Introduction to Parallel Sysplex Performance

Joan Kelley
IBM Corp
Poughkeepsie, NY

Trademarks and Disclaimers

The following are trademarks of the International Business Machines Corportation:

S/390
CICS
IMS/TM
IMSDB
VTAM
RACF
DFSMS
VSAM
MVS/ESA
Parallel Sysplex
RMF
z/OS

The following is a registered trademark:

DB2

Performance numbers were achieved in a controlled laboratory environment and may vary based on customer environments

Overview

1. General Sysplex Tuning

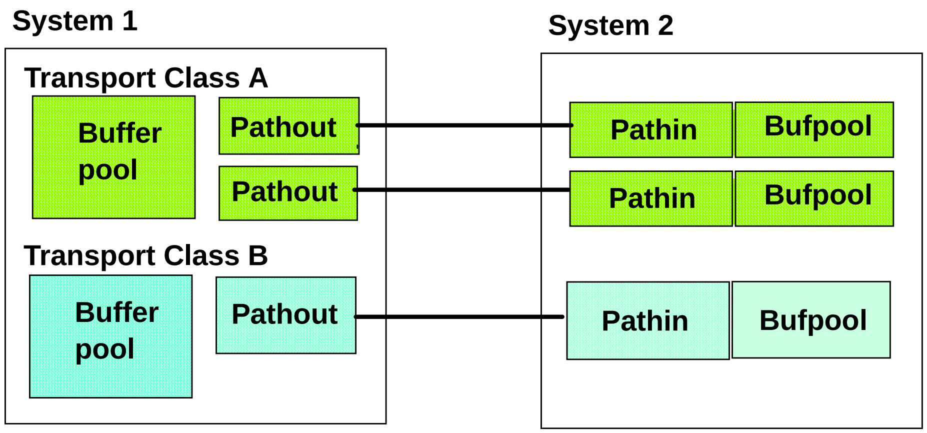
2. Tuning Coupling Facilities

3. Other Parallel Sysplex resources

4. LPAR Considerations

XCF Tuning

Resources used by XCF communication



Transport class definitions group messages by

■ Group name ■ Message size

► Pool resources by defining a minimum number of transport classes based on message size

Message Buffers

CLASSLEN defines buffer size

- If too small, XCF will expand (and contract) buffers, generating extra internal signals

```
                 XCF  USAGE  BY  SYSTEM
 ----------------------------------------------------------
                                             REMOTE  SYSTEMS
 ----------------------------------------------------------
                           OUTBOUND  FROM  JF0
 ----------------------------------------------------------
                                          ----- BUFFER -----
TO         TRANSPORT  BUFFER         REQ    %    %    %    %
SYSTEM     CLASS      LENGTH         OUT  SML  FIT  BIG  OVR
JA0        DEFAULT    20,412      15,449  100    0   <1  100
           DEFSMALL      956      88,960    0  100    0    0
           DEF8K       8,124       3,827   74   26    0    0
```

%BIG should be small (<10%)

► Increase CLASSLEN for largest transport class

JdK

Message Buffer Space

Fixed real and expanded storage

MAXMSG defines upper limit for various resources

- If too small, request could be rejected

```
                                 XCF  USAGE  BY  SYSTEM
 -----------------------------------------------------------------------------
               OUTBOUND  FROM  JF0                       INBOUND  TO  JF0
 --------------------------------------------------    -----------------------

TO       TRANSPORT  BUFFER          REQ       REQ    FROM          REQ     REQ
SYSTEM   CLASS      LENGTH   ...    OUT  ... REJECT  SYSTEM         IN  REJECT
JA0      DEFAULT    20,412       15,449           0  JA0        61,475       0
         DEFSMALL      956       88,960           0
         DEF8K       8,124        3,827           0
```

► Let MAXMSG default

If REQ REJECT>0, increase MAXMSG for that resource

JdK

Signaling Paths

- Insufficient number of paths
  - Messages will queue up

```
                           XCF PATH STATISTICS
--------------------------------------------------------------------------------
                           OUTBOUND FROM JF0
--------------------------------------------------------------------------------
        T FROM/TO
TO      Y DEVICE, OR      TRANSPORT        REQ   AVG Q
SYSTEM  P STRUCTURE       CLASS            OUT   LNGTH    AVAIL    BUSY    RETRY
JA0     S IXCPLEX_PATH1   DEFAULT       15,449    0.01   15,449       0        0
        S IXCPLEX_PATH2   DEFSMALL      43,853    0.02   43,468     385        0
        S IXCPLEX_PATH3   DEF8K          3,827    0.00    3,827       0        0
        C C600 TO C654    DEFSMALL       2,288    0.01    2,074     214        0
        C C601 TO C655    DEFSMALL       4,119    0.01    3,806     313        0
        C C602 TO C656    DEFSMALL      38,906    0.02   38,656     250        0
```

If AVG Q LNGTH > 1.00

- ► Consider adding more paths, or a different type of path
  - TYP indicates CF Structure(S) or CTC(C)
  - CF structures equivalent to CTCs with Hyperlinks (ISC2), faster with ICBs

JdK

Signaling Paths

- No paths

  Messages will be rerouted to another class

```
                         XCF USAGE BY SYSTEM
-----------------------------------------------------------
                                          REMOTE SYSTEMS
-----------------------------------------------------------
                        OUTBOUND FROM SY03
-----------------------------------------------------------
                                                         ALL
TO        TRANSPORT  BUFFER        REQ                 PATHS
SYSTEM    CLASS      LENGTH        OUT               UNAVAIL
SY04      DEF8K       8,124      2,564  ...            2,564
          DEF1K         956    164,158                     0
          DEF4K       4,028          1                     1
          ....
                                  ----------
        TOTAL                        289,260
```

If ALL PATHS UNAVAIL >0

- ► Verify Path Definition
- ► Check physical connection

JdK

Managing XCF traffic

1. Eliminate unnecessary traffic
   - Tune XCF
   - Place shared resources on system with heaviest usage
   - Reduce lock contention
   - WLM Dyn Alias Mgmt APAR - OW50276

2. Improve Response time
   - Measuring XCF response time
   - Performance comparison of various types of XCF paths

3. Provide more capacity
   - How determine when more XCF paths are needed
   - Increase structure size if increasing number of systems in sysplex

Placement of shared resources

OAM - Object Access Method

DFSMS 1.5.0 allows shared optical devices

HFS - Shared Hierarchical File Systems

OS/390 R9 allows simultaneous R/W access of HFS

Both use XCF to pass data, so depending on workload, can cause a LOT of XCF traffic

- ► Put shared resource on system with heaviest workload
- ► Add paths/transport classes if needed
- ► See http://www.ibm.com/servers/eserver/pseries/unix Performance for latest HFS performance tips

Reduce Lock contention

- Lock manager sends XCF signals to resolve contention (check for increased traffic in Group name IXCO0xxx)
  - ►Use D XCF command to associate group with lock

```
D XCF,STR,STRNAME=IRLMLOCK1
IXC360I  11.33.11  DISPLAY XCF       FRAME  1     F
STRNAME: IRLMLOCK1
 STATUS: ALLOCATED
  POLICY SIZE    : 64000 K
  SYSTEM-MANAGED PROCESS LEVEL: 8
  ...
  XCF GRPNAME    : IXCLO009
```

- ►Determine / correct cause of contention
  - – False contention -> increase structure size
  - – GRSSTAR contention -> clean up enqueues
  - – Datasharing -> more frequent checkpoints on long running jobs

Display Command

D XCF, PI, STRNM=ALL gives additional info

```
IXC356I  14.46.26  DISPLAY XCF 783
STRNAME          REMOTE   PATHIN      UNUSED                   LAST  MXFER
PATHIN           SYSTEM   STATUS      PATHS    RETRY  MAXMSG RECVD TIME
IXCPLEX_PATH1             WORKING        84      100    1000     -     -
                 JF0      WORKING                             7540  1307
                 JG0      WORKING                             6806   777
                 J80      WORKING                             6961   917
                 J90      WORKING                             7711   642

STRNAME          REMOTE    PATHIN    DELIVRY BUFFER  MSGBUF SIGNL
PATHIN     LIST  SYSTEM    STATUS    PENDING LENGTH  IN USE NUMBR NOBUF
IXCPLEX_PATH1
             46  JF0       WORKING         0  20412       0  7540     0
            153  JG0       WORKING         0  20412       0  6806     5
             38  J80       WORKING         0  20412       0  6961     0
            111  J90       WORKING         0  20412      22  7711     0
```

Measuring XCF response time

OW38138 adds Mean Transfer time to D XCF display

```
D XCF,PI,STRUCTURE=ALL,STATUS=WORKING
IXC356I  03.06.04  DISPLAY XCF 295
STRNAME         REMOTE   PATHIN     UNUSED                   LAST   MXFER
PATHIN          SYSTEM   STATUS     PATHS    RETRY   MAXMSG  RECORD TIME
IXCPLEX_PATH2            WORKING        0      100     1000      -      -
                JA0      WORKING                               5747   1251
                JA0      WORKING                               5871   1264
```

```
D XCF,PI,DEVICE=ALL,STATUS=WORKING
IXC356I  03.14.02  DISPLAY XCF      FRAME  1     F      E   SYS=Z0
LOCAL DEVICE    REMOTE   PATHIN     REMOTE                   LAST   MXFER
PATHIN          SYSTEM   STATUS     PATHOUT RETRY   MAXMSG  RECORD  TIME
C604            JA0      WORKING    C450      100     1000   85256  1754
C605            JA0      WORKING    C451      100     1000   80613  3189
C606            JA0      WORKING    C452      100     1000   73662  1582
```

OW41317 stores this data in SMF 74.2 record

JdK

Displaying SMF fields - ERBSCAN - 2.6.0

Use ISPF 3.4 to display the SMF datasets

```
  -----------------------------------------------------------------
 ERBSCAN   SMFDATA.SMFTPN.G6540V00
           SMFDATA.SMFTPN.G6541V00
```

Select SMF record

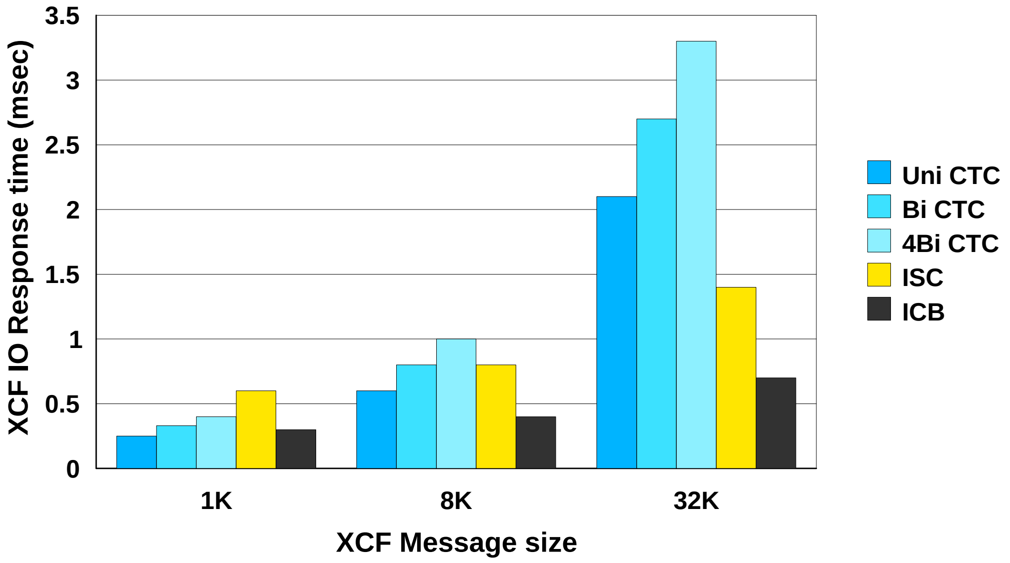
```
Command ===> ERBSHOW 388
 388 074.002 32716 2000.236 12.10.01 2000.236 12.05.00 05.00.000 TPN
 389 074.002 13700 2000.236 12.10.01 2000.236 12.05.00 05.00.000 TPN
 390 074.003   456 2000.236 12.10.01 2000.236 12.05.00 05.00.000 TPN
```

Find desired section and field

```
Record Number 924: SMF Record Type 74(2) - RMF XCF Activity
===========================================================
  -> Path Data Section (168)
 ==========================
    #26:  +0000:  E3D7D540 40404040 40404040 00800300  *TPN
          +0010:  D1C1F040 40404040 40404040 20000000  *JA0
          +0020:  00000064 00000000 000003E8 000000CF  *          Y
          +0030:  00000000 00000000 00000000 00000000  *
          +0040:  40404040 40404040 C9E7C3D7 D3C5E76D  *        IXCP
          +0050:  D7C1E3C8 F1404040 00003745           *PATH1
```

JdK

XCF - Path Comparison



JdK

More XCF paths?

Minimum of two physical paths for availability

Check response time with D XCF command

RMF AVG Q LNGTH is not a good indicator for CTCs

- Message rate capacity depends of
  - Size of message
  - How paths are defined (ex, UNI, BI, Multiple BI)
  - Other users of path (ex, VTAM)

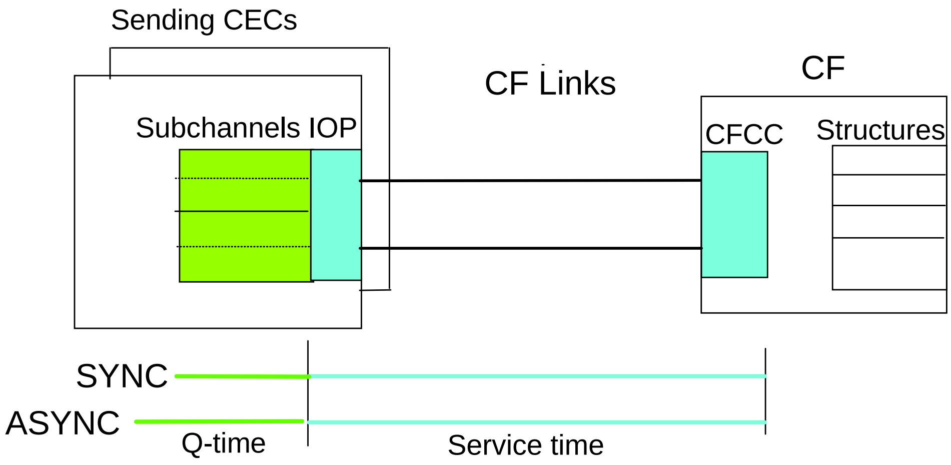
Max capacity

CTC	HiPerLink	ICB
1000-5000/sec	4000/sec	9000/sec

See WSC Flash 10011 for complete XCF tuning story

JdK

Tuner's view of CF resources



Sending CECs

Subchannels IOP

CF Links

CF

CFCC

Structures

SYNC

ASYNC

Q-time

Service time

- 2 (COMPAT) or 7 (PEER) subchannels for each CF link
- IOP handles I/O, CTC and ASYNC CF requests
- CF links can be shared (EMIFed) if multiple MVS images

JdK

What are the best service times I can expect?

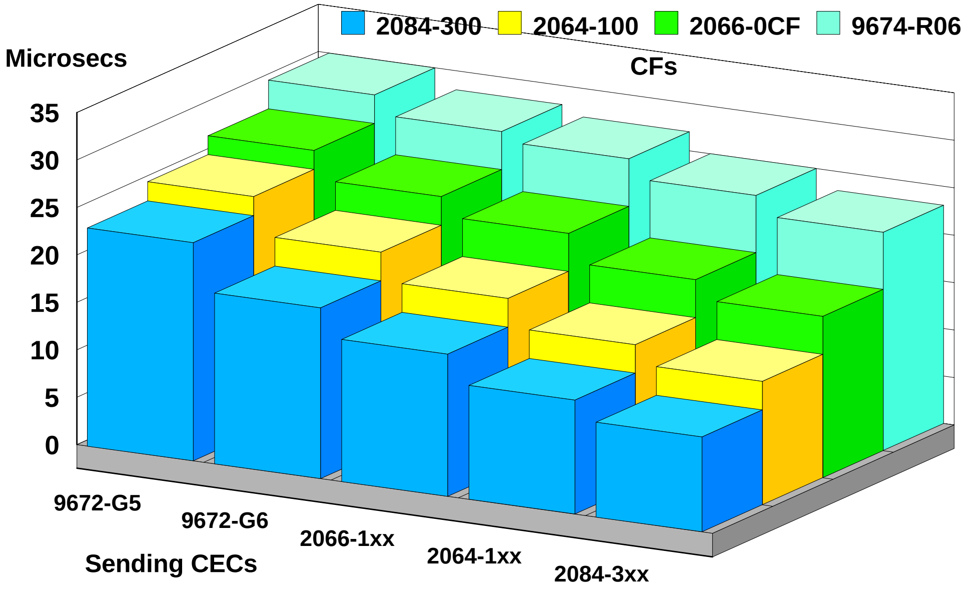
	9672-Rx6 to R06	2064-114 to 1xx	2084-3xx to 3xx
SYNC	30 - 60	15 - 30	10 -25
ASYNC	300 - 900	150 - 450	100-350

1. Range to account for amount of data being transferred
   - Low end - no data (ex. GRSLOCK)
   - High end - largest data transfer allowed
     - SYNC - 4K
     - ASYNC - 64K

2. Assumes fastest CF link technology available on that processor and a well-tuned sysplex

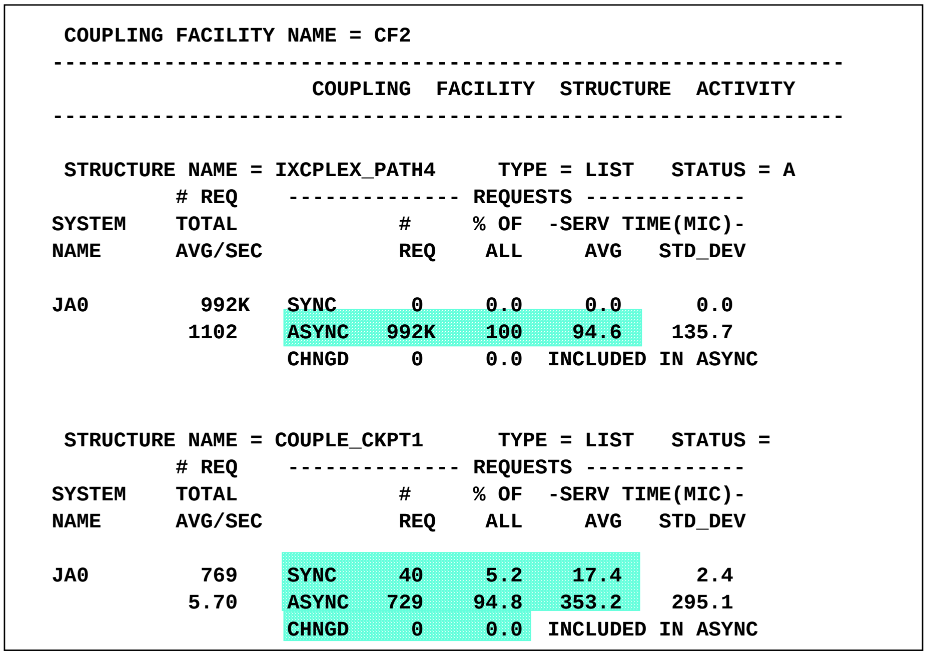
JdK

Lock Service Times



JdK

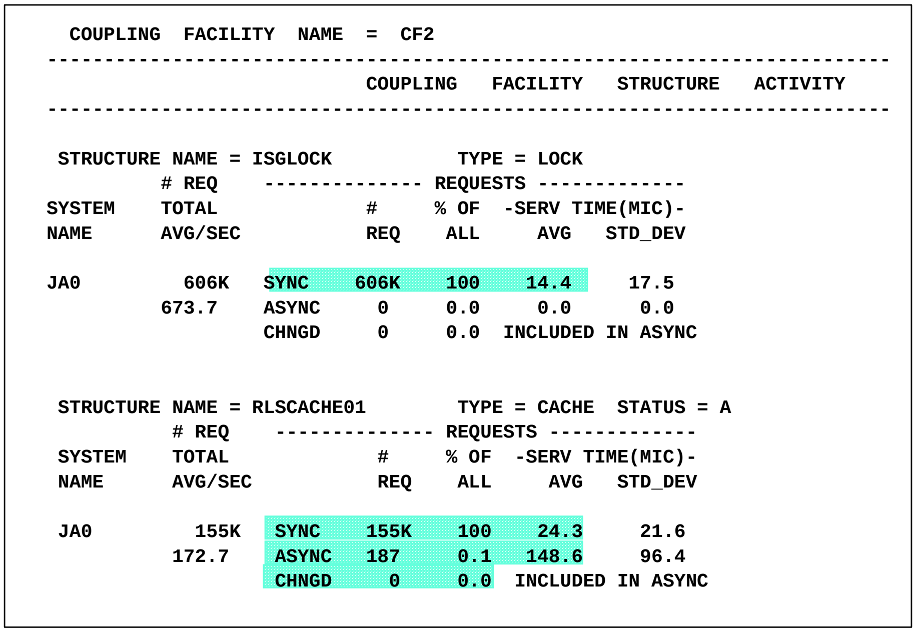
RMF PP - CF Structure Activity



```
 COUPLING FACILITY NAME = CF2
----------------------------------------------------------------
                     COUPLING  FACILITY  STRUCTURE  ACTIVITY
----------------------------------------------------------------

 STRUCTURE NAME = IXCPLEX_PATH4     TYPE = LIST   STATUS = A
          # REQ    -------------- REQUESTS -------------
SYSTEM    TOTAL            #     % OF  -SERV TIME(MIC)-
NAME      AVG/SEC          REQ    ALL     AVG   STD_DEV

JA0         992K   SYNC      0    0.0     0.0      0.0
            1102   ASYNC   992K   100    94.6    135.7
                   CHNGD     0    0.0  INCLUDED IN ASYNC


 STRUCTURE NAME = COUPLE_CKPT1      TYPE = LIST   STATUS =
          # REQ    -------------- REQUESTS -------------
SYSTEM    TOTAL            #     % OF  -SERV TIME(MIC)-
NAME      AVG/SEC          REQ    ALL     AVG   STD_DEV

JA0          769   SYNC     40    5.2    17.4      2.4
            5.70   ASYNC   729   94.8   353.2    295.1
                   CHNGD     0    0.0  INCLUDED IN ASYNC
```

JdK

RMF PP - CF Structure Activity

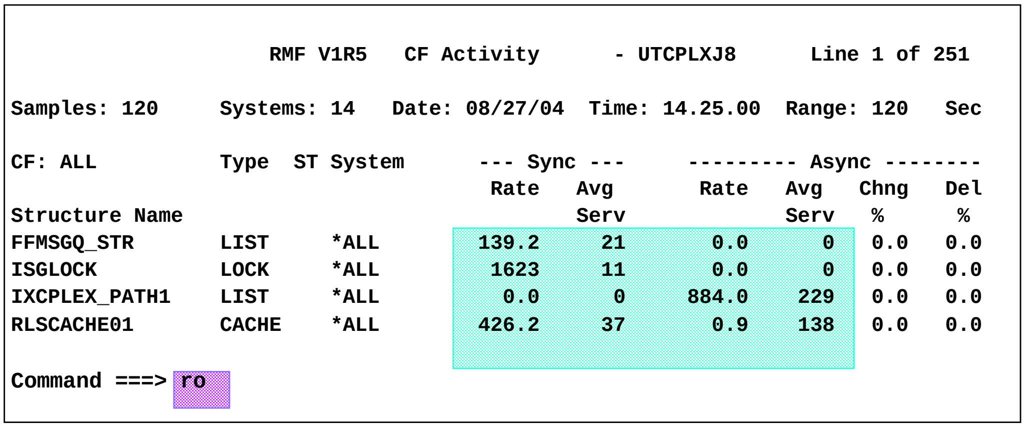


```
  COUPLING  FACILITY  NAME  =  CF2
--------------------------------------------------------------------------
                            COUPLING   FACILITY   STRUCTURE   ACTIVITY
--------------------------------------------------------------------------

 STRUCTURE NAME = ISGLOCK            TYPE = LOCK
          # REQ    -------------- REQUESTS ------------
SYSTEM    TOTAL             #      % OF  -SERV TIME(MIC)-
NAME      AVG/SEC           REQ    ALL      AVG   STD_DEV

JA0         606K   SYNC    606K    100     14.4     17.5
          673.7    ASYNC      0    0.0      0.0      0.0
                   CHNGD      0    0.0  INCLUDED IN ASYNC


 STRUCTURE NAME = RLSCACHE01         TYPE = CACHE  STATUS = A
           # REQ    -------------- REQUESTS ------------
 SYSTEM    TOTAL             #      % OF  -SERV TIME(MIC)-
 NAME      AVG/SEC           REQ    ALL      AVG   STD_DEV

 JA0         155K   SYNC    155K    100     24.3     21.6
           172.7    ASYNC   187     0.1    148.6     96.4
                    CHNGD      0    0.0  INCLUDED IN ASYNC
```

JdK

RMF Mon III - CF Structure Activity

Select S.7 - Coupling Facility Activity



```
                      RMF V1R5   CF Activity      - UTCPLXJ8      Line 1 of 251

Samples: 120      Systems: 14   Date: 08/27/04  Time: 14.25.00  Range: 120   Sec

CF: ALL           Type  ST System      --- Sync ---      --------- Async --------
                                        Rate    Avg       Rate    Avg   Chng   Del
Structure Name                                  Serv              Serv   %      %
FFMSGQ_STR        LIST     *ALL        139.2     21        0.0      0   0.0    0.0
ISGLOCK           LOCK     *ALL         1623     11        0.0      0   0.0    0.0
IXCPLEX_PATH1     LIST     *ALL          0.0      0      884.0    229   0.0    0.0
RLSCACHE01        CACHE    *ALL        426.2     37        0.9    138   0.0    0.0

Command ===> ro
```

Use 'ro' to change options: Detail yes shows each image

For more info - put cursor under data field and hit enter

JdK

Potential sources of delay

If exceed guidelines, possible causes are:

- Insufficient CF capacity
- IOP Contention
- Shortage of CF subchannels
- Contention for CF paths

Insufficient CF Capacity

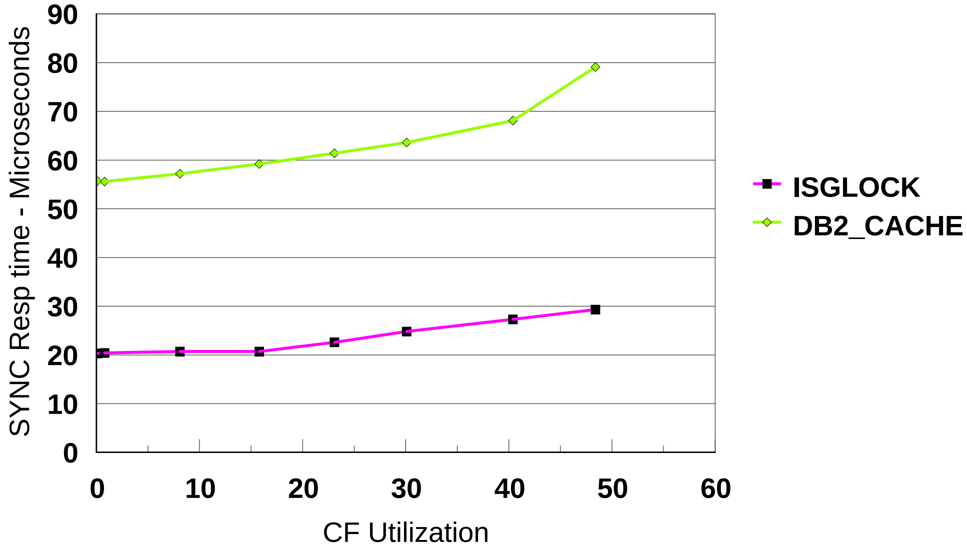
R.O.T - Best response time if CF Util <50%

If CF Util. > 50% for

- ► Verify all CPs are operational
  Check LOGICAL PROCESSORS DEFINED
- ► Verify the CF CP resource is what you expected
  Check LOGICAL PROCESSORS EFFECTIVE
- ► If one CF is much busier than the other, redistribute the structures based on ALLOC SIZE and # REQ
- ► Upgrade CF - more CPs, faster CPs

Service time as a function of CF utilization

CF1 - z990, 2 Ded CPs, ISC links, Sender - z990



JdK

CF Configuration Options

Many combinations

1. Standalone CF (ex. 2066 - 0CF, 2084 - 300)
   - Dedicated CPs - best choice for production
   - **Shared CPs**

2. Internal CF (ex. 2064 - 108)
   - Dedicated CPs (expensive - added into S/W license costs)
   - Dedicated ICFs - good choice for production if...
   - **CPs shared with MVS images**
   - **CPs shared with other CF images**

JdK

CF - Dedicated CPs

- Standalone CF with dedicated CPs
  - ► Best choice for primary production CF

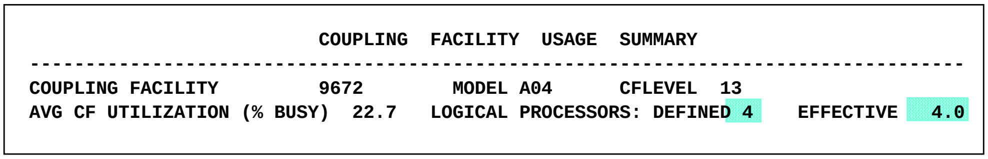
- Internal CF with dedicated CPs (ICF)
  - ► On G6, internal CF is only version available
  - ► Best suited to
    - ● CF for a single CEC sysplex
    - ● CF which is not part of this sysplex
    - ● Structure which don't need a local copy from the failing system to rebuild
  - ► Sug - Define on a CEC which is not using structures which need a local copy to rebuild

JdK

How many CPs really assigned to the CF?

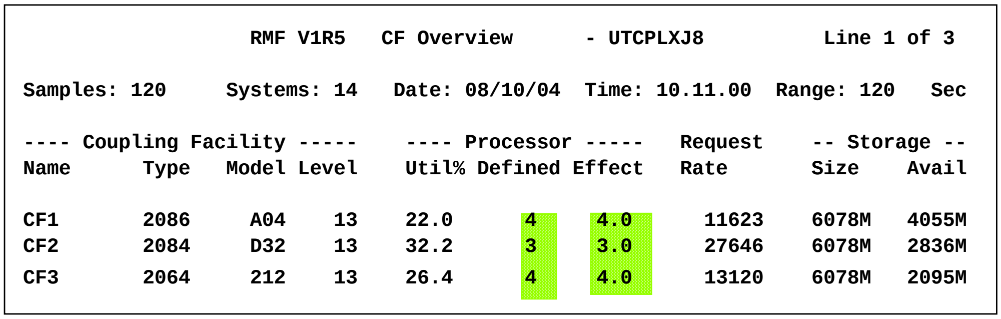
1. Standalone CF - No operating system so only RMF CF reports

Mon I - Post processor



```
                        COUPLING  FACILITY  USAGE  SUMMARY
 ------------------------------------------------------------------------------
 COUPLING FACILITY         9672        MODEL A04      CFLEVEL  13
 AVG CF UTILIZATION (% BUSY)  22.7   LOGICAL PROCESSORS: DEFINED 4    EFFECTIVE    4.0
```

Mon III - Real time



```
                  RMF V1R5   CF Overview      - UTCPLXJ8          Line 1 of 3

Samples: 120     Systems: 14   Date: 08/10/04  Time: 10.11.00  Range: 120   Sec

---- Coupling Facility -----   ---- Processor -----   Request     -- Storage --
Name       Type   Model Level   Util% Defined Effect   Rate        Size    Avail

CF1        2086    A04    13    22.0       4     4.0    11623      6078M   4055M
CF2        2084    D32    13    32.2       3     3.0    27646      6078M   2836M
CF3        2064    212    13    26.4       4     4.0    13120      6078M   2095M
```

JdK

How many CPs really assigned to the CF?

2. Internal CFs - RMF CF report

```
                           COUPLING  FACILITY  USAGE  SUMMARY
---------------------------------------------------------------------------------------
COUPLING FACILITY          2064        MODEL 106      CFLEVEL  12
AVG CF UTILIZATION (% BUSY)  8.3    LOGICAL PROCESSORS: DEFINED 1     EFFECTIVE    0.7
```

RMF Partitioned Activity Report

```
Number of Physical Processors            8
                   CP                    6
                   ICF                   2
...
                     ---MSU---   Cap  Processor-  Logical Processors -- Physical Processors --
Name  S   Wgt  Def    Act  Def  WLM%  Num   Type  Effective    Total  LPAR Mgt  Effect.  Total
 ---  -   --- ----   ----  ---  -----  ---   ----  ---------    -----  ---------  -------  -----
CF1A  A    10    0    N/A  NO    0.0    1   ICF      66.23    66.27      0.02    33.12  33.13
CF2A  A    10    0    N/A  NO    0.0    1   ICF      66.24    66.27      0.02    33.12  33.14
CF2B  A    10    0    N/A  NO    0.0    1   ICF      66.24    66.27      0.02    33.12  33.14
*PHYSICAL*                                                               0.58             0.58
TOTAL                                                                    0.63    99.35  99.99
```

JdK

Implications of Sharing CF CPs

A. CF request response times

If a request to CF cannot be executed because the CF is timesliced out, the request waits:

1. If it's a SYNC request, the sender waits
   - Service times go up -> more sender cycles used
   - Heuristic algorithm provides some relief by changing SYNC to ASYNC
2. Subchannel is held longer -> more channel utilization

A. Processor utilization

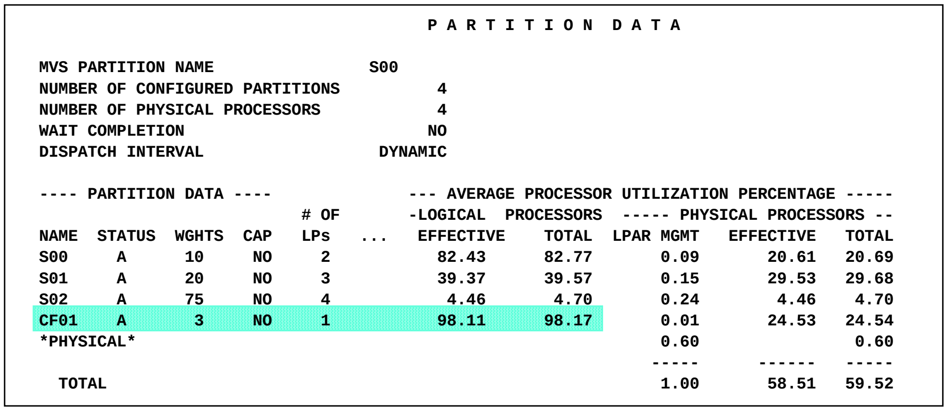
- LIC codes runs in an 'active wait' polling for work, so LPAR sees the CF image as 100% busy and give the CF all the processor resources available, even at very low CF rates.

- CP resource is apportioned by LPAR weight, **BUT**...LPAR gives each CF image control every 125 microsecs, so a CF image with low weight gets more resource than expected.

JdK

CF- Shared CPs

CF partition will use all the CP resource it can get



```
                                        P A R T I T I O N  D A T A

MVS PARTITION NAME                S00
NUMBER OF CONFIGURED PARTITIONS          4
NUMBER OF PHYSICAL PROCESSORS            4
WAIT COMPLETION                         NO
DISPATCH INTERVAL                  DYNAMIC

---- PARTITION DATA ----              --- AVERAGE PROCESSOR UTILIZATION PERCENTAGE -----
                           # OF       -LOGICAL  PROCESSORS  ----- PHYSICAL PROCESSORS --
NAME  STATUS  WGHTS  CAP   LPs   ...   EFFECTIVE    TOTAL  LPAR MGMT   EFFECTIVE   TOTAL
S00     A      10     NO     2            82.43     82.77       0.09       20.61   20.69
S01     A      20     NO     3            39.37     39.57       0.15       29.53   29.68
S02     A      75     NO     4             4.46      4.70       0.24        4.46    4.70
CF01    A       3     NO     1            98.11     98.17       0.01       24.53   24.54
*PHYSICAL*                                                      0.60                0.60
                                                               -----      ------   -----
  TOTAL                                                         1.00       58.51   59.52
```

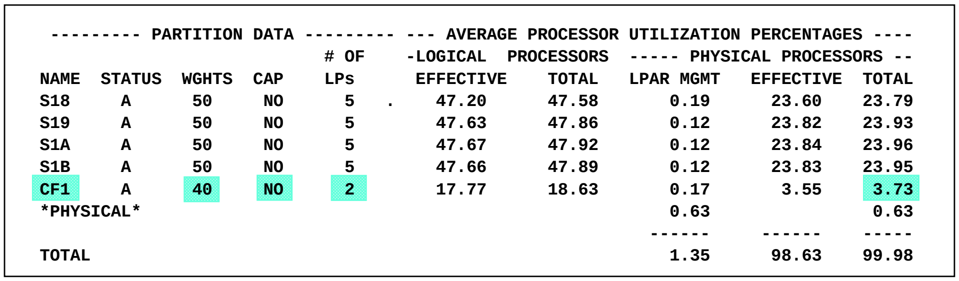
- Contention will limit CF CP resource to wgt. defined
- If the CF is sharing CPs, do not CAP the partition and give it a respectable weight (at least 50%)
- Anything less than a CP will elongate service time

JdK

CF - Shared CPs on Sending CEC

Dynamic CF Dispatching - allows tradeoff between CF response time and CP Utilization

- At low utilization, CFCC suspended for short periods
  - More CP resource for other partitions, but CF requests delayed
- As utilization increases, less CFCC suspension
  - Less CP resource for other partitions, but faster CF requests



```
 --------- PARTITION DATA --------- --- AVERAGE PROCESSOR UTILIZATION PERCENTAGES ----
                            # OF    -LOGICAL  PROCESSORS  ----- PHYSICAL PROCESSORS --
NAME  STATUS  WGHTS  CAP    LPs      EFFECTIVE    TOTAL   LPAR MGMT   EFFECTIVE  TOTAL
S18     A      50     NO      5   .    47.20      47.58       0.19      23.60    23.79
S19     A      50     NO      5        47.63      47.86       0.12      23.82    23.93
S1A     A      50     NO      5        47.67      47.92       0.12      23.84    23.96
S1B     A      50     NO      5        47.66      47.89       0.12      23.83    23.95
CF1     A      40     NO      2        17.77      18.63       0.17       3.55     3.73
*PHYSICAL*                                                    0.63                0.63
                                                            ------      ------   -----
TOTAL                                                         1.35      98.63    99.98
```

At low utilization, less CP resource used but...

JdK

Dynamic CF Dispatching

But CF response time increases....

```
                         COUPLING  FACILITY  USAGE  SUMMARY
 -------------------------------------------------------------------------------
AVG. CF UTIL. (%BUSY)   23.6%   LOGICAL PROCESSORS:  DEFINED  1   EFFECTIVE  0.0

                         COUPLING  FACILITY  STRUCTURE  ACTIVITY
 -------------------------------------------------------------------------------
STRUCTURE NAME = CFTWDB2_LOCK1     TYPE = LIST
           # REQ    -------------- REQUESTS ------------    ...
 SYSTEM    TOTAL             #     % OF   -SERV TIME(MIC)-
 NAME      AVG/SEC          REQ    ALL      AVG    STD_DEV

 J90         122    SYNC     54    3.6%  1219.6    1055.6
             2.03   ASYNC    68    4.5%  2004.2    2441.7
                    CHNGD     0    0.0%  INCLUDED IN ASYNC
```

As activity in the test CF partition increases, more CP resource is used and CF response time improves.

JdK

IOP Contention

```
           z/OS V1R2                 SYSTEM ID JG0
TOTAL SAMPLES =  17528   IODF = D8   CR-DATE: 02/06/2002
       - INITIATIVE QUEUE -   ------- IOP UTILIZATION -------
 IOP      ACTIVITY    AVG Q    % IOP   I/O START    INTERRUPT
            RATE      LNGTH    BUSY      RATE         RATE
  00       901.544     8.55    100.0    901.557      1425.612

LCU    CONTROL UNITS    DCM GROUP  CHAN     CHPID    % DP   % CU  CONTENTION
                       MIN MAX DEF PATHS    TAKEN    BUSY   BUSY     RATE
0031   BD80                         15      6.250    87.34  0.00
                                    47      5.980    87.87  0.00
                                     *     12.231    87.61  0.00     46.906
```

IOP handles

- I/O to DASD
- CTC traffic
- ASYNC CF requests

AVG Q LNGTH should be less than 1.0

JdK

CF Options - CF links

G2	G3	G4	G5	G6	z900
9672-Rx2/3	9672-Rx4	9672-Rx5	9672-Rx6	9672-Rx7	2064-1xx
C02/C03	C04	C05	R06		100*
ISC	ISC				*
	ISC-2 Hyperlink	ISC-2 Hyperlink	ISC-2 Hyperlink	ISC-2 Hyperlink	ISC-3 Compat
					ISC-3 Peer
			ICB	ICB	ICB Compat
					ICB-3 Peer
			IC	IC	IC-3 Peer
ICMF	ICMF	ICMF	ICMF	ICMF	

Details on valid z900 link combos in z/Series 900 - System Overview

JdK

ISC links

Optical fiber - for distances greater that 10 meters

- Now available in lengths up to 100K
  - Each 1K in length adds 10 microsecs to service time
  - May need additional links to handle traffic

```
STRUCTURE NAME = THRLSTCQS_1       TYPE = LIST   STATUS = ACTIVE
         # REQ    -------------- REQUESTS -----------    -------------- DELAYED REQUESTS -----------
SYSTEM   TOTAL              #     % OF  -SERV TIME(MIC)-    REASON     #    % OF  --- AVG TIME(MIC) ----
NAME     AVG/SEC           REQ     ALL      AVG   STD_DEV               REQ     REQ   /DEL    STD_DEV   /ALL

S08        251K   SYNC     16K     4.3     56.0      51.1    NO SCH 1712     0.7   27.1       76.9    0.2
         836.0    ASYNC   233K    64.3    289.8     487.7    PR WT     60     0.0   10.0        1.5    0.0
                  CHNGD   1999     0.6  INCLUDED IN ASYNC    PR CMP     0     0.0    0.0        0.0    0.0
                                                             DUMP       0     0.0    0.0        0.0    0.0

S09        112K   SYNC     632     0.2   1110.6     120.6    NO SCH 2398     2.2  487.9      703.1   10.7
         373.4    ASYNC   109K    30.0   1377.8     719.9    PR WT      0     0.0    0.0        0.0    0.0
                  CHNGD   2393     0.7  INCLUDED IN ASYNC    PR CMP     0     0.0    0.0        0.0    0.0
                                                             DUMP       0     0.0    0.0        0.0    0.0
```

JdK

Integrated Cluster Bus (ICB)

Copper cable plugged directly into STI (i.e, no link adapters). Maximum distance 7M (cable is 10M).

Link	Link Speed MB/sec	Link Mode	z990 Connectivity
ICB-2	250	Compat	G5,G6
ICB-3	500	Peer	z-series
ICB-4	1500	Peer	z990

Internal CF Links (IC)

Microcode CF Links

IC - G5, G6 - COMPAT mode

- Replacing Hiperlinks with ICs could produce a 50-60% improvement in SYNC service times - can result in 1-4% improvement in coupling overhead
- Much more efficient than ICMF, no LPAR interrupts
- Can be combined with physical links

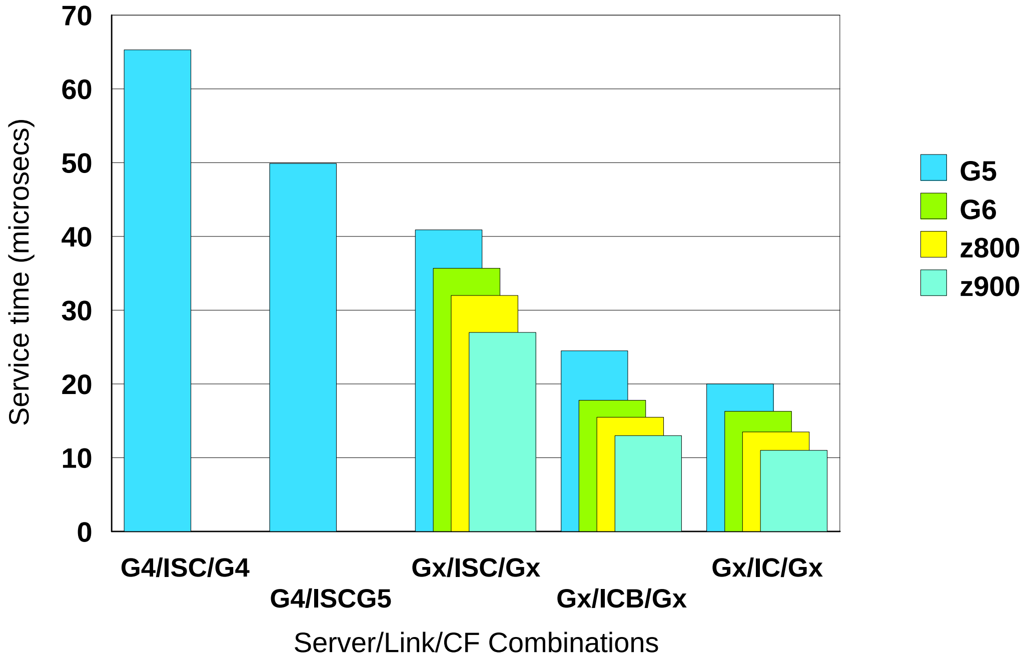
IC-3 - z800, z9xx - PEER mode

- Data rate twice IC

Two IC links are usually plenty

- Recommend a limit of 1 < Total #CP on the CEC
- Ex. z800-1C3 with 1 ICF -> Maximum of 3 IC links

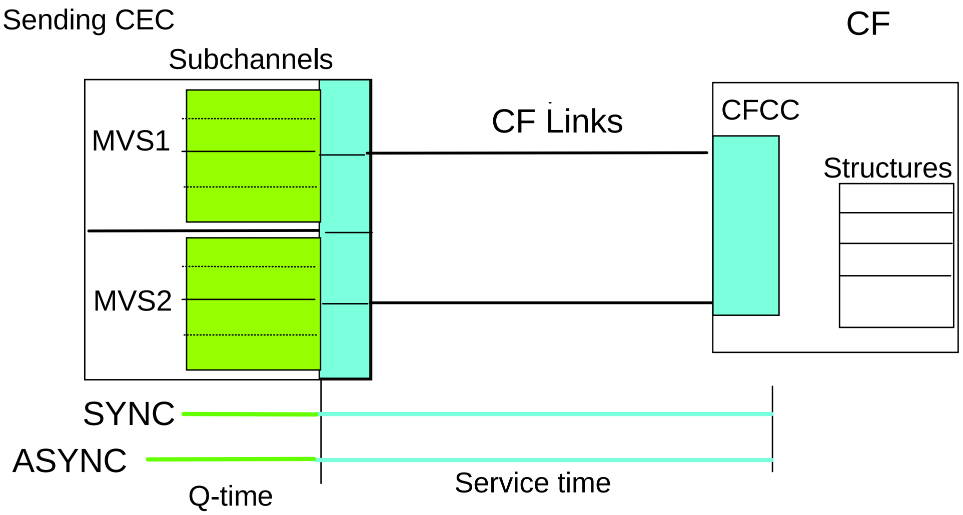
Example - ISGLOCK Structure



Jdk

JdK

Potential CF link delays



- MVS matches subchannels to links to avoid busy conditions
  But if links are shared (multiple MVS images), path busy's occur
- <z/OS 1.2, SYNC reqs retried immediately / ASYNC requeued
  z/OS 1.2, all reqs retried immediately.
- Path busy retry time included in service time in all cases.

JdK

Shortage of CF subchannels

Determine how many requests encounter subchannel busy

- SYNC requests - impact capacity
- ASYNC requests - impact response time to sender

```
                 SUBCHANNEL  ACTIVITY
------------------------------------------------------------------------------------
NAME      ----------- REQUESTS -----------    ------------- DELAYED REQUESTS ----------
SYSTEM             #    -SERVICE TIME(MIC)-           #    % OF   ----- AVG TIME(MIC)---
NAME             REQ      AVG     STD_DEV           REQ     REQ    /DEL  STD_DEV    /ALL

JG0 ...  SYNC    808546   101.0      57.2    SYNC    65    0.0%    90.3    214.5     0.0
         ASYNC   221733   462.9     471.5    ASYNC   11K   4.8%   300.2    254.4    14.5
         CHANGED     99  INCLUDED IN ASYNC   TOTAL   11K   1.0%
         UNSUCC       0     0.0       0.0

JI0 ...  SYNC      2445   148.0      60.1    SYNC     4    0.2%    88.0     54.4     0.1
         ASYNC    13401  1180.3      1915    ASYNC 2614   19.5%   804.7     1090   156.9
         CHANGED      3  INCLUDED IN ASYNC   TOTAL 2618   16.5%
         UNSUCC       0     0.0       0.0
```

Guideline - Total % of REQ delayed should be less than 10%

To assess overall impact of delay - /ALL

- ► Consider adding more subchannels (CF links)

JdK

Shortage of CF subchannels (CF links)

RMF report changed in z/OS 1.2

Now shows number of links of each type and number of subchannels (2 / COMPAT - 7 / PEER)

```
                                 SUBCHANNEL  ACTIVITY
--------------------------------------------------------------------------
          # REQ                              ----------- REQUESTS -----------
SYSTEM    TOTAL    -- CF LINKS --   PTH               #    -SERVICE TIME(MIC)-
NAME      AVG/SEC  TYPE   GEN  USE  BUSY             REQ       AVG      STD_DEV
JA0         4099K  CBP      2    2   276   SYNC      2980K    70.1        97.1
          2277.4   CFP      2    2         ASYNC     1116K   307.2        1623
                   SUBCH   28   28         CHANGED      79   INCLUDED IN ASYNC
                                           UNSUCC        0     0.0         0.0
Z0          4171K  CBP      2    2     0   SYNC     356165    47.0        81.3
          2317.1   CFP      4    4         ASYNC     3818K   139.4       377.5
                   SUBCH   42   42         CHANGED       0   INCLUDED IN ASYNC
                                           UNSUCC        0     0.0         0.0
```

Check that all subchannels are functioning

USE should equal GEN

- ► Check connection

JdK

zSeries CF Links - Peer /Compat

zSeries	Connecting To	Subchannels	Can be shared by sender and receiver
PEER	zSeries	7	yes
COMPAT	non-zSeries	2	no

```
  z/OS V1R2                                    SUBCHANNEL  ACTIVITY
---------------------------------------------------------------------------
          # REQ                                ----------- REQUESTS -----------
 SYSTEM   TOTAL    -- CF LINKS --  PTH                  #   -SERVICE TIME(MIC)-
 NAME     AVG/SEC TYPE   GEN  USE  BUSY                REQ     AVG     STD_DEV
J80         3453K CBS      1    1   11K  SYNC          2987K   51.0      145.0
           1918.2 CFS      2    2        ASYNC        455143  210.7      797.4
                  SUBCH    6    6        CHANGED        1620  INCLUDED IN ASYNC
                                         UNSUCC            0    0.0        0.0
J90         2397K CBP      2    2     0  SYNC          1946K   45.1      141.9
           1331.8 ICP      4    4        ASYNC        448128  393.7       3129
                  SUBCH   42   42        CHANGED           0  INCLUDED IN ASYNC
```

JdK

Contention for CF Paths

CF paths can be shared (EMIFed by multiple MVS images on the same processor

```
                                                   SUBCHANNEL  ACTIVITY
-----------------------------------------------------------------------------
         # REQ                                   ----------- REQUESTS -------MIC)-
SYSTEM   TOTAL                    --BUSY--                #   -SERVICE TIME(M_DEV
NAME     AVG/SEC -- CONFIG --     -COUNTS-               REQ     AVG     STD_
JF0        51566 SCH GEN     8    PTH     0     SYNC    12167   161.1      72.3
            28.6 SCH USE     8    SCH     0     ASYNC   35607  1432.4      1680
                 SCH MAX     8                  CHANGED     0  INCLUDED IN ASYNC
                 PTH         4                  UNSUCC      0     0.0       0.0
JG0        1035K SCH GEN     8    PTH  2275     SYNC   808546   101.0      57.2
           575.0 SCH USE     6    SCH    65     ASYNC  221733   462.9     471.5
                 SCH MAX     6                  CHANGED    99  INCLUDED IN ASYNC
                 PTH         3                  UNSUCC      0     0.0       0.0
```

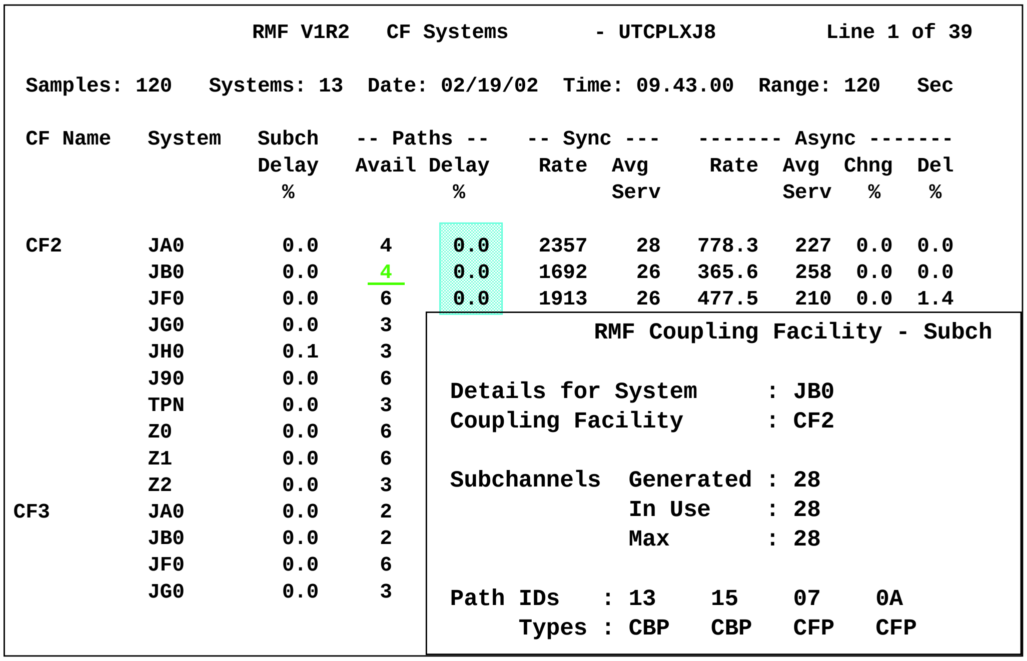
Guideline - less than 10% of requests encounter PTH BUSY

- ►Consider dedicating paths or additional paths

►Tune PTH BUSY first - may correct other conditions

JdK

RMF Mon III - CF Path Activity



```
                 RMF V1R2   CF Systems      - UTCPLXJ8       Line 1 of 39

Samples: 120   Systems: 13  Date: 02/19/02  Time: 09.43.00  Range: 120   Sec

CF Name   System   Subch   -- Paths --  -- Sync ---   ------- Async -------
                   Delay   Avail Delay  Rate  Avg     Rate  Avg  Chng  Del
                     %           %            Serv          Serv   %    %

CF2       JA0        0.0     4    0.0   2357   28    778.3   227  0.0  0.0
          JB0        0.0     4    0.0   1692   26    365.6   258  0.0  0.0
          JF0        0.0     6    0.0   1913   26    477.5   210  0.0  1.4
          JG0        0.0     3
          JH0        0.1     3
          J90        0.0     6
          TPN        0.0     3
          Z0         0.0     6
          Z1         0.0     6
          Z2         0.0     3
CF3       JA0        0.0     2
          JB0        0.0     2
          JF0        0.0     6
          JG0        0.0     3
```

```
                RMF Coupling Facility - Subch

Details for System       : JB0
Coupling Facility        : CF2

Subchannels   Generated  : 28
              In Use     : 28
              Max        : 28

Path IDs    : 13    15    07    0A
     Types  : CBP   CBP   CFP   CFP
```

SYNC changed to ASYNC

Long running SYNC CF requests use more CPU on sender.

Prior to z/OS1.2, XES changed some LIST/CACHE SYNC requests to ASYNC based on preset rules. Factors included

1. Request type
2. Sender and receiver processor type
3. Amount of data being sent

In z/OS 1.2, CF response time for SYNC requests is monitored for every request type and compared to threshold so all/only long requests (for whatever reason) are converted.

- Different thresholds for simplex/duplex and lock/non-lock are based on ASYCN pathlenght and normalized by processor type
- Thresholds are not externally adjustable
  - OW51813 for the latest threshold adjustment

SYNC -> ASYNC, cont.

Requests which are changed from SYNC to ASYNC based on the Heuristic Algorithm are counted as ASYNC

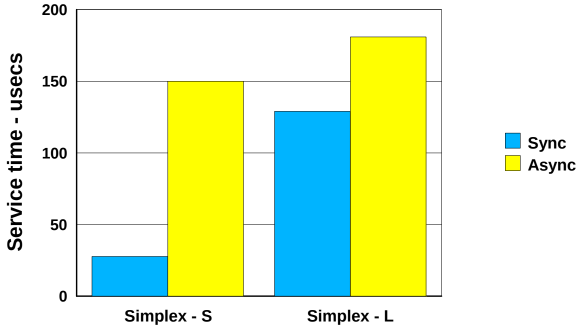
- not included in the CHNGD counts

```
STRUCTURE NAME = DSNDB1G_LOCK1     TYPE = LOCK   STATUS = ACTIVE
        # REQ    -------------- REQUESTS -------------    ------- DELAYED REQUESTS ------
 SYSTEM  TOTAL            #     % OF  -SERV TIME(MIC)-    REASON    #   % OF   AVG TIME(MIC)
 NAME    AVG/SEC         REQ     ALL     AVG    STD_DEV            REQ   REQ   /DEL   STD_DEV

 JA0       641K   SYNC    641K   20.8    33.0     132.7    NO SCH    4   0.0   21.8      4.9
        355.9     ASYNC     0     0.0     0.0       0.0    PR WT     0   0.0    0.0      0.0
                  CHNGD     0     0.0  INCLUDED IN ASYNC   PR CMP    0   0.0    0.0      0.0

 JE0      1073K   SYNC   1072K   34.8    34.5     134.0    NO SCH  114   0.0  104.2    241.9
        596.1     ASYNC    502    0.0   128.2     224.9    PR WT     0   0.0    0.0      0.0
                  CHNGD     0     0.0  INCLUDED IN ASYNC   PR CMP    0   0.0    0.0      0.0
```

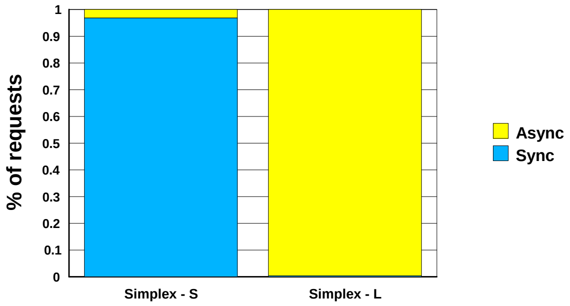
The decision is continuously reevaluated by allowing every nth SYNC request to be issued unchanged and comparing it with the thresholds.

DB2 Lock Structure - Long Link

Service times increase about 100 μsec - as expected for 10K links



For long links, most SYNC requests are converted to ASYNC



Value of Sync ► Async heuristic

New heuristic tries to limit the impact of

- DISTANCE
- Technology mismatch
- High CF utilization

- Benchmark results
  - CICS/DB2 data sharing workload
  - z900 host and CF technology

Distance between CFs	Cost of d.s. pre z/OS 1.2	Cost of d.s. z/OS 1.2
5 m	10%	10%
10 km	20%	14%

Misc Updates

1. CPENABLE recommendation - G5, G6, zSeries

- LPAR MVS images with shared CPs

CPENABLE(0,0)

- Improves I/O Response times in all cases
- Slight cost in response time

2. Performance improvements during system failure recovery and cleanup

- APAR **OW48624**
  - Only one system initiates cleanup
  - Confirmation process more efficient
  - CFRM I/O processing reduced for user sync point (IXLUSYNC) event processing.

Change to CF storage

CFLevel 12 - no distinction between Control storage and Data storage

```
                                              COUPLING  FACILITY  USAGE  SUMMARY
-------------------------------------------------------------------------------------------
 STORAGE SUMMARY  - CFLEVEL 11
-------------------------------------------------------------------------------------------
 TOTAL CF STORAGE SIZE                      6082M

  ...                                       ALLOC     % ALLOCATED
                                            SIZE

 TOTAL CONTROL STORAGE DEFINED              2027M       28.9
 TOTAL DATA STORAGE DEFINED                 4096M       49.6
```

```
 STORAGE SUMMARY  - CFLEVEL 12
-------------------------------------------------------------------------------------------

TOTAL CF STORAGE SIZE                      6082M

                                            ALLOC     % ALLOCATED
                                            SIZE

 TOTAL CONTROL STORAGE DEFINED              6082M       55.6
 TOTAL DATA STORAGE DEFINED                    0K        0.0
```

Sizing Structures

CFSizer on Parallel Sysplex Website
http://www.ibm.com/servers/eserver/pseries/pso

☐ XCF List Structure XCFHelp

# Systems	CLASSLEN
4	956

...

Click here to size structure

Structure Sizing Results

Function	Type	NAME	Size
XCF	LIST	IXC...	8704K
XCF	LIST	IXC1..	8704K

OW43778 - Handles size differences during rebuild

General concepts - structure size

- In CFRM policy can specify
  - INITSIZE and SIZE
  - If no INITSIZE, SIZE value is used
- If INITSIZE is specified,
  - Two attributes can be changed without a REBUILD
    1. Structure size - changed by command or IXLALTER
    2. Entry/Element ratio - changed by IXLALTER
  - Changing other attributes (like size of lock table, castout class, etc) requires a REBUILD
- Don't overestimate SIZE (see INFO APAR II10608)

For initial size estimate, use

- CF Structure Sizer on Parallel Sysplex website
- Parallel Sysplex Cookbook

Additional Information

- Websites www.s390.ibm.com/servers/eserver/zseries
  - Parallel sysplex (CF sizer, CFLevel description) .. /pso
  - RMF (tools, presentations, newsletters) .. /rmf
- WSC FLASHs
  - Flash10011 XCF Performance Considerations
  - Flash10159 New Heuristic Algorithm for CF Request Conversion
  - W99037 Performance Impacts of Using Shared ICF CPs
- Publications
  - Setting up a Sysplex (SA22-7625-06)
  - z/Series 900 System Overview (SA22-1027-03b)
  - z/Series 990 System Overview (SA22-1032-00a)
  - Processor Resource/System Manager Planning Guide (SB10-7036-01)