Diagnosing Sysplex Problems

Mark A Brooks
IBM

Wednesday, February 6, 2013: 3:00 PM-4:00 PM
Session Number 13402

mabrook@us.ibm.com
The following are trademarks of the International Business Machines Corporation in the United States and/or other countries.

IBM® IBM® (logo)
ibm.com® AIX® BladeCenter®
CICS® DS4000®
CICSPlex® DS6000™
DB2® POWER7®
eServer™ ProtecTIER®
ESCON® Redbooks®
FICON® Resource Measurement Facility™
GDPS® zSeries®
HyperSwap® Flash Express®
IMS™ System Storage®
IMS/ESA® System x®

The following are trademarks or registered trademarks of other companies.

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.

Cell Broadband Engine is a trademark of Sony Computer Entertainment, Inc. in the United States, other countries, or both and is used under license there from.

Java and all Java-based trademarks are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

InfiniBand is a trademark and service mark of the InfiniBand Trade Association.

Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

ITIL is a registered trademark, and a registered community trademark of the Office of Government Commerce, and is registered in the U.S. Patent and Trademark Office.

IT Infrastructure Library is a registered trademark of the Central Computer and Telecommunications Agency, which is now part of the Office of Government Commerce.

* All other products may be trademarks or registered trademarks of their respective companies.

Notes:
Performance is in Internal Throughput Rate (ITR) ratio based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput improvements equivalent to the performance ratios stated here.

IBM hardware products are manufactured from new parts, or new and serviceable used parts. Regardless, our warranty terms apply.

All customer examples cited or described in this presentation are presented as illustrations of the manner in which some customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics will vary depending on individual customer configurations and conditions.

This publication was produced in the United States. IBM may not offer the products, services or features discussed in this document in other countries, and the information may be subject to change without notice. Consult your local IBM business contact for information on the product or services available in your area.

All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

Information about non-IBM products is obtained from the manufacturers of those products or their published announcements. IBM has not tested those products and cannot confirm the performance, compatibility, or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

Prices subject to change without notice. Contact your IBM representative or Business Partner for the most current pricing in your geography.
Downloads available from SHARE website

- **Presented Slides**
  - Material to be presented during the session
  - “See handout” is a reference to slides in the handout that will not be presented

- **Handout Slides**
  - All of the above plus additional slides that will not be presented
    - Has a more complete treatment of key concepts
    - Includes examples of relevant messages, display output, and reports to illustrate what you would be looking at when diagnosing problems
  - If following along with the handout during the presentation, you can use the slide numbers in the lower left corner to keep pace (there will be lots of skipped slides)
Sysplex environment

Couple data sets

Signalling paths

Common Time Reference

Applications
Middleware
z/OS Components

XES

XCF

DASD

Shared data

CF
Context

- In the sysplex, we have instances of z/OS, middleware, and applications running on individual systems.

- These individual instances cooperate with one another to perform some function.

- A given sysplex function may rely on the services of other sysplex functions, sometimes in rather non-obvious, even circular ways.
Sysplex Problems

- Sympathy sickness (hangs)
  - If an instance of a sysplex function is unresponsive, it may:
    - Hold serialization on shared resource
    - Stop sending replies
    - Other instances may hang as a result

- Performance issues (delays)
  - Instance is responsive, but “slow”
  - May impact response time of individual requests, which at high request rates, can lead to significant throughput problems and/or large queue effects that also give the appearance of hangs
Signs that a sysplex might be sick

- IXC102A “reply down when system reset”
- IXC426D “sending signals but not updating status, how to proceed?”
- IXC431I “member stalled”
- IXC631I “member stalled and impacting other systems”
- IXC633I “member is impaired”
- IXL040E, IXL041E “structure connector not responding”
- IXL045E “having trouble getting connector SRBs to run”
- IXL044I “errors communicating with CF”
- ISG361A “waiting for list lock”
- ISG178E “ring disruption”
- D GRS,C
  - ISG343I output
- IOS071I “MIH timeout”
Signs that a sysplex might be sick ...

- IXC426D “sending signals, not updating status. How to proceed?”
- IXC427A “sending signals, not updating status. SFM will act”
- IXC446I “sending signals, not updating status. SFM will act”
- IXC256A “cannot remove CDS until these systems respond”

But the real question is: **Why is my sysplex sick?**
And our concern is that the symptoms are not always so obvious
Getting to root cause of problem can be hard

- You don't have a “sysplex problem”, you have a “problem”
- The problem is more complex because of its sysplex context
  - Lots of “pieces” scattered throughout the sysplex
  - Any given symptom may have a root cause elsewhere in the sysplex
  - As problem persists, its impact tends to spread, which induces more symptoms, which makes the root cause harder to find
  - Lack expertise to understand relationships and dependencies
- But ultimately there is likely some one thing on some one system that is causing the problem
- How to find that thing?
My thesis

- We do not have the expertise needed to do sysplex diagnosis
- We do not know:
  - Implementations, interactions, and dependencies
  - Causes, relationships, or relevance of various symptoms
  - How a given problem might impact the sysplex
  - All potential causes of a given symptom
- But we do know:
  - Sysplex application instances interact with one another
  - Most exploit sysplex services to do so
  - These services and their usage can be observed and analyzed
- So despite our imperfect knowledge, there is hope
  - We can make sure the sysplex infrastructure is sound
  - If applications can readily communicate and share data, we likely have a
    problem that will yield to traditional single system diagnosis
  - Hopefully our analysis will point us towards the vicinity of the culprit
Objectives

- Describe a small corner of the universe of sysplex infrastructure problems and their symptoms
- Provide a methodology to reliably discover root causes
  - Or at least get closer to the real source of the problem
- As a side benefit, you should be better able to:
  - Prevent problems
  - Assess risks
  - Apply appropriate remedies to resolve problems
- And if you need help, you should be better able to:
  - Provide the data needed for diagnosis and resolution
  - Identify the trouble spot, and thereby start your service call with the right set of IBM component experts
How Component Experts Diagnose Problems

Step 1: Investigate “my component”
Step 2: Look at everything in “my component’s” space
  - Configuration
  - Defect
  - Workload
  - Changes
Step 3: Share findings with customers
Step 4: Customer will try suggestions
Step 5: If sufficient relief attained then DONE. Otherwise, continue to Step 6.
Step 6: If it’s not “my component” consider what else it can be???
  - Maybe the “correct” next component is selected
Step 7: Engage another component or transfer the PMR to the next component
Step 8: Go to Step 1
Bridging the Gap

- You know your systems and your workload (I hope), but lack component expertise

- IBM has component expertise, but does not know your shop

- I hope this material helps to bridge the gap
  - Use knowledge of your shop
  - Identify relevant symptoms
  - Deduce potential causes
  - Engage the right component experts

- So that problem resolution is timely and effective
What is normal? What changed?

- Understanding differences between past and current (problem) behavior is often helpful for diagnosis
  - Configuration
  - Workload
  - Utilization
  - Request rates
  - Maintenance, either software or hardware
  - Migration, either software or hardware

- Good change activity logs can be invaluable for helping to pinpoint what and when (why?) a potentially relevant change was made
  - Although sometimes the change that leads to a problem was long ago and far away
Problem Taxonomy

- Dead System
- Sick System
- Sysplex Fabric
- Sysplex Componentry
  - Coupling Facility
  - Signalling Service
  - Couple Data Sets
  - External Time Reference
- Configuration / Capacity
- Software Issues
Dead System

- A dead system can’t participate in anything
  - Dead System = wait state and not removed from sysplex
  - The root cause of sympathy sickness often turns out to be a dead system that no one noticed!

- **So this should always be the first thing you check**

- **Symptoms**
  - IXC402D “system looks dead”
  - IXC102A “tell me when system is reset”
  - Other messages (see handout)
Dead System Symptoms: XCF messages

- IXC101I “removing system from sysplex”
- IXC105I “removed system from sysplex”
- IXC102A “reset system and reply down”
- IXC108I “fencing system”
- IXC109I “fencing completed (or failed)”
- IXC256A “cannot remove CDS until these systems respond”
- IXC409D “lost signal connectivity, how to proceed?”
- IXC426D “sending signals but not updating status, how to proceed?”
- IXC427A “sending signals, not updating status. SFM will act”
- IXC602I “SFM will take this action if system looks dead”
- IXC800I “ARM could not restart elements elsewhere”
Dead system must be dealt with ASAP

- Failure to deal with a dead system in a timely manner can easily induce sympathy sickness that expands into a morass of extraneous and confusing symptoms because nearly every sysplex application is likely to be impacted sooner or later somewhere in the sysplex.

- So you really want automatic removal
  - SFM policy with ISOLATETIME (see handout)
  - SFM with BCPii

- Or you have to rely on vigilant operators
Fencing isolates a system so that it cannot access shared data, thus making it safe for the survivors to release serialization of the shared resources.

A command is sent via a CF to the target CEC. The target image will not be able to initiate any new I/O and ongoing I/O will be terminated.
SFM with ISOLATETIME

“Status missing” means both:
- Not updating status in sysplex CDS
- Not sending signals

Proven to be a reliable indicator
SFM With BCPii

XCF uses BCPii to
- Obtain identity of an image
- Query status of remote CPC and image
- Reset an image

z/OS Images
(not VM guests)

Requires operational SE and HMC network
z/OS 1.11 SFM with BCPii

- Expedient removal of unresponsive or failed systems is essential to high availability in sysplex
- XCF exploits BCPii services to:
  - Detect failed systems
  - Reset systems
- Benefits:
  - Improved availability by reducing duration of sympathy sickness
  - Eliminate manual intervention in more cases
  - Potentially prevents human error that can cause data corruption
SFM with BCPii

Last status update

Status Missing condition exceeds system FDI value.

Isolation of failed system begins

Periodically use BCPii to query state of system

FID = Failure Detection Interval

IXC101I System removed

IXC102A

Operator

IXC108I Fails

Fence system

IXC108I Fails

Reset system

IXC106I System demised

Being removed

Fence system

Fails

Reset system

Fails

IXC105I
z/OS 1.11 SFM with BCPii

SFM will automatically exploit BCPii and as soon as the required configuration is established:

- Pairs of systems running z/OS 1.11 or later
- BCPii configured, installed, and available
- XCF has security authorization to access BCPii defined FACILITY class resources
- z10 GA2 with appropriate MCL’s, or z196, or z114, or zEC12
- **New version of sysplex CDS is primary in sysplex**
  - Toleration APAR OA26037 for z/OS 1.9 and 1.10
  - Does NOT allow systems to use new SSD function or protocols

If you have the appropriate environment, SET THIS UP!
You will likely eliminate this entire class of “dead system” issues
Manual intervention may still be needed!

- Use of BCPii and SFM policies that permit the sysplex to automatically recover from dead systems will eliminate many of the cases where manual intervention is required.
- XCF falls back on manual intervention if these fail
  - Need to be prepared for that possibility
    - Likely fewer opportunities for operators to maintain skills
  - So even if you exploit these technologies, checking for a dead system should still be first thing to consider

Suggestion: Page system programmer when IXC102A is issued
Problem Taxonomy

- Dead System
- **Sick System**
- Sysplex Fabric
- Sysplex Componentry
  - Coupling Facility
  - Signalling Service
  - Couple Data Sets
  - External Time Reference
- Configuration / Capacity
- Software Issues
Why Discuss Sick System?

- Why discuss “single system diagnosis” in a “sysplex diagnosis” presentation?

- Many are fooled into thinking they have a sysplex problem when in fact they have a system problem
  - *So problem resolution is inefficient and takes longer than needed*

- A sick system can induce sympathy sickness
  - *So the root cause of a sysplex problem may well be one that is single system in scope*
  - *If the local components of the sysplex infrastructure are not running on a solid foundation, the sysplex will likely suffer*
Sick System: Typical Root Causes

- Storage constraints
  - Not enough resource for system to run reasonably
  - Run away application
  - Defect

- CPU constraints
  - Not enough resource for system to run reasonably
  - SRBs looping

- Contention
  - ENQ, latches, local lock, spin locks

- DASD I/O issues

- Sick components
Run Time Diagnostics (RTD)

- Reviews critical messages in the log
- Analyzes contention
  - GRS ENQ
  - GRS Latches
  - z/OS UNIX file system latches
- Examines address spaces with high CPU usage
- Looks for an address space that might be in a loop
- Evaluates local lock conditions
- Performs additional analysis based on what is found
  - For example, if XES reports a connector as unresponsive, RTD investigates the appropriate address space

RTD can find many sick system issues. Give it a try.
Runtime Diagnostics

- Allows installation to quickly analyze a system experiencing “sick but not dead” symptoms
- Looks for evidence of “soft failures”
- Reduces the skill level needed when examining z/OS for “unknown” problems where the system seems “sick”
- Provides timely, comprehensive analysis at a critical time period with suggestions on how to proceed

- Runs as a started task in z/OS V1R12
  - $ S $ HZR
- Starts at IPL in z/OS V1R13
  - $ F $ HZR,ANALYZE command initiates report

- Message HZR0200I contains the report
Sick System: Storage constraints

- System may not run well if storage constrained
- Shortages tend to induce paging
  - Resolving page faults introduces delay and system overhead
  - Swap out of address spaces implies programs are not running

- Typical Causes
  - You need more memory
  - Runaway application
  - Page pack performance (see “Sick System: DASD I/O issues”)
  - Defects
Detecting Storage Constraints

- What changed?

- Who is consuming storage?
  - How much?
  - May need application knowledge to determine whether reasonable
  - But, is this typical for you?
Detecting Storage Constraints

- Messages (see handout)
  - Auxiliary storage shortages
  - Pageable storage shortages
  - Frame shortages

- ABENDs
  - XCF 00C rsn xxxx006C implies frame shortage

- IPCS RSMDATA SUMMARY to see frame usage

- IPCS ANALYZE RESOURCE
  - Look for “ANY REAL FRAME”, presence implies storage constraint

- IPCS SYSTRACE ALL
  - Find “AVQ”. If low, implies RSM needs to replenish frames and requests for virtual storage may hang
Sick system: Storage constraint messages

- IRA200E “auxiliary storage shortage”
- IRA201E “critical auxiliary storage shortage”
- IRA202I “auxiliary storage shortage relieved”
- IRA203I “consumer of auxiliary storage”
- IRA204E “consumer of auxiliary storage”
- IRA205I “consumed half of auxiliary storage”
- IRA206I “consumer of auxiliary storage”
- IRA210E “consumer set non-dispatchable”
- IRA211I “consumer set dispatchable”
- IRA220I “critical auxiliary storage shortage”
- IRA221D “show more or cancel consumer”
- IRA222I “consumer not canceled”
- IEE787A “pageadd command, how to proceed?”
Sick system: Storage constraint messages

- IRA400E “pageable storage shortage”
- IRA401E “critical pageable storage shortage”
- IRA402I “pageable storage shortage relieved”
- IRA403E “swapped out address space”
- IRA404I “report storage consumed by address space”
- IRA405I “report high % fixed frames”
- IRA410E “set non-dispatchable”
- IRA411I “set dispatchable”
- IRA420I “consumers of fixed frames”
- IRA421D “show more or cancel consumer”
- IRA422I “consumer not canceled”
Sick system: Storage constraint messages

- IRA100E “SQA shortage”
- IRA101E “critical SQA shortage”
- IRA102I “SQA shortage relieved”
- IRA103I “SQA expanded into CSA”
- IRA104I “SQA expansion into CSA relieved”
- IRA110E “high shared virtual storage shortage”
- IRA111E “critical high shared virtual storage shortage”
- IRA112I “high shared virtual storage shortage relieved”
- IRA120E “large frame shortage”
- IRA121E “critical large frame shortage”
- IRA122I “large frame shortage relieved”
- IRA130E “high common storage shortage”
- IRA131E “critical high common storage shortage”
- IRA132I “high common storage shortage relieved”
Sick System: CPU Constraints

- “Something” will not be running
- Usually the “something” is deemed to be less important, so it may seem that you are getting the desired result
  - Test system
  - Discretionary work
- But failure to run the less important “something” can induce sympathy sickness for the important work if it:
  - Holds serialization for which there is contention
  - Holds resources, or worse, accumulates them while not running
    - For example, XCF Signal buffers
  - Not participating in cooperative processes
CPU Constraint Considerations

- **Latent demand?**
  - If LPAR busy less than MVS busy, the physical processor is being ripped away despite the fact that MVS has work to do

- **Blocked work?**
  - If ready work is only getting “trickles”, the box is pretty much saturated

- **Running high importance interactive/transaction oriented workloads at more than 90-92% busy** is asking for trouble

- **Did something change?**
  - Configuration (upgrade, number CPs, capping, CF, ...)
  - Workload
  - CPU utilization by some job/space
Detecting CPU Constraints

- RMF partition data reports
  - CPC Capacity
  - Distribution of IN-READY Queue

- RMF Monitor III
  - CPC Report

- Will likely need a timely dump to determine if there is a loop
  - Run Time Diagnostics might be able to detect it
## CPU Activity

**System Information:**
- **z/OS Version:** V1R11
- **System ID:** PP1A
- **Date:** 05/27/2012
- **Interval:** 15.00.001
- **Conversion:** Converted to z/OS V1R12 RMF
- **Time:** 19.00.00
- **Cycle:** 1.000 seconds
- **Model:** 750
- **Change Reason:** N/A
- **HiperDispatch:** Yes
- **H/W Model:** E56

### CPU Activity Table

<table>
<thead>
<tr>
<th>NUM</th>
<th>TYPE</th>
<th>ONLINE</th>
<th>LPAR BUSY</th>
<th>MVS BUSY</th>
<th>PARKED</th>
<th>SHARE %</th>
<th>RATE</th>
<th>% VIA TPI</th>
<th>LOG PROC</th>
<th>--I/O INTERRUPTS--</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>CP</td>
<td>100.00</td>
<td>40.99</td>
<td>42.37</td>
<td>0.00</td>
<td>100.0</td>
<td>HIGH</td>
<td>233.4</td>
<td>28.92</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>CP</td>
<td>100.00</td>
<td>64.68</td>
<td>64.35</td>
<td>0.00</td>
<td>100.0</td>
<td>HIGH</td>
<td>1113</td>
<td>27.85</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>CP</td>
<td>100.00</td>
<td>15.31</td>
<td>15.28</td>
<td>0.00</td>
<td>100.0</td>
<td>HIGH</td>
<td>56.55</td>
<td>27.45</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>CP</td>
<td>100.00</td>
<td>18.61</td>
<td>18.57</td>
<td>0.00</td>
<td>100.0</td>
<td>HIGH</td>
<td>74.50</td>
<td>28.44</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>CP</td>
<td>100.00</td>
<td>42.72</td>
<td>42.62</td>
<td>0.00</td>
<td>100.0</td>
<td>HIGH</td>
<td>268.6</td>
<td>28.69</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>CP</td>
<td>100.00</td>
<td>51.47</td>
<td>51.19</td>
<td>0.00</td>
<td>100.0</td>
<td>HIGH</td>
<td>884.9</td>
<td>27.19</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>CP</td>
<td>100.00</td>
<td>13.54</td>
<td>13.51</td>
<td>0.00</td>
<td>100.0</td>
<td>HIGH</td>
<td>52.71</td>
<td>27.76</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>CP</td>
<td>100.00</td>
<td>16.32</td>
<td>16.29</td>
<td>0.00</td>
<td>100.0</td>
<td>HIGH</td>
<td>66.75</td>
<td>28.83</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>CP</td>
<td>100.00</td>
<td>54.93</td>
<td>54.82</td>
<td>0.00</td>
<td>100.0</td>
<td>HIGH</td>
<td>309.3</td>
<td>28.63</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>CP</td>
<td>100.00</td>
<td>31.54</td>
<td>31.50</td>
<td>0.00</td>
<td>100.0</td>
<td>HIGH</td>
<td>114.5</td>
<td>28.00</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>CP</td>
<td>100.00</td>
<td>18.08</td>
<td>18.05</td>
<td>0.00</td>
<td>100.0</td>
<td>HIGH</td>
<td>61.62</td>
<td>28.28</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>CP</td>
<td>100.00</td>
<td>24.52</td>
<td>24.49</td>
<td>0.00</td>
<td>100.0</td>
<td>HIGH</td>
<td>87.97</td>
<td>28.65</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>CP</td>
<td>100.00</td>
<td>51.44</td>
<td>51.35</td>
<td>0.00</td>
<td>100.0</td>
<td>HIGH</td>
<td>247.8</td>
<td>28.48</td>
<td></td>
</tr>
</tbody>
</table>

**etc.**
Post Processor: CPU Activity Report

<table>
<thead>
<tr>
<th>NUMBER OF ADDR SPACES (%)</th>
<th>0</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
<th>90</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;= N</td>
<td>100.0</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>= N + 1</td>
<td>0.0</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>= N + 2</td>
<td>0.0</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>= N + 3</td>
<td>0.0</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>&lt;= N + 5</td>
<td>0.0</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>&lt;= N + 10</td>
<td>0.0</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>&lt;= N + 15</td>
<td>0.0</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>&lt;= N + 20</td>
<td>0.0</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>&lt;= N + 30</td>
<td>0.0</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>&lt;= N + 40</td>
<td>0.0</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>&lt;= N + 60</td>
<td>0.0</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>&lt;= N + 80</td>
<td>0.0</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>&gt;  N + 80</td>
<td>0.0</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>

N = NUMBER OF PROCESSORS ONLINE UNPARKED (48.0 ON AVG)

The more lines you have with the >>>>> arrows, the worse the performance is.
### RMF Monitor III CPC Capacity

**RMF V1R11 CPC Capacity**

<table>
<thead>
<tr>
<th>Samples: 100</th>
<th>System: SD0</th>
<th>Date: 08/07/12</th>
<th>Time: 10.36.40</th>
<th>Range: 100 Sec</th>
</tr>
</thead>
</table>

**Partition:** SD0 2817 Model 750

**CPC Capacity:** 4300  
Weight % of Max: 10.0  
4h Avg: 1  
Group: N/A

**Image Capacity:** 1376  
WLM Capping %: 0.0  
4h Max: 33  
Limit: N/A

<table>
<thead>
<tr>
<th>Partition</th>
<th>--- MSU ---</th>
<th>Cap</th>
<th>Proc</th>
<th>Logical Util % - Physical Util % -</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Def</td>
<td>Act</td>
<td>Def</td>
</tr>
</tbody>
</table>

| *CP       | 88.0        | 0.3  | 4.8  | 5.1 |
| CSK       | 0 2 NO 8.0  | 0.3  | 0.3  | 0.0 | 0.0 | 0.1 |
| SA0       | 0 39 NO 8.0 | 5.6  | 5.7  | 0.0 | 0.9 | 0.9 |
| SB0       | 0 16 NO 8.0 | 2.2  | 2.3  | 0.0 | 0.4 | 0.4 |
| SC0       | 0 43 NO 8.0 | 6.0  | 6.2  | 0.0 | 1.0 | 1.0 |
| SD0       | 0 3 NO 8.0  | 0.4  | 0.4  | 0.0 | 0.1 | 0.1 |
| SJ0       | 0 3 NO 8.0  | 0.4  | 0.4  | 0.0 | 0.1 | 0.1 |
| etc.      |             |      |      |     |     |      |      |      |      |
## RMF Monitor III Processor Delay

<table>
<thead>
<tr>
<th>Jobname</th>
<th>CX</th>
<th>Class</th>
<th>CPU</th>
<th>DLY</th>
<th>USG</th>
<th>EAppl</th>
<th>Service</th>
<th>% Name</th>
<th>% Name</th>
<th>% Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>WWS07</td>
<td>O</td>
<td>OMVS</td>
<td>CP</td>
<td>11</td>
<td>46</td>
<td>59.4</td>
<td>% Name</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WSP1S2FS</td>
<td>SO</td>
<td>WASCR</td>
<td>CP</td>
<td>4</td>
<td>12</td>
<td>42.5</td>
<td>% Name</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AAP</td>
<td>6</td>
<td>12</td>
<td>39.4</td>
<td>% Name</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WSP1S6FS</td>
<td>SO</td>
<td>WASCR</td>
<td>CP</td>
<td>0</td>
<td>0</td>
<td>5.3</td>
<td>% Name</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AAP</td>
<td>6</td>
<td>0</td>
<td>5.3</td>
<td>% Name</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DBS3DBM1</td>
<td>S</td>
<td>DB2HIGH</td>
<td>CP</td>
<td>2</td>
<td>6</td>
<td>111.0</td>
<td>% Name</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WSP1S6F</td>
<td>SO</td>
<td>WASCR</td>
<td>CP</td>
<td>0</td>
<td>0</td>
<td>5.3</td>
<td>% Name</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AAP</td>
<td>6</td>
<td>0</td>
<td>3.0</td>
<td>% Name</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U078069</td>
<td>O</td>
<td>OMVS</td>
<td>CP</td>
<td>2</td>
<td>0</td>
<td>8.0</td>
<td>% Name</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WSP1S4F</td>
<td>SO</td>
<td>WASCR</td>
<td>CP</td>
<td>0</td>
<td>0</td>
<td>1.0</td>
<td>% Name</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AAP</td>
<td>6</td>
<td>0</td>
<td>1.0</td>
<td>% Name</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>% Name</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

© 2013 IBM Corporation
Sick System: Contention

- Arises when the various ENQs, latches, or locks used for serialization are not immediately available to one or more work units because they are already held by another work unit.

- The requesting work units then suffer delay.

- I think there are two categories:
  - Persistent – held too long (something is not working)
  - Competitive – lots of interest (workload)

- Generally need to stay focused on the holder, not the waiters.
Persistent Contention

- Work unit holds serialization for unreasonably long time
- Perhaps lots of work to do while serialized
  - Spikes?
  - Increase in volume?
- Perhaps work items are taking longer than normal
  - Is work unit running? enough?
  - Impacted by delays in some service it uses?
  - Hardware issues? (device recovery, retry, timeouts, ....)
  - Software issues? (error recovery, dumping, …)
- Perhaps a little of both
  - Workload changes create more work and this application comes up short on resources needed to run effectively
Competitive Contention

- Work is running, but response/transaction times impacted
- In effect, the pieces don't play well together
- Tends to arise as the result of workload changes

May be hard to detect and/or diagnose
- The contention does not last long enough to see it, or
- Constantly being released and acquired but always looks held when observed
  - Without data describing duration of the acquisition, this might seem to be persistent contention
- Likely needs application specific understanding to figure out how the various pieces interact with each other
  - Are design changes needed?
Sick System: Contention

- **ENQ contention**
  - Often arises as the result of submitting some new batch job that serializes same resource as existing work/job, or
  - Submitting multiple batch jobs that need to use the same data sets

- **Latch contention**
  - Often arises as the result of work load changes
  - USS, RRS, Logger are examples of users of latches
  - Latches are local but can induce sympathy sickness:
    - Get latch, send message, <delay?>, get response, release latch
    - <delay?> Get latch, formulate response, send response, release latch
  - Can give rise to a tangled web of sympathy sickness which can be quite difficult to diagnose (stay focused on the holder)
Sick System: Contention ...

- **Local Lock**
  - Some applications use local lock for serialization
  - Many system services often need it as well
    - GETMAIN, ATTACH, ...
  - Is holder running? enough?
  - Long queue effects?

- **Spin Locks**
  - If RSM lock, could be sign of real storage shortage
  - Hardware errors?
  - Long queue effects?
  - Software defect?
Detecting contention

- Run Time Diagnostics
  - D GRS,C "Most common command to display contention"
  - D GRS,AN,BLOCKER “To find out the trouble makers”
  - D GRS,AN,WAITER “To find the victims”
  - D GRS,RES=(major_name,minor name) “To focus on a particular resource”

- RMF Serialization Delay Report
  - ENQ
  - Latches
  - Local lock (others as well)

- IPCS ANALYZE RESOURCE

- Messages
  - IEE601E “excessive spin loop”
  - IEE331A “excessive spin loop, how to proceed?”
  - IEE178I “spin loop recovery action”
# D GRS,C Sample Output

ISG343I 05.26.32 GRS STATUS 604

S=SYSTEM  SYSTIOT

<table>
<thead>
<tr>
<th>SYSNAME</th>
<th>JOBNAME</th>
<th>ASID</th>
<th>TCBADDR</th>
<th>EXC/SHR</th>
<th>STATUS</th>
</tr>
</thead>
<tbody>
<tr>
<td>MVSS</td>
<td>JES2</td>
<td>0031</td>
<td>00AC5D48</td>
<td>SHARE</td>
<td>OWN</td>
</tr>
<tr>
<td>MVSS</td>
<td>JES2</td>
<td>0031</td>
<td>00AC5598</td>
<td>SHARE</td>
<td>OWN</td>
</tr>
<tr>
<td>MVSS</td>
<td>JES2</td>
<td>0031</td>
<td>00ABD170</td>
<td>EXCLUSIVE</td>
<td>WAIT</td>
</tr>
<tr>
<td>MVSS</td>
<td>JES2</td>
<td>0031</td>
<td>00ABBD58</td>
<td>EXCLUSIVE</td>
<td>WAIT</td>
</tr>
<tr>
<td>MVSS</td>
<td>JES2</td>
<td>0031</td>
<td>00ABB058</td>
<td>EXCLUSIVE</td>
<td>WAIT</td>
</tr>
<tr>
<td>MVSS</td>
<td>JES2</td>
<td>0031</td>
<td>00AC5AB8</td>
<td>SHARE</td>
<td>WAIT</td>
</tr>
<tr>
<td>MVSS</td>
<td>JES2</td>
<td>0031</td>
<td>00ABD4C0</td>
<td>SHARE</td>
<td>WAIT</td>
</tr>
<tr>
<td>MVSS</td>
<td>JES2</td>
<td>0031</td>
<td>00AC5828</td>
<td>SHARE</td>
<td>WAIT</td>
</tr>
<tr>
<td>MVSS</td>
<td>JES2</td>
<td>0031</td>
<td>00ABB2B8</td>
<td>SHARE</td>
<td>WAIT</td>
</tr>
<tr>
<td>MVSS</td>
<td>JES2</td>
<td>0031</td>
<td>00ABBAC8</td>
<td>SHARE</td>
<td>WAIT</td>
</tr>
</tbody>
</table>

NO REQUESTS PENDING FOR ISGLOCK STRUCTURE

NO LATCH CONTENTION EXISTS
### RMF Mon III ENQ Delays

<table>
<thead>
<tr>
<th>Jobname</th>
<th>%</th>
<th>%</th>
<th>STAT Major/Minor Names (Scope)</th>
<th>%</th>
<th>Name/SYS STAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>DUMPSRV</td>
<td>1</td>
<td>1</td>
<td>EW SYSZDAE (SYSS)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Samples: 100  System: SD0  Date: 08/07/12  Time: 10.41.40  Range: 100 Sec

DLY -------------- Resource Waiting -------------- ---- Holding ----

DATASET                                   /SE0
Identify owner of offending spin lock

- **PSACLHS and PSACLHSE indicate which spin locks are held**
  - Use IHAPSA mapping macro to decode the lock held strings
  - Comments identify z/OS component that uses the lock
  - Each CPU has its own PSA

- **Need to get a dump and use IPCS**
  - IP STATUS shows PSACLHS for each CPU
  - Use IP LIST PSAn to format PSA for CPU n
    - PSACLHS is at offset x2F8
    - PSACLHSE is at offset x4C4
  - Message IEE178I identifies offending CPU

If you identify relevant spin lock, you can better route the PMR. If RSM lock, you might look for storage constraints.
Contestion considerations

- Contention is often induced by other problems
- Experience suggests:
  - Changing mix of batch jobs often induces ENQ contention
  - Spinning on RSM locks often implies real storage shortages
  - Workload changes often induce competitive contention
  - Lack of dispatch time can induce persistent contention
  - USS latch contention often arises due to issues on peer system

- So at this point in our methodology, if root cause is not obvious
  - *As it might be for batch jobs or real storage shortages*

- Take note of the contention, but continue diagnosis
  - *Who holds the resource? Why aren't they making progress?*
Sick System: DASD I/O issues

- Performance, response times, throughput, even functionality may be impacted if applications encounter errors or delays while accessing data on DASD.

- Experience suggests that potential for such delays exists with:
  - Synchronous mirroring
  - Slow DASD
  - Workload changes
    - Changes in request rates, device contention, ...

- Of particular concern for sysplex are impacts to
  - Couple Data Sets
  - Data sets used by Logger
  - Page packs
Detecting DASD I/O Issues

- Messages (see handout)
  - IOS complaints regarding I/O errors and timeouts
  - XCF complaints about couple data sets
  - Logger complaints

- RMF DASD I/O reports. Response time issues?
  - Direct Access DASD Activity – post processor report

- RESERVES
  - IOS071I issued if delayed due to RESERVE
  - Run ENQ/DEQ Monitor to proactively identify any RESERVES (ISGAUDIT)

- IPCS ANALYZE RESOURCE
  - Indicate any outstanding paging I/O?

Best Practice: Eliminate all RESERVES. Run GRS STAR mode. Convert all RESERVES to global resources.
Sick System: DASD I/O messages (IOS)

- IOS001E “some paths to device are inoperative”
- IOS002A “no paths to device”
- IOS050I “channel detected error”
- IOS051I “channel timeout”
- IOS052I “channel detected error, recovered and logged it”
- IOS071I “MIH timeout” (adjust MIH? HW error?)
- IOS079I “deleted queued request due to timeout”
- IOS075E “device has recurring MIH condition”
- IOS076E “MIH timeout” (various reasons, likely HW error)
- IOS077E “MIH timeout” (various reasons, likely HW error)
- IOS100I “boxed device”
- IOS101I “device boxed or forced offline”
- IOS102I “device boxed or forced offline”
- IOS107I “deferring boxing of device”
- IOS431I “someone holding reserve on device”
Sick System: DASD I/O messages (XCF)

- IXC244E “cannot use this sysplex CDS”
- IXC246E “CDS experiencing I/O delays”
- IXC255I “cannot use this function CDS”
- IXC259I “I/O error on CDS”
- IXC267E “processing without alternate CDS”
Sick System: DASD I/O messages (Logger)

- IXG114A “offload not making progress”
- IXG115A “fix offload problem. Terminate task?”
- IXG271I “logger experiencing delays”
- IXG272E “logger task delayed, what to do?”
- IXG310I “offload not making progress”
- IXG311I “offload not making progress”
- IXG312E “offload delayed, what to do?”
### Post Processor: Direct Access Device Activity

**DIRECT ACCESS DEVICE ACTIVITY**

z/OS V1R11  
SYSTEM ID NICF  
START 07/18/2012-18.46.00  INTERVAL 000.15.00  
RPT VERSION V1R11 RMF  
END 07/18/2012-19.01.00  CYCLE 0.500 SECONDS  

TOTAL SAMPLES = 1,800  IODF = 00  CR-DATE: 04/04/2012  CR-TIME: 13.06.35  ACT: ACTIVATE  

<table>
<thead>
<tr>
<th>GROUP</th>
<th>NUM</th>
<th>TYPE</th>
<th>OF CYL</th>
<th>SERIAL</th>
<th>VOLUME</th>
<th>PAV</th>
<th>LCU</th>
<th>ACTIVITY</th>
<th>RESP</th>
<th>IOSQ</th>
<th>CMR</th>
<th>DB</th>
<th>PEND</th>
<th>DISC</th>
<th>CONN</th>
<th>DEV</th>
<th>DEV</th>
<th>DEV</th>
<th>NUMBER</th>
<th>ANY</th>
<th>MT</th>
</tr>
</thead>
<tbody>
<tr>
<td>NF34</td>
<td>0B00</td>
<td>33909</td>
<td></td>
<td></td>
<td>32760</td>
<td>MX0518</td>
<td>1.0H</td>
<td>0007</td>
<td>0.519</td>
<td>2.60</td>
<td>.000</td>
<td>.136</td>
<td>1.60</td>
<td>0.08</td>
<td>0.13</td>
<td>0.0</td>
<td>18.3</td>
<td>100.0</td>
<td>0.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NF34</td>
<td>0B01</td>
<td>33909</td>
<td></td>
<td></td>
<td>32760</td>
<td>MX0519</td>
<td>1.0H</td>
<td>0007</td>
<td>0.051</td>
<td>1.70</td>
<td>.000</td>
<td>.134</td>
<td>1.29</td>
<td>0.00</td>
<td>0.01</td>
<td>0.0</td>
<td>27.6</td>
<td>100.0</td>
<td>0.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NF34</td>
<td>0B02</td>
<td>33909</td>
<td></td>
<td></td>
<td>32760</td>
<td>MX0520</td>
<td>1.0H</td>
<td>0007</td>
<td>0.044</td>
<td>1.27</td>
<td>.000</td>
<td>.122</td>
<td>.246</td>
<td>0.00</td>
<td>0.01</td>
<td>0.0</td>
<td>17.3</td>
<td>100.0</td>
<td>0.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NF34</td>
<td>0B03</td>
<td>33909</td>
<td></td>
<td></td>
<td>32760</td>
<td>MX0521</td>
<td>1.0H</td>
<td>0007</td>
<td>0.031</td>
<td>4.00</td>
<td>.000</td>
<td>.146</td>
<td>3.61</td>
<td>0.00</td>
<td>0.01</td>
<td>0.0</td>
<td>7.3</td>
<td>100.0</td>
<td>0.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NF44</td>
<td>0B04</td>
<td>33909</td>
<td></td>
<td></td>
<td>32760</td>
<td>NF4588</td>
<td>1.0H</td>
<td>0007</td>
<td>0.001</td>
<td>.256</td>
<td>.000</td>
<td>.128</td>
<td>.128</td>
<td>0.00</td>
<td>0.00</td>
<td>0.0</td>
<td>1.0</td>
<td>100.0</td>
<td>0.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NF44</td>
<td>0B05</td>
<td>33909</td>
<td></td>
<td></td>
<td>32760</td>
<td>NF4589</td>
<td>1.0H</td>
<td>0007</td>
<td>0.001</td>
<td>.256</td>
<td>.000</td>
<td>.128</td>
<td>.128</td>
<td>0.00</td>
<td>0.00</td>
<td>0.0</td>
<td>0.0</td>
<td>100.0</td>
<td>0.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NF34</td>
<td>0B06</td>
<td>33909</td>
<td></td>
<td></td>
<td>32760</td>
<td>MX0563</td>
<td>1.0H</td>
<td>0007</td>
<td>1.734</td>
<td>10.4</td>
<td>.000</td>
<td>.147</td>
<td>3.83</td>
<td>6.44</td>
<td>1.12</td>
<td>1.78</td>
<td>11.8</td>
<td>100.0</td>
<td>0.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NF34</td>
<td>0B07</td>
<td>33909</td>
<td></td>
<td></td>
<td>32760</td>
<td>MX0526</td>
<td>1.0H</td>
<td>0007</td>
<td>1.262</td>
<td>10.7</td>
<td>.000</td>
<td>.149</td>
<td>3.77</td>
<td>6.77</td>
<td>0.85</td>
<td>1.33</td>
<td>10.3</td>
<td>100.0</td>
<td>0.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NFX7</td>
<td>0B08</td>
<td>33909</td>
<td></td>
<td></td>
<td>10017</td>
<td>NFX825</td>
<td>1.0H</td>
<td>0007</td>
<td>0.001</td>
<td>.128</td>
<td>.000</td>
<td>.128</td>
<td>.000</td>
<td>0.00</td>
<td>0.00</td>
<td>0.0</td>
<td>1.0</td>
<td>100.0</td>
<td>0.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NFX7</td>
<td>0B09</td>
<td>33909</td>
<td></td>
<td></td>
<td>10017</td>
<td>NFX826</td>
<td>1.0H</td>
<td>0007</td>
<td>0.001</td>
<td>.128</td>
<td>.000</td>
<td>.128</td>
<td>.000</td>
<td>0.00</td>
<td>0.00</td>
<td>0.0</td>
<td>2.0</td>
<td>100.0</td>
<td>0.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NFX7</td>
<td>0B0A</td>
<td>33909</td>
<td></td>
<td></td>
<td>10017</td>
<td>NFX827</td>
<td>1.0H</td>
<td>0007</td>
<td>0.001</td>
<td>.256</td>
<td>.000</td>
<td>.128</td>
<td>.128</td>
<td>0.00</td>
<td>0.00</td>
<td>0.0</td>
<td>2.0</td>
<td>100.0</td>
<td>0.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NFX7</td>
<td>0B0B</td>
<td>33909</td>
<td></td>
<td></td>
<td>10017</td>
<td>NFX828</td>
<td>1.0H</td>
<td>0007</td>
<td>0.014</td>
<td>2.43</td>
<td>.000</td>
<td>.148</td>
<td>2.08</td>
<td>0.207</td>
<td>0.00</td>
<td>0.00</td>
<td>0.0</td>
<td>6.0</td>
<td>100.0</td>
<td>0.0</td>
<td></td>
</tr>
</tbody>
</table>
### RMF Mon III Device Delays

**RMF V1R11 Device Delays**

**Samples:** 100  
**System:** SD0  
**Date:** 08/07/12  
**Time:** 14.00.00  
**Range:** 100 Sec

<table>
<thead>
<tr>
<th>Jobname</th>
<th>C</th>
<th>Class</th>
<th>%</th>
<th>%</th>
<th>%</th>
<th>VOLSER</th>
<th>%</th>
<th>VOLSER</th>
<th>%</th>
<th>VOLSER</th>
<th>%</th>
<th>VOLSER</th>
</tr>
</thead>
<tbody>
<tr>
<td>IXGLOGR</td>
<td>S</td>
<td>SYSTEM</td>
<td>1</td>
<td>12</td>
<td>12</td>
<td>1</td>
<td>LGR10Q</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Experience suggests that the system will not be running well if the following components are having issues (no particular order)

- Logger
- RACF
- JES3
- JES2
- RRS
- Unix System Services
- Consoles
- GRS
- SMF

Certainly others, but these tend to be most prevalent with respect to having single system issues that impact the sysplex.
Detecting Sick Components

- Diagnostic Data
  - LOGREC entries
  - Dumps

- RMF Monitor III
  - Job Delay
  - Processor Delay

- Component messages (see handout)
- XCF Messages (see handout)
- XES Messages (see handout)

The XCF/XES messages point at components who may be sick. They generally do not imply that XCF/XES is sick.
Detecting Sick Components: Dumps

- IEA045I “SVC dump started”
- IEA145E “messages and commands lost, may need to reissue”
- IEA611I “finished a dump”
- IEA794I “captured SVC dump”
- IEA799I “unable to automatically allocate SVC dump data set”
- IEA911E “finished a dump”
- IEE711I “unable to take dump”

*If a component is creating dumps, it detected some sort of problem. It suggests that the component is sick. Likely a good place to focus.*
RMF Monitor III: Job Delay

RMF V1R11 Job Delays

Samples: 100     System: SD0   Date: 08/07/12   Time: 14.28.20   Range: 100   Sec

Job: XCFAS     Primary delay: Job is waiting to use the processor.

Probable causes: 1) Higher priority work is using the system.  
2) Improperly tuned dispatching priorities.

------------------------- Jobs Holding the Processor -------------------------

Job: XCFAS     Holding: 1%     PROD Using: 1%     DEV Using: 0%

Job: CID1GA02   Holding: 1%     PROD Using: 2%     DEV Using: 0%

Job: CID1GA03   Holding: 1%     PROD Using: 3%     DEV Using: 0%

--------------------------- Job Performance Summary ---------------------------

Service       WFL -Using%- DLY IDL UKN ---- % Delayed for ---- Primary
CX ASID Class  P Cr %  PRC DEV % % % PRC DEV STR SUB OPR ENQ Reason
S 0006 SYSTEM  1  50 1 0 1 0 99 1 0 0 0 0 0 XCFAS
## RMF Monitor III Delay Report

Samples: 100
System: SD0
Date: 08/07/12
Time: 13.05.00
Range: 100 Sec

<table>
<thead>
<tr>
<th>Service</th>
<th>WFL USG DLY IDL UKN</th>
<th>% Delayed for</th>
<th>Primary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>CX Class</td>
<td>Cr % % % % %</td>
<td>PRC DEV STR SUB OPR ENQ Reason</td>
</tr>
<tr>
<td>------------</td>
<td>---------------------</td>
<td>---------------</td>
<td>---------</td>
</tr>
<tr>
<td>NFAGEN</td>
<td>T TSO</td>
<td>0 0 17 15 2 0 0 0 17 0 0</td>
<td>HSM</td>
</tr>
<tr>
<td><em>MASTER</em></td>
<td>S SYSTEM</td>
<td>0 0 1 0 99 0 1 0 0 0 0</td>
<td>HSMCDS</td>
</tr>
<tr>
<td>CID1GT01</td>
<td>SO CICSRGN</td>
<td>63 5 3 0 93 3 0 0 0 0 0</td>
<td>CID1GA03</td>
</tr>
<tr>
<td>IXGLOGR</td>
<td>S SYSTEM</td>
<td>92 11 1 0 88 0 1 0 0 0 0</td>
<td>LGR10M</td>
</tr>
<tr>
<td>SMSVSAM</td>
<td>S SYSTEM</td>
<td>100 1 0 0 99 0 0 0 0 0 0</td>
<td></td>
</tr>
<tr>
<td>VTAM44ST</td>
<td>S SYSSTC</td>
<td>100 1 0 0 99 0 0 0 0 0 0</td>
<td></td>
</tr>
<tr>
<td>DFHSM</td>
<td>S SYSSTC</td>
<td>100 1 0 0 99 0 0 0 0 0 0</td>
<td></td>
</tr>
<tr>
<td>TCPCST</td>
<td>SO SYSSTC</td>
<td>100 2 0</td>
<td></td>
</tr>
<tr>
<td>CID1GA01</td>
<td>SO CICSRGN</td>
<td>100 4 0</td>
<td></td>
</tr>
<tr>
<td>CID1GA02</td>
<td>SO CICSRGN</td>
<td>100 3 0</td>
<td></td>
</tr>
<tr>
<td>CID1GA03</td>
<td>SO CICSRGN</td>
<td>100 1 0</td>
<td></td>
</tr>
<tr>
<td>CID1GA04</td>
<td>SO CICSRGN</td>
<td>100 1 0</td>
<td></td>
</tr>
</tbody>
</table>

**User NFAGEN delayed waiting for HSM to recall datasets.**
Sick System Components: Logger

- IXG114A “offload not making progress”
- IXG115A “fix offload problem. Terminate task?”
- IXG261E “logger CDS running out of space”
- IXG262A “logger CDS has run out of space”
- IXG271I “logger experiencing delays”
- IXG272E “logger task delayed, what to do?”
- IXG310I “offload not making progress”
- IXG311I “offload not making progress”
- IXG312E “offload delayed, what to do?”
Sick System Components: RACF

- IRRX004A
- IRRX017I
- IRRC022I
- IRRC023I
- IRRC024I
- IRRC025I
- IRRC026I
- IRRC032I
- IRRC033I
- IRLI013I
- IRLI081I
- IRRN081I
- IRRR081I
- IRRQ081I

Sorry, I did not get these interpreted yet
Sick System Components: RACF ....

- ICH501I “RACF not active”
- ICH505A “RACF initialization failed”
- ICH600A “How should RACF proceed?”
  - ICH586A “RACF DB will be corrupted if mixed usage”
  - ICH588A “RACF DB will be corrupted if mixed usage”
  - ICH589A “RACF DB will be corrupted if mixed usage”
  - ICH590A “RACF DB will be corrupted if being shared”
  - ICH591A “RACF DB will be corrupted if being shared”
Sick System Components: JES3

- IAT1105 “file directories in use exceeds threshold”
- IAT7134 “console buffers in use exceeds threshold”
- IAT1016 “spool partition full, sysout processing suspended”
- IAT1017 “spool partition full, job selection suspended”
- IAT1018 “FYI, spool partition full”
- IAT8054 “queue being held”
- IAT6368 “checkpoint data set too small”
- IAT2008 “start cmd fails, likely due to storage or paging constraints”
- IAT6341 “running out of job numbers”
- IAT6395 “jobs are waiting”
- IAT6850 “too many WTO’s”
Sick System Components: JES2

- $HASP9207 “checkpoint lock held for a long time”
- $HASP292 “waiting for response during checkpoint write”
Sick System Components: RRS

- ATR247E “severe RRS error”
- ATR248E “RRS waiting for Logger to recover”
- ATR249E “RRS waiting for Logger to recover given logstream”
- ATR202D “gap in logstream, how should RRS proceed?”
- ATR210E “gap in RM DATA logstream, may need to cold start RRS members”
- ATR225D “cancel delayed, how should RRS proceed?”
- ATR226D “memterm delayed, how should RRS proceed?”
- ATR227D “cancel delayed, how should RRS proceed?”
- ATR228D “memterm delayed, how should RRS proceed?”
- ATR229D “cancel delayed, how should RRS proceed?”
- ATR230D “memterm delayed, how should RRS proceed?”
- ATR231D “cancel delayed, how should RRS proceed?”
- ATR232D “memterm delayed, how should RRS proceed?”
- ATR233D “cancel delayed, how should RRS proceed?”
- ATR234D “memterm delayed, how should RRS proceed?”
Sick System Components:
Unix System Services

- BPXB001E
- BPXF002I
- BPXF006I
- BPXF008I
- BPXF014D
- BPXF019I
- BPXF020I
- BPXF021I BPXF029E BPXF032D BPXF034I BPXF039I BPXF044I BPXF045A BPXF075I
  BPXF076I BPXF077S BPXF078W BPXF079S BPXF080I BPXF083I BPXF213E
- BPXF214E “unable to access BPXMCDS couple data set”
- BPXF215E “unable to access BPXMCDS couple data set”
- BPXF216E “file system partition cleanup delayed”
- BPXF217E “file system partition cleanup failed”
- BPXF218I BPXF221I BPXF222E BPXF226E BPXF230I BPXF242I BPXF243E BPXF244E
  BPXF245I BPXF247I BPXF249I BPXF252I BPXF253E BPXF254I BPXF255I BPXF256I BPXF257I
  BPXF259I BPXI004I BPXI005I BPXI016I BPXI017I BPXI018I BPXI019E BPXI026I BPXI031E
  BPXI027I BPXI028E BPXI029I BPXI030I BPXI032E BPXI033E BPXI035E BPXI036E AVAILABL
  BPXI039I BPXI040I BPXI043E BPXI055I BPXI056E BPXI058I BPXI068I BPXI060I BPXI061E
  BPXI062I BPXI064E BPXI065E BPXI066E BPXI067E BPXI068I BPXI075E BPXI076E BPXI077I
  BPXI078D BPXI082E BPXI083D BPXI084E BPXI085D BPXM048I BPXM054SE BPXM050E
  BPXM055D BPXM057E BPXM120D BPXM056E BPXN002I BPXN003E BPXO043I BPXP003E
  BPXP004E BPXP006E BPXP022E BPXP022I BPXP001I BPXP007E BPXP008E

Sorry, I did not get these interpreted yet
Sick System Components: Consoles

- CNZ2202E
- CNZ3003I
- CNZ3004E
- CNZ3005A
- CNZ3007I
- CNZ3009E
- CNZ3010I
- CNZ3014I
- CNZ4200I
- CNZ4201E
- CNZZ002E
- CNZZ007E
- CNZZ009E
- CNZZ014E
- CNZZ031E
- CNZZ033E

Sorry, I did not get these interpreted yet
Sick System Components: Consoles

- IEA145E “messages and commands lost, may need to reissue”
- IEA230E “WTOR buffer shortage”
- IEA231A “WTOR buffer shortage”
- IEA359E “retained action message buffer shortage”
- IEA360A “retained action message buffer shortage”
- IEA367A “multiple console support inoperative, please acknowledge”
- IEA404A “WTO buffer shortage”
- IEA405E “WTO buffer shortage”
- IEA555E “unable to restart delayed WTOR processor”
- IEA556I “peer system’s console support inoperative”
- IEA557A “need operator to respond to reply ID 0 WTOR”
- IEA652A “discarding messages due to WTO buffer shortage”
- IEA654A “sympathy sickness due to WTO buffer shortage on peer system”
Sick System Components: Consoles

- IEE141A “no master console, no alternates either”
- IEE624I “need to enable system console”
- IEE765E “syslog task failed”
- IEE767A “syslog buffer is full”
- IEE769E “system error in syslog”
- IEE775E “no storage for syslog buffer”
- IEE795I “syslog data loss”
- IEE806A “exceeded command limit”
- IEE824E “communication task failed”
Sick System Components: GRS

- ISG361A “waiting for list lock”
- ISG362I “finally got the list lock”

With “critical member” support in z/OS V1R12, the system may well be removed from the sysplex before it gets a chance to complain about the list lock

- White paper on Diagnosing GRS issues
  - z/OS GRS: Performance Considerations
  - www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP101725
Sick System Components: SMF

- IEE986E “SMF buffer space exceeds threshold”
Detecting Sick Components: XCF messages

- IXC430E “system has stalled group members”
- IXC431I “group member is stalled”
- IXC440E “stalled group member impacting other systems”
- IXC615I “terminating group member”
- IXC631I “stalled group member impacting peer system”
- IXC640E “peer stalled group member impacting me”

*Generally these messages suggest that someone other than XCF is sick. XCF however, can be impacted by these problems, which in turn can induce sympathy sickness. Stay focused on the stalled group member and the system where the member resides.*
Detecting Sick Components: XES messages

- IXL040E “structure connector not responding”
- IXL041E “structure connector not responding”
- IXL045E “having trouble getting connector SRBs to run”

Generally these messages suggest that someone other than XES is sick. XES itself is unlikely to be impacted by these problems, but they can in turn induce sympathy sickness. Generally stay focused on the stalled connector and the system where it resides. However, some connectors will appear to be unresponsive because they are “waiting” for a peer connector to complete some task. In such cases, one really needs application specific knowledge to determine whether the unresponsive connector is the culprit, or a victim. You want to stay focused on the culprit.
## Correlating messages and components

<table>
<thead>
<tr>
<th>Component</th>
<th>Prefix</th>
<th>Jobname</th>
<th>XCF Group</th>
<th>CDS</th>
<th>CF Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>XCF</td>
<td>IXC</td>
<td>XCFAS</td>
<td>SYSXCF</td>
<td>Sysplex ARM, SFM</td>
<td>IXC…</td>
</tr>
<tr>
<td>XES</td>
<td>IXL</td>
<td>(user's job)</td>
<td>IXCLO…</td>
<td>CFRM</td>
<td></td>
</tr>
<tr>
<td>GRS</td>
<td>ISG</td>
<td>GRS</td>
<td>SYSGRS</td>
<td>Sysplex</td>
<td>ISGLOCK (if star mode)</td>
</tr>
<tr>
<td>RACF</td>
<td>ICH</td>
<td></td>
<td>IRRXCF00</td>
<td>IRR</td>
<td></td>
</tr>
<tr>
<td>Consoles</td>
<td>CNZ</td>
<td>CONSOLES</td>
<td>SYSMCS</td>
<td></td>
<td>(via logger)</td>
</tr>
<tr>
<td>Logger</td>
<td>IXG</td>
<td>IXGLOGR</td>
<td>LOGR</td>
<td>lots</td>
<td></td>
</tr>
<tr>
<td>JES</td>
<td>HASP or IAT</td>
<td>JES2</td>
<td>JESXCF</td>
<td></td>
<td>checkpoint</td>
</tr>
<tr>
<td>RRS</td>
<td>ATR</td>
<td>RRS (maybe)</td>
<td>SYSATR</td>
<td></td>
<td>(via logger)</td>
</tr>
<tr>
<td>Unix System Services</td>
<td>BPX</td>
<td></td>
<td>SYSBPX</td>
<td>BPXMCDS</td>
<td></td>
</tr>
</tbody>
</table>
At this point in our sysplex diagnosis methodology, we may have determined that a particular component appears to be sick. *But the component might have dependencies or interactions with other systems in the sysplex.* *Without component knowledge, we may not be able to tell whether the sick component is actually suffering from sympathy sickness.*

So if root cause is not obviously a local issue, take note of the component and continue diagnosis.
Problem Taxonomy

- Dead System
- Sick System
- Sysplex Fabric
  - Sysplex Componentry
    - Coupling Facility
    - Signalling Service
    - Couple Data Sets
    - External Time Reference
- Configuration / Capacity
- Software Issues
Sysplex Fabric

- Consists of the various cables, links, channel paths, CHPIDs, and adapter cards that provide physical access to the sysplex componentry
- Along with the parmlib members and policies that govern logical access to the sysplex componentry
- Performance, response time, throughput, even functionality may be impacted if access to the sysplex componentry is impeded by
  - Error prone connections
  - Loss of connectivity
  - Outright loss or lack of access
- So now we make sure that the expected sysplex componentry exists and is accessible to each system in the sysplex
Sysplex Fabric: Couple Data Sets

- **Physical access**
  - See: “Sick System: DASD I/O issues”
  - IODF

- **Logical access**
  - COUPLExx COUPLE statement for sysplex CDS
  - COUPLExx DATA statements for function CDS's
  - SETXCF COUPLE command

- **Detecting Fabric Problems**
  - D XCF,COUPLE,TYP=ALL (see handout)
  - XCF Messages (see handout)
  - IOS messages (see handout)
  - Exploiter Messages (see handout)

Do the various CDS exist? Have an alternate? Accessible to all systems?

Using right COUPLExx? What changed?
You run risk of a syplex outage if you don't have an alternate Sysplex CDS.
Sysplex Fabric: Couple Data Sets
D XCF,COUPLE,TYPE=xxx ...

ARM COUPLE DATA SETS
PRIMARY  DSN: UTCXCF.SVPLEX1.ARMR14.PRI
           VOLSER: X1CPLA   DEVN: B010
           FORMAT TOD   MAXSYSTEM
           04/11/2003 21:02:00   32
           ADDITIONAL INFORMATION:
           FORMAT DATA
           VERSION 1, HBB7707 SYMBOL TABLE SUPPORT
           POLICY(8) MAXELEM(400) TOTELEM(600)

ALTERNATE  DSN: UTCXCF.SVPLEX2.ARMR14.PRI
            VOLSER: X1CPLA   DEVN: B010
            FORMAT TOD   MAXSYSTEM
            04/11/2003 21:02:00   32
            ADDITIONAL INFORMATION:
            FORMAT DATA
            VERSION 1, HBB7707 SYMBOL TABLE SUPPORT
            POLICY(8) MAXELEM(400) TOTELEM(600)

ARM IN USE BY ALL SYSTEMS

Should have an alternate for availability

Should normally be accessible by all systems
 Sysplex Fabric: Couple Data Sets

- IXC244E “cannot use this sysplex CDS”
- IXC246E “CDS experiencing I/O delays”
- IXC253I “CDS removed from service”
- IXC255I “cannot use this function CDS”
- IXC256A “cannot remove CDS until these systems respond”
- IXC259I “I/O error on CDS”
- IXC267E “processing without alternate CDS”
Sysplex Fabric: Couple Data Sets (IOS messages if relevant to CDS)

- See “Sick System: DASD I/O Messages (IOS)” for list of IOS messages that might be relevant to DASD I/O Issues

- You need to know where the CDS resides so that you can consider only those IOS messages related to the channel paths and devices that are relevant to the various CDS of interest
Sysplex Fabric: Function Couple Data Sets

- ARM
  - IXC807I “rejected use of alternate CDS”
  - IXC808I “no access to CDS”
  - IXC809I “lost access to CDS”
  - IXC810I “unable to use CDS”
Sysplex Fabric: Function Couple Data Sets

- **BPXMCDS**
  - BPXF214E “unable to access CDS”
  - BPXF215E “unable to access CDS”
  - BPXF226E “rejected use of CDS”
  - BPXF230I “rejected use of alternate CDS”
  - BPXI046I “unable to initialize new primary CDS”
  - BPXF050I “I/O errors accessing CDS”
  - BPXF058I “lost access to CDS”
Sysplex Fabric: Function Couple Data Sets

- **CFRM**
  - IXC220W “lost access to CDS”
  - IXC520I “not using CDS”

- **LOGR**
  - IXG047I “unable to access CDS”
  - IXG054A “no CDS available”

- **SFM**
  - IXC610I “unable to use CDS”

- **WLM**
  - IWM047E “unable to access CDS”
  - IWM048E “no CDS”
Sysplex Fabric: Coupling Facility

- Physical Access
  - CF
  - Links

- Logical Access
  - CFRM Policy

- Detecting Fabric Issues
  - D CF
  - D XCF,CF
  - D XCF,POLICY,TYP=CFRM
  - XCF Messages (see handout)
  - IXLERLOG LOGREC entries for link issues (see handout)
  - Check for flashing icons on the HMC

  - HMC command to display CF

Does the CF exist?
Accessible from all systems?
Using the right CFRM policy?

physically connected?
logically connected?
right policy?
hardware issues?
CF exist? Does it respond?
CF becomes inaccessible to z/OS if:
- CF image is reset
- CF aborts
- CF suffers power outage
- CEC on which CF resides goes down
- Loses connectivity to the CF

z/OS issues messages IXL157I and IXC517I if unable to access the CF. But two possible scenarios:
- Could be due to loss of connectivity
- Could be due to CF going down

They both look the same to z/OS, but root cause is different
Sysplex Fabric: Coupling Facility

- If CF does not respond to z/OS within 2 seconds, z/OS recycles the link under the assumption that there has been some sort of communication issue.

- If z/OS recycles all the links at the same time, connectivity to the CF is lost.
  - The CF may be up, but z/OS “disconnected” so to speak.
D CF, CFNAME=xxxx   ..... 

If CF is not accessible, D CF may not show the CF at all. Alternatively, one or more sections of the output will not be available.

| COUPLING FACILITY HAS ONLY ONE ONLINE CF | CF REQUEST TIME ORDERING: REQUIRED |
| COUPLING FACILITY SPACE CONFIGURATION |  |
| IN USE | FREE | TOTAL |  |
| CONTROL SPACE | 0 M | 0 M | 0 M |  |
| NON-CONTROL SPACE | 0 M | 0 M | 0 M |  |

**NO COUPLING FACILITY SPACE DATA AVAILABLE**

<table>
<thead>
<tr>
<th>PATH</th>
<th>PHYSICAL</th>
<th>LOGICAL</th>
<th>CHANNEL TYPE</th>
<th>AID</th>
<th>PORT</th>
</tr>
</thead>
<tbody>
<tr>
<td>03</td>
<td>CIB</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>C0</td>
<td>NOT OPERATIONAL</td>
<td>ONLINE</td>
<td>CIB</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>C1</td>
<td>0738</td>
<td>ONLINE</td>
<td>CIB</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

**NO PATH STATUS AVAILABLE**

<table>
<thead>
<tr>
<th>COUPLING FACILITY SUBCHANNEL STATUS</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>TOTAL: 96</td>
<td>IN USE: 7</td>
</tr>
</tbody>
</table>

**NO COUPLING FACILITY DEVICE STATUS AVAILABLE**

<table>
<thead>
<tr>
<th>20E  / 25E0</th>
<th>20F0 / 25E0</th>
<th>23E0 / 23ED</th>
<th>25E0 / 25ED</th>
<th>25F0 / 25ED</th>
</tr>
</thead>
<tbody>
<tr>
<td>F432 / 25ED</td>
<td>F433 / 25EE</td>
<td>F434 / 25EF</td>
<td>F435 / 25F0</td>
<td></td>
</tr>
</tbody>
</table>
CF will not be accessible if “request time ordering” is required, but not enabled. Unless you are really down level on hardware, I would expect the function is installed. Thus “not enabled” likely implies that z/OS image and CF are either not in the same CTN, or are having other ETR related issues.
D XCF,CF,CFNAME=SVT1CF1

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>IXC362I</td>
<td>22.06.42 DISPLAY XCF 411</td>
</tr>
<tr>
<td>CFNAME:</td>
<td>SVT1CF1</td>
</tr>
<tr>
<td></td>
<td>COUPLING FACILITY:</td>
</tr>
<tr>
<td></td>
<td>002817.IBM.02.0000000F8E66</td>
</tr>
<tr>
<td></td>
<td>PARTITION: 09 CPCID: 00</td>
</tr>
<tr>
<td></td>
<td>SITE: SITE1</td>
</tr>
<tr>
<td></td>
<td>POLICY DUMP SPACE SIZE: 6000 K</td>
</tr>
<tr>
<td></td>
<td>ACTUAL DUMP SPACE SIZE: 6 M</td>
</tr>
<tr>
<td></td>
<td>STORAGE INCREMENT SIZE: 1 M</td>
</tr>
<tr>
<td></td>
<td>CONNECTED SYSTEMS:</td>
</tr>
<tr>
<td></td>
<td>N64   N65   N66   N67</td>
</tr>
<tr>
<td></td>
<td>MONITORING SYSTEM: N64</td>
</tr>
<tr>
<td></td>
<td>STRUCTURES:</td>
</tr>
<tr>
<td></td>
<td>DSVPLX1_SCA</td>
</tr>
<tr>
<td></td>
<td>IRRXCF00_B001</td>
</tr>
<tr>
<td></td>
<td>IRRXCF00_B002</td>
</tr>
<tr>
<td></td>
<td>ISTGENERIC(NEW)</td>
</tr>
<tr>
<td></td>
<td>IXCplex_PATH2</td>
</tr>
<tr>
<td></td>
<td>IXCplex_PATH4</td>
</tr>
<tr>
<td></td>
<td>LOGGER_STR1</td>
</tr>
<tr>
<td></td>
<td>SYSZWLM_8E662817(NEW)</td>
</tr>
</tbody>
</table>

Shows CFs defined in the active CFRM policy. If CF not shown, you have a policy issue.

Right policy?

CF defined in the policy?

With the right Node ID?
D XCF,POLICY

IXC364I 22.40.25 DISPLAY XCF 927
TYPE: ARM
POLICY NOT STARTED
TYPE: BPXMCDS
    NOT SUPPORTED BY
TYPE: CFRM
    POLNAME: IXCPOLX1
    STARTED: 07/19/2012 21:19:19
    LAST UPDATED: 07/19/2012 11:24:45
TYPE: LOGR
    NOT SUPPORTED BY DISPLAY XCF,POLICY
TYPE: SFM
    POLNAME: IXCSFMX1
    STARTED: 04/17/2012 19:24:49
    LAST UPDATED: 04/17/2012 19:24:08

SYSPLEX FAILURE MANAGEMENT IS ACTIVE
TYPE: WLM
    NOT SUPPORTED BY DISPLAY XCF,POLICY

CFRM Policy started? Right name? Right instance?
Sysplex Fabric: Coupling Facility

- IXC501A “use this CF or not?”
- IXC517I “system using this CF” was the problem before this?
- IXC518I “system not using this CF” but should it be?
- IXC519E “coupling facility is damaged”
- IXC512I “CFRM policy change pending”
- IXL051I “CF dump was taken for hardware support to review”

CF went down if dump was disruptive

- IXL044I “experiencing repeated IFCCs on path to CF”
- IXL157I “path to CF now operational”
- IXL158I “path to CF is not operational”

Is CF connected to same ETR as z/OS images?

- IXL159E “detected error with notification vectors”

- IXL160E “not using CF due to lack of request time ordering”
- IXL162E “not using CF due to lack of request time ordering”
IXC501A “Use This CF or Not?”

- Operations must be very, very careful with IXC501A
- System programmer must be very careful with CFRM policy
- z/OS is asking operator to confirm whether this sysplex is supposed to be using the indicated CF
- The message is issued when it appears that the CF might be in use by some other sysplex
- If the operator responds “yes”, this sysplex will take over ownership of the CF
- If the CF is actually in use by some other live sysplex, that other sysplex will lose access to the CF
  - That other sysplex could potentially suffer a sysplex wide outage as a result of losing its CF
Sysplex Fabric: Coupling Facility
Symptom Record in LOGREC

Physical link errors may arise due to:
- Bad or pinched cables
- Cables not well seated
- Dirt

Look for IXLERLOG
Scroll down to find issue IFCC if CF link issues
Other issues possible.
HMC: Operations System Messages For CF

2011034 15:59:20 CF0010I Coupling Facility is active with:
  1 CP
  14 CF Receiver Channels
  10 CF Sender Channels
  101956 MB of allocatable storage

2011040 15:50:52 => help
2011040 15:58:52 CF0400I CF commands:
CONFIGURE - take CHPID on or off line.
CP - take CF on or off line.
DISPLAY - show resources.
HELP <command> - command specific help.
MODE - set volatility mode.
RIDEOUT - set power failure rideout time.
SHUTDOWN - terminate CF operation.
TIMEZONE - set timezone offset.
TRACE - set trace control.
DYNDISP - turn Dynamic CF Dispatching On or Off.
MT0 - turn mt0 table on.
CFDUMP - force non-disruptive dump.
NDDUMP - nddump_command.

2011040 15:59:00 => cf dump
2011040 15:59:00 CF00000I A non-disruptive dump was taken by the CF.
Sysplex Fabric: CF Structures

- **Physical Access**
  - Do expected structures exist?

- **Logical Access**
  - CFRM Policy
  - IXLCONN

- **Detecting Fabric Issues**
  - D XCF,CF
  - D XCF,POLICY,TYP=CFRM
  - RMF CF Activity Report
  - XCF/XES Messages (see handout)

Which structures do you normally run with?
Do they exist?
Running with expected CFRM policy?
Any failed structure allocations?
Do your structures exist? In the expected/desired CF?

**Sysplex Fabric: CF Structures**

D XCF, CF, CFNAME=SVT1CF1

**IXC362I 22.06.42 DISPLAY XCF 411**

**CFNAME: SVT1CF1**

**COUPLING FACILITY:** 002817.IBM.02.00000000F8E66

**PARTITION:** 09  **CPCID:** 00

**SITE:** SITE1

**POLICY DUMP SPACE SIZE:** 6000 K

**ACTUAL DUMP SPACE SIZE:** 0 K

**STORAGE INCREMENT SIZE:**

**CONNECTED SYSTEMS:**

N64  N65  N66  N67

**MONITORING SYSTEM:** N64

**STRUCTURES:**

DBSVPLX1_SCA  IRRXCF00_B001  IRRXCF00_B002

ISTGENERIC(NEW)  IXCplex_PATH2  IXCplex_PATH4

LOGGER_STR1  SYSZWLM_8E662817(NEW)
Sysplex Fabric: CF Structures
D XCF,POLICY

D XCF,POLICY,TYPE=ALL
IXC364I 22.40.25 DISPLAY XCF 927
TYPE: ARM
POLICY NOT STARTED
TYPE: BPXMCDS
NOT SUPPORTED BY
TYPE: CFRM
POLNAME: IXCPOLX1
STARTED: 07/19/2012 21:19:19
LAST UPDATED: 07/19/2012 11:24:45
TYPE: LOGR
NOT SUPPORTED BY DISPLAY XCF,POLICY
TYPE: SFM
POLNAME: IXCSFMX1
STARTED: 04/17/2012 19:24:49
LAST UPDATED: 04/17/2012 19:24:08

SYSPLEX FAILURE MANAGEMENT IS ACTIVE
TYPE: WLM
NOT SUPPORTED BY DISPLAY XCF,POLICY

CFRM Policy started?
Right name?
Right instance?
Sysplex Fabric: CF Structures
Messages

- IXL013I “application unable to connect to structure”
- IXL014I “application unable to connect to structure”
  - An application must connect in order to use a structure. The failure text might indicate that the structure could not be created at all.
- IXC453I “not enough signaling paths”
- IXC454I “unable to establish signaling connectivity”
- IXC455D “reply with interval or retry to re-initialize XCF”
  - You tend to see these messages on an IPLing system. Many root causes, but frequently implies that XCF signalling structures are not accessible.
- Applications may well issue their own messages to complain if their structures are not accessible
Sysplex Fabric: Signalling Paths

- **Physical Access**
  - See “Sysplex Fabric: CF Structures”
  - CTC devices

- **Logical Access**
  - COUPLExx PATHOUT statements, or SETXCF START,PATHOUT
  - COUPLExx PATHIN statements, or SETXCF START,PATHIN

- **Detecting Fabric Issues**
  - D XCF,PO
  - D XCF,PI
  - Messages (see handout)
  - Apply “Sysplex Fabric: CF Structures” to signal structures
  - Check for IOS messages related to signal path CTC devices

Are CF structures used for signalling accessible?
Are CTC devices used for signalling online and operational?
Sysplex Fabric: Signalling Paths

D XCF,PO – summary view

D XCF,PO
IXC355I 22.32.57 DISPLAY XCF 867
PATHOUT TO SYSNAME: ???????? - PATHS NOT CONNECTED TO OTHER SYSTEMS
DEVICE (LOCAL/REMOTE): C140/???? C141/???? C142/???? C150/????
C151/???? C152/???? D140/???? D141/????
D142/???? D150/???? D151/???? D152/????

PATHOUT TO SYSNAME: N64
DEVICE (LOCAL/REMOTE): C110/C144 C111/C145 C112/C146 D110/D144
D111/D145 D112/D146
STRNAME:
IXCPLEX_PATH1 IXCPLEX_PATH2
IXCPLEX_PATH3 IXCPLEX_PATH4

PATHOUT TO SYSNAME: N65
DEVICE (LOCAL/REMOTE): C120/C144
D121/D145
STRNAME:
IXCPLEX_PATH
IXCPLEX_PATH

PATHOUT TO SYSNAME: N66
DEVICE (LOCAL/REMOTE): C130/C144

Shows perspective of local system
Paths not connected but should be?
Missing expected paths?
Paths to every other system?

“connected” does not imply “working”
### Sysplex Fabric: Signalling Paths

#### D XCF,PO – detail

<table>
<thead>
<tr>
<th>LOCAL DEVICE</th>
<th>REMOTE SYSTEM</th>
<th>PATHOUT STATUS</th>
<th>REMOTE PATHIN RETRY</th>
<th>MAXMSG CLASS</th>
<th>TRANSPORT</th>
</tr>
</thead>
<tbody>
<tr>
<td>C110</td>
<td>N64</td>
<td>WORKING</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Local** | **Remote** | **Pathout** | **Pathin** | **Status** | **Transport**
---|---|---|---|---|---|
C110 | C144 | N64 | WORKING | | |

- Shows perspective of local system
- “working” likely good, but not conclusive.
- (If doubts, use detail D XCF,PI to verify flow from the perspective of target system)
- Nonzero “pending” suggests delay or spike.

### D XCF,PO,STRNAME=IXCPELEX_PATH1

<table>
<thead>
<tr>
<th>STRNAME</th>
<th>REMOTE PATHOUT SYSTEM STATUS</th>
<th>TRANSFER BUFFER MSGBUF SIGNAL MXFER</th>
</tr>
</thead>
<tbody>
<tr>
<td>IXCPELEX_PATH1</td>
<td>N64 WORKING</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N65 WORKING</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N66 WORKING</td>
<td></td>
</tr>
</tbody>
</table>

**List** | **System** | **Status** | **Transfer Buffer** | **Msgbuf** | **Signal** | **Mxfer**
---|---|---|---|---|---|---|
IXCPELEX_PATH1 | N64 | WORKING | 0 | 8124 | 110 | 239 | 1667 |
| N65 | WORKING | 0 | 8124 | 110 | 299 | 1012 |
| N66 | WORKING | 0 | 8124 | 56 | 4496 | 2703 |
To check flow, issue detail pathin on the receiving side of signal path. Successive displays should show changes (unless path unused).

Pending delivery of 4 is typical for CTC. 0 typical for list path. Bigger suggests msg exit delays.

Signal numbers on pathout side and pathin side should be close. Suggests transfer delay if not.
Sysplex Fabric: Signalling Paths

- IXC458I “stopped signalling path”
- IXC459I “stopped signalling path unconditionally”
- IXC467I “restarting or stopping or rebuilding signalling path”

- IXC453I “not enough signaling paths”
- IXC454I “unable to establish signalling connectivity”

- Other messages
  - IOS for CTC devices used for signal paths
  - XCF/XES messages for CF or structures used for signal paths
Sysplex Fabric: External Time Reference

- **Physical Access**
  - Coordinated Time Network
  - Timing links

- **Logical Access**
  - CTN ID

- **Detecting Fabric Issues**
  - D XCF, SYSPLEX, ALL
  - D ETR, DATA
  - Messages (see handout)
  - HMC

If z/OS image loses access to ETR, the system is in a wait-state.

So my “fabric detection” is either
- Proactive prevention
- Post mortem analysis

If lose ETR, “live” investigation is via the HMC and other systems that still have ETR access

Timer links operational?
Is the CTN ID correct?
PTS operational? BTS? Arbiter?
Is CTS the one you want?
Sysplex Fabric: External Time Reference

- IEA394A “lost access to ETR (STP)”
- IEA015A “lost access to ETR (sysplex timer)”

- IXC406I “not connected to same ETR”
- IXC438I “new or changed CTNID”
- IXL160E “required request time ordering not enabled”
- IXL162E “required request time ordering will never be enabled”

- IEA031I “STP alert was issued to HMC”
- IEA395I “switched to backup time server”

- And of course, the HMC
Methodology Recap

- We have eliminated any local issues that might be root cause
  - Wait-states, constraints, device errors
  - Explainable contention
  - Lack of access to sysplex componentry

- We may be reserving judgment on local issues that might arise as the result of sympathy sickness
  - Unexplained contention
  - Sick components

- Or we might not have any potential culprits identified, but still believe that a problem exists
Problem Taxonomy

- Dead System
- Sick System
- Sysplex Fabric
- Sysplex Componentry
  - Coupling Facility
  - Signalling Service
  - Couple Data Sets
  - External Time Reference
- Configuration / Capacity
- Software Issues
Sysplex Componentry Considerations

- The sysplex components are rather intertwined and mutually dependent on each other. For example:
  - A CF structure might be used for signalling
  - But signals need to be sent to manage the structures in the CF
  - Management of the CF needs access to CDS
  - But signals need to be sent to manage the CDS
  - And the CDS defines who the signalling service can talk to

- I am largely ignoring this complexity and pretending that we can look at the components in isolation
  - But that could be a vast oversimplification if you happen to be rebuilding a signal structure while in the midst of performing a PSWITCH to an alternate CFRM CDS
Methodology Recap

- At this point, we should have the following conditions:
  - All systems are operational
  - All sysplex componentry is accessible via error free connections

- We suspect there may be issues with sysplex componentry
  - Performance issues
  - Configuration or capacity issues

- So we need a detailed understanding of the how the sysplex componentry works so that we can determine
  - How such problems might arise
  - How they might be observed
  - How they might be resolved
Coupling Facility Request Processing

1) Application issues request
   Sync vs Async?
   Pick subchannel (queue or spin?)
2) Initiate operation
3) CSS picks path, sends operation
4) CF receives and processes operation. Sends results.
5) z/OS sees operation completion
6) Application gets request results
CF Service Time Considerations

- Delay Time is time spent waiting for a subchannel
- Service Time is delta between sending operation and observing its completion
- Service time impacted by
  - Coupling Facility
    - Technology
    - Utilization
    - Contention
  - CF Links
    - Technology
    - Path busy conditions
    - Distance
  - Completion recognition

You get one number that encompasses all these factors.
One or more could be the issue.
Coupling Facility Request Processing

1. Delay Time
2. Service Time
3. Operation Completion
4. Request Completion
5. CF responsive?
6. CF technology?
7. CFCC issues?
CF Responsive?

- LPAR configuration must allow the CF to be sufficiently responsive so that the polling loop can run to receive commands in a timely manner
  - Dedicated CPs recommended
  - Shared CPs need sufficient weight
  - >50% for simplex or user managed duplexing
  - >95% for system managed duplexing
- Dynamic Dispatch can give rise to erratic, elongated response times
- Sharing CPs between z/OS and CF can impact response time

- CF utilization guidelines
  - <30% busy if single CP, otherwise <50%
  - Ensure sufficient capacity to handle structures that might be rebuilt into the CF as the result of failures or maintenance on a peer CF
CF Technology Issues

- Our primary concern is the impact that CF service time has on the z/OS image that is accessing the CF
  - Faster CF request service time reduces overhead for z/OS image
- Type of CF processor determines how quickly an operation can be performed (but just one factor among many)
- Rule of thumb: CF processor should be no more than one generation behind the processor that hosts the z/OS images that use the CF
CFCC Issues

- Stay current with maintenance

- Internal Contention
  - Very hard to detect, and seldom seen. I mention it to be complete.
  - Contention issues within the CF generally arise from usage patterns for particular structures, and are isolated to those structures
  - Do not typically see “global” contention issues in the CF
Detecting Coupling Facility Issues

- First resolve existence and accessibility issues
  - See “Sysplex Fabric: Coupling Facility”

- Performance
  - RMF CF Activity Report
  - RMF Partition Data Report
  - RMF Monitor III – CF Overview
  - RMF Monitor III – Sysplex CF Views

  Getting enough physical CP?
  Timely dispatch?
  CF utilization within guidelines?
RMF CF Activity Report

- CF Utilization within guidelines?
- Dynamic dispatching?
- Effective CPs?

**PROCESSOR SUMMARY**

<table>
<thead>
<tr>
<th>COUPLING FACILITY</th>
<th>2817</th>
<th>MODEL M80</th>
<th>CFLEVEL 17</th>
<th>DYNDISP OFF</th>
</tr>
</thead>
<tbody>
<tr>
<td>AVERAGE CF UTILIZATION (% BUSY)</td>
<td>30.7</td>
<td>LOGICAL PROCESSORS:</td>
<td>Defined 2</td>
<td>Effective 2.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Shared 0</td>
<td>Avg Weight 0.0</td>
</tr>
</tbody>
</table>

*May need to look at data over several intervals
What utilization is normal for you?*
### RMF Partition Data Report

#### z/OS V1R13
- **SYSTEM ID**: SYS1
- **DATE**: 07/12/2012
- **INTERVAL**: 15.00.294

---

#### PARTITION DATA

<table>
<thead>
<tr>
<th>NAME</th>
<th>S</th>
<th>WGT</th>
<th>DEF</th>
<th>ACT</th>
<th>DEF</th>
<th>WLM%</th>
<th>NUM</th>
<th>TYPE</th>
<th>EFFECTIVE</th>
<th>TOTAL</th>
<th>EFFECTIVE</th>
<th>TOTAL</th>
<th>LPAR</th>
<th>MGMT</th>
<th>EFFECTIVE</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>CF1</td>
<td>A</td>
<td>DED</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ICF</td>
<td>00.15.00.269</td>
<td>00.15.00.275</td>
<td>100.0</td>
<td>100.0</td>
<td>0.00</td>
<td>50.00</td>
<td>50.00</td>
<td></td>
</tr>
<tr>
<td>CF2</td>
<td>A</td>
<td>DED</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ICF</td>
<td>00.15.00.170</td>
<td>00.15.00.176</td>
<td>99.99</td>
<td>99.99</td>
<td>0.00</td>
<td>49.99</td>
<td>49.99</td>
<td></td>
</tr>
<tr>
<td><em>PHYSICAL</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### AVERAGE PROCESSOR UTILIZATION PERCENTAGES

<table>
<thead>
<tr>
<th>LOGICAL PROCESSORS</th>
<th>EFFECTIVE</th>
<th>TOTAL</th>
<th>PHYSICAL PROCESSORS</th>
<th>EFFECTIVE</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>00.30.00.440</td>
<td>00.30.00.500</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

**Shared CP (suspect), or Dedicated CP (good) ?**

**This is not CF Utilization. The logical CP looks 100% busy due to polling loop.**

**Getting enough physical CP ?**
### RMF Monitor III: CF Overview

**RMF V1R11  CF Overview  - MABPLX  Line 1 of 3**

Samples: 100     Systems: 9    Date: 08/07/12  Time: 14.00.00  Range: 100   Sec

CF Policy: CFRMPOL1    Activated at: 07/24/12 18.28.37

<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>Model</th>
<th>Lvl</th>
<th>Dyn</th>
<th>Util%</th>
<th>Def</th>
<th>Shr</th>
<th>Wgt</th>
<th>Effect</th>
<th>Rate</th>
<th>Size</th>
<th>Avail</th>
</tr>
</thead>
<tbody>
<tr>
<td>CFRP</td>
<td>2817</td>
<td>M66</td>
<td>17</td>
<td>OFF</td>
<td>6.6</td>
<td>2</td>
<td>0</td>
<td>2.0</td>
<td>9441</td>
<td>20G</td>
<td>17G</td>
<td></td>
</tr>
<tr>
<td>CF1A</td>
<td>2094</td>
<td>S54</td>
<td>15</td>
<td>OFF</td>
<td>4.6</td>
<td>2</td>
<td>0</td>
<td>2.0</td>
<td>3123</td>
<td>20G</td>
<td>17G</td>
<td></td>
</tr>
<tr>
<td>CF1C</td>
<td>2817</td>
<td>M32</td>
<td>17</td>
<td>OFF</td>
<td>5.6</td>
<td>2</td>
<td>0</td>
<td>2.0</td>
<td>9095</td>
<td>20G</td>
<td>16G</td>
<td></td>
</tr>
</tbody>
</table>
### RMF Monitor III: CF System View

**RMF V1R11**  
**CF Systems**  
**- MABPLXS**  
**Line 1 of 27**

**Samples:** 100  
**Systems:** 9  
**Date:** 08/07/12  
**Time:** 14.00.00  
**Range:** 100  
**Sec**

<table>
<thead>
<tr>
<th>CF Name</th>
<th>System</th>
<th>Subchannel</th>
<th>-- Paths --</th>
<th>-- Sync ---</th>
<th>--------</th>
<th>Async -------</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Delay %</td>
<td>Busy %</td>
<td>Avail Delay %</td>
<td>Rate Serv</td>
<td>Avg Chng Del %</td>
</tr>
<tr>
<td>CFRP</td>
<td>CSK</td>
<td>0.0</td>
<td>0.0</td>
<td>4</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>22.8</td>
<td>344 0.0 0.0</td>
</tr>
<tr>
<td>SA0</td>
<td></td>
<td>0.0</td>
<td>2.1</td>
<td>4</td>
<td>0.0</td>
<td>112.6 11 2051</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>20.6</td>
<td>279 11.2 0.5</td>
</tr>
<tr>
<td>SB0</td>
<td></td>
<td>0.0</td>
<td>0.8</td>
<td>4</td>
<td>0.1</td>
<td>104.6 7 1345</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.7</td>
<td>166 0.0 0.0</td>
</tr>
<tr>
<td>SC0</td>
<td></td>
<td>0.0</td>
<td>2.6</td>
<td>4</td>
<td>0.0</td>
<td>120.4 11 3519</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.1</td>
<td>204 0.0 0.0</td>
</tr>
<tr>
<td>SD0</td>
<td></td>
<td>0.0</td>
<td>0.2</td>
<td>4</td>
<td>0.0</td>
<td>108.8 6 664.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6.5</td>
<td>81 0.0 0.0</td>
</tr>
<tr>
<td>SE0</td>
<td></td>
<td>0.0</td>
<td>0.5</td>
<td>2</td>
<td>0.0</td>
<td>43.8 24 440.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.6</td>
<td>161 0.0 0.0</td>
</tr>
<tr>
<td>SF0</td>
<td></td>
<td>0.0</td>
<td>0.4</td>
<td>2</td>
<td>0.0</td>
<td>38.7 25 381.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.8</td>
<td>157 0.0 0.0</td>
</tr>
<tr>
<td>SG0</td>
<td></td>
<td>0.0</td>
<td>0.1</td>
<td>4</td>
<td>0.0</td>
<td>&lt;0.1 72 &lt;0.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.2</td>
<td>389 0.0 0.0</td>
</tr>
<tr>
<td>SH0</td>
<td></td>
<td>0.0</td>
<td>0.2</td>
<td>2</td>
<td>0.0</td>
<td>&lt;0.1 66 &lt;0.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.1</td>
<td>412 0.0 0.9</td>
</tr>
<tr>
<td>CF1A</td>
<td>CSK</td>
<td>0.0</td>
<td>0.0</td>
<td>4</td>
<td>0.0</td>
<td>14.5 13 19.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.4</td>
<td>387 0.0 0.0</td>
</tr>
<tr>
<td>SA0</td>
<td></td>
<td>0.0</td>
<td>0.3</td>
<td>4</td>
<td>0.0</td>
<td>140.6 24 446.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.4</td>
<td>164 0.5 0.2</td>
</tr>
<tr>
<td>SB0</td>
<td></td>
<td>0.0</td>
<td>0.2</td>
<td>4</td>
<td>0.0</td>
<td>53.3 19 388.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.5</td>
<td>135 0.0 0.0</td>
</tr>
<tr>
<td>SC0</td>
<td></td>
<td>0.0</td>
<td>0.2</td>
<td>4</td>
<td>0.0</td>
<td>156.5 23 475.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.5</td>
<td>416 0.0 1.1</td>
</tr>
<tr>
<td>SD0</td>
<td></td>
<td>0.0</td>
<td>0.3</td>
<td>4</td>
<td>0.0</td>
<td>58.7 20 484.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.6</td>
<td>146 0.0 0.0</td>
</tr>
<tr>
<td>SE0</td>
<td></td>
<td>0.0</td>
<td>0.6</td>
<td>2</td>
<td>0.0</td>
<td>33.9 32 348.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.6</td>
<td>224 0.0 0.0</td>
</tr>
<tr>
<td>SF0</td>
<td></td>
<td>0.0</td>
<td>0.3</td>
<td>2</td>
<td>0.0</td>
<td>7.0 30 258.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.7</td>
<td>184 0.0 0.0</td>
</tr>
<tr>
<td>SG0</td>
<td></td>
<td>0.0</td>
<td>0.1</td>
<td>4</td>
<td>0.0</td>
<td>12.4 6 44.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.1</td>
<td>348 0.0 0.0</td>
</tr>
<tr>
<td>SH0</td>
<td></td>
<td>0.0</td>
<td>0.2</td>
<td>2</td>
<td>0.0</td>
<td>38.9 13 43.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.3</td>
<td>480 0.0 2.3</td>
</tr>
</tbody>
</table>

e tc.
Coupling Facility Request Processing

- Link technology
- Distance
- No subchannel
- Path busy
CF Link Considerations

- **Link technology and distance**
  - Faster links improve transmit portion of service time
  - Distance increases service time by 10 mics/km

- **No subchannel conditions**
  - Bursts?  Sustained load?  Tuned due to path busy?

- **Path busy conditions**
  - Number of requests exceeds number of available link buffers
  - XES may tune number of subchannels to avoid this condition
  - Distance (link buffers in use for longer)

- **Configured correct number of subchannels/CHPID?**
  - 32 for HCA2-O LR or HCA3-O LR, otherwise 7
Detecting CF link issues

- First eliminate all physical link errors

- First resolve or eliminate CF responsiveness issues
  - Unresponsive CF can induce link problems
  - Link buffers “linger”, which can induce path busy conditions

- RMF Report of CF Activity
  - Delayed requests implies “no subchannel”
  - Reports “path busy” conditions

- D CF,CFNAME
  - Shows configured links
  - How many subchannels available? Being used?
### RMF CF Activity: Subchannel Activity

**SUBCHANNEL ACTIVITY**

<table>
<thead>
<tr>
<th>Requests</th>
<th>Delayed Requests</th>
</tr>
</thead>
<tbody>
<tr>
<td># REQ</td>
<td>AVG TIME(MIC)</td>
</tr>
<tr>
<td></td>
<td>STD_DEV</td>
</tr>
<tr>
<td>Sync</td>
<td>151301</td>
</tr>
<tr>
<td>Async</td>
<td>8388K</td>
</tr>
<tr>
<td>Changed</td>
<td>6627K INCLUDED IN async</td>
</tr>
<tr>
<td>Unsucc</td>
<td>0</td>
</tr>
<tr>
<td>Sync</td>
<td>96288</td>
</tr>
<tr>
<td>Async</td>
<td>8444K</td>
</tr>
<tr>
<td>Changed</td>
<td>6636K INCLUDED IN async</td>
</tr>
<tr>
<td>Unsucc</td>
<td>0</td>
</tr>
</tbody>
</table>

*Know your workload.*

*What is normal for you.*

*What changed?*
### RMF CF Activity: Subchannel Activity

This was a test intended to drive path busy conditions (it worked).

Know your workload.
What is normal for you.
What changed?
D CF,CFNAME=xxxx    .....

<table>
<thead>
<tr>
<th>PATH</th>
<th>PHYSICAL</th>
<th>LOGICAL</th>
<th>CHANNEL TYPE</th>
<th>AID</th>
<th>PORT</th>
</tr>
</thead>
<tbody>
<tr>
<td>03 / 0727</td>
<td>NOT OPERATIONAL</td>
<td>ONLINE</td>
<td>CIB</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>C0 / 0737</td>
<td>NOT OPERATIONAL</td>
<td>ONLINE</td>
<td>CIB</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>C1 / 0738</td>
<td>ONLINE</td>
<td>ONLINE</td>
<td>CIB</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

**COUPLING FACILITY SUBCHANNEL STATUS**
TOTAL: 96   IN USE: 7   NOT USING: 0   NOT USABLE: 89

Operational devices / subchannels:
- F42F / 025E
- F42F / 025F
- F43F / 035E
- F43F / 035F
- F43F / 035E
- F43F / 035F

- **Configured links with 32 SCH/CHPID but link only supports 7**
- **Nonzero implies XES tuning to avoid path busy**
- **Configuration error? Hardware problem?**
Methodology Concern

- Subchannel and path busy conditions imply that there are not enough subchannels and link buffers to satisfy the workload.

- Typically resolved by:
  - Adding links, or CHPIDs (for IFB)
  - Upgrading link technology

- But …
  - Has workload grown?
  - Experiencing a spike? Just tolerate it?
  - Runaway application?

Know your workload.
What is normal for you.
What changed?
Coupling Facility Request Processing

- Delay Time
- Service Time
- Subchannels
- Operation Completion
- Request Completion
- Sync vs async
- Back end z/OS issues
Asynchronous operation completion

**z/OS**

**Dispatcher**
- If global summary
  - Loop:
    - If local summary[i]
      - Schedule SCN SRB[i]

**SCN SRB[i]**
- Loop:
  - If subchannel vector[j]
    - STCK( T2 )
  - If XCF Signal, call CE
  - Else Schedule CE

**Completion Exit SRB**
- Store results, free CB
- Select user mode
  - When exit: Call CE
  - When ECB: Post
  - When token: n/a

**CSS**

- Global Summary
- Subchannel Vectors
- Local Summary
- Subchannels

**SCN = Subchannel Completion Notification**
**CE = User Completion Exit**
Request Completion Issues

- **Mismanagement of summary bits and subchannel vectors**
  - (Extremely rare, only seen this once. I mention to be complete.)
  - Subchannel completion vector issues are likely “repaired” by synchronous requests, or an internal monitor

- **Loss of physical processor**
  - May elongate asynchronous CF service times
  - May elongate application response times

- **Low loads**
  - MVS may ask LPAR to take it out of its no work wait less often
  - If dispatcher runs less often, takes longer to notice completion
  - Which increases async service times
Request Completion Issues ...

z/OS Dispatching Issues

- Is application address space getting dispatched enough?
  - Won't show up in CF service time measurements
  - But application might appear to be sluggish, and
  - Increases dwell time of XES control blocks (common storage)

- XCF runs at high dispatch priority, so not usually an issue
  - Unless there is a more global issue
    • Storage constraints, spin loops, …
    • But we already eliminated those issues, right?
What does your “one number” tell you?

- Time z/OS spends waiting for subchannels
- Time spent resolving path busy conditions
- Time spent transmitting request from z/OS to CF
- Time spent waiting for physical dispatch of CF to receive request
- Time CF spends preparing, processing, and completing the request
- Time spent transmitting results from CF to z/OS
- Time spent waiting for physical dispatch of z/OS to receive results
- Time spent in back end application completion processing
Methodology

- Having eliminated issues for CF requests in general, there could be issues that are unique to specific structures.

- Might use one or more of the following approaches to decide which ones to look at. You might review structures that:
  - Are known to be critical to the sysplex
    - XCF signalling paths, ISGLOCK, ...
  - Have service times or request rates out of line with past behavior
  - Are known to be used by applications under suspicion

  *Bear in mind that the application could be using a service that exploits its own CF structure. You might miss the root cause if you only look at the structures you think are relevant to the suspicious application.*

- If you do identify something abnormal, you may need application specific expertise for deeper diagnosis.
CF Structure Specific Concerns

- Properly sized?
  - Always resize after CF/hardware upgrade
  - Resize as workload changes
  - Is CFRM policy change still pending for new size?
  - For lock structures: Contention? False contention?

- Hitting full thresholds?
  - Whether an issue depends on application

- Being altered?

- Being rebuilt?

- Newly allocated?

- Application design issues
  - Perhaps the implementation induces contention/queueing in CF

See “z/OS Hot Topics” Issue 26 Lead article on CF Sizer

Sizing issues often arise long after CF upgrade for which structure was not resized

Contention implies delay, and induces signalling activity. Can fix false contention by increasing structure size appropriately.

Activity is quiesced (delayed) while in rebuild
CF Structure Specific Diagnosis
Be Prepared

- Know your workload

- Periodically review structure usage
  - Request rates
  - Service times
  - Storage consumption (within the structure)
  - For lock structures, contention and false contention rates

- Relate changes in structure usage to workload variations
  - Track any “odd” behavior

- Maintain a list of which applications use which structures
Diagnosing CF Structure Specific Issues

- What changed?
  - CFCC MCL upgrades
  - Processor upgrades
  - Software maintenance upgrades
  - Software migrations
  - Workload changes

- D XCF,STR

- RMF CF Activity Reports
  - Request rates and service times for each structure, each system

- RMF Monitor III – Sysplex CF Views
D XCF,STR,STRNAME=xxx

D XCF,STR,STRNAME=IXCPLEX_PATH1
IXC360I  22.30.28  DISPLAY XCF 852
STRNAME:  IXCPLEX_PATH1
  STATUS:  ALLOCATED
  EVENT MANAGEMENT:  POLICY-BASED
  TYPE:  LIST
  POLICY INFORMATION:
    POLICY SIZE :  278016 K
    POLICY INITSIZE: N/A
    POLICY MINSIZE :  0 K
    FULLTHRESHOLD :  80
    ALLOWAUTOALT : NO
    REBUILD PERCENT: N/A
    DUPLEX : DISABLED
    ALLOWREALLOCATE: YES
    PREFERENCE LIST:  SVT1  SVT1CF2  SVT1CF1
    ENFORCEORDER : NO
    EXCLUSION LIST IS EMPTY
D XCF,STR,STRNAME=xxxx .....
**COUPLING FACILITY NAME = CFMARK**

**TOTAL SAMPLES (AVG) = 845**  **(MAX) = 861  ** **(MIN) = 828**

### STRUCTURE SUMMARY

<table>
<thead>
<tr>
<th>TYPE</th>
<th>STRUCTURE NAME</th>
<th>STATUS</th>
<th>CHG</th>
<th>ALLOC</th>
<th>% OF CF</th>
<th>STOR</th>
<th>% OF ALL</th>
<th>% OF CF</th>
<th>CF</th>
<th>AVG</th>
<th>LST/DIR</th>
<th>DATA ELEMENTS</th>
<th>LOCK</th>
<th>DIR REC/ XI'S</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST</td>
<td>MARK_LIST1</td>
<td>ACTIVE</td>
<td></td>
<td>20M</td>
<td>0.0</td>
<td>857662</td>
<td>1.2</td>
<td>0.4</td>
<td>952.96</td>
<td>24K</td>
<td>21K</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MARK_LIST2</td>
<td>ACTIVE</td>
<td></td>
<td>121M</td>
<td>0.1</td>
<td>18336</td>
<td>0.0</td>
<td>0.0</td>
<td>20.37</td>
<td>143K</td>
<td>285K</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MBROOKS_LOCK1</td>
<td>ACTIVE</td>
<td></td>
<td>768M</td>
<td>0.4</td>
<td>11206K</td>
<td>15.8</td>
<td>4.8</td>
<td>12452</td>
<td>1808K</td>
<td>0</td>
<td>134M</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MBROOKS_LOCK1</td>
<td>ACTIVE</td>
<td></td>
<td>500M</td>
<td>0.2</td>
<td>2453K</td>
<td>3.5</td>
<td>1.0</td>
<td>2726.0</td>
<td>702K</td>
<td>0</td>
<td>134M</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MBROOKS_GBP0</td>
<td>ACTIVE</td>
<td></td>
<td>49M</td>
<td>0.0</td>
<td>94</td>
<td>0.0</td>
<td>0.0</td>
<td>0.10</td>
<td>41K</td>
<td>8287</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>STRUCTURE TOTALS</strong></td>
<td></td>
<td></td>
<td></td>
<td>96G</td>
<td>48.3</td>
<td>70907K</td>
<td>100</td>
<td>100</td>
<td>78785</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Post Processor: CF Activity Report – PART 3

<table>
<thead>
<tr>
<th>STRUCTURE NAME = ISGLOCK</th>
<th>TYPE = LOCK</th>
<th>STATUS = ACTIVE</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>SYSTEM NAME</th>
<th>AVG/SEC</th>
<th>TOTAL</th>
<th>#</th>
<th>% OF</th>
<th>-SERV TIME(MIC)-</th>
<th>REASON</th>
<th>#</th>
<th>% OF</th>
<th>---- AVG TIME(MIC) ----</th>
<th>EXTERNAL REQUEST</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>AVG/SEC</td>
<td>REQ</td>
<td>ALL</td>
<td>AVG</td>
<td>STD_DEV</td>
<td>REQ</td>
<td>REQ</td>
<td>/DEL</td>
<td>STD_DEV</td>
</tr>
<tr>
<td>BROOKS1</td>
<td>317K</td>
<td>351.7</td>
<td>317K</td>
<td>18.1</td>
<td>9.9</td>
<td>2.3</td>
<td>NO SCH</td>
<td>59</td>
<td>0.0</td>
<td>19.7</td>
</tr>
<tr>
<td>CHNGD</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>PR CMP</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th>SYSTEM NAME</th>
<th>AVG/SEC</th>
<th>TOTAL</th>
<th>#</th>
<th>% OF</th>
<th>-SERV TIME(MIC)-</th>
<th>REASON</th>
<th>#</th>
<th>% OF</th>
<th>---- AVG TIME(MIC) ----</th>
<th>EXTERNAL REQUEST</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>AVG/SEC</td>
<td>REQ</td>
<td>ALL</td>
<td>AVG</td>
<td>STD_DEV</td>
<td>REQ</td>
<td>REQ</td>
<td>/DEL</td>
<td>STD_DEV</td>
</tr>
<tr>
<td>TOTAL</td>
<td>1746K</td>
<td>1940</td>
<td>1584K</td>
<td>90.8</td>
<td>10.5</td>
<td>4.2</td>
<td>NO SCH</td>
<td>91</td>
<td>0.0</td>
<td>20.1</td>
</tr>
<tr>
<td>CHNGD</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>PR CMP</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>
### RMF Monitor III: CF System View

**RMF V1R11 CF Activity - MABPLX**  
Line 1 of 154

- **Samples:** 100  
- **Systems:** 9  
- **Date:** 08/07/12  
- **Time:** 14.00.00  
- **Range:** 100 Sec

**CF: ALL**

<table>
<thead>
<tr>
<th>Structure Name</th>
<th>Type</th>
<th>ST</th>
<th>System</th>
<th>CF</th>
<th>Util %</th>
<th>Rate</th>
<th>Avg Serv</th>
<th>--- Sync</th>
<th>--- Async</th>
<th>Chng</th>
<th>Del %</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADSW_DFHJ01</td>
<td>LIST</td>
<td>AP</td>
<td>*ALL</td>
<td>0.0</td>
<td>0.0</td>
<td>0</td>
<td>0.8</td>
<td>305</td>
<td>1.3</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LIST</td>
<td>AS</td>
<td>*ALL</td>
<td>0.0</td>
<td>0.0</td>
<td>0</td>
<td>0.8</td>
<td>290</td>
<td>1.3</td>
<td>1.3</td>
<td></td>
</tr>
<tr>
<td>ADSW_DFHJ02</td>
<td>LIST</td>
<td>A</td>
<td>*ALL</td>
<td>0.0</td>
<td>0.8</td>
<td>6</td>
<td>&lt;0.1</td>
<td>20</td>
<td>50.0</td>
<td>50.0</td>
<td></td>
</tr>
<tr>
<td>ADSW_DFHJ03</td>
<td>LIST</td>
<td>A</td>
<td>*ALL</td>
<td>0.0</td>
<td>3.5</td>
<td>16</td>
<td>0.7</td>
<td>991</td>
<td>1.5</td>
<td>1.5</td>
<td></td>
</tr>
<tr>
<td>ADSW_DFHJ04</td>
<td>LIST</td>
<td>A</td>
<td>*ALL</td>
<td>0.0</td>
<td>0.7</td>
<td>25</td>
<td>0.0</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>ADSW_DFHJ05</td>
<td>LIST</td>
<td>A</td>
<td>*ALL</td>
<td>0.0</td>
<td>0.7</td>
<td>7</td>
<td>0.0</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>ADSW_DFHJ06</td>
<td>LIST</td>
<td>A</td>
<td>*ALL</td>
<td>0.0</td>
<td>&lt;0.1</td>
<td>18</td>
<td>0.0</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>ADSW_DFHJ07</td>
<td>LIST</td>
<td>A</td>
<td>*ALL</td>
<td>0.0</td>
<td>0.7</td>
<td>25</td>
<td>0.0</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>ADSW_DFHLGLOG1</td>
<td>LIST</td>
<td>AP</td>
<td>*ALL</td>
<td>0.0</td>
<td>0.0</td>
<td>0</td>
<td>0.0</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LIST</td>
<td>AS</td>
<td>*ALL</td>
<td>0.0</td>
<td>0.0</td>
<td>0</td>
<td>0.0</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>APPCLOG</td>
<td>LIST</td>
<td>A</td>
<td>*ALL</td>
<td>0.0</td>
<td>0.0</td>
<td>0</td>
<td>&lt;0.1</td>
<td>880</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>CI1_DFLOG</td>
<td>LIST</td>
<td>A</td>
<td>*ALL</td>
<td>13.0</td>
<td>192.1</td>
<td>25</td>
<td>227.9</td>
<td>226</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>CI1_DFLOG1</td>
<td>LIST</td>
<td>A</td>
<td>*ALL</td>
<td>4.2</td>
<td>236.0</td>
<td>22</td>
<td>197.3</td>
<td>208</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>CI1_DFLOG2</td>
<td>LIST</td>
<td>A</td>
<td>*ALL</td>
<td>2.2</td>
<td>116.5</td>
<td>22</td>
<td>108.3</td>
<td>211</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>CI1_DFSHUNT</td>
<td>LIST</td>
<td>A</td>
<td>*ALL</td>
<td>0.0</td>
<td>0.0</td>
<td>0</td>
<td>0.0</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>CI1_DFSHUNT1</td>
<td>LIST</td>
<td>A</td>
<td>*ALL</td>
<td>0.0</td>
<td>0.0</td>
<td>0</td>
<td>0.0</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>CI1_DFSHUNT2</td>
<td>LIST</td>
<td>A</td>
<td>*ALL</td>
<td>0.0</td>
<td>0.0</td>
<td>0</td>
<td>0.0</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
</tbody>
</table>

etc.
CF Structure Specific Diagnosis

XCF messages

- IXC512I “CFRM policy change pending”
- IXC521I “rebuild starting / rebuild complete”
- IXC522I “stopping rebuild of structure”
- IXC538I “duplexing rebuild not started (or stopped)”
- IXC552I “duplexed structures not failure isolated from each other”
- IXC553E “duplexed structures not failure isolated from each other”
- IXC573I “system managed rebuild failed”
- IXC585E “structure exceeds full threshold”
- IXC586I “structure below full threshold”
- IXC588I “altering structure”
- IXC589I “altering ended”
- IXC590I “auto alter completed, indicates if targets obtained or not”
Problem Taxonomy

- Dead System
- Sick System
- Sysplex Fabric
- Sysplex Componentry
  - Coupling Facility
  - **Signalling Service**
  - Couple Data Sets
  - External Time Reference
- Configuration / Capacity
- Software Issues
Signalling Overview

Application Receives Message

Schedule MsgExit
Replenish path

IXCMSGIX
Copy data
Recycle buffer

Outbound Buffer Pool

IXCMSGGOX
Select pool
Get buffer
Copy data

Application Sends Message

Inbound Buffer Pool

Msg

Pathout

Pathin

Msg

Msg

Msg

Select path
Transfer Data
Recycle buffer

Msg

Msg

Msg

Msg

Msg

<>

<>

<>

<>

Msg

Msg
Message Exit Concerns: Application Perspective

- Responsiveness
- Throughput

Many applications pass messages to tasks for processing
- Message delivery to the message exit may be timely
- But the application can suffer if the tasks are lagging
- XCF cannot detect such delays

- Might the volume of message exit SRBs be starving the tasks?
- Local lock or latch contention between the SRBs and the tasks inducing delays for one or the other?
Message Exit Concerns: Sysplex Perspective

- Member delays/bursts can impede signal transfers, which can induce delays for other applications

- Long queue effects
  - Storage and buffer consumption
  - Processing time to run the queue
Message Exit Delays

Loop for N:
Prepare to deliver
Call MsgExit

Prepare to receive
IXCMSGIX to receive

Copy msg
Recycle buffer

Process/Queue msg
Return

Recycle Buffer
(if needed)

Inbound Buffer Pool

Not Dispatched

Schedule

Impacts:
• Tie up signal buffers
• Tie up SRBs
• Delay message
• Delay next message

Member Delivery Queue
Detecting Member Message Exit Issues

- XCF member stalled messages (see handout)
- D XCF,GROUP (see handout)
- IPCS COUPLE SIGNAL DETAIL report (see handout)

These are not perfect in that they take a relatively long time to declare a stall condition.

We seem to be encountering more and more cases where short duration stalls impact the sysplex. Very hard to see.
## Signalling Sympathy Sickness Indicators

<table>
<thead>
<tr>
<th>Impacted System</th>
<th>Culprit System</th>
</tr>
</thead>
<tbody>
<tr>
<td>D XCF,G… shows stalls</td>
<td>D XCF,G… shows stalls</td>
</tr>
<tr>
<td>IXC467I Restart stalled I/O</td>
<td>IXC431I member stalled</td>
</tr>
<tr>
<td><strong>Stalled Members</strong></td>
<td>ABEND 00C 020F0006</td>
</tr>
<tr>
<td>IXC440E impacted</td>
<td>IXC430E stalled members</td>
</tr>
<tr>
<td><strong>Sympathy Sickness</strong></td>
<td>IXC631I member causing SS</td>
</tr>
<tr>
<td></td>
<td>IXC640E if/when to act</td>
</tr>
<tr>
<td></td>
<td>ABEND 00C 020F000C</td>
</tr>
<tr>
<td>If SFM allowed to take action</td>
<td><strong>If SFM allowed to take action</strong></td>
</tr>
<tr>
<td></td>
<td>ABEND 00C 020F000D</td>
</tr>
<tr>
<td></td>
<td>IXC615I terminating</td>
</tr>
<tr>
<td></td>
<td>• ABEND 00C 00000160</td>
</tr>
<tr>
<td></td>
<td>• Wait State 0A2 rsn 160</td>
</tr>
</tbody>
</table>
## Timeline for stalled member messages

<table>
<thead>
<tr>
<th>Event</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>IXC430E</strong></td>
<td>System has stalled members</td>
</tr>
<tr>
<td><strong>IXC431I</strong></td>
<td>Member stalled</td>
</tr>
<tr>
<td><strong>IXC601I</strong></td>
<td>Stall condition</td>
</tr>
<tr>
<td><strong>IXC631I</strong></td>
<td>Member causing sympathy sickness</td>
</tr>
<tr>
<td><strong>IXC640E</strong></td>
<td>Taking action at time T</td>
</tr>
<tr>
<td><strong>IXC631I</strong></td>
<td>Indeterminate</td>
</tr>
<tr>
<td><strong>IXC615I</strong></td>
<td>XCF Terminating member to resolve sympathy sickness</td>
</tr>
<tr>
<td><strong>IXC647I</strong></td>
<td>Path restart</td>
</tr>
<tr>
<td>Monitor pops</td>
<td></td>
</tr>
</tbody>
</table>

Stall condition = At least 1 exit stalled for 30 seconds or work item on head of queue for 30 seconds

Time $T = \text{"now"} + \text{MEMSTALLTIME}$
Stalled Member Drilling

D XCF, G

IXC331I  09.43.27  DISPLAY XCF 297

* INDICATES PROBLEM, ! INDICATES SEVERE PROBLEM

GROUPS(SIZE):

*A0000002(2)       COFVLFNO(2)       CTTXNGRP(2)
EZBTCPCS(2)         ISTCFS01(2)       ISTXCF(2)
IXCLO02A(2)         IXCLO02B(2)       SYSCNZMG(2)
SYSDAE(4)           SYSENF(2)         SYSGRS(2)
SYSGRS2(1)          SYSIEFTS(2)       SYSIGW00(3)
SYSIGW01(3)         SYSIKJBC(2)       SYSIOSPX(2)
SYSIOS01(1)         SYSIOS02(1)       SYSJES(3)
SYSJ2$XD(3)         SYSMCS(7)         SYSMCS2(43)
SYSTTRC(2)          SYSWLM(3)         SYSXCF(2)
Stalled Member Drilling …

D XCF,G,A0000002

IXC332I  09.43.27  DISPLAY XCF 300
* INDICATES PROBLEM, ! INDICATES SEVERE PROBLEM
GROUP A0000002:  *MEMBER1  MEMBER2
Stalled Member Drilling …

D XCF,G,A0000002, MEMBER1

IXC333I  09.43.27  DISPLAY XCF 303
  INFORMATION FOR GROUP A0000002
  * INDICATES PROBLEM, ! INDICATES SEVERE PROBLEM
  MEMBER NAME:        SYSTEM:     JOB ID:    STATUS:
  *MEMBER1             SY1         XCAC3Z07   ACTIVE

INFO FOR GROUP A0000002 MEMBER MEMBER1 ON SYSTEM SY1
  * INDICATES STALLS

FUNCTION: TESTCASE XCJC3Z07
MEMTOKEN: 01000064 001E0001      ASID: 0025       SYSID: 0100000B
  INFO: CURRENT COLLECTED: 09/10/2010 09:47:03.876399

ATTRIBUTES
  JOIN TASK ASSOCIATION
  CRITICAL MEMBER
  LOCAL CLEANUP NOT NEEDED
  TERMLEVEL IS TASK
  MEMSTALL RESOLUTION IS JOIN TASK TERMINATION AFTER 603 SECONDS
  EXITS DEFINED: MESSAGE, GROUP
## Stalled Member Drilling …

<table>
<thead>
<tr>
<th>SIGNALLING SERVICE</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSGO ACCEPTED: 23 NOBUFFER: 0</td>
</tr>
<tr>
<td>MSGO XFER CNT: 0 LCL CNT: 23 BUFF LEN: 956</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SENDPND RESPPND COMPLTD MOSAVED MISAVED</th>
</tr>
</thead>
<tbody>
<tr>
<td>MESSAGE TABLE: 0 0 0 0 0 0</td>
</tr>
</tbody>
</table>

| MSGI RECEIVED: 23 PENDINGQ: 6 |
| MSGI XFER CNT: 0 XFERTIME: N/A |

<table>
<thead>
<tr>
<th>IO BUFFERS DREF PAGEABLE CRITICAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSGI PENDINGQ: 0 0 5 0</td>
</tr>
<tr>
<td>SYMPATHY SICK: 0</td>
</tr>
</tbody>
</table>

*ITEM 02053020: 09/10/2010 09:42:53.787346 ME SEQ: 19
*ITEM 02116050: 09/10/2010 09:46:27.096268 ME SEQ: 24
*EXIT 01F8CF00: 09/10/2010 09:42:53.784022 ME RUNNING
*EXIT 01F8D100: 09/10/2010 09:42:53.786051 ME RUNNING
*EXIT 01F8D300: 09/10/2010 09:43:26.842027 ME RUNNING
*EXIT 01F8D500: 09/10/2010 09:43:59.885911 ME RUNNING
*EXIT 01F8D700: 09/10/2010 09:46:28.024502 MV 00:00:00.000061
*EXIT 02003100: 09/10/2010 09:42:53.767349 ME RUNNING
*EXIT 02003500: 09/10/2010 09:42:53.769113 ME RUNNING

**Delta Pending Preparing Running**
IPCS COUPLE SIGNAL DETAIL as well

**Detail of signal exit SRBs for group:** SYSGRS  member: SYSK

<table>
<thead>
<tr>
<th>SRB Addr</th>
<th>TOD When Service Called</th>
<th>FC</th>
<th>Duration</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>08703600</td>
<td>06/17/2011 10:39:21.144506</td>
<td>OM *00:02:44.229274</td>
<td>Running</td>
<td></td>
</tr>
<tr>
<td>08704400</td>
<td>06/17/2011 10:39:21.147527</td>
<td>OM *00:02:44.226253</td>
<td>Running</td>
<td></td>
</tr>
<tr>
<td>089E8900</td>
<td>06/17/2011 10:39:21.148002</td>
<td>NA *00:02:44.225778</td>
<td>Pending</td>
<td></td>
</tr>
<tr>
<td>08705200</td>
<td>06/17/2011 10:39:21.148003</td>
<td>NA *00:02:44.225777</td>
<td>Pending</td>
<td></td>
</tr>
</tbody>
</table>

**Detail of work items queued for group:** SYSGRS  member: SYSK

<table>
<thead>
<tr>
<th>ItemAddr</th>
<th>TOD When Queued</th>
<th>Duration as of</th>
<th>ItemSeq#</th>
</tr>
</thead>
<tbody>
<tr>
<td>02799110</td>
<td>06/17/2011 10:39:21.144763</td>
<td>OM *00:02:44.229017</td>
<td>000000538</td>
</tr>
<tr>
<td>02BB7038</td>
<td>06/17/2011 10:39:21.145037</td>
<td>OM *00:02:44.228743</td>
<td>000000539</td>
</tr>
<tr>
<td>02BB7138</td>
<td>06/17/2011 10:39:21.145308</td>
<td>OM *00:02:44.228472</td>
<td>00000053A</td>
</tr>
<tr>
<td>0299D0E8</td>
<td>06/17/2011 10:39:21.145768</td>
<td>OM *00:02:44.228012</td>
<td>00000053B</td>
</tr>
<tr>
<td>0299D1E8</td>
<td>06/17/2011 10:39:21.146210</td>
<td>OM *00:02:44.227570</td>
<td>00000053C</td>
</tr>
</tbody>
</table>
Typical Causes of Message Exit Delays

- **CPU Constraints**
  - Higher priority work winning
  - Insufficient LPAR weight

- **Storage Constraints**
  - Page faults
  - Exhausted private
  - Fragmentation

- **Contention**
  - Waiting for ENQ, local lock, or latch

- **Signal Volume**

- **Defects**

Without a timely dump, virtually impossible to diagnose stalls, particularly those of short duration

Try Run Time Diagnostics
Signalling Overview

- **Outbound Buffer Pool**
  - Select pool
  - Get buffer
  - Copy data
- **IXCMSGOX**
- **Pathout**
- **Pathin**
- **IXCMSGIX**
  - Copy data
  - Recycle buffer
- **Application**
  - Sends Message
- **Signalling Paths**
  - Schedule MsgExit
  - Replenish path
  - MsgExit Replenish path

- **Signalling Overview**
  - Application Receives Message
  - Select path
  - Transfer Data
  - Recycle buffer

- **Msg**
  - <>
  - <>
  - <>

- **Inbound Buffer Pool**
- **Msg**
Signal Path Concerns

- Target system operational?
- Path operational?
  - Restart? Stop? Rebuild?
- Inbound “no buffer” conditions
- Transfer time
Signal Path Concerns: Inbound Buffers

- Stalled members consuming buffers?

- Lack of Transport Class Segregation on sending side?
  - *Every transport class must have a path for every target system*
  - *Signals may not be flowing via the expected paths*

- Transport Class Length relative to PATHIN MAXMSG?
  - *Number of buffers for given MAXMSG decreases as class length increases*
  - *If you define/modify class length, you may need to modify PATHIN MAXMSG to maintain number of buffers*

- PATHIN MAXMSG too small?
  - *May need more buffers if they tend to dwell in member message delivery*

The z/OS Health Checker looks for these problems
Transport Class Specifications

- COUPLE
  - MAXMSG
- CLASSDEF(s)
  - CLASSLEN
  - GROUP
- PATHOUT
  - CLASS
  - How many paths?
- PATHIN
  - MAXMSG

You must consider impact to target as well

Sending System

Target System(s)
Transport Class Must Have Paths

CLASS SMALL

Pathout

Pathout

CLASS BIG

Big msg sent via paths in small class
Pathin buffer pool resized for larger msgs
So has fewer buffers for given MAXMSG
So pathin may get nobuff condition
So both big and small may get delays
Detecting Signal Path Issues

- D XCF,PI,DEV=xxxx
- D XCF,PI,STRNAME=xxxx
- RMF Reports of XCF Activity
  - Path Statistics (inbound paths)

- D XCF,PO,DEV=xxxx
- D XCF,PO,STRNAME=xxxx
- D XCF,CLASSDEF
- RMF Reports of XCF Activity
  - Usage by System (outbound side)

Primary concern is "no buffer" conditions on the inbound side of the path. Signals will not flow if the inbound side has no buffers with which to receive signals.
XCF Pathin Display

D XCF, PI, DEVICE=ALL

IXC356I 09.18.51 DISPLAY XCF

| LOCAL | REMOTE | REMOTE | PATHIN | DELIVERY | BUFFER | MSGBUF | SIGNL | LOCAL | REMOTE | REMOTE | PATHIN | DELIVERY | BUFFER | MSGBUF | SIGNL | LOCAL | REMOTE | REMOTE | PATHIN | DELIVERY | BUFFER | MSGBUF | SIGNL | LOCAL | REMOTE | REMOTE | PATHIN | DELIVERY | BUFFER | MSGBUF | SIGNL | LOCAL | REMOTE | REMOTE | PATHIN | DELIVERY | BUFFER | MSGBUF | SIGNL |
|-------|--------|--------|--------|----------|--------|--------|------|-------|--------|--------|--------|--------|----------|--------|--------|------|-------|--------|--------|--------|--------|-------|--------|--------|--------|--------|----------|--------|--------|------|-------|--------|--------|--------|--------|-------|--------|--------|--------|--------|----------|--------|--------|------|
|       |        |        |        |          |        |        |      |       |        |        |        |        |          |        |        |      |       |        |        |        |        |       |        |        |        |        |          |        |        |      |       |        |        |        |        |       |        |        |        |        |          |        |        |      |
| 2101  | 2401   | SYSA   | WORKING| 4        | 956    |        |      |      | 2101  | 2401   | SYSA   | WORKING| 4        | 956    |        |      |      | 2101  | 2401   | SYSA   | WORKING| 4        | 956    |        |      |      |
| 2201  | 2401   | SYSB   | WORKING| 4        | 4028   |        |      |      | 2201  | 2401   | SYSB   | WORKING| 4        | 4028   |        |      |      | 2201  | 2401   | SYSB   | WORKING| 4        | 4028   |        |      |      |
| 2301  | 2401   | SYSC   | WORKING| 4        | 8124   |        |      |      | 2301  | 2401   | SYSC   | WORKING| 4        | 8124   |        |      |      | 2301  | 2401   | SYSC   | WORKING| 4        | 8124   |        |      |      |
| 3101  | 3401   | SYSA   | WORKING| 4        | 956    |        |      |      | 3101  | 3401   | SYSA   | WORKING| 4        | 956    |        |      |      | 3101  | 3401   | SYSA   | WORKING| 4        | 956    |        |      |      |
| 3201  | 3401   | SYSB   | WORKING| 4        | 956    |        |      |      | 3201  | 3401   | SYSB   | WORKING| 4        | 956    |        |      |      | 3201  | 3401   | SYSB   | WORKING| 4        | 956    |        |      |      |
| 3301  | 3401   | SYSC   | WORKING| 4        | 956    |        |      |      | 3301  | 3401   | SYSC   | WORKING| 4        | 956    |        |      |      | 3301  | 3401   | SYSC   | WORKING| 4        | 956    |        |      |      |

These NOBUF counts are lifetime for the path and sending system instance.
RMF XCF Path Statistics (inbound)

<table>
<thead>
<tr>
<th>INBOUND TO</th>
<th>SYSB</th>
</tr>
</thead>
<tbody>
<tr>
<td>T FROM/TO</td>
<td></td>
</tr>
<tr>
<td>FROM</td>
<td>Y DEVICE, OR</td>
</tr>
<tr>
<td>SYSTEM</td>
<td>P STRUCTURE</td>
</tr>
<tr>
<td>SYSA</td>
<td>S IXCSTR1</td>
</tr>
<tr>
<td></td>
<td>S IXCSTR2</td>
</tr>
<tr>
<td></td>
<td>S IXCSTR3</td>
</tr>
<tr>
<td></td>
<td>S IXCSTR4</td>
</tr>
</tbody>
</table>

These NOBUF counts are for the indicated reporting interval
Inbound No Buffer Caveats

- CTC: wanted a buffer, but was signal in flight?
- List: wanted a buffer because signal is in flight
- Periodic retry could inflate “no buffer” count
  - Up to 200 millisecond delay, then every 10 milliseconds
  - So nonzero count indicates potential delay, but not necessarily “quantity”
Signal transfers times are interesting in that they may indicate delays. But signal transfers are generally always good unless you have other problems (stalled members, no buffer conditions, message flooding, CF service time issues, etc). If you resolve those problems, you won't have transfer time issues.

So the handout has plenty of details to explain the technical details. But generally there is nothing to see here. Move along..
CTC Path Signal Transfer

Sending System

Next Msgo

Available

Send System

Queue to path Chain CCWs (Resume I/O)

SLIH

SLIH

WRITE

CTC

Available

I/O

I/O

TOD

TOD

In Delivery

Recycle

(Resume I/O)

(Recycle)

(Replenish) (Resume I/O)

Available

3

4

5

6

7

suspend

suspend

© 2013 IBM Corporation
CTC Path Timing

- Signals immediately chained for I/O, so would expect queue time to be quite small

- High queue time might be caused by:
  - Host issues between request and start I/O
  - Resend via alternate path after restart/stop of selected path

- High transfer times might be caused by:
  - Inbound buffer shortage
  - Host issues on target system
  - Device issues
  - Bursts, perhaps (long channel program chain)
Signal Structure

List Structure

- List Controls
  - List Limit
  - XCF controls

- List Entry
  - LEC
  - Adjunct
  - Data Entry
    - data
    - data
List Path Signal Transfer

Sending System

Recieving System
List Path Timing

- Signal immediately written if in current batch, in which case expect small queue time (like CTC)

- High queue time might be caused by:
  - Host issues between request and start I/O
  - Host issues completing back end of current batch
  - List full conditions (CFSIZER ?)

- High transfer times might be caused by:
  - Inbound buffer shortage
  - Host issues on target system
  - Delayed empty to non-empty notifications
  - Issues getting to the CF (subchannel/path busy)
  - Issues in the CF (shared CP, contention, …)
  - Bursts, perhaps (long queue in CF)
Queue Time

- “start I/O” minus “requested”
  - Includes time spent copying message to I/O buffer (page fault?)
  - Could lose CPU after accepted but before queued
  - Includes time spent initiating sends to peer targets for broadcast message
  - Could include time spent waiting for “no buffer” or “no path” conditions to clear
  - Could include time spent doing path restart (or stop) protocol to figure out that signal needed to be resent

  New “started I/O” time taken when signal is redriven
Transfer Time

- “arrived” minus “start I/O”
  - For list path, could include list transition processing
    - Driving transition exit
    - Initiating request to read signal from CF
  - Includes z/OS time to get to back end routine
    - Interrupt processing for CTC path
    - Async command completion for list path
  - Includes XCF time to verify signal arrived intact
## Signal Timings

### Timings for Recent Signal Transfers

#### CTC device

<table>
<thead>
<tr>
<th>TOD When MSGO Requested</th>
<th>Queue Time</th>
<th>Transfer Time</th>
<th>TOD When Arrived</th>
</tr>
</thead>
<tbody>
<tr>
<td>06/17/2011 10:35:40.031364</td>
<td>00:00:00.000001</td>
<td>00:00:00.000799</td>
<td>06/17/2011 10:35:40.032166</td>
</tr>
<tr>
<td>06/17/2011 10:35:44.229656</td>
<td>00:00:00.000000</td>
<td>00:00:00.000774</td>
<td>06/17/2011 10:35:44.230432</td>
</tr>
<tr>
<td>06/17/2011 10:35:49.477406</td>
<td>00:00:00.000000</td>
<td>00:00:00.000768</td>
<td>06/17/2011 10:35:49.478175</td>
</tr>
<tr>
<td>06/17/2011 10:35:51.576329</td>
<td>00:00:00.000002</td>
<td>00:00:00.000789</td>
<td>06/17/2011 10:35:51.577120</td>
</tr>
<tr>
<td>06/17/2011 10:35:57.873469</td>
<td>00:00:00.000000</td>
<td>00:00:00.000769</td>
<td>06/17/2011 10:35:57.874240</td>
</tr>
</tbody>
</table>

#### List Path

<table>
<thead>
<tr>
<th>TOD When MSGO Requested</th>
<th>Queue Time</th>
<th>Transfer Time</th>
<th>TOD When Arrived</th>
</tr>
</thead>
<tbody>
<tr>
<td>06/17/2011 10:30:52.133303</td>
<td>00:00:00.000000</td>
<td>00:00:00.002652</td>
<td>06/17/2011 10:30:52.135956</td>
</tr>
<tr>
<td>06/17/2011 10:31:55.838311</td>
<td>00:00:00.000001</td>
<td>00:00:00.001375</td>
<td>06/17/2011 10:31:55.839687</td>
</tr>
<tr>
<td>06/17/2011 10:32:58.806705</td>
<td>00:00:00.000000</td>
<td>00:00:00.001941</td>
<td>06/17/2011 10:32:58.808647</td>
</tr>
<tr>
<td>06/17/2011 10:34:02.092010</td>
<td>00:00:00.000000</td>
<td>00:00:00.001946</td>
<td>06/17/2011 10:34:02.093958</td>
</tr>
<tr>
<td>06/17/2011 10:34:24.103905</td>
<td>00:00:00.000055</td>
<td>00:00:00.000956</td>
<td>06/17/2011 10:34:24.104917</td>
</tr>
</tbody>
</table>

Take dump. Run IPCS COUPLE SIGNAL DETAIL report

Recent timings reported for both paths and members
Delivery Time

- “called msg exit” minus “arrived”
  - But we do not currently gather these

- Msg exit “returned” minus “called”
  - Available for residual SRBs
  - But no historical tracking
Looking for Delay on Inbound Side

D XCF,PI,DEVICE=ALL

IXC356I  09.17.26  DISPLAY XCF 676

<table>
<thead>
<tr>
<th>LOCAL DEVICE</th>
<th>REMOTE</th>
<th>PATHIN</th>
<th>REMOTE PATHOUT</th>
<th>LAST RETRY</th>
<th>MAXMSG</th>
<th>RECVD TIME</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOCAL PATHIN</td>
<td>REMOTE SYSTEM PATHIN</td>
<td>STATUS</td>
<td>PATHOUT SYSTEM</td>
<td>STATUS</td>
<td>PENDING LENGTH</td>
<td>IN USE NUMBR NOBUF</td>
</tr>
<tr>
<td>2101 SYSA</td>
<td>WORKING</td>
<td>2301</td>
<td>20 2000</td>
<td>2076 379</td>
<td>4 956 8 2076</td>
<td>0</td>
</tr>
<tr>
<td>3101 SYSA</td>
<td>WORKING</td>
<td>3301</td>
<td>20 2000</td>
<td>10099 405</td>
<td>4 956 8 10099</td>
<td>0</td>
</tr>
</tbody>
</table>

STALL-IOPND
STALL-INOP
STALL-SS?
STALL-SS
Looking for Delay on Outbound Side

```
D XCF,PO,DEVICE=ALL

IXC356I 09.17.11 DISPLAY XCF

<table>
<thead>
<tr>
<th>LOCAL DEVICE</th>
<th>REMOTE PATHOUT</th>
<th>REMOTE PATHIN</th>
<th>REMOTE RETRY</th>
<th>MAXMSG</th>
<th>CLASS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SYSTEM STATUS</td>
<td>SYSTEM STATUS</td>
<td>PATHOUT</td>
<td>20</td>
<td>DEFAULT</td>
</tr>
<tr>
<td>2301</td>
<td>SYSC WORKING</td>
<td>2101</td>
<td>20</td>
<td>2000</td>
<td>DEFAULT</td>
</tr>
<tr>
<td>3301</td>
<td>SYSC WORKING</td>
<td>3101</td>
<td>20</td>
<td>2000</td>
<td>DEFAULT</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LOCAL PATHOUT</th>
<th>REMOTE PATHIN</th>
<th>REMOTE SYSTEM</th>
<th>PATHOUT STATUS</th>
<th>TRANSFR PENDING</th>
<th>BUFFER LENGTH</th>
<th>MSGBUF IN USE</th>
<th>NUMBR TIME</th>
<th>TIME</th>
</tr>
</thead>
<tbody>
<tr>
<td>2301</td>
<td>2101</td>
<td>SYSC</td>
<td>WORKING</td>
<td>0</td>
<td>956</td>
<td>10</td>
<td>1642</td>
<td>357</td>
</tr>
<tr>
<td>3301</td>
<td>3101</td>
<td>SYSC</td>
<td>WORKING</td>
<td>0</td>
<td>956</td>
<td>10</td>
<td>10041</td>
<td>315</td>
</tr>
</tbody>
</table>

STALL-IOPND
STALL-INOP
STALL-SS?
STALL-SS
```
Sanity Check Structure Size and Use

From RMF Coupling Facility Usage Summary Report

<table>
<thead>
<tr>
<th>STRUCTURE</th>
<th>ALLOC</th>
<th>LST/DIR</th>
<th>DATA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NAME</td>
<td>ENTRIES</td>
<td>ELEMENTS</td>
</tr>
<tr>
<td>LIST</td>
<td>DSNZPLEX_SCA</td>
<td>38K, 76K</td>
<td>38K, 76K</td>
</tr>
<tr>
<td></td>
<td>IXCSTR1</td>
<td>118, 614</td>
<td>118, 614</td>
</tr>
<tr>
<td></td>
<td>IXCSTR3</td>
<td>5255, 5229</td>
<td>5255, 5229</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1, 18</td>
<td>1, 26</td>
</tr>
</tbody>
</table>

Expect mostly empty. Potential for list full conditions?
Signalling Overview

Transport Classes

Outbound Buffer Pool

Select path
Transfer Data
Recycle buffer

IXCMSGOX

Select pool
Get buffer
Copy data

IXCMSGIX

Copy data
Recycle buffer

Application Sends Message

Application Receives Message

Pathout

Pathin
Transport Class Concerns

- Appropriate segregation
  - Class Length
  - Group (?)

- Signalling Paths
  - How many (must be > 0)
  - Type

- Signal Buffers

  Make one class for 956 byte messages. Perhaps one, or two, maybe three classes for bigger message sizes.

  Put at least two paths in each class for each target system.

  Make sure the PATHIN MAXMSG values are reasonable for the chosen transport class length.
Segregation by Size

CLASSDEF CLASS(SML) CLASSLEN(956)
CLASSDEF CLASS(MED) CLASSLEN(8000)
CLASSDEF CLASS(BIG) CLASSLEN(32000)

SML = 0....956
MED = 957...8124
BIG = 8125..62464

Msgs smaller than defined size tend to waste space
Msgs larger than defined size use next bigger class or require PO/PI size adjustments

Notation "<nnK" means nnK-68 bytes, indicates length of longest message that fits in buffer
### Size Distributions by Class

**SYS3 d xcf, classdef, class=all**

SYS3 IXC344I 13.00.45 DISPLAY XCF

<table>
<thead>
<tr>
<th>TRANSPORT</th>
<th>CLASS</th>
<th>DEFAULT</th>
<th>ASSIGNED</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLASS</td>
<td>LENGTH</td>
<td>MAXMSG</td>
<td>GROUPS</td>
</tr>
<tr>
<td>DEFAULT</td>
<td>956</td>
<td>3000</td>
<td>UNDESIG</td>
</tr>
</tbody>
</table>

**DEFAULT TRANSPORT CLASS USAGE FOR SYSTEM SYS1**

<table>
<thead>
<tr>
<th>SUM MAXMSG:</th>
<th>12000</th>
<th>IN USE:</th>
<th>18</th>
<th>NOBUFF:</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEND CNT:</td>
<td>34523</td>
<td>BUFFLEN (FIT):</td>
<td>956</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEND CNT:</td>
<td>256</td>
<td>BUFFLEN (BIG):</td>
<td>4028</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEND CNT:</td>
<td>7</td>
<td>BUFFLEN (BIG):</td>
<td>8124</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEND CNT:</td>
<td>1</td>
<td>BUFFLEN (BIG):</td>
<td>24508</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**DEFAULT TRANSPORT CLASS USAGE FOR SYSTEM SYS2**

<table>
<thead>
<tr>
<th>SUM MAXMSG:</th>
<th>12000</th>
<th>IN USE:</th>
<th>14</th>
<th>NOBUFF:</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEND CNT:</td>
<td>9929</td>
<td>BUFFLEN (FIT):</td>
<td>956</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEND CNT:</td>
<td>357</td>
<td>BUFFLEN (BIG):</td>
<td>4028</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEND CNT:</td>
<td>6</td>
<td>BUFFLEN (BIG):</td>
<td>8124</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**DEFAULT TRANSPORT CLASS USAGE FOR SYSTEM SYS3**

<table>
<thead>
<tr>
<th>SUM MAXMSG:</th>
<th>3000</th>
<th>IN USE:</th>
<th>4</th>
<th>NOBUFF:</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEND CNT:</td>
<td>3292</td>
<td>BUFFLEN (FIT):</td>
<td>956</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEND CNT:</td>
<td>196</td>
<td>BUFFLEN (BIG):</td>
<td>4028</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Size Distributions by Member

D XCF,G,SYMSCS,ALL
IXC333I  15.45.54  DISPLAY XCF 303
<br>INFO FOR GROUP SYSMCS MEMBER SY1 ON SYSTEM SY1

FUNCTION: Not Specified
MEMTOKEN: 01000003 00070006    ASID: 0009    SYSID: 0100000B
INFO: CURRENT    COLLECTED: 09/14/2010 15:45:54.275742

ATTRIBUTES
JOIN TASK ASSOCIATION
LASTING MEMBER
SYSTEM CLEANUP PARTICIPANT
LOCAL CLEANUP NOT NEEDED
TERMLEVEL IS TASK
MEMSTALL RESOLUTION IS NO ACTION
EXITS DEFINED: MESSAGE, GROUP

SIGNALLING SERVICE

| MSGO ACCEPTED: | 6820 | NOBUFFER: | 0 |
| MSGO XFER CNT: | 3351 | LCL CNT: | 2532 | BUFF LEN: | 956 |
| MSGO XFER CNT: | 399  | LCL CNT: | 419  | BUFF LEN: | 4028 |
| MSGO XFER CNT: | 109  | LCL CNT: | 10   | BUFF LEN: | 8124 |
## RMF XCF Usage By System (outbound)

### (5 minute interval)

#### OUTBOUND FROM SYSA

<table>
<thead>
<tr>
<th>TO SYSTEM</th>
<th>TRANSPORT</th>
<th>BUFFER</th>
<th>OUT</th>
<th>SML</th>
<th>FIT</th>
<th>BIG</th>
<th>OVR</th>
<th>UNAVAIL</th>
<th>REJECT</th>
<th>PATHS</th>
<th>REQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>SYSB</td>
<td>DEFAULT</td>
<td>956</td>
<td>1,038</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>LARGE</td>
<td>16,316</td>
<td>50</td>
<td>60</td>
<td>0</td>
<td>40</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>SMALL</td>
<td>4,028</td>
<td>1</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SYSC</td>
<td>DEFAULT</td>
<td>956</td>
<td>1,568</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>LARGE</td>
<td>16,316</td>
<td>46</td>
<td>93</td>
<td>0</td>
<td>7</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>SMALL</td>
<td>4,028</td>
<td>125</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SYSD</td>
<td>DEFAULT</td>
<td>956</td>
<td>1,013</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>LARGE</td>
<td>16,316</td>
<td>33</td>
<td>91</td>
<td>0</td>
<td>9</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>SMALL</td>
<td>4,028</td>
<td>1</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

---

**TOTAL** 3,875
### Appropriate Number of Paths?

(5 minute interval)

<table>
<thead>
<tr>
<th>TO</th>
<th>Y DEVICE, OR</th>
<th>TRANSPORT</th>
<th>CLASS</th>
<th>REQ</th>
<th>AVG Q LNGTH</th>
<th>AVAIL</th>
<th>BUSY</th>
<th>RETRY</th>
</tr>
</thead>
<tbody>
<tr>
<td>SYSB</td>
<td>S IXCSTR1</td>
<td>DEFAULT</td>
<td>843</td>
<td>0.00</td>
<td>843</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S IXCSTR2</td>
<td>SMALL</td>
<td>1</td>
<td>0.00</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S IXCSTR3</td>
<td>LARGE</td>
<td>50</td>
<td>0.00</td>
<td>47</td>
<td>3</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S IXCSTR4</td>
<td>DEFAULT</td>
<td>214</td>
<td>0.00</td>
<td>214</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>SYSC</td>
<td>S IXCSTR1</td>
<td>DEFAULT</td>
<td>357</td>
<td>0.00</td>
<td>357</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S IXCSTR2</td>
<td>SMALL</td>
<td>125</td>
<td>0.00</td>
<td>125</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S IXCSTR3</td>
<td>LARGE</td>
<td>46</td>
<td>0.00</td>
<td>46</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S IXCSTR4</td>
<td>DEFAULT</td>
<td>1,231</td>
<td>0.00</td>
<td>1,231</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>SYSD</td>
<td>S IXCSTR1</td>
<td>DEFAULT</td>
<td>408</td>
<td>0.00</td>
<td>408</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S IXCSTR2</td>
<td>SMALL</td>
<td>1</td>
<td>0.00</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S IXCSTR3</td>
<td>LARGE</td>
<td>33</td>
<td>0.00</td>
<td>33</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S IXCSTR4</td>
<td>DEFAULT</td>
<td>623</td>
<td>0.00</td>
<td>623</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

----------

TOTAL 3,932

You do not need to drive BUSY to zero.
I tend not to worry about avail vs busy.
As Q Length approaches 1 or more, then assuming no other issues impacting transfers, you might add a path.
Detecting Transport Class Issues

- Run z/OS Health Checker

- D XCF,CLASSDEF
- D XCF,PI
- D XCF,GROUP
- COUPLE SIGNAL DETAIL
- RMF reports
“XCF Lost My Signal”

- Preposterous!
  - XCF does not lose signals
  - But... signals can be delayed indefinitely
  - Usually a self-inflicted problem by exploiter

- Where might the signal be?
  - Pending transfer on the sending side
    - Signals flowing?
    - Any path in restart? rebuild?
  - Pending delivery to the message exit on the target system
    - Queued to the member delivery queue
    - Scheduled for delivery
  - (similarly for the response, except reverse roles of systems)

- Need timely dumps from both systems to investigate
XCF Signal Floods

- We are seeing issues in the field where some application floods XCF with signals and others suffer as a result
  - Generally they can't keep up with their own flood
  - Or we get long queues of signals pending over the paths
  - Impact varies with who the “other” is

- Band Aids
  - Increase MAXMSG
  - Add signalling paths

- But understanding where the influx came from is critical
  - Know your workload
  - Use RMF Reports to see who is sending the signals

- Need timely dumps from both systems to investigate
  - Why sending? Why so many? Why not keeping up?
Problem Taxonomy

- Dead System
- Sick System
- Sysplex Fabric
- Sysplex Componentry
  - Coupling Facility
  - Signalling Service
  - **Couple Data Sets**
  - External Time Reference
- Configuration / Capacity
- Software Issues
Couple Data Set Concerns

- Connectivity
  - Access (see “Sysplex Fabric: Couple Data Sets”)
  - Implications for partial connectivity
- Capacity
  - Be sure all formatted for same number of systems
  - Otherwise needs to be formatted to meet needs of exploiter
    • Up to them to indicate when too small
- Performance
  - Response time
  - Contention
- Regression
Couple Data Sets and Partial Connectivity

- **Sysplex**
  - System needs connectivity to IPL (except XCF-Local mode)
  - Systems that lose access wait-state

- **ARM** – system cannot use ARM

- **BPXMCDS** - functions may hang until access is restored

- **CFRM** - systems that lose access wait-state

- **LOGR** – lose logger functions that need CDS

- **SFM** - full sysplex-wide access required for CONNFAIL=YES

- **WLM** – lose functions that need CDS
Sysplex Couple Data Set Capacity

- IXC202I “sysplex CDS is full (systems)”
- IXC700E “sysplex CDS is full (groups or members)”
- XCF_SYSPLEX_CDS_CAPACITY health check
- D XCF,COUPLE

<table>
<thead>
<tr>
<th>SYSPLEX COUPLE DATA SETS</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRIMARY</td>
</tr>
<tr>
<td>DSN: UTCXCF.SVPLEX1.COUPLE.PRI</td>
</tr>
<tr>
<td>VOLSER: X1CPLP</td>
</tr>
<tr>
<td>DEVN: 1C6E</td>
</tr>
<tr>
<td>FORMAT TOD</td>
</tr>
<tr>
<td>MAXSYSTEM MAXGROUP(PEAK)</td>
</tr>
<tr>
<td>MAXMEMBER(PEAK)</td>
</tr>
<tr>
<td>32   120 (110) 503 (56)</td>
</tr>
</tbody>
</table>
Function couple data sets should be formatted to support at least as many systems as supported by the sysplex couple data sets.

### Sysplex Couple Data Sets

<table>
<thead>
<tr>
<th>PRIMARY</th>
<th>DSN: UTCXCF.SVPLEX1.COUPLE.PRI</th>
</tr>
</thead>
<tbody>
<tr>
<td>VOLSER: X1CPLP</td>
<td>DEVN: 1C0E</td>
</tr>
<tr>
<td>FORMAT TOD</td>
<td>MAXSYSTEM MAXGROUP (PEAK) MAXMEMBER (PEAK)</td>
</tr>
</tbody>
</table>

### BPXMCDs Couple Data Sets

<table>
<thead>
<tr>
<th>PRIMARY</th>
<th>DSN: UTCXCF.SVPLEX1.QMVS1.R14.PRI</th>
</tr>
</thead>
<tbody>
<tr>
<td>VOLSER: X1CPLP</td>
<td>DEVN: 1C0E</td>
</tr>
<tr>
<td>FORMAT TOD</td>
<td>MAXSYSTEM</td>
</tr>
<tr>
<td>04/11/2003 21:00:01</td>
<td>8</td>
</tr>
</tbody>
</table>

### CFRM Couple Data Sets

<table>
<thead>
<tr>
<th>PRIMARY</th>
<th>DSN: UTCXCF.SVPLEX1.CFRMR18P.PRI</th>
</tr>
</thead>
<tbody>
<tr>
<td>VOLSER: X1CPLA</td>
<td>DEVN: B010</td>
</tr>
<tr>
<td>FORMAT TOD</td>
<td>MAXSYSTEM</td>
</tr>
<tr>
<td>12/01/2010 12:08:18</td>
<td>32</td>
</tr>
</tbody>
</table>

Oops
Function Couple Data Set Capacity

- **DISPLAY XCF,COUPLE,TYPE=ALL**
  - Shows format parameters (limits) for all but WLM

- **ARM**
  - D XCF,ARMSTATUS for registered elements vs TOTELEM
  - Create of policy fails if MAXELEM exceeded

- **BPXMCDS**
  - MODIFY BPXOINIT,FILESYS=DISPLAY,GLOBAL
  - BPXI043E “approaching mount table limit”
  - AMTRULES – no way to tell limit reached?
Function Couple Data Set Capacity

- **CFRM**
  - IXL013I, if “connection failure” = “max permitted connectors”
  - Create policy fails if exceed POLICY, CF, or STR limits
  - D XCF, STR will indicate need for larger CDS for reconciliation
  - D XCF, CF will indicate need for larger CDS for reconciliation
  - IXC502I “need bigger CDS for reconciliation”
  - IXC503I “need bigger CDS for reconciliation”
  - IXC514I “need bigger CDS for reconciliation”
Function Couple Data Set Capacity

- **LOGR**
  - IXG010E “CDS too small”
  - IXG261E “CDS becoming full”
  - IXG262A “CDS is essentially full”
  - IXG270I “CDS becoming full”

- **SFM**
  - All capacity issues are detected when creating policies

- **WLM**
  - All capacity issues are detected when installing your service definitions
Couple Data Set Performance

- Various CDS generally have relatively low request rates
  - Though may depend on application design/behavior
  - Sysplex monitors run amok?
    * Be careful how you run RMF and similar tools

- But good performance is needed to avoid
  - Removal of CDS
  - Application delays

- First resolve access issues if any
  - See “Sick System: DASD I/O Issues”
  - See “Sysplex Fabric: Couple Data Sets”
Typical sources of CDS performance issues

- **DASD**
  - Old and slow?
  - Device caching enabled?
  - Synchronous mirroring?

- **Record Size**
  - CDS Formatted with too much white space?

- **Contention**
  - Using MSGBASED processing for CFRM?
  - Reserves?
  - Other data sets with high request rates on same volume?
    - Increased request rates due to workload changes
Detecting CDS Performance Issues

- Likely need baselines for comparisons
  - Changes to request rates
  - Changes to DASD response times

- RMF DASD I/O reports.
  - Direct Access DASD Activity – post processor report
Couple Data Set Regression

- When IPLing the sysplex, your couple data sets should either be:
  - The ones most recently used by the sysplex (typical and best), OR
  - Freshly formatted, never before used (atypical, usually DR site)

- Couple data sets often contain status and configuration data

- Regressing to an older CDS is risky because the data in that CDS may not be consistent with:
  - The current configuration
  - Data recorded in other (non-regressed) couple data sets
  - Related application data recorded in other data sets

You might get some messages prompting operator “Should we use this CDS?” when IPLing. Once the sysplex is up, I don't know of any reliable way to detect that regressed CDS was used.
Couple Data Set Regression ...

- Regression typically occurs when you use:
  - A backup copy of the CDS
  - A former primary/alternate CDS no longer actively in use
  - A CDS that was previously in use by some other sysplex

- NOTE: Bringing an old CDS into use via ACOUPLE is safe

- Resetting the TOD clock can also wreak havoc since time stamps recorded in the CDS may suddenly seem to be:
  - In the distant past
  - From the future
Problem Taxonomy

- Dead System
- Sick System
- Sysplex Fabric
- Sysplex Componentry
  - Coupling Facility
  - Signalling Service
  - Couple Data Sets
  - **External Time Reference**
- Configuration / Capacity
- Software Issues
External Time Reference

- System is in a wait-state if not time synchronized with sysplex
  - See “Sysplex Fabric: External Time Reference”

- Server Time Protocol (STP)
  - Must provide and maintain a resilient configuration as there is potential for sysplex-wide outage!
  - Understand Primary, Backup, and Arbiter roles

- Operational issues
  - Reassigning roles after failure (or planned outage) of any server that has a special role
  - Dealing with STP events
  - Dealing with loss of time synchronization
External Time Reference

- D ETR,DATA shows STP info for local system
- Redundant timing links?
- Expected role?

```
D ETR,DATA
IEA386I 22.41.27 TIMING STATUS 932
SYNCHRONIZATION MODE = STP
   THIS SERVER IS A STRATUM 2
   CTN ID = POKSTP
   THE STRATUM 1 NODE ID = 002097.E40.IBM.00.000000699DF
   NUMBER OF USABLE TIMING LINKS = 23
   THIS STP NETWORK HAS NO SERVER TO ACT AS ARBITER
```
External Time Reference

- IEA031I “STP alert was issued to HMC”
- IEA382I “have single point of failure”
- IEA383I “have single point of failure”
- IEA388I “not connected to backup time server (or arbiter)”
- IEA389I “no server to act as backup timer (or arbiter)”
- IEA395I “switched to backup time server (or preferred)”

- IXC406I “not connected to same ETR”
- IEA394A “lost access to ETR (STP)”
- IEA015A “lost access to ETR (sysplex timer)”
Operator Alerts for STP Events

- Operator alerts sent to z/OS console as well as HMC for STP related hardware & timing events:
  - Dial-out time service outside allowable tracking range
  - Dial-out access failure
  - NTP server failure
  - NTP servers not synchronized

- IEA031I STP ALERT RECEIVED. ALERT CODE = nn

- Available in z/OS 1.11
- Also available at z/OS 1.9 and 1.10 with OA28323
- Supported by z10 and z9 servers with MCL driver 76
External Time Reference

- **Redbooks**
  - *Server Time Protocol Implementation Guide*
    www.redbooks.ibm.com/abstracts/sg247281.html
  - *Server Time Protocol Planning Guide*
    www.redbooks.ibm.com/abstracts/sg247280.html
  - *Server Time Protocol Recovery Guide*
    www.redbooks.ibm.com/abstracts/sg247380.html

*Need to be very careful with CTN. These books are excellent references. Exposed to suffering outages during migration to CTN, changing configurations, or when operational errors are made.*
Problem Taxonomy

- Dead System
- Sick System
- Sysplex Fabric
- Sysplex Componentry
  - Coupling Facility
  - Signalling Service
  - Couple Data Sets
  - External Time Reference
- Configuration / Capacity
- Software Issues
Configuration and Capacity Concerns

- Resiliency

- Workload changes
  - Growth
  - Reconfiguration

- That might lead to more:
  - XCF signals
  - CF requests
  - CDS accesses
Preventing Configuration Issues

- Adhere to best practices for availability
  - No single points of failure
  - Ensuring redundancy and fail-over capability often enough to mitigate configuration mistakes (perhaps until failures occur)

- Use IBM Health Checker for z/OS
  - My anecdotal experience leads me to believe that the number of multi-system outages has dropped significantly since the introduction of the health checker

www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP101966
Preventing Capacity Issues

- Good capacity planning
- Resize structures
- White space and head room for failover
RMF Report of CF Activity

<table>
<thead>
<tr>
<th>COUPLING FACILITY USAGE SUMMARY</th>
</tr>
</thead>
<tbody>
<tr>
<td>STORAGE SUMMARY</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ALLOC SIZE</th>
<th>% OF CF STORAGE</th>
<th>% IN USE</th>
<th>MAX % REQUESTED</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOTAL CF STORAGE USED BY STRUCTURES</td>
<td>1430M</td>
<td>9.5</td>
<td></td>
</tr>
<tr>
<td>TOTAL CF DUMP STORAGE</td>
<td>6M</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>TOTAL CF STORAGE AVAILABLE</td>
<td>13573M</td>
<td>90.4</td>
<td></td>
</tr>
</tbody>
</table>

- Bigger CF?
- Allow for failover? Maintenance?
- Structures properly sized?
- Force structures?
Methodology Concern

- We might have what appears to be a sysplex configuration and capacity issue
- That is really an application design issue
- For example:
  - An application might choose to share data by sending copies of the data to all systems in the sysplex instead of sharing a single copy of the data in a CF that is accessible to all
- I proceed on the assumption that you are more likely to change the configuration than the application
  - But one can make a case for looking at application issues before considering configuration and capacity issues
Typical causes of increased activity

- **Signals**
  - Merge of systems
  - Increase work
  - Application changes
  - GRS contention
  - Migration / Maintenance

- **CF requests**
  - Merge of systems
  - Increase in work
  - Change in type of work
  - Application changes
  - Processor upgrade or MCL application

- **CDS access**
  - Systems joining / leaving sysplex
  - Recovery
  - XCF groups joining / leaving
Track impact on sysplex componentry as workload changes

- RMF Reports of XCF Activity
  - Member sending signals
  - Usage by System
  - Signal path usage

- RMF Reports of CF activity
  - CF as a whole
  - Structure by Structure

- RMF Reports of DASD activity
  - As related to Couple Data Sets

Know your workload
Loss of capacity

- Expect failures. Plan for them.
  - Provide spare capacity and head room for fail over
  - Reduce MTTR by automating restart of failed applications and systems

- See “z/OS MVS: Planning Operations” for information on using Auto-IPL to have failed system automatically IPLed back into the sysplex

Problem Taxonomy

- Dead System
- Sick System
- Sysplex Fabric
- Sysplex Componentry
  - Coupling Facility
  - Signalling Service
  - Couple Data Sets
  - External Time Reference
- Configuration / Capacity
- Software Issues
Methodology Recap

- **At this point:**
  - We have neither dead systems nor sick systems
  - Sysplex componentry is accessible, operational, and performing
  - Configuration is sufficient to support the workload

- **If still have a problem, likely have an application specific issue**
Design decisions for sysplex applications

- Communication
  - XCF Signalling Service
  - CF List Structure

- Data sharing
  - Distributed
  - Centralized

- Serialization
  - ENQ
  - XES Locking Services
  - “Owner”

- Topology
  - Peer to peer
  - Master (fixed or dynamic)
  - Ring
  - Star

- Coordination
  - Signals
  - XES Sync points
  - ENQ, locks, latches
Those design decisions have consequences

- Scalability
- Performance characteristics
- Failure modes

- All of which impact problem diagnosis
  - Hard to make progress if don't understand how its put together
  - Will likely need application specific expertise
Preventing software issues

- Stay current with maintenance
- Stay current with releases

IBM Best Practice: Apply z/OS maintenance RSU regularly and HIPERs more often
  - http://www-03.ibm.com/support/techdocs/atsmastr.nsf/PubAllNum/Flash10106

IBM Best Practice: Subscribe to Red Alerts
  - http://www-947.ibm.com/support/entry/portal/Overview/Software/Software_support_(general)?pgel=wspace
Summary

- Sysplex problems can be difficult to diagnose
  - Complex inter-dependencies
  - Symptoms may not be directly related to the root cause

- Prevention is highly recommended
  - **Enable SFM with BCPii** to resolve dead system issues
  - Exploit SFM policy to resolve sick but not dead issues
  - Use z/OS Health Checker to verify configuration

- Use the “Problem Taxonomy” to guide your diagnosis efforts. I believe it provides a disciplined approach that can help identify the likely area for root cause, even in the face of imperfect knowledge of the system
For more information

- Setting Up a Sysplex
- Problem Determination Guide
- Initialization and Tuning Reference
- z/OS MVS Data Areas
- MVS Sysplex Services Guide
- MVS Sysplex Services Reference
- RMF Users Guide
Questions?

Diagnosing Sysplex Problems
Session 13402

Please fill out the online session evaluation at either:
• SHARE.org/SanFranciscoEval, or
• Aim your smartphone at this QR code below:
Appendix
Non-disruptive CF dump

SETXCF DUMPCF,
{CFNAME=cfname[,TYPE=NONDISRUPTIVE|DISRUPTIVE][,UNCOND=NO|YES]}
{STRNAME=strname}
CFNAME=cfname
Allows the operator to specify the CF to be dumped.

TYPE=DISRUPTIVE|NONDISRUPTIVE
Allows the operator to optionally request a disruptive CF dump.
Default: Nondisruptive

UNCOND=YES|NO
Allows the operator to bypass a confirmation if a disruptive CF dump is requested.
Default: No

STRNAME=strname
Allows the operator to request a CF dump by specifying a structure name. The CF(s) in which the structure resides will be dumped and the dump requested will be non-disruptive.

OA35342 – Introduced the z/OS operator command to collect a non-disruptive serialized CF dump. CF dumps are reviewed by IBM hardware support.
Getting sysplex dumps

- If CF issue, activate SYSXES Ctrace
  - TRACE CT,16M,COMP=SYSXES
  - R XX,OPTIONS=(HWLAYER,REQUEST,CONFIG,CONNECT,RECOVERY,LOCKMGR),END

- If XCF issue, activate SYSXCF ctrace
  - TRACE CT,ON,COMP=SYSXCF
  - R xx,OPTIONS=(GRPNAME=(GEOPLEX),SIGNAL,GROUP,SFM),END

- Place dump command in IEADMCxx parmlib member:
  - JOBNAME=(XCFAS),DSPNAME=('XCFAS'.*),
  - SDATA=(ALLNUC,CSA,PSA,LPA,LSQA,NUC,RGN,SQA,SUM,SWA,TRT,XESDATA,COUPLE),
  - REMOTE=(SYSLIST=*(XCFAS'),DSPNAME,SDATA)),END
  - Initiate the dump on one system in the sysplex:
    - DUMP COMM=(meaningful dump title),PARMLIB=xx