#SHAREorg

SHARE

z/OS Parallel Sysplex z/OS 2.1 Update

Mark A. Brooks: mabrook@us.ibm.com
IBM Corporation

Diana M. Henderson: dmhender@us.ibm.com
IBM Corporation

Tuesday August 13, 2013; 11:00 AM – 12:15 PM
Session Number 14231

SHARE in Boston

Trademarks

The following are trademarks of the International Business Machines Corporation in the United States and/or other countries.

AIX*	DB2*	HiperSockets*	MQSeries*	PowerHA*	RMF	System z*	zEnterprise*	z/VM*
BladeCenter*	DFSMS	HyperSwap	NetView*	PR/SM	Smarter Planet*	System z10*	z10	z/VSE*
CICS*	EASY Tier	IMS	OMEGAMON*	PureSystems	Storwize*	Tivoli*	z10 EC	
Cognos*	FICON*	InfiniBand*	Parallel Sysplex*	Rational*	System Storage*	WebSphere*	z/OS*	
DataPower*	GDPS*	Lotus*	POWER7*	RACF*	System x*	XIV*		

* Registered trademarks of IBM Corporation

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 therefrom.
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.
IT Infrastructure Library is a registered trademark of the Central Computer and Telecommunications Agency which is now part of the Office of Government Commerce.
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.
Java and all Java based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.
Linear Tape-Open, LTO, the LTO Logo, Ultrium, and the Ultrium logo are trademarks of HP, IBM Corp. and Quantum in the U.S. and
Linux is a registered trademark of Linus Torvalds 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.
OpenStack is a trademark of OpenStack LLC. The OpenStack trademark policy is available on the OpenStack website.
TEALEAF is a registered trademark of Tealeaf, an IBM Company.
Windows Server and the Windows logo are trademarks of the Microsoft group of countries.
Worklight is a trademark or registered trademark of Worklight, an IBM Company.
UNIX is a registered trademark of The Open Group in the United States and other countries.
* Other product and service names might be trademarks of IBM or other 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.

This information provides only general descriptions of the types and portions of workloads that are eligible for execution on Specialty Engines (e.g, zIIPs, zAAPs, and IFLs) ("SEs"). IBM authorizes customers to use IBM SE only to execute the processing of Eligible Workloads of specific Programs expressly authorized by IBM as specified in the "Authorized Use Table for IBM Machines" provided at www.ibm.com/systems/support/machine_warranties/machine_code/aut.html ("AUT"). No other workload processing is authorized for execution on an SE. IBM offers SE at a lower price than General Processors/Central Processors because customers are authorized to use SEs only to process certain types and/or amounts of workloads as specified by IBM in the AUT.

Trademarks

The following are trademarks of the International Business Machines Corporation in the United States and/or other countries.

IBM®	MQSeries®	S/390®	z9®
ibm.com®	MVS™	Service Request Manager®	z10™
CICS®	OS/390®	Sysplex Timer®	z/Architecture®
CICSPlex®	Parallel Sysplex®	System z®	zEnterprise™
DB2®	Processor Resource/Systems Manager™	System z9®	z/OS®
eServer™	PR/SM™	System z10®	z/VM®
ESCON®	RACF®	System/390®	z/VSE®
FICON®	Redbooks®	Tivoli®	zSeries®
HyperSwap®	Resource Measurement Facility™	VTAM®	
IMS™	RETAIN®	WebSphere®	
IMS/ESA®	GDPS®		
	Geographically Dispersed Parallel Sysplex™		

The following are trademarks or registered trademarks of other companies.

IBM, z/OS, Predictive Failure Analysis, DB2, Parallel Sysplex, Tivoli, RACF, System z, WebSphere, Language Environment, zSeries, CICS, System x, AIX, BladeCenter and PartnerWorld are registered trademarks of IBM Corporation in the United States, other countries, or both.
DFSMShsm, z9, DFSMSrmm, DFSMSdfp, DFSMSdss, DFSMS, DFS, DFSORT, IMS, and RMF are trademarks of IBM Corporation in the United States, other countries, or both.
Java and all Java-based trademarks are trademarks of Sun Microsystems, Inc. in the United Sta/tes, 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.
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.

* 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.

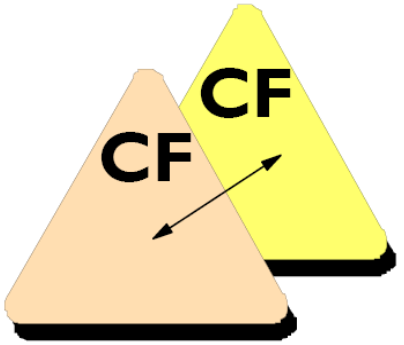
Prices subject to change without notice. Contact your IBM representative or Business Partner for the most current pricing in your geography.

Agenda

- Hardware Updates
  - CFCC Level 19
  - CFCC Level 18
  - Parallel Sysplex Coupling Links
  - Server Time Protocol (STP)
- Software Updates
  - z/OS V2R1
  - z/OS V1R13
  - z/OS V1R12
- Summary

Agenda

- Hardware Updates
  - **CFCC Level 19**
  - CFCC Level 18
  - Parallel Sysplex Coupling Links
  - Server Time Protocol (STP)
- Software Updates
  - z/OS V2R1
  - z/OS V1R13
  - z/OS V1R12
- Summary



CFLEVEL 19

- IBM zEnterprise™ EC12 GA2 (zEC12) and BC12 (zBC12)
  - Available September, 2013
- Performance Improvements and Resiliency Improvements
- Prerequisites
  - z/OS V2.1 or later with PTFs for OA38734 and OA38781
  - z/VM V6.3 or later with PTFs for guest exploitation

Dedicated vs. Shared CF Engines

Dedicated Physical CF Engines

Shared Physical CF Engines

ICF ICF

ICF ICF ICF ICF ICF ICF ICF ICF ICF ICF

CFLP1	CFLP2	CFLP3	CFLP4	CFLP5
Logical ICF	Logical ICF	Logical ICF	Logical ICF	Logical ICF
Logical ICF	Logical ICF	Logical ICF	Logical ICF	Logical ICF
	Logical ICF	Logical ICF		Logical ICF
	Logical ICF	Logical ICF		
		Logical ICF		
		Logical ICF		

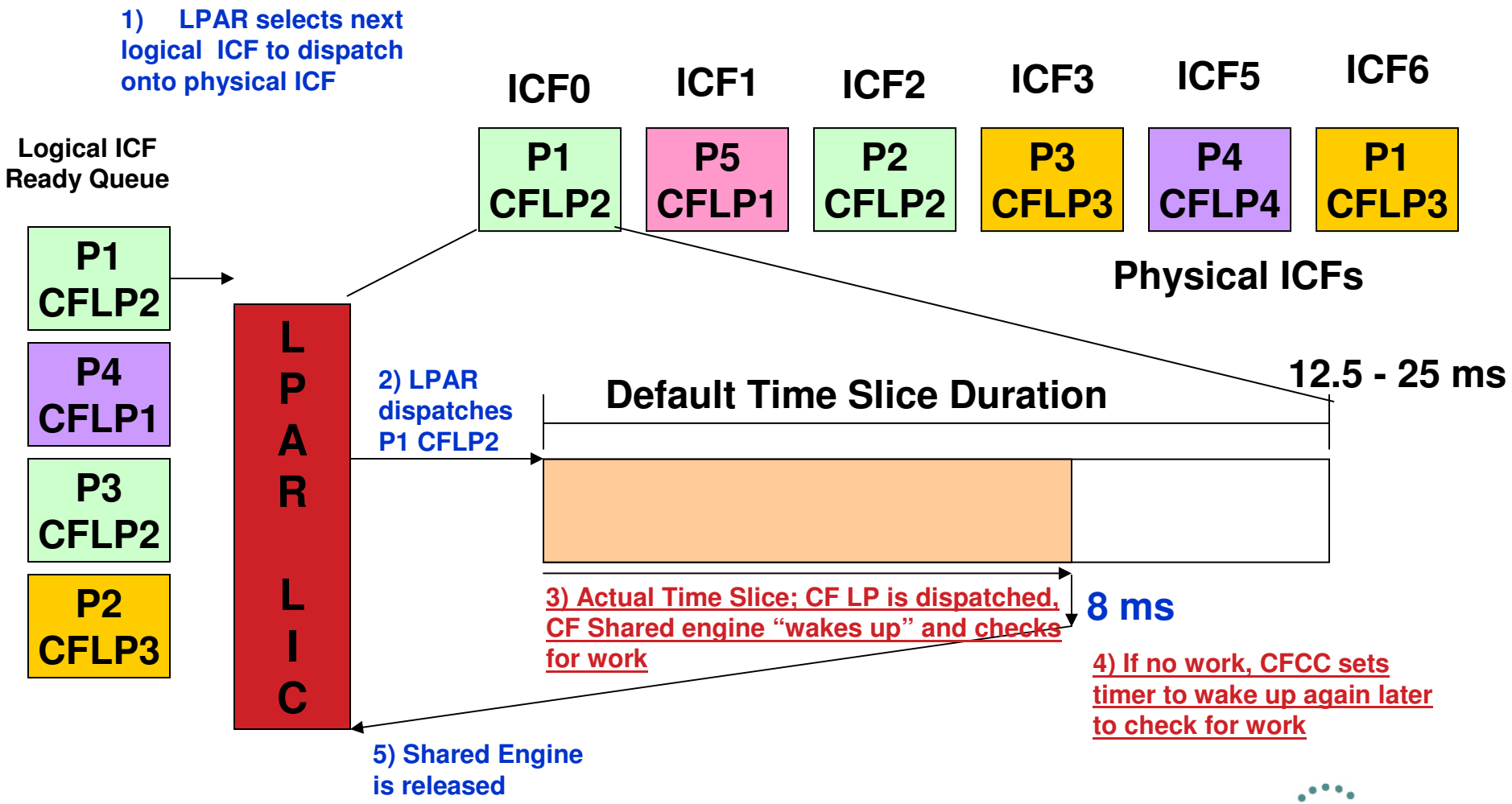
CFCC Polling Model

- Coupling Facilities (CFs) are polling-based mechanisms
  - Internal work dispatching in the CF polls for and then processes work
- Polling provides consistently good service times **only** when there is a **dedicated engine** assigned to the CF image
- **Shared Engine CFs** used with Shared Processors can yield **unpredictable, variable, and High CF service times**
  - High Sync-to-Asnych request conversion

Dynamic CF Dispatching (DCFD)

- An enhancement to the Polling Mechanism with **a time-based algorithm used to share engines more effectively** in a shared engine CF
- Enables the CF to heuristically determine when to give up control of the physical processor when no CF commands are queued for execution at observed arrival rates
- DCFD=NO: the CF runs the polling model, generating good response times
  - The CF will not release its shared engine until the engine is taken away by PR/SM LPAR and given to another CF LP
- DCFD=YES: CF performs its own time-based time slicing at more granular level than PR/SM LPAR
  - CF LP is dispatched, CFCC “wakes up” and checks for waiting work
  - If there’s no work, CFCC sets a timer, and goes back to sleep, releasing the shared engine
    - Little work, lots of sleep; The shared CP is made available to other LPs
    - Lots of work, little sleep; ~Polling

Dynamic CF Coupling Dispatching



How does DCFD differ from LPAR Event-Driven Dispatching?

- **Dynamic CF Dispatching**
  - DCFD=NO, the polling loop mechanism is still in effect
  - The engine driving the CF partition will do nothing until its request has completed
  - It's essentially still TIME-driven dispatching – with heuristically shortened time slices compared LPAR time slicing with the traditional Polling Model

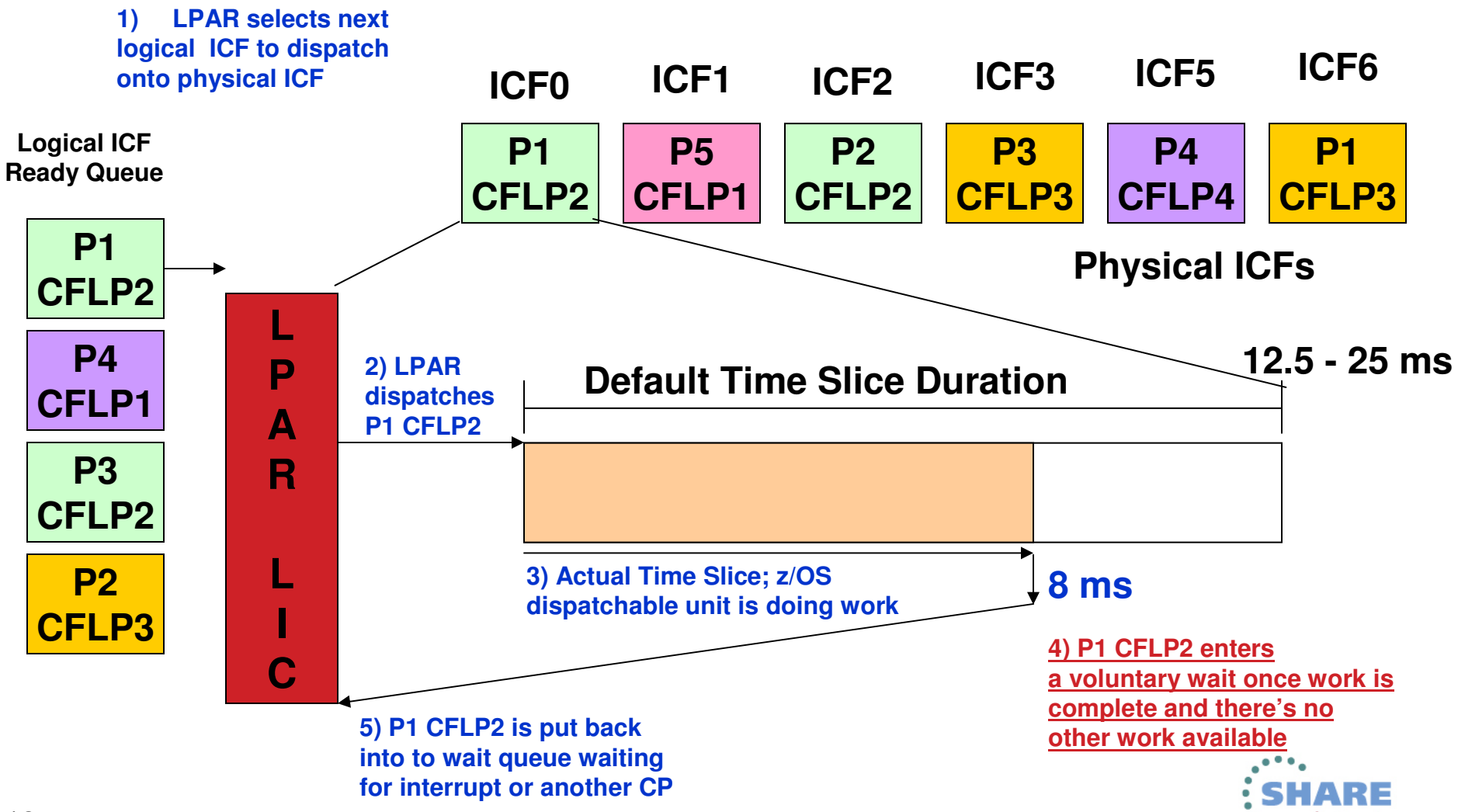
- **LPAR Event-Driven Dispatching**
  - PR/SM LPAR can take a physical CP away from a logical CP BEFORE its LPAR time slice ends because that CP is no longer doing any productive work
    - This is an Interrupt
    - Interrupts have shown to generate the most efficient use of CP resources

Coupling Thin Interrupts

Goal: Expedite the dispatching of the partition

- Generate Coupling Thin Interrupts to **wake up and dispatch a shared processor in a timely fashion to service work** as opposed to having the processor wait for PR/SM LPAR to perform its natural processing until that work can be processed
- Once the CF image gets dispatched, the existing “poll for work” logic in both CFCC and z/OS can be used to locate and process the work
- CF will give up control when work is exhausted (or when LPAR kicks it off the shared processor)

Coupling Thin Interrupts



Polling / DCFD / Coupling Thin Interrupts DYNDISP Modes

CFCC controls Coupling Thin Interrupts under the CF's Dynamic CF Dispatching engine (DYNDISP)

CF Polling	Dynamic CF Dispatching	Coupling Thin Interrupts
DYNDISP=NO	**DYNDISP=YES**	**DYNDISP=THININTERRUPT**
LPAR Time slicing	CF time-based algorithm for CF engine sharing	CF releases shared engine if no work left to be done
- CF does not "play nice" with other shared images sharing the processor - CF controls processor long after work is exhausted	- CF does own time slicing - More Effective engine sharing than polling - Blind to presence or absence of work to do - No Interrupt Available	- Event-Driven Dispatching -CF relies on generation of thin interrupt to dispatch processor when new work arrives - Now the most effective use of shared engines across multiple CF images

Coupling Thin Interrupt Exploitation When can this be used?

- Provides hardware, firmware, and software support for Coupling Facility “thin interrupts” to be generated when events such as the following occur:

- **On the CF side**: (CFLEVEL 19) – DYNDISP=THININTERRUPT
  - a CF command is received by a shared-engine CF image (e.g. arrival of a primary CF command that needs to be processed)
  - a CF signal is received by a shared-engine CF image (e.g. arrival of a secondary message duplexing signal that needs to be processed)
  - Completion of a secondary message sent by the CF (e.g. completion of a secondary message duplexing signal sent by the image)

Customer Value

- **Elapsed time performance improvements for CF-based messaging functions** (e.g. MQ, XCF signaling, IMS SMQ) and any CF- exploiting functions operating in environments with significant amounts of asynchronous CF accesses
  - Improved performance and throughput
- **Faster and far more consistent CF service times** will be achieved by shared-engine CF images
  - Should enable shared-engine CF images acceptable for use in a broader range of configurations and uses (e.g. test versus development versus production)
- **Multiple test and/or production CF images can be aggregated onto a single shared CF engine**, with reasonably good performance
  - Avoids the need to dedicate an ICF engine (or several) to each CF image
- **Configuration simplification**
  - Makes shared-engine CFs as simple and natural to use as shared-engine z/OS images are today

Dedicated-engine CF performance will still provide best CF image performance

CFLEVEL 19 – CF Flash Exploitation

- Currently, the CF is a pure "real memory" system – all CF structures are allocated and backed entirely by real memory in the CF image
  - No paging, no virtual storage, no disk I/O at all

- Adding relatively slower Flash memory to a CF structure, therefore, cannot speed anything up
  - CF Flash exploitation is

    NOT a performance enhancement item

Statement of Direction: General Availability Target: 1H2014

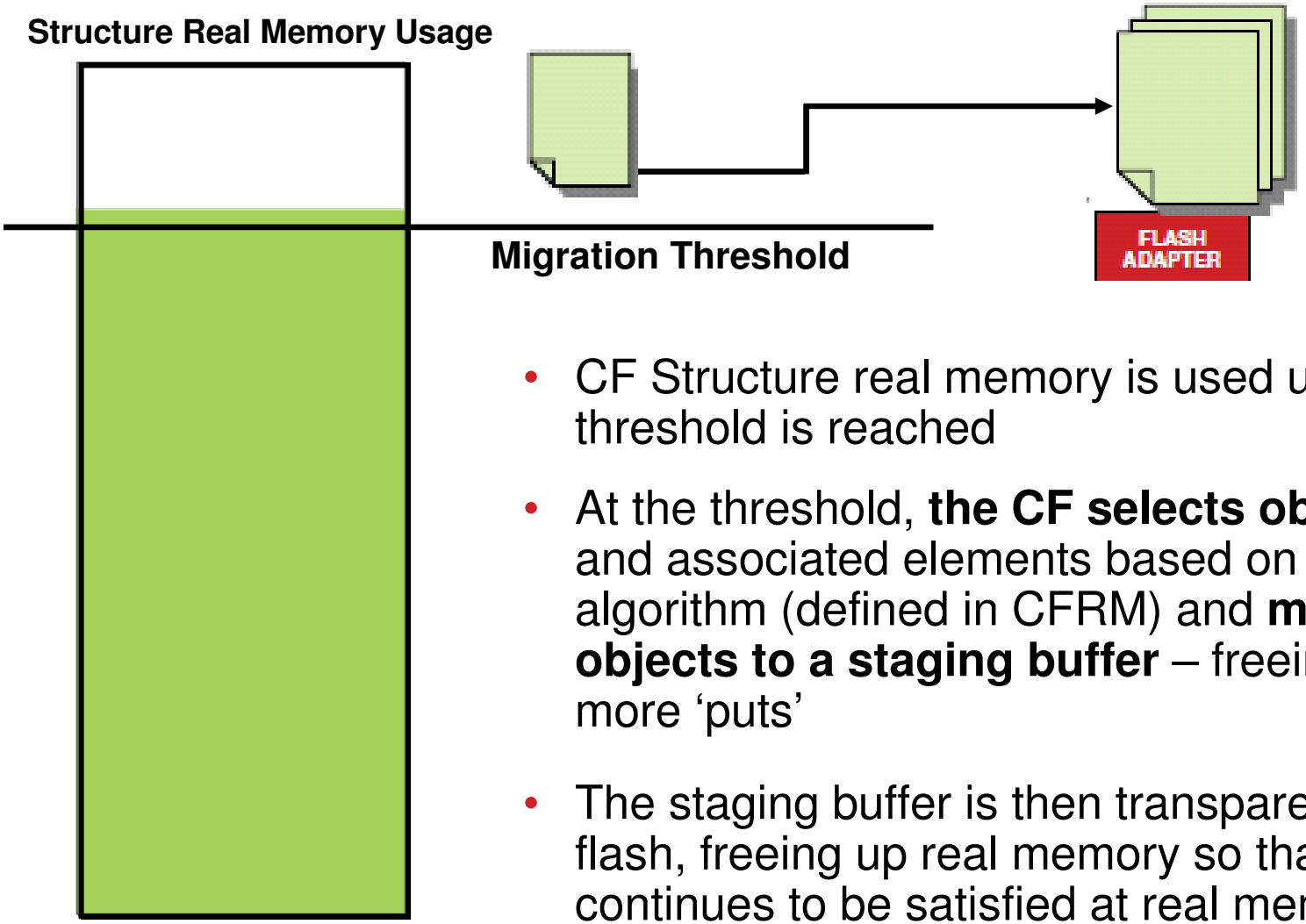
CF Flash Initial Exploitation

- Initial CF Flash exploitation is targeted to MQ shared queues application structures
- **Provides emergency capacity to handle MQ shared queue buildups during abnormal situations**, such as where “putters” are putting to the shared queue, but “getters” are transiently not getting from the shared queue – or other such transient producer/consumer mismatches on the queue
- Requirements:
  - CFCC support planned for CFLEVEL 19
  - z/OS support planned for 2.1 and potentially 1.13 (with PTFs)
  - No new level of MQ required

CF Flash – Structure Characteristics

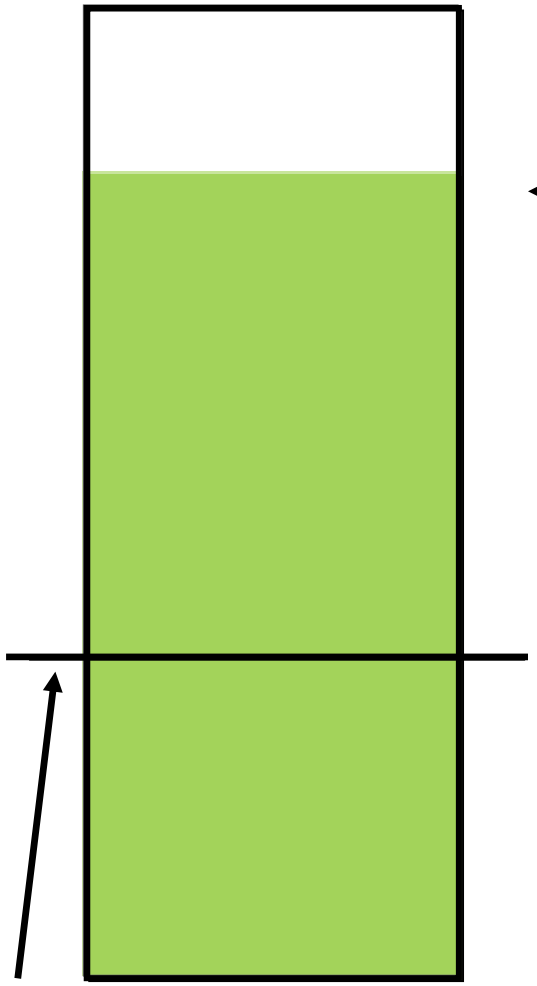
- Flash Memory Assignment
  - Flash memory **exists on a flash card** in the CPC and is **assigned** to a CF partition via hardware definition panels, **just like it is assigned to the z/OS partitions**
- Flash Permit
  - Once Assigned, the **CFRM Policy in z/OS is used to permit certain amount of memory** to be used by the structure, size defined at a structure level granularity
- No Pre-Assignment
  - Flash memory **amounts can be permitted but not pre-assigned**
- Overhead
  - **Structure size requirements will increase** for those structures assigned flash memory due to **additional control objects in the CF**
  - **CFSIZER has been updated** to account for these larger structures during capacity planning exercises

CF Flash: How Does It Work? MIGRATION



- CF Structure real memory is used until a **migration** threshold is reached
- At the threshold, **the CF selects objects** (list entries and associated elements based on a tailored algorithm (defined in CFRM) and **migrates those objects to a staging buffer** – freeing memory for more 'puts'
- The staging buffer is then transparently moved to flash, freeing up real memory so that write activity continues to be satisfied at real memory speed

CF Flash: How Does It Work? PREFETCH



- When the **structure's real memory utilization goes BELOW a prefetch threshold**, or when certain other list-specific criteria are reached, the **CF starts to prefetch** objects (list entries and associated data elements) **FROM Flash memory** back into a staging buffer in real memory
- Again, a tailored algorithm, defined in CFRM on a per structure basis, is used to determine which objects to retrieve
- The structure is then repopulated from the staging buffer so the objects can be retrieved on "get" references at real memory speeds

CF Flash Considerations - Migration

- If migration to flash keeps up, structure real memory never fills; thus, write activity continues to be satisfied at real memory speeds
  - If migration does NOT keep up, there is the potential that structure real memory will fill up, at which point, writes would be delayed until migration catches up enough to free up sufficient real memory objects to permit the new writes to succeed.

CF Flash Considerations - PreFetch

- If prefetching keeps up, the “get” references to objects never have to wait for anything to be brought back into real; thus these references continue to be satisfied at real memory speeds
  - If prefetching does not keep up, there is the potential that an object would be requested or referenced that is still on flash. At this point a “**flash fault**” would occur, and the request would be delayed until the necessary objects were brought back in to real memory.
  - Truly random references to objects can result in Flash faults too (like a Page Fault) – delaying those references until the Flash fault is resolved
  - NOTE: CF requests NEVER wait synchronously for Flash accesses to take place; the delayed requests always return to z/OS, who redrives them to the CF in hopes that the CF will have resolved the Flash issue that caused the redrive to occur.

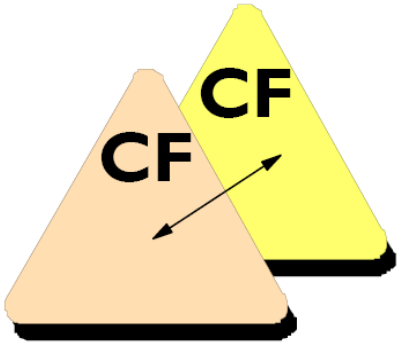
CF Flash Considerations

- CF Flash memory is assigned *dynamically* to structures – *not statically*

- There are both structure sizing (CFSIZER) and CF real storage planning implications associated with CF Flash

Agenda

- Hardware Updates
  - CFCC Level 19
  - **CFCC Level 18**
  - Parallel Sysplex Coupling Links
  - Server Time Protocol (STP)
- Software Updates
  - z/OS V2R1
  - z/OS V1R13
  - z/OS V1R12
- Summary



CFLEVEL 18

- IBM zEnterprise™ EC12 (zEC12)
  - Available September, 2012
- Serviceability and Performance Enhancements
- Requirements
  - z/OS V1.13 and V1.12 with PTFs
  - z/VM 5.4 with PTFs for guest exploitation

CFLEVEL 18 - Performance Enhancements
Cache Write Around

- Enhancements to the IXLCACHE macro interface and CFCC allow exploiters to optionally request that writes to CF Cache be suppressed if:
  - The data is not currently stored in the CF Cache structure, and
  - Only the local cache has registered interest
- Can intelligently decide which entries should be written to the cache and which should just be “written around” directly to disk
  - Helps preserve application “working set”
  - Suppressing writes reduces work that CF must perform
- Requires application exploitation, and:
  - APAR OA37550 at z/OS V1R12 and up
  - z/VM 5.4 with PTFs for guest exploitation
- Also note: Roll back to CFCC Release 17 (MCL12)

CFCC Level 18 - Performance Enhancements
Cache Write Around …

- This support is planned for future exploitation by DB2 during batch update/insert processing to conditionally write to group buffer pool (GBP)
  - Helps avoid over running cache structures with directory entries and changed data that are not part of the normal working set
  - Avoids thrashing the cache through LRU processing
  - Avoids castout processing backlogs and delays
- Intended to improve DB2 batch performance

CFCC Level 18 - Performance Enhancements
Cache Write Around …

- Online transactions may encounter less delay during large concurrent batch updates
- Requires
  - CFCC support and z/OS support (previous slide)
  - DB2 exploitation is planned for future release

CFCC Level 18 - Performance Enhancements Internal CFCC Changes for Cache Structures

- Elapsed time improvements when dynamically altering cache structure
  - Entry / Element ratio
  - Size
- CF Storage Class and Castout Class contention avoidance
  - Changes the way serialization is performed on individual storage class and castout class queues
  - Reduces storage class and castout class latch contention.
- Throughput enhancements for parallel cache castout processing.

CFLEVEL 18 – Resiliency and Performance Delete Name Extensions for Cache Structures

- Halt on Changed
  - Allows exploiter to redrive cast out processing when changed data is unexpectedly encountered
  - Helps avoid the accidental deletion of directory entries which might lead to data corruption
- Optional suppression of cross invalidate signals
  - Helps improve delete name performance, particularly at distance
  - But local vector will not reflect validity of locally cached data
- Requires
  - OA38419 at z/OS V1R12 and up
  - Exploitation
    - DB2 is planned for future release
- Also note
  - Rolling back to CFCC Release 17 (MCL13)
  - Rolling back to CFCC Release 16 is also expected

CFLEVEL18 – Resiliency
Register Attach Validation

- Verification of local cache controls for a Coupling Facility cache structure connector.
  - Performed when connection registers interest in data
  - System gathers diagnostics if discrepancies detected
  - System proactively takes steps to mitigate problem before data corruption can occur

- Also note
  - Rolled back to CFCC Release 17 (MCL12)

CFLEVEL 18 – RAS Enhancements

- Background structure deallocation
  - XES task freed to perform other work and requests instead of redriving the deallocate command in the foreground
- This change also allows for better structure dumps
  - Extended dumping can be performed when structure damage is detected allowing for the capturing of more content for error analysis
  - In the past, foreground deallocation could cause structure dumps to be truncated

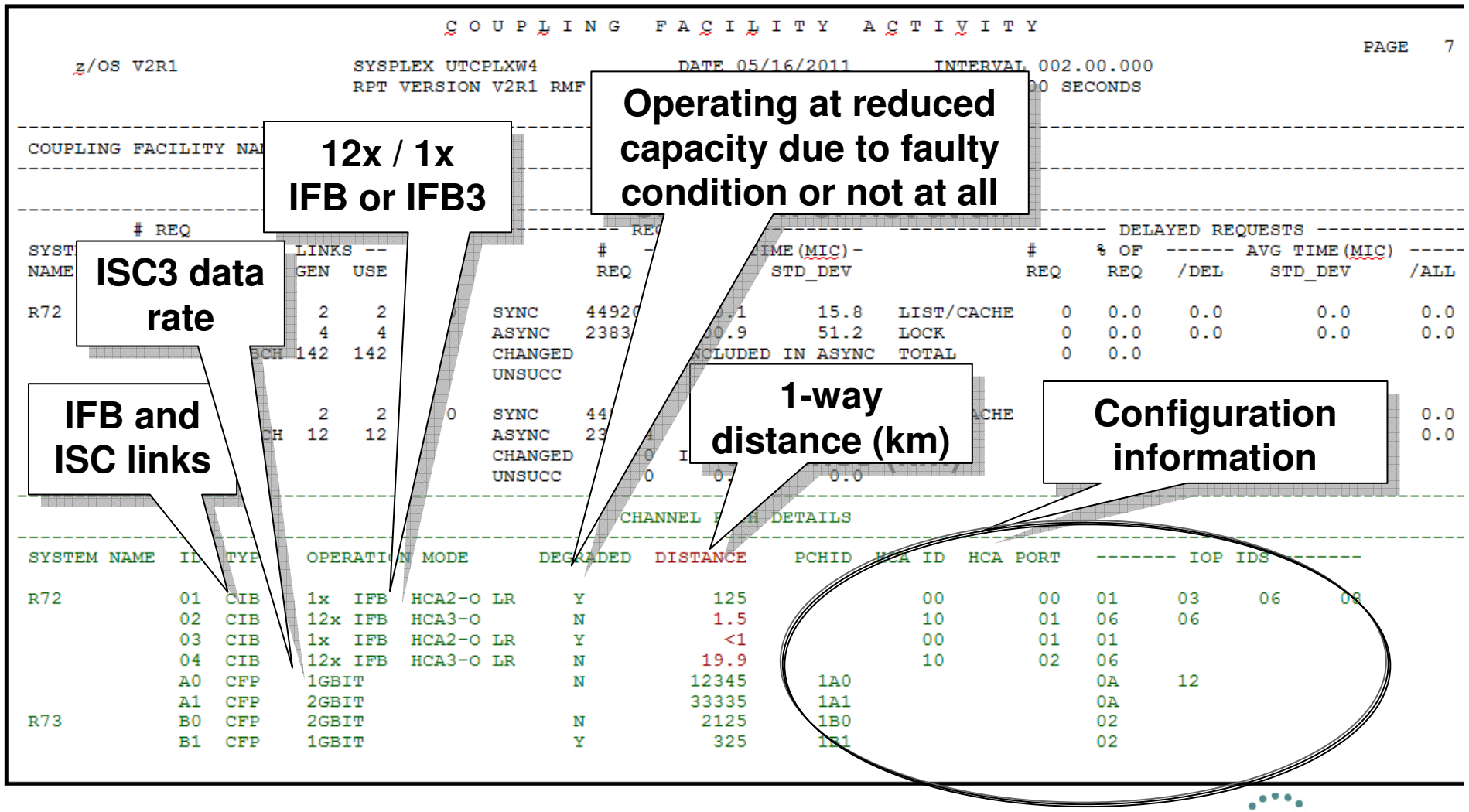
CFCC Level 18 – RAS Enhancements

- Enhanced CFCC tracing support
  - Significantly enhanced trace points, especially in troublesome areas
    - Latching (CP and suspend),
    - Locate queue and suspend queue management/dispatching,
    - Duplexing protocols (especially suppression and clear-off processing),
    - Sublist notification,
    - Alter/ECR,
    - Castout processing
    - RCC cursors, etc.
  - Quantity gathered
    - Trace buffer size increase
  - Trace buffer granularity –
    - Special trace buffers for specific types of traces (e.g. Alter/ECR)
  - Controls
    - Default/detail/exception levels of tracing, activated via OPERMSG commands

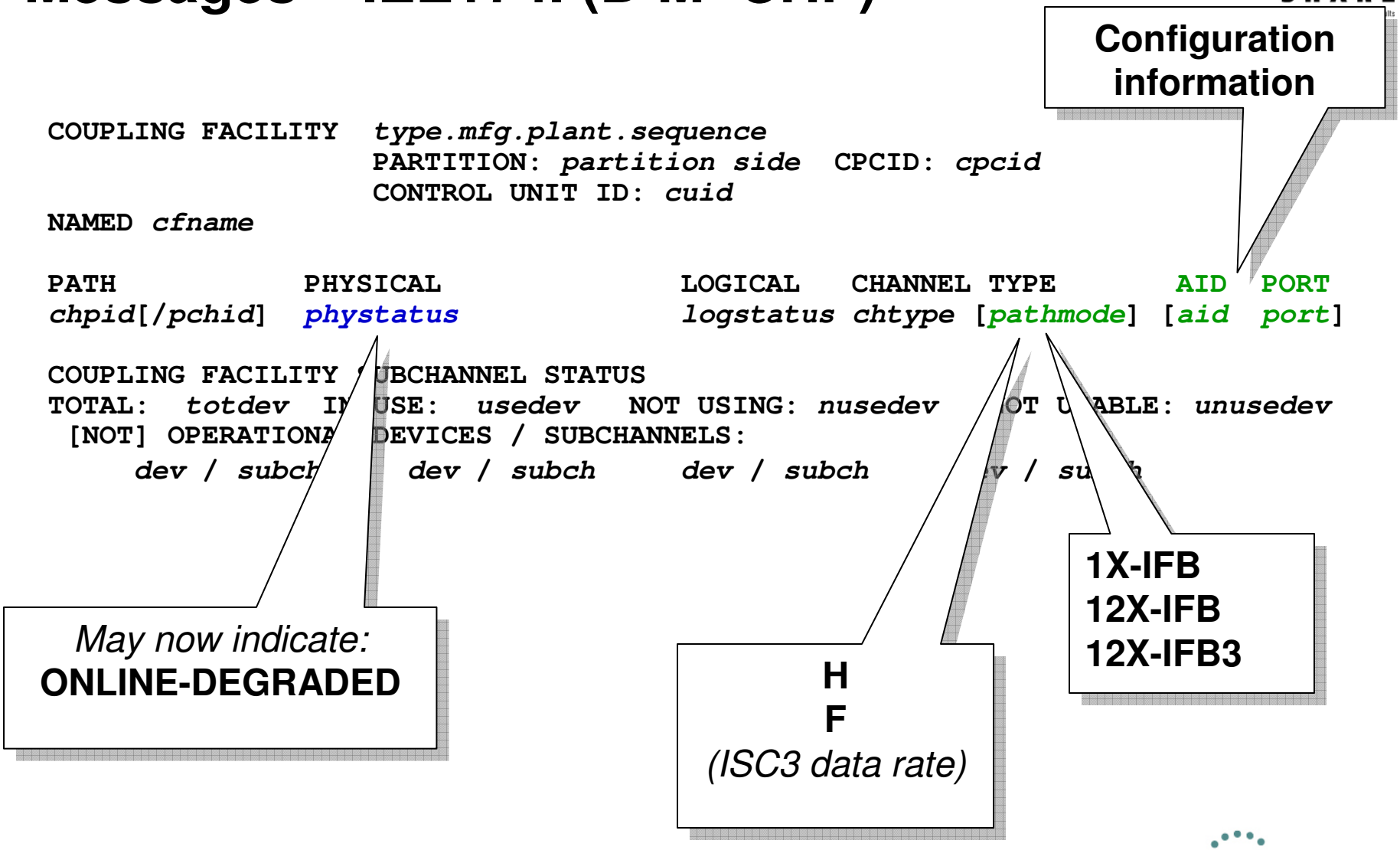
CFCC Level 18 – RMF Channel Path Details

- Provides enhanced reporting of channel path characteristics for Parallel Sysplex Coupling Facility CIB or CFP links
- Helps understand link performance, response times and coupling overheads
  - Channel path ID
  - Channel path operation mode
  - Channel path degraded status
  - Channel path distance number
  - Accessible I/O processors
  - Channel path type acronym
  - Physical channel path ID
  - Host channel adapter ID
  - Host channel adapter port
- New Channel Path Details section
  - RMF Coupling Facility Postprocessor Report
  - RMF Monitor III CFSYS Report
  - XML report
- New display commands
- APAR OA38312 for support on z/OS V1R12 and up
- APAR OA37826 for RMF support

CFCC Level 18 – RMF Channel Path Details



Messages – IEE174I (D M=CHP)



Messages – IXL150I (DISPLAY CF output)

```
IXL150I hh.mm.ss DISPLAY CF
COUPLING FACILITY type.mfg.plant.sequence
                  PARTITION: partition side  CPCID: cpcid
                  LP NAME: lparname   CPC NAME: cpcname
                  CONTROL UNIT ID: cuid
NAMED cfname
. . .
    DYNAMIC CF DISPATCHING: ON|OFF]
    COUPLING FACILITY IS standalonestate
. . .
PATH            PHYSICAL              LOGICAL   CHANNEL TYPE        AID  PORT
chpid[/pchid]   phystatus             logstatus chtype [pathmode]   aid  port
. . .
REMOTELY CONNECTED COUPLING FACILITIES
       CFNAME               COUPLING FACILITY
       --------             -------------------------
       rfcfname             rftype.rfmfg.rfplant.rfsequence
                            PARTITION: partition rfside CPCID: rfcpcid
                            CHPIDS ON cfname CONNECTED TO REMOTE FACILITY
                            RECEIVER: CHPID    TYPE
                                      rfchpid  rfchtype [rfpmode]
                            SENDER:   CHPID    TYPE
                                      rfschpid rfschtype [rfspmode]
                     [* = PATH OPERATING AT REDUCED CAPACITY]
```

CFLEVEL 18 Summary

- CF Cache Write Around
- Internal CFCC Changes for Cache Structures
- Delete Name Extensions for Cache Structures
- Register Attach Validation
- RAS Enhancements
- Enhanced RMF Channel Path Reporting

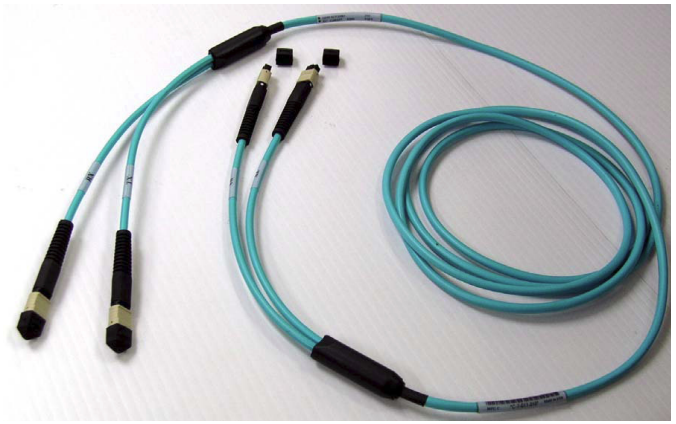
CFCC Level 18 or 19 - Migration

- In general, get to most current LIC levels
- Use CF Sizer website to check/update structure sizes:
  - CF structure sizes may increase when migrating to CFCC Level 18 or 19 from earlier levels due to additional CFCC controls
  - Improperly sized structures can lead to outages !
- Minimum CFCC image size is 512MB

www.ibm.com/systems/support/z/cfsizer/

Agenda

- Hardware Updates
  - CFCC Level 19
  - CFCC Level 18
  - **Parallel Sysplex Coupling Links**
  - Server Time Protocol (STP)
- Software Updates
  - z/OS V2R1
  - z/OS V1R13
  - z/OS V1R12
- Summary



Coupling Link Choices - Overview

- **ISC (Inter-System Channel / HCA2-C / CPC-to-IO Fanout)**
  - Fiber optics
  - I/O Adapter card
  - 10km, 20km support with RPQ 8P2197 as carry forward only, and longer distances with qualified WDM solutions
- **PSIFB (1x IFB / HCA2-O LR or HCA3-O LR / CPC-to-CPC Fanout)**
  - Fiber optics – uses same cabling as ISC
  - 10km and longer distances with qualified WDM solutions
  - Supports multiple CHPIDs per physical link
  - Multiple CF partitions can share physical link
- **PSIFB (12x IFB / HCA2-O or HCA3-O / CPC-to-CPC Fanout)**
  - 150 meter max distance optical cabling
  - Supports multiple CHPIDs per physical link
  - Multiple CF partitions can share physical link
- **IC (Internal Coupling Channel / Internal CPC)**
  - Microcode - no external connection
  - Only between partitions on same processor

System z – CF Link Connectivity Distances and Rates (Peer Mode only)

Connectivity Options	zEC12 / zBC12 / z196 / z114 ISC-3	zEC12 / zBC12 / z196 / z114 1x InfiniBand	zEC12 / zBC12 / z196 / z114 12x InfiniBand
zEC12 / zBC12 / z196 / z114 / z10 ISC-3 (RPQ 8P2197 – 20 km)	1 Gbps	N/A	N/A
zEC12 / zBC12 / z196 / z114 / z10 ISC-3 (10/100 km)	2 Gbps	N/A	N/A
zEC12 / zBC12 / z196 / z114/ z10 1x-InfiniBand (10/100 km)	N/A	5 Gbps	N/A
zEC12 / zBC12/ z196 / z114 /z10 12x InfiniBand (150 m)	N/A	N/A	6 GBps

Coupling Technology vs Host Processor Speed

Host effect with primary application involved in data sharing

Chart based on 9 CF ops/Mi - may be scaled linearly for other rates

Host / CF	z10 BC	z10 EC	z114	z196	zBC12	zEC12
z10 BC ISC3	16%	18%	17%	21%	19%	24%
z10 BC 1x IFB	13%	14%	14%	17%	18%	19%
z10 BC 12x IFB	12%	13%	13%	15%	15%	17%
z10 BC ICB4	10%	11%	NA	NA	NA	NA
z10 EC ISC3	16%	17%	17%	21%	19%	24%
z10 EC 1x IFB	13%	14%	14%	17%	17%	19%
z10 EC 12x IFB	11%	12%	12%	14%	14%	16%
z10 EC ICB4	10%	10%	NA	NA	NA	NA
z114 ISC3	16%	18%	17%	21%	19%	24%
z114 1x IFB	13%	14%	14%	17%	17%	19%
z114 12x IFB	12%	13%	12%	15%	15%	17%
z114 12x IFB3	NA	NA	10%	12%	12%	13%
z196 ISC3	16%	17%	17%	21%	19%	24%
z196 1x IFB	13%	14%	13%	16%	16%	18%
z196 12x IFB	11%	12%	11%	14%	14%	15%
z196 12x IFB3	NA	NA	9%	11%	10%	12%
zBC12 ISC3	16%	17%	17%	21%	19%	24%
zBC12 1x IFB	14%	15%	14%	18%	17%	20%
zBC12 12x IFB	13%	13%	12%	15%	14%	17%
zBC12 12x IFB3	NA	NA	10%	11%	11%	12%
zEC12 ISC3	16%	17%	17%	21%	19%	24%
zEC12 1x IFB	13%	13%	13%	16%	16%	18%
zEC12 12x IFB	11%	11%	11%	13%	13%	15%
zEC12 12x IFB3	NA	NA	9%	10%	10%	11%

With z/OS V1.2 and above, synch-> asynch conversion caps values in the table at about 18%
IC links scale with the speed of the host technology and would provide an 8% effect in each case

System z – Maximum Coupling Links and CHPIDs

Server	1x IFB (HCA3-O LR)	12x IFB & 12x IFB3 (HCA3-O)	1x IFB (HCA2-O LR*)	12x IFB (HCA2-O*)	IC	ICB-4	ISC-3*	Max External Links	Max Coupling CHPIDs
zEC12	64 H20 – 32** H43 – 64**	32 H20 – 16** H43 – 32**	32* H20 – 16** H43 – 32**	32* H20 – 16** H43 – 32**	32	N/A	48*	104 (1)	128
zBC12	H13 – 32** H06 – 16**	H13 – 16** H06 – 8**	H13 – 12** H06 – 8**	H13 – 16** H06 – 8**	32	N/A	32 (4)	H13 (2) H06 (3)	128
z196	48 M15 – 32**	32 M15 – 16** M32 – 32**	32 M15 – 16** M32 – 32**	32 M15 – 16** M32 – 32**	32	N/A	48	104 (1)	128
z114	M10 – 32** M05 – 16**	M10 – 16** M05 – 8**	M10 – 12** M05 – 8**	M10 – 16** M05 – 8**	32	N/A	48	M10 (2) M05 (3)	128
z10 EC	N/A	N/A	32 E12 – 16*	32 E12 – 16**	32	16 (32/RPQ)	48	64	64
z10 BC	N/A	N/A	12	12	32	12	48	64	64

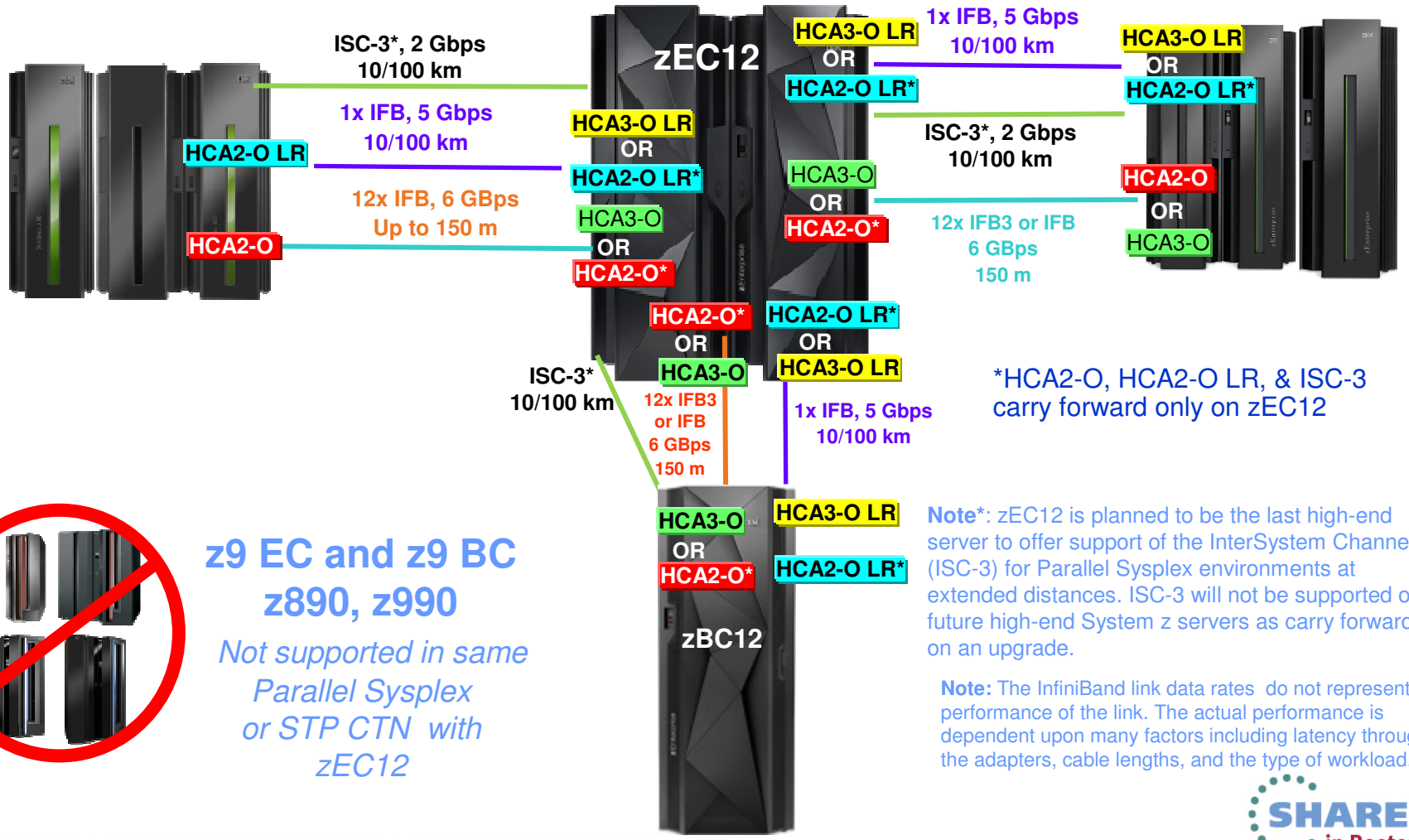
zEC12/zBC12 Parallel Sysplex Coupling Connectivity

z10 EC and z10 BC

12x IFB, 1x IFB & ISC-3

z196 and z114

12x IFB, 12x IFB3, 1x IFB, & ISC-3



z9 EC and z9 BC
z890, z990

Not supported in same Parallel Sysplex or STP CTN with zEC12

*HCA2-O, HCA2-O LR, & ISC-3 carry forward only on zEC12

Note*: zEC12 is planned to be the last high-end server to offer support of the InterSystem Channel-3 (ISC-3) for Parallel Sysplex environments at extended distances. ISC-3 will not be supported on future high-end System z servers as carry forward on an upgrade.

Note: The InfiniBand link data rates do not represent the performance of the link. The actual performance is dependent upon many factors including latency through the adapters, cable lengths, and the type of workload.

For More Information

- "IBM System z Connectivity Handbook" (SG24-5444)
- "Implementing and Managing InfiniBand Coupling Links on System z" *(*SG24-7539)
  - Available at www.redbooks.ibm.com
- www.ibm.com/systems/z/advantages/pso/whitepaper.html
  - CF Configuration Options White Paper

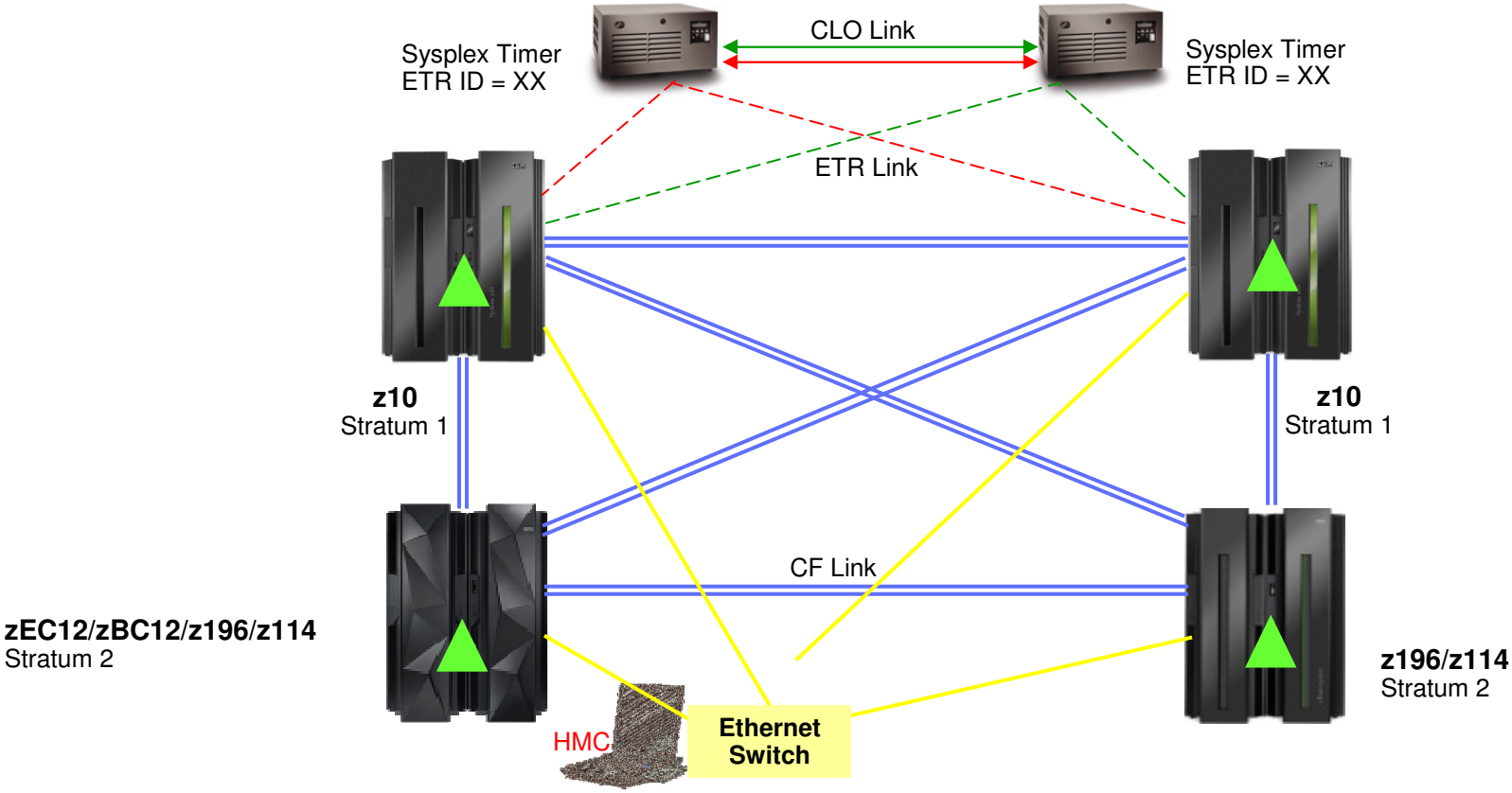
Agenda

- Hardware Updates
  - CFCC Level 19
  - CFCC Level 18
  - Parallel Sysplex Coupling Links
  - **Server Time Protocol (STP)**
- Software Updates
  - z/OS V2R1
  - z/OS V1R13
  - z/OS V1R12
- Summary

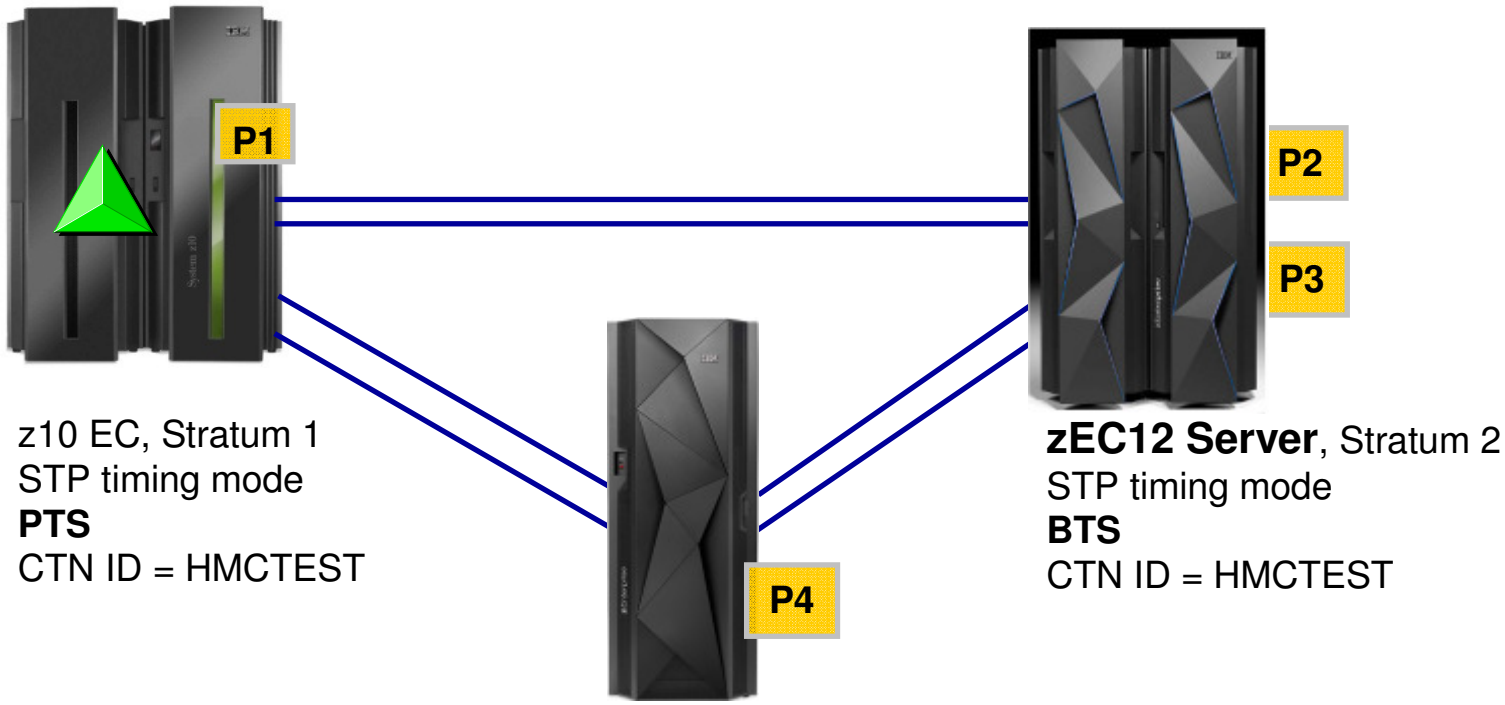
Glossary for System z Server Time Protocol (STP)

Acronym	Full name	Comments
Arbiter	Arbiter	Server assigned by the customer to provide additional means for the Backup Time Server to determine whether it should take over as the Current Time Server.
BTS	Backup Time Server	Server assigned by the customer to take over as the Current Time Server (stratum 1 server) because of a planned or unplanned reconfiguration.
CST	Coordinated Server Time	The Coordinated Server Time in a CTN represents the time for the CTN. CST is determined at each server in the CTN.
CTN	Coordinated Timing Network	A network that contains a collection of servers that are time synchronized to CST.
CTN ID	Coordinated Timing Network Identifier	Identifier that is used to indicate whether the server has been configured to be part of a CTN and, if so, identifies that CTN.
CTS	Current Time Server	A server that is currently the clock source for an STP-only CTN.
	Going Away Signal	A reliable unambiguous signal to indicate that the CPC is about to enter a check stopped state.
PTS	Preferred Time Server	The server assigned by the customer to be the preferred stratum 1 server in an STP-only CTN.

No Support for ETR with zEC12/zBC12/z196/z114 – Use Mixed CTN



STP-only CTN Example with System zEC12/zBC12 Servers



z10 EC, Stratum 1
STP timing mode
PTS
CTN ID = HMCTEST

zEC12 Server, Stratum 2
STP timing mode
BTS
CTN ID = HMCTEST

zBC12, Stratum 2
STP timing mode
Arbiter
CTN ID = HMCTEST

zEC12/zBC12
Server Time Protocol Enhancements

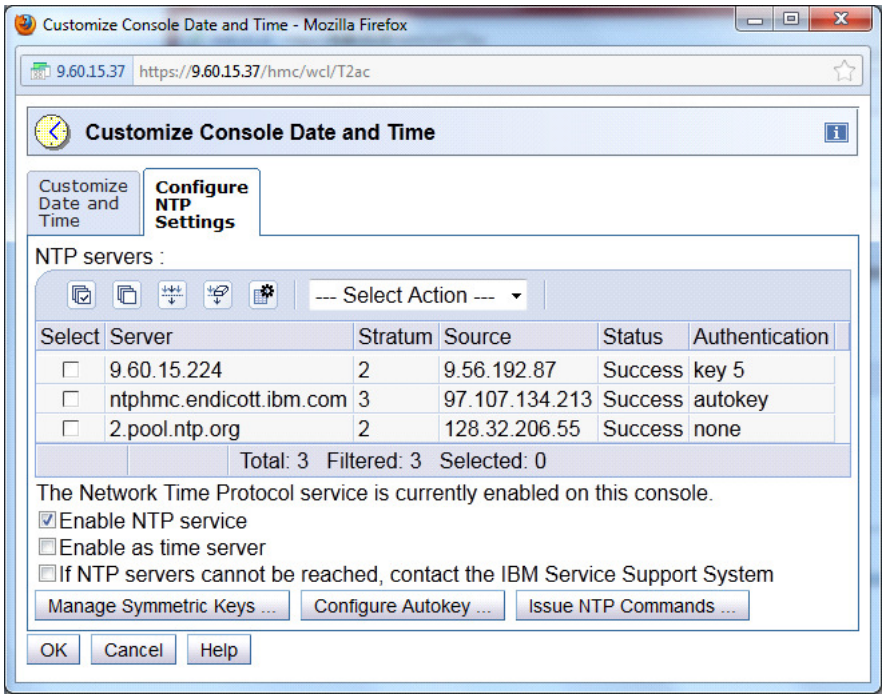
- Improved SE Time Accuracy
  - Optionally, the SE can be configured to connect to an external time source periodically to maintain highly accurate time that can be used, if required, to initialize CTN time during POR.
- Broadband Security Improvements for STP
  - Authenticates NTP servers when accessed by the HMC client through a firewall
  - Authenticates NTP clients when the HMC is acting as an NTP server
  - Provides symmetric key (NTP V3-V4) and autokey (NTP V4) authentication (Autokey is not supported if Network Address Translation is used)
  - This is the highest level of NTP security available

zEC12/zBC12
Server Time Protocol Enhancements

- Improved NTP Commands panel on HMC/SE
  - Shows command response details
- Telephone modem dial out to an STP time source is no longer supported
  - All STP dial functions are still supported by broadband connectivity
  - zEC12 HMC LIC no longer supports dial modems (Fulfills the Statement of Direction in Letter 111-167, dated October 12, 2011)

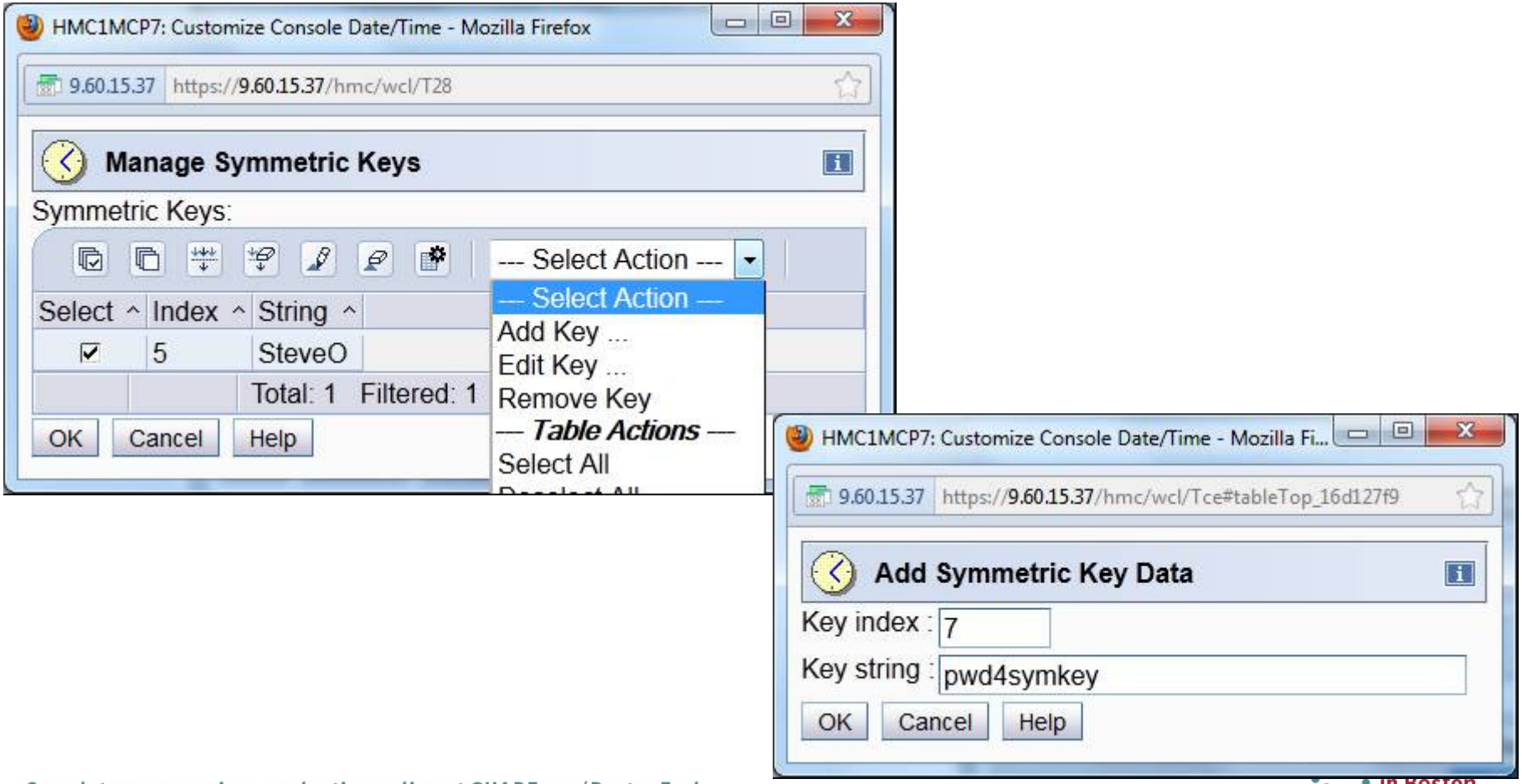
NTP Broadband Authentication Support for zEC12/zBC12

- Highest level of NTP security available
- Panels accept and generate key information to be configured into HMC NTP configuration.
- Autokey authentication not available with network address translating (NAT) firewall. Symmetric key still supports NAT.
- Autokey availability based on MCP level. First supported at zEC12 MCP level.

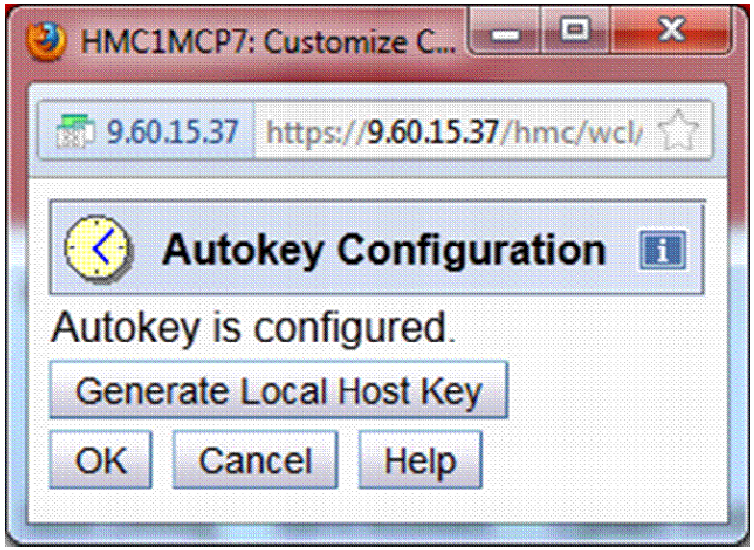


NTP Authentication - Symmetric Key

- Symmetric key encryption uses the same key for both encryption and decryption. Users exchanging data keep this key to themselves. Message encrypted with a secret key can be decrypted only with the same secret key.



NTP Authentication - AutoKey

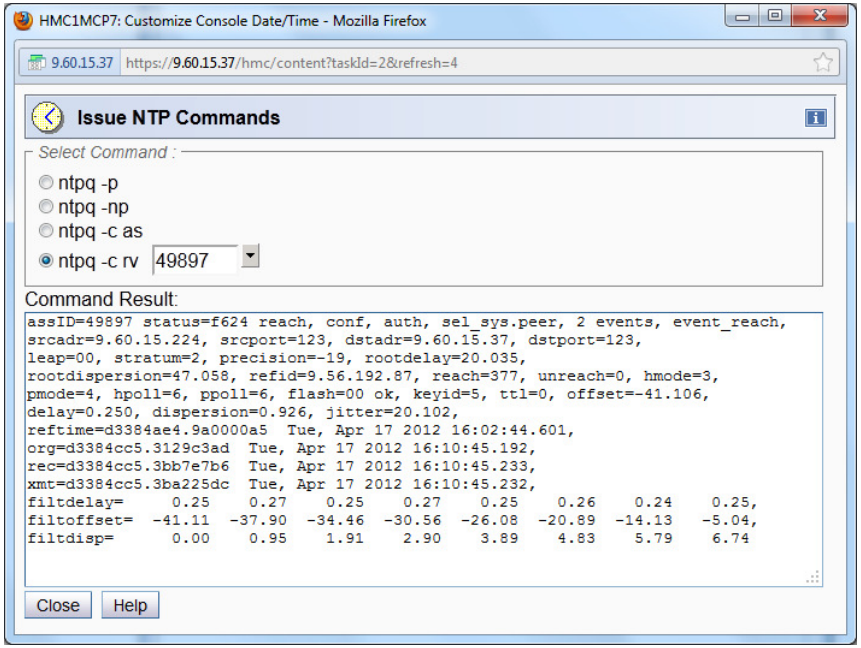


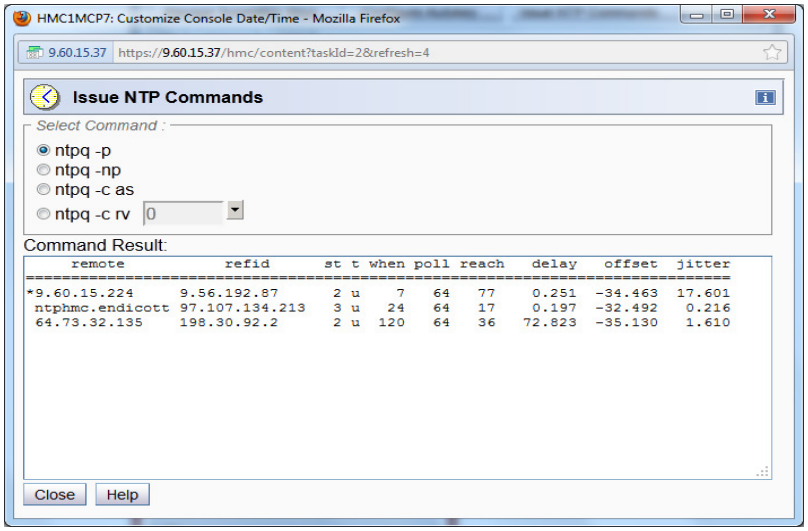
- An autokey cipher (also known as the autoclave cipher) is a cipher which incorporates the message (the plaintext) into the key
- http://www.eecis.udel.edu/~mills/ntp/html/autokey.html

NTP Authentication – Issue NTP Commands Panel

- One customer complaint in regard to the HMC NTP server panels, as they stand today, has been the status of the connection to target NTP servers.

- With the addition of NTP authentication, this display will aid in the determination of failures during configuration.





The explanation for the ntpq commands are all located on the following link: http://www.eecis.udel.edu/~mills/ntp/html/ntpq.html

STP References for Additional Information

- **Redbooks**
  - Server Time Protocol Planning Guide, SG24-7280
    - http://www.redbooks.**ibm.com**/redpieces/abstracts/sg247280.html
  - Server Time Protocol Implementation Guide, SG24-7281
    - http://www.redbooks.**ibm.com**/redpieces/abstracts/sg247281.html
- **TechDocs**
  - http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP102019 (new)
  - http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP102037 (new)
  - http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/TD105103 (updated)
  - http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP102081 (new)
  - http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/PRS2398 (updated)
- **Education**
  - Introduction to Server Time Protocol (STP)
    - Available on Resource Link at General Availability (GA)
    - www.**ibm.com**/servers/resourcelink/hom03010.nsf?OpenDatabase
- **STP Web site**
  - www.**ibm.com**/systems/z/pso/stp.html
- **Systems Assurance**
  - The IBM team is required to complete a Systems Assurance Review (SAPR Guide, SA06-012) and to complete the Systems Assurance Confirmation Form via Resource Link
  - http://w3.**ibm.com**/support/assure/assur30i.nsf/WebIndex/SA779
- **For further information on NTP and the NTP Public services project, refer to the Web sites:**
  - http://www.ntp.org
  - http://support.ntp.org
  - http://www.faqs.org/rfcs/rfc1305.html

Statement of Direction

(Previously announced)

- **The z196 and z114 would be the last System z servers to:**
  - Offer ordering of ESCON channels
  - Offer ordering of ISC-3
  - Support dial-up modem
- **And so the zEC12:**
  - Does not support ESCON
  - ISC-3 links available only via the I/O Carry Forward feature
    - *Will not be able to Carry Forward after zEC12*
  - HMC LIC no longer supports dial modems
- **Implications**
  - If using CTC devices for XCF signaling paths, need to migrate to FICON from ESCON or use Signaling Structures Only.
  - Migrate from ISC-3 coupling links to infiniband
  - Migrate to alternatives for dial-up time services

Statements of Direction

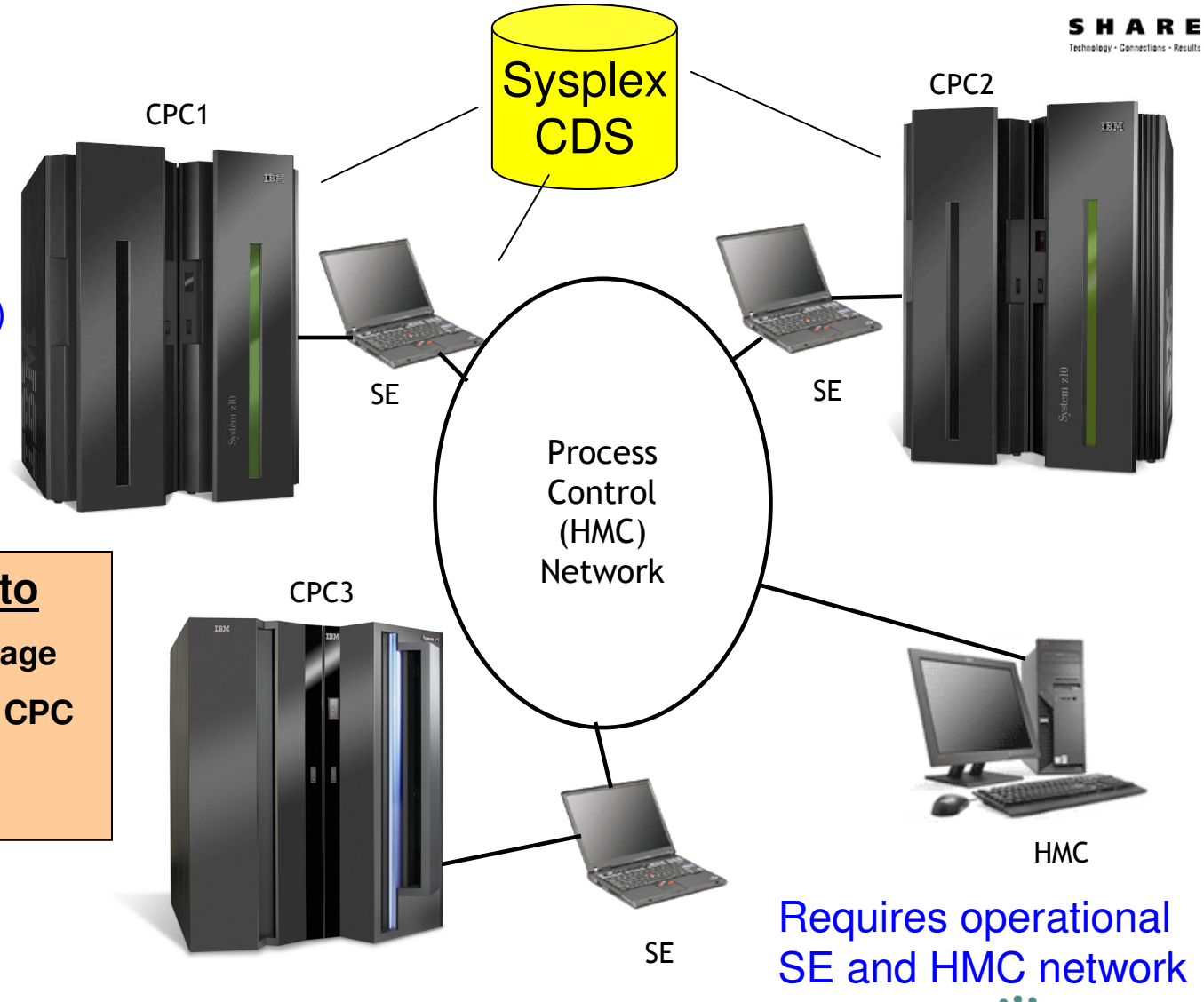
Future System z processors will NOT support:

- Connections to an STP Mixed CTN
  - Need to be migrating to STP-only timing network
- ISC-3 links
  - Need to migrate to HCA3-O LR 1x IFB Infiniband Coupling Links
- HCA2-O 12x or HCA2-O LR 1x IFB Fanouts
  - Need to migrate to HCA3-O 12x IFB and HCA3-O LR 1x IFB Infiniband Coupling Link Fanouts

Agenda

- Hardware Updates
  - CFCC Level 19
  - CFCC Level 18
  - Parallel Sysplex Coupling Links
  - Server Time Protocol (STP)
- Software Updates
  - z/OS V2R1
  - z/OS V1R13
  - z/OS V1R12
- Summary

z/OS V1R11 - SFM with BCPii



XCF uses BCPii to

- **Obtain identity of an image**
- **Query status of remote CPC and image**
- **Reset an image**

Requires operational SE and HMC network

z/OS V1R11 - SFM with BCPii

- Expedient removal of unresponsive or failed systems is essential to high availability in sysplex
- XCF exploits new BCPii services to:
  - Detect failed systems
  - Reset systems
- **Benefits**:
  - Improved availability by reducing duration of sympathy sickness
    - No waiting for FDI to expire
  - Eliminate manual intervention in more cases
    - Avoid IXC102A, IXC402D, IXC409D
  - Potentially prevent human error that can cause data corruption
    - Validate “down”

z/OS V1R11 - SFM with BCPii

- SFM will automatically exploit BCPii 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 or TRUSTED attribute
  - z10 GA2, or z196, z114, zEC12, zBC12
  - 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

Enabling SFM to use BCPii will have a big impact on availability. Make it happen !

z/OS V2R1 - Summary

• **Serial Rebuild**	Better MTTR
• **Thin Interrupts**	Better async service time
• **Sync/Async Thresholds**	Response time vs CPU
• **XCF Note Pads**	Simpler API for List Str
• D XCF STR,STATUS	New filters
• XCF Signal Throughput	100K/sec
• Couple Data Set Accessibility Verification	Avoid outages
• Cache Vector Corruption Detection	Avoid data corruption

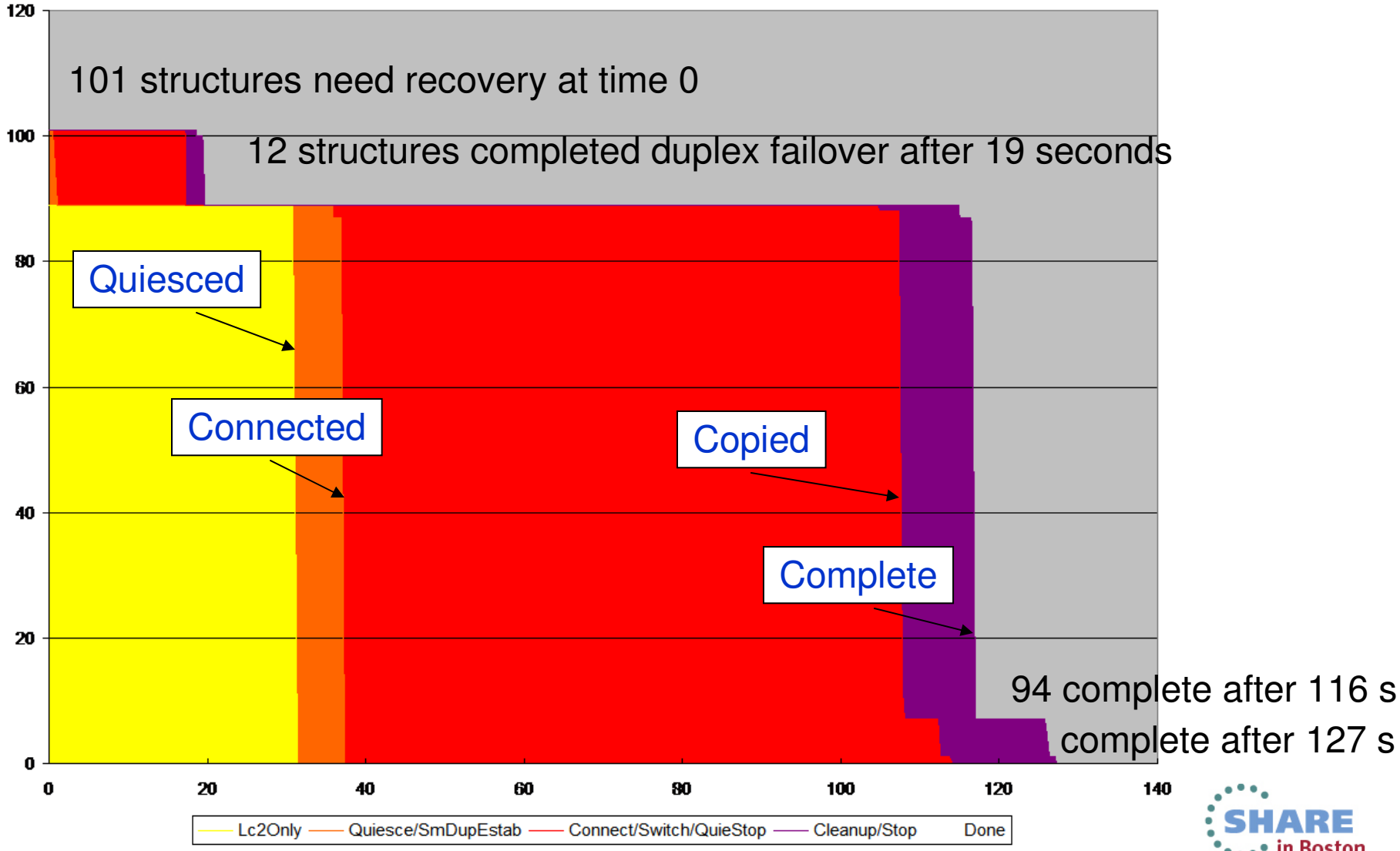
Due to time restrictions, only the topic in bold will be discussed.
Slides for the remaining topics are included in the handout.

z/OS V2R1 – Serial Rebuild

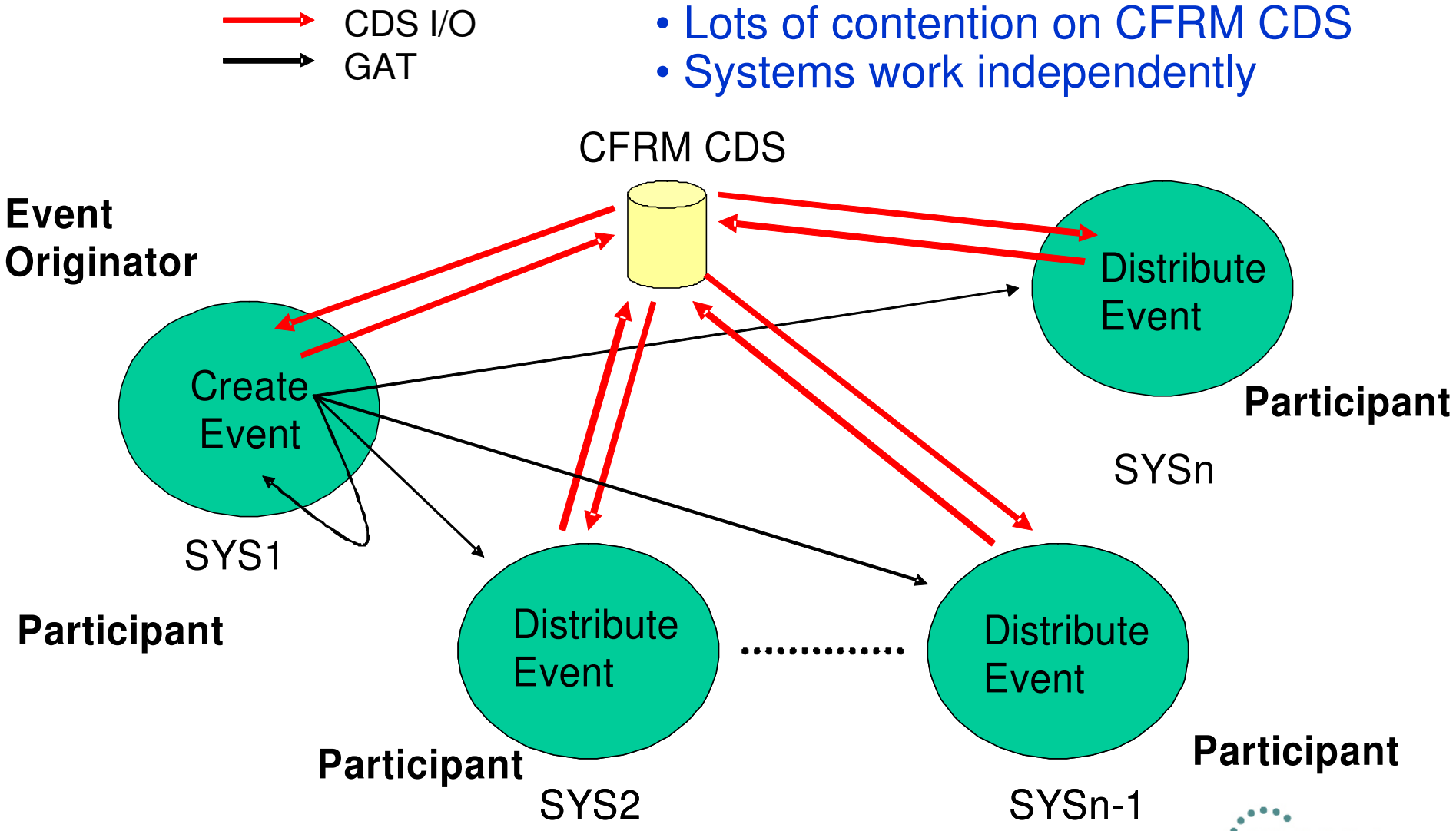
- When a system loses connectivity to a coupling facility the sysplex tries to recover the structures by:
  - Failing over to the accessible copy of a duplexed structure
  - Rebuilding the structure in some other coupling facility
- During the recovery, the structure is unavailable for use by the workload, so applications tend to hang for the duration
- A coupling facility could have dozens, perhaps hundreds of structures to recover
- Today, there is one massive burst of recovery processing launched in parallel for all of the impacted structures …

Parallel Rebuild - Test Results
Policy Based Event Processing

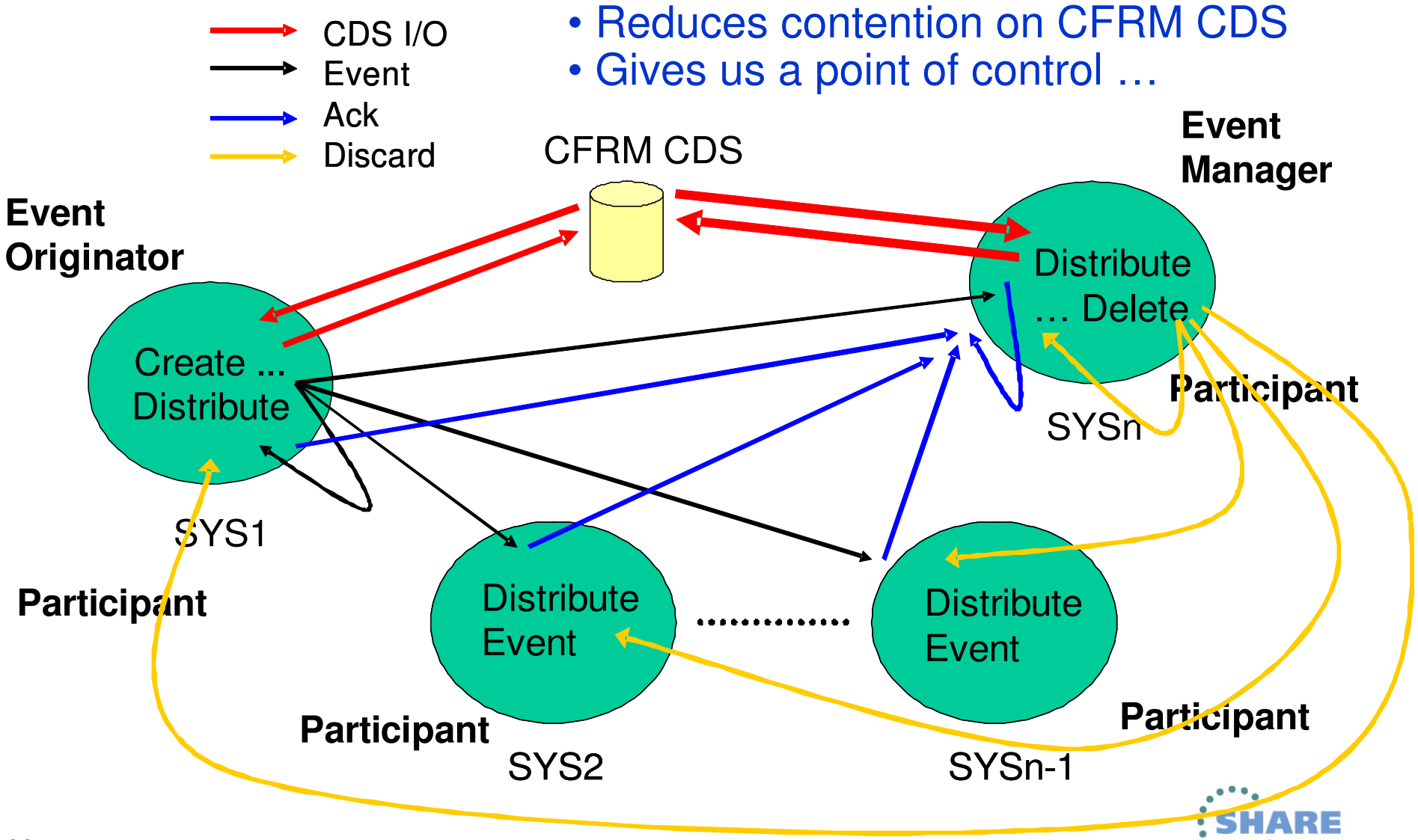
Not formal performance test measurements



Policy Based Event Processing



Message Based Event Processing

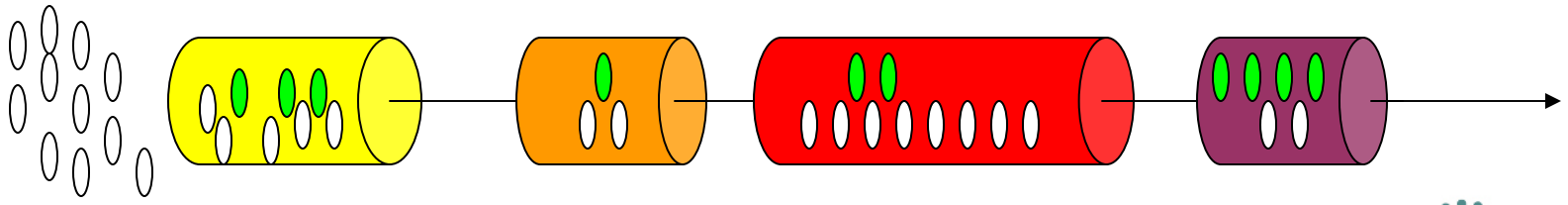


Serial Rebuild Improves Availability by Reducing MTTR

- Recovery is actually faster when done with less parallelism since there is less resource contention in several areas:
  - CFRM CDS
  - Coupling Facility
  - Participating systems
- Faster recovery means
  - Structures will be inaccessible for shorter periods, so
  - Applications are down for shorter periods
- Finishing recovery of the “important” structures sooner can reduce the business impact of the failure

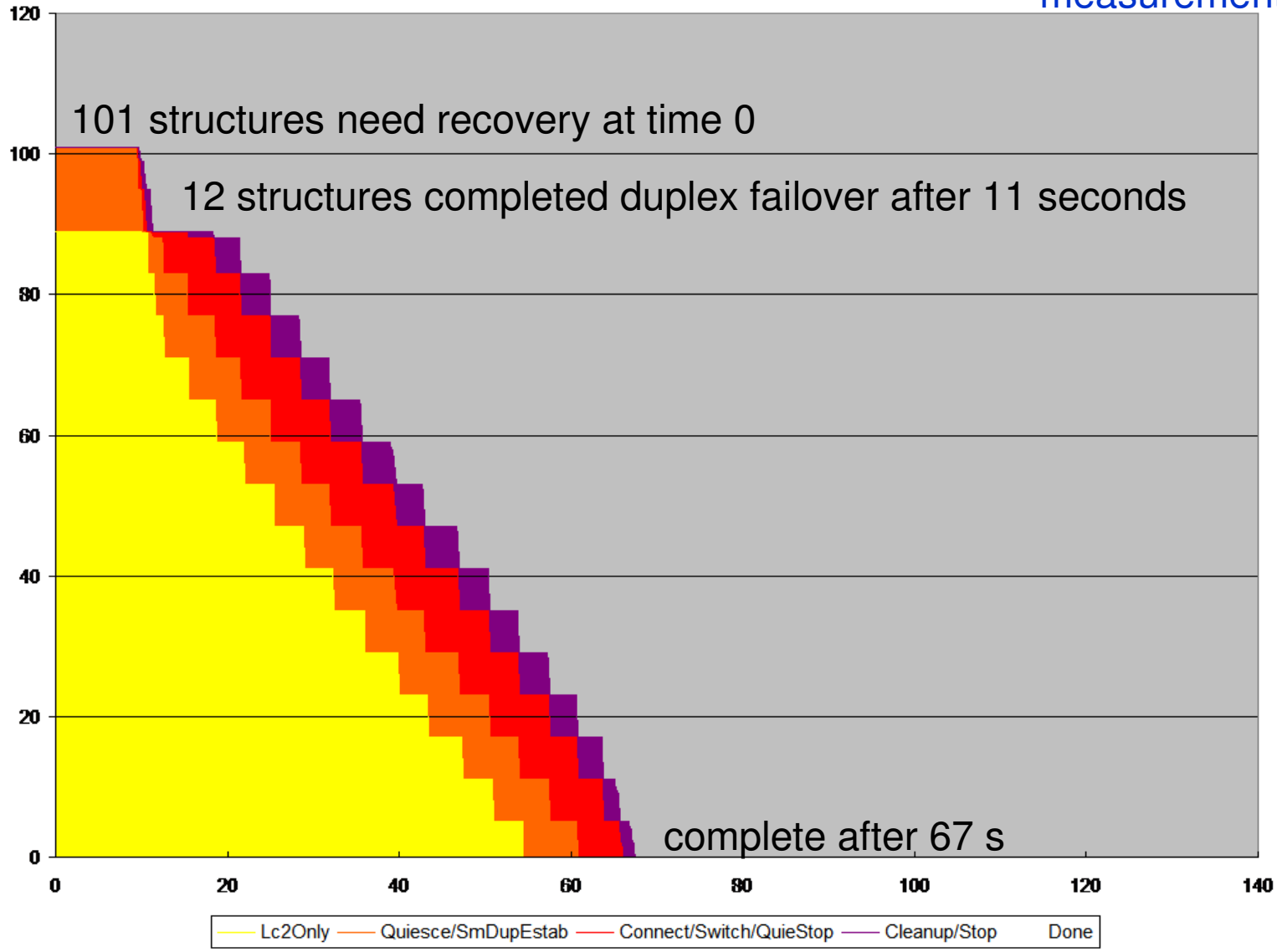
Serial Rebuild – New Behavior

- Issue message IXC568I “starting recovery”
- Sort structures according to recovery criteria
- Prime the pipe by initiating work to do rebuild or duplex failover for a batch of highest priority structures
- Do until done:
  - As work higher in the pipe completes a phase, push finite amount of completed work lower in the pipe ahead to the next phase
  - If no progress is being made in a phase, pull in finite amount of work from a lower phase
  - Use priority to determine what work to move to next phase
- Issue message IXC568I “finished recovery”



Serial Rebuild - Test Results

Not formal performance test measurements



Serial Rebuild – Influencing Priority of Rebuild

- You have some input as to what order structures will be selected for processing during LossConn Recovery
  - Presumably this would relate to the order in which you want your business applications to be restored to service
- Optional RECPRTY(value) specification in CFRM Policy
  - Decimal value in the range 1 to 4
  - Low numbers imply rebuild sooner, high numbers later
  - Takes effect immediately when policy activated
- Default order: “System” structures, locks, the rest.

Serial Rebuild – Concerns About Rebuild Order

- There may be unknown dependencies or relationships between the structures
  - The rebuild of one structure might not be able to complete until the rebuild of a second structure has completed if …
  - The rebuild process for the first structure calls a service that needs access to the second structure
- What if the rebuild priorities are reversed?
- Serial Rebuild will not deadlock
  - Pulls in more work if not enough progress being made
  - So all structures will eventually complete
  - But “eventually” might be longer than necessary

Serial Rebuild – Exploitation

- All systems in the sysplex must be z/OS V2R1
- Primary CFRM CDS must be formatted for MSGBASED
- The new XCF CFLCRMGMT switch must be ENABLED on all systems in the sysplex
  - COUPLExx PARMLIB member at IPL, or
  - Dynamically using SETXCF
- Consider use of RECPRTY keyword in CFRM policy

Another Undesirable Old Behavior

- Scenario:
  - CF is reset. Looks like LossConn to all the systems.
  - Duplex structures fail over and are available in simplex mode
  - Remaining structures in massive wave of rebuilds
  - CF reboots and connectivity is restored
  - CFRM immediately tries to re-duplex the structures
  - Re-duplexing effort bogged down in massive wave of rebuilds
- Well this is rather annoying
  - Duplex structures quickly restored to service after initial failure
  - And now unavailable for duration of the non-duplex rebuilds

Sysplex-wide CFRM Processing is Now Prioritized

- Most important
  - LossConn Recovery (disconnect, rebuild, or duplex failover)
  - REALLOCATE/POPULATECF structure evaluation/action
  - Policy-initiated STOP CF structure duplexing for policy change
  - Policy-initiated START CF structure duplexing for DUPLEX(ENABLED)
- Least important

- Work of lesser importance is deferred if there is more important work …

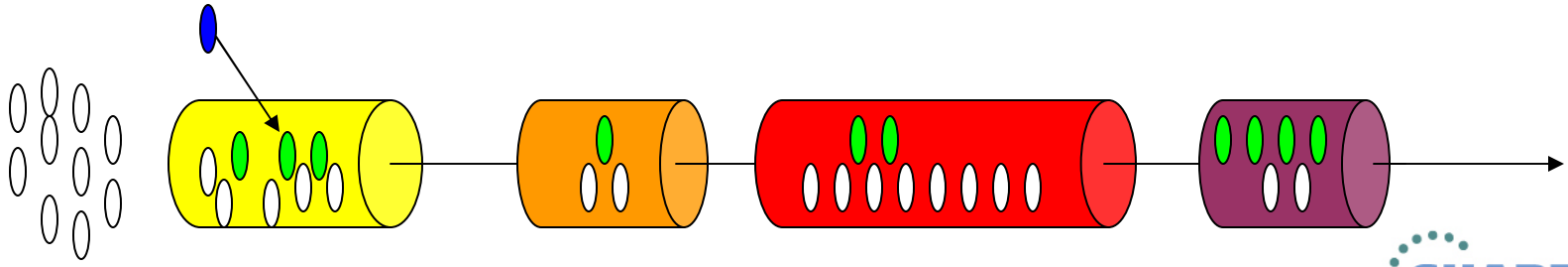
New Behavior – Serial Duplexing

- Re-duplexing effort is deferred until after the LossConn recovery is completed
  - The duplex structures are quickly restored to service as they fail over to simplex mode as part of the LossConn recovery
  - Let the other structures complete their recovery before launching a new duplexing effort
- When launched, CFRM will re-duplex the structures:
  - Sequentially, one at a time
  - In a system determined order
    - Recovery priority is not used

What About Other Rebuild Requests

- During LossConn Recovery, the rebuild processing is being carefully managed
- An externally initiated rebuild request could arrive
  - SETXCF
  - Application initiated

- The new rebuild request is immediately initiated, but is otherwise managed along with all the others

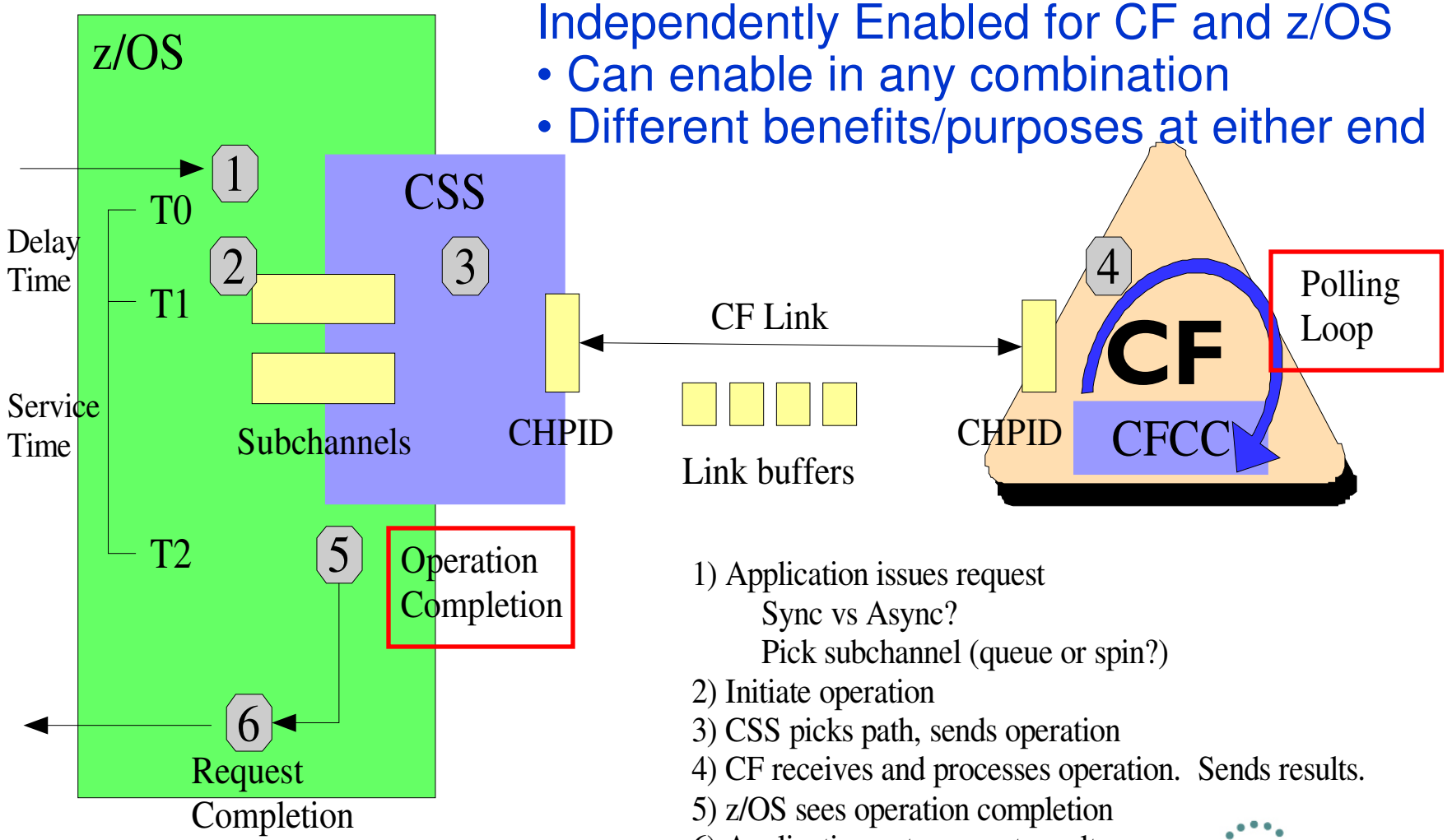


z/OS V2R1 - Thin Interrupts

- For z/OS, improves service time for asynchronous requests by reducing latency of completion processing

- So relevant workload should have:
  - Shorter elapsed time
  - More throughput

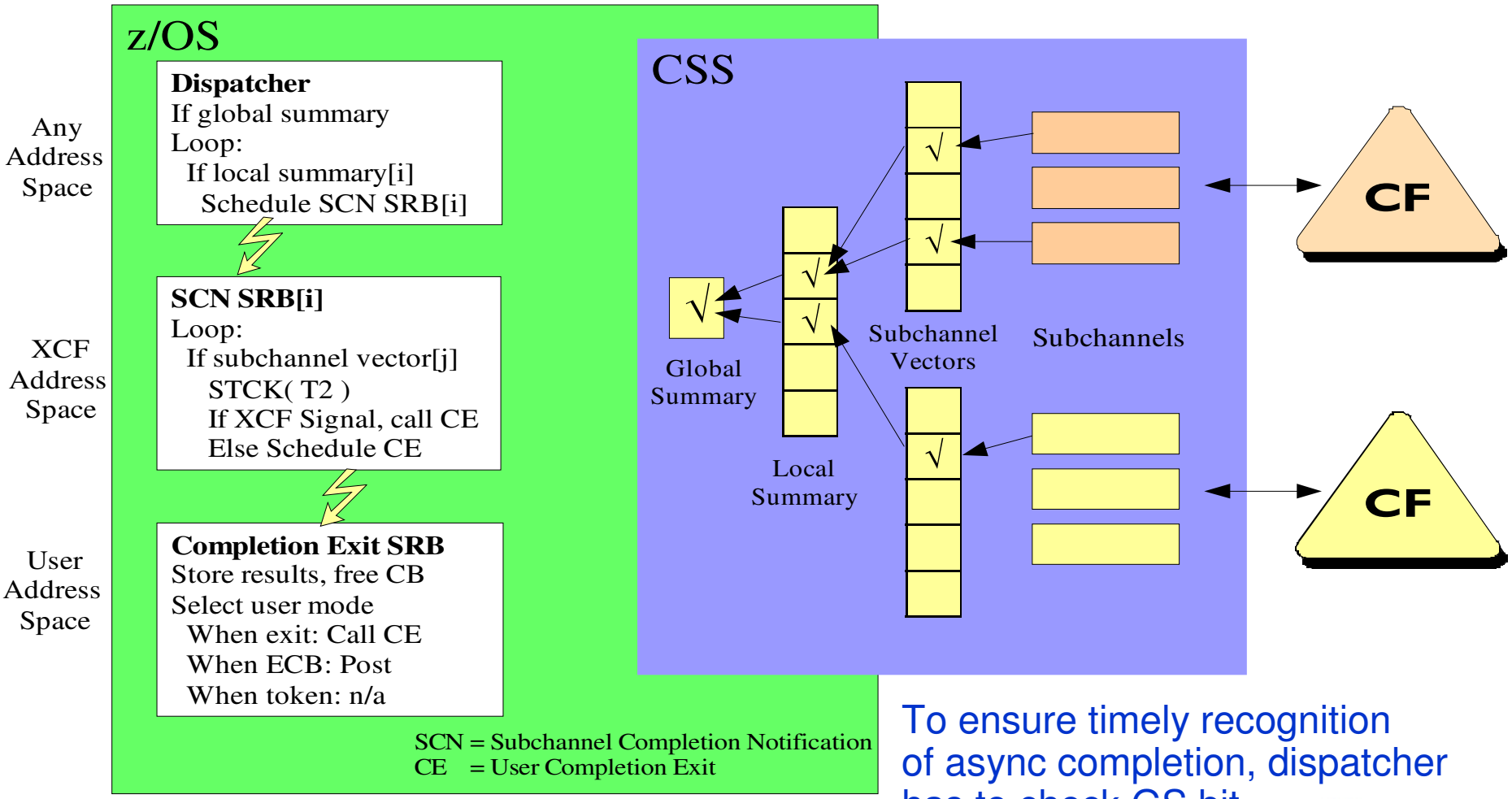
z/OS V2R1 – Thin Interrupts



Independently Enabled for CF and z/OS
- Can enable in any combination
- Different benefits/purposes at either end

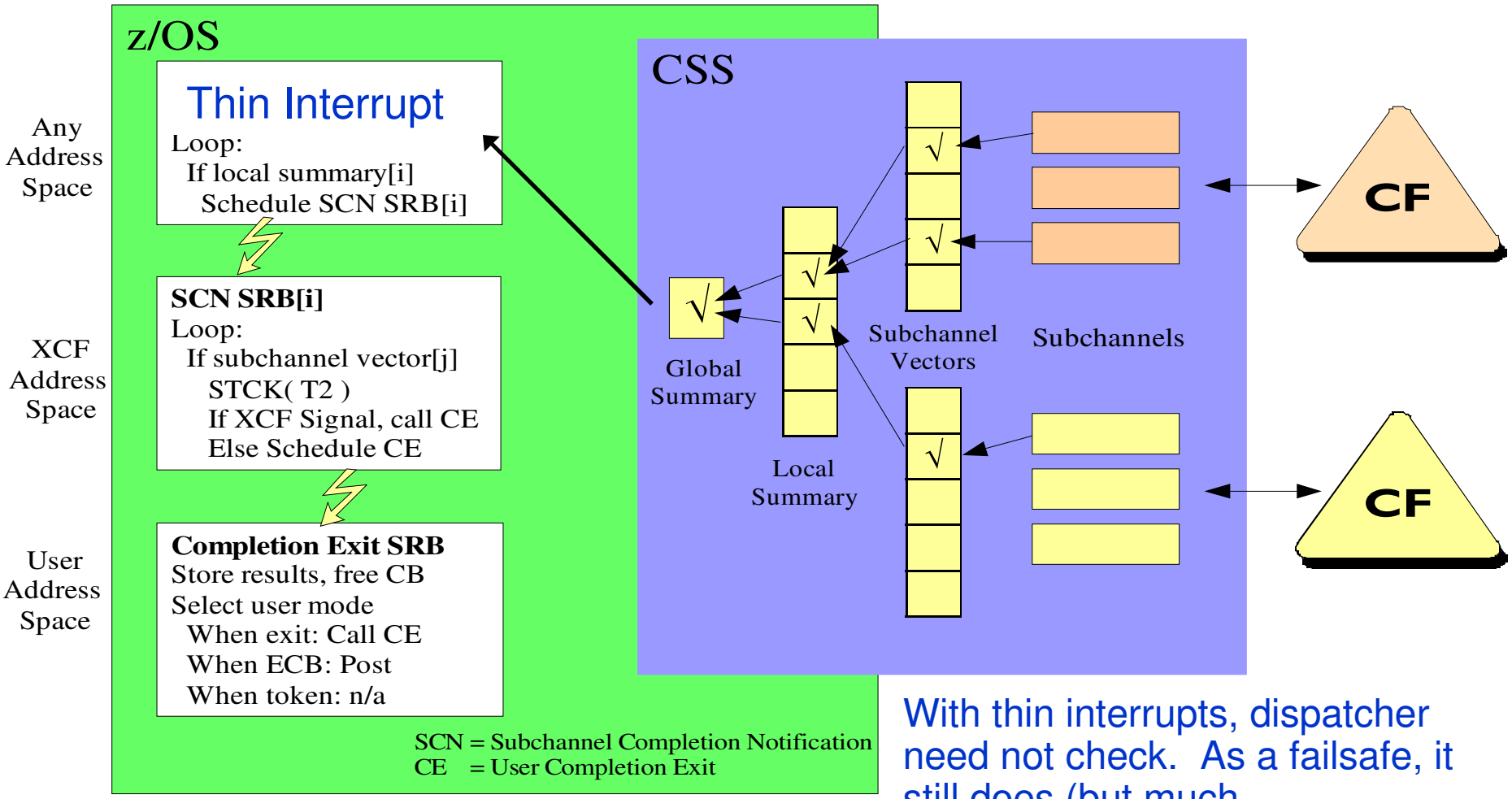
1) Application issues request
   - Sync vs Async?
   - Pick subchannel (queue or spin?)
2) Initiate operation
3) CSS picks path, sends operation
4) CF receives and processes operation. Sends results.
5) z/OS sees operation completion
6) Application gets request results

Asynchronous Operation Completion - Today



To ensure timely recognition of async completion, dispatcher has to check GS bit frequently

Asynchronous Operation Completion - With Thin Interrupt



With thin interrupts, dispatcher need not check. As a failsafe, it still does (but much less frequently).

Thin Interrupts - Configuration

- Software Requirements (for z/OS exploitation)
  - z/OS V2R1, or
  - z/OS V1R12 and V1R13 with the following service installed:
    - APAR OA38734 (XES)
    - APAR OA37186 (Supervisor)
    - APAR OA38781 (IOS)
- Hardware Requirements
  - zEC12 GA2 or BC12 for z/OS coupling thin interrupts
  - CFCC CF Level 19 for CFCC coupling thin interrupts

Thin Interrupts – Switch

- XCF FUNCTIONS switch to enable or disable use of thin interrupts on the z/OS side
  - Does not change the CF behavior at all
  - Default is for COUPLINGTHININT to be ENABLED
- If enabled (and hardware supports it)
  - CSS is told to drive thin interrupts when asynchronous operation completes
  - Dispatcher checks global summary bit occasionally
- If disabled
  - CSS is told to not generate thin interrupts (if hardware supports it)
  - Dispatcher checks global summary bit frequently

Thin Interrupts – Messages

- D XCF,C
  - Reports COUPLETHININT switch setting
- D XCF,CF
  - CF DYNDISP setting – add “thin interrupts”
  - Are thin interrupts supported and/or enabled on the CF CEC
- New Messages
  - IXL163I – XES could not enable/disable thin interrupts
  - IXL164I – enabled thin interrupts
- Health Check updated
  - Should have thin interrupts if using shared CPs
  - Configuration data includes thin interrupt information

z/OS V2R1 – Sync/Async Thresholds

- XES Heuristic Algorithm
  - Measures synchronous service times for each kind of CF operation
  - Determines whether the opportunity cost of running the operation synchronously exceeds the overhead of running the request asynchronously
  - If so, the request will be processed asynchronously
- Helps optimize use of CPU resources so as to maximize the amount of work performed
- At the expense of increasing the elapsed time of the request

Some Prefer a Different Tradeoff

- Some customers would rather sacrifice CPU efficiency in order to maintain shorter service times
  - The longer service times impact their business objectives
  - They can tolerate getting less total work done
- To accommodate this need, you can now set the conversion thresholds used by the heuristic algorithm to tailor the tradeoff between CPU cost and service time.
  - Also available on z/OS V1R13 and V1R12 via APAR OA41661

Setting Conversion Threshold

- COUPLExx parmlib member
  - On new SYNCASYNC statement specify: keyword(value)
- SETXCF MODIFY,SYNCASYNC,keyword=value
- “keyword” is one of the following thresholds
  - SIMPLEX - for simplex list and cache requests
  - DUPLEX - for duplexed list and cache requests
  - LOCKSIMPLEX - for simplex lock requests
  - LOCKDUPLEX - for duplexed lock requests
- “value” can be:
  - Numeric value in range 1 to 10000 (microseconds)
  - DEFAULT – to use the system determined threshold value

Some Cautions

- The default threshold value is dynamically computed to account for factors that would affect the opportunity cost
  - Speed of the processor
  - Number of CPs
- When you set the threshold
  - It is a fixed value
  - Never adjusted for dynamic changes that might occur

- Playing with these knobs could significantly impact your workload. Be careful.

What Are My Threshold Values?

- D XCF,COUPLE

```
SYNC/ASYNC CONVERSION    THRESHOLD    -SOURCE-    DEFAULT
          SIMPLEX              350     SETXCF          413
           DUPLEX              457     SYSTEM       IN USE
     LOCK SIMPLEX              413     SYSTEM       IN USE
      LOCK DUPLEX              551     SYSTEM       IN USE
```

Which threshold

Current value being used

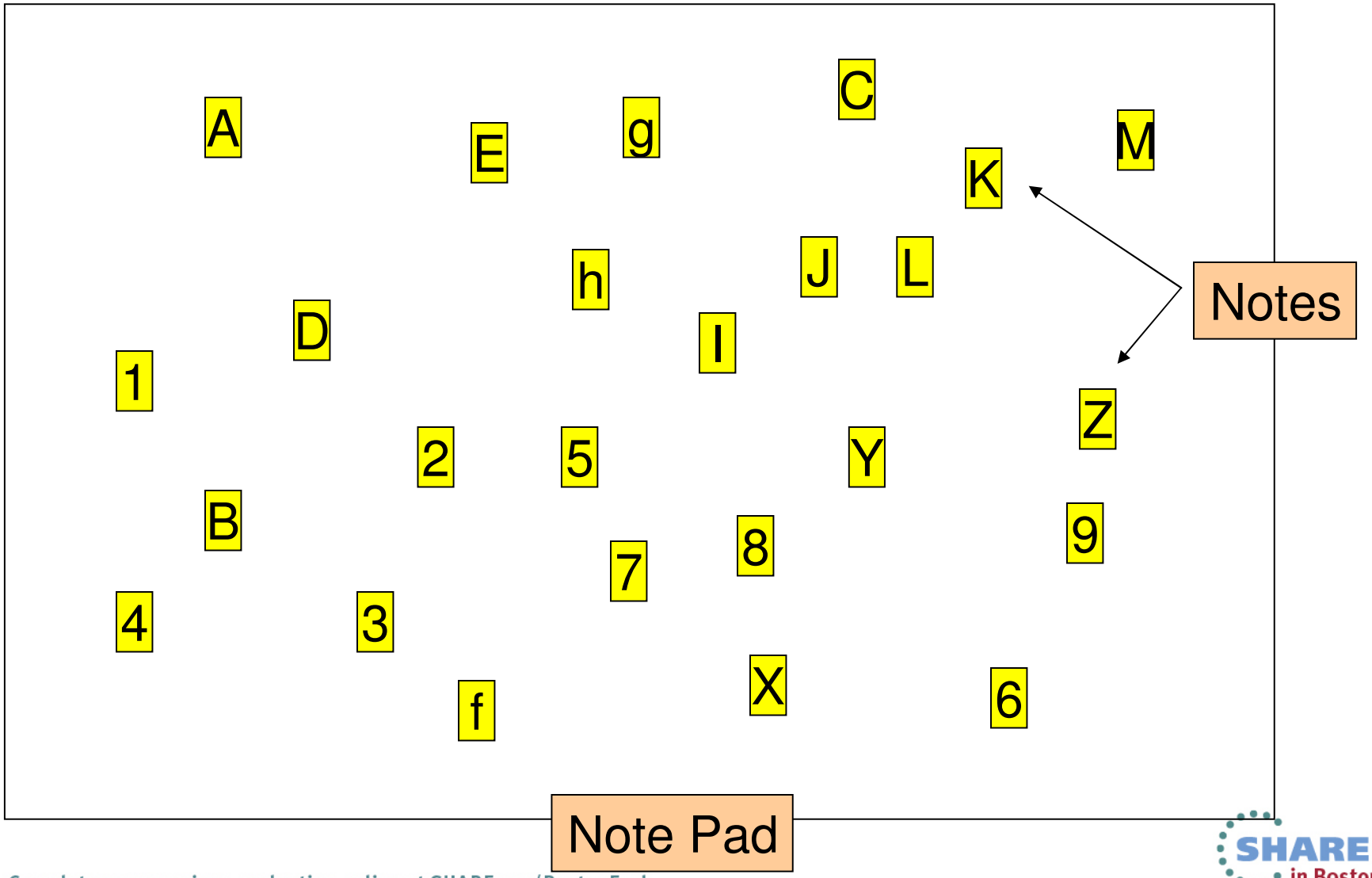
How was it set?

Current system default if not already in use

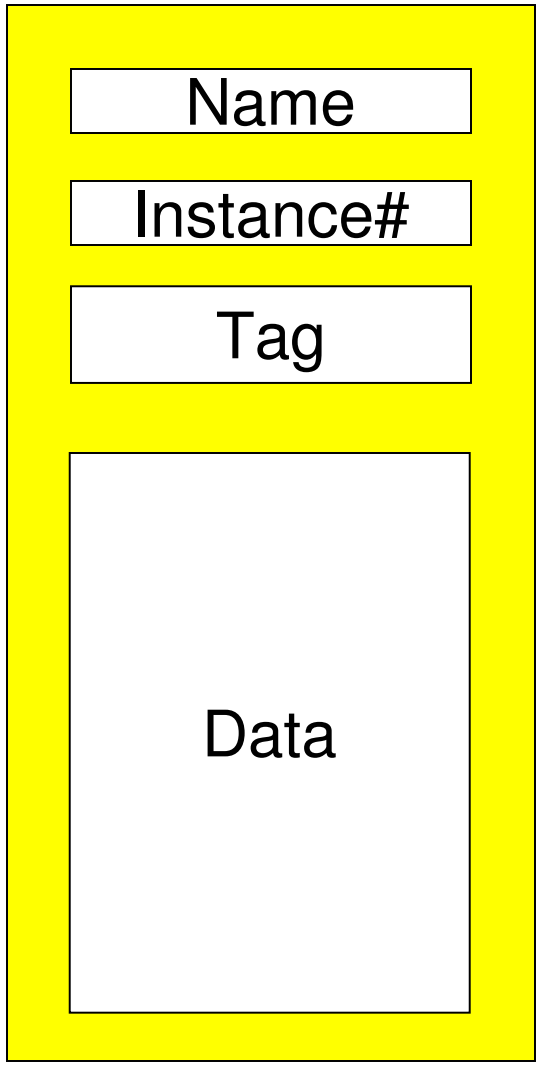
z/OS V2R1 – XCF Note Pad Service (IXCNOTE macro)

- Programs can read and write notes (list entries) in an XCF note pad (CF list structure)
  - Supports unauthorized callers
  - One or more note pads can reside in the same list structure
  - Each note pad can contain finite number of 1K notes
- Useful for applications that can exploit the “note pad” model
  - High performance access to (state) data from any system
  - Not useful for message passing or work flow
  - Does not expose full functionality of list structure
- XCF connects to CF structure and deals with various XES exits and protocols
- Simplifies development, reduces complexity, decreases implementation and support costs by masking most of the traditional CF exploitation overhead

Note Pad



Note in a Note Pad



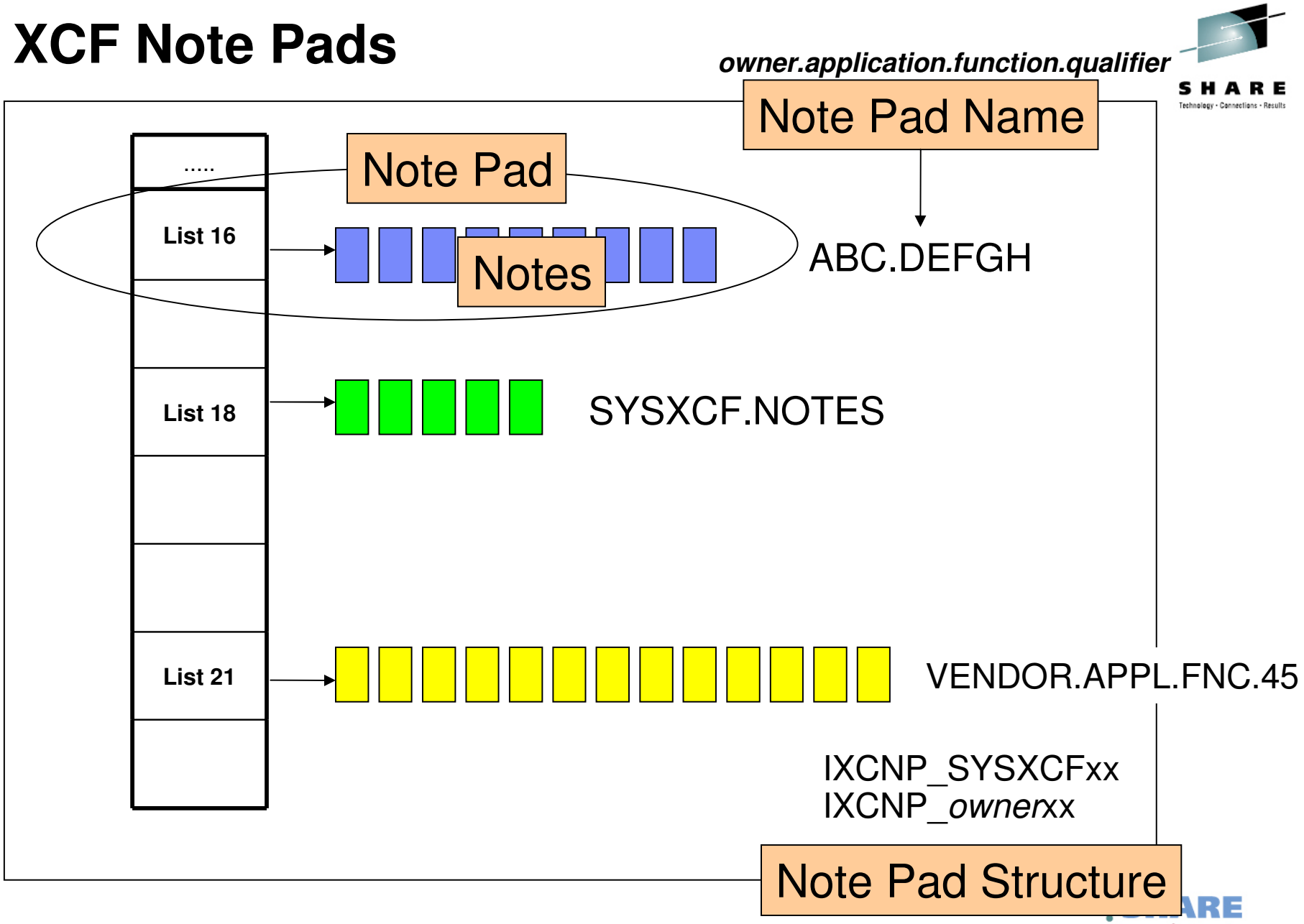
8 byte user note name

8 byte XCF instance number

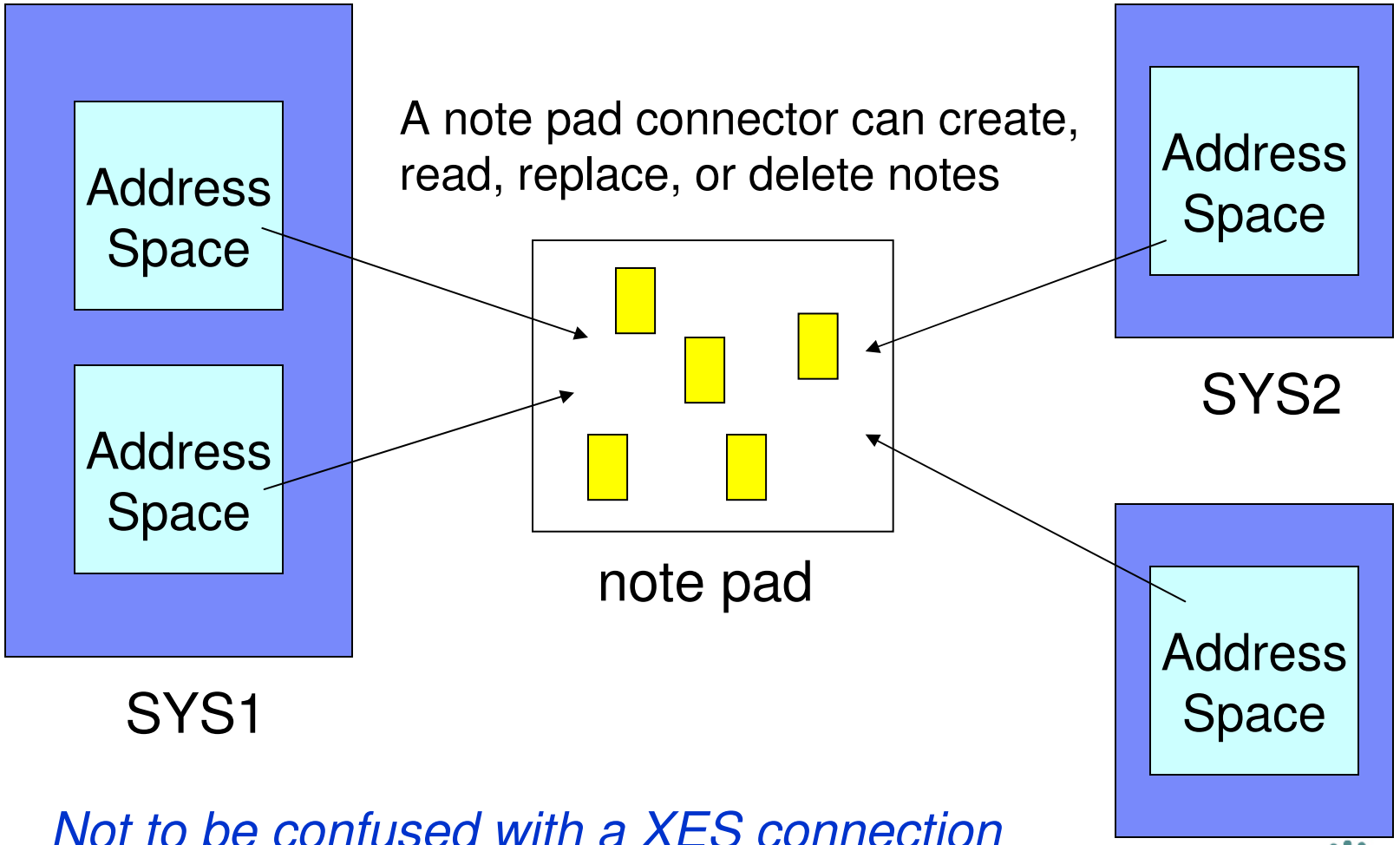
16 bytes of user metadata

1024 bytes of user data
(or none)

XCF Note Pads

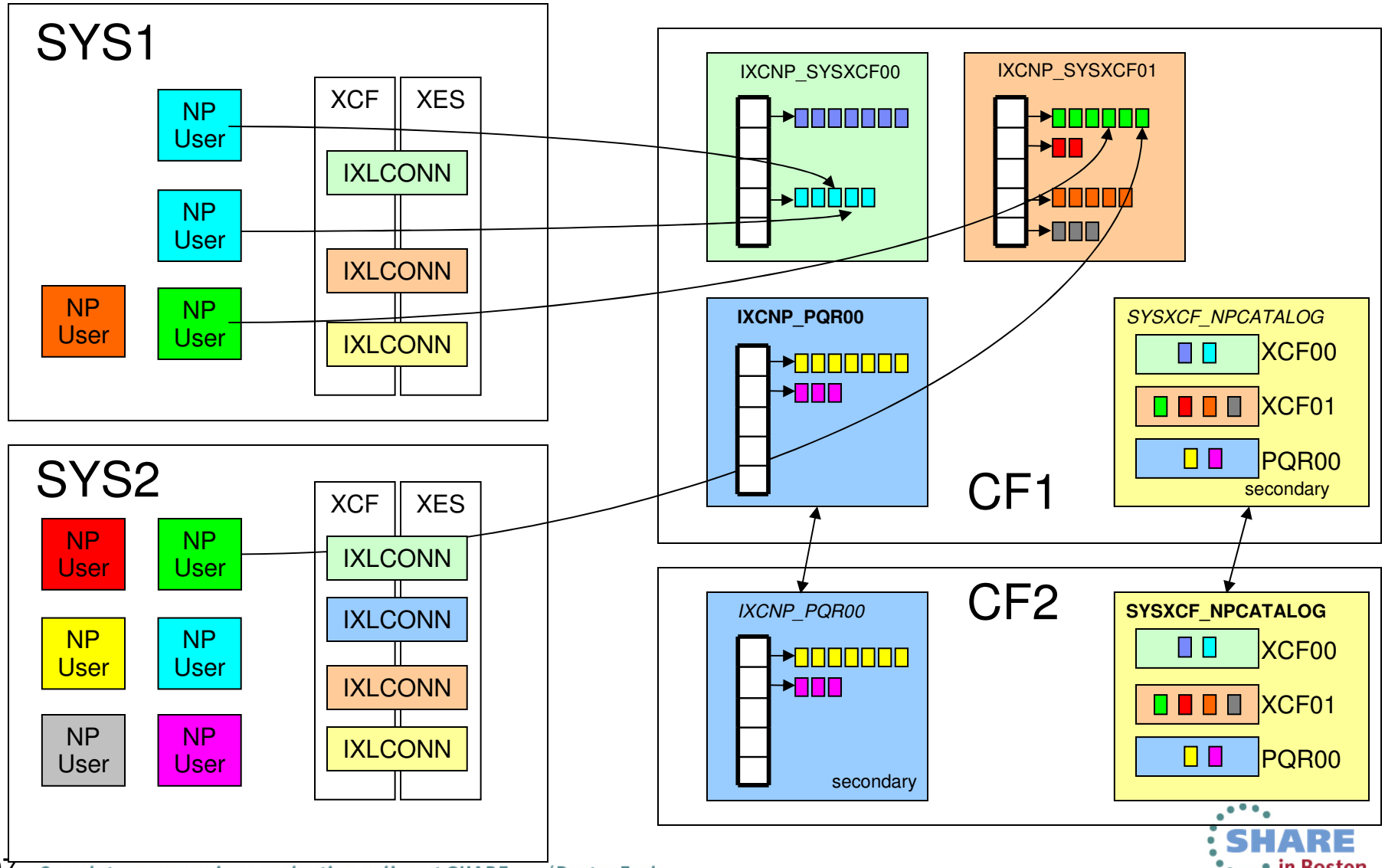


Connections to a Note Pad



Not to be confused with a XES connection to the note pad structure

XCF Note Pads in the Sysplex



XCF Note Pad
System Programmer Perspective

- **Requirements**
  - z/OS V2R1, or
  - z/OS 1.13 with OA38450
  - CFLEVEL 9 or later
- **Note Pad Catalog**
  - Size
  - Duplex
- **Note Pad Structure(s)**
  - Names
  - Size
  - Simplex or duplex ?
- **Security**
  - Note pads
  - Structures
- **Management**
  - D XCF,NP
  - Messages
  - Delete Utility
  - Delete Structures
  - Measurement
- Diagnostics
  - XCF CTRACE options

D XCF,STR – New Status Filters

```
D XCF,STR,STAT=?
IXC352I DISPLAY XCF SYNTAX ERROR, COULD NOT RECOGNIZE:
?.  ONE OF THE FOLLOWING WAS EXPECTED:
  ( ALLOCATED NOTALLOCATED REBUILD
  STRDUMP DEALLOCPENDING POLICYCHANGE LARGERCFRMDS
  FPCONN NOCONN ALTER INCLEANUP
  DUPREBUILD DUPMISMATCH LOSSCONN RBPROC
  RBPEND DUPENAB DUPALLOW
```

- DUPMISMATCH
  - Allocated but DUPLEXED state does not match policy – start or stop duplexing pending
- LOSSCONN
  - A connector has lost connectivity to the structure
- RBPROC
  - Structure in rebuild processing (other than duplex established)
- RBPEND
  - POPCF or REALLOCATE evaluation pending
- DUPENAB/DUPALLOW
  - Structure with policy DUPLEX specification of ENABLED or ALLOWED, respectively

D XCF,STR – Use of New Filters

- Did z/OS duplex all my DUPLEX(ENABLED) structures?
  - D XCF,STR,STAT=DUPENAB
  - If not, maybe delayed for more important work, rebuild processing, or stop duplex
    - D XCF,STR,STAT=(DUPMISMATCH,RBPROC,RBPEND)

      CF LOSSCONN RECOVERY MANAGEMENT IS IN PROGRESS.

      THE REALLOCATE PROCESS IS IN PROGRESS.

      POPULATECF REBUILD PENDING

      REBUILD IN PROGRESS

- Did z/OS resolve all duplexing mismatches?
  - D XCF,STR,STAT=DUPMISMATCH
  - If not, maybe delayed for more important work or rebuild processing
    - D XCF,STR,STAT=(RBPROC,RBPEND)

      CF LOSSCONN RECOVERY MANAGEMENT IS IN PROGRESS.

      THE REALLOCATE PROCESS IS IN PROGRESS.

      POPULATECF REBUILD PENDING

      REBUILD IN PROGRESS

D XCF,STR – Use of New Filters …

- Did REALLOCATE (or POPCF) complete?
  - D XCF,STR,STAT=RBPEND

    THE REALLOCATE PROCESS IS IN PROGRESS.

    POPULATECF REBUILD PENDING

    POPULATECF REBUILD IN PROGRESS
  - If not, maybe delayed for more important work

    CF LOSSCONN RECOVERY MANAGEMENT IS IN PROGRESS.
- Did CF LOSSCONN RECOVERY complete?
  - D XCF,STR,STAT=LOSSCONN,STRNM=*

    CF LOSSCONN RECOVERY MANAGEMENT IS IN PROGRESS.

XCF Signal Throughput Improvement

- XCF manages message exit SRBs to provide for:
  - Responsive message delivery without …
  - Starving tasks in the target member address space
    - Many members drop messages off to tasks for processing
- To do so, XCF controls
  - Number of SRBs
  - Number of signals an SRB can process
  - Frequency with which SRBs are scheduled
- We have adjusted these controls to permit delivery of signals in the neighborhood of 100,000 signals/second
  - Roughly 4X improvement over prior releases
  - Assuming the target member can keep up

Couple Data Set Accessibility Verification

Problem

- A loss of DASD power at one of two sysplex sites can cause loss of all couple data sets (CDSes) and a sysplex-wide outage.

Solution

- XCF processing to remove a CDS will now attempt to verify the accessibility of the remaining CDS. If XCF can determine that both CDSes of a given type have been lost simultaneously, it will refrain from sending the signals that can trigger the sysplex outage.
- So when a subset of sysplex systems have lost both CDSes of a given type, only those systems will be required to remove both CDSes from service. The remaining systems may lose only one CDS or neither.

Also available with APAR OA38311 at z/OS V1R12 and V1R13

Cache Vector Corruption Detection

- Had an issue in the field which caused buffer invalidation (XI) signals to be missed by DB2 for a Group Buffer Pool
- Unable to determine source of problem
  - DB2? XES? Links? CFCC?
- So we added support in the CF to:
  - Detect possible occurrences
  - Gather timely diagnostic data
  - Fail in a way that avoids data corruption
- And it seems to have worked ….

Cache Vector Corruption Detection …

- XES is a guilty party
  - Timing window where cleanup of vector used to manage XI signals was not properly handled
- APAR OA42519

Agenda

- Hardware Updates
  - CFCC Level 19
  - CFCC Level 18
  - Parallel Sysplex Coupling Links
  - Server Time Protocol (STP)
- Software Updates
  - z/OS V2R1
  - z/OS V1R13
  - z/OS V1R12
- Summary

z/OS V1R13 - Summary

- D XCF,SYSPLEX – Revised output
- CF Structure Placement – more explanation
- ARM – New timeout for application cleanup

- New XCF API for sending signals – IXCSRVR,IXCSEND,IXCRECV
- SETXCF MODIFY - Disable structure alter processing
- SDSF – Sysplex wide data gather without MQ
- Runtime Diagnostics – Detects more contention
- Zfs – Sysplex wide direct access to shared files

Due to time restrictions, I can only summarize.
Slides with more information on each topic are included in the Appendix

z/OS V1R12 Summary

- REALLOCATE – TEST and REPORT options
- Critical Members – handle sympathy sickness
- CFSTRHANGTIME – handle sympathy sickness
- Support for CFLEVEL 17 - Larger, more structures
- Health Checks - Related to XCF, CF
- Auto Reply - Automate during IPL
- XCF API's – 64 bit msgs; join parms

Due to time restrictions, I can only summarize.
Slides with more information on each topic are included in the Appendix

Agenda

- Hardware Updates
  - CFCC Level 19
  - CFCC Level 18
  - Parallel Sysplex Coupling Links
  - Server Time Protocol (STP)
- Software Updates
  - z/OS V2R1
  - z/OS V1R13
  - z/OS V1R12
- Summary

Highlights

- CFLEVEL 19 for zEC12 GA2, zBC12
- CFLEVEL 18
  - Serviceability and performance improvements
- STP
  - Enhanced security
- Thin Interrupts:
  - On CF for configuration flexibility
  - On z/OS for better async CF service time
- Serial Rebuild for better MTTR

z/OS Publications

- *MVS Setting Up a Sysplex*
- *MVS Initialization and Tuning*
- *MVS Systems Commands*
- *MVS Diagnosis: Tools and Service Aids*
- *z/OS V2R1 Migration*
- *z/OS V2R1 Planning for Installation*
- *z/OS MVS Programming: Callable Services for High Level Languages*
  - Documents BCPii Setup and Installation and BCPii APIs

Sysplex-related Redbooks

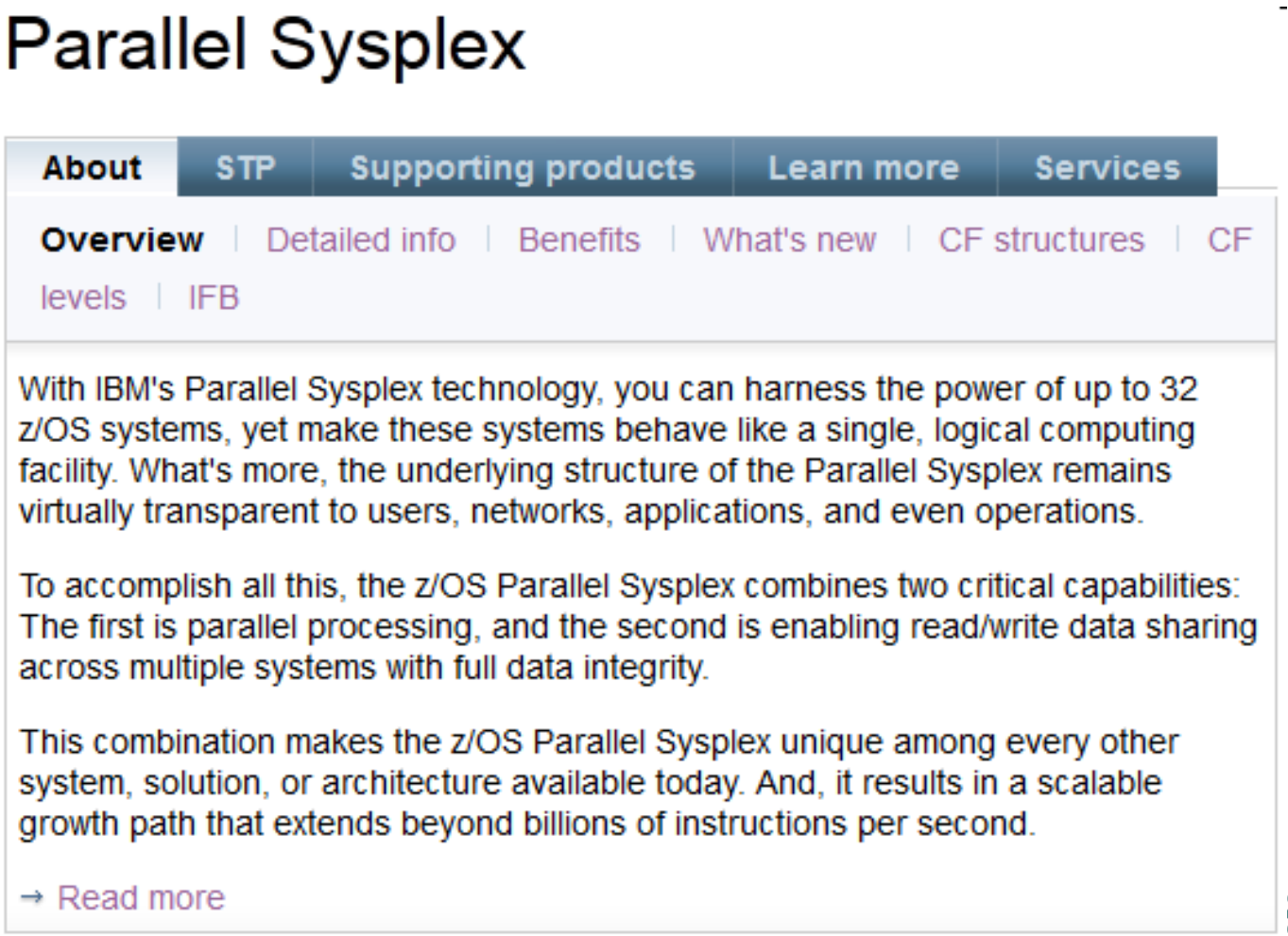
- System z Parallel Sysplex Best Practices, SG24-7817
- Considerations for Multi-Site Sysplex Data Sharing, SG24-7263
- Server Time Protocol Planning Guide, SG24-7280
- Server Time Protocol Implementation Guide, SG24-7281
- Server Time Protocol Recovery Guide, SG24-7380

- Exploiting the IBM Health Checker for z/OS Infrastructure, REDP-4590

- Available at www.redbooks.ibm.com

Parallel Sysplex Web Site

http://www.ibm.com/systems/z/advantages/pso/index.html



Appendix

Sysplex Related Highlights From
z/OS V1R13

Appendix – z/OS V1R13

- D XCF,SYSPLEX – Revised output
- CF Structure Placement – more explanation
- ARM – New timeout parameter for application cleanup

- New XCF Client/Server API for sending signals
- SETXCF MODIFY - Disable structure alter processing
- SDSF – Sysplex wide data gathering without MQ
- Runtime Diagnostics – Detects more contention
- zFS – Direct access to shared files throughout sysplex

z/OS V1R13 - DISPLAY XCF,SYSPLEX

- D XCF,SYSPLEX command is a popular command used to display the systems in the sysplex
- But, prior to z/OS V1R13:
  - Output not as helpful for problem diagnosis as it could be
  - Much useful system and sysplex status information is kept by XCF, but not externalized in one central place
- So z/OS V1R13 enhances the output
  - You can still get the same output (perhaps with new msg #)
  - And you can get more details than before

z/OS V1R13 – D XCF,SYSPLEX,ALL

	z/OS 1.12
D XCF,S,ALL	IXC335I 12:55:00 DISPLAY XCF FRAME LAST F E SYS=SY1 SYSPLEX PLEX1 SYSTEM TYPE SERIAL LPAR STATUS TIME SYSTEM STATUS SY1 4381 9F30 N/A 04/22/2011 12:55:00 ACTIVE TM=SIMETR SY2 4381 9F30 N/A 04/22/2011 12:54:56 ACTIVE TM=SIMETR SY3 4381 9F30 N/A 04/22/2011 12:54:56 ACTIVE TM=SIMETR SYSTEM STATUS DETECTION PARTITIONING PROTOCOL CONNECTION EXCEPTIONS: SYSPLEX COUPLE DATA SET NOT FORMATTED FOR THE SSD PROTOCOL
	z/OS 1.13
D XCF,S,ALL	IXC337I 12.29.36 DISPLAY XCF FRAME LAST F E SYS=SY1 SYSPLEX PLEX1 MODE: MULTISYSTEM-CAPABLE SYSTEM SY1 STATUS: ACTIVE TIMING: SIMETR NETID: 0F STATUS TIME: 05/04/2011 12:29:36.000218 JOIN TIME: 05/04/2011 10:31:08.072275 SYSTEM NUMBER: 01000001 SYSTEM IDENTIFIER: AC257038 01000001 SYSTEM TYPE: 4381 SERIAL: 9F30 LPAR: N/A NODE DESCRIPTOR: SIMDEV.IBM.PK.D13ID31 PARTITION: 00 CPCID: 00 RELEASE: z/OS 01.13.00 SYSTEM STATUS DETECTION PARTITIONING PROTOCOL CONNECTION EXCEPTIONS: SYSPLEX COUPLE DATA SET NOT FORMATTED FOR THE SSD PROTOCOL

z/OS V1R13 – CF Structure Placement

- Why did it put my structure in that CF ?
  - A dark art, often a mystery to the observer
- Existing messages updated to help explain
  - IXL015I: Initial/rebuild structure allocation
    - Also has “CONNECTIVITY=” insert
  - IXC347I: Reallocate/Reallocate test results
  - IXC574I: Reallocate processing, system managed rebuild processing, or duplexing feasibility

z/OS V1R13 – CF Structure Placement …

```
IXL015I STRUCTURE ALLOCATION INFORMATION FOR
STRUCTURE THRLST01, CONNECTOR NAME THRLST0101000001,
CONNECTIVITY=SYSPLEX
 CFNAME        ALLOCATION STATUS/FAILURE REASON
 --------      ----------------------------------------
 LF01          ALLOCATION NOT PERMITTED
               COUPLING FACILITY IS IN MAINTENANCE MODE
 A             STRUCTURE ALLOCATED CC007B00
 TESTCF        PREFERRED CF ALREADY SELECTED CC007B00
               PREFERRED CF HIGHER IN PREFLIST
 LF02          PREFERRED CF ALREADY SELECTED CC007300
               EXCLLIST REQUIREMENT FULLY MET BY PREFERRED CF
 SUPERSES      NO CONNECTIVITY 98007800
```

Automatic Restart Management (ARM)

- If you have an active ARM policy, then:
  - After system failure, ARM waits up to two minutes for survivors to finish cleanup processing for the failed system
  - If cleanup does not complete within two minutes, ARM proceeds to restart the failed work anyway
- Problem: Restart may fail if cleanup did not complete
- Issue: Two minutes may not be long enough for the applications to finish their cleanup processing

z/OS V1R13 – New ARM Parameter

- CLEANUP_TIMEOUT
  - New parameter for the ARM policy specifies how long ARM should wait for survivors to cleanup for a failed system
  - Specified in seconds, 120..86400 (2 min to 24 hours)
- If parameter not specified
  - Defaults to 300 seconds (5 minutes, not 2)
  - Code 120 if you want to preserve old behavior
- If greater than 120:
  - Issues message IXC815I after two minutes to indicate that restart is being delayed
  - If the timeout expires, issues message IXC815I to indicate restart processing is continuing despite incomplete cleanup
- Available for z/OS V1R10 and up with APAR OA35357

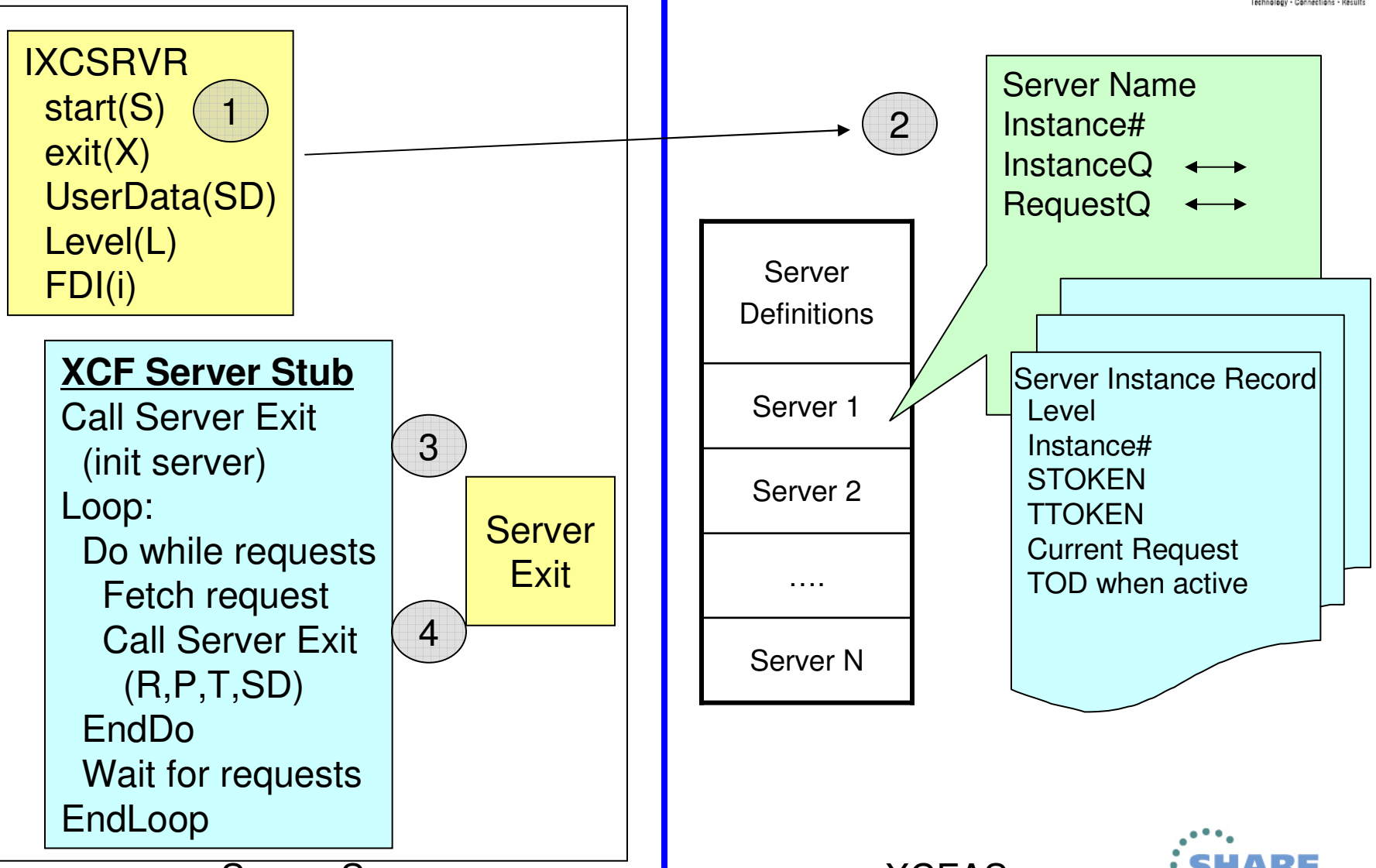
z/OS V1R13 – New XCF API for Message Passing (XCF Client/Server)

- Allows authorized programs to send and receive signals within a sysplex **without** joining an XCF Group
- XCF does communication and failure handling
- Simplifies development, reduces complexity, implementation and support costs by eliminating some of the XCF exploitation costs
- Messages delivered to a **task** instead of an SRB
  - **Server** is collection of tasks identified by **server name**
  - Server exit routine (instead of message exit routine)
  - Various server selection criteria (routing options)
- Response processing
  - Occurs under thread of application's choosing
  - Blocking or non-blocking

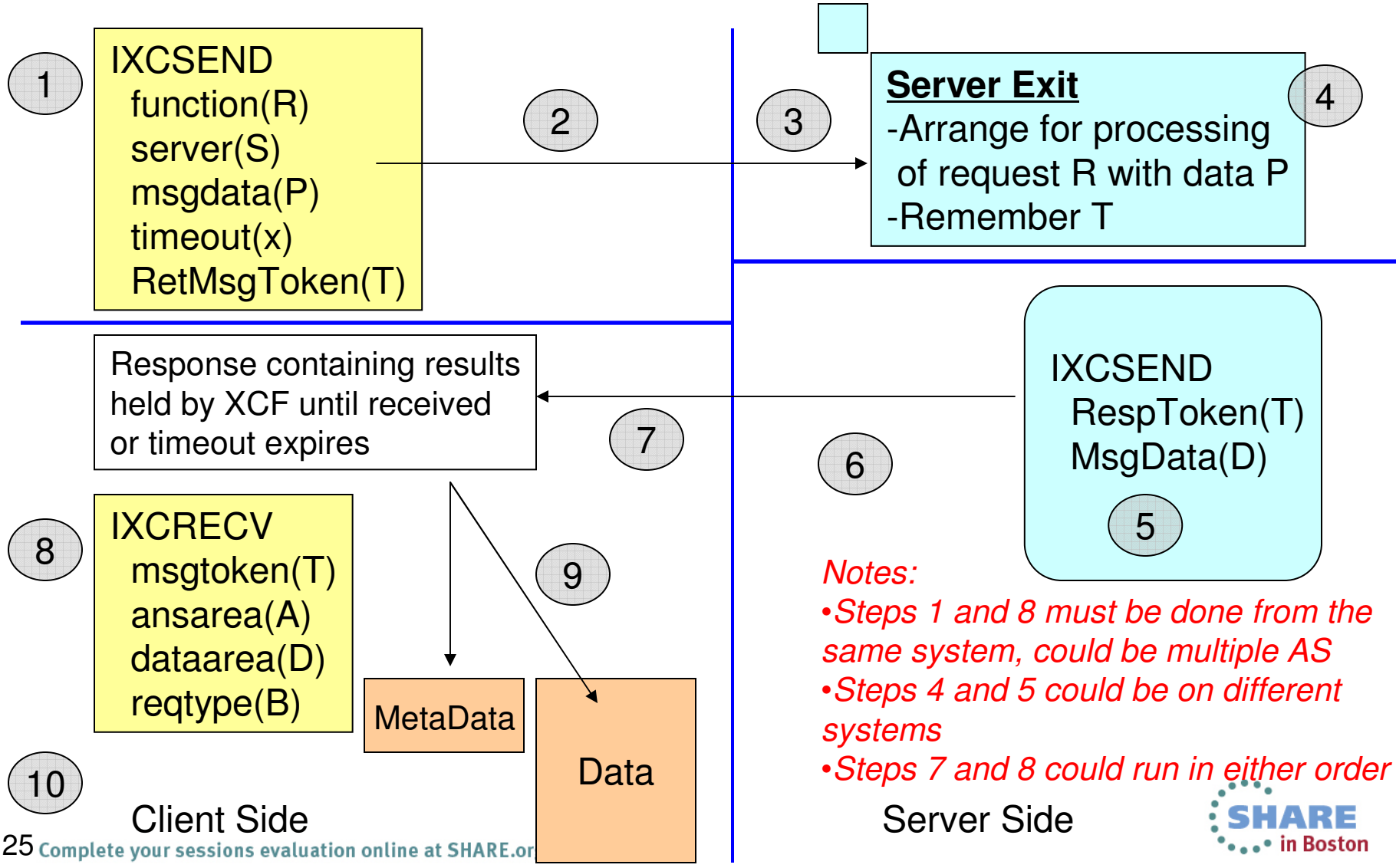
z/OS V1R13 XCF Client/Server

- IXCSEND – send request to one or more servers
- IXCSRVR – start or stop a server instance
- IXCSEND – send response to client request
- IXCRECV – receive response(s) from server(s)
- IXCYSRVR – data mappings

XCF Client/Server – Server Task Overview



XCF Client/Server - Send/Receive Overview



DISPLAY XCF,SERVER

- The DISPLAY XCF command was extended to display information about servers, server instances, and queued work

```
D XCF, { SERVER | SRV }
            [ ,{SYSNAME | SYSNM}={sysname | (sysname [,sysname]. . .) }  ]
            [ ,{SERVERNAME | SRVNAME | SRVNM}={servername}  ]
            [ ,SCOPE={ {SUMMARY | SUM} | {DETAIL | DET} } ]
            [, TYPE=NAME [, STATUS=(STALLED)] |
                        {INSTANCE | INST}
                                [, STATUS=( [{WORKING | WORK}] [, STALLED] ) ]
                                [, {INSTNUM | INST#}=inst# ] ]
```

CF Structure Alter Processing

- CF Structure Alter processing is used to dynamically reconfigure storage in the CF and its structures to meet the needs of the exploiting applications
  - Size of structures can be changed
  - Objects within structures can be reapportioned
- Alter processing can be initiated by the system, the application, or the operator
- There have been occasional instances, either due to extreme duress or error, where alter processing has contributed to performance problems
- Want an easy way to inhibit alter processing ….

z/OS V1R13 – Enable/Disable Start Alter Processing

- SETXCF MODIFY,STRNAME=pattern,ALTER=DISABLED
- SETXCF MODIFY,STRNAME=pattern,ALTER=ENABLED
  - STRNAME=strname
  - STRNAME=strprfx*
  - STRNAME=ALL | STRNAME=*
- D XCF,STRUCTURE, ALTER={ENABLED|DISABLED}
- Only systems with support will honor ALTER=DISABLED indicator in the active policy
  - So you may not get the desired behavior until the function is rolled around the sysplex
  - But fall back is trivial since downlevel code ignores it
- APAR OA34579 for z/OS V1R10 and up
  - OA37566 as well

z/OS V1R13 - SDSF

- SDSF provides sysplex view of panels:
  - Health checks; processes; enclaves; JES2 resources
- Data gathered on each system using the SDSF server
- Consolidated on client for display so user can see data from all systems
- Previously used MQ series to send and receive requests
  - Requires configuration and TCP/IP, instance of MQ queue manager on each system
- z/OS V1R13 implementation uses XCF Client/Server
  - No additional configuration requirements

z/OS V1R13 – 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

z/OS V1R13 – Runtime Diagnostics …

Does what you might do manually today:

- Review critical messages in the log
- Analyze contention
  - GRS ENQ
  - GRS Latches
  - z/OS UNIX file system latches
- Examine address spaces with high CPU usage
- Look for an address space that might be in a loop
- Evaluate local lock conditions
- Perform additional analysis based on what is found
  - For example, if XES reports a connector as unresponsive, RTD will investigate the appropriate address space

z/OS V1R13 - zFS

- Full read/write capability from anywhere in the sysplex for shared file systems
  - Better performance for systems that are not zFS owner
  - Reduced overhead on the owner system
- Expected to improve performance of applications that use zFS services
  - z/OS UNIX System Services
  - WebSphere® Application Server

Appendix

Sysplex Related Highlights From
z/OS V1R12

z/OS V1R12 Summary

- REALLOCATE
- Critical Members
- CFSTRHANGTIME
- Support for CFLEVEL 17
- Health Checks
- Auto Reply
- XCF Programming Interfaces

Background - REALLOCATE

- **SETXCF START,REALLOCATE**
  - Puts structures where they belong
- **Well-received, widely exploited for CF structure management**
- **For example, to apply "pure" CF maintenance:**
  - SETXCF START,MAINTMODE,CFNAME=cfname
  - SETXCF START,REALLOCATE to move structures out of CF
  - Perform CF maintenance
  - SETXCF STOP,MAINTMODE,CFNAME=cfname
  - SETXCF START,REALLOCATE to restore structures to CF

Background - REALLOCATE

But…

- Difficult to tell what it did
  - Long-running process
  - Messages scattered all over syslog
  - Difficult to find and deal with any issues that arose
- And people want to know in advance what it will do

z/OS V1R12 - REALLOCATE

- DISPLAY XCF,REALLOCATE,option

- TEST option
  - Provides detailed information regarding what REALLOCATE would do if it were to be issued
  - Explains why an action, if any, would be taken
- REPORT option
  - Provides detailed information about what the most recent REALLOCATE command actually did do
  - Explains what happened, but not why

z/OS V1R12 – REALLOCATE …

Caveats for TEST option

- Actual REALLOCATE could have different results
  - Environment could change
  - For structures processed via user-managed rebuild, the user could make “unexpected” changes
  - Capabilities of systems where REALLOCATE runs differ from the system where TEST ran
    - For example, connectivity to coupling facilities
- TEST cannot be done:
  - While a real REALLOCATE (or POPCF) is in progress
  - If there are no active allocated structures in the sysplex

z/OS V1R12 – REALLOCATE …

Caveats for REPORT option

- Can be done during or after a real REALLOCATE, but not before a real REALLOCATE is started
- A REPORT is internally initiated by XCF if a REALLOCATE completes with exceptions

z/OS V1R12 - Critical Members

- A system may appear to be healthy with respect to XCF system status monitoring, namely:
  - Updating status in the sysplex CDS
  - Sending signals
- But is the system actually performing useful work?
  - There may be critical functions that are non-operational
  - Which in effect makes the system unusable, and perhaps induces sympathy sickness elsewhere in the sysplex
- Action should be taken to restore the system to normal operation OR it should be removed to avoid sympathy sickness

z/OS V1R12 - Critical Members …

- **A Critical Member is a member of an XCF group that Identifies itself as "critical" when joining its group**
- **If a critical member is "impaired" for long enough, XCF will eventually terminate the member**
  - Per the member's specification: task, space, or system
  - SFM parameter MEMSTALLTIME determines "long enough"

- **GRS is a "system critical member"**
  - XCF will remove a system from the sysplex if GRS on that system becomes "impaired"

z/OS V1R12 - Critical Members …

- New Messages
  - IXC633I “member is impaired”
  - IXC634I “member no longer impaired”
  - **IXC635E “system has impaired members”**
  - IXC636I “impaired member impacting function”
- Changed Messages
  - IXC431I “member stalled” (includes status exit)
  - IXC640E “going to take action”
  - IXC615I “terminating to relieve impairment”
  - IXC333I “display member details”
  - IXC101I, IXC105I, IXC220W “system partitioned”

z/OS V1R12 - Critical Members …

- **Coexistence considerations**
  - Toleration APAR OA31619 for systems running z/OS V1R10 and z/OS V1R11 should be installed before IPLing z/OS V1R12
  - The APAR allows the down level systems to understand the new sysplex partitioning reason that is used when z/OS V1R12 system removes itself from the sysplex because a system critical component was impaired
  - If the APAR is not installed, the content of the IXC101I and IXC105I messages will be incorrect

z/OS V1R12 - Critical Members …

- **Potential migration action**
  - Evaluate, perhaps change MEMSTALLTIME parameter

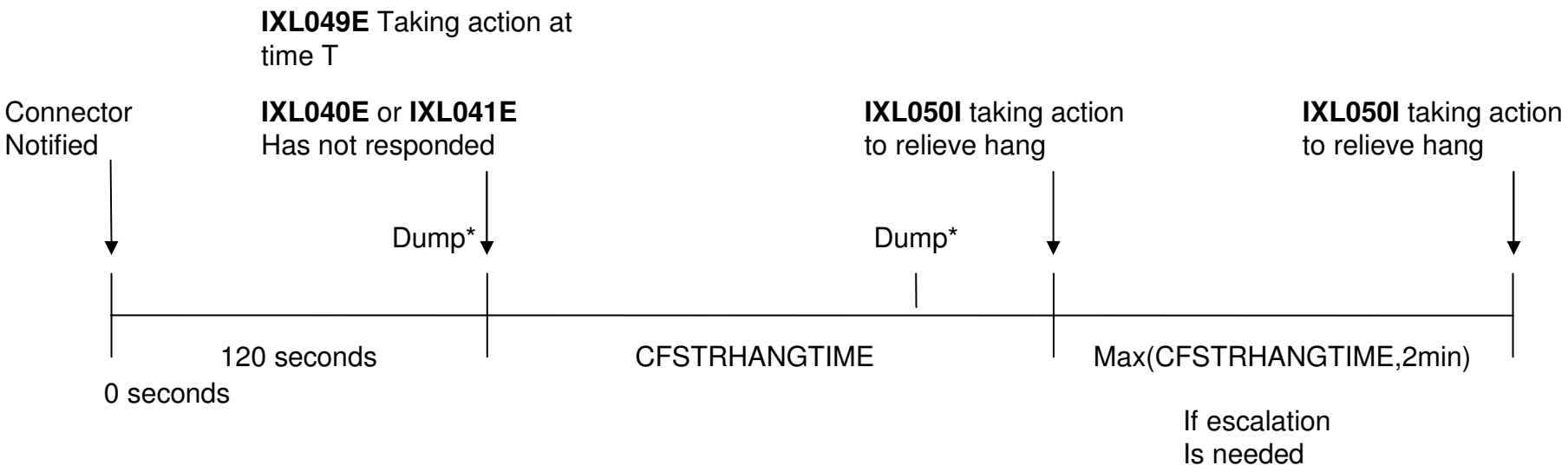
XES Connector Hang Detection

- Connectors to CF structures need to participate in various processes and respond to relevant events
- XES monitors the connectors to ensure that they are responding in a timely fashion
- If not, XES issues messages (IXL040E, IXL041E) to report the unresponsive connector
- Users of the structure may hang until the offending connector responds or is terminated
  - Impact: sympathy sickness, delays, outages
- Need a way to resolve this automatically …

z/OS 1VR12 – CFSTRHANGTIME …

- CFSTRHANGTIME
  - A new SFM Policy specification
  - Indicates how long the system should allow a structure hang condition to persist before taking corrective action(s) to remedy the situation
- Corrective actions may include:
  - Stopping rebuild
  - Forcing the user to disconnect
  - Terminating the connector task, address space, or system

z/OS V1R12 – CFSTRHANGTIME Processing



Dump* = Base release, dump is taken either when hang is announced or just prior to termination.
With OA34440, dump taken only when hang is announced

z/OS 1.12 – CFSTRHANGTIME …

New Messages

IXL049E HANG RESOLUTION ACTION FOR CONNECTOR NAME: conname
TO STRUCTURE strname, JOBNAME: jobname, ASID: asid:
actiontext

IXL050I CONNECTOR NAME: conname TO STRUCTURE strname,
JOBNAME: jobname, ASID: asid
HAS NOT PROVIDED A REQUIRED RESPONSE AFTER noresponsetime SECONDS.
TERMINATING termtarget TO RELIEVE THE HANG.

z/OS V1R12 – CFSTRHANGTIME …

- Coexistence
  - Toleration APAR OA30880 for z/OS V1R10 and z/OS V1R11 makes reporting of the CFSTRHANGTIME keyword with IXCMIAPU utility possible on those releases.
  - However the capability to take action to resolve the problem is not rolled back to previous releases

z/OS 1.12 – Support for CFLEVEL 17

- Large CF Structures
  - Increased CF structure size supported by z/OS to 1TB
  - Usability enhancements for structure size specifications
    - CFRM policy sizes
    - Display output
- More CF Structures can be defined
  - New z/OS limit is 2048 (CF limit is 2047)
- More Structure Connectors (CF limit is 255)
  - Lock structure – new limit is 247
    Serialized list – new limit is 127
    Unserialized list – new limit is 255

z/OS 1.12 – Support for CFLEVEL 17 …

- A new version of the CFRM CDS is needed to define more than 1024 structures in a CFRM policy

- May need to roll updated software around the sysplex for any exploiter that wants to request more than 32 connectors to list and lock structures
  - Not aware of any at this point (so really just positioning for future growth)

z/OS 1.12 – Support for CFLEVEL 17 …

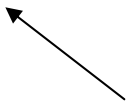
- z/OS requests non-disruptive CF dumps as appropriate
- Coherent Parallel-Sysplex Data Collection Protocol
  - Exploited for duplexed requests
  - Triggering event will result in non-disruptive dump from both CFs, dumps from all connected z/OS images, and capture of relevant link diagnostics within a short period
  - Prerequisites:
    - Installation must ENABLE the XCF function DUPLEXCFDIAG
    - z/OS 1.12
    - z/OS 1.10 or 1.11 with OA31392 (IOS) and OA31387 (XES)
  - Note that full functionality requires that:
    - z/OS image initiating the CF request reside on a z196
    - CF that “spreads the word” reside on a z196

z/OS 1.12 Health Checks

- XCF_CF_PROCESSORS
  - Ensure CF CPU's configured for optimal performance
- XCF_CF_MEMORY_UTILIZATION
  - Ensure CF storage is below threshold value
- XCF_CF_STR_POLICYSIZE
  - Ensure structure SIZE and INITSIZE values are reasonable

z/OS 1.12 Health Checks …

- XCF_CDS_MAXSYSTEM
  - Ensure function CDS supports at least as many systems as the sysplex CDS
- XCF_CFRM_MSGBASED
  - Ensure CFRM is using desired protocols
- XCF_SFM_CFSTRHANGTIME
  - Ensure SFM policy using desired CFSTRHANGTIME specification



Initially complained if more than 300 (5 minutes). APAR OA34439 changed it to 900 (15 minutes) to allow more time for operator intervention and more time for all rebuilds to complete after losing connectivity to a CF

z/OS 1.12 Auto-Reply

- Fast, accurate, knowledgeable responses can be critical
- Delays in responding to WTOR's can impact the sysplex
- Parmlib member defines a reply value and a time delay for a WTOR. The system issues the reply if the WTOR has been outstanding longer than the delay

- Very simple automation
- **Can be used during NIP !**

z/OS 1.12 Auto-Reply ….

- For example:

```
IXC289D REPLY U TO USE THE DATA SETS LAST USED FOR
typename OR C TO USE THE COUPLE DATA SETS SPECIFIED
IN COUPLExx
```

- **The message occurs when the couple data sets specified in the COUPLExx parmlib member do not match the ones in use by the sysplex (as might happen when the couple data sets are changed dynamically via SETXCF commands to add a new alternate or switch to a new primary)**
- **Most likely always reply “U”**

z/OS 1.12 - XCF Programming Interfaces

- IXCMSGOX
  - 64 bit storage for sending messages
  - Duplicate message toleration
  - Message attributes: Recovery, Critical
- IXCMSGIX
  - 64 bit storage for receiving messages
- IXCJOIN
  - Recovery Manager
  - Critical Member
  - Termination level