work on a particular system.

- Application Environments were introduced to enable a performance administrator to control the number of address spaces in a multi-address space server subsystem such as Webserver, DSOM, SQL stored procedures and MQ Series.

- To enable more granular management of transactions, preemptible SRBs were introduced. This allows WLM to manage work at a level other than the address space level. There are two types:

  1. Client SRBs which are units of work which are run in one address space on behalf of another address space. Priority and account, such as CPU time, comes from the client address space and is accounted back to the client address space.

  2. Enclave SRBs are units of work where the priority and account comes from an enclave, rather than from the associated address space.

  These are used by DDF in DB2 V4.1 and above. They can of course be used by other applications such as Component Broker. The ITSO redbook *OS/390 Workload Manager Implementation and Exploitation*, SG24-5326, contains further detailed information on preemptible SRBs.

- Another feature, introduced in OS/390 V2R8, was the ability to manage device aliases for IBM 2105 devices. 2105's provide the ability to effectively have more than one UCB per device (known as "aliases"), allowing more than one concurrent I/O to a device. Systems in WLM goal mode have the ability to move an alias from one device to another if such a change will help service classes meet their objectives or reduce overall IOS queueing.

- OS/390 V2R9 introduced the ability to manage enclaves across multiple systems in a sysplex. With this support, enclaves can now run in address spaces on multiple systems within a Parallel Sysplex cluster. Work begun on one system, for instance, can be split into parallel tasks that can then be exported to other systems for parallel processing. Multisystem enclaves provide the capability to manage and report on work requests that are executed in a parallel environment on multiple OS/390 images as single entities. Enclaves can be managed across the Parallel Sysplex cluster with this new function. WLM is able to change the control of the system as the units of work changes so you can ensure the work receives appropriate resources no matter where it is running. Consistent goal-oriented management is the result. This function requires CFLEVEL=9 coupling facility control code (CFCC) support.

Note: WLM operates no matter what type of sysplex you run, including Parallel Sysplex and monoplex. All the systems in a sysplex running in goal mode use the same WLM policy. This is a significant departure from compatibility mode where each system can have its own (potentially different) IEAICS and IEAIPS members.

A.3.1.1 Migration from Compatibility Mode to Goal Mode

It is not a difficult task to migrate from compatibility mode to goal mode. It should take several weeks, if normal planning has taken place.

A number of enhancements were introduced by OS/390 V2R10 that remove features that some customers found impeded their ability to migrate to goal mode. Briefly, these enhancements are:

- Enhanced CPU management to provide protection for critical work across major workload shifts. Some customers have very stringent response time goals for critical subsystems, and want to ensure that those subsystems *always* have higher dispatching priorities than other less important work—this new support adds the ability to be able to achieve this.
- It is now possible to use transaction response time goals for just a subset of CICS or IMS regions, should you so wish. This provides the ability, for example, to manage production regions to production response time goals, while letting test regions run with velocity goals.
- Storage Isolation for critical regions over long, idle periods. This ensures that critical subsystems that have very quick ramp-up of workloads will have the storage they need to handle that workload, even if they been idle for an extended amount of time.
- Classification by system group, system name, or "subsystem collection" name (for example, JES2 MAS name) support for heterogeneous Parallel Sysplex clusters. This gives you the ability to treat similarly named workloads differently, depending on where the work is running.

More information about these enhancements is available on the WLM Home Page.

There is a document on the WLM Web site, entitled “The Washington Systems Center's WLM Migration Guide and Checklist,”

that provides a checklist of things to consider before starting the move to goal mode. This checklist was created by the Washington Systems Center based on their experience with actual customer migrations. It is *highly* recommended to review this list before embarking on a migration to goal mode. The document can be downloaded in PDF format from the WLM Web site at:
`http://www.s390.ibm.com/wlm/documents/documents.html`

To assist in migration from WLM compatibility mode to WLM goal mode, there are two sample policies available. There is one created by the Washington System Center, and another created by Cheryl Watson. Both can be reached by links from the WLM Home Page. These quick-start policies allow you to move to WLM goal mode using a generic type of policy. Whichever policy you choose will need to be fine-tuned to your environment.

One point to note, is that you may need to check your accounting packages or routines to ensure that they reflect any measurement changes that the WLM goal mode introduces. Also, packages that use performance groups or reporting performance groups will need to be changed to use the new SMF fields that contain the service class or reporting service class instead.

The WLM Web site is available at:
`http://www.ibm.com/s390/wlm/`

For more information on implementing and exploiting WLM, refer to the ITSO redbook *OS/390 Workload Manager Implementation and Exploitation*, SG24-5326.

A.3.1.2 Parallel Sysplex Considerations for WLM Goal Mode

The following comments apply when configuring Parallel Sysplex for running in goal mode:

- It is not mandatory to operate your systems in Parallel Sysplex in goal mode. It is, however, strongly recommended.
- You can run your Parallel Sysplex systems with some of them in goal mode and some of them in compatibility mode. The ones in goal mode form a WLMplex. In the WLMplex, the goals described in the service policy have a WLMplex scope; that is, they refer to the transactions running in those systems. The ones in compatibility mode are excluded from the WLMplex and have a single system scope.

  This is illustrated in Figure 8 on page 47 in Volume 1: Overview, SG24-5637.
- It is strongly recommended to switch off the generation of SMF type 99 records to avoid a high CP and I/O cost. SMF type 99 contains very detailed information on what actions SRM took to address service class goals. This information is not required in normal situations and can be compared to other type of traces. Even when SMF Type 99 records are switched off in the SMFPRMxx member, 15 minutes worth of the records are buffered in storage and are included in the dump if you dump the WLM address space.
- The WLM couple data set is very lightly used and does not need any specific recommendation for placement. However, you should not place the WLM couple data sets on any devices that will experience reserves. See A.4,

"Operations Management Products and Parallel Sysplex" on page 139 for a discussion on couple data sets and reserves.

- CICSPlex SM may use the WLM service policy to obtain the goal of the arriving transaction. Based on the goal, the transaction can be routed to the most suitable AORs in the Parallel Sysplex.
- CICS, IMS/DC, and DB2 DDF use the WLM interface to pass the external properties of arriving transactions, to allow WLM to derive the service class. DB2 further provides the following WLM support:
  - DB2 servers on OS/390 provide sysplex routing services for DRDA requesters. When a DRDA requester (single user or gateway) connects to a DB2 server in a data sharing group, the server returns a list of available members of the group together with weighting information supplied by WLM to enable the requester to select the best member of the group to which to direct the request. The DRDA requester must be enabled to support this feature.
  - DB2 stored procedures, when called, have their dispatching priority tied to the priority of the thread, which has been established in the WLM service class. WLM dynamically manages the quantity of address spaces serving the DB2 stored procedures workloads.

For more information about performance of WLM when running in goal mode, refer to the ITSO redbook *S/390 MVS/ESA SP V5 WLM Performance Studies*, SG24-4532.

A.3.2 OS/390 Resource Measurement Facility (RMF) in Parallel Sysplex

This topic assumes that the reader is familiar with RMF concepts and terminology. For more basic information about RMF, review the manual *OS/390 RMF User's Guide*.

Some objectives for RMF are to support:

- The Workload Manager in goal mode
- The concept of a single system image for performance management in a Parallel Sysplex
- The CF by providing reports that cover CFs, CF structures, and CF link usage
- Datasharing subsystems by identifying contention and locking problems

RMF consists of a number of components:

- Two online monitors that provide short-term and snapshot information.
- A postprocessor for producing reports in batch.
- A sysplex data server.
- Two PC components. One is used for creating long-term trend reports, the other is for online performance monitoring of one or more systems.

These components are described in more detail in the following sections.

A.3.2.1 RMF Online Monitors

RMF provides two online monitors. The first, Monitor II (previously called with the RMFMON command), provides snapshot information, using an ISPF front end.

RMF Monitor II reports can now contain data from any system in the sysplex. One important Monitor II report is the IRLM Long Lock Detection report. This report will assist you in identifying lock contention situations. An example of this report can be found in B.1.7, "IRLM Long Lock Detection Report" on page 211 in Volume 2: Cookbook, SG24-5638.

The other online monitor is Monitor III (previously called with the RMFWDM command). Monitor III samples data over a variable amount of time and stores data in storage and optionally in a set of VSAM data sets—this provides the ability to interactively investigate performance problems after the fact. Monitor III is also responsible for creating some of the SMF records that are subsequently used by the RMF Postprocessor.

Using Monitor III, it is possible to create reports either at the system or sysplex level. The reporter can potentially run on any or all of the systems in the sysplex. This makes it possible to monitor the performance of all systems in the sysplex without having to be logged on to multiple systems, although if there are multiple OS/390 releases in the sysplex, you should log on to the one that is running the highest release. The Sysplex Data Index report shows the time range covered in the Monitor III data gatherer buffers for every RMF instance. For more information about the RMF Monitor III reports related to Parallel Sysplex, refer to Appendix B, "RMF Reporting in Parallel Sysplex" on page 189 in Volume 2: Cookbook, SG24-5638.

A.3.2.2 RMF Postprocessor

The RMF Postprocessor produces local system reports and sysplex-wide reports. The data comes from SMF records that are created by Monitor I, Monitor II, and Monitor III. The type of reports that are produced include:

- Activity of CF structures
- Availability of CF logical processors
- Reasons for delays on requests to the CF

Appendix B, "RMF Reporting in Parallel Sysplex" on page 189 in Volume 2: Cookbook, SG24-5638 contains examples of the RMF Postprocessor sysplex reports, including a CF Activity Report.

A.3.2.3 RMF Sysplex Data Server

The RMF Sysplex Data Server is a distributed RMF function. It is started as an RMF subtask on each system of the sysplex. Each copy of the data server communicates with all other copies in the sysplex using XCF services. RMF uses this sysplex communication method to provide access to RMF measurement data from any member of the sysplex.

The RMF Sysplex Data Server provides performance data, collected locally by the distinct RMF data gatherers, to any RMF reporter instance in the sysplex. (Performance data includes VSAM records produced by the Monitor III data gatherer, and potentially any type of SMF records, including ones not produced by RMF.)

The RMF Sysplex Data Server becomes active when RMF is started. Its main task is to manage requests from other systems for SMF records or Monitor II or

III data. In addition to this, it also manages RMF's SMF data buffer. This is a wrap around buffer residing in a data space whose size may vary from 1 MB to 2 GB. (The default is 32 MB.) It can hold any type of SMF record as long as that record type is not suppressed in the SMFPRMxx parmlib member. By default, it holds the records written by Monitor I (70 to 78).

A.3.2.4 RMF Distributed Data Server

The RMF Distributed Data Server (DDS) is used by the PM of OS/390 component (described in A.3.2.5, "RMF PC Components"). The DDS runs as a started task and uses TCP to distribute RMF information to clients. Only one DDS must be run per sysplex, and a client may communicate with more than one DDS to gather performance information from many sysplexes. The DDS should run on whichever system in the sysplex is running the highest release of OS/390.

RMF DDS is also used by the Tivoli Distributed Monitoring Agent to retrieve "traditional" performance data such as cache, CPU, storage, and other information.

A.3.2.5 RMF PC Components

RMF provides two PC components. The first of these provides support for viewing and manipulating RMF data in a spreadsheet. The RMF Spreadsheet Converter imports data from RMF online and Postprocessor reports into a spreadsheet. The RMF Spreadsheet Reporter provides a set of sample macros to manipulate and chart the data that has been imported from RMF. Both of these tools are designed for trend reporting.

The Performance Monitoring of OS/390 (PM of OS/390) component provides a Java interface to allow online monitoring of one or more OS/390 sysplexes on a PC workstation. You can design your own reports, and can have reports that monitor multiple systems in the same report. PM of OS/390 runs on Windows 95, Windows NT, and OS/2 Warp 4 (requires APAR OW44855 for the Java 1.1.8 OS/2 Client). PM of OS/390 uses the RMF Distributed Data Server to collect performance information from RMF on OS/390.

A.3.2.6 RMF and Automation

A facility of RMF that can be used either with automation, or for online alerting of key personnel, is RMF Client/Server Enabling (RMFCS). RMFCS uses the client/server concept to support performance management for OS/390 systems without an active TSO/TCAS subsystem on the host.

You can access Monitor II and Monitor III reports with RMFCS by exploiting the ISPF Batch GUI feature. This way, RMFCS combines the advantages of a single point of control for OS/390 performance management with a state-of-the-art user front end. Although the RMF manual only mentions OS/2 clients, RMFCS in fact works with both OS/2 and Windows clients.

RMFCS supports event-driven monitoring. That is, predefined events on the MVS hosts can be configured to initiate performance monitoring. These events may be either specific system messages, or selected performance data counters that exceed predefined Monitor III exception thresholds.

As well as a GUI front end, RMFCS can also be used to send messages when specified thresholds are exceeded. These messages can be intercepted by your automation package, and acted upon, before the problem escalates and starts impacting service.

Information about setting up RMFCS, and examples of how it can be used with automation is described in the chapter "RMF Client/Server Enabling" in *OS/390 RMF User's Guide*, SC28-1949.

A.3.2.7 RMF Synchronization in Parallel Sysplex

To create sysplex reports, RMF requires that all gatherers are running in sync on each of the systems. The best choice is to use the default (SYNC(SMF)) that synchronizes the interval RMF writes records with SMF. This makes it possible to compare RMF records with SMF records from other sources. Furthermore, it is recommended that the same interval length be used on each system. This makes reporting more consistent.

Requirement to Synchronize RMF in Parallel Sysplex

To produce Parallel Sysplex reports, all the RMF Monitor I instances in Parallel Sysplex *must* have the same interval and synchronization values.

Some sample reports are provided in Appendix B, "RMF Reporting in Parallel Sysplex" on page 189 in Volume 2: Cookbook, SG24-5638. Some of the fields in the RMF reports related to the Parallel Sysplex are described in B.1.1, "Parallel Sysplex Reports" on page 189 in Volume 2: Cookbook, SG24-5638.

A.3.2.8 CF Report

The data for this report is gathered individually by each instance of the Monitor III data gatherer in the Parallel Sysplex and stored in one SMF record (74.4) per CF per interval. A Postprocessor running in one image accesses this data through Parallel Sysplex data server services. The report is produced for each CF attached to the Parallel Sysplex and has three sections:

- CF Usage Summary
- CF Structure Activity
- CF Subchannel Activity

This set of reports assist the installation in managing the CFs. They are very helpful for:

- Optimizing each structure placement across multiples CFs
- Tuning the structure allocation policy in the couple data set
- Doing capacity planning globally and by workload type

Refer to the B.1.1, "Parallel Sysplex Reports" on page 189 in Volume 2: Cookbook, SG24-5638 to get the description of the fields in these reports.

There is also an item in the RMF Newsletter about the use of RMF in a Parallel Sysplex. The newsletter can be viewed on the Web at:
`http://www.s390.ibm.com/products/rmf/rmfhtmls/rmfnws11.htm`

A.3.2.9 Support for Tivoli Distributed Monitoring Agent

A recent announcement adds support for an OS/390 Monitoring Agent for Tivoli Distributed Monitoring. The monitoring agent runs under OS/390 UNIX System Services and collects performance data about USS. It also collects data from RMF about the following items:

- OS/390 sysplex resources
- OS/390 sysplex workloads
- OS/390 system resources

- OS/390 system workloads

The Distributed Monitoring Agent for OS/390 is included in the base code for Tivoli Distributed Monitoring (5698-EMN) to maximize the availability of computing resources in a distributed environment.

More information is available on the Web at:
http://www.ibmlink.ibm.com/usalets&parms=H_200-199

A.3.3 Performance Reporter for OS/390 and Parallel Sysplex

Performance Reporter for OS/390, formerly known as Enterprise Performance Data Manager/MVS (EPDM), collects detailed systems management data recorded by many parts of the Parallel Sysplex, stores it in a standard DB2 database, and presents it as either tabular or graphical reports.

There are many questions that need to be answered when doing performance analysis in Parallel Sysplex. Examples include:

- What are the expected data volumes for SMF, IMS logs, and other data sources?
- What data should be recorded or avoided?
- How do you track IMS and CICS response times and performance?
- What are the typical run times for data reduction?
- How do you manage all the log data sets produced?

Performance Reporter for OS/390 can provide meaningful information from the processed logs in Parallel Sysplex.

A.3.3.1 Processing the Log Data

The volumes of log data produced in Parallel Sysplex depend on the number of images and the load and the type of applications executing. In a Parallel Sysplex environment, SMF still records individually on each system. This means that the number of SMF log data sets increases in proportion to the number of images. Managing many data sets is time consuming and error-prone. Therefore automated procedures are beneficial in such circumstances. It is important to be aware of the expected run times for postprocessing log data in batch.

Performance Reporter for OS/390 has tailoring functions that provide the user with a flexible tool for achieving the trade-off between the cost of processing large amounts of data, and the value of detailed information. The more granular the information, the longer the run times.

A.3.3.2 Reports from the Data Processed

Performance Reporter for OS/390 can derive, from the vast amounts of collected data, management information for tracking the performance of:

- Application response times
- System throughput, measured as volumes of work (transactions, jobs) per unit of time
- Availability of the subsystem or application used by the end user

The list of supplied reports with Performance Reporter for OS/390 is long, and there are a number of reports that are relevant both to a sysplex and a Parallel Sysplex environment. The following list is a sample from the supplied reports:

- OS/390 Workload by Workload Type, Monthly Trend This report shows consumed CP time (TCB and SRB) per workload type, in millions of service units. The report also shows the total CPC capacity, which is the number of service units that would be consumed if CPC usage is 100%.
- OS/390 Sysplex Overview, Monthly Trend

  This report gives you an overview of the activity for all systems in a sysplex.
- OS/390 Workload by Resource Group, Daily/Monthly Overview

  These reports show unweighted TCB and SRB service units used for resource groups and the minimum/maximum capacity specified for the resource groups.
- OS/390 Availability for a Sysplex, Daily/Monthly Trend
- OS/390 Response Time (Goal in %), Daily/Monthly Trend.
- OS/390 Average Coupling Facility Busy Profile, Hourly Trend.
- OS/390 Execution Velocity (Goal in %), Daily/Monthly Trend

A.3.4 System Measurement Facility (SMF)

The function of SMF is not in itself affected by Parallel Sysplex configuration. Note however:

- RMF uses SMF record type 74, subtype 2 for XCF performance-related data, and subtype 4 for CF performance-related data.
- RMF has a distributed server function for SMF records. This function is called the RMF sysplex data server. Through the RMF sysplex data server, any application can obtain SMF records produced in other systems in a Parallel Sysplex. Refer to A.3.2.3, "RMF Sysplex Data Server" on page 134 to get more information about the RMF sysplex data server.
- Record type 88 is produced in response to ENF signal 37, which indicates that the SMF global recording interval has ended. Each record reports system logger activity for one log stream or structure. Record values from multiple systems can be summed to give the sysplex view of system logger activity. There is a sample report program, IXGRPT1, provided in SYS1.SAMPLIB.
- The generation of type 99 SMF records should be avoided during the production shift when operating in goal mode. It consumes CP resources and requires significant DASD space in the SMF data sets. It can be switched on in specific cases where IBM requires information about WLM decisions.
- Automatic restart management (ARM) introduces some changes in SMF records, as follows:

  The *type 30* (common address space work) and *type 90* (system status) SMF records have been enhanced to provide information relating to ARM activity, elements, and policy management.
- DB2 can create SMF type 100, 101, and 102 records. These record types can be post-processed by the DB2 PM product.

The details can be found by referencing *OS/390 MVS System Management Facilities (SMF)*.

A.4 Operations Management Products and Parallel Sysplex

Operations management activities that are affected by the introduction of Parallel Sysplex are data and storage backup and recovery, data management, automation, consoles, and scheduling.

A.4.1 Data and Storage Backup and Recovery

This area does not differ very much from the considerations in a multi-image environment.

Couple Data Sets and Reserve/Release

Couple data sets, particularly the sysplex couple data sets but also other types of couple data sets as well, *must not* be placed on a volume that is subject to any reserve/release activity. Many sysplex functions rely upon unimpeded access to the couple data sets (CDSs), and the sysplex (or a subset of systems or subsystems operating in the sysplex) may be adversely impacted by such reserve/release activity. Be careful when making backups of data sets that are placed on volumes having CDSs or other vital sysplex data sets.

A.4.1.1 DFSMSdss Considerations for Parallel Sysplex

When DFSMSdss is doing a full volume backup, it issues a reserve against the volume. If the volume contains data sets such as a couple data set, the volume backup activity could cause a status update missing condition, as the couple data set could not be accessed during the backup activity.

Planning to switch from a primary to an alternate couple data set in time for scheduled backups could be difficult to manage and may lead to disruption within the sysplex.

A possible solution to this issue is to change the reserve activities for the volumes concerned into global ENQs. This can be done by ensuring the following statement is included in the active GRSRNLxx member of parmlib, for each appropriate volume:

```
RNLDEF RNL(CON) TYPE(SPECIFIC)
QNAME(SYSVTOC) RNAME(cdsvol)   /*  CONVERT VTOC of CDS VOLUME  */
```

Before implementing the solution, installations should check that converting reserve activity to global ENQ does not impose performance problems.

Review *OS/390 MVS Planning: Global Resource Serialization*, before implementing reserve conversion for other possible implications in your environment.

Another solution would be to use Concurrent Copy or SnapShot, which results in a reserve being held on the volume for a much shorter time. DFSMS V1R1 and appropriate hardware support in the DASD controller are required for these functions.

A.4.1.2 DFSMShsm Considerations

When DFSMShsm starts any of its functions, it issues an ENQ to guarantee that no other HSMs in the complex is doing the same task at the same time. The format of the resource name is common to all HSM systems. As a result, if there are two separate HSMplexes in one sysplex, there will be false contention if they both try to run the same function at the same time.

To eliminate this inhibitor, DFSMS V1R5 introduced a new function known as Single GRSplex Serialization. This new function adds a new keyword to the HSM startup. When you specify this new keyword, RNAMEDSN, the format of the resource name will change to include the name of the related HSM CDS. Assuming that the CDS name for the different HSMplexes are different, this now gives you the ability to run the same function at the same time in each HSMplex, without incurring this false contention.

Another new function introduced in DFSMS V1R5 is called Secondary Host Promotion. In a HSMplex, there are certain functions that are only performed by the Primary host. If the Primary host is not available, those functions will not be carried out. Secondary Host Promotion gives you the ability to define one or more Secondary HSM hosts. HSM joins an XCF group, and uses XCF functions to get notified if the Primary HSM has failed. If this happens, one of the Secondary hosts will take over the role of Primary host, ensuring that the Primary host-unique functions continue to be performed.

Both of these functions, and other enhancements, are discussed in a new sysplex chapter in *DFSMShsm Implementation and Customization Guide*.

A.4.2 Data Management

This section describes some important aspects and products related to data management in Parallel Sysplex.

A.4.2.1 DFSMS/MVS in Parallel Sysplex

Parallel Sysplex support is provided by DFSMS/MVS V1.2 and upwards through the VSAM local shared resource (LSR) cross-invalidate function and through the support for more than eight systems in a SMSplex environment. DFSMS/SMS V1.3 furthers the support for Parallel Sysplex data sharing by providing VSAM record-level sharing (RLS) as exploited by CICS TS for OS/390 R1 and additional system name support (which is discussed later in this section).

VSAM LSR Cross-Invalidate: VSAM local shared resources cross-invalidate (VSAM LSR XI) is an enhancement to VSAM LSR. This enhancement uses the OS/390 XES services, in conjunction with the CF, to achieve local buffer coherency in a DL/1 VSAM data sharing environment.

DL/1 also continues to provide the locking required in the data sharing environment. This support includes obtaining an exclusive lock on a VSAM control interval before changing the contents of the control interval.

Note: DL/1 still provides the local buffer coherency function for OSAM by using the Integrated Resource Lock Manager (IRLM) to broadcast buffer invalidation requests.

VSAM LSR XI works for both central storage and hiperspace buffers.

For further information regarding VSAM LSR XI use of CF structures, refer to 2.7, "IMS Structures" on page 98 in Volume 2: Cookbook, SG24-5638.

CICS/VSAM RLS: CICS/VSAM RLS is discussed in detail in 4.3.3, "CICS/VSAM Record Level Sharing Considerations" on page 209 in Volume 1: Overview, SG24-5637. Information regarding the CF structures used to implement CICS/VSAM RLS is provided in 2.0, "A Structured View of CF Structures" on page 59 in Volume 2: Cookbook, SG24-5638.

A.4.2.2 SMSplex

An SMSplex is a system (an MVS image) or a collection of systems that shares a common SMS configuration. The systems in an SMSplex share a common ACDS and COMMDS pair.

It is recommended that the SMSplex *matches* the Parallel Sysplex for better manageability of your data. This is illustrated in Figure 12 on page 55 in Volume 1: Overview, SG24-5637.

DFSMS/MVS V1.2 introduced the concept of SMS system group names that allow the specification of a system group as a member of a SMSplex. A *system group* consists of all systems that are part of the same sysplex and are running SMS with the same configuration, minus any systems in the sysplex that are explicitly defined in the SMS configuration.

The SMS system group name must match the sysplex name defined in the COUPLExx parmlib member, and the individual system names must match system names in IEASYSxx parmlib member. When a system group name is defined in the SMS configuration, all systems in the named sysplex are represented by the same name in the SMSplex. The SMSplex does not have to mirror a Parallel Sysplex; you can choose to configure individual systems and Parallel Sysplexes into an SMSplex configuration.

Note: Note that JES3 does not support SMS system group names.

DFSMS/MVS V1.3 and Group Name Support

DFSMS/MVS V1.3 introduces support for up to 32 system names or group names, or a combination of both. This support now enables you to support more than eight systems in a JES3 SMS complex, where system group names cannot be used. For more information, refer to *DFSMS/MVS V1.3 General Information.*

DFSMShsm Considerations: DFSMShsm support for migration, backup, and dump processing supports system group names.

The enhanced DFSMShsm processes storage groups for automatic space management, data availability management, and automatic dump processing in the following ways:

- A storage group may be processed by any system if the system name in the storage group is blank.
- A storage group may be processed by a subset of systems if the system name in the storage group is a system group name.
- A storage group may be processed by a specific system if the system name in the storage group is a system name.

DFSMShsm can optionally use CICS/VSAM RLS as a means to share its control data sets in DFSMS/MVS 1.4.

SMSplex and Coexistence: DFSMS/MVS V1.2 may coexist with previous DFSMS/MVS or MVS/DFP releases that support SMS, with the following restrictions:

- System group names are not used. If system names only are defined in the configuration, DFSMS/MVS V1.2 is SMS-compatible with previous releases.
- All pre-DFSMS/MVS V1.2 systems are explicitly defined in the configuration, and only DFSMS/MVS V1.2 systems are represented by system group names. The system group names are treated as system names by pre-DFSMS/MVS V1.2 systems.
- For DFSMS/MVS V1.3, you can run SMS in either 32-name mode or in compatibility mode. When your system runs in 32-name mode, the SMS configuration can only be shared with other DFSMS/MVS 1.3 systems running in 32-name mode. When your system runs in compatibility mode, only eight system names, system group names, or a combination of system and system group names are supported in the SMS configuration. In compatibility mode, the configuration can be shared with systems running down level releases of DFSMS/MVS or DFP, and with systems also running in compatibility mode.

A.4.3 System Automation for OS/390

System Automation for OS/390 is an integrated package of operations management program products and services. It includes a workstation-based graphical user interface that provides a single point of control for monitoring and controlling hardware and software resources in a Parallel Sysplex. The user interface is implemented through the cooperative interaction of the products that make up System Automation for OS/390. This level of integration permits transitions across product boundaries, as well as exception-based design that enables monitoring the status of all important systems resources through color-coded icons. Figure 37 on page 143 provides sample screens to illustrate the System Automation for OS/390 operations interface. System Automation for OS/390 integrates three systems management products:

- AOC/MVS
- TSCF
- ESCON Manager

System Automation for OS/390 provides I/O exception monitoring, additional point-and-click sysplex command support, and integrated automation features for CICS, IMS, and OPC. For more information regarding System Automation for OS/390, refer to *System Automation for OS/390 General Information*.

A.4.3.1 Considerations for System Automation in Parallel Sysplex

This section describes aspects of system automation that may need to change in a Parallel Sysplex environment. The focus of the section is primarily on NetView-based automation.

In a Parallel Sysplex, you may have many system images and applications. System messages are issued from many systems, and may be visible on many systems. The recommendation is to have a single *managing system* that acts as a single point of control (SPOC). This system may be either inside or outside the Parallel Sysplex.

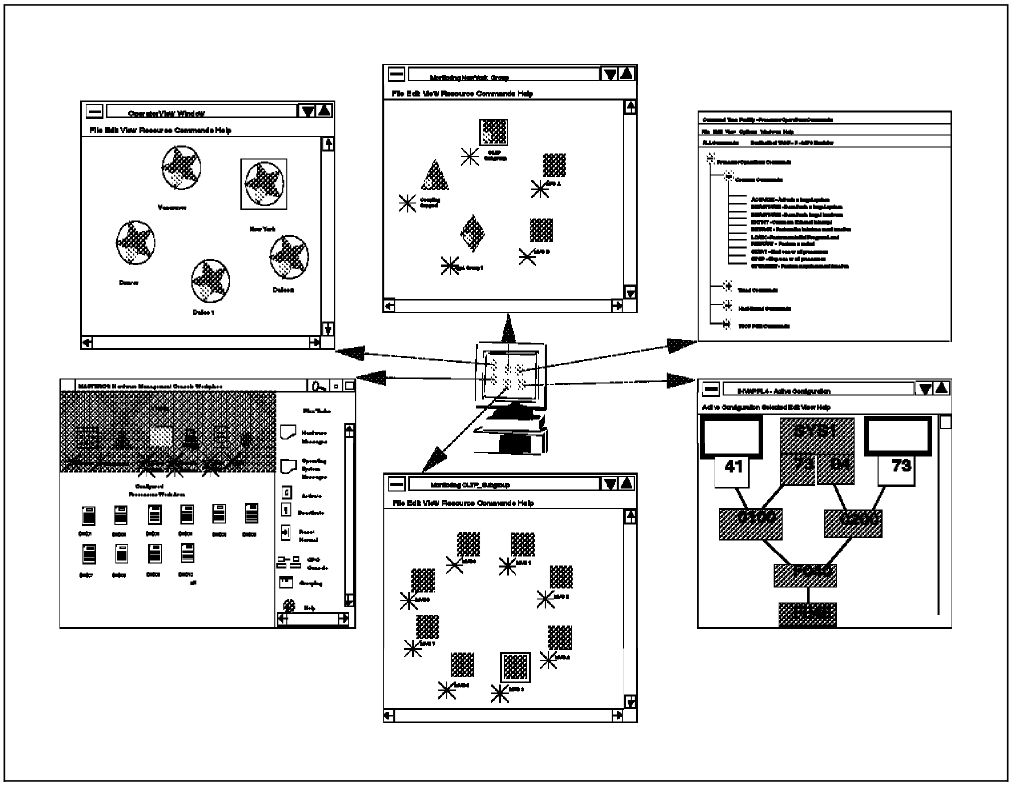


Figure 37. System Automation for OS/390: Operations Interface

System automation in a Parallel Sysplex needs to handle single-system resources as well as Parallel Sysplex resources. Parallel Sysplex resources include:

- CFs
- Sysplex Timers
- CDSs
- HMCs

Automation in Parallel Sysplex is very similar to automation in a multisystem environment. System Automation for OS/390 extends NetView to provide automated operations facilities that improve system control and enhance operator productivity.

In multisystem environments today, the recommendation is to have one of the systems act as a *focal* point system to provide a single point of control where the operator is notified of exception conditions only. This is still the case in Parallel Sysplex.

Another recommendation is to automate as close to the source as possible. This means having both NetView and System Automation for OS/390 installed on all systems in the Parallel Sysplex to enable automation to take place on *the system where the condition or message occurs.*

System Automation for OS/390 provides support for the Automatic Restart Management (ARM) function. Both System Automation for OS/390 and ARM provide facilities to restart applications. System Automation for OS/390 provides for coordination of these facilities, allowing you to develop a restart strategy that takes advantage of both. System Automation for OS/390 coordinates with system images to:

- Determine which facility is responsible for restarting a specific application.
- Avoid possible conflicts and duplicate application recovery attempts.
- Allow you to take full advantage of the ARM recovery capabilities for applications running on sysplexes. System Automation for OS/390 will continue to automate an application after ARM has moved it to a secondary system, provided System Automation for OS/390 is installed and customized on that system. If it is not installed on the secondary system, System Automation for OS/390 is still aware that the application is active within the sysplex and will not attempt to restart it on its primary system, if that system is restarted within the sysplex.

Additional enhancements include:

- Improved status coordination between focal point and target systems. The status of monitored resources can be accurately reflected on the System Automation for OS/390 graphical interface.
- Fast query capability is available, providing a quick means of accessing automation control file information.
- Additional icons for tracking outstanding WTORs on a system or sysplex.
- Policy dialogs now have support for the use of numeric system names.

Refer to *System Automation for OS/390 General Information*, for more information.

In Figure 38, the association between OS/390, NetView, and System Automation for OS/390 is shown.

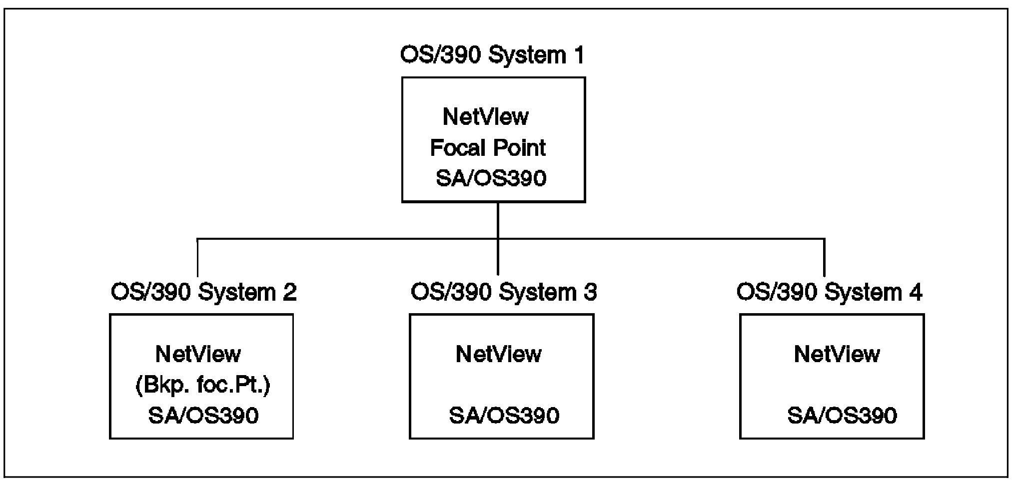


Figure 38. NetView and System Automation for OS/390 Association in a Sysplex

To allow for contingency, thus supporting the goal of continuous availability, a backup focal point NetView should also be planned for your configuration. This will allow the current focal point system to be taken out of the Parallel Sysplex for planned outages. In addition, unplanned outages on the focal point will not render the Parallel Sysplex inoperable.

Additional support for management of Parallel Sysplex resources was added to System Automation for OS/390 via APAR OW39485. Additional information about the support added by this APAR is available on the Web at:
`http://www.s390.ibm.com/products/sa/sainfos.html#SysplexSPE`

Additional support for Tivoli OPS/ESA V2.3 was added via APAR OW40864. One of the new functions added by this APAR is that jobs whose WLM service class has been changed by OPC/ESA get highlighted on SDF. More information about this APAR is available on the Web at:
`http://www.s390.ibm.com/products/sa/nopc23.html`

Refer to the ITSO redbooks *Automation for S/390 Parallel Sysplex*, SG24-4549, and *Automation Using Tivoli NetView OS/390 V1R3 and System Automation OS/390 V1R3*, SG24-5515, for more information.

Automation Is Key

To ensure the success of a Parallel Sysplex migration project, it is vital that you include the automation staff as part of the project team from day one. This will ensure that required modifications to existing automation infrastructure and routines are clearly understood and accounted for early in the project.

A successful Parallel Sysplex automation project is easier to obtain if you use:

- Single point of control functions wherever possible.
- Symmetrically configured systems, with the ability to run any applications on any system. If an application can execute on multiple systems, make it either identical or parameterize the system-unique parts according to simple and consistent naming standards in order to ease cloning and provide the ability to easily move the application to a backup system.
- Carefully established standards for applications, systems and other resources to facilitate policy sharing and cloning.
- System Automation for OS/390 sysplex functions and integration capabilities.
- A flexible, carefully designed LAN configuration to implement your controlling workstations as well as some of the managed resources.

A.4.4 Tivoli Business Systems Manager

Tivoli has announced a product called Tivoli Business Systems Manager (TBSM), which will integrate the features and functions of Tivoli Global Enterprise Manager (GEM) and Tivoli Manager for OS/390 (TM/390) into a single product. At the time of writing, availability of TBSM is planned for 3Q 2000.

The initial release of Tivoli Business Systems Manager will have the following characteristics:

- A Microsoft Windows technology-based graphical user interface (GUI) that allows customers to use a drag-and-drop feature to define customized line of business views. Business impact and hyperbolic resource views are also supported.

- An integrated CIM-compliant data model supporting managed resources, line-of-business tree views, and event propagation and history.
- Full report capability of both current and historical data. This data will be available for reports on availability, trend analysis, capacity planning, and more.
- Full host and distributed command and control.
- An enhanced set of application instrumentation and business system enablers, which allows customers to create, customize, and maintain their own business systems. This includes the ability to create a line-of-business dynamically and nest business systems within other business systems.
- Support for multiple forms of instrumentation, including the Tivoli Application Management Specification instrumentation for Tivoli Ready products, the Tivoli Manager product family, and customer-developed applications.
- Correlation of alarms; fast root cause analysis; and integration with problem management solutions.
- Enhanced ability to document mission-critical applications.

The two constituent products, GEM and TM/390, are described below.

A.4.4.1 Tivoli Global Enterprise Manager

Tivoli GEM combines and builds on the capabilities of IBM SystemView and the Tivoli Management Environment (TME) to achieve real cross-platform management integration. It allows management function to be performed in either environment (host or distributed), or a combination of both.

Tivoli GEM Management Integration Services support the key operations of your host and distributed environments. Specifically, Management Integration Services functions are provided for the following:

- Events and automation
- Commands
- Topology
- Security management
- Problem management
- Storage management
- Performance
- Job scheduling

Additionally, Tivoli GEM features the Applications Policy Manager, a special application view server with Java clients that gives a business-system level view of any networked enterprise.

A.4.4.2 Tivoli Manager for OS/390

Tivoli Manager for OS/390 (TM390) is Tivoli's operational management tool for OS/390. With TM390, customers can actively monitor and control systems management events across a multi-datacenter, multi-platform environment. Tivoli Manager for OS/390 accommodates a number of existing IBM, Tivoli, and third-party systems management tools to help automate, schedule, and monitor specific resources. System resources are stored in an open, structured query language (SQL)-compliant database and are logically grouped and graphically displayed in line- of-business views.

A.4.5 Job Scheduling in Parallel Sysplex

This topic describes a typical configuration for Tivoli OPC in a Parallel Sysplex. As with OS/390 automation in Parallel Sysplex, the Tivoli OPC setup is not very different from today's multi-system environment. Tivoli OPC is designed to have a single *active* controller in a Parallel Sysplex and a *tracker* on each image where Tivoli OPC is to schedule work to be run. For continuous operations, at least one standby controller should also be defined in your Tivoli OPC configuration.

Note: Tivoli OPC is capable of having a *tracker* agent on OS/390 (including OS/390 UNIX), OS/400, OS/2, Windows NT, DEC Open/VMS, AIX/6000, HP-UX, SunOS, Sun Solaris, Digital UNIX, DEC Open/MVS, Pyramid MIPS ABI workloads. Support is also provided for SAP/R3 jobs through the exploitation of Tivoli Workload Scheduler Extended Agent technology.

A.4.5.1 Tivoli OPC Exploitation of Parallel Sysplex

Tivoli OPC has enhanced the use of Parallel Sysplex facilities compared to OPC/ESA, which used cross coupling facility (XCF) services for communication between the controller and the trackers.

Multiple copies of the controller can be started on each member of the XCF group, but only one becomes the active controller. If the active controller fails, then another one, running in another OS/390 image of the sysplex, automatically switches from Standby to Active status and takes over.

Switching controllers is simplified by cloning using a shared parmlib. Cloning applies to both Controller and Tracker initialization statements as well as started task JCL. Variable substitution, using system symbols, is used in the installation dialog and job streams (EQQJOBS) for unique identification.

The concept of a Single System Image has been introduced, which removes the restriction for OPC TSO users (ISPF dialog) and PIF users (batch and user-written applications) to reside on the same MVS image of the OPC controller. This is based on APPC end user local/remote transparency. In a Parallel Sysplex, the VTAM function, Model Application Definitions, is used. A specific OPC server address space, running on the same system as the OPC controller address space, is responsible for receiving remote requests and forwarding them to a local SSI. The redbook *Batch Processing in a Parallel Sysplex*, SG24-5329, contains a sample configuration and all the related definition statements.

Tivoli OPC interfaces with the Workload Manager (WLM) function of OS/390 to move a specific critical batch job into a predefined WLM service class. This service class has higher performance goals than the normal service class a job runs in. When OPC decides that a job needs greater performance to meet its deadline, OPC calls a documented WLM interface to move the job into the predefined service class. This new function is documented in the chapter entitled “How Tivoli OPC Selects Work to be Assisted by WLM” in *Tivoli Operations Planning and Control V2R3 Planning and Scheduling the Workload*, and in the DOC APAR PQ37471.

There is also a recent OPC redbook that describes the recent enhancements to OPC: *End to End Scheduling With OPC and TWS*, SG24-6013. You can also get information from the OPC Web site:
`http://www.tivoli.com/products/index/opc/`

Note: Tivoli OPC supports workload restart and workload reroute. Refer to *TIVOLI OPC Installation Guide* for a description of these functions.

A.5 Change Control and Configuration Management

This section looks at some issues that arise when configuring and changing your installation in a Parallel Sysplex environment. Some of the topics covered are:

- SMP/E configuration guidelines
- Hardware Configuration Definition (HCD) considerations
- Sharing the SYSRES volume
- Sharing the master catalog volume
- Sysplex couple data sets
- Naming conventions and the use of cloning to replicate systems in the Parallel Sysplex

Many of the topics discussed in this section are covered in much greater detail in the redbook *Parallel Sysplex - Managing Software for Availability*, SG24-5451.

Available IBM Services

Note that in your country, there is likely to be a range of Parallel Sysplex implementation and migration services. Check with your country service providers to find out what is available. It is recommended to include a service provider that has prior experience with Parallel Sysplex migrations as part of any Parallel Sysplex implementation project.

A.5.1 SMP/E Configuration Guidelines

In this topic, we take a look at current recommendations for SMP/E usage in a Parallel Sysplex environment. For maximum flexibility, we recommend that there should be an SMP/E target zone for each set of target data sets. This may or may not increase the number of SMP/E environments you have to maintain. MVS/ESA SP V5 provided the ability to share both your sysres volume and your Master Catalog among a number of systems. As a result, is unlikely that you will have (or need) a SYSRES for every system. Using a combination of shared SYSRESes and one SMP/E target zone per SYSRES gives you the ability to apply updates (either new products or maintenance) to each of your target environments independently of each other.

With Parallel Sysplex, you also have the opportunity to update each *image* in the Parallel Sysplex independently. This will require an additional set of target libraries, which you can introduce into the Parallel Sysplex one system at a time. Consider the Parallel Sysplex environment in Figure 39 on page 149.

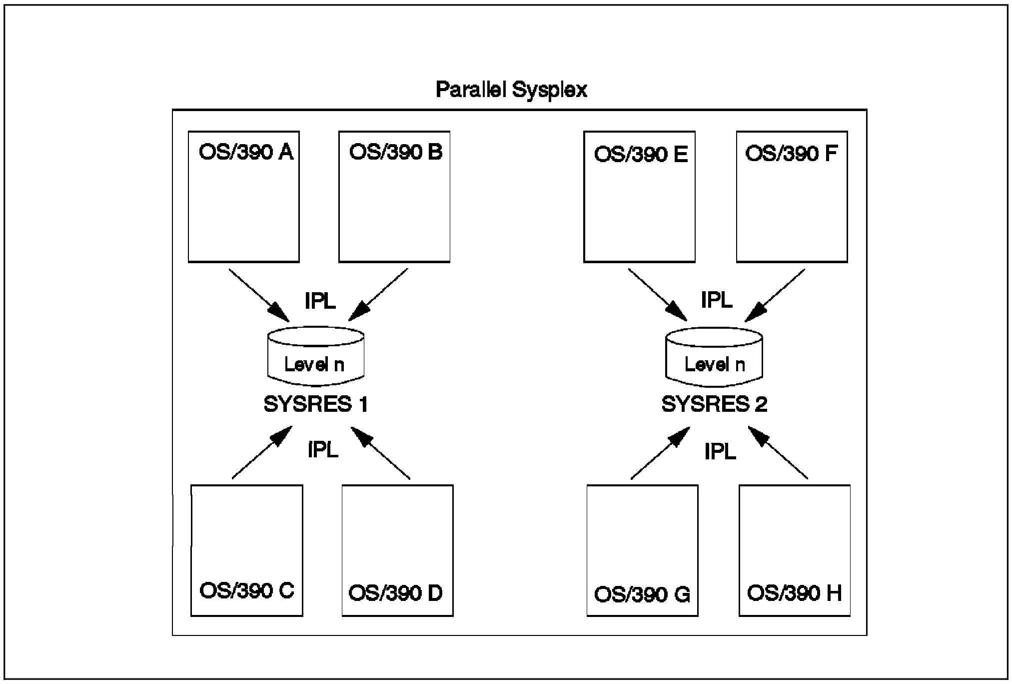


Figure 39. Example of a Shared SYSRES Parallel Sysplex Environment

In this Parallel Sysplex, there are eight OS/390 images, with two SYSRES volumes. To introduce a new level of the system, we must introduce a third SYSRES (see Figure 40 on page 150).

In Figure 40, we see that one of the system images (A) is now IPLed from the new SYSRES M, where this new SYSRES is the new level of system to be introduced into the Parallel Sysplex. After this new level has been "proven" in Parallel Sysplex, then system B, system C, and system D can have IPLs scheduled to bring them up to the n+1 level. Once this is done, SYSRES 1 can then be cloned to be at n+1 also. Using this new SYSRES 1, you can then IPL systems E, F, G, and H, one by one, onto this new n+1 level.

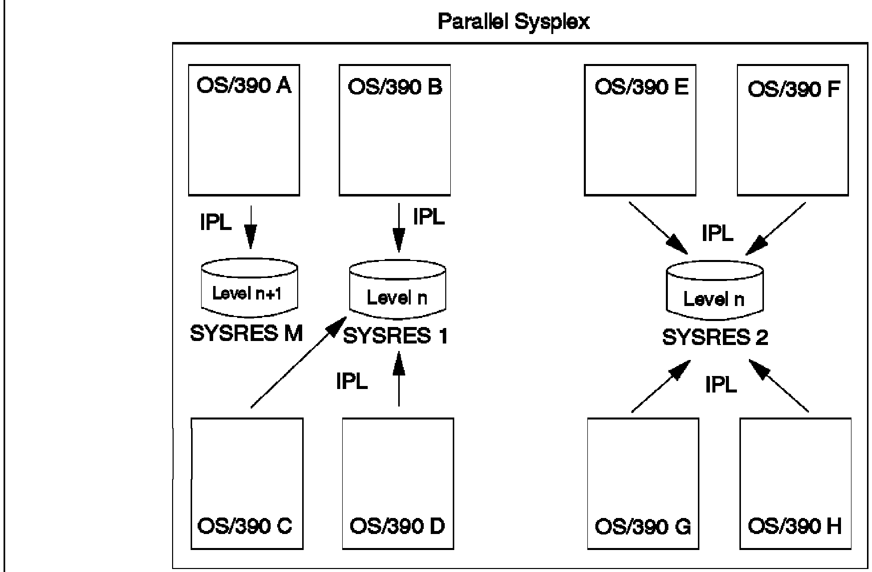


Figure 40. Introducing a New Software Level into Parallel Sysplex

Note: Each SYSRES volume should have a backup. Once all of the systems have been brought up to the n+1 level, the old backup SYSRES data sets should then be updated to the n+1 level also. Finally, the old SYSRES 2 can then be returned to the maintenance system to become the "new" SYSRES M.

OS/390 N, N+3 Coexistence

In a multisystem sysplex configuration, four consecutive releases of OS/390 can coexist, providing compatibility and flexibility in migration. Effective from OS/390 V2R9, MVS/ESA V5 can no longer coexist with that, or later, releases of OS/390. More information about the coexistence policy is available in the White Paper entitled *Planning Guide for Multisystem Customers: OS/390 Coexistence and Planning Considerations Through the Year 2000*, available on the Web at:
`http://www.s390.ibm.com/stories/year2000/coexist.html`
and in the WSC FLASH 10034, available on the Web at:
`http://www.ibm.com/support/techdocs/atsmastr.nsf/PubAllNum/Flash10034`

At least two consecutive releases of key subsystems running on OS/390, such as DB2, CICS, and IMS, can coexist within a multisystem sysplex. Depending on the installed release, more than two consecutive releases may be supported in some cases. See 2.3.1.2, "System Software Level Considerations" on page 30 in Volume 1: Overview, SG24-5637 for more information about the coexistence support for these products.

A.5.1.1 SMP/E Data Set Placement

If you have one target zone per sysres, the most sensible place for the target CSI is on the SYSRES volume itself.

The SMP/E global CSI data set should be on a separate volume with, for example, the SMPPTS data set. For a discussion about SMP/E and software management in a Parallel Sysplex, refer to the redbook *Parallel Sysplex - Managing Software for Availability*, SG24-5451.

A.5.2 Hardware Configuration Definition (HCD) Considerations

This section is not intended to give detailed information on installing or tailoring HCD. The aim is to focus on HCD in a Parallel Sysplex environment. With OS/390, HCD is required to build an I/O definition file (IODF).

A.5.2.1 Parallel Sysplex Considerations for HCD

The key message for the use of HCD in a Parallel Sysplex environment is: keep your configuration *symmetrical*. There are two main reasons for this:

- Operational simplicity

  Symmetry provides operational simplicity. If a given device has the same address on every system, it is much easier to automate the handling of that device, and it also will result in far fewer operational mistakes resulting from the wrong device address being specified. It will also speed up problem determination and resolution, as there will be no confusion about which physical device a message refers to.

- Ease of cloning

  Symmetry provides ease of cloning. If all devices are symmetrical, then cloning a system image may not require *any* modification to the IODF. You may want to change some image-specific information, such as the eligible device table (EDT).

Note: In the Parallel Sysplex environment, *every* CPC and CF must be defined in the *same* IODF. If this is not done, then it will not be possible to define the CF links correctly between the CPC and the CF.

A very useful enhancement introduced in OS/390 V2R5 is the ability to create the contents for your CONFIGxx member. Parallel Sysplex configurations are often large and complex, and as a result, many installations do not create a CONFIGxx member. Using this enhancement, the CONFIGxx member can be very easily created, and can then be used to check that the active configuration actually matches the configuration in HCD. This new facility is described in *OS/390 Hardware Configuration Definition User's Guide*.

A.5.3 Hardware Configuration Manager (HCM)

The Hardware Configuration Manager (HCM) is a PC-based client/server interface to HCD. HCM provides a graphical front end to HCD as well as adding functionality. It allows you to define the logical and physical configuration information in one common place.

For example, HCM allows you to define the following physical (or labelling) information as you create the configuration:

- Serial numbers of all the hardware
- Physical connection information

- Patchports and panels
- Crossbar switches
- Cables and cable numbers

The physical and logical information match because they are produced from the same source.

HCM provide three major advantages over using HCD alone:

- The configuration is created graphically using objects.

  You can see you configuration diagram as it is created.

- You can label your configuration as it is created

- HCM produces accurate configuration diagrams

  Your diagrams are always kept synchronized with your configuration. This is because your configuration and diagrams are created at the same time by the same process.

OS/390 only requires the information currently contained in tie IODF. HCM builds it own configuration file which contains the extra information such as labelling. HCM synchronizes its file with an IODF to ensure tat each time a configuration is opened, the HCM PC file matches the IODF obtained from HCD.

A.5.3.1 Simplified Definitions

HCM simplifies the definition process, for example, for CTCs. The CTC is done in a few easy steps and the only information you provide is:

- Source CPC and CHPID number
- LPs for the source CPC if in an LPAR environment
- Target CPC and CHPID number
- LPs for the target CPC if in an LPAR environment
- Device address

  See the section entitled “ESCON Channel-to-Channel (CTC) Considerations” in *OS/390 MVS Parallel Sysplex Configuration Volume 3: Connectivity*, SG24-2077 for a discussion of an ESCON CTC addressing scheme for XCF signalling paths.

New functions in OS/390 V2R9 HCM provide visual displays of the I/O operations status for the active OS/390 I/O configuration. HCM now cooperates with the I/O Operations component of System Automation for OS/390 (formerly ESCON Manager) to display color-coded status information and active switch paths directly on the existing HCM configuration diagram. OS/390 operators also can retrieve detailed node element information and issue I/O Operations connectivity commands while working at the HCM workstation.

For more information, refer to *OS/390 HCM User's Guide* and to the HCM Home Page:
`http://www.ibm.com/s390/hcm`

A.5.4 Naming Conventions

As systems work together in a Parallel Sysplex to process work, multiple copies of a software product in the Parallel Sysplex need to appear as a single system image. This single system image is important for managing systems, the JES2 MAS configuration or JES3 complex, transaction managers, database managers, and ACF/VTAM. However, the notion of sysplex scope is important to any software product that you replicate to run on systems in the Parallel Sysplex. To manage these separate parts as a single image, it is necessary to establish a naming convention for systems and subsystems that run in the Parallel Sysplex. The ITSO redbook *Getting the Most Out of a Parallel Sysplex*, SG24-2073, provides an excellent discussion of this topic.

A.5.4.1 Developing Naming Conventions

You need to develop a flexible, consistent set of names for systems and subsystems in the sysplex.

- Develop consistent names for CICS, IMS TM, IMS DB, IRLM, DB2, and VTAM for use in data sharing. For detailed information and recommendations on application subsystem naming conventions, refer to *OS/390 Parallel Sysplex Application Migration*. There is also a discussion about naming conventions for a sysplex, especially the relationship between the names of different components, in the redbook *Parallel Sysplex - Managing Software for Availability*, SG24-5451.
- Specify the same names for the following on each system:
  - System name
  - SMF system identifier (SID)
  - JES2 member name
- Keep system names short (for example, three to four characters). Short system names are easy for operators to remember and use, and they reduce the chance of operator error.

  Consider the following examples of system names for a sysplex:

  ```
  System 1 : S01 or SY01
  System 2 : S02 or SY02
  System 3 : S03 or SY03
     ...      ...       ....
  System 32: S32 or SY32
  ```
- Develop consistent and usable naming conventions for the following system data sets that systems in the sysplex *cannot* share:
  - LOGREC data sets
  - STGINDEX data sets
  - Page/Swap data sets
  - SMF data sets

OS/390 allows these non-shareable data sets to be defined with system symbols (variable) so that OS/390 can substitute the system name for each image. This allows you to potentially use a single IEASYSxx member for every system in the sysplex.

For example, specifying the LOGREC data set name as SYS1.&SYSNAME..LOGREC in IEASYSxx means that every system can use the same IEASYS member, but still have a unique LOGREC data set (assuming the data sets are all called SYS1.SY01.LOGREC, SYS1.SY02.LOGREC, and so on).

Examples of working naming conventions can be found in the *OS/390 Parallel Sysplex Test Report*. A further example is contained in the *Hints and Tips Checklist for Parallel Sysplex*, which is available from your IBM representative through the CHKLIST package on MKTTOOLS.

A.5.4.2 Sharing System Definitions in a Multisystem Environment

System symbols provide two or more systems with the ability to share a single set of system definitions in a multisystem environment. The use of system symbols is key to sharing definitions in a multisystem environment. All systems can use the same commands, dynamic allocations, parmlib members, and job control language (JCL) for started tasks while retaining unique values where required.

A system symbol acts like a variable in a program; it can take on different values, based on the input to the program. Picture a situation where several systems require the same definition, but one value within that definition must be unique. You can have all the systems share the definition, and use a system symbol as a "place holder" for the unique value. When each system processes the shared definition, it replaces the system symbol with the unique value it has defined to the system symbol. If all systems in a multisystem environment can share definitions, you can view the environment as a single system image with one point of control.

Sharing resource definitions has the following benefits:

- It helps *maintain* meaningful and consistent naming conventions for system resources.
- It provides a single place to *change* installation definitions for all systems in a multisystem environment. For example, you can specify a single SYS1.PARMLIB data set that all systems share.
- It *reduces* the number of installation definitions by allowing systems to share definitions that require unique values.
- It allows you to *ensure* that systems specify unique values for commands or jobs that can flow through several systems.

When the system symbols are defined in a member that is shared by two or more systems, each system substitutes its own unique defined values for those system symbols. There are two types of system symbols:

- Static system symbols have substitution texts that remain fixed for the life of an IPL.
- Dynamic system symbols have substitution texts that can change during an IPL.

System symbols can be used in:

- Dynamic allocations
- JES2 initialization statements and commands
- JES3 commands
- JCL for started tasks and TSO/E logon procedures
- Most parmlib members
- Most system commands

If your installation wants to substitute text for system symbols in other interfaces, such as application or vendor programs, it can call a service to perform symbolic substitution.

OS/390 provides the following static system symbols:

&SYSNAME The system name

&SYSPLEX The sysplex name

&SYSCLONE The last 2 characters of the substitution text for the &SYSNAME symbol

&SYSR1 The IPL volume serial number

Other symbols Up to 99 additional system symbols that your installation can define

You can also define the &SYSPLEX system symbol in the LOADxx member, so that it is available right at the beginning of system initialization. The early processing of &SYSPLEX allows you to use its defined substitution text in other parmlib members.

Symbolic Parmlib Processor: The Symbolic Parmlib Processor lets you interactively determine what symbol substitutions will take place in parmlib before use, thus potentially saving system IPLs. This also allows you to pre-process symbolics defined in a concatenated parmlib. Error detection and correction can now be done in advance of the IPL from an ISPF session.

See *OS/390 MVS Initialization and Tuning Reference* for information about how to set up support for system symbols. For information about how to use system symbols, see the books listed in Table 9.

Table 9. References Containing Information on the Use of System Symbols

Use System Symbol Substitution Service in:	**Reference**
Application programs	*OS/390 MVS Programming: Assembler Services Guide*
The DYNALLOC macro for dynamic allocations	*OS/390 MVS Programming: Authorized Assembler Services Guide*
JCL for started tasks	*OS/390 MVS JCL Reference*
JES2 commands	*OS/390 MVS JES2 Commands*
JES2 initialization statements	*OS/390 MVS JES2 Initialization and Tuning Reference*
JES3 commands in	*OS/390 MVS JES3 Commands*
Parmlib members	*OS/390 MVS Initialization and Tuning Reference*
VTAM definitions in SYS1.VTAMLST	*VTAM for MVS/ESA V4.4 Release Guide*
System messages and commands	*OS/390 MVS System Commands*
TSO/E REXX and CLIST variables	*OS/390 MVS TSO/E V2 User's Guide* and *OS/390 MVS TSO/E V2 CLISTs*
TSO/E logon procedures	*OS/390 MVS TSO/E V2 Customization*
General usage information	*Parallel Sysplex - Managing Software for Availability*

A.5.5 Shared Master Catalog

The master catalog is a critical data set, that you are likely to share across the systems in the sysplex. OS/390 provides functions that make sharing the master catalog easy, either across all or some of the systems in the sysplex.

OS/390 allows an installation that has multiple images to share a master catalog and an IPL volume among multiple images. OS/390 provides a number of symbols which assist in sharing a master catalog. To aid in the sharing, the previously fixed named, non-shareable system data sets SYS1.LOGREC and SYS1.STGINDEX can now have an installation-specified data set name. In addition, a system symbol, &SYSNAME, was introduced and may be used as part of the data set name specifications for some parameters in Parmlib. When you use &SYSNAME, the data set name specification becomes flexible, and you do not need a separate parameter specification for each system in the sysplex.

A.5.5.1 Extended Alias Support

Extended Alias support is a new function introduced in DFSMS V1R5. It was introduced to address a restriction that arose relating to the use of aliases with shared master catalogs. An alias can be used to reference different levels of product libraries. For example, if we define an alias

```
DEFINE ALIAS (NAME(SYS1.PRODUCT) RELATE(SYS1.V1R3M0.PRODUCT))
```

new versions of the product can be tested by changing the alias to point to another set of libraries. Some installations used this capability to have the same JCL on all systems, but have different libraries used, depending on which system the job ran on.

In a sysplex environment we may have systems at different levels of a product. However, with a shared master catalog, we can only have one alias for all the systems that share that catalog.

To address this situation, a new function called the Symbolic Alias Facility was introduced. Using this new support, we can now define the alias as follows:

```
 DEFINE ALIAS (NAME(SYS1.PRODUCT) -
            SYMBOLICRELATE('SYS1.&PRODVR..PRODUCT'))
```

The symbol &PRODVR. is resolved to the value that has been defined on each individual OS/390 system. On system SY01, for example, this may resolve to V1R3M0, while on system SY02, it resolves to V1R4M0. There is only one alias definition in the master catalog, and a single shared catalog, but you still have the ability to have an alias refer to different data sets, depending the which system the job runs on. For further information, refer to *DFSMS MVS Managing Catalogs*.

A.5.5.2 Enhanced Catalog Sharing

Enhanced Catalog Sharing is another new function introduced by DFSMS V1R5. This function moves some of the catalog sharing information from the VVDS into a CF structure. It provides a substantial performance improvement for shared user catalogs by placing data in a structure in the CF—the performance obtained through the use of this feature approaches that achieved with non-shared catalogs. For further information, refer to 2.11, "DFSMS Enhanced Catalog Sharing (ECS) Structures" on page 134 in Volume 2: Cookbook, SG24-5638, and to the redbook *Enhanced Catalog Sharing and Management*, SG24-5594.

A.5.6 Shared System Residence (SYSRES) Volume

Today's DASD is typified by high performance and high reliability. This, coupled with the ability to use more than one I/O definition (with HCD) with any particular system residence (SYSRES) volume, has made sharing SYSRES volumes a valuable software management tool.

A.5.6.1 What Is a Shared SYSRES?

A shared SYSRES is used as the IPL volume for more than one OS/390 image at a time. It contains the target libraries for OS/390 and its related products. Although it will be discussed here as though it is a single volume, more than one volume might be needed to contain your target libraries.

The number of images that can share a SYSRES is limited predominately by the number of paths to the DASD device on which it resides. Features like VLF and LLA have helped to significantly reduce the number of I/Os to a typical SYSRES, making it viable to share SYSRES among systems without performance issues. The only time that SYSRES performance could become an issue is if many systems are IPLing from that volume at the same time—hopefully a rare event!

Extended Indirect Volser Support: Indirect cataloging by using VOLUMES(******) on an IDCAMS non-VSAM define job has long been used to support shared SYSRES and Master Catalog. As more and more products are integrated with OS/390, it is probable that the SYSRES data sets will extend over more than one physical volume. To allow the continued use of indirect cataloging in a multi-volume SYSRES environment, OS/390 R3 introduced indirect volser cataloging support.

OS/390 provides a system symbol, &SYSR1, that contains the volser of the IPL volume. Using this information, system symbols can be created to contain the volsers of the overflow SYSRES volumes. These symbols can then be used in Parmlib members and IDCAMS DEFINE statements to create indirect catalog entries for the data sets on the overflow volumes.

For more information, see *DFSMS/MVS Access Method Services for ICF Catalogs*.

A.5.6.2 Availability Considerations for Shared SYSRES

Before choosing to implement a shared SYSRES in your environment, you should consider the implications for availability. Given the availability characteristics of modern DASD subsystems (RAID, mirroring, and so on), most installations are not overly concerned about a single SYSRES being a single point of failure. However, you should consider the implications of such a move before proceeding.

Statistically, the availability of all the systems in the sysplex is improved with a shared SYSRES volume—the chance of one SYSRES failing is less than the chance that one of a number of SYSRESes will fail (if you have a separate SYSRES for each system). The availability of individual systems within the sysplex is unaffected by whether you share the SYSRES or not.

There are times when the availability of the entire sysplex is less important than the availability of some combination of systems in the sysplex. When you implement a shared SYSRES, it becomes a single critical resource for all the systems that share it. In this way, it becomes a resource analogous to other single critical resources in your installation, such as the RACF database.

There are situations where you may wish to avoid sharing the SYSRES. For instance, you might need to limit the number of single critical resources for two or more systems. (For example, a single SYSRES probably should not be shared between an IMS system and the system its XRF backup runs on. Either the primary IMS system or its backup should have a dedicated system residence volume.)

In this and similar cases, the more important consideration is the probability of both systems failing at the same time. This probability is increased in a shared SYSRES environment. You might also want to avoid a shared SYSRES environment between images that require frequent maintenance and those that do not.

A.5.6.3 XCF and XES Considerations for Parallel Sysplex

For more information about XCF/XES Parallel Sysplex configuration options, refer to 2.4.1.1, "XCF and XES Considerations" on page 75 in Volume 1: Overview, SG24-5637.

A.5.6.4 XCF Signalling Paths Considerations for Parallel Sysplex

For more information about XCF signalling path configuration options, refer to 2.4.1.2, "XCF Signalling Path Considerations" on page 76 in Volume 1: Overview, SG24-5637.

A.5.7 Considerations for Couple Data Sets

When implementing a Parallel Sysplex, it is necessary for a number of couple data sets to be shared by some (or every) system in the Parallel Sysplex. Those couple data sets are:

- Sysplex couple data sets (also known as XCF couple data sets)
- CF resource manager (CFRM) couple data sets
- Sysplex failure management (SFM) couple data sets
- Workload Manager (WLM) couple data sets
- Automatic restart manager (ARM) couple data sets
- System logger (LOGR) couple data sets

The sysplex CDS must be shared by all systems in the Parallel Sysplex. The other couple data sets must be shared by the systems using this function.

When planning for the couple data sets, the following considerations should be taken into account. These considerations are applicable not only to the sysplex (or XCF) couple data sets, but to the couple data sets for CFRM, SFM, WLM, ARM and LOGR policy data, as well.

- An alternate couple data set.

  To avoid a single point of failure in the sysplex, all couple data sets should also have an alternate couple data set on a different device, control unit, and channel from the primary.

Attention: IPLing with Just One Sysplex Couple Data Set

It is possible, but not recommended, to IPL with one sysplex couple data set, if only the primary sysplex couple data set is specified in the COUPLExx member.

If *both* the primary and alternate sysplex couple data sets are specified, APAR OW30658 is required to let the system IPL if both couple data sets are not available. Prior to this APAR, it was necessary to have an alternate COUPLExx member in order to be able to IPL in this situation.

Because of the exposure when running without an alternate, with this APAR, message IXC267E is issued at every IPL whenever there is a primary but no alternate CDS.

- A spare couple data set.

  When the alternate couple data set replaces the primary, the original primary data set is deallocated and there is no longer an alternate couple data set. Because it is highly recommended to always have an alternate couple data set available to be switched, consider formatting three data sets before IPL. For example:

  – SYS1.XCF.CDS01 - specified as primary couple data set
  – SYS1.XCF.CDS02 - specified as alternate couple data set
  – SYS1.XCF.CDS03 - spare

  Then, if the alternate (CDS02) becomes the primary, you can issue the SETXCF COUPLE,ACOUPLE command to make the spare data set (CDS03) the alternate. Details of the command are found in *OS/390 MVS System Commands*.

  A couple data set can be switched by the operator through use of the SETXCF command, and by the system because of error conditions.

  The SETXCF command can be used to switch from the primary couple data set to the alternate couple data set.

- A couple data set cannot span volumes.

  XCF does not support multi-volume data sets.

- A multiple extent couple data set is not supported.

  For the sysplex couple data set, the format utility determines the size of the data set based on the number of groups, members, and systems specified, and allocates space on the specified volume for the data set. There must be enough contiguous space available on the volume for the couple data set.

  For the couple data sets that support administrative data (for example CFRM and SFM), the format utility determines the size of the data sets based on the number of parameters within the policy type that is specified.

- A couple data set is used by only one sysplex.

  The name of the sysplex for which a data set is intended must be specified when the couple data set is formatted. The data set can be used only by systems running in the sysplex whose name matches that in the couple data set.

  Each sysplex must have a unique name, and each *system* in the sysplex must have a unique name. Each couple data set for a sysplex, therefore, must be formatted using the sysplex name for which the couple data set is intended.

- The couple data set must not exist prior to formatting.

  The format utility cannot use an existing data set. This prevents the accidental reformatting of an active couple data set. You must delete an existing couple data set before reformatting it.

- Couple data set placement.

  Couple data sets should be placed on volumes that do not have high I/O activity. The I/O activity to the volume by XCF is not great; however, it is essential that XCF be able to get to the volume whenever it has to. For the same reason, you should not place the couple data set on a volume that is subject to reserves, has page data sets, or may have an SVC dump data set allocated and written on that volume.

  Attention: Status Missing Update Condition

  If SFM is active for status update missing conditions, and such a condition occurs because of the I/O being disrupted, it is possible that systems could be partitioned from the sysplex.

  The system health check method is enhanced with APAR OW30926. In addition to checking the other system status record in sysplex couple data sets, SFM uses XCF signaling to check the other system's health. This enhancement reduces the chance of SFM automatically partitioning a healthy system out of the sysplex.

  Refer to 3.6.1, “Sysplex Failure Management (SFM)” on page 119 in Volume 1: Overview, SG24-5637, for more information about OW30926.

  If the volume that the couple data set resides on is one for which DFDSS does a full volume backup, you will have to take this into consideration and possibly plan to switch the primary to the alternate during the backup to avoid a status update missing condition due to the reserve against the volume by DFDSS.

  See A.4.1.1, “DFSMSdss Considerations for Parallel Sysplex” on page 139 for a discussion of a possible solution to this issue.

  When selecting a volume for an alternate couple data set, use the same considerations as described for the primary. When XCF writes to the couple data set, it first writes to the primary, waits for a successful completion, and then writes to the alternate. Not until the write to the alternate is successful is the operation complete.

- Performance and availability considerations.

  The placement of couple data sets can improve performance, as well as availability. For maximum performance and availability, each couple data set should be on its own volume. However, this may be an expensive approach. The following examples provide the lowest cost approach that is workable:

  – Do not place the primary sysplex couple data set on the same volume as the primary CFRM couple data set. All other primary couple data sets can reside on the same volume, and all other alternate couple data sets can reside on a different volume. In the following table, (P) represents the primary and (A) represents the alternate couple data sets:

```
Volume X Couple Data Sets      Volume Y Couple Data Sets
-------------------------      -------------------------
     Sysplex (P)                    Sysplex (A)
     CFRM (A)                       CFRM (P)
     SFM (P)                        SFM (A)
     WLM (P)                        WLM (A)
     ARM (P)                        ARM (A)
     LOGR (P)                       LOGR (A)
```

Note: In this example, if access to volume X fails for any reason and a switch is made to the alternate CDSes, then it will be noted that the sysplex and CFRM CDSes are now on the same volume. It may be advantageous to put the CFRM (or sysplex) pair on two completely different volumes to ensure this situation does not arise under any circumstances.

– Place couple data sets on volumes that are attached to cached control units with the DASD fast write (DFW) feature. This recommendation applies to all couple data sets in any size sysplex. The couple data sets most affected by this are the sysplex couple data set and the CFRM couple data set. The recommendation becomes more important as the number of systems in the sysplex increases.

– Place couple data sets on volumes that are not subject to reserve/release contention or significant I/O contention from sources not related to couple data sets. This is true even if the I/O contention is sporadic.

Attention: Effect of Over-Defining CDS Size

Avoid specifying large parameter values when formatting the couple data set. This leads to wasted space and degraded couple data set performance. A couple data set formatted with the maximum values will be more than eight hundred 3390 cylinders in size. A couple data set of this size can take over 60 minutes to initialize when placed behind a control unit without DASD Fast Write (DFW).

The size of the couple data set can always be *increased* by using PSWITCH to switch to an alternate couple data set that is formatted with parameters greater than the current couple data set. It is not possible to PSWITCH to a smaller couple data set.

- MIH considerations for couple data sets.

  The interval for missing interrupts is specified on the DASD parameter of the MIH statement in the IECIOSxx parmlib member. The default time is 15 seconds. If there is little or no I/O contention on the DASD where the couple data sets reside, consider specifying a lower interval (such as 7 seconds) to be used by MIH in scanning for missing interrupts. A lower value alerts the image earlier to a problem with a couple data set.

  XCF has been changed to exploit the I/O Timing facility with APAR OW26649. XCF will set a time value for the completion of each I/O. This value will be 15 seconds, or one-third of the system's failure detection interval, whichever is smaller. Refer to 3.6.1, "Sysplex Failure Management (SFM)" on page 119 in Volume 1: Overview, SG24-5637, for more information about OW26649. Also refer to WSC Flash 98038, *XCF Service Recommendations to maximize Sysplex Availability*.

The recommended MIH value of some DASD subsystems are larger than 3390, for example the value of RAMAC virtual array subsystem is 300 seconds. It is recommended to specify the IOTIMING value in the IECIOSxx parmlib member or install APAR OW26649.

For more information about the placement of control data sets, especially in relation to large capacity storage subsystems, refer to *OS/390 Critical Dataset Placement Considerations*, WSC FLASH 10017.

- Security considerations for couple data sets.

  Consider RACF-protecting the couple data sets with the appropriate level of security. If you are using RACF, you want to ensure that XCF has authorization to access RACF-protected sysplex resources. The XCF STC must have an associated RACF user ID defined in the RACF started task procedure table. The started procedure name is XCFAS.

A.5.7.1 Couple Data Set Sizing

The size of the couple data set is a function of the number of systems, number of XCF groups, and number of members within the groups:

$$CDS\ Size = f(No.\ Systems,\ No.\ XCF\ Groups,\ No.\ Members\ Per\ Group)$$

For all practical purposes, the size is primarily a function of the number of groups, and secondarily a function of the number of members per group. The number of systems in the sysplex has a smaller effect on the size. Figure 41 on page 163 gives an indication of the size of the sysplex couple data set that is formatted using the IXCL1DSU utility. It indicates the number of 3390-3 tracks that will be formatted based on the number of systems, XCF groups and group members, and the use of the GRS star configuration.

If GRS star is not used, then the indicated size is reduced by seven tracks in all combinations. If the number of systems exceeds 16, then the indicated size is increased by two tracks. If the maximums are taken for all parameters, then the resulting size is 12,291 tracks (approximately 800 cylinders).

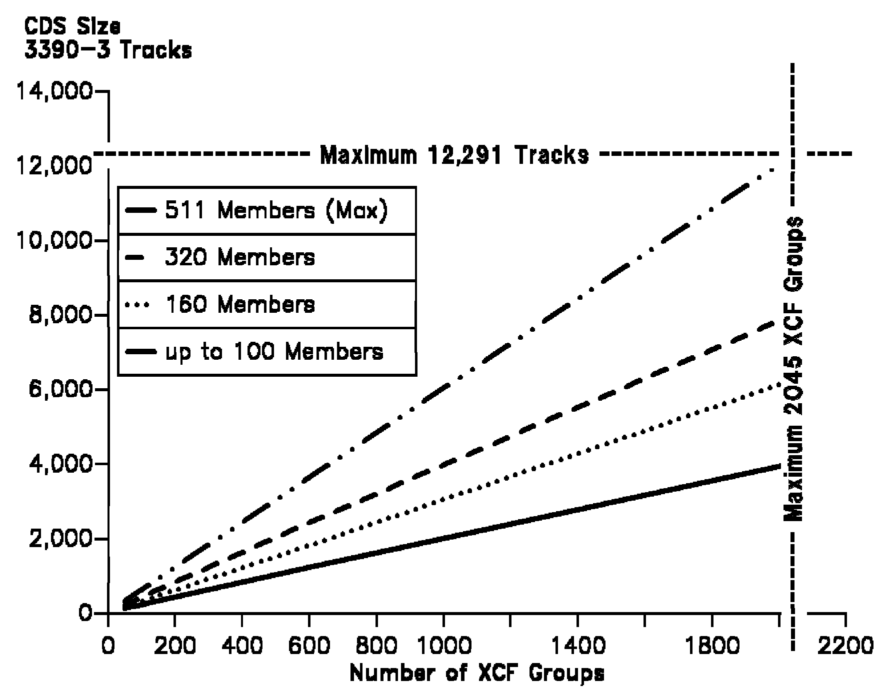


Figure 41. Sysplex Couple Data Set Size - Maximum Configuration

Figure 42 shows an excerpt from Figure 41 that represents likely configurations.

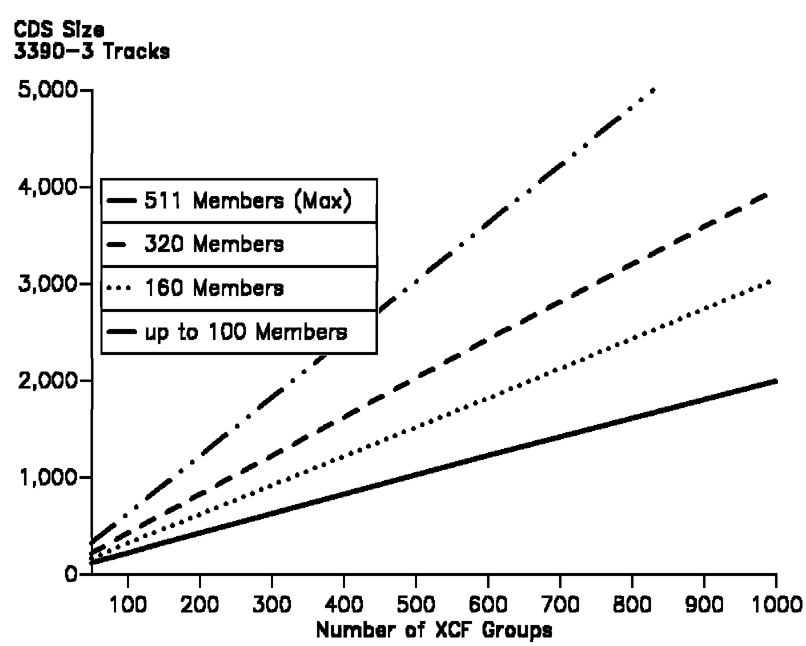


Figure 42. Sysplex Couple Data Set Size - Likely Configuration

More important than the waste of DASD space is the impact on performance. This is the main reason for recommending that you do not grossly over-size your control data sets. Remember, if the control data set is too small, it is easy to allocate a new, larger one and switch to that. However, a sysplex-wide IPL is required if you wish to switch to a *smaller* control data set.

For more information about other aspects of couple data sets, see *OS/390 MVS Setting Up a Sysplex*.

A.5.8 Rebuild for CF Maintenance (Rebuild Enhancement)

When a CF (CF1) is to be taken down for maintenance or upgrade, you must rebuild all structures which are allocated in that CF to another CF (CF2). In this case, you may use the following command to rebuild all structures that support rebuild, except for the XCF structures.

```
SETXCF START,REBUILD,CFNAME=CF1,LOC=OTHER
```

For XCF structures, you must specifically name the structure, as in the following command:

```
SETXCF START,REBUILD,STRNM=XCF structure name,LOC=OTHER
```

Once CF1 is brought back into the configuration, however, there is no good way to restore the structures that belong in CF1 back into CF1. You could use the following command:

```
SETXCF START,REBUILD,CFNAME=CF2,LOC=NORMAL
```

However, this command causes *all* structures (that support rebuild) to be rebuilt, with the following results:

- Some will rebuild in place in CF2.
- Some will move to their preferred location in CF1.

This is not an attractive option, because the structures that are *supposed* to be in CF2 should not have to go through rebuild processing.

A.5.8.1 REBUILD POPULATECF Command

In OS/390 V2R6, a new rebuild request type was introduced:

```
SETXCF START,REBUILD,POPULATECF=cfname
```

This designates a target CF which is to be populated with structures rebuilt from other CF locations, based on the preference list defined in the CFRM policy.

Any allocated structure (in any CF) is marked "pending rebuild" if the POPULATECF appears higher in that preference list than the CF in which the structure is currently allocated. Then OS/390 initiates the structure rebuild serially, one structure at a time, for structures that have been marked "pending rebuild." As each structure is rebuilt, OS/390 applies its structure allocation weighting criteria, which is calculated using the REBUILDPERCENT parameter in the CFRM policy, and the WEIGHT parameter in the SFM policy, to determine if the CF specified on the POPULATECF command is really a more preferred location than where the CF is currently allocated, and:

- If so, the rebuild proceeds to rebuild the structure into the designated POPULATECF target CF.
- If not, the rebuild is automatically stopped and the structure does not move from its current structure.

The POPULATECF may be useful for the following cases:

- Re-populating a CF after it has been temporarily removed from the configuration and then restored.

- Initially populating a new CF that has been added to the configuration.

Another advantage of POPULATECF is that it can be used to move the XCF structures along with everything else, whereas doing a rebuild using LOC=OTHER or LOC=NORMAL requires that the XCF structures are moved individually. This substantially reduces systems management and operational complexity.

For more detail information about rebuild characteristics, refer to 2.0, “A Structured View of CF Structures” on page 59 in Volume 2: Cookbook, SG24-5638.

A.5.8.2 REBUILD POPULATECF Command Output Example

Figure 43 shows the output resulting from the use of the POPULATECF option. In this example, IEFAUTOS, DSNDSGC_LOCK1 and DSNDSGC_SCA .A. are all originally allocated to CFT2. Since their preference list is (CFT1,CFT2), it is determined that they should be rebuilt in CFT1. The other structures .b. are not rebuilt because they are already in the populate CF (CFT2). The structure IEFAUTOS rebuild is initiated at first .c., and after completing the rebuild, the next rebuild is initiated.

```
SETXCF START,REBUILD,POPCF=CFT1
IXC521I REBUILD FOR STRUCTURE IEFAUTOS 276
HAS BEEN STARTED
IXC540I POPULATECF REBUILD FOR CFT1 REQUEST ACCEPTED. 277
THE FOLLOWING STRUCTURES ARE PENDING REBUILD:   .A
IEFAUTOS               DSNDSGC_LOCK1          DSNDSGC_SCA
THE FOLLOWING WILL NOT BE REBUILT:   .b
JES2CKPT_1      :  STRUCTURE IS ALLOCATED IN A MORE PREFERRED CF
ISGLOCK         :  STRUCTURE IS ALLOCATED IN A MORE PREFERRED CF
IXC_DEFAULT_1   :  STRUCTURE IS ALLOCATED IN A MORE PREFERRED CF
IXC_DEFAULT_2   :  STRUCTURE ALREADY ALLOCATED IN THE POPULATECF
IRRXCF00_P001   :  STRUCTURE ALREADY ALLOCATED IN THE POPULATECF
IRRXCF00_B001   :  STRUCTURE IS ALLOCATED IN A MORE PREFERRED CF
:
IXC367I THE SETXCF START REBUILD REQUEST FOR COUPLING FACILITY 278
CFT1 WAS ACCEPTED.
IEF265I AUTOMATIC TAPE SWITCHING: REBUILD IN PROGRESS BECAUSE 299
THE OPERATOR REQUESTED IEFAUTOS REBUILD.   .c
:
IXC526I STRUCTURE IEFAUTOS IS REBUILDING FROM 602
COUPLING FACILITY CFT2 TO COUPLING FACILITY CFT1.
REBUILD START REASON: OPERATOR INITIATED
INFO108: 00000059 00000059.
IXL014I IXLCONN REBUILD REQUEST FOR STRUCTURE IEFAUTOS 603
WAS SUCCESSFUL.  JOBNAME: ALLOCAS ASID: 0013
CONNECTOR NAME: IEFAUTOSSC42 CFNAME: CFT1
REBUILD START REASON: OPERATOR INITIATED
INFO108: 00000059 00000059.
IXL014I IXLCONN REBUILD REQUEST FOR STRUCTURE IEFAUTOS 603
WAS SUCCESSFUL.  JOBNAME: ALLOCAS ASID: 0013
CONNECTOR NAME: IEFAUTOSSC42 CFNAME: CFT1
IXL015I REBUILD NEW STRUCTURE ALLOCATION INFORMATION FOR 604
STRUCTURE IEFAUTOS, CONNECTOR NAME IEFAUTOSSC42
 CFNAME     ALLOCATION STATUS/FAILURE REASON
 --------   --------------------------------
 CFT1       STRUCTURE ALLOCATED
IXL014I IXLCONN REQUEST FOR STRUCTURE IEFAUTOS WAS SUCCESSFUL. 904
:
:
```

Figure 43. REBUILD POPULATECF Command Example

A.6 Parallel Sysplex Education

Parallel Sysplex is one of the first changes to OS/390 in a long time that can have a significant impact on the Operations department. There are now more components to manage, and more messages to be familiar with. Most important is the fact that events on one system are likely to have more of an impact on other systems in the sysplex than was the case in the past.

For this reason, it is important that education for the operators is considered as an integral part of migrating to, and *maintaining*, a Parallel Sysplex.

IBM and other education providers provide a number of Parallel Sysplex-related courses for operators and system programmers. In addition to this, however, there is now also a certification program available.

A.6.1 Parallel Sysplex Certification Program

This program was developed jointly by IBM and a number of Parallel Sysplex customers, and is designed to provide an independent measure of the Parallel Sysplex skills of the candidate. There are three levels of certification available:

- IBM Certified Specialist - Parallel Sysplex Operator
- IBM Certified Specialist - Parallel Sysplex Operator - Data Sharing
- IBM Certified Specialist - Parallel Sysplex System Programmer

More information is available on the Web at:
`http://www.s390.ibm.com/pso/rfa.html`

In addition, a sample test is also available on the Web at:
`http://www.ibm.com/education/certify/tests/sam320.phtml`

Appendix B. Special Notices

This publication is intended to help large systems customers configure a Parallel Sysplex.

References in this publication to IBM products, programs or services do not imply that IBM intends to make these available in all countries in which IBM operates. Any reference to an IBM product, program, or service is not intended to state or imply that only IBM's product, program, or service may be used. Any functionally equivalent program that does not infringe any of IBM's intellectual property rights may be used instead of the IBM product, program or service.

Information in this book was developed in conjunction with use of the equipment specified, and is limited in application to those specific hardware and software products and levels.

IBM may have patents or pending patent applications covering subject matter in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to the IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785.

Licensees of this program who wish to have information about it for the purpose of enabling: (i) the exchange of information between independently created programs and other programs (including this one) and (ii) the mutual use of the information which has been exchanged, should contact IBM Corporation, Dept. 600A, Mail Drop 1329, Somers, NY 10589 USA.

Such information may be available, subject to appropriate terms and conditions, including in some cases, payment of a fee.

The information contained in this document has not been submitted to any formal IBM test and is distributed AS IS. The information about non-IBM ("vendor") products in this manual has been supplied by the vendor and IBM assumes no responsibility for its accuracy or completeness. The use of this information or the implementation of any of these techniques is a customer responsibility and depends on the customer's ability to evaluate and integrate them into the customer's operational environment. While each item may have been reviewed by IBM for accuracy in a specific situation, there is no guarantee that the same or similar results will be obtained elsewhere. Customers attempting to adapt these techniques to their own environments do so at their own risk.

Any pointers in this publication to external Web sites are provided for convenience only and do not in any manner serve as an endorsement of these Web sites.

Any performance data contained in this document was determined in a controlled environment, and therefore, the results that may be obtained in other operating environments may vary significantly. Users of this document should verify the applicable data for their specific environment.

This document contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples contain the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

Reference to PTF numbers that have not been released through the normal distribution process does not imply general availability. The purpose of including these reference numbers is to alert IBM customers to specific information relative to the implementation of the PTF when it becomes available to each customer according to the normal IBM PTF distribution process.

The following terms are trademarks of the International Business Machines Corporation in the United States and/or other countries:

ACF/VTAM	AD/Cycle
ADSTAR	Advanced Function Printing
Advanced Peer-to-Peer Networking	AFP
AIX	AnyNet
APPN	BatchPipes
BatchPipeWorks	BookManager
C/370	CBIPO
CBPDO	CICS
CICS/ESA	CICS/MVS
CICS/VSE	CICSPlex
CUA	DATABASE 2
DB2	DB2/2
DB2 Connect	DB2 Universal Database
DFSMS	DFSMS/MVS
DFSMSdfp	DFSMSdss
DFSMShsm	DFSORT
Distributed Relational Database Architecture	DRDA
eNetwork	Enterprise System/9000
Enterprise Systems Architecture/390	Enterprise Systems Connection Architecture
ES/3090	ES/9000
ESA/390	ESCON XDF
ESCON	Extended Services
GDDM	Hardware Configuration Definition
Hiperbatch	Hipersorting
Hiperspace	HPR Channel Connectivity
IBM	IBMLink
IMS	IMS/ESA
Intelligent Miner	Language Environment
LSPR	LSPR/PC
MERVA	MQ
MQSeries	Multiprise
MVS/DFP	MVS/ESA
MVS/SP	MVS/XA
Netfinity	Net.Data
NetView	NTune
Nways	OPC
OpenEdition	Operating System/2
OS/2	OS/390
OS/400	Parallel Sysplex
Personal System/2	Powered by S/390
PowerPC	PowerPC 604
PR/SM	Presentation Manager
Print Services Facility	Processor Resource/Systems Manager
PROFS	PS/2
PSF	QMF
RACF	RAMAC
Resource Measurement Facility	RETAIN

RISC System/6000	RMF
RRDF	RS/6000
S/370	S/390
S/390 Parallel Enterprise Server	SAA
SecureWay	SNAP/SHOT
SQL/DS	Sysplex Timer
System/370	System/390
SystemPac	Systems Application Architecture
Systems Validation Services	SystemView
S/390 Parallel Enterprise Server	VisualAge
VM/ESA	VM/XA
VSE/ESA	VTAM
WebSphere	WIN-OS/2
3090	400

The following terms are trademarks of other companies:

Tivoli, Manage. Anything. Anywhere.,The Power To Manage., Anything. Anywhere., TME, NetView, Cross-Site, Tivoli Ready, Tivoli Certified, Planet Tivoli, and Tivoli Enterprise are trademarks or registered trademarks of Tivoli Systems Inc., an IBM company, in the United States, other countries, or both. In Denmark, Tivoli is a trademark licensed from Kjøbenhavns Sommer - Tivoli A/S.

C-bus is a trademark of Corollary, Inc. in the United States and/or other countries.

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Sun Microsystems, Inc. in the United States and/or other countries.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States and/or other countries.

PC Direct is a trademark of Ziff Communications Company in the United States and/or other countries and is used by IBM Corporation under license.

ActionMedia, LANDesk, MMX, Pentium and ProShare are trademarks of Intel Corporation in the United States and/or other countries.

SET, SET Secure Electronic Transaction, and the SET logo are trademarks owned by SET Secure Electronic Transaction LLC.

UNIX is a registered trademark in the United States and other countries licensed exclusively through The Open Group.

Other company, product, and service names may be trademarks or service marks of others.

Appendix C. Related Publications

The publications listed in this section are considered particularly suitable for a more detailed discussion of the topics covered in this redbook.

Note that some of these publications are only available in softcopy—either on one of the CD collections or on the Internet.

C.1 IBM Redbooks

For information on ordering these ITSO publications see “How to get IBM Redbooks” on page 181.

The publications are listed in alphabetical order.

- *A Performance Study of Web Access to CICS*, SG24-5748
- *Accessing CICS Business Applications from the World Wide Web*, SG24-4547
- *Accessing DB2 for OS/390 Data from the World Wide Web*, SG24-5273
- *Automation for S/390 Parallel Sysplex*, SG24-4549
- *Batch Processing in a Parallel Sysplex*, SG24-5329
- *CICS and VSAM RLS: Implementation Guide*, SG24-4766 (available in softcopy only)
- *CICS and VSAM RLS: Planning Guide*, SG24-4765 (available in softcopy only)
- *CICS and VSAM RLS: Recovery Considerations*, SG24-4768 (available in softcopy only)
- *CICS TS for OS/390 V1R2 Implementation Guide*, SG24-2234
- *CICS TS for OS/390: V1.3 Implementation Guide*, SG24-5274
- *CICS TS: Web Interface and 3270 Bridge*, SG24-5243
- *CICS Workload Management Using CICSPlex SM and the MVS/ESA Workload Manager*, GG24-4286
- *Connecting IMS to the World Wide Web: A Practical Guide to IMS Connectivity*, SG24-2220
- *Consolidating UNIX Systems onto OS/390*, SG24-2090
- *Continuous Availability S/390 Technology Guide*, SG24-2086
- *Continuous Availability Systems Design Guide*, SG24-2085
- *Data Warehousing with DB2 for OS/390*, SG24-2249
- *DB2 for MVS/ESA Version 4 Data Sharing Implementation*, SG24-4791
- *DB2 for MVS/ESA V4 Data Sharing Performance Topics*, SG24-4611
- *DB2 for OS/390 Application Design Guidelines for High Performance*, SG24-2233
- *DB2 for OS/390 Capacity Planning*, SG24-2244
- *DB2 for OS/390 V5 Performance Topics*, SG24-2213
- *DB2 on the MVS Platform: Data Sharing Recovery*, SG24-2218
- *DB2 UDB for OS/390 and Continuous Availability*, SG24-5486
- *DB2 UDB for OS/390 Version 6 Performance Topics*, SG24-5351
- *DFSMS Optimizer Usage Guide*, SG24-2235
- *Disaster Recovery Library: Database Recovery*, GG24-3993 (available in softcopy only)
- *Enhanced Catalog Sharing and Management*, SG24-5594
- *Getting Started with DB2 Stored Procedures: Give Them a Call through the Network*, SG24-4693
- *Getting the Most Out of a Parallel Sysplex*, SG24-2073
- *HCD and Dynamic I/O Reconfiguration Primer*, SG24-4037
- *IMS/ESA Data Sharing in a Parallel Sysplex*, SG24-4303

- *IMS e-business Connect Using the IMS Connectors*, SG24-5427
- *IMS/ESA Multiple Systems Coupling in a Parallel Sysplex*, SG24-4750
- *IMS/ESA Shared Queues: A Planning Guide*, SG24-5257
- *IMS/ESA Sysplex Data Sharing: An Implementation Case Study*, SG24-4831
- *IMS/ESA V6 Shared Queues*, SG24-5088
- *IMS/ESA V6 Parallel Sysplex Migration Planning Guide*, SG24-5461
- *IMS/ESA V6 Guide*, SG24-2228
- *IMS Version 7 Release Guide*, SG24-5753
- *Inside APPN: The Essential Guide to the Next-Generation SNA*, SG24-3669
- *JES3 in a Parallel Sysplex*, SG24-4776
- *Modeling Host Environments using SNAP/SHOT*, SG24-5314
- *Net.Commerce for OS/390*, SG24-5154
- *Open Systems Adapter 2 Implementation Guide*, SG24-4770
- *OS/390 eNetwork Communications Server V2R7 TCP/IP Implementation Guide Volume 1: Configuration and Routing*, SG24-5227
- *OS/390 eNetwork Communications Server V2R7 TCP/IP Implementation Guide Volume 2: UNIX Applications*, SG24-5228
- *OS/390 MVS Multisystem Consoles Implementing MVS Sysplex Operations*, SG24-4626
- *OS/390 Parallel Sysplex Application Considerations Presentation Guide*, SG24-4743
- *OS/390 MVS Parallel Sysplex Capacity Planning*, SG24-4680
- *OS/390 MVS Parallel Sysplex Performance Healthcheck Case Study: DMdata/Danske Data*, SG24-5373
- *OS/390 R5 Implementation*, SG24-5151
- *OS/390 Workload Manager Implementation and Exploitation*, SG24-5326
- *Parallel Sysplex CF Online Monitor: Installation and User's Guide*, SG24-5153
- *Parallel Sysplex Automation Guidelines*, SG24-5441
- *Parallel Sysplex Automation: Using System Automation for OS/390*, SG24-5442
- *Parallel Sysplex Continuous Availability Case Studies*, SG24-5346
- *Parallel Sysplex - Managing Software for Availability*, SG24-5451
- *Parallel Sysplex Operational Scenarios*, SG24-2079
- *Planning for IBM Remote Copy*, SG24-2595
- *RACF V2.2 Installation and Implementation Guide*, SG24-4580
- *RAMAC Virtual Array: Implementing Peer-to-Peer Remote Copy*, SG24-5338
- *Revealed! CICS Transaction Gateway and More CICS Clients Unmasked*, SG24-5277
- *Revealed! Architecting Web Access to CICS*, SG24-5466
- *Securing Web Access to CICS*, SG24-5756
- *Selecting a Server - The Value of S/390*, SG24-4812
- *SNA in a Parallel Sysplex Environment*, SG24-2113
- *S/390 G3: Enterprise Server Complex Systems Availability and Recovery Configuration Guidelines Student Handout*, SG24-4943 (available in softcopy only)
- *S/390 G3: Enterprise Server Complex Systems Availability and Recovery Configuration Guidelines*, SG24-4927 (available in softcopy only)
- *S/390 G3: Enterprise Server Complex Systems Availability and Recovery Exercise Guide*, SG24-4913 (available in softcopy only)
- *S/390 G3: Enterprise Server Complex Systems Availability and Recovery Exercise Installation and Run-Time Procs*, SG24-4912 (available in softcopy only)
- *S/390 G3: Enterprise Server Complex Systems Availability and Recovery Presentation Guide*, SG24-4911 (available in softcopy only)
- *S/390 G3: Enterprise Server Complex Systems Availability and Recovery Student Handout*, SG24-4942 (available in softcopy only)

- *S/390 MVS Parallel Sysplex Batch Performance*, SG24-2557 (available in softcopy only)
- *S/390 MVS Parallel Sysplex Continuous Availability Presentation Guide*, SG24-4502
- *S/390 MVS Parallel Sysplex Continuous Availability SE Guide*, SG24-4503
- *S/390 MVS Parallel Sysplex Migration Paths*, SG24-2502
- *S/390 MVS Parallel Sysplex Performance*, SG24-4356
- *S/390 MVS/ESA SP V5 WLM Performance Studies*, SG24-4352
- *S/390 OSA-Express Gigabit Ethernet Implementation Guide*, SG245443
- *S/390 OSA-Express Implementation Guide*, SG245948
- *S/390 Parallel Sysplex: Resource Sharing*, SG24-5666
- *TCP/IP in a Sysplex*, SG24-5235
- *Using VTAM Generic Resources with IMS*, SG24-5487
- *World Wide Web Access to DB2*, SG24-4716
- *WOW! DRDA Supports TCP/IP: DB2 Server for OS/390 and DB2 Universal Database*, SG24-2212

C.2 IBM Redbooks collections

Redbooks are also available on the following CD-ROMs. Click the CD-ROMs button at `http://www.redbooks.ibm.com/` for information about all the CD-ROMs offered, updates and formats.

CD-ROM Title	Collection Kit Number
System/390 Redbooks Collection	SK2T-2177
Networking and Systems Management Redbooks Collection	SK2T-6022
Transaction Processing and Data Management Redbooks Collection	SK2T-8038
Lotus Redbooks Collection	SK2T-8039
Tivoli Redbooks Collection	SK2T-8044
AS/400 Redbooks Collection	SK2T-2849
Netfinity Hardware and Software Redbooks Collection	SK2T-8046
RS/6000 Redbooks Collection (BkMgr Format)	SK2T-8040
RS/6000 Redbooks Collection (PDF Format)	SK2T-8043
Application Development Redbooks Collection	SK2T-8037
IBM Enterprise Storage and Systems Management Solutions	SK3T-3694

C.3 Other Resources

These publications are also relevant as further information sources.

The publications are listed in alphabetical order.

- *A Technical Overview: VTAM V4R2, NCP V6R3, V7R1 and V7R2*, GG66-3256
- *BatchPipes for OS/390 V2 R1 BatchPipeWorks Reference*, SA22-7456
- *BatchPipes for OS/390 V2 R1 BatchPipeWorks User Guide*, SA22-7457
- *BatchPipes for OS/390 V2 R1 Users Guide and Reference*, SA22-7458
- *BatchPipes for OS/390 V2 R1 Introduction*, GA22-7459
- *CICS Transaction Affinities Utility Guide*, SC33-1777
- *CICS IMS Database Control Guide*, SC33-1700
- *CICSPlex Systems Manager V1 Concepts and Planning*, GC33-0786
- *CICS TS for OS/390 Installation Guide*, GC33-1681
- *CICS TS for OS/390 Intercommunication Guide*, SC33-1695
- *CICS TS for OS/390 Migration Guide*, GC34-5353
- *CICS TS for OS/390 Operations and Utilities Guide*, SC33-1685
- *CICS TS for OS/390 Recovery and Restart Guide*, SC33-1698
- *CICS TS for OS/390 Release Guide*, GC33-1570

- *CICS TS for OS/390 System Definition Guide*, SC33-1682
- *DB2 UDB for OS/390 V6 Administration Guide*, SC26-9003
- *DB2 UDB for OS/390 V6 Data Sharing: Planning and Administration*, SC26-9007
- *DB2 UDB for OS/390 V6 Release Guide*, SC26-9013
- *DFSMS/MVS V1.5 Access Method Services for ICF Catalogs*, SC26-4906
- *DFSMS/MVS V1.5 DFSMSdfp Storage Administration Reference*, SC26-4920
- *DFSMS/MVS V1.5 General Information*, GC26-4900
- *DFSMS/MVS V1.5 Managing Catalogs*, SC26-4914
- *DFSMS/MVS V1.5 Planning for Installation*, SC26-4919
- *DFSMS/MVS Remote Copy Guide and Reference*, SC35-0169
- *ES/9000 and ES/3090 PR/SM Planning Guide*, GA22-7123
- *Fiber Optic Link Planning*, GA23-0367
- *Hardware Configuration Manager User's Guide*, SC33-6469
- *IBM RAMAC Scalable Array Storage Introduction*, GC26-7212
- *IBM RAMAC Virtual Array Storage Introduction*, GC26-7168
- *IBM RAMAC Virtual Array Storage Operation and Recovery*, GC26-7171
- *IBM RAMAC Virtual Array Storage Planning, Implementation and Usage*, GC26-7170
- *IBM RAMAC Virtual Array Storage Reference*, GC26-7172
- *IBM Token-Ring Network Introduction and Planning Guide*, GA27-3677
- *IBM 3990/9390 Storage Control Planning, Installation and Storage Administration Guide*, GA32-0100
- *IBM 3990/9390 Reference for Model 6*, GA32-0274
- *IBM 3990/9390 Operations and Recovery Guide*, GA32-0253
- *IBM 9397 RAMAC Electronic Array Storage Introduction*, GC26-7205
- *IMS/ESA V6 Administration Guide: Database Manager*, SC26-8725
- *IMS/ESA V6 Administration Guide: System*, SC26-8730
- *IMS/ESA V6 Administration Guide: Transaction Manager*, SC26-8731
- *IMS/ESA V6 Database Recovery Control (DBRC) Guide and Reference*, SC26-8733
- *IMS/ESA V6 Installation Volume 2: System Definition & Tailoring*, SC26-8737
- *IMS/ESA V6 Operations Guide*, SC26-8741
- *IMS/ESA V6 Release Planning Guide*, GC26-8744
- *IMS/ESA V6 Utilities Reference: Database Manager*, SC26-8769
- *IMS/ESA V6 Utilities Reference: System*, SC26-8770
- *Large Systems Performance Reference*, SC28-1187
- *NCP V7.8 Resource Definition Guide*, SC31-6223
- *Network Products Reference*, GX28-8002
- *Nways Multiprotocol Access Services Configuration Reference Vol 1*, SC30-3884
- *Nways Multiprotocol Access Services Configuration Reference Vol 2*, SC30-3885
- *OS/390 MVS An Introduction to OS/390*, GC28-1725
- *OS/390 MVS Conversion Notebook*, GC28-1747
- *OS/390 MVS Hardware Configuration Definition: User's Guide*, SC28-1848
- *OS/390 MVS Initialization and Tuning Guide*, SC28-1751
- *OS/390 MVS Initialization and Tuning Reference*, SC28-1752
- *OS/390 MVS Parallel Sysplex Application Migration*, GC28-1863
- *OS/390 MVS Parallel Sysplex Hardware and Software Migration*, GC28-1862
- *OS/390 MVS Parallel Sysplex Overview*, GC28-1860
- *OS/390 MVS Parallel Sysplex Test Report*, GC28-1963 (available in softcopy only)
- *OS/390 MVS Planning: APPC/MVS Management*, GC28-1807
- *OS/390 MVS Planning: Global Resource Serialization*, GC28-1759

- *OS/390 MVS Planning: Operations*, GC28-1760
- *OS/390 MVS Planning: Security*, GC28-1920
- *OS/390 MVS Planning: Workload Management*, GC28-1761
- *OS/390 MVS Programming: Assembler Services Guide*, GC28-1762
- *OS/390 MVS Programming: Assembler Services Reference*, GC28-1910
- *OS/390 MVS Programming: Authorized Assembler Services Guide*, GC28-1763
- *OS/390 MVS Programming: Sysplex Services Reference*, GC28-1496
- *OS/390 MVS Programming: Sysplex Services Guide*, GC28-1495
- *OS/390 MVS Recovery and Reconfiguration Guide*, GC28-1777
- *OS/390 MVS System Commands*, GC28-1781
- *OS/390 MVS System Management Facilities (SMF)*, GC28-1783
- *OS/390 MVS RMF Performance Management Guide*, SC28-1951
- *OS/390 MVS RMF Report Analysis*, SC28-1950
- *OS/390 MVS RMF User's Guide*, SC28-1949 (available in softcopy only)
- *OS/390 MVS Security Server System Programmer's Guide*, SC28-1913
- *OS/390 MVS Setting Up a Sysplex*, GC28-1779
- *Planning for S/390 Open Systems Adapter Feature*, GC23-3870
- *S/390 9672/9674/2003 PR/SM Planning Guide*, GA22-7236
- *S/390 9672/9674 Hardware Management Console Guide*, GC38-0453
- *S/390 9672/9674 Managing Your Processors*, GC38-0452
- *S/390 9672/9674 Programming Interfaces: Hardware Management Console Appication*, SC28-8141
- *Sysplex Timer Planning*, GA23-0365
- *System Automation for OS/390 General Information*, GC28-1541
- *Planning for the 9037 Model 2*, SA22-7233
- *Using the IBM 9037 Model 2 Sysplex Timer*, SA22-7230
- *TME 10 OPC Installation Guide*, SH19-4379
- *OS/390 IBM Communications Server: IP Configuration Guide*, SC31-8725
- *OS/390 IBM Communications Server: SNA Migration*, SC31-8622
- *OS/390 IBM Communications Server: SNA Network Implementation Guide*, SC31-8563
- *OS/390 IBM Communications Server: SNA Resource Definition Reference*, SC31-8565
- *2216 Nways Multiaccess Connector Hardware Installation Guide*, GA27-4106
- *2216 Nways Multiaccess Connector Planning Guide*, GA27-4105
- *3745/3746 Overview*, GA33-0180
- *3745 Communications Controller Models A, 3746 Nways Multiprotocol Controller Models 900 and 950 Planning Guide*, GA33-0457
- *3746 Model 900 Migration and Planning Guide*, GA33-0183
- *3746 Model 900/950 Network Node Migration and Planning Guide*, GA33-0349

C.4 Packages

Most of the following packages are available on MKTTOOLS. Some packages come from other tools disks as noted.

MKTTOOLS is accessed by IBM employees in multiple ways:

- OMNIDISK: A VM facility that provides easy access to IBM TOOLS and conference disks
- TOOLCAT: A VM EXEC that provides a catalog of TOOLS disks and easy access to the packages that reside on the disks
- Direct CMS commands (useful if you already know the package name and where it resides)

Many of the packages listed here are available for customer usage; ask your IBM representative for more information.

The publications are listed in alphabetical order.

- ARMPAPER package: *The Automatic Restart Manager*[4]
- ARMWRAP package: *ARM Wrapper*
- BV390 package: *S/390 Parallel Sysplex Business Value*
- BWATOOL package: *Batch Workload Analysis Tool*
- CD13 package: *CICS MVS/ESA CMOS Processor Utilization Tool*[5]
- CHKLIST package: *Hints and Tips Checklist for Parallel Sysplex*
- CICSE41P package: *Performance Comparison of CICS/ESA V4.1 and V3.3*
- CICSTS01 package: *CICS TS for OS/390*
- CICSTS02 package: *CICS TS for OS/390 Presentation Guide*
- CICSTS03 package: *CICS TS for OS/390 Migration Guidance*
- CICSTS05 package: *CICS TS for OS/390 Questions and Answers*
- CICSTS06 package: *CICS TS for OS/390 Support for Parallel Sysplex*
- CICSTS07 package: *CICS TS a Technical Introduction presentation*
- CICSTS09 package: *CICS TS for OS/390 Performance Report*
- CICSTS12 package: *CICS TS Performance Report for CPSM 1.3.0*
- CICSTS13 package: *CICS TS and RLS Performance*
- 6950-22S package: *IBM Performance Management and Capacity Planning Svcs*
- CP2KOVER package: *An Overview of CP2000*[6]
- CP2000 package: *CP2000 OS/2 PC Tool*[7]
- CVRCAP package: *CICS/VSAM RLS Capacity Planning Tool*
- EMEAHELP package: *Parallel Sysplex focal points in EMEA*
- EPSOEMEA package: *EPSO Reference Guide and Reference Card plus EPSO Services Perform Guide and Statement of Work for EMEA*
- FICONPG package: *S/390 FICON Overview (FIPG) -- Connectivity Leadership*
- GBOF7101 package: *S/390 Parallel Sysplex Application Briefs*
- GDPS package: *S/390 Geographically Dispersed Parallel Sysplex Pguide*
- GF225005 package: *A Strategy Report to Customers*
- GF225008 package: *S/390 Technology Leadership White Paper*
- GF225009 package: *Parallel Sysplex Overhead: A Reality Check*
- GF225042 package: *S/390 Coupling Facility Configuration Alternatives*
- GF225114 package: *GDPS S/390 Multi-Site Application Availability Solution*
- GF225115 package: *IBM System/390 Value of Resource Sharing WP*
- G2219017 package: *IMS Workload Router - S/390*
- G3260594 package: *S/390 Software Pricing Reference Guide*
- G3263007 package: *SAP R/3 and S/390 - A Powerful Combination*
- G3263025 package: *S/390 Parallel Sysplex White Paper*
- G3263026 package: *e-business Powered by S/390 White Paper*
- G3263027 package: *S/390 SAP R/3 Whitepaper*
- G3263030 package: *Decision Support White Paper*
- G3263056 package: *S/390 Cost of Scalability Report*
- LPARCE package: *LPAR Capacity Estimator (LPAR/CE)*[8]
- LSPRPC package: *Large Systems Performance Reference (LSPR)*

---

4 TOOLS SENDTO KGNVMC TOOLSMVS MVSTOOLS GET ARMPAPER PACKAGE

5 TOOLS SENDTO WINVMB TOOLS TXPPACS GET CD13 PACKAGE

6 TOOLS SENDTO WSCVM TOOLS CPSTOOLS GET CP2KOVER PACKAGE

7 TOOLS SENDTO WSCVM TOOLS CPSTOOLS GET CP2000 PACKAGE

8 TOOLS SENDTO WSCVM TOOLS CPSTOOLS GET LPARCE PACKAGE

- OPSYSPLX package: *Why Implement an S/390 Operational Single Image (Sysplex)?*
- OSAPERF package: *Performance for S/390 Open Systems Adapter*
- OS390PSR package: *OS/390 Parallel Sysplex Recovery*
- OS390PRF package: *OS/390 Performance Studies*
- OS390TST package: *S/390 Parallel Sysplex Test Report*
- OS390T97 package: *OS/390 Parallel Sysplex Test Report*
- OS390T98 package: *OS/390 Parallel Sysplex Test Report*
- OS390T99 package: *OS/390 Parallel Sysplex Test Report*
- TLLBPAR package: *IBM S/390 Parallel Sysplex Cluster Technology P-Guide*
- TLLBOS79 package:*S/390 Tech Ldrshp Tech Library - OS/390 Version 2 R9*
- PCR package: *CP2000 Processor Capacity Reference (PCR)*[9]
- PSLCU package: *Parallel Sysplex License Charge (US Version)*
- QCBTRACE package: *Queue Control Block Trace*[10]
- QPS390 package: *Quick Pricer for S/390 Support Family*
- RLSLKSZ package: *RLS Lock Structure Sizing Tool*
- RMFTREND package: *RMF Trend Monitor*[11]
- SAFOS390 package: *System Automation for OS/390 Presentations and Infos*
- SA227403 package: *S/390 Open Systems Adapter-Express Customer's Guide*
- SG245176 package: *Introduction to FICON*
- SOFTCAP package: *Software Migration Capacity Planning Aid*
- SPSG390 package: *S/390 services from PSS - package list*
- SPSSZR package: *CP2000 S/390 Parallel Sysplex Quick-Sizer*[12]
- SWPRICER package: *S/390 PSLC Software Pricer (US Version)*
- 6942-08D package: *Operational Support - Parallel Sysplex Exploitation*
- 9037MIGR package: *Migration Planning for the 9037 Model 2 Sysplex Timer*
- 9729 Package: *IBM 9729 Optical Wavelength Division Multiplexer Timer*

C.5 Other References

If you would like to receive a weekly e-mail of IBM announcements tailored to your interest areas, you can subscribe at:
`http://www.ibm.com/isource`

Articles/White Papers

The publications listed are sorted in alphabetical order.

- *Coordination of Time-Of-Day Clocks among Multiple Systems, IBM Journal of Research and Development, Volume 36 No. 4*, G322-0181
- *Coupling Facility Configuration Options*

  To view or print this S/390 white paper:
  `http://www.s390.ibm.com/marketing/gf225042.html`
- *Five Nines/Five Minutes - Achieving Near Continuous Availability*

  To view or print this S/390 white paper:
  `http://www.s390.ibm.com/marketing/390avail.html`

9 TOOLS SENDTO WSCVM TOOLS CPSTOOLS GET PCR PACKAGE

10 TOOLS SENDTO KGNVMC TOOLSMVS MVSTOOLS GET QCBTRACE PACKAGE

11 http://www.s390.ibm.com/rmf

12 TOOLS SENDTO WSCVM TOOLS CPSTOOLS GET SPSSZR PACKAGE

- *Geographically Dispersed Parallel Sysplex: The S/390 Multi-Site Application Availability Solution*

  To view or print this S/390 white paper:
  `http://www.s390.ibm.com/ftp/marketing/position/gf225063.pdf`

- *IBM Systems Journal special issue on Parallel Sysplex Clustering Technology*

  To view or print this S/390 white paper:
  `http://www.research.ibm.com/journal/sj36-2.html`

- *IBM's Parallel Sysplex Overhead: A Reality Check*

  To view or print this S/390 white paper:
  `http://www.s390.ibm.com/ftp/marketing/position/gf225009a.pdf`

- *Network Implications of the Parallel Sysplex*

  To view or print this S/390 white paper:
  `http://www.s390.ibm.com/pso/image/syspwp.pdf`

- *OS/390 Maintenance Recommendations for Improving Availability in a Parallel Sysplex Environment*

  To view or print this S/390 white paper:
  `http://www.s390.ibm.com/marketing/psos390maint.html`

- *Improve Your Availability with Sysplex Failure Management*

  To view or print this S/390 white paper:
  `http://www.s390.ibm.com/pso/image/sfm.pdf` G321-0127

- *S/390 Parallel Sysplex Cluster Technology: IBM's Advantage*

  To view or print this S/390 white paper:
  `http://www.s390.ibm.com/ftp/marketing/position/gf225015.pdf`

Diskette kit:

- *Hardware Management Console Tutorial*, SK2T-1198 (SK2T-1198 is not orderable on its own, but is part of SK2T-5843).

Washington System Center Flashes (IBMLink/Hone)

The following flashes are available from the Washington Systems Center Web site at:
`http://www.ibm.com/support/techdocs/atsmastr.nsf`

- *Dynamic ICF Expansion*, WSC Flash 98028
- *Integrated Coupling Migration Facility (ICMF) Dispatching Enhancement*, WSC Flash 95028
- *Locating OS/390 Information on the Internet*, WSC Flash 97044
- *LPAR Management Time Performance Update*, WSC Flash 99048
- *MVS Performance Capacity Planning Considerations For 9672-Rxx Processors*, WSC Flash 95005
- *MVS/ESA Parallel Sysplex Performance LPAR Performance Considerations For Parallel Sysplex Environments*, WSC Flash 96009
- *MVS/ESA Parallel Sysplex Performance*, WSC Flash 97031
- *MVS/ESA Parallel Sysplex Performance XCF Performance Considerations*, WSC Flash 10011
- *MVS/ESA V5.1 Performance Information V5 Release To Release Migration Software Performance Impact*, WSC Flash 94041
- *OS/390 Performance Information OS/390 Increased Virtual Storage Requirements*, WSC Flash 96013

- *OS/390 R2 GRS Star APAR/PTF and Lock Structure Sizing*, WSC Flash 97015
- *Parallel Sysplex Configuration Planning for Availability*, WSC Flash 98029
- *Parallel Sysplex Interactive Tools*, WSC Flash 99054
- *Parallel Sysplex Operational Procedures: An Update*, WSC Flash 98022
- *Performance Impacts of Using Shared ICF CPs*, WSC Flash 99037
- *SFM Functions and the Impact of OW30814 for Parallel Sysplex*, WSC Flash 98025B
- *Using a Coupling Facility for the JES2 Checkpoint*, WSC Flash 98048B
- *XCF Service Recommendations to Maximize Sysplex Availability*, WSC Flash 98038

How to get IBM Redbooks

This section explains how both customers and IBM employees can find out about IBM Redbooks, redpieces, and CD-ROMs. A form for ordering books and CD-ROMs by fax or e-mail is also provided.

- **Redbooks Web Site** http://www.redbooks.ibm.com/

  Search for, view, download, or order hardcopy/CD-ROM Redbooks from the Redbooks Web site. Also read redpieces and download additional materials (code samples or diskette/CD-ROM images) from this Redbooks site.

  Redpieces are Redbooks in progress; not all Redbooks become redpieces and sometimes just a few chapters will be published this way. The intent is to get the information out much quicker than the formal publishing process allows.

- **E-mail Orders**

  Send orders by e-mail including information from the redbook fax order form to:

In United States:	e-mail address: usib6fpl@ibmmail.com
Outside North America:	Contact information is in the "How to Order" section at this site: http://www.elink.ibmlink.ibm.com/pbl/pbl/

- **Telephone Orders**

United States (toll free)	1-800-879-2755
Canada (toll free)	1-800-IBM-4YOU
Outside North America	Country coordinator phone number is in the "How to Order" section at this site: http://www.elink.ibmlink.ibm.com/pbl/pbl/

- **Fax Orders**

United States (toll free)	1-800-445-9269
Canada	1-403-267-4455
Outside North America	Fax phone number is in the "How to Order" section at this site: http://www.elink.ibmlink.ibm.com/pbl/pbl/

This information was current at the time of publication, but is continually subject to change. The latest information may be found at the Redbooks Web site.

IBM Intranet for Employees

IBM employees may register for information on workshops, residencies, and Redbooks by accessing the IBM Intranet Web site at http://w3.itso.ibm.com/ and clicking the ITSO Mailing List button. Look in the Materials repository for workshops, presentations, papers, and Web pages developed and written by the ITSO technical professionals; click the Additional Materials button. Employees may access MyNews at http://w3.ibm.com/ for redbook, residency, and workshop announcements.

IBM Redbooks fax order form

Please send me the following:

Title	Order Number	Quantity

First name Last name

Company

Address

City Postal code Country

Telephone number Telefax number VAT number

□ Invoice to customer number

□ Credit card number

Credit card expiration date Card issued to Signature

We accept American Express, Diners, Eurocard, Master Card, and Visa. Payment by credit card not available in all countries. Signature mandatory for credit card payment.

Glossary

Explanations of Cross-References:

The following cross-references are used in this glossary:

Contrast with. This refers to a term that has an opposed or substantively different meaning.

See. This refers the reader to multiple-word terms in which this term appears.

See also. This refers the reader to terms that have a related, but not synonymous meaning.

Synonym for. This indicates that the term has the same meaning as a preferred term, which is defined in the glossary.

If you do not find the term you are looking for, see the IBM Software Glossary at the URL: `http://www.networking.ibm.com/nsg/nsgmain.htm`

A

abend. Abnormal end of task.

ACF/VTAM. Advanced Communications Function for the Virtual Telecommunications Access Method. Synonym for *VTAM*.

active IRLM. The IRLM supporting the active IMS subsystem in an XRF complex.

active service policy. The service policy that determines workload management processing if the sysplex is running in goal mode. See *goal mode*.

adapter. Hardware card that allows a device, such as a PC, to communicate with another device, such as a monitor, a printer, or other I/O device. In a LAN, within a communicating device, a circuit card that, with its associated software and/or microcode, enables the device to communicate over the network.

affinity. A connection or association between two objects.

alternate IRLM. The IRLM supporting the alternate IMS subsystem in an XRF complex.

alternate site. Another site or facility, such as a commercial hot site or a customer-owned second site, that will be a recovery site in the event of a disaster.

ambiguous cursor. A database cursor that is not declared with either the clauses FOR FETCH ONLY or FOR UPDATE OF, and is not used as the target of a WHERE CURRENT OF clause on an SQL UPDATE or DELETE statement. The package processes dynamic SQL statements.

architecture. A logical structure that encompasses operating principles including services, functions, and protocols. See *computer architecture, network architecture*, and *Systems Network Architecture (SNA)*.

asynchronous. Without regular time relationship. Unexpected or unpredictable with respect to the program's instructions, or to time. Contrast with synchronous.

authorized program analysis report (APAR). A request for correction of a problem caused by a defect in a current release of a program unaltered the user.

availability. A measure of how much (often specified as a percentage) the data processing services are available to the users in a specified time frame.

B

base or basic sysplex. A base or basic sysplex is the set of one or more OS/390 systems that is given a cross-system coupling facility (XCF) name and in which the authorized programs can then use XCF coupling services. A base sysplex does not include a CF. See also *Parallel Sysplex* and *sysplex*.

basic mode. A central processor mode that does not use logical partitioning. Contrast with *logically partitioned (LPAR) mode*.

batch checkpoint/restart. The facility that enables batch processing programs to synchronize checkpoints and to be restarted at a user-specified checkpoint.

batch environment. The environment in which non-interactive programs are executed. The environment schedules their execution independently of their submitter.

batch message processing (BMP) program. An IMS batch processing program that has access to online databases and message queues. BMPs run online, but like programs in a batch environment, they are started with job control language (JCL).

batch-oriented BMP program. A BMP program that has access to online databases and message queues while performing batch-type processing. A batch-oriented BMP does not access the IMS message queues for input or output. It can access online databases, GSAM databases, and OS/390 files for both input and output.

batch processing program. An application program that has access to databases and OS/390 data management facilities but does not have access to the IMS control region or its message queues. See also *batch message processing program* and *message processing program.*

block level sharing. A kind of data sharing that enables application programs in different IMS systems to update data concurrently.

block multiplexer channel. A channel that transmits blocks of data to and from more than one device by interleaving the record blocks. Contrast with *selector channel*.

BMP program. See *batch message processing program*.

buffer. (1) A portion of storage used to hold input or output data temporarily. (2) A routine or storage used to compensate for a difference in data rate or time of occurrence of events, when transferring data from one device to another.

buffer invalidation. A technique for preventing the use of invalid data in a Parallel Sysplex data sharing environment. The technique involves marking all copies of data in DB2 or IMS buffers invalid once a sharing DBMS subsystem has updated that data.

buffer pool. A set of buffers that contains buffers of the same length. See also *buffer*, *buffer invalidation*, and *group buffer pool*.

byte multiplexer channel. A multiplexer channel that interleaves bytes of data. See also *block multiplexer channel*. Contrast with *selector channel*.

C

cache structure. A CF structure that enables high-performance sharing of cached data by multisystem applications in a sysplex. Applications can use a cache structure to implement several different types of caching systems, including a store-through or a store-in cache. As an example, DB2 uses data sharing group cache structures as GBPs. See also *group buffer pool*, *castout*, and *cache structure services*.

cache structure services. OS/390 services that enable applications in a sysplex to perform operations such as the following on a CF cache structure:

- Manage cache structure resources.
- Store data into and retrieve data from a cache structure.
- Manage accesses to shared data.
- Determine when shared data has been changed.
- Determine whether a local copy of shared data is valid.

card-on-board (COB) logic. The type of technology that uses pluggable, air-cooled cards.

CAS. Coordinating address space.

castout. The DB2 process of writing changed pages from a GBP to DASD.

catalog. A data set that contains extensive information required to locate other data sets, to allocate and deallocate storage space, to verify the access authority of a program or operator, and to accumulate data set usage statistics.

central processing unit (CPU). The part of a computer that includes the circuits that control the interpretation and execution of instructions.

central processor (CP). The part of the computer that contains the sequencing and processing facilities for instruction execution, initial program load, and other machine operations. See also *central processor complex*, *central electronic complex* and *PU*.

central processor complex (CPC). A physical collection of hardware that includes central storage, one or more CPs, timers, and channels.

central storage. Storage that is an integral part of the processor unit. Central storage includes both main storage and the hardware system area.

CF. Coupling Facility. See also *Coupling Facility*.

CFCC. Coupling Facility Control Code. See also *Coupling Facility Control Code*.

CFRM policy. A declaration regarding the allocation rules for a CF structure. See also *structure*.

channel. (1) A functional unit, controlled by a S/390 CPC that handles the transfer of data between processor storage and local peripheral equipment. (2) A path along which signals can be sent. (3) The portion of a storage medium that is accessible to a given reading or writing station. (4) In broadband transmission, a designation of a frequency band 6 MHz wide.

channel subsystem (CSS). A collection of subchannels that directs the flow of information between I/O devices and main storage, relieves the processor of communication tasks, and performs path management functions.

channel-to-channel (CTC). Refers to the communication (transfer of data) between programs on opposite sides of a channel-to-channel adapter (CTCA).

channel-to-channel adapter (CTCA). A hardware device that can be used to connect two channels on the same computing system or on different systems.

CICSPlex. (1) The largest set of CICS systems to be monitored and controlled as a single entity. (2) In a CICSPlex SM environment, the user-defined name, description, and configuration information for a CICSPlex. A CICSPlex can be made up of CICS systems or CICS system groups. See also *CICS system* and *CICS system group*.

CICSPlex SM. CICSPlex System Manager.

CICSPlex SM address space (CMAS). A CICSPlex SM component that is responsible for managing a CICSPlex. A CMAS provides the single system image for a CICSPlex by serving as the interface to other CICSPlexes and external programs. There must be at least one CMAS for each OS/390 image on which you are running CICSPlex SM. A single CMAS can manage CICS systems within one or more CICSPlexes. See also *coordinating address space (CAS)* and *managed address space (MAS)*.

CICSPlex System Manager (CICSPlex SM). An IBM CICS systems management product that provides single system image and single point of control for one or more CICSPlexes, including CICSPlexes on heterogeneous operating systems.

classification. The process of assigning a service class and, optionally, a report class to a work request. Subsystems, together with workload management services, use classification rules to assign work to a service class when it enters a sysplex.

classification rules. The rules workload management and subsystems use to assign a service class and, optionally, a report class to a work request. A classification rule consists of one or more of the following work qualifiers: subsystem type, subsystem instance, user ID, accounting information, transaction name, transaction class, source LU, NETID, and LU name.

CMAS. CICSPlex SM address space. See also *CICSPlex SM address space (CMAS)*.

CMAS link. A communications link between one CICSPlex SM address space (CMAS) and another CMAS or a remote managed address space (MAS). CMAS links are defined when CICSPlex SM is configured.

CNC. Mnemonic for an ESCON channel-attached to an ESCON-capable device.

command. An instruction that directs a control unit or device to perform an operation or a set of operations.

commit. In data processing the point at which the data updates are written to the database in a way which is irrevocable.

compatibility mode. A mode of processing in which the SRM parmlib members IEAIPSxx and IEAICSxx determine system resource management. See also *goal mode*.

complementary metal-oxide semiconductor (CMOS). A technology that combines the electrical properties of positive and negative voltage requirements to use considerably less power than other types of semiconductors.

component. (1) Hardware or software that is part of a functional unit. (2) A functional part of an operating system; for example, the scheduler or supervisor.

computer architecture. The organizational structure of a computer system, including hardware and software.

configuration. The arrangement of a computer system or network as defined by the nature, number, and chief characteristics of its functional units. More specifically, the term *configuration* may refer to a hardware configuration or a software configuration. See also *system configuration.*

connectivity. A term used to describe the physical interconnections of multiple devices/computers/networks employing similar or different technology and/or architecture together to accomplish effective communication between and among connected members. It involves data exchange and/or resource sharing.

console. A logical device that is used for communication between the user and the system. See also *service console*.

construct. A collective name for data class, storage class, management class, and storage group.

continuous availability. The elimination or masking of both planned and unplanned outages, so that no system outages are apparent to the end user.

Continuous availability can also be stated as the ability to operate 24 hours/day, 7 days/week, with no outages apparent to the end user.

continuous operations. The elimination or masking of planned outages. A system that delivers continuous operations is a system that has no scheduled outages.

control interval (CI). A fixed-length area of direct access storage in which VSAM creates distributed free space and stores records. Also, in a key-sequenced data set or file, the set of records pointed to by an entry in the sequence-set index record. The control interval is the unit of information that VSAM transmits to or from direct access storage. A control interval always comprises an integral number of physical records.

control region. The OS/390 main storage region that contains the IMS control program.

control unit. A general term for any device that provides common functions for other devices or mechanisms. Synonym for *controller*.

coordinating address space (CAS). An OS/390 subsystem that provides ISPF end-user access to the CICSPlex. There must be at least one CAS for each OS/390 image on which you are running CICSPlex SM. See also *CICSPlex SM address space (CMAS)* and *managed address space (MAS)*.

couple data set. A data set that is created through the XCF couple data set format utility and, depending on its designated type, is shared by some or all of the OS/390 systems in a sysplex. See also *Sysplex couple data set* and *XCF couple data set*.

Coupling Facility (CF). A special LP that provides high-speed caching, list processing, and locking functions in Parallel Sysplex. See also *Coupling Facility channel*, *Coupling Facility white space*, and *coupling services*.

Coupling Facility channel (CF link). A high bandwidth fiber optic channel that provides the high-speed connectivity required for data sharing between a CF and the CPCs directly attached to it.

Coupling Facility Control Code (CFCC). The Licensed Internal Code (LIC) that runs in a CF LP to provide shared storage management functions for a sysplex.

Coupling Facility Data Tables (CFDT). CFDT enables user applications, running in different CICS regions that reside in one or more OS/390 images, within a Parallel Sysplex, to share working data with update integrity.

Coupling Facility white space. CF storage set aside for rebuilding of structures from other CFs, in case of failure.

coupling services. In a sysplex, the functions of XCF that transfer data and status between members of a group residing on one or more OS/390 systems in the sysplex.

CPU service units. A measure of the task control block (TCB) execution time multiplied by an SRM constant that is CPC-model-dependent. See also *service unit*.

CP TOD. In a CPC with more than one CP, each CP can have a separate TOD clock, or more than one CP might share a clock, depending on the model. In all cases each CP has access to a single clock also called a CPC TOD clock

common queue server (CQS). A server that receives, maintains, and distributes data objects from a shared queue on a CF list structure for its clients.

cross-system coupling facility (XCF). XCF is a component of OS/390 that provides functions to support cooperation between authorized programs running within a sysplex.

cross-system extended services (XES). Provides services for OS/390 systems in a sysplex to share data on a CF.

cryptographic. Pertaining to the transformation of data to conceal its meaning.

Customer Information Control System (CICS). An IBM-licensed program that enables transactions entered at remote terminals to be processed concurrently by user-written application programs. It includes facilities for building, using, and maintaining databases.

CVC. Mnemonic for an ESCON channel-attached to a IBM 9034 (ESCON Converter).

D

daemon. A task, process, or thread that intermittently awakens to perform some chores and then goes back to sleep (software).

data entry database (DEDB). A direct-access database that consists of one or more areas, with each area containing both root segments and dependent segments. The database is accessed using VSAM media manager.

Data Facility Hierarchical Storage Manager (DFHSM). An IBM-licensed program used to back up, recover, and manage space on volumes.

Data Language/I (DL/I). The IMS data manipulation language, a common high-level interface between a user application and IMS. DL/I calls are invoked from application programs written in languages such as PL/I, COBOL, VS Pascal, C, and Ada. It can also be invoked from assembler language application programs by subroutine calls. IMS lets the user define data structures, relate structures to the application, load structures, and reorganize structures.

data link. (1) Any physical link, such as a wire or a telephone circuit, that connects one or more remote terminals to a communication control unit, or connects one communication control unit with another. (2) The assembly of parts of two data terminal equipment (DTE) devices that are controlled by a link protocol and the interconnecting data circuit, and that enable data to be transferred from a data source to a data link. (3) In SNA, see also *link*.

Note: A telecommunication line is only the physical medium of transmission. A data link includes the physical medium of transmission, the protocol, and associated devices and programs; it is both physical and logical.

data set. The major unit of data storage and retrieval, consisting of a collection of data in one of several prescribed arrangements and described by control information to which the system has access.

data sharing. In Parallel Sysplex, the ability of concurrent subsystems (such as DB2 or IMS database managers) or application programs to directly access and change the same data while maintaining data integrity. See also *Sysplex data sharing* and *data sharing group*.

data sharing group. A collection of one or more subsystems that directly access and change the same data while maintaining data integrity. See also *DB2 data sharing group* and *IMS DB data sharing group*.

data sharing Parallel Sysplex. A Parallel Sysplex where data is shared at the record level across more than one system, using a CF structure to guarentee cross-system integrity.

database. (1) A set of data, or a part or the whole of another set of data, that consists of at least one file and is sufficient for a given purpose or for a given data-processing system. (2) A collection of data fundamental to a system. See also *Database Control (DBCTL)*, *data entry database (DEDB)*, *data sharing*, and *data sharing group*.

Database Control (DBCTL). An environment allowing full-function databases and DEDBs to be accessed from one or more transaction management subsystems.

DB2 data sharing group. A collection of one or more concurrent DB2 subsystems that directly access and change the same data while maintaining data integrity.

DDF. Distributed Data Facility (DB2). DB2 subsystem running in an address space that supports VTAM communications with other DB2 subsystems and supports the execution of distributed database access requests on behalf of remote users. This provides isolation of remote function execution from local function execution.

DEDB. See *data entry database*.

delay monitoring services. The workload management services that monitor the delays encountered by a work request.

device. (1) A mechanical, electrical, or electronic contrivance with a specific purpose. (2) An input/output unit such as a terminal, display, or printer.

direct access storage device (DASD). A physical device like IBM's 3390 in which data can be permanently stored and subsequently retrieved using licensed products like IMS and DB2, or using IBM-supported access methods like VSAM in operating system environments like OS/390.

directory. A list of files that are stored on a disk or diskette. A directory also contains information about the file, such as size and date of last change.

disaster. An event that renders IT services unavailable for an extended period. Often the IT facilities must be moved to another site in the event of a disaster.

DNS. See *Domain Name System*.

domain name. In the Internet suite of protocols, a name of a host system. A domain name consists of a sequence of subnames separated by a delimiter character. For example, if the fully qualified domain name (FQDN) of host system is ralvm7.vnet.ibm.com, each of the following is a domain name: ralvm7.vnet.ibm.com, vnet.ibm.com, ibm.com.

domain name server. In the Internet suite of protocols, a server program that supplies name-to-address translation by mapping domain names to IP addresses. Synonymous with *name server*.

Domain Name System (DNS). In the Internet suite of protocols, the distributed database system used to map domain names to IP addresses.

dynamic. Pertaining to an operation that occurs at the time it is needed rather than at a predetermined or fixed time. See also *dynamic connection*, *dynamic connectivity*, *dynamic reconfiguration*, *dynamic reconfiguration management*, and *dynamic storage connectivity*.

dynamic CF dispatching. With dynamic CF dispatching, the CF will monitor the request rate that is driving it and adjust its usage of CP resource accordingly. If the request rate becomes high enough, the CF will revert back to its original dispatching algorithm, constantly looking for new work. When the request rate lowers, the CF again becomes more judicious with its use of CP resource. See also *dynamic ICF expansion*.

dynamic connection. In an ESCON director, a connection between two ports, established or removed by the ESCD and that, when active, appears as one continuous link. The duration of the connection depends on the protocol defined for the frames transmitted through the ports and on the state of the ports.

dynamic connectivity. In an ESCON director, the capability that allows connections to be established and removed at any time.

dynamic ICF expansion. Dynamic ICF expansion provides the ability for a CF LP that is using a dedicated ICF to expand into the pool of shared ICFs or shared CPs. At low request rates, the resource consumption of the shared PU should be 1% to 2%. As the request rate increases, the resource consumption will increase, up to the point where the LP will consume its full share of the shared PU as defined by the LPAR weights. See also *dynamic CF dispatching*.

dynamic reconfiguration. Pertaining to a processor reconfiguration between a single-image (SI) configuration and a physically partitioned (PP) configuration when the system control program is active.

dynamic reconfiguration management. In OS/390, the ability to modify the I/O configuration definition without needing to perform a power-on reset (POR) of the hardware or an initial program load (IPL).

dynamic storage reconfiguration. A PR/SM LPAR function that allows central or expanded storage to be added or removed from a logical partition without disrupting the system control program operating in the logical partition.

E

ECS. Enhanced Catalog Sharing (DFSMS/MVS V1.5)

EMIF. Enhanced multiple image facility (formerly ESCON multiple image facility). A facility which allows the sharing of FICON or ESCON channels between LPs.

emitter. In fiber optics, the source of optical power.

end node. A type 2.1 node that does not provide any intermediate routing or session services to any other node. For example, APPC/PC is an end node.

enhanced catalog sharing. By using a CF cache structure instead of DASD to store catalog sharing control information, shared catalog performance in sysplex environment is improved. This sharing method, called enhanced catalog sharing (ECS), eliminates a reserve, dequeue, and I/O request to the VVDS on most catalog calls.

Enhanced Multiple Image Facility (EMIF). See EMIF.

enhanced sysplex. An enhanced sysplex is a sysplex with one or more CFs. See also *base sysplex* and *sysplex*.

enterprise. A business or organization that consists of two or more sites separated by a public right-of-way or a geographical distance.

Enterprise Systems Connection (ESCON). A set of products and services that provides a dynamically connected environment using optical cables as a transmission medium. See also *ESCD*, *ESCM*, and *ESCON channel*.

Environmental Services Subsystem (ESSS). A component of CICSPlex SM that owns all the data spaces used by the product in an OS/390 image. The ESSS executes at initialization and remains in the OS/390 image for the life of the IPL to ensure that the data spaces can survive the loss of a CICSPlex SM address space (CMAS).

ESA/390. Enterprise Systems Architecture/390.

ESCD. Enterprise Systems Connection (ESCON) Director. See also *ESCD console*, *ESCD console adapter*, and *ESCM*.

ESCD console. The ESCON director input/output device used to perform operator and service tasks at the ESCD.

ESCD console adapter. Hardware in the ESCON director console that provides the attachment capability between the ESCD and the ESCD console.

ESCM. See *ESCON Manager*.

ESCON channel. A channel having an Enterprise Systems Connection channel-to-control-unit I/O interface that uses optical cables as a transmission medium. Contrast with *parallel channel*.

ESCON director (ESCD). A device that provides connectivity capability and control for attaching any two links to each other.

ESCON Extended Distance Feature (ESCON XDF). An ESCON feature that uses laser/single-mode fiber optic technology to extend unrepeated link distances up to 20 km. LPs in an ESCON environment.

ESCON Manager (ESCM). A licensed program that provides S/390 CPC control and intersystem communication capability for ESCON director connectivity operations.

ESCON multiple image facility (EMIF). A facility that allows channels to be shared among PR/SM logical partitions in an ESCON environment.

ESCON XDF. ESCON extended distance feature.

Ethernet. A local area network that was originally marketed by Xero Corp. The name is a trademark of Xerox Corp.

ETR. See *External Time Reference*.

ETR offset. The time zone offset identifies your system location within a network of other systems. This offset is the difference between your local time and Universal Time Coordinate (UTC). See also *Universal Time Coordinate*.

ETS. See *External Time Source*.

exclusive lock. A lock that prevents concurrently executing application processes from reading or changing data. Contrast with *shared lock.*

expanded storage. (1) Optional integrated high-speed storage that transfers 4KB pages to and from central storage. (2) Additional (optional) storage that is addressable by the system control program.

extended Parallel Sysplex. This name is sometimes used to refer to Parallel Sysplexes that exploit data sharing and future enhancements for ultra high availability and disaster recovery.

Extended Recovery Facility (XRF). Software designed to minimize the effect of failures in OS/390, VTAM, the S/390 CPC CP, or IMS/VS on sessions between IMS/VS and designated terminals. It provides an alternate subsystem to take over failing sessions.

External Time Reference (ETR). This is how OS/390 documentation refers to the 9037 Sysplex Timer. An ETR consists of one or two 9037s and their associated consoles.

External Time Source (ETS). An accurate time source used to set the time in the Sysplex Timer. The accurate time can be obtained by dialing time services or attaching to radio receivers or time code generators.

F

false lock contention. A contention indication from the CF when multiple lock names are hashed to the same indicator and when there is no real contention.

Fast Path. IMS functions for applications that require good response characteristics and that may have large transaction volumes. Programs have rapid access to main-storage databases (to the field level), and to direct-access data entry databases. Message processing is grouped for load balancing and synchronized for database integrity and recovery. See also *MSDB* and *DEDB.*

Fast Path databases. Two types of databases designed to provide high data availability and fast processing for IMS applications. They can be processed by the following types of programs: MPPs, BMPs, and IFPs. See also *main storage database* and *data entry database.*

feature. A part of an IBM product that can be ordered separately by the customer.

FICON channel. FIbre CONnection. A S/390 channel which uses industry standard Fibre Channel Standard (FCS) as a base.

file system. The collection of files and file management structures on a physical or logical mass storage device such as a disk.

format. (1) A specified arrangement of things, such as characters, fields, and lines, usually used for displays, printouts, or files. (2) To arrange things such as characters, fields, and lines.

forward recovery. Reconstructing a file or database by applying changes to an older version (backup or image copy) with data recorded in a log data set. The sequence of changes to the restored copy is in the same order in which they were originally made.

frame. For an S/390 microprocessor cluster, a frame may contain one or more CPCs, support elements, and AC power distribution.

frequency. The rate of signal oscillation, expressed in hertz (cycles per second).

full function databases. Hierarchic databases that are accessed through Data Language I (DL/1) call language and can be processed by all four types of application programs: IFP, MPPs, BMPs, and batch. Full function databases include HDAM, HIDAM, HSAM, HISAM, SHSAM, and SHISAM.

G

generic resource name. A name used by VTAM that represents several application programs that provide the same function in order to handle session distribution and balancing in a sysplex.

gigabytes. One billion (10^9) bytes.

global locking (DB2). For data consistency in a data sharing environment, locks must be known and respected between all members. DB2 data sharing uses global locks ensure that each member is aware of all members' locks.

Two locking mechanisms are used by DB2 data sharing to ensure data consistency, logical locks and physical locks.

The two types can be briefly compared as follows:

1. Logical locks

   Logical locks are used to control concurrent access from application processes, such as transactions or batch programs.

2. Physical locks

   Physical locks are used by DB2 members to control physical resourses
   - Page set physical locks are used to track the level of interest in a particular page set or partition and thus determine the nee GBP coherency controls.
   - Page physical locks are used to preserve the physical consistency of pages.

See also *P-lock*.

global resource serialization (GRS). A component of OS/390 used for sharing system resources and for converting DASD reserve volumes to data set ENQueues.

global resource serialization complex (GRSplex). One or more OS/390 systems that use global resource serialization to serialize access to shared resources (such as data sets on shared DASD volumes).

GMT. See *Greenwich Mean Time*.

goal mode. A mode of processing where the active service policy determines system resource management. See also *compatibility mode*.

Greenwich Mean Time (GMT). Time at the time zone centered around Greenwich, England.

group buffer pool. A CF cache structure used by a DB2 data sharing group to cache data and to ensure that the data is consistent for all members. See also *buffer pool*.

group services. Services for establishing connectivity among the multiple instances of a program, application, or subsystem (members of a group running on OS/390) in a sysplex. Group services allow members of the group to coordinate and monitor their status across the systems of a sysplex.

H

Hardware Management Console. A console used to monitor and control hardware such as the 9672 CPCs.

hardware system area (HSA). A logical area of central storage, not addressable by application programs, used to store Licensed Internal Code and control information.

highly parallel. Refers to multiple systems operating in parallel, each of which can have multiple processors. See also *n-way*.

high-speed buffer. A cache or a set of logically partitioned blocks that provides significantly faster access to instructions and data than that provided by central storage.

HiPerLink. A HiPerLink provides improved CF link efficiency and response times in processing CF requests, compared to previous CF link configurations. With HiPerLinks, current data sharing overheads are reduced and CF link capacity is improved.

host (computer). (1) In a computer network, a computer that provides end users with services such as computation and databases and that usually performs network control functions. (2) The primary or controlling computer in a multiple-computer installation.

HSA. See *hardware system area*.

I

IBF. See *Internal Battery Feature*.

IC. See *Internal Coupling Link*.

ICB. See *Integrated Cluster Bus*.

ICF. See *Internal Coupling Facility*.

ICF. Integrated Catalog Facility.

importance level. An attribute of a service class goal that indicates the importance of meeting the goal relative to other service class goals, in five levels: lowest, low, medium, high, and highest.

IMS DB data sharing group. A collection of one or more concurrent IMS DB subsystems that directly access and change the same data while maintaining data integrity. The components in an IMS DB data sharing group include the sharing IMS subsystems, the IRLMs they use, the IRLM, OSAM, and VSAM structures in the CF, and a single set of DBRC RECONs.

IMS system log. A single log made up of online data sets (OLDSs) and write-ahead data sets (WADSs).

in-doubt period. The period during which a unit of work is pending during commit processing that involves two or more subsystems. See also *in-doubt work unit*.

in-doubt work unit. In CICS/ESA and IMS/ESA, a piece of work that is pending during commit processing; if commit processing fails between the polling of subsystems and the decision to execute the commit, recovery processing must resolve the status of any work unit that is in doubt.

indirect CMAS. A CICSPlex SM address space (CMAS) that the local CMAS can communicate with through an adjacent CMAS. There is no direct CMAS-to-CMAS link between the local CMAS and an indirect CMAS. Contrast with *adjacent CMAS*. See also *local CMAS*.

initial microcode load (IML). The action of loading the operational microcode.

initial program load (IPL). The initialization procedure that causes an operating system to start operation.

input/output support processor (IOSP). The hardware unit that provides I/O support functions for the primary support processor (PSP). It also provides maintenance support function for the processor controller element (PCE).

installed service definition. The service definition residing in the couple data set for WLM. The installed service definition contains the active service policy information.

interactive. Pertaining to a program or system that alternately accepts input and then responds. An interactive system is conversational; that is, a continuous dialog exists between user and system. Contrast with *batch*.

interface. A shared boundary. An interface might be a hardware component to link two devices or it might be a portion of storage or registers accessed by two or more computer programs.

Integrated Cluster Bus channel (ICB). The Integrated Cluster Bus channel uses the Self Timed Interface to perform the S/390 coupling communication. The cost of coupling is reduced by using a higher performing (Approximately 280 MB/sec) but less complex transport link suitable for the relatively short distances (The cable is 10 meters; the distance between CPCs is approximately 7 meters).

Integrated Offload Processor (IOP). The processor in the interconnect communication element that detects, initializes, and ends all channel subsystem operations.

Integrated Coupling Migration Facility (ICMF). A PR/SM LPAR facility that emulates CF links for LPs (CF LPs and OS/390 LPs) running on the same CPC to assist in the test and development of data sharing applications.

internal battery feature (IBF). The internal battery feature (IBF) provides the function of a local uninterruptible power source (UPS). This feature may increase power line disturbance immunity for S/390 CPCs.

Internal Coupling channel (IC). The Internal Coupling channel emulates the coupling facility functions in microcode between images within a single CPC. It is a high performance channel transferring data at up to 6Gb/sec. Internal Coupling implementation is a totally logical channel requiring no channel or even cable hardware. However, a CHPID number must be defined in the IOCDS. A replacement for ICMF.

Internal Coupling Facility (ICF). The Internal Coupling Facility (ICF) uses up to two spare PUs on selected S/390 CPCs. The ICF may use CF links or emulated links (ICMF). It can be used initially as an entry configuration into Parallel Sysplex and then maintained as a backup configuration in the future.

interrupt. (1) A suspension of a process, such as execution of a computer program caused by an external event, and performed in such a way that the process can be resumed. (2) To stop a process in such a way that it can be resumed. (3) In data communication, to take an action at a receiving station that causes the sending station to end a transmission. (4) To temporarily stop a process.

invalidation. The process of removing records from cache because of a change in status of a subsystem facility or function, or because of an error while processing the cache image of the set of records. When such a cache image is invalidated, the corresponding records cannot be accessed in cache and the assigned cache space is available for allocation.

IOCDS. I/O configuration data set.

IOCP. I/O configuration program.

I/O service units. A measure of individual data set I/O activity and JES spool reads and writes for all data sets associated with an address space.

J

JES (Job Entry Subsystem). A system facility for spooling, job queuing, and managing I/O.

jumper cable. In an ESCON environment, an optical cable, having two conductors, that provides physical attachment between two devices or between a device and a distribution panel. Contrast with *trunk cable*.

L

latency. The time interval between the instant at which an instruction control unit initiates a call for data and the instant at which the actual transfer of data starts.

leap second. Corrections of exactly one second inserted into the UTC time scale since January 1, 1972. This adjustment occurs at the end of a UTC month, normally on June 30 or December 31. Seconds are occasionally added to or subtracted from the UTC to compensate for the wandering of the earth's polar axis and maintain agreement with the length of the solar day. See also *Universal Time Coordinate(UTC)*.

LIC. See *Licensed Internal Code*.

Licensed Internal Code (LIC). Software provided for use on specific IBM machines and licensed to customers under the terms of IBM's Customer Agreement. Microcode can be Licensed Internal Code and licensed as such.

link. The combination of physical media, protocols, and programming that connects devices.

list structure. A CF structure that enables multisystem applications in a sysplex to share information organized as a set of lists or queues. A list structure consists of a set of lists and an optional lock table, which can be used for serializing resources in the list structure. Each list consists of a queue of list entries.

list structure services. OS/390 services that enable multisystem applications in a sysplex to perform operations such as the following on a CF list structure:

- Read, update, create, delete, and move list entries in a list structure.
- Perform serialized updates on multiple list entries in a list structure.
- Monitor lists in a list structure for transitions from empty to non-empty.

local cache. A buffer in local system storage that might contain copies of data entries in a CF cache structure.

local CMAS. The CICSPlex SM address space (CMAS) that a user identifies as the current context when performing CMAS configuration and management tasks.

local MAS. A managed address space (MAS) that resides in the same OS/390 image as the CICSPlex SM address space (CMAS) that controls it and that uses the Environmental Services Subsystem (ESSS) to communicate with the CMAS.

lock resource. Data accessed through a CF structure.

lock structure. A CF structure that enables applications in a sysplex to implement customized locking protocols for serialization of application-defined resources. The lock structure supports shared, exclusive, and application-defined lock states, as well as generalized contention management and recovery protocols. See also *exclusive lock, shared lock,* and *false lock contention*.

lock structure services. OS/390 services that enable applications in a sysplex to perform operations such as the following on a CF lock structure:

- Request ownership of a lock.
- Change the type of ownership for a lock.
- Release ownership of a lock.
- Manage contention for a lock.
- Recover a lock held by a failed application.

logical connection. In a network, devices that can communicate or work with one another because they share the same protocol.

logical control unit. A group of contiguous words in the HSA that provides all of the information necessary to control I/O operations through a group of paths that are defined in the IOCDS. Logical control units represent to the channel subsystem a set of control units that attach common I/O devices.

logical partition (LP). In LPAR mode, a subset of the processor unit resources that is defined to support the operation of a system control program (SCP). See also *logically partitioned (LPAR) mode*.

logical unit (LU). In *VTAM*, the source and recipient of data transmissions. Data is transmitted from one logical unit (LU) to another LU. For example, a terminal can be an LU, or a CICS or IMS system can be an LU.

logically partitioned (LPAR) mode. A CPC power-on reset mode that enables use of the PR/SM feature and allows an operator to allocate CPC hardware resources (including CPs, central storage, expanded storage, and channel paths) among logical partitions. Contrast with *basic mode*.

loosely coupled. A multisystem structure that requires a low degree of interaction and cooperation between multiple OS/390 images to process a workload. See also *tightly coupled*.

LP. See *logical partition*.

LPAR. See *logically partitioned (LPAR) mode*.

LU. See *logical unit.*

M

m-image. The number (m) of OS/390 images in a sysplex. See also *n-way*.

main storage. A logical entity that represents the program addressable portion of central storage. All user programs are executed in main storage. See also *central storage.*

main storage database (MSDB). A root-segment database, residing in main storage, which can be accessed to a field level.

mainframe (S/390 CPC). A large computer, in particular one to which other computers can be connected so that they can share facilities the S/390 CPC provides; for example, an S/390 computing system to which personal computers are attached so that they can upload and download programs and data.

maintenance point. A CICSPlex SM address space (CMAS) that is responsible for maintaining CICSPlex SM definitions in its data repository and distributing them to other CMASs involved in the management of a CICSPlex.

managed address space (MAS). A CICS system that is being managed by CICSPlex SM. See also *local MAS* and *remote MAS*.

MAS. Managed address space.

MAS agent. A CICSPlex SM component that acts within a CICS system to provide monitoring and data collection for the CICSPlex SM address space (CMAS). The level of service provided by a MAS agent depends on the level of CICS the system is running under and whether it is a local or remote MAS. See also *CICSPlex SM address space (CMAS), local MAS,* and *remote MAS*.

massively parallel. Refers to thousands of processors in a parallel arrangement.

mega-microsecond. A carry out of bit 32 of the TOD clock occurs every 2^{20} microseconds (1.048576 seconds). This interval is sometimes called a "mega-microsecond" (Mµs). This carry signal is used to start one clock in synchronism with another, as part of the process of setting the clocks. See also *time-of-day clock*.

member. A specific function (one or more modules or routines) of a multisystem application that is defined to XCF and assigned to a group by the multisystem application. A member resides on one system in the sysplex and can use XCF services to communicate (send and receive data) with other members of the same group. See *XCF group*, and *multisystem application*.

memory. Program-addressable storage from which instructions and other data can be loaded directly into registers for subsequent execution or processing. Synonymous with *main storage*.

microcode. (1) One or more microinstructions. (2) A code, representing the instructions of an instruction set, that is implemented in a part of storage that is not program-addressable. (3) To design, write, and test one or more microinstructions.

microprocessor. A processor implemented on one or a small number of chips.

migration. Installing a new version or release of a program when an earlier version or release is already in place.

mixed complex. A global resource serialization complex in which one or more of the systems in the global resource serialization complex are not part of a multisystem sysplex.

monitoring environment. A record of execution delay information about work requests kept by the workload management services. A monitoring environment is made up of one or more performance blocks. See also *performance block*.

monoplex. A one system sysplex with sysplex couple data sets that XCF prevents any other system from joining. See also *multisystem sysplex*.

MP. Multiprocessor.

MSDB. See *main storage database.*

MSU. Millions of Service Units. The unit used in IBM PSLC pricing as an estimate of CPC capacity. It is used in this book as an estimate of CPC capacity for CPC and 9674 capacity planning purposes.

multifiber cable. An optical cable that contains two or more fibers. See also *jumper cable, optical cable assembly,* and *trunk cable*.

multimode optical fiber. A graded-index or step-index optical fiber that allows more than one bound mode to propagate. Contrast with *single-mode optical fiber*.

Multi-Node Persistent Session (MNPS). MNPS extends persistent sessions capability across multiple CPCs connected through the CF. MNPS provides for the recovery of VTAM, OS/390, hardware or application failures by restarting the application on another host in the Parallel Sysplex without requiring users to re-logon.

Multiple Systems Coupling (MSC). An IMS facility that permits multiple IMS subsystems to communicate with each other.

multiprocessing. The simultaneous execution of two or more computer programs or sequences of instructions. See also *parallel processing*.

multiprocessor (MP). A CPC that can be physically partitioned to form two operating processor complexes.

multisystem application. An application program that has various functions distributed across OS/390 images in a multisystem environment.

Examples of multisystem applications are:

- CICS
- Global resource serialization (GRS)
- Resource Measurement Facility (RMF)
- OS/390 Security Server (RACF)
- Workload manager (WLM)

See *XCF group*.

multisystem environment. An environment in which two or more OS/390 images reside in one or more processors, and programs on one image can communicate with programs on the other images.

multisystem sysplex. A sysplex in which two or more OS/390 images are allowed to be initialized as part of the sysplex. See also *single-system sysplex*.

Mμs. See *mega-microsecond*.

N

named counter server (CICS). CICS provides a facility for generating unique sequence numbers for use by applications in a Parallel Sysplex environment (for example, to allocate a unique number for orders or invoices). This facility is provided by a named counter server, which maintains each sequence of numbers as a named counter. Each time a sequence number is assigned, the corresponding named counter is incremented automatically so that the next request gets the next number in sequence. This facility uses a CF list structure to hold the information.

NCP. (1) Network Control Program (IBM-licensed program). Its full name is Advanced Communications Function for the Network Control Program. Synonymous with *ACF/NCP*. (2) Network control program (general term).

network. A configuration of data processing devices and software connected for information interchange. See also *network architecture* and *network control program (NCP)*.

network architecture. The logical structure and operating principles of a computer network.

node. (1) In SNA, an endpoint of a link or junction common to two or more links in a network. Nodes can be distributed to S/390 CPC CPs, communication controllers, cluster controllers, or terminals. Nodes can vary in routing and other functional capabilities.

n-way. The number (n) of CPs in a CPC. For example, a 6-way CPC contains six CPs.

O

OLDS. *See* online log data set.

online log data set (OLDS). A data set on direct access storage that contains the log records written by an online IMS system.

open system. (1) A system with specified standards and that therefore can be readily connected to other systems that comply with the same standards. (2) A data communications system that conforms to the standards and protocols defined by open systems interconnection (OSI). Synonym for *node*.

Operational Single Image. Multiple operating system images being managed as a single entity. This may be a basic sysplex, standard Parallel Sysplex or extended Parallel Sysplex.

optical cable. A fiber, multiple fibers, or a fiber bundle in a structure built to meet optical, mechanical, and environmental specifications. See also *jumper cable* and *trunk cable*.

optical receiver. Hardware that converts an optical signal to an electrical logic signal. Contrast with *optical transmitter*.

optical repeater. In an optical fiber communication system, an opto-electronic device or module that receives a signal, amplifies it (or, for a digital signal, reshapes, retimes, or otherwise reconstructs it) and retransmits it.

optical transmitter. Hardware that converts an electrical logic signal to an optical signal. Contrast with *optical receiver*.

OS/390 image. A single occurrence of the OS/390 operating system that has the ability to process work.

OS/390 system. An OS/390 image together with its associated hardware, which collectively are often referred to simply as a system, or OS/390 system.

P

P-lock. There are times when a P-lock must be obtained on a page to preserve physical consistency of the data between members. These locks are known as page P-locks. Page P-locks are used, for example, when two subsystems attempt to update the same page of data and row locking is in effect. They are also used for GBP-dependent space map pages and GBP-dependent leaf pages for type 2 indexes, regardless of locking level. IRLM P-locks apply to both DB2 and IMS DB data sharing.

Page set P-locks are used to track inter-DB2 read-write interest, thereby determining when a page set has to become GBP-dependent. When access is required to a page set or partition through a member in a data sharing group, a page set P-lock is taken. This lock is always propagated to the lock table on the CF and is owned by the member. No matter how many times the resource is accessed through the member, there will always be only one page set P-lock for that resource for a particular member. This lock will have different modes depending on the level (read or write) of interest the member has in the resource. See also *global locking*.

parallel. (1) Pertaining to a process in which all events occur within the same interval of time, each handled by a separate but similar functional unit; for example, the parallel transmission of the bits of a computer word along the lines of an internal bus. (2) Pertaining to the concurrent or simultaneous operation of two or more devices or to concurrent performance of two or more activities in a single device. (3) Pertaining to the concurrent or simultaneous occurrence of two or more related activities in multiple devices or channels. (4) Pertaining to the simultaneity of two or more processes. (5) Pertaining to the simultaneous processing of the individual parts of a whole, such as the bits of a character and the characters of a word, using separate facilities for the various parts. (6) Contrast with *serial*.

parallel processing. The simultaneous processing of units of work by many servers. The units of work can be either transactions or subdivisions of large units of work (batch).

Parallel Sysplex. A Parallel Sysplex is a sysplex with one or more CFs. See also *base sysplex*, *sysplex*, *extended Parallel Sysplex*, and *standard Parallel Sysplex*.

partition. An area of storage on a fixed disk that contains a particular operating system or logical drives where data and programs can be stored.

partitionable CPC. A CPC can be divided into two independent CPCs. See also *physical partition, single-image mode, MP*, and *side*.

partitioned data set (PDS). A data set in DASD storage that is divided into partitions, called *members*, each of which can contain a program, part of a program, or data.

performance. For a storage subsystem, a measurement of effective data processing speed against the amount of resource that is consumed by a complex. Performance is largely determined by throughput, response time, and system availability. See also *performance administration*, *performance block*, *performance management*, and *performance period*.

performance administration. The process of defining and adjusting workload management goals and resource groups based on installation business objectives.

performance block. A piece of storage containing workload management's record of execution delay information about work requests.

performance management. The process workload management uses to decide how to match resources to work according to performance goals and processing capacity.

performance period. A service goal and importance level assigned to a service class for a specific duration. You define performance periods for work that has variable resource requirements.

persistent connection. A connection to a CF structure with a connection disposition of KEEP. OS/390 maintains information about the connection so that when the connection terminates abnormally from a CF structure, OS/390 places the connection in a failed-persistent state, and the connection can attempt to reconnect to the structure.

persistent session. (1) In the NetView program, a network management session that remains active even though there is no activity on the session for a specified period of time. (2) An LU-LU session that VTAM retains after the failure of a VTAM application program. Following the application program's recovery, the application program either restores or terminates the session.

persistent structure. A structure allocated in the CF with a structure disposition of KEEP. A persistent structure keeps its data intact across system or sysplex outages, regardless of whether any users are connected to the structure.

physical partition. Part of a CPC that operates as a CPC in its own right, with its own copy of the operating system.

physically partitioned (PP) configuration. A system configuration that allows the processor controller to use both CPC sides as individual CPCs. The A-side of the processor controls side 0, and the B-side controls side 1. Contrast with *single-image (SI) mode*.

policy. A set of installation-defined rules for managing sysplex resources. The XCF PR/SM policy and sysplex failure management policy are examples of policies.

power-on reset. The state of the machine after a logical power-on before the control program is IPLed.

preference list. An installation list of CFs, in priority order, that indicates where OS/390 is to allocate a structure.

processing unit (PU). The part of the system that does the processing, and contains processor storage. On a 9672 CPC the PU may be assigned as either a CP, SAP, ICF or act as a spare PU.

processor. A processing unit, capable of executing instructions when combined with main storage and channels. See also *processor complex*, *processor controller*, and *processor controller element (PCE* and *CPC)*.

processor complex. A physical collection of hardware that includes main storage, one or more processors, and channels.

processor controller. Hardware that provides support and diagnostic functions for the CPs.

processor controller element (PCE). Hardware that provides support and diagnostic functions for the processor unit. The processor controller communicates with the processor unit through the logic service adapter and the logic support stations, and with the power supplies through the power thermal controller. It includes the primary support processor (PSP), the initial power controller (IPC), the input/output support processor (IOSP), and the control panel assembly.

Processor Resource/Systems Manager (PR/SM). A function that allows the processor unit to operate several system control programs simultaneously in LPAR mode. It provides for logical partitioning of the real machine and support of multiple preferred guests. See also *LPAR.*

program specification block (PSB). The control block in IMS that describes databases and logical message destinations used by an application program.

PR/SM. See *Processor Resource/Systems Manager*.

PSB. See *program specification block.*

public network. A communication common carrier network that provides data communication services over switched, non-switched, or packet-switching lines.

R

RAS. Reliability, availability, and serviceability.

receiver. In fiber optics, see *optical receiver.*

reconfiguration. (1) A change made to a given configuration in a computer system; for example, isolating and bypassing a defective functional unit or connecting two functional units by an alternative path. Reconfiguration is effected automatically or manually and can be used to maintain system integrity.
(2) The process of placing a processor unit, main storage, and channels offline for maintenance, and adding or removing components.

Record Level Sharing (RLS). RLS is an access mode for VSAM data sets supported by DFSMS 1.3 and later releases. RLS enables VSAM data to be shared, with full update capability, between many applications running in many CICS regions across the Parallel Sysplex.

recovery. To maintain or regain system operation after a failure occurs. Generally, to recover from a failure is to identify the failed hardware, to de-configure the failed hardware, and to continue or restart processing.

recovery control (RECON) data sets. Data sets in which Database Recovery Control stores information about logging activity and events that might affect the recovery of databases.

relative processor power (RPP). A unit used to express processor capacity. RPP is a measured average of well defined workload profiles ITR-ratios. ITR (Internal Throughput Rate) is measured in transactions/CPU second. LSPR (Large Systems Performance Reference) measurements predict RPP values for processors running certain releases of operating systems.

remote MAS. A managed address space (MAS) that uses MRO or LU6.2 to communicate with the CICSPlex SM address space (CMAS) that controls it. A remote MAS may or may not reside in the same OS/390 image as the CMAS that controls it.

remote operations. The ability to perform operations tasks from a remote location.

remote site recovery. The ability to continue or resume processing of the critical workload from a remote site.

report class. A group of work for which reporting information is collected separately. For example, you can have a WLM report class for information combining two different service classes, or a report class for information on a single transaction.

request. A service primitive issued by a service user to call a function supported by the service provider.

request for price quotation (RPQ). A custom feature for a product.

resource group. An amount of processing capacity across one or more OS/390 images, assigned to one or more WLM service classes.

Resource Sharing. S/390 Resource Sharing provides the following functionality:

- XCF Signalling - providing multisystem signaling with reduced cost/management
- GRS Star - multisystem resource serialization for increased performance, recoverability and scalability
- JES Checkpointing - multisystem checkpointing for increased simplicity and reduced cost
- Shared Tape - multisystem tape sharing for reduced duplication cost
- Merged Operations Log - multisystem log for single system image/management
- Merged LOGREC - multisystem log for single system image/management

- Shared Catalog - multisystem shared master catalogs/user catalogs for increased performance/simplicity and reduced cost

response time. The amount of time it takes after a user presses the enter key at the terminal until the reply appears at the terminal.

routing. The assignment of the path by which a message will reach its destination.

RPP. See *relative processor power*.

RPQ. See *request for price quotation*.

S

secondary host promotion. Secondary host promotion allows one DFSMShsm system to automatically assume the unique functions of another DFSMShsm system that has failed.

selector channel. An I/O channel that operates with only one I/O device at a time. Once the I/O device is selected, a complete record is transferred one byte at a time. Contrast with *block multiplexer channel.*

serial. (1) Pertaining to a process in which all events occur one after the other; for example, serial transmission of the bits of a character according to V24 CCITT protocol. (2) Pertaining to the sequential or consecutive occurrence of two or more related activities in a single device or channel. (3) Pertaining to the sequential processing of the individual parts of a whole, such as the bits of a character or the characters of a word, using the same facilities for successive parts. (4) Contrast with *parallel*.

serialized list structure. A CF list structure with a lock table containing an array of exclusive locks whose purpose and scope are application-defined. Applications can use the lock table to serialize on parts of the list structure, or resources outside the list structure.

server. A device, program, or code module on for example a network dedicated to a specific function.

server address space. Any address space that helps process work requests.

service administration application. The online ISPF application used by the service administrator to specify the workload management service definition.

service class. A subset of a workload having the same service goals or performance objectives, resource requirements, or availability requirements. For workload management, you assign a service goal and optionally a resource group to a service class.

service console. A logical device used by service representatives to maintain the processor unit and to isolate failing field replaceable units. The service console can be assigned to any of the physical displays attached to the input/output support processor.

service definition. An explicit definition of the workloads and processing capacity in an installation. A service definition includes workloads, service classes, systems, resource groups, service policies, and classification rules.

service definition coefficient. A value that specifies which type of resource consumption should be emphasized in the calculation of service rate. The types of resource consumption are CPU, IOC, MSO, and SRB.

service policy. A named set of performance goals and, optionally, processing capacity boundaries that workload management uses as a guideline to match resources to work. See also *active service policy*.

service request block (SRB) service units. A measure of the SRB execution time for both local and global SRBs, multiplied by an SRM constant that is CPU model dependent.

service unit. The amount of service consumed by a work request as calculated by service definition coefficients and CPU, SRB, I/O, and storage service units.

session. (1) A connection between two application programs that allows them to communicate. (2) In SNA, a logical connection between two network addressable units that can be activated, tailored to provide various protocols, and deactivated as requested. (3) The data transport connection resulting from a call or link between two devices. (4) The period of time during which a user of a node can communicate with an interactive system; usually it is the elapsed time between logon and logoff. (5) In network architecture, an association of facilities necessary for establishing, maintaining, and releasing connections for communication between stations.

shared. Pertaining to the availability of a resource to more than one use at the same time.

shared lock. A lock that prevents concurrently executing application processes from changing, but not from reading, data. Contrast with *exclusive lock.*

side. A part of a partitionable PC that can run as a physical partition and is typically referred to as the A-side or the B-side.

single-image (SI) mode. A mode of operation for a multiprocessor (MP) system that allows it to function as one CPC. By definition, a uniprocessor (UP) operates in single-image mode. Contrast with *physically partitioned (PP) configuration*.

single-mode optical fiber. An optical fiber in which only the lowest-order bound mode (which can consist of a pair of orthogonally polarized fields) can propagate at the wavelength of interest. Contrast with *multimode optical fiber*.

single-OS/390 environment. An environment that supports one OS/390-image. See also *OS/390 image*.

single point of control. The characteristic a sysplex displays when you can accomplish a given set of tasks from a single workstation, even if you need multiple IBM and vendor products to accomplish that particular set of tasks.

single point of failure. An essential resource for which there is no backup.

single GRSplex serialization. Single GRSplex serialization allows several HSMplexes, within a single GRSplex, to operate without interfering with any other HSMplex.

single-system image. The characteristic a product displays when multiple images of the product can be viewed and managed as one image.

single-system sysplex. A sysplex in which only one OS/390 system is allowed to be initialized as part of the sysplex. In a single-system sysplex, XCF provides XCF services on the system but does not provide signalling services between OS/390 systems. See also *multisystem complex* and *XCF-local mode.*

SMS communication data set (COMMDS). The primary means of communication among systems governed by a single SMS configuration. The SMS communication data set (COMMDS) is a VSAM linear data set that contains the current utilization statistics for each system-managed volume. SMS uses these statistics to help balance space usage among systems.

SMS configuration. The SMS definitions and routines that the SMS subsystem uses to manage storage.

SMS system group. All systems in a sysplex that share the same SMS configuration and communications data sets, minus any systems in the sysplex that are defined individually in the SMS configuration.

SSP. See *system support programs*.

standard. Something established by authority, custom, or general consent as a model or example.

standard Parallel Sysplex. A non-data sharing Parallel Sysplex.

STI. Self-Timed Interconnect.

storage. A unit into which recorded data can be entered, in which it can be retained and processed, and from which it can be retrieved.

storage management subsystem (SMS). An operating environment that helps automate and centralize the management of storage. To manage storage, SMS provides the storage administrator with control over data class, storage class, management class, storage group, and ACS routine definitions.

structure. A construct used to map and manage storage in a CF. See *cache structure, list structure*, and *lock structure.*

subarea. A portion of the SNA network consisting of a subarea node, any attached peripheral nodes, and their associated resources. Within a subarea node, all network addressable units, links, and adjacent link stations (in attached peripheral or subarea nodes) that are addressable within the subarea share a common subarea address and have distinct element addresses.

subarea node. In SNA, a node that uses network addresses for routing and whose routing tables are therefore affected by changes in the configuration of the network. Subarea nodes can provide gateway function, and boundary function support for peripheral nodes. Type 4 and type 5 nodes are subarea nodes.

subsystem. A secondary or subordinate system, or programming support, that is usually capable of operating independently of or asynchronously with a controlling system.

support element. A hardware unit that provides communications, monitoring, and diagnostic functions to a central processor complex (CPC).

symmetry. The characteristic of a sysplex where all systems, or certain subsets of the systems, have the same hardware and software configurations and share the same resources.

synchronous. (1) Pertaining to two or more processes that depend on the occurrences of a specific event such as common timing signal.
(2) Occurring with a regular or predictable timing relationship.

sysplex. A set of OS/390 systems communicating and cooperating with each other through certain multisystem hardware components and software services to process customer workloads. There is a distinction between a base sysplex and a Parallel Sysplex. See also *OS/390 system, base sysplex, enhanced sysplex*, and *Parallel Sysplex*.

sysplex couple data set. A couple data set that contains sysplex-wide data about systems, groups, and members that use XCF services. All OS/390 systems in a sysplex must have connectivity to the sysplex couple data set. See also *couple data set*.

sysplex data sharing. The ability of multiple IMS subsystems to share data across multiple OS/390 images. Sysplex data sharing differs from two-way data sharing in that the latter allows sharing across only two OS/390 images.

sysplex failure management. The OS/390 function that minimizes operator intervention after a failure occurs in the sysplex. The function uses installation-defined policies to ensure continued operations of work defined as most important to the installation.

sysplex management. The functions of XCF that control the initialization, customization, operation, and tuning of OS/390 systems in a sysplex.

sysplex partitioning. The act of removing one or more systems from a sysplex.

sysplex query parallelism. Businesses have an increasing need to analyze large quantities of data, whether to validate a hypothesis or to discover new relationships between data. This information is often critical to business success, and it can be difficult to get the information in a timely manner. DB2 V4 and later releases lets you split and run a single query within a DB2 subsystem. With sysplex query parallelism, DB2 V5 and later releases extends parallel processing to allow a single query to use all the CPC capacity of a data sharing group.

Sysplex query parallelism is when members of a data sharing group process a single query. DB2 determines an optimal degree of parallelism based on estimated I/O and processing costs. Different DB2 members processes different ranges of the data. Applications that are primarily read or and are processor-intensive or I/O-intensive can benefit from sysplex query parallelism. A query can split into multiple parallel tasks that can run in parallel across all images (up to 32) in a Sysplex. It can run in parallel on up to 344 CPs within a Parallel Sysplex of 32 systems with 12 CPs each.

sysplex sockets. Socket applications are written generally to communicate with a partner on any platform. This means that the improved performance and scalability Parallel Sysplex is not exploited, unless some application-specific protocol is used; this is not always possible.

The sysplex sockets function provides a standard way to discover information about the connected partner which can then be used to make decisions that can exploit the value of the Parallel Sysplex where applicable.

Sysplex Timer. An IBM unit that synchronizes the time-of-day (TOD) clocks in multiple processors or processor sides. External Time Reference (ETR) is the OS/390 generic name for the IBM Sysplex Timer (9037).

system. In data processing, a collection of people, machines, and methods organized to accomplish a set of specific functions.

system configuration. A process that specifies the devices and programs that form a particular data processing system.

system control element (SCE). The hardware that handles the transfer of data and control information associated with storage requests between the elements of the processor unit.

System Support Programs (SSP). An IBM-licensed program, made up of a collection of utilities and small programs, that supports the operation of the NCP.

systems management. The process of monitoring, coordinating, and controlling resources within systems.

S/390 microprocessor cluster. A configuration that consists of CPCs and may have one or more CFs.

S/390 partners in development. Membership in S/390 Partners in Development is open to companies and organizations developing or planning to develop commercially marketed software executing in an IBM S/390 environment under OS/390, VM or VSE operating systems.

Offerings include low-cost application development and porting platforms, answer to technical questions, and information on IBM's trends, directions and latest technology.

- Special offers on development machines and software
- Bulletin-board Q&A support - Free
- Fee-based technical support
- Choice of S/390 platform library (CD-ROM) - Free
- S/390 Rainbow collection CD-ROM - Free
- Porting and technology workshops
- Remote access to IBM systems for application porting and development
- Solution Partnership Centers for test driving IBM Technologies
- Access to early IBM code
- OS/390, VM and VSE technical disclosure meetings
- Information Delivery - Members get funneled pertinent S/390 information

T

takeover. The process by which the failing active subsystem is released from its extended recovery facility (XRF) sessions with terminal users and replaced by an alternate subsystem.

TCP/IP. Transmission control protocol/Internet protocol. A public domain networking protocol with standards maintained by the U.S. Department of Defense to allow unlike vendor systems to communicate.

Telnet. U.S. Department of Defense's virtual terminal protocol, based on TCP/IP.

throughput. (1) A measure of the amount of work performed by a computer system over a given period of time, for example, number of jobs per day. (2) A measure of the amount of information transmitted over a network in a given period of time.

tightly coupled. Multiple CPs that share storage and are controlled by a single copy of OS/390. See also *loosely coupled* and *tightly coupled multiprocessor*.

tightly coupled multiprocessor. Any CPC with multiple CPs.

time-of-day (TOD) clock. A 64-bit unsigned binary counter with a period of approximately 143 years. It is incremented so that 1 is added into bit position 51 every microsecond. The TOD clock runs regardless of whether the processing unit is in a running, wait, or stopped state.

time-to-live (TTL). In the context of a TCP/IP DNS nameserver, the time-to-live is the time that a DNS nameserver will retain resource records in its cache for resources for which it is not the authoritative name server.

TOD. See time-of-day (TOD) clock.

Token-Ring. A network with a ring topology that passes tokens from one attaching device (node) to another. A node that is ready to send can capture a token and insert data for transmission.

transaction. In an SNA network, an exchange between two programs that usually involves a specific set of initial input data that causes the execution of a specific task or job. Examples of transactions include the entry of a customer's deposit that results in the updating of the customer's balance, and the transfer of a message to one or more destination points.

transmission control protocol/Internet protocol (TCP/IP). A public domain networking protocol with standards maintained by the U.S. Department of Defense to allow unlike vendor systems to communicate.

transmitter. In fiber optics, see *optical transmitter*.

trunk cable. In an ESCON environment, a cable consisting of multiple fiber pairs that do not directly attach to an active device. This cable usually exists between distribution panels and can be located within, or external to, a building. Contrast with *jumper cable*.

TSO. See *TSO/E*.

TSO/E. In OS/390, a time-sharing system accessed from a terminal that allows user access to OS/390 system services and interactive facilities.

TTL. See *time-to-live*

tutorial. Online information presented in a teaching format.

type 2.1 node (T2.1 node). A node that can attach to an SNA network as a peripheral node using the same protocols as type 2.0 nodes. Type 2.1 nodes can be directly attached to one another using peer-to-peer protocols. See *end node*, *node*, and *subarea node*.

U

uniprocessor (UP). A CPC that contains one CP and is not partitionable.

universal time coordinate (UTC). UTC is the official replacement for (and is generally equivalent to) the better known “Greenwich Mean Time.”

UP. See *uniprocessor (UP)*.

UTC. See *Universal Time Coordinate*.

V

validity vector. On a CPC, a bit string that is manipulated by cross-invalidate to present a user connected to a structure with the validity state of pages in its local cache.

velocity. A service goal naming the rate at which you expect work to be processed for a given service class or a measure of the acceptable processor and storage delays while work is running.

W

VTS. See *Virtual Tape Server*.

warm start. Synonymous with normal restart.

white space. CF storage set aside for rebuilding of structures from other CFs, in case of planned reconfiguration or failure.

workload. A group of work to be tracked, managed and reported as a unit. Also, a group of service classes.

workload management mode. The mode in which workload management manages system resources on an OS/390 image. Mode can be either *compatibility mode* or *goal mode*.

work qualifier. An attribute of incoming work. Work qualifiers include: subsystem type, subsystem instance, user ID, accounting information, transaction name, transaction class, source LU, NETID, and LU name.

write-ahead data set (WADS). A data set containing log records that reflect completed operations and are not yet written to an online log data set.

X

XCF. See *cross-system coupling facility*.

XCF couple data set. The name for the sysplex couple data set prior to MVS SP V5.1. See *sysplex couple data set*.

XCF dynamics. XCF Dynamics uses the Sysplex Sockets support that is introduced in OS/390 V2R7 IP. Sysplex Sockets allows the stacks to communicate with each other and exchange information like VTAM CPNames, MVS SYSCLONE value and IP addresses. Dynamic XCF definition is activated by coding the IPCONGFIG DYNAMICXCF parameter in TCPIP.PROFILE.

XCF group. A group is the set of related members defined to XCF by a multisystem application in which members of the group can communicate (send and receive data) between OS/390 systems with other members of the same group. A group can span one or more of the systems in a sysplex and represents a complete logical entity to XCF. See *Multisystem application*.

XCF-local mode. The state of a system in which XCF provides limited services on one system and does not provide signalling services between OS/390 systems. See also *single-system sysplex*.

XCF PR/SM policy. In a multisystem sysplex on PR/SM, the actions that XCF takes when one OS/390 system in the sysplex fails. This policy provides high availability for multisystem applications in the sysplex.

XES. See *cross-system extended services*.

XRF. See *Extended recovery facility*.

Y

Year 2000 test datesource facility. The Year 2000 test datesource facility allows you to define several LPs on a single CPC that can enter a sysple with a time and date other than that of the production system. This eliminates the need for dedicating an entire CPC to do year 2000 testing in a multi-member sysplex.

List of Abbreviations

ABARS	aggregate backup and recovery support
ac	alternating current
ACB	access control block
ACDS	alternate control data set
ACDS	active control data set (SMS)
ACS	automatic class selection (SMS)
ACTS	automated computer time service
ADM	asynchronous data mover
ADMF	asynchronous data mover function
ADMOD	auxiliary DC power module
ADSM	ADSTAR distributed storage manager
ALA	link address (ESCON)
AMD	air moving device
AMODE	addressing mode
AMRF	action message retention facility
ANR	automatic networking routing
AOR	application owning region (CICS)
APA	all points addressable
API	application program interface
APPC	advanced program-to-program communication
APPLID	application identifier
APPN	advanced peer-to-peer networking
ARF	automatic reconfiguration facility
ARM	Automatic Restart Manager
AS	address space
AS	auto switchable (Tape Sharing)
ASKQ	ask questions (HONE)
ASYNC	asynchronous
ATL	automated tape library
AVG	average
AWM	alternate wait management (OS/390)
AXM	authorized cross-memory
BAN	boundary access node
BBU	battery backup
BCDS	backup control data set (DFSMShsm)
BCP	basic control program
BCU	basic configurable unit
BCS	basic catalog structure
BDAM	basic direct access method
BGP	border gateway protocol
BIX	build index
BL	block (block-multiplexer channel)
BMP	batch message process (IMS)
BNN	boundary network node (SNA)
BOC	battery operated clock
BOF	bill of forms
BP	buffer pool
BRS	bandwidth reservation
BSC	binary synchronous communication
BSDS	bootstrap data set
BSN	bus switch network
BUF	buffer
BUFND	buffer number data (VSAM)
BUFNI	buffer number index (VSAM)
BUFNO	buffer number
BUFSP	buffer space (VSAM)
BWAT	batch workload analysis tool
BWO	backup while open
BY	byte (byte-multiplexor channel)
CA	Computer Associates
CA	continuous availability
CAA	cache analysis aid
CADS	channel adapter data streaming
CAS	catalog address space
CAS	coordinating address space
CAU	CICS affinity utility
CB	Component Broker
CBIC	control blocks in common
CBIPO	custom-built installation process offering
CBPDO	custom-built product delivery offering
CBU	capacity backup
CBY	ESCON byte multiplexer channel
CCU	communications control unit
CCU	central control unit
CDRM	cross-domain resource manager
CDSC	catalog data space cache
CDS	central directory server
CDS	couple data set
CEC	central electronics complex
CF	Coupling Facility
CFCC	Coupling Facility Control Code
CFCC LP	Coupling Facility Logical Partition
CFDT	Coupling Facility data table
CFFP	Coupling Facility failure policy (OS/390)
CFIA	component failure impact analysis
CFR	Coupling Facility receiver
CFRM	Coupling Facility resources management
CFS	Coupling Facility sender
CFVM	Coupling Facility Virtual Machine (VM)
CHG	changed
CHKP	check point
CHKPT	check point
CHPID	channel path identifier
CI	control interval (VSAM)
CICS	Customer Information Control System
CIU	CICS Interdependencies Utility
CKPT	checkpoint
CLI	call level interface
CLIST	command list
CLO	control link oscillator
CMAS	CICSPlex SM address space
CMC	communications management configuration
CMF	CICS monitoring facility
CMOS	complementary metal oxide semiconductor
CNC	ESCON channel attached to an ESCON-capable device
COB	card-on-board
COMDS	communications data set
CP	control program
CP	central processor
CP/SM	CICSPlex systems manager (CICSPlex SM)
CPC	central processing complex

CPF	command prefix facility
CPU	central processing unit
CQS	common queue server
CR	control region (CICS)
CR	capture ratio
CRA	Configuration Reporting Architecture (S/390)
CDRSC	cross-domain resources
CS	cursor stability (DB2)
CS	central storage
CSA	common systems area
CSAR	complex systems availability and restart
CSECT	control section
CSI	consolidated software inventory (SMP/E)
CSS	channel subsystem
CSTOR	control storage (central storage for CF)
CSU	customer setup
CTC	channel-to-channel
CTCA	channel-to-channel adapter
CU	control unit
CUA	common user access (SAA)
CUoD	Capacity Upgrade on Demand
CV	Central Version (IDMS/CA)
CVC	conversion channel (ESCON)
CVOL	control volume
CWA	common work area
DADSM	direct access device space management
DAE	dump analysis elimination
DASD	direct access storage device
DB	database
DBD	database definition (IMS)
DBRC	database recovery control (IMS)
DCAF	distributed control access facility
DCB	data control block
DCCF	disabled console communication facility
DCE	distributed computing environment
DCFD	Dynamic CF Dispatching
DD	data definition
DDSR	Dynamic Database Session Routing (IDMS/CA)
DDF	distributed data facilities (DB2)
DEDB	data entry database (IMS)
DEQ	dequeue
DFHMS	Data Facility Hierarchical Storage Manager
DFR	deferred
DFSMS	Data Facility Storage Management Subsystem
DFW	DASD fast write
DIB	data in block
DIM	data in memory
DL/1	Data Language 1
DLC	data link control
DLISAS	DL/1, separate address space
DLM	Distributed Lock Manager (Oracle)
DLS	data link switch
DLSw	data link switching
DLUR	dependent LU requester/server
DMA	direct memory access
DMA	dynamic memory array
DNS	Domain Name Server
DP	dispatching priority
DPL	distributed program link (CICS)
DRDA	distributed remote database access
DRDA	distributed relational database architecture
DSG	data sharing group
DSM	distributed security manager
DSNT	data set name table
DSP	data space
DSR	dynamic storage reconfiguration
DTE	data terminal equipment
EC	engineering change
ECB	event control block
ECC	error correction code
ECI	external call interface (CICS)
ECL	emitter coupled logic
ECS	Enhanced Catalog Sharing
EDR	enqueue residency value
EDT	eligible device table
EMC	event monitoring control
EMCS	extended multiple console support
EMH	expedited message handler (IMS)
EMHQ	expedited message handler queue (IMS)
EMIF	ESCON multiple image facility
EMIF	enhanced multiple image facility
EN	end node
ENF	event notification facility
ENQ	enqueue
ENTR	Ethernet/Token-Ring
EP	Emulation Program (3745)
EPDM	Enterprise Data Manager
EPSO	Enhanced S/390 Parallel Sysplex Offering
ERS	enqueue residency value (GRS)
ES	expanded storage
ES	Enterprise Systems
ESCA	ESCON adapter
ESCD	ESCON director
ESCON	Enterprise Systems Connection
ESO	expanded storage only (hiperspace)
ESTOR	non-control storage (expanded storage for CF)
ETOD	Extended time-of-day
ETR	external time reference
ETS	external time source
EXCI	external CICS interface (CICS)
EV	execution velocity (WLM)
FAX	facsimile
FDBR	fast database recovery (IMS)
FF	fast forward
FF	fox fox (hexadecimal)
FF	full function
FICON	FIbre CONnection
FOR	file-owning region (CICS)
FRAD	frame relay access device
FRU	field replaceable unit
FSS	functional subsystem
FTP	file transfer program
FTP	file transfer protocol
FTS	fiber transport services
GA	general availability
GAC	global access checking
GB	gigabyte
GbE	Gigabit Ethernet

GBP group buffer pool (DB2)
GDG generation data group
GEM global enterprise manager
GENCB generate control block
GMLC graduated monthly license charge
GMMA Goal Mode Migration Aid
GMT Greenwich mean time
GR generic resource
GRG generic resource group (IMS)
GRECP group buffer pool recovery pending (DB2)
GRG generic resource group
GRS global resource serialization
GSR global shared resources
GTF Generalized Trace Facility
GUI graphical user interface
GW gateway
HCD hardware configuration definition
HDAM hierarchic direct access method
HFS hierarchical file system (UNIX)
HIDAM hierarchic indexed direct access method
HISAM hierarchic indexed sequential access method
HLQ high level qualifier
HMC hardware management console
HOD host-on-demand
HONE hands on network environment
HPDT high performance data transfer (APPC)
HPR high performance routing (APPN)
HRDW hardware
HSA hardware service area
HSB high speed buffer
HSC hardware system console
HSSI high speed serial interface
HW hardware
HWMCA hardware management console application
Hz hertz
I/O input/output
IBB internal bus buffer
IBF internal battery facility
IBM International Business Machines Corporation
IC internal coupling
ICB integrated cluster bus
ICE I/O controller element
ICF Integrated Catalog Facility
ICF Internal Coupling Facility
ICMF Integrated Coupling Migration Facility
ICP Interconnect Controller Program (IBM program product)
ICRF Integrated Cryptographic Feature
ICS installation control specifications
IFP IMS fast path
IIOP Internet Inter-Object Request Block Protocol
IML initial microcode load
IMS Information Management System
IMS DB Information Management System Database Manager
IMS TM Information Management System Transaction Manager
IOBF I/O buffer (IMS)
IOC I/O count (SRM)
IOCP input output configuration program
IODF I/O definition file
IOP I/O processor
IOQ I/O queue
IOSP input/output support processor
IPA IP assist
IP Internet protocol
IPC initial power controller
IPCS interactive problem control system
IPL initial program load
IPS installation performance specification
IPsec IP security protocol
IPX Internet packet exchange
IRC interregion communications
IRLM integrated resource lock manager
ISC inter-system communications (in CICS and IMS)
ISC inter-system coupling (CF link type)
ISD internal disk (channel path)
ISMF Integrated Storage Management Facility
ISPF Interactive System Productivity Facility
ISO integrated services offering
ISR intermediate session routing
ISV independent software vendor
ITR internal throughput rate
ITRR internal throughput rate ratio
ITSC International Technical Support Center
ITSO International Technical Support Organization
ITU International Telecommunication Union
JCL job control language
JECL JES control language
JES Job entry subsystem
JMF JES3 monitoring facility
JOE job output element (JES)
KB kilobyte
KGTV Korhonen George Thorsen Vaupel
km kilometer
LAN local area network
LASER light amplification by stimulated emission of radiation
LCU logical control unit
LE language environment
LED light emitting diode
LFS LAN file server
LIC licensed internal code
LOB large object (DB2)
LOC locate
LOC location
LP logical partition
LPAR logically partitioned mode
LPCTL logical partition controls (frame)
LPL logical page list
LSCD large scale computing division
LSPR large systems performance reference
LSR local shared resources
LST load module temporary store (SMP/E)
LTERM logical terminal
LU logical unit (SNA)
LUPS local UPS
LX long wavelenght

MAC medium access control
MAS multiaccess spool
MAU multistation access unit
MAX maximum
MB megabyte
Mbps megabits per second
MB/Sec megabytes per second
MCCU multisystem channel communications unit
MCDS migration control data set (DFSMShsm)
MCL maintenance change level
MCM multichip module
MCS multiple console support
MDB message data block
MDH migration data host
MES miscellaneous equipment specification
MIB management information base (OSI)
MICR magnetic ink character recognition/reader
MIH missing interrupt handler
MIM Multi-Image Manager
MIPS millions of instructions per second
MLSLEC maximum list-set-element count
ML1 migration level 1 (DFSMShsm)
ML2 migration level 2 (DFSMShsm)
MODCB modify control block
MM multimode
MNPS multinode persistent session
MP multiprocessor
MPC+ multipath channel+
MPF message suppressing facility
MPG multiple preferred guests
MPL multiprogramming level
MPNP multiprotocol network program
MPR message processing region (IMS)
MPTN multiprotocol transport networking
MRNS multiprotocol routing network services
MRO multiregion operation
MSC multisystems communication (IMS)
MSDB main storage database (IMS)
MSGQ shared message queue (IMS)
MSO main storage occupancy (SRM)
MSS multiprotocol switched services
MSU millions of service units
MTS macro temporary store (SMP/E)
MTU maximum transmission unit (Ethernet)
MTW mean time to wait
MUF multi-user facility (Datacom/CA)
MULC measured usage license charge
MVS Multiple Virtual Storage
N/A not applicable
NAU network addressable unit
NDF non-deferred
NFS network file system
NIB node initialization block
NIP nucleus initialization process
NLP network layer packets
NN network node (SNA)
NNP network node processor
NNS named counter server (CICS)
NO number
NOOVLY no overlay
NSS national system support
NSSS networking and system services and support
NVS nonvolatile storage
OAS OSA address table
OCDS offline control data set (DFSMShsm)
OCF operations command facility
OCR optical character recognition/reader
ODBC open database connectivity
OEMI original equipment manufacturer information/interface
OLDS online log data set (IMS)
OLS offline sequence
OMF operations management facility
ONC open network computing
OO object-oriented
OPC operations planning and control
OS operating system
OSA open systems architecture
OSAM overflow sequential access method
OSI open systems interconnection
OSPF open shortest path first
OVLY overlay
PAF Processor Availability Facility
PCE processor controller element
PCM plug-compatible manufacturers
P/DAS peer-to-peer dynamic address switching
PD program directory
PDS partitioned data set
PDSE partitioned data set enhanced
PEL picture element
PEP Partitioned Emulation Program (3745)
PES Parallel Enterprise Server
PET platform evaluation test
PI performance index
PI path in
PI program isolation
PIF program interface
PLET product announcement letter
PLO perform locked operation
PM Presentation Manager (OS/2)
PMIO performance management input/output
PMOS performance management offerings and services
PO path out
POR power-on reset
PP physical partitioned (mode of operation)
PPP point-to-point protocol
PPRC peer-to-peer remote copy
PPS pulse-per-second
PR/SM processor resource/system manager
PROFS professional office system
PSB program specification block
PSLC Parallel Sysplex license charge
PSMF Parallel Sysplex Management Facility
PSO Parallel Server Option (Oracle)
PSP preventive service planning
PSP primary support processor
PTF program temporary fix
PTH path
PTS parallel transaction server
PU physical unit (SNA)
PU processing unit (9672)

QDIO Queued Direct Input/Output
QOR queue-owning region (CICS)
RACF Resource Access Control Facility
RAMAC Random Access Method of Accounting and Control
RAS reliability availability serviceability
RBA relative block address
RBA relative byte address
RCD read configuration data
RDS restructured database (RACF)
REA RAMAC electronic array
REQ required
RETAIN Remote Technical Assistance and Information Network
RG resource group
RIOC relative I/O content
RIP routing information protocol
RIT RECON initialization timestamp (IMS)
RJP remote job processor
RLL row-level locking
RLS record-level sharing (VSAM)
RLSWAIT RLS wait
RMF Resource Measurement Facility
RMODE residency mode
RNL resource name list (GRS)
ROT rules of thumb
RPC remote procedure call
RPL request parameter list
RPP relative processor performance
RPQ request for price quotation
RR repeatable read (DB2)
RRDF Remote Recovery Data Facility
RRMS Recoverable Resource Management Services (OS/390)
RRS resource recovery services
RRSF RACF Remote Sharing Facility
RSA RAMAC scalable array
RSA ring system authority (GRS)
RSM real storage manager
RSR remote site recovery
RSU recommended service upgrade
RTA real time analysis
RTM recovery termination manager
RTP rapid transport protocol
RVA RAMAC virtual array
SA OS/390 System Automation for OS/390
SAA systems application architecture
SAE single application environment
SAF System Authorization Facility
SAP system assist processor
SAPR systems assurance product review
SCA shared communication area (DB2)
SCDS save control data set (SMP/E)
SCDS source control data set (SMS)
SCE system control element
SCH subchannel
SCKPF Store Clock Programmable Field
SCP system control program
SCS source control data set
SCSI small computer system interface
SCTC ESCON CTC control unit
SCV software-compatible vendor
SDEP sequential dependent (IMS)
SDT shared data tables (CICS)
SE support element
SEC system engineering change
SECS seconds
SETI SE technical information (HONE)
SFM sysplex failure management
SHISAM simple hierarchic indexed sequential access method
SI single image (mode of operation)
SID system identifier
SIE start interpretive execution (instruction)
SIGP signal processor (instruction)
SIMETR simulated external time reference (OS/390)
SLA service level agreement
SLIP serviceability level indicator processing
SLM system lock manager (XCF)
SLMH single-link multihost (IBM 3174 with ESCON interface)
SLSS system library subscription service
SM single-mode
SMF Systems Management Facility
SMOL sales manual online (HONE)
SMP/E system modification program/extended
SMPLOG SMP/E log
SMPLTS SMP/E load module temporary store
SMPMTS SMP/E macro temporary store
SMPSCDS SMP/E save control data set
SMPSTS SMP/E source temporary store
SMQ shared message queue (IMS)
SMS system managed storage
SMSVSAM system managed storage VSAM
SNA systems network architecture
SNAP/SHOT system network analysis program/ simulation host overview technique
SNI SNA network interconnect
SNMP simple network management protocol
SP system product
SPE small programming enhancement
SPOC single point of control
SPOF single point of failure
SPOOL simultaneous peripheral operation online
SPUFI SQL processor using file input (DB2)
SQG shared queues group (IMS)
SQL structured query language (DB2)
SQL/DS structured query language/data system (VM)
SRB service request block
SRDS structure recovery data set (CQS)
SROD shared read-only database (IMS)
SRM System Resources Manager
S/S start/stop
SSCH start subchannel
SSI single system image
SSI subsystem interface
SSL secure socket layer
SSM secondary space management (DFSMShsm)
SSP subsystem storage protect
STC started task control
STCK store clock (instruction)

STCKE	store clock extended (instruction)
STI	self-timed interconnect (S/390)
STS	source temporary store
STSI	Store System Information Instruction (S/390)
SUBSYS	subsystem
SC	Shared Cache (IDMS/CA)
SVC	supervisor call (instruction)
SVS	solutions validation services
SW	software
SX	short wavelength
SYNC	synchronous
SYSAFF	system affinity
SYSCONS	system consoles
SYSID	system identifier
SYSPLEX	systems complex
SYSRES	system residence volume (or IPL volume)
TAI	French for International Atomic Time
TCB	task control block (OS/390)
TCM	thermal conduction Module
TCP/IP	Transmission Control Protocol/Internet Protocol
TCU	terminal control unit
TESTCB	test control block
TG	transmission group
TKE	trusted key entry
TM	transaction manager
TME	Tivoli Management Environment
TMM	tape mount management
TOD	time of day
TOR	terminal-owning region
TP	transaction program (APPC)
TPF	Transaction Processing Facility
TPNS	Teleprocessing Network Simulator
TS	temporary storage (CICS)
TS	transaction server (CICS)
TSC	TOD synchronization compatibility
TSO	Time Sharing Option
TSOR	temporary storage-owning region (CICS)
TSQ	temporary storage queue (CICS)
TTL	time-to-live (TCP/IP)
UBF	user buffering
UCW	unit control word
UD	undeliverable (message)
UDB	universal database (DB2)
UDF	update-time-order (DB2)
UDP	user diagram protocol
UOW	unit of work
UP	uniprocessor
UPS	uninterruptible power supply/system
UR	uncommitted read (DB2)
URL	universal resource locator
UTC	universal time coordinate
VF	vector facility
VGA	video graphics array/adapter
VIO	virtual I/O
VIPA	virtual IP addressing (TCP/IP)
VLF	virtual lookaside facility
VM	Virtual Machine
VPD	vital product data
VR	virtual route (SNA)
VRTG	virtual-route-based transmission group (SNA)
VSAM	Virtual Storage Access Method
VSO	virtual storage option (IMS)
VTAM	Virtual Telecommunications Access Method
VTOC	volume table of content
VTS	virtual tape server
VVDS	VSAM volume data set
WAITRBLD	wait for rebuild (IMS)
WAN	wide area networks
WLM	Workload Manager
WLR	IMS/ESA workload router
WQE	write-to-operator-queue-element
WSC	Washington System Center
WSS	working set size
WTAS	world trade account system
WTO	write-to-operator
WTOR	write-to-operator-with-reply
WWQA	world wide question & answer
WWW	World Wide Web
XCA	external communications adapter
XCF	Cross-System Coupling Facility
XDF	Extended Distance Feature
XES	Cross-System Extended Services
XI	cross-invalidate
XJS	extended job entry subsystem
XRC	extended remote copy
XRF	Extended Recovery Facility

Index

Special Characters

Numerics

A

B

C

D

E

F

G

J

L

N

O

P

Q

R

S

T

X

IBM Redbooks evaluation

OS/390 Parallel Sysplex Configuration Connectivity
SG24-5639-00

Your feedback is very important to help us maintain the quality of IBM Redbooks. **Please complete this questionnaire and return it using one of the following methods:**

- Use the online evaluation form found at `http://www.redbooks.ibm.com/`
- Fax this form to: USA International Access Code + 1 914 432 8264
- Send your comments in an Internet note to `redbook@us.ibm.com`

Which of the following best describes you?
__Customer __Business Partner __Solution Developer __IBM employee
__None of the above

Please rate your overall satisfaction with this book using the scale:
(1 = very good, 2 = good, 3 = average, 4 = poor, 5 = very poor)

Overall Satisfaction ____________

Please answer the following questions:

Was this redbook published in time for your needs? Yes____ No____

If no, please explain:

____

____

____

____

What other redbooks would you like to see published?

____

____

____

Comments/Suggestions: (THANK YOU FOR YOUR FEEDBACK!)

____

____

____

____

____

SG24-5639-00
Printed in the U.S.A.

OS/390 Parallel Sysplex Configuration Connectivity

SG24-5639-00