How to Relocate a Massive Sysplex with Minimal Service Disruption

A New Datacenter Move Success Story

Carles Arís
itnow/CaixaBank
caris@silk.es

SHARE Session: 12121

August 2012

Agenda

❑ Introduction
❑ The Move Concept
❑ GDPS was Key for the Move
❑ Input/Output, Connectivity & Other zOS Issues
❑ Disk Replication
❑ FICON Directors Fabrics
❑ Network & Coms Server
❑ Questions & Comments

Introduction

CaixaBank Background

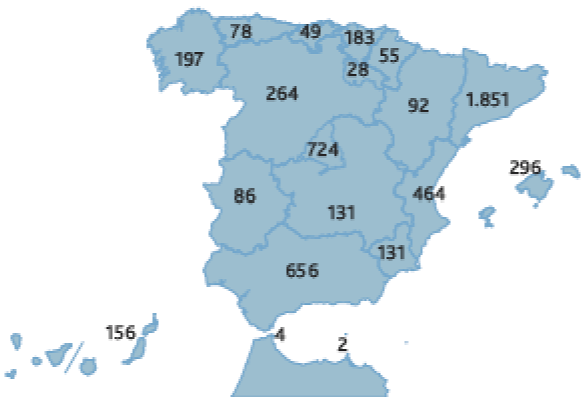
- Established in 1990 as a merger of the first and third largest savings and loan banks in Spain, originally founded in 1844 and 1904. It became a commercial bank in 2011.
- Biggest bank in Spain and one of the largest in Europe. Headquarters located in Barcelona.
- Universal banking model where deposits and mortgage lending are still the core business.
- Main share holder is la Caixa. Non profit social-financial institution, privately managed by laws approved by the regional government of Catalonia.
- An important part of the total net income is given to ”la Caixa” foundation, which devotes its budget to social, educational, scientific and cultural projects.

Some Business Figures

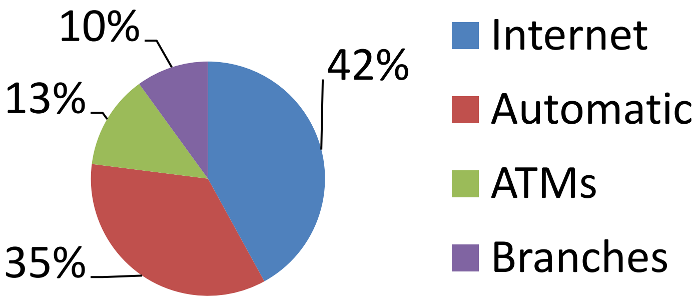
Big numbers

- 28,000 employees
- 10.5 million customers
- More than 5,200 branch offices
- Close to 8,000 ATMs
- 6 million online banking customers
- 4 billion operations per year

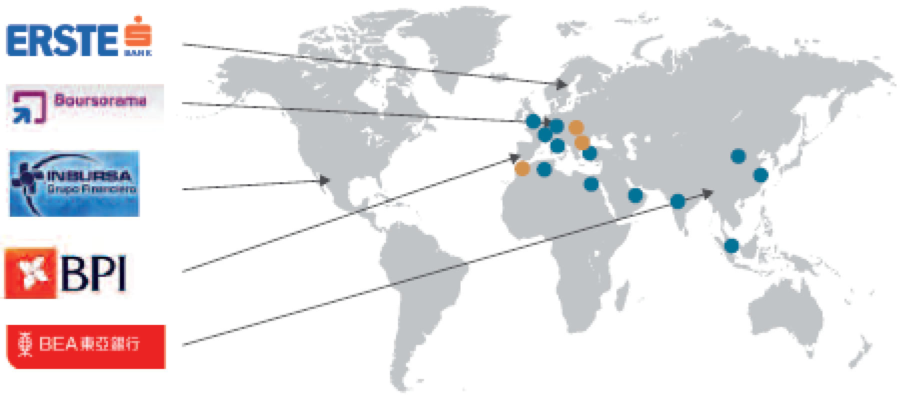
Branches in 2011



Operations



International



● ● Representation and Branch Offices, respectively

IMS Transaction Volumes (Peak Hour)

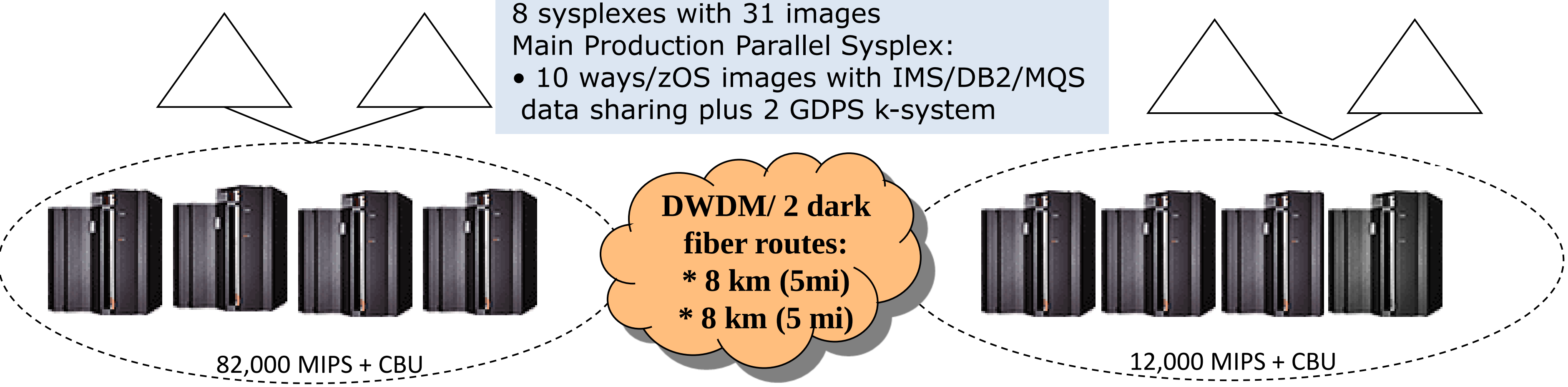
Business Channel	April 2012
Branch Offices	791 Tran / sec.
ATMs	111 Tran / sec.
Home banking transactions	608 Tran / sec.
POS transactions	20 Tran / sec.
Total arrival transactions	1,632 Tran / sec.
Total processed transactions	2,114 Tran / sec.

Mainframe Infrastructure

DC1: Cerdanyola (new DC!)

DC2: Sant Cugat

8 sysplexes with 31 images
Main Production Parallel Sysplex:
• 10 ways/zOS images with IMS/DB2/MQS data sharing plus 2 GDPS k-system



DWDM/ 2 dark fiber routes:
* 8 km (5mi)
* 8 km (5 mi)

82,000 MIPS + CBU

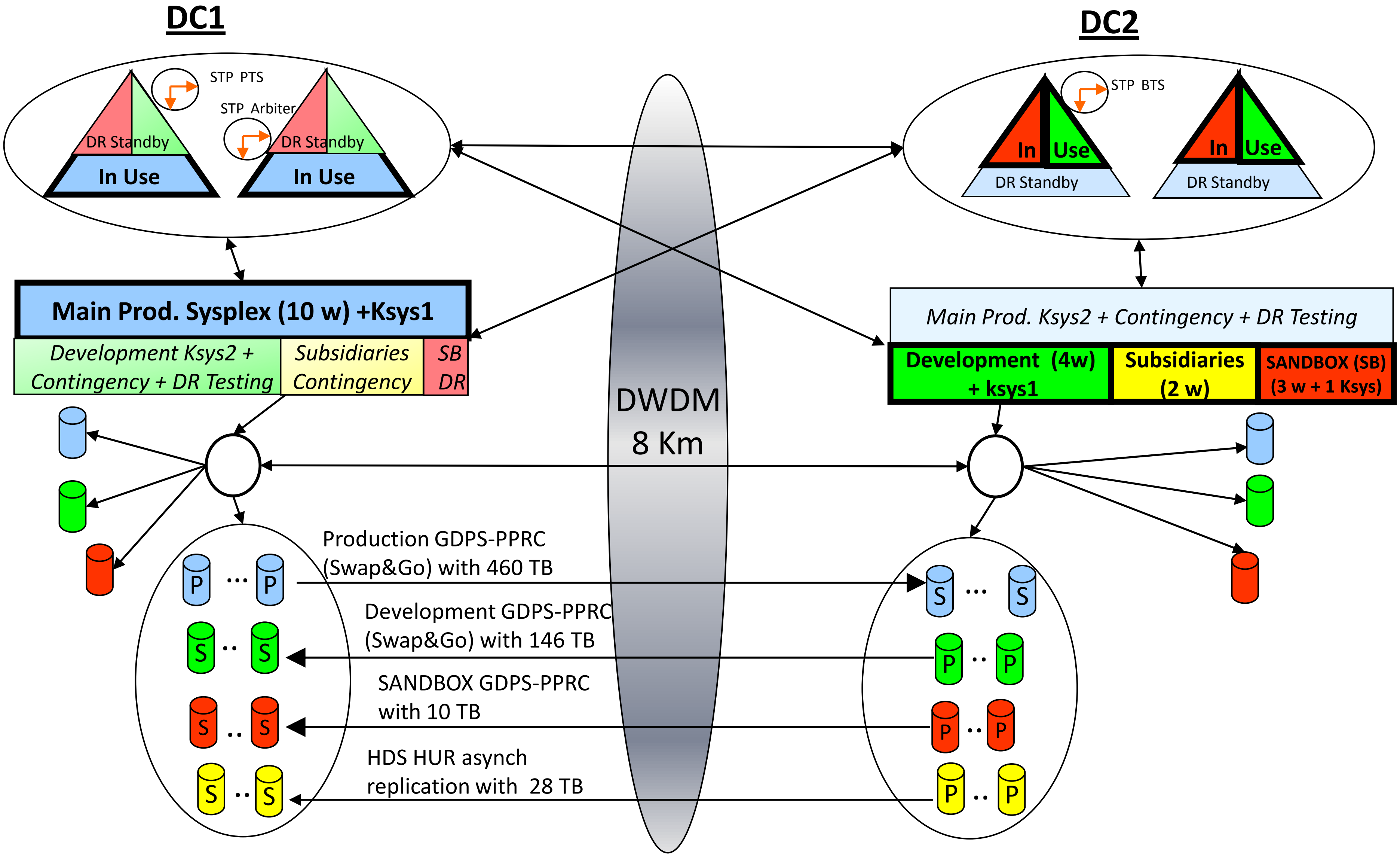
12,000 MIPS + CBU

- Main Production Parallel Sysplex plus Subsidiaries and Development Disaster Recovery
  - 4 x z196 zOS CECs
  - 2 x z196 Stand Alone Coupling Facilities
- WAS environment on zOS.
  - Old Terminal Branches application
  - 1 x z9 zOS CEC

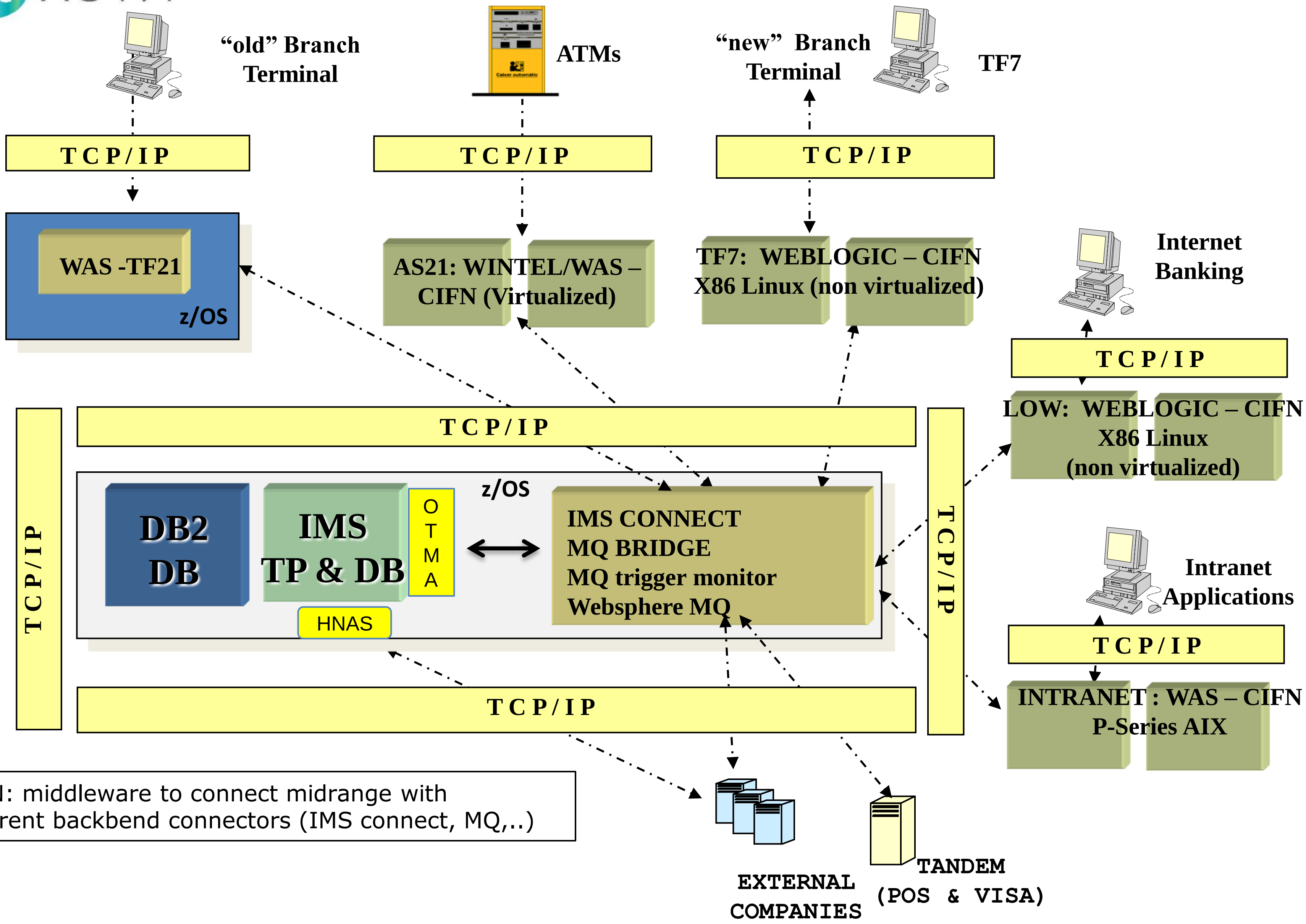
- Development and Subsidiaries Sysplexes plus Main Production Disaster Recovery (GDPS)
  - 3 x z10 & 1 x 196 zOS CECs
  - 2 x z196 Stand Alone Coupling Facilities
- WAS environment on zOS.
  - Old Terminal Branches application.
  - 1 x z9 zOS CEC

➢DASD Mainframe in HDS with 1.4 PB
➢TAPES Mainframe in IBM robotics & VTS 7740 (18 grids)
➢FICON DIRECTORS: CISCO and Brocade

Mainframe Main Picture


DC1
DC2
STP PTS
STP Arbiter
DR Standby
In Use
DR Standby
In Use
STP BTS
In Use
DR Standby
In Use
DR Standby
Main Prod. Sysplex (10 w) +Ksys1
Development Ksys2 + Contingency + DR Testing
Subsidiaries Contingency
SB DR
Main Prod. Ksys2 + Contingency + DR Testing
Development (4w) + ksys1
Subsidiaries (2 w)
SANDBOX (SB) (3 w + 1 Ksys)
DWDM
8 Km
P … P
S .. S
S .. S
S .. S
S … S
P .. P
P .. P
P .. P
Production GDPS-PPRC (Swap&Go) with 460 TB
Development GDPS-PPRC (Swap&Go) with 146 TB
SANDBOX GDPS-PPRC with 10 TB
HDS HUR asynch replication with 28 TB

Business Channels: Global Picture



CIFN: middleware to connect midrange with different backbend connectors (IMS connect, MQ,..)

The Move

General Context

Requirements

- No service disruption
- Minimize time without contingency (weekend)
- Use proven technology (GDPS, Hyperswap, Multisite Sysplex,...)

To take into account

- Minimize costs
- Avoid changes and new technology throughout the move (as much as possible)

HW Strategy

- HW cloning for zOS CECs, CFs, DASD and Ficon Directors (FD)
    - 3rd DASD copy
- Move (reuse) TAPES (VTS) with their FDs
- Try to use the same IODF!

Global Move Concept

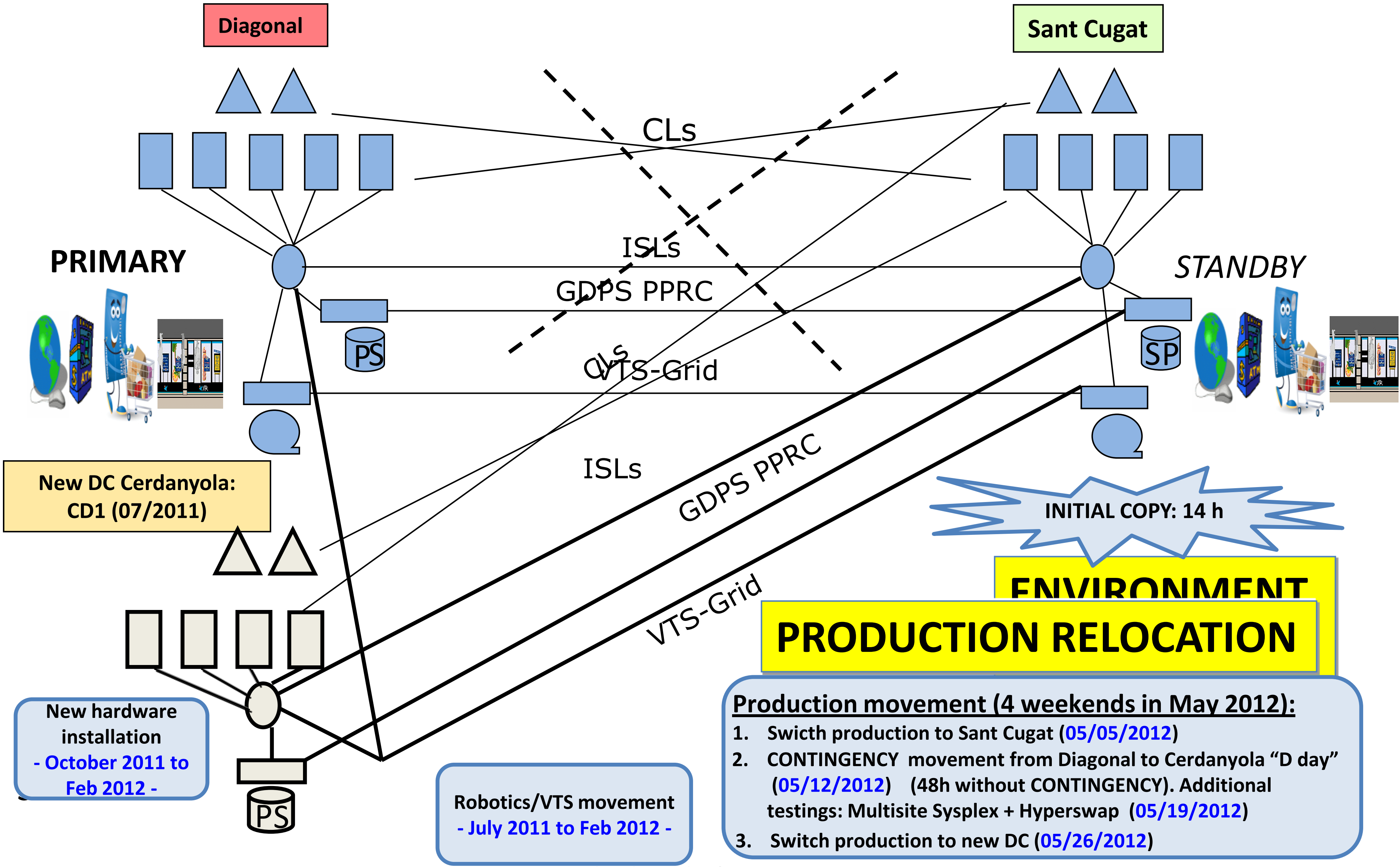
Basic rules

- Move one sysplex at a time, from less to more critical
- Always move the contingency part
  - Whenever necessary perform a site switch before the move
- GDPS/PPRC procedures were used
  - 3rd disk copy in new DC became PPRC secondary (Initial Copy)

Move Strategy

- Populate new DC with necessary HW
- Try it in an isolated way
- Leave old DC to be moved with only contingency (standby) functions
  - Perform as many sysplex site switches as necessary
- On weekends, D days, move contingency to new DC and verify it
  - It's necessary to perform a PPRC initial copy
- Some time later, once new DC fully verified, perform site switches to it

Move: Visual Approach



New DC Cerdanyola: CD1 (07/2011)

New hardware installation
- October 2011 to Feb 2012 -

Robotics/VTS movement
- July 2011 to Feb 2012 -

INITIAL COPY: 14 h

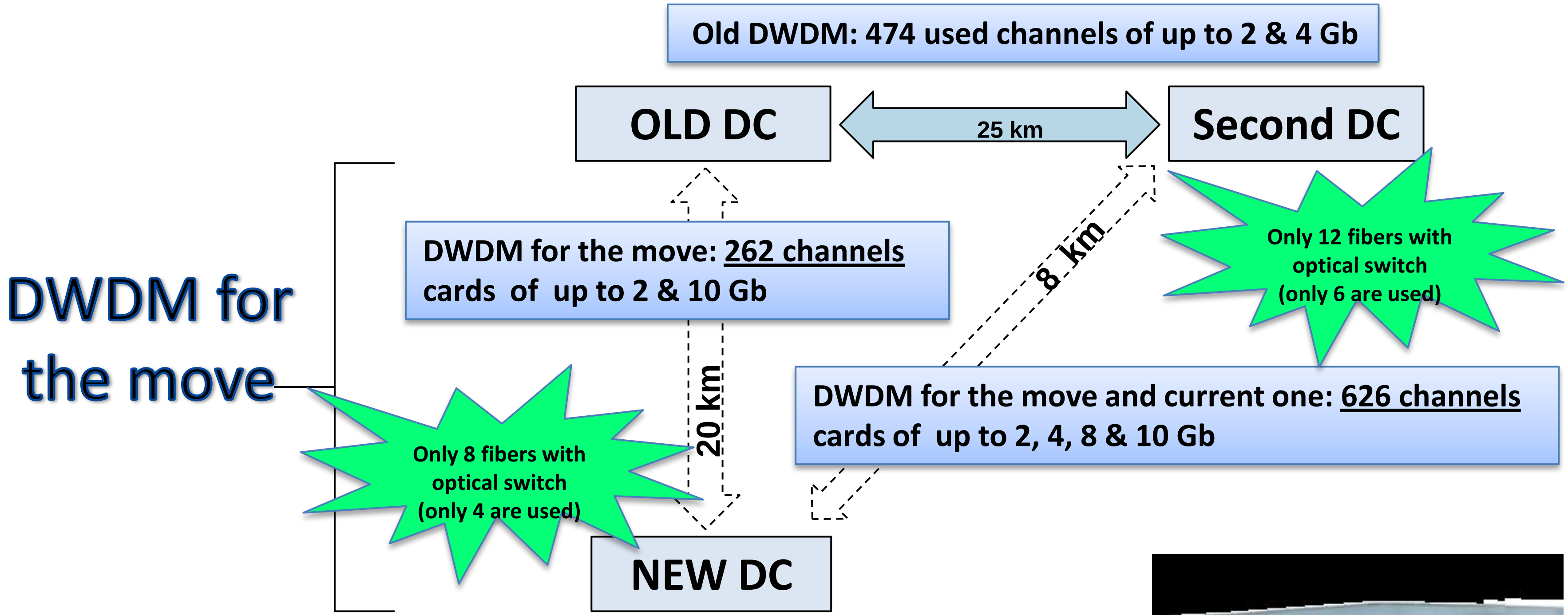
ENVIRONMENT

PRODUCTION RELOCATION

Production movement (4 weekends in May 2012):

1. Swicth production to Sant Cugat (05/05/2012)
2. CONTINGENCY movement from Diagonal to Cerdanyola “D day” (05/12/2012) (48h without CONTINGENCY). Additional testings: Multisite Sysplex + Hyperswap (05/19/2012)
3. Switch production to new DC (05/26/2012)

DWDM Infrastructure



Old DWDM: 474 used channels of up to 2 & 4 Gb

DWDM for the move

DWDM for the move: <u>262 channels</u> cards of up to 2 & 10 Gb

DWDM for the move and current one: <u>626 channels</u> cards of up to 2, 4, 8 & 10 Gb

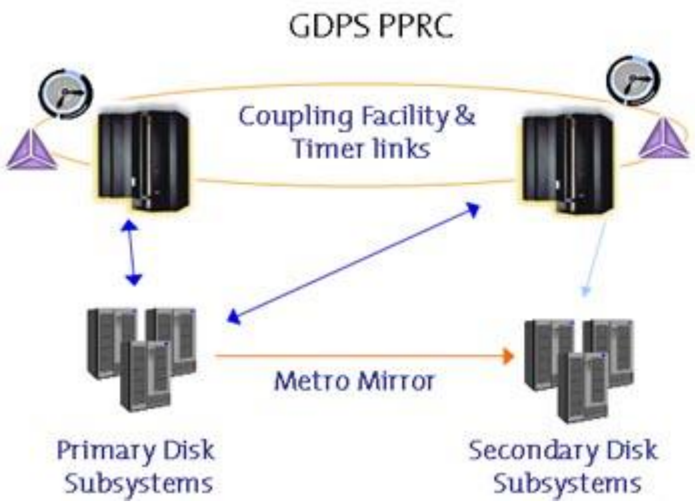
Only 12 fibers with optical switch (only 6 are used)

Only 8 fibers with optical switch (only 4 are used)

- Up to 96 user channels in a single dark fiber (C and L bands)
- Optical switches to have dark fiber high availability
- DWDM cards can operate at different speeds and protocols
  - 1 GbE and 10 GbE
  - 2, 4 and 8 Gb FC

FiberNet

GDPS Was Key for the Move



GDPS Technology Used for the Move

- **GDPS** is a powerful DR solution and it **can also help for a new DC move**
- In May, over four weekends, we moved the main production Sysplex using GDPS procedures

May 5th & 6th: First site switch

- Site switch from old DC to be moved to second DC. **Done with GDPS**

May 12th & 13th: Contingency relocation to new DC

- Contingency is relocated from old DC to new DC
- **GDPS is used for**:
  - **Multisite** IPLs to new DC to try it
  - Planned **Hyperswap** to try new DC disks

May 19th & 20th: Carefully testing new DC

- We tried new DC infrastructure. **GDPS is used for:**
  - IPL critical Systems in new DC to try them (**Multisite**)
  - **Unplanned Hyperswap** (to settle SWAP&GO policy)

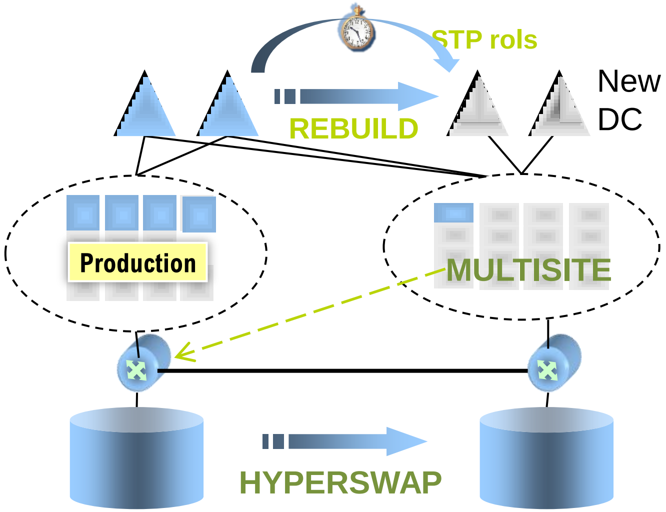
May 26th: Final site switch to new DC

- Final site switch from second DC to new DC. **Done with GDPS**

GDPS sessions: 11661-GDPS 3.9 Update and 11663-GDPS Active/Active Sites Update

May 2012

Mo	Tu	We	Th	Fr	Sa	Su
	1	2	3	4	**5**	6
7	8	9	10	11	**12**	**13**
14	15	16	17	18	**19**	**20**
21	22	23	24	25	**26**	27
28	29	30	31			



10,607 PPRC pairs in 61 LCUs

Input/Output, Connectivity & Other zOS Issues

Cabling

Deploying a new DC is really hard in terms of data cabling preparation.
We had to settle all this connectivity:

Kind of Connection	New Connections
Channels to new DC CECs	616
Channels with old DCs (DWDM)	824
Robotics and VTS move	624
Switch and FD to CUs	586
Network (OSAs, VTS grid,...)	150
TOTAL	**2,800**

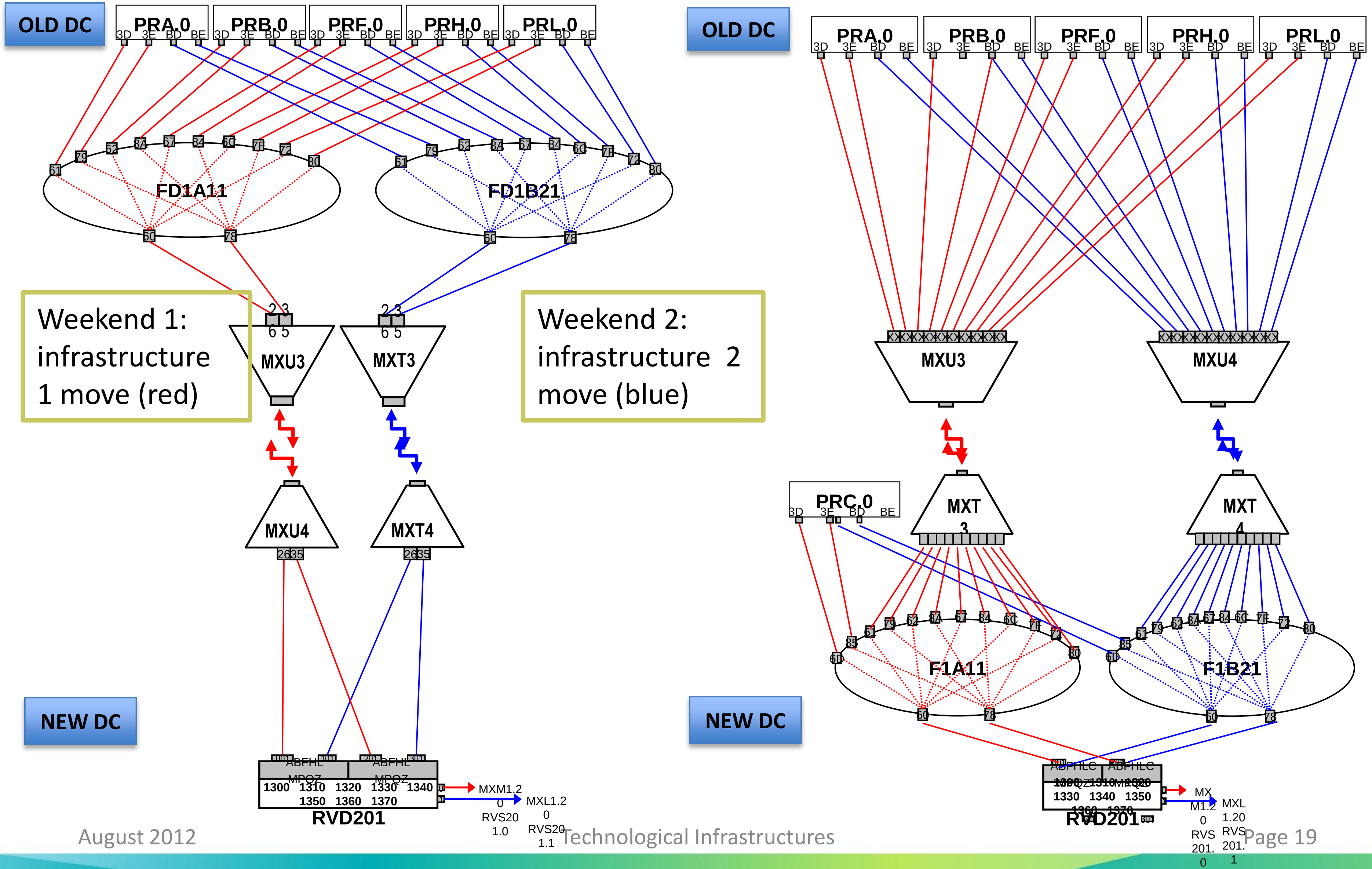
Our experience...

- Cabling people measure attenuation levels → this doesn't prevent you from getting IOerrors
- In total we had 133 incidents with new connectivity (bad cables, lasers, connectors. and even channel cards). Some of them took weeks to be fixed
- We had to use a starter system to try new connections as much as possible. We also became familiar with HMC toggle off/on to reset channels → time consuming task!
  - We had to schedule 90 windows of 4-6 hours each to try new connectivity

Cabling: One Example

FICON DIRECTORS TO BE MOVED FOR VTS

NEW CABLING....



Weekend 1: infrastructure 1 move (red)

Weekend 2: infrastructure 2 move (blue)

Avoiding Q4 Contention

As we did the move, we had to bring many channels offline. Usually, they weren't the last ones going to a control unit/device. However, the system gets Q4 exclusively. This can be really problematic when there is heavy activity on the system (Batch and Tapes/VTS allocation).

Procedure to avoid Q4 in exclusive when configuring channels offline

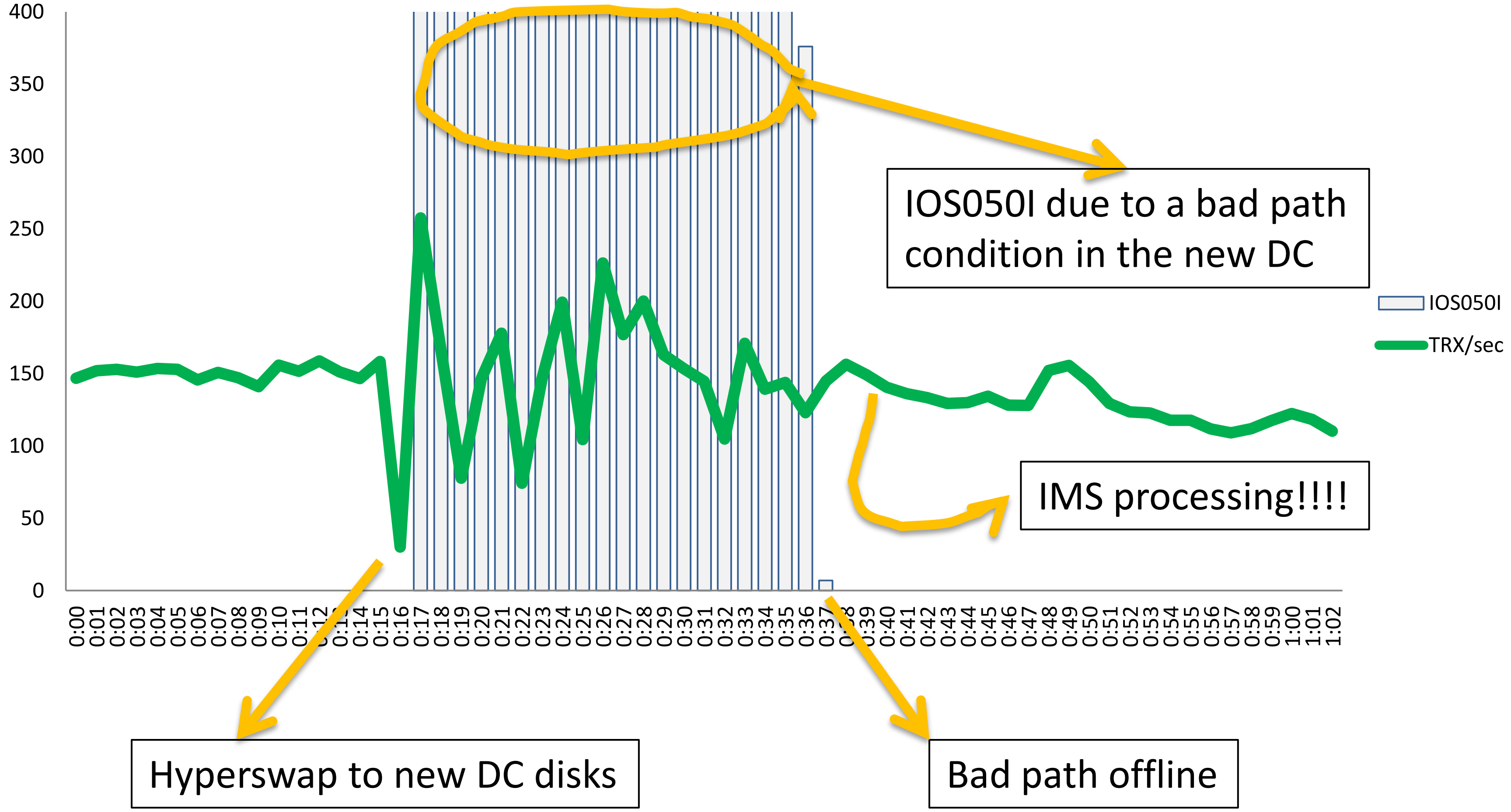
1 V PATH(dddd, cc),offline for every device on the CHPID to be brought offline
    - On z/OS 1.12 you could issue VARY CU(dddd),PP(cc),OFFLINE for every CU on the CHPID
    - No Q4 obtained with OA15006 on vary path offline.

2. Then Issue CF CHP(cc),OFFLINE,FORCE
    - Since the path is already offline, no risk in terminating an active I/O request (could be a problem for XES and JES2 channel programs recovery)
    - With FORCE option Q4 isn't obtained either
    - There is a confirmation WTO that you will need to automate the answer to "YES"

I/O Recovery

OLTP systems are extremely sensitive to I/O recovery

IMS Processed Transactions/sec versus IOS050I message count



IOS050I due to a bad path condition in the new DC

IMS processing!!!!

Hyperswap to new DC disks

Bad path offline

How to Mitigate I/O Recovery Impact

- Limited recovery time
  - Look into OA22573
  - Speed up I/O recovery
  - Ability to reduce I/O recovery timeout from 15 to 2 seconds for Disk
    - RECOVERY,LIMITED_RECTIME=2,DEV=DASD
  - It's possible that fewer retries are done
  - Doesn't work for IOS050I and I/O recovery continues being at device level (really slow!)

- Path recovery improvements in 1.13
  - Look into HOT TOPICS issue 25 pg. 44 article *(The path to recovery: Realizing improved channel path recovery)*
  - If in a given interval (PATH_INTERVAL) you get a certain number of IOerrors (PATH_THRESHOLD) in a bad path, you can bring it offline for all CUs using it (PATH_SCOPE=CU)
  - It covers also IOS050I

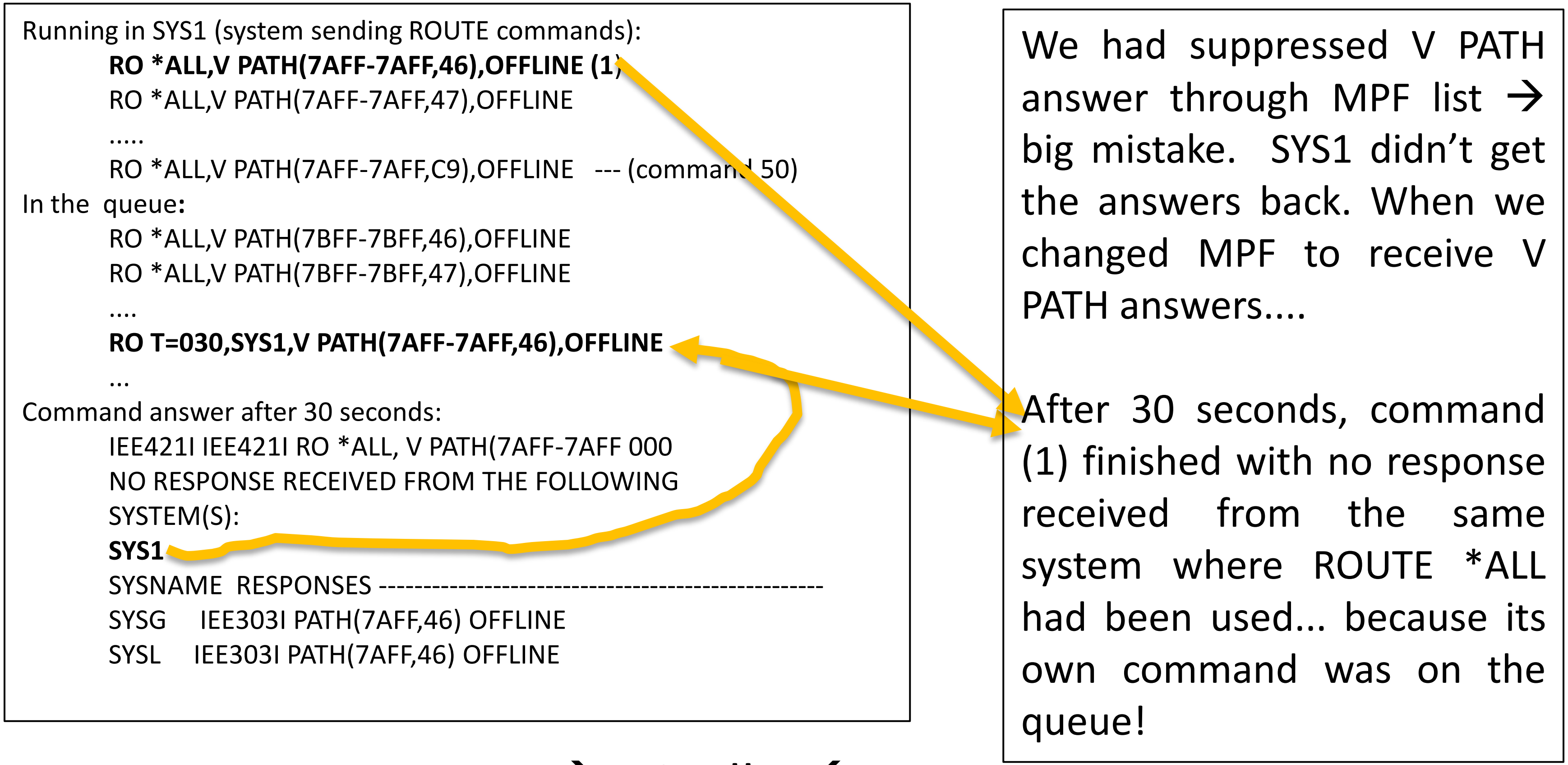
ROUTE *ALL (1)

Don’t use ROUTE *ALL commands to send hundreds of V PATH OFFLINE/ONLINE and/or V ONLINE/OFFLINE to all the systems in the sysplex
To move to new DC disks we had to execute 528 V PATH OFFLINE commands in 11 systems and we tried ROUTE *ALL V PATH OFFLINE…

```
From  SYS1 - ROUTE *ALL,V PATH(E200-E2D8,E3),OFFLINE
        RO T=030,SYS1,V PATH(E200-E2D8,E3),OFFLINE
        RO T=030,SYST,V PATH(E200-E2D8,E3),OFFLINE
        RO T=030,SYSN,V PATH(E200-E2D8,E3),OFFLINE
        RO T=030,SYSX,V PATH(E200-E2D8,E3),OFFLINE
        .......
```

- By default only 50 commands can be executed simultaneously, and then are queued up (class C3)
- In reality, it seems that only one command is executed at a time due to SYSZMCS serialization used by ROUTE command
- Look into “OA11161: WHEN MULTIPLE ROUTE *ALL ARE ISSUED THE RESPONSES ARE DELAYED” for additional information

ROUTE *ALL (2)



Running in SYS1 (system sending ROUTE commands):

```
        RO *ALL,V PATH(7AFF-7AFF,46),OFFLINE (1)
        RO *ALL,V PATH(7AFF-7AFF,47),OFFLINE
        .....
        RO *ALL,V PATH(7AFF-7AFF,C9),OFFLINE   --- (command 50)
In the  queue:
        RO *ALL,V PATH(7BFF-7BFF,46),OFFLINE
        RO *ALL,V PATH(7BFF-7BFF,47),OFFLINE
        ....
        RO T=030,SYS1,V PATH(7AFF-7AFF,46),OFFLINE
        ...
Command answer after 30 seconds:
        IEE421I IEE421I RO *ALL, V PATH(7AFF-7AFF 000
        NO RESPONSE RECEIVED FROM THE FOLLOWING
        SYSTEM(S):
        SYS1
        SYSNAME  RESPONSES ------------------------------------------------
        SYSG     IEE303I PATH(7AFF,46) OFFLINE
        SYSL     IEE303I PATH(7AFF,46) OFFLINE
```

We had suppressed V PATH answer through MPF list → big mistake. SYS1 didn't get the answers back. When we changed MPF to receive V PATH answers....

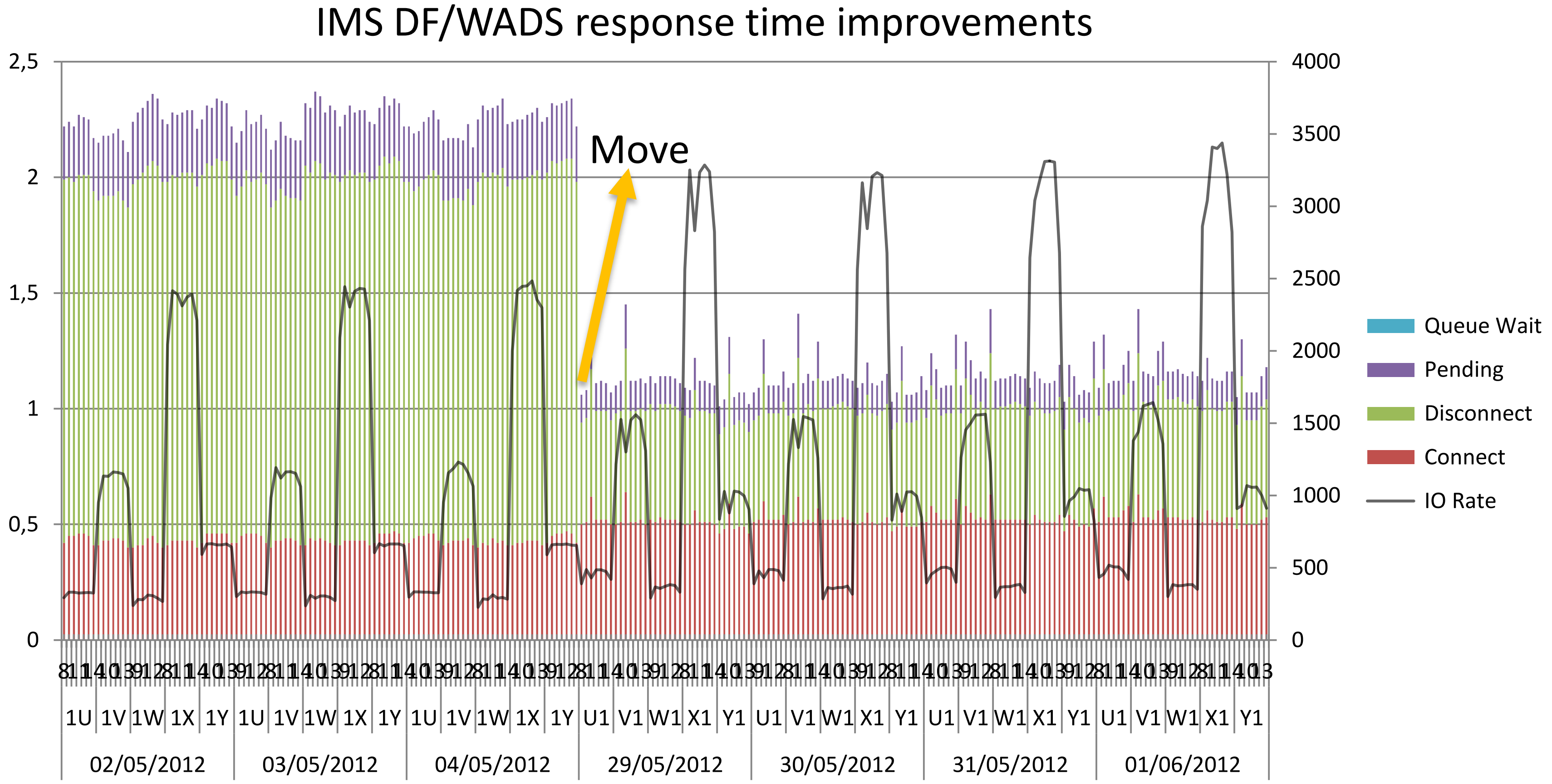
After 30 seconds, command (1) finished with no response received from the same system where ROUTE *ALL had been used... because its own command was on the queue!

→ Finally ←

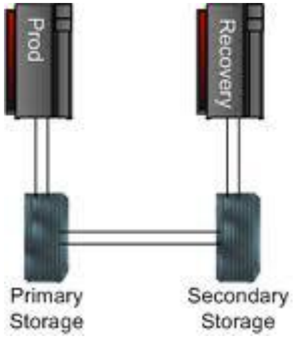
- All commands on the queue were removed CMDS REMOVE,CLASS=C3
- We waited patiently for the 50 ones running to finish and sent V PATH commands directly to every system on the sysplex

Disk Response Time Improvements

- Distance between two DCs is shorter than before (8Km vs 26km) → less impact on disk replication
- New Disk Technology in New DC



Disk Replication



Initial Copy

For every sysplex to be moved, we did two initial copies against the new DC disks:

1) to test them
2) to do the move

We hit several issues:

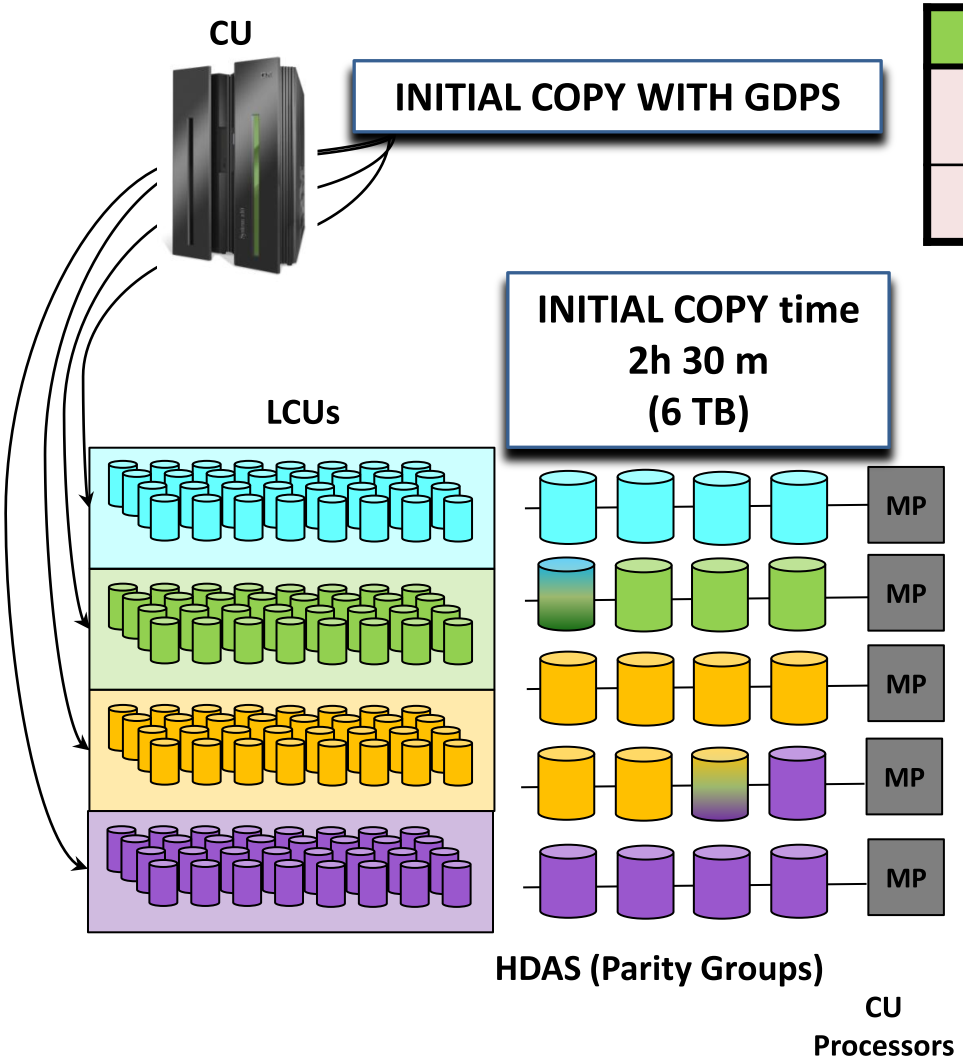
- Initial Copy procedure efficiency
- Performance impact when starting initial copy
- Initial Copy elongation due to
  - PPRC flapping links
  - Fabric for replication congestion

Initial Copy Efficiency (1)

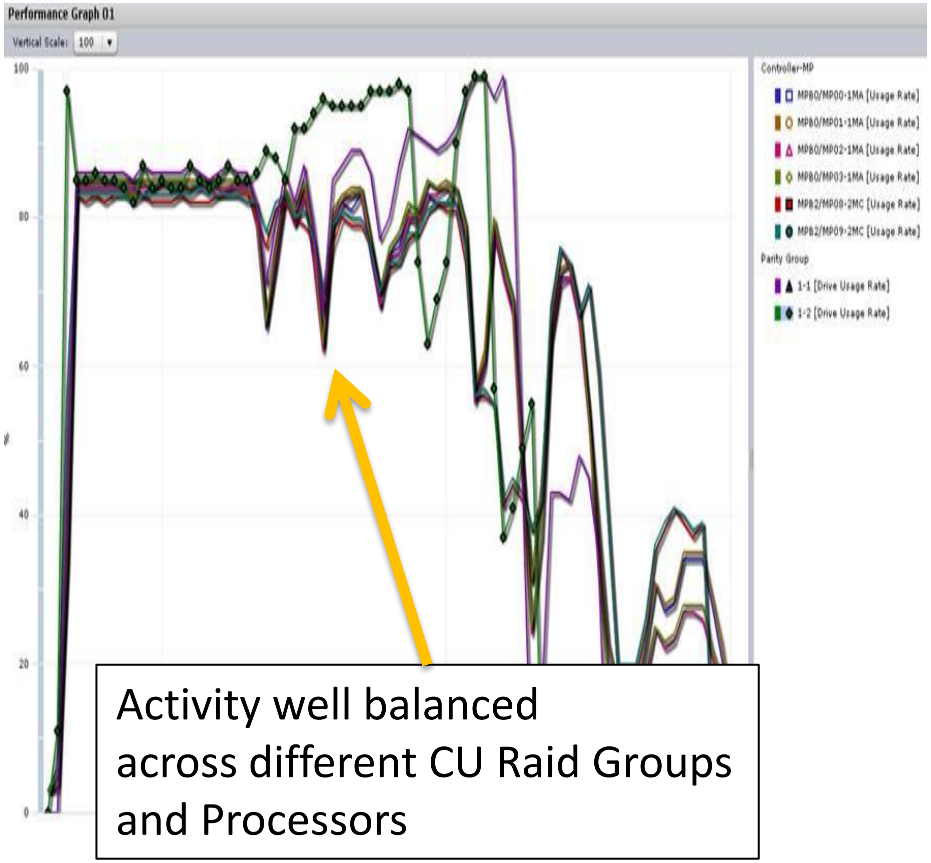
Disk units that we use can run concurrently, depending on internal settings, up to 64 initial copy processes (64 volumes being copied at the same time).

- GDPS performs the initial copy in all the LCUs at the same time. The 64 concurrent initial copies are spread reasonably well along the physical configuration of the CU (Raid Groups). So, initial copy performance is maximized. However, we couldn't use GDPS to start initial copy because we were performing the move and the necessary cross site connectivity wasn't available yet
- We planned to use our self made procedure to start initial copy. We tried one task per CU. However, the 64 initial copy processes belonged to a few Raid Groups which collapsed, and the overall initial copy process was elongated
- To avoid this issue we improved our self made procedure to execute all the initial copy commands, CESTPAIR, on each CU taking into account how the volumes were spread in the different Raid Groups
  - ❑ With this improvement we could reduce Initial Copy time to be even lower than the one that we get with GDPS

Initial Copy Efficiency (2)

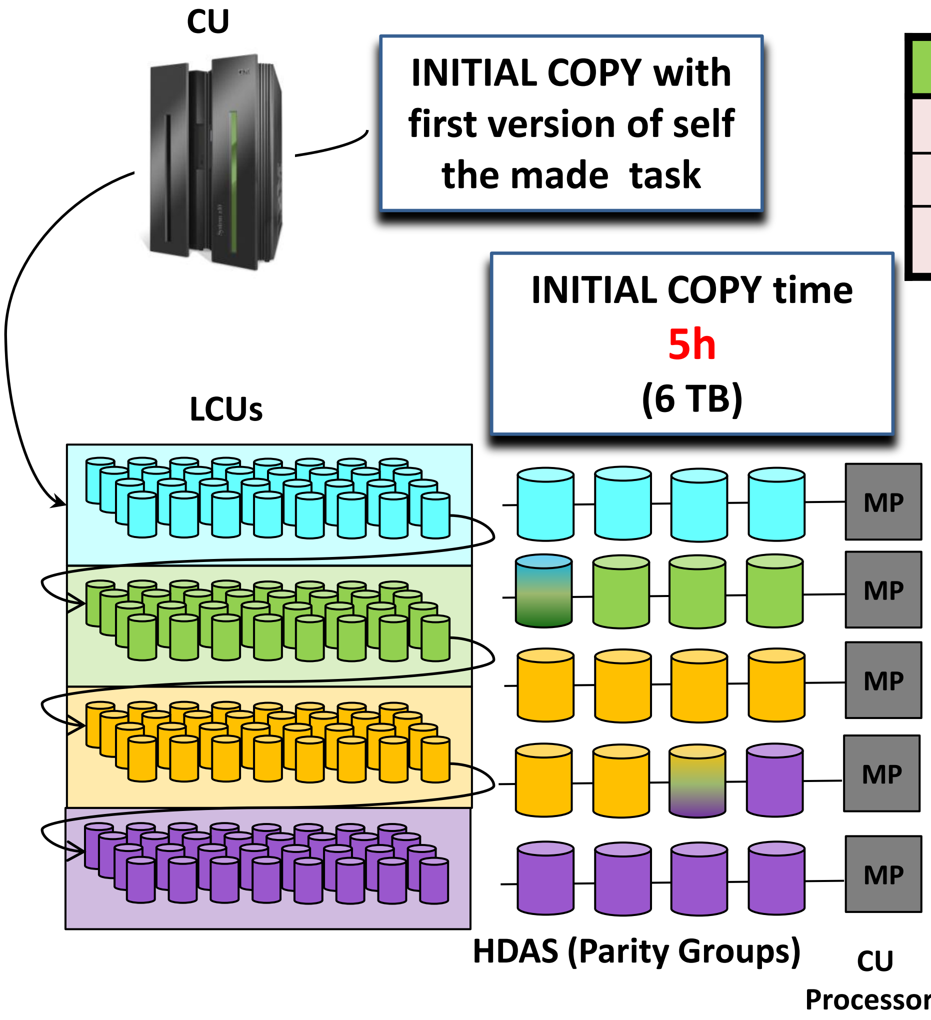


INITIAL COPY with GDPS
We couldn't use it for the move because cross site connectivity wasn't available yet
Commands are executed in parallel at LCU level



Activity well balanced across different CU Raid Groups and Processors

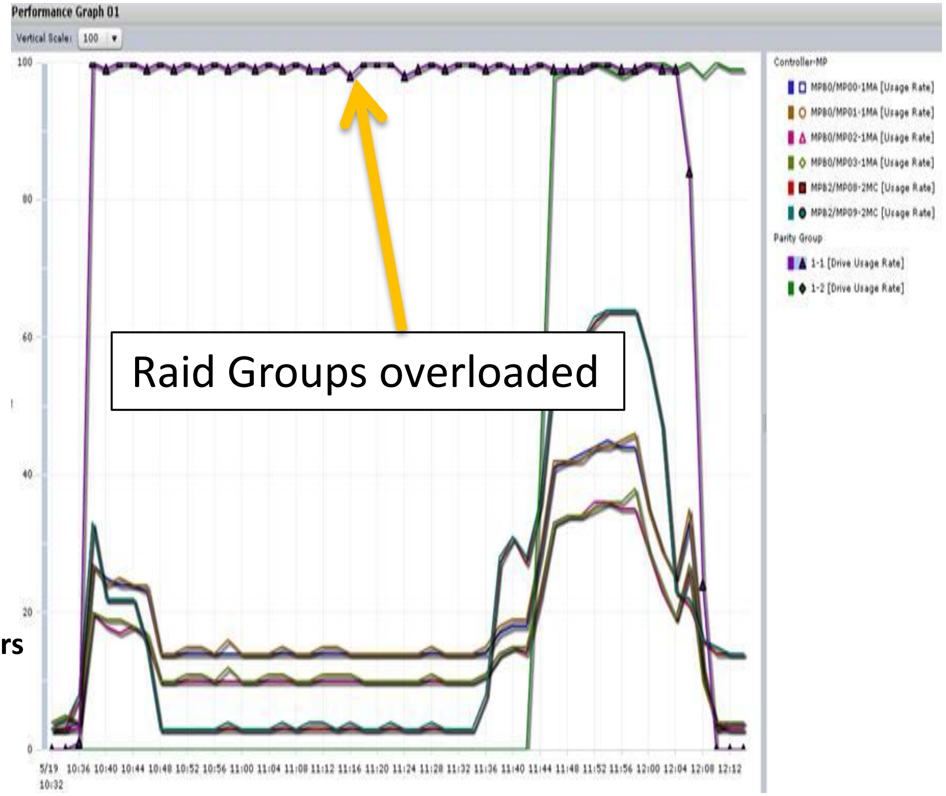
Initial Copy Efficiency (3)



INITIAL COPY with first version of self the made task

INITIAL COPY time
5h
(6 TB)

INITIAL COPY with self made task version 1
It doesn't need cross site connectivity for zOS (only PPRC links)
1 STC per CU
Commands are executed in sequence from first device to last device in CU.

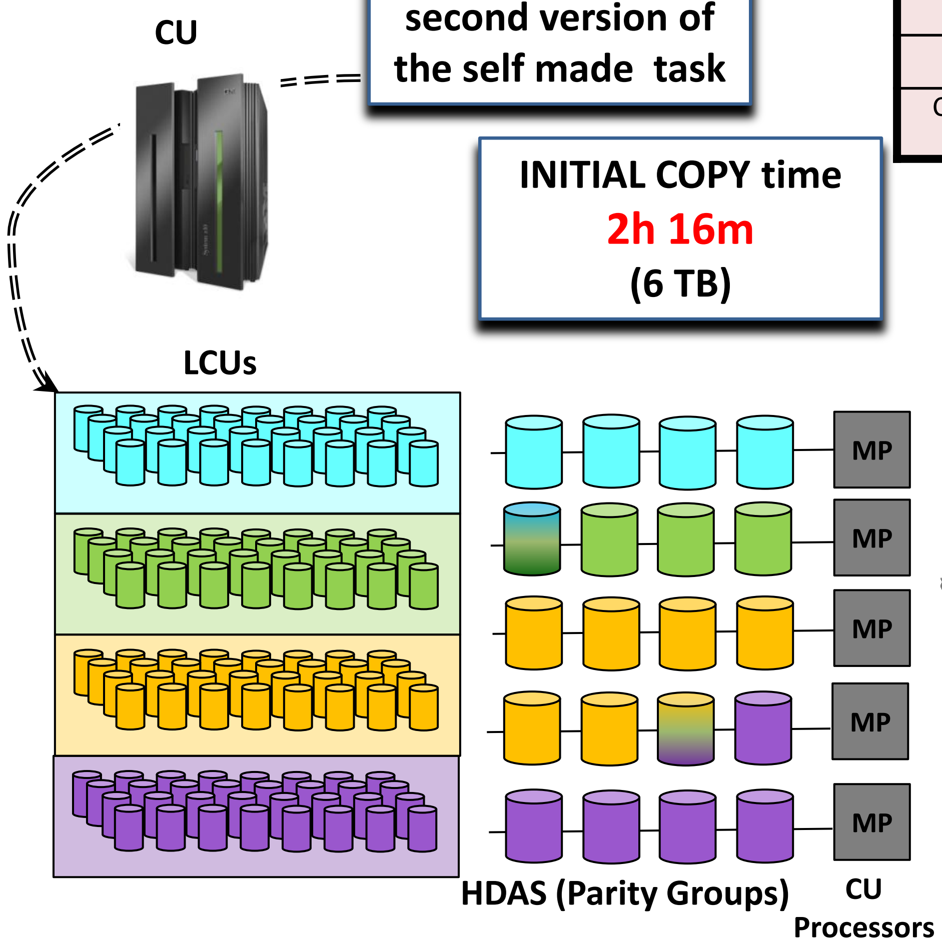


Raid Groups overloaded

Initial Copy Efficiency (4)

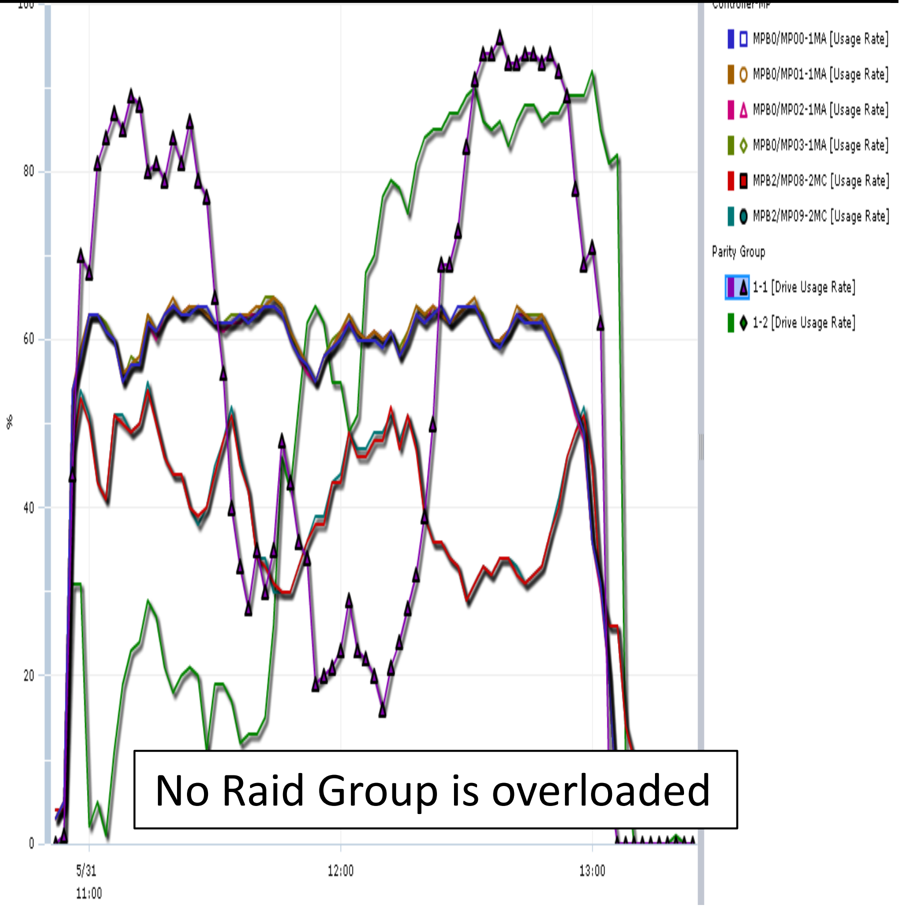
INITIAL COPY with second version of the self made task

INITIAL COPY time 2h 16m (6 TB)



14 minutes less than with GDPS!!!

INITIAL COPY with self made task version 1
It doesn't need cross site connectivity for zOS (only PPRC links)
1 STC per **CU**
Commands are executed in an order that takes into account the physical distribution inside the CU



No Raid Group is overloaded

Initial Copy Efficiency (5)

CESTPAIR COMMANDS sorted by raid group

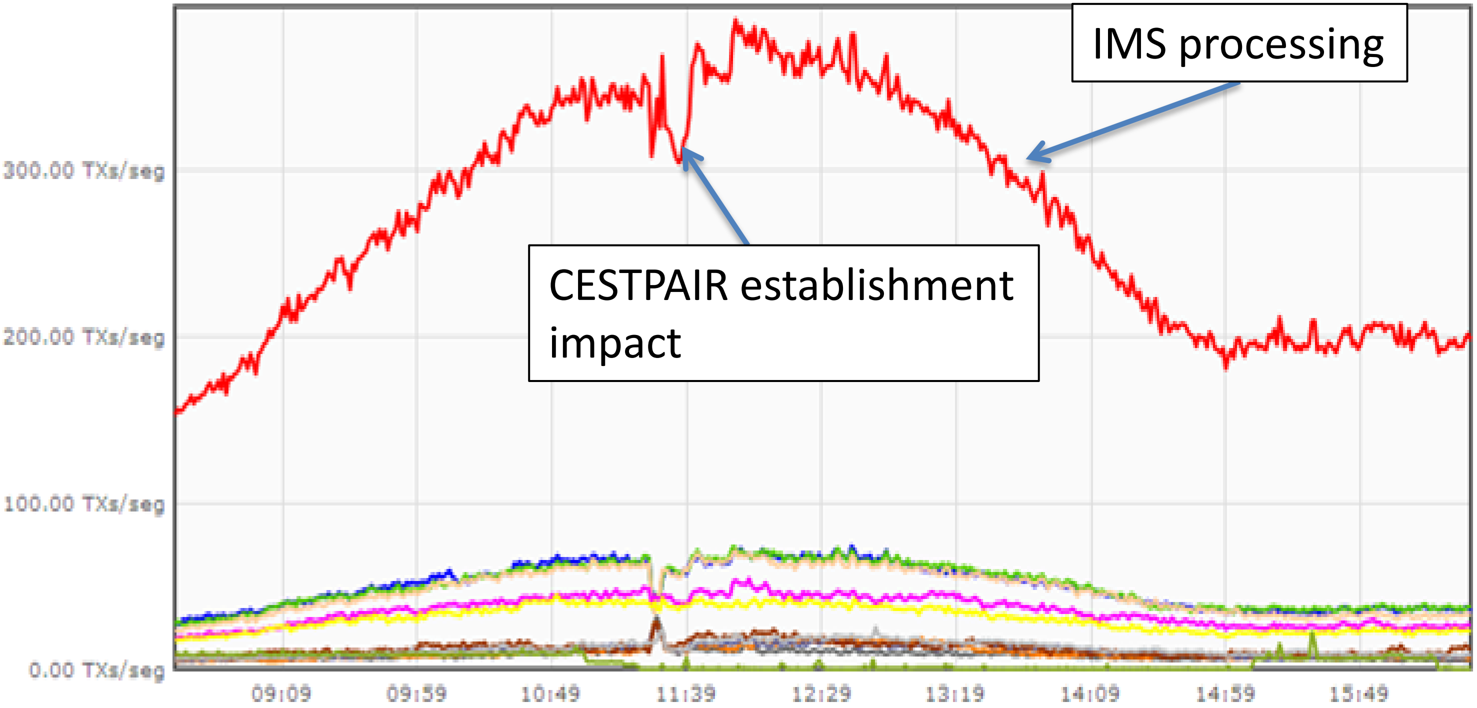
```
CESTPAIR DEVN(X'2201')  PRIM(X'12B0' 62111 X'01' X'00')  SEC(X'92B0' 22359 X'01' X'00') MODE(COPY) CRIT(NO) PACE(1)
CESTPAIR DEVN(X'2210')  PRIM(X'12B0' 62111 X'10' X'00')  SEC(X'92B0' 22359 X'10' X'00') MODE(COPY) CRIT(NO) PACE(1)
CESTPAIR DEVN(X'2222')  PRIM(X'12B0' 62111 X'22' X'00')  SEC(X'92B0' 22359 X'22' X'00') MODE(COPY) CRIT(NO) PACE(1)
CESTPAIR DEVN(X'2234')  PRIM(X'12B0' 62111 X'34' X'00')  SEC(X'92B0' 22359 X'34' X'00') MODE(COPY) CRIT(NO) PACE(1)
CESTPAIR DEVN(X'2246')  PRIM(X'12B0' 62111 X'46' X'00')  SEC(X'92B0' 22359 X'46' X'00') MODE(COPY) CRIT(NO) PACE(1)
CESTPAIR DEVN(X'2208')  PRIM(X'12B0' 62111 X'08' X'00')  SEC(X'92B0' 22359 X'08' X'00') MODE(COPY) CRIT(NO) PACE(1)
CESTPAIR DEVN(X'2219')  PRIM(X'12B0' 62111 X'19' X'00')  SEC(X'92B0' 22359 X'19' X'00') MODE(COPY) CRIT(NO) PACE(1)
CESTPAIR DEVN(X'222B')  PRIM(X'12B0' 62111 X'2B' X'00')  SEC(X'92B0' 22359 X'2B' X'00') MODE(COPY) CRIT(NO) PACE(1)
CESTPAIR DEVN(X'223D')  PRIM(X'12B0' 62111 X'3D' X'00')  SEC(X'92B0' 22359 X'3D' X'00') MODE(COPY) CRIT(NO) PACE(1)
CESTPAIR DEVN(X'2300')  PRIM(X'12B1' 62111 X'00' X'01')  SEC(X'92B1' 22359 X'00' X'01') MODE(COPY) CRIT(NO) PACE(1)
CESTPAIR DEVN(X'2310')  PRIM(X'12B1' 62111 X'10' X'01')  SEC(X'92B1' 22359 X'10' X'01') MODE(COPY) CRIT(NO) PACE(1)
CESTPAIR DEVN(X'2322')  PRIM(X'12B1' 62111 X'22' X'01')  SEC(X'92B1' 22359 X'22' X'01') MODE(COPY) CRIT(NO) PACE(1)
CESTPAIR DEVN(X'2334')  PRIM(X'12B1' 62111 X'34' X'01')  SEC(X'92B1' 22359 X'34' X'01') MODE(COPY) CRIT(NO) PACE(1)
CESTPAIR DEVN(X'2346')  PRIM(X'12B1' 62111 X'46' X'01')  SEC(X'92B1' 22359 X'46' X'01') MODE(COPY) CRIT(NO) PACE(1)
CESTPAIR DEVN(X'2308')  PRIM(X'12B1' 62111 X'08' X'01')  SEC(X'92B1' 22359 X'08' X'01') MODE(COPY) CRIT(NO) PACE(1)
```

10,607 CESTPAIR commands in 9 Disk Subsystems with 61 LCUs for production

CESTPAIR Performance Impact (1)

When you send a bunch of CESTPAIR commands (10,000 of them spread across 9 Disk Subsystems for our Production Sysplex), they are accepted in a few minutes (5) and the Initial Copy takes place (with a maximum parallelism of 64 volumes being copied per Disk Subsystem)

- As the 10,000 commands are processed, Disk Subsystems must perform certain validations and, meanwhile (5 minutes), disk response time can be severely impacted.



CESTPAIR Performance Impact (2)

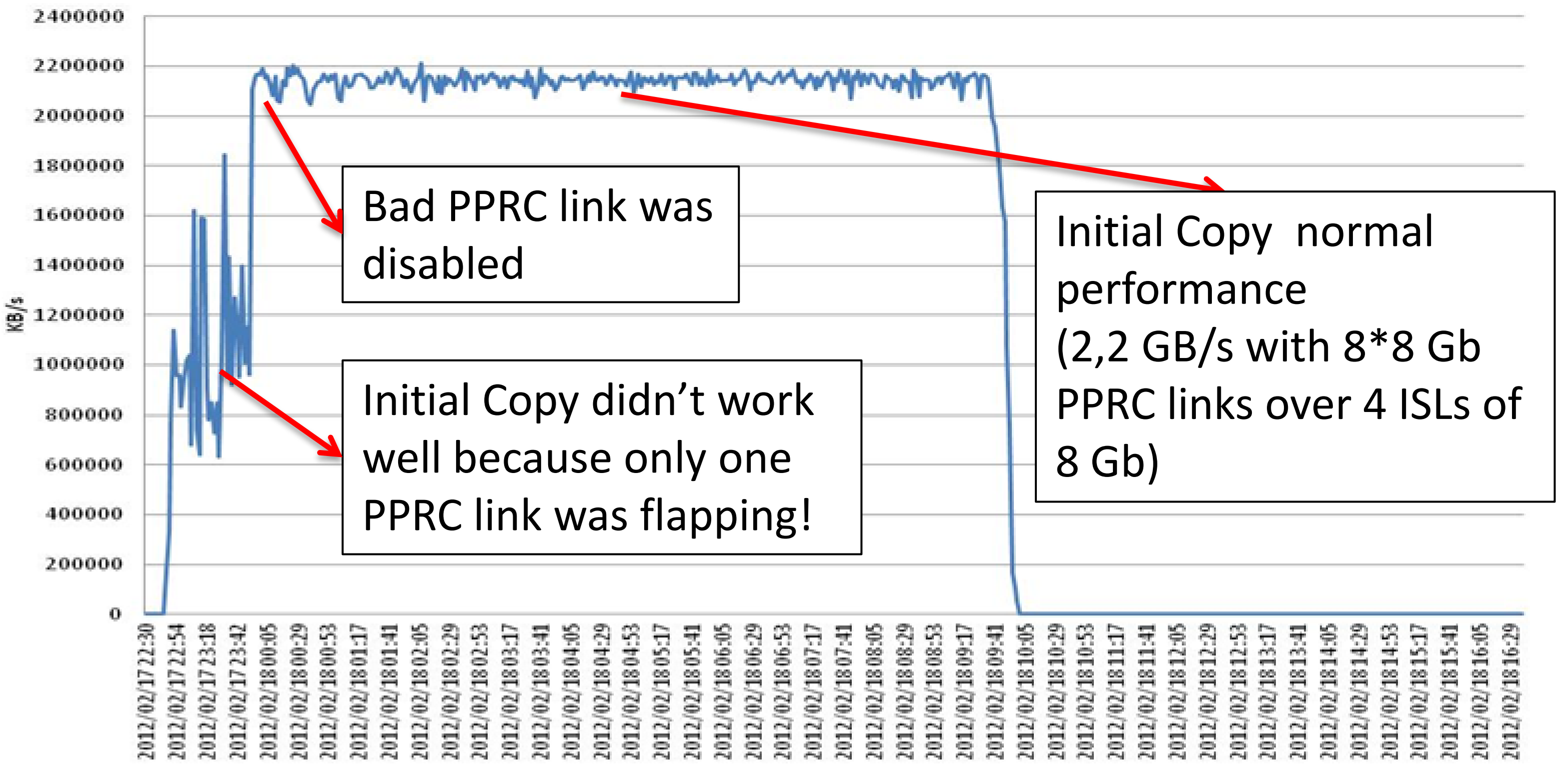
We did two things to mitigate this impact

- Start initial copy at a very low activity period
- Send CESTPAIR commands for one Disk Subsystem, wait for completion (5 minutes) and continue with the next one. **Don’t process all CESTPAIR commands at the same time**
  - There could be an elongation of 45 minutes in the total Initial Copy process duration but was worth it. We reduced these 45 minutes by starting Initial Copy in the biggest disk subsystem first

Initial Copy Flapping Links

When a PPRC link has IOerrors, Initial copy performance drops dramatically! Even if you have many other available links. Many improvements could be done in this area...

Initial Copy Performance for one Disk Subsystem



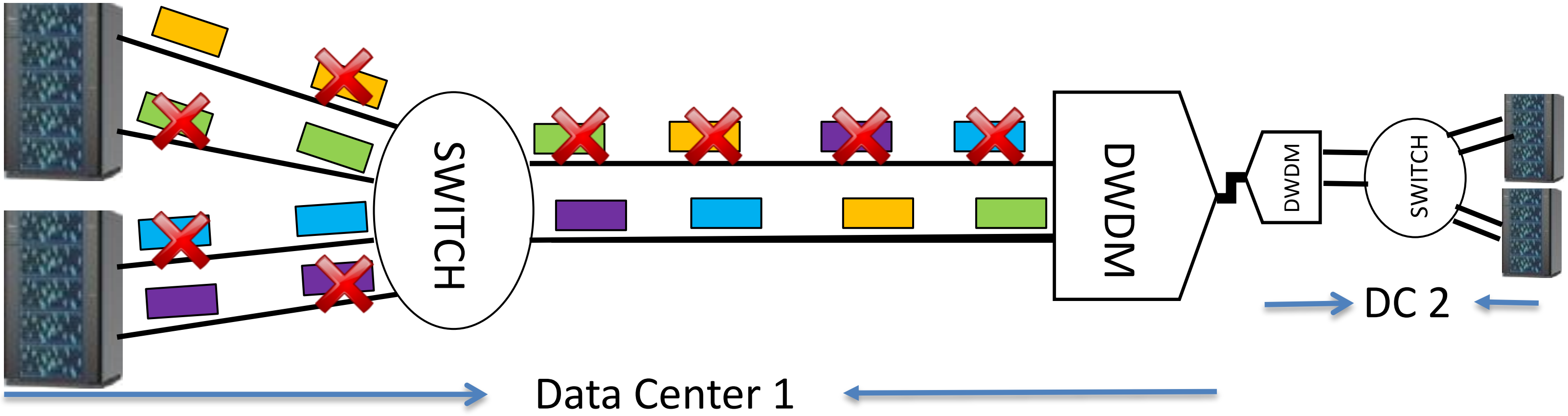
Bad PPRC link was disabled

Initial Copy didn't work well because only one PPRC link was flapping!

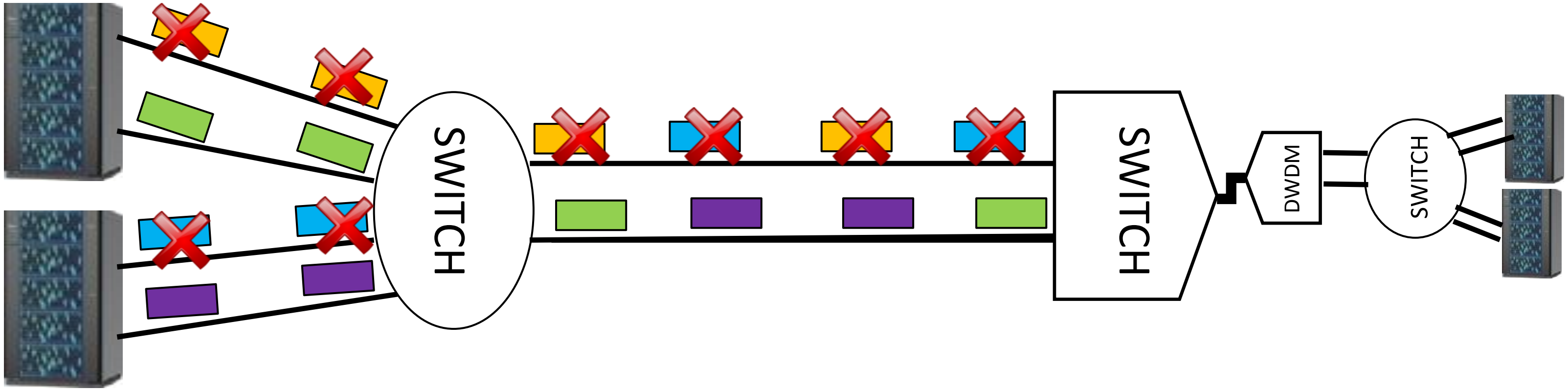
Initial Copy normal performance (2,2 GB/s with 8*8 Gb PPRC links over 4 ISLs of 8 Gb)

Fabric for Replication Congestion (1)

Be careful with the Fabrics for replication. We used dynamic load balancing at frame level, but problems in one ISL were propagated to many adapters/CUs sharing the Fabric

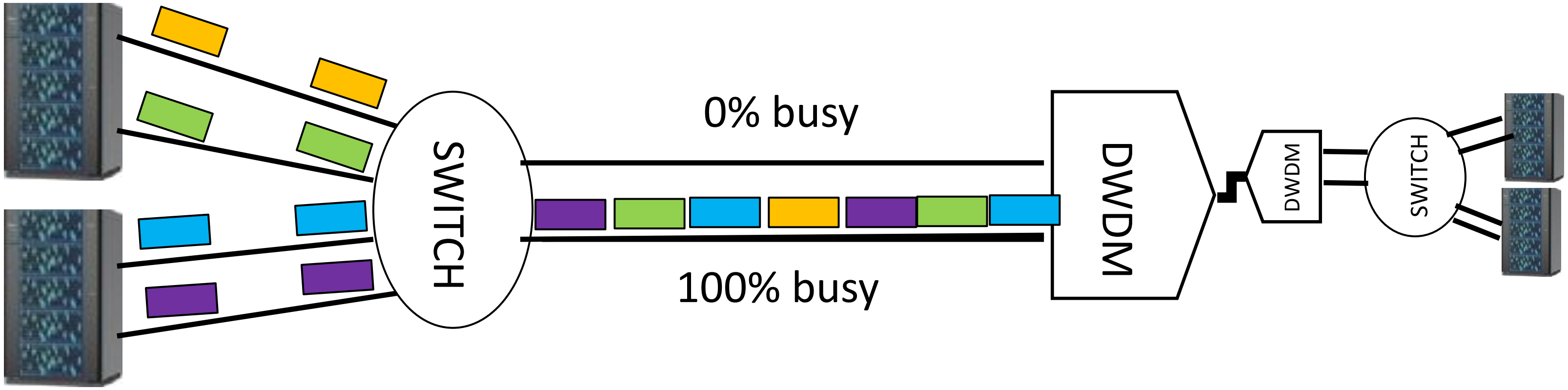


To minimize the impact of ISLs/DWDM cards failures we changed to static load balancing (round robin at connection level). Errors are propagated to fewer adapters/CUs



Fabric for Replication Congestion (2)

However, depending on initialization events, with static load balancing (per connection), you could run into a situation where one ISL does nothing and the other is overloaded



This happened to us in our development Sysplex and Initial Copy for 29 TB lasted 11 h instead of 5 h.

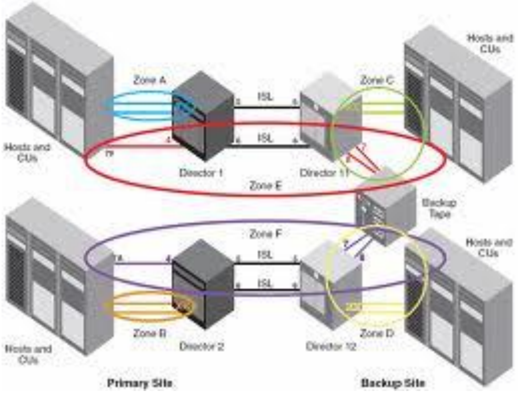
Commands to see which connections (fcid source and target) are using a given ISL (CISCO switches 9134 – CISCO 9148)

- *sh loadbalancing module 1 vsan # (source id FCid) (Destination id FCid)*

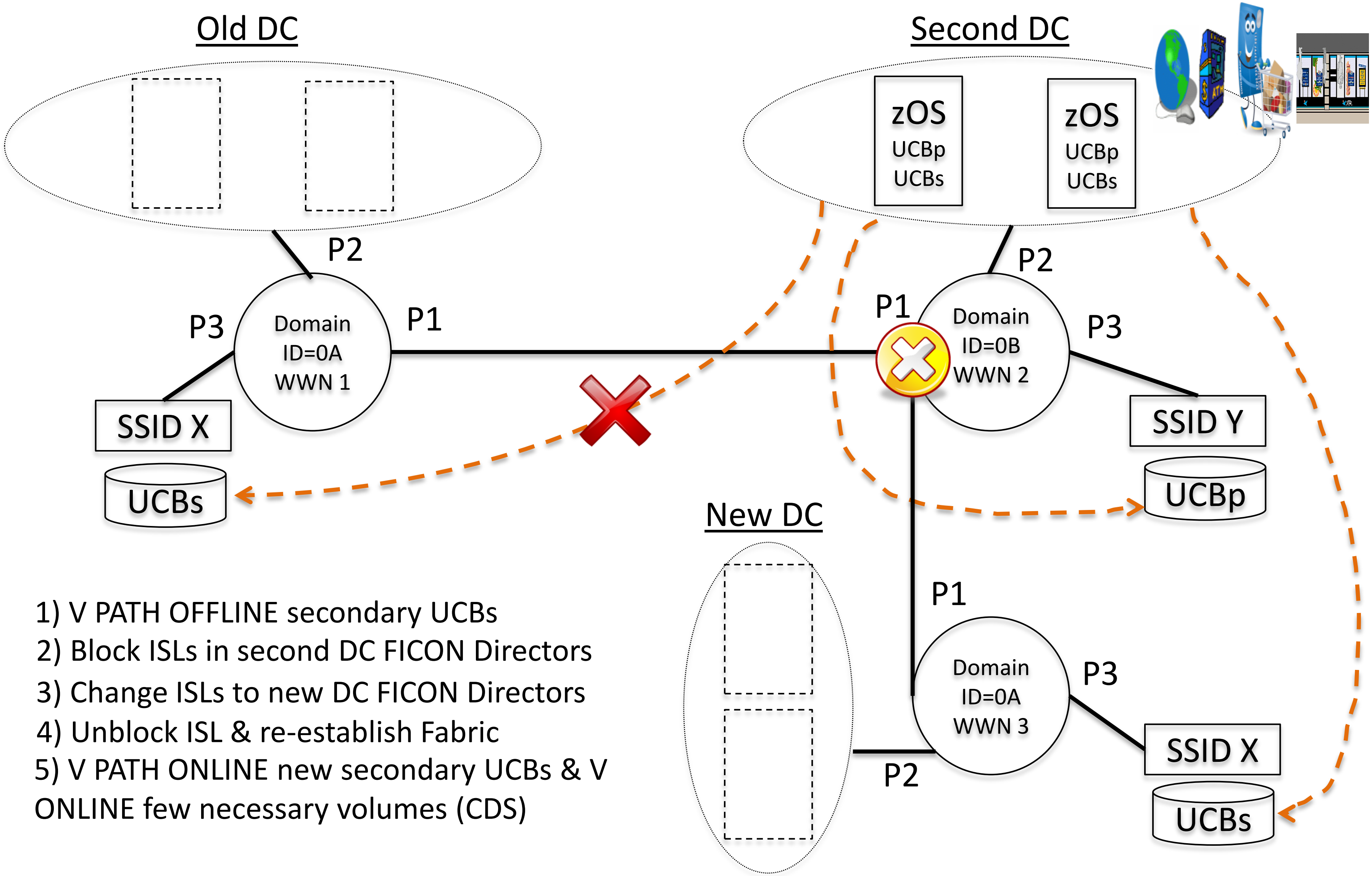
Commands to clean fcids to load balance in an even way

- *purge fcdomian fcid vsan #*

Ficon Directors and Fabrics

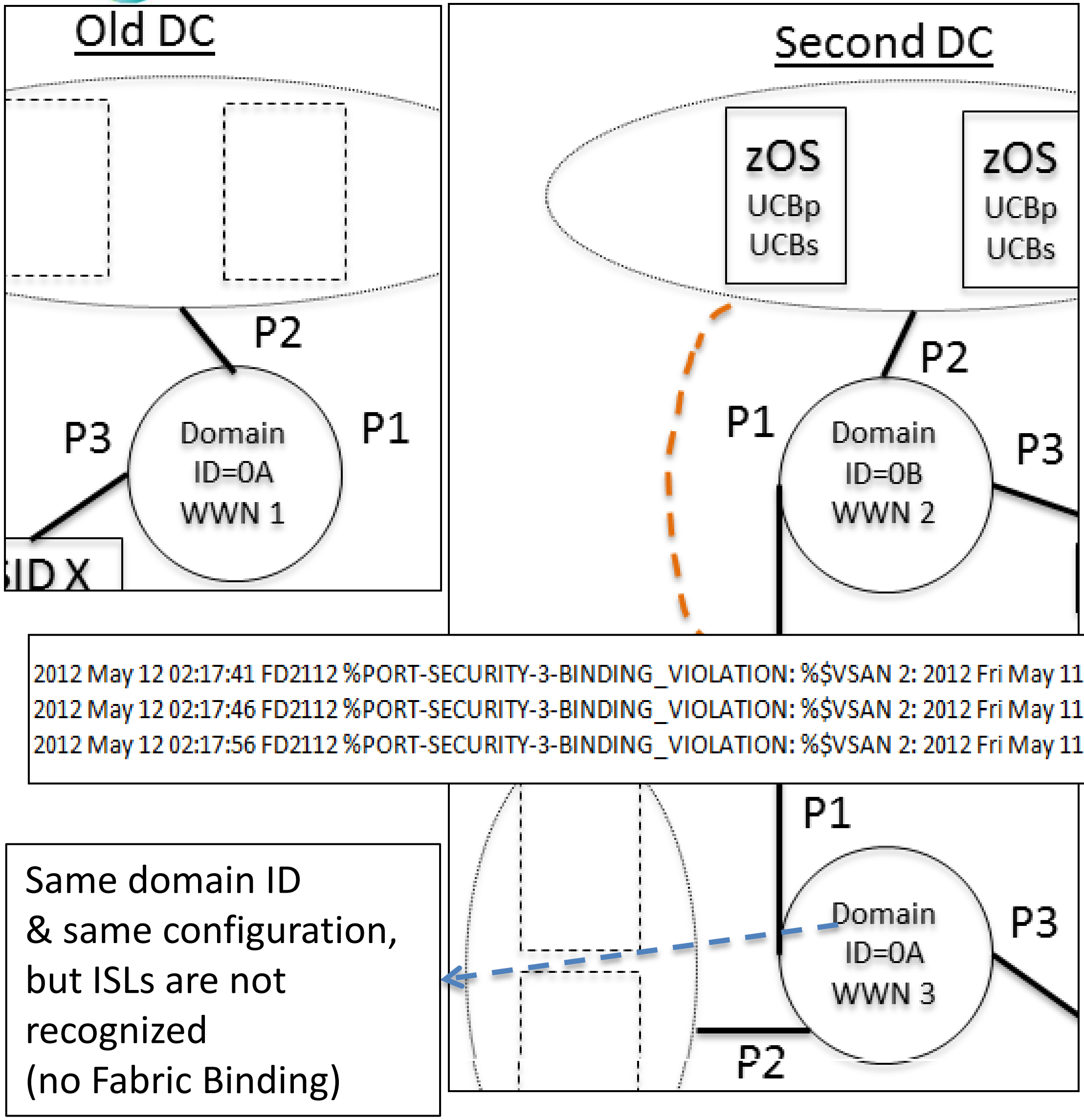


Changing Fabric to New DC



1) V PATH OFFLINE secondary UCBs
2) Block ISLs in second DC FICON Directors
3) Change ISLs to new DC FICON Directors
4) Unblock ISL & re-establish Fabric
5) V PATH ONLINE new secondary UCBs & V ONLINE few necessary volumes (CDS)

New Fabric Binding Failure



Same domain ID
& same configuration,
but ISLs are not
recognized
(no Fabric Binding)

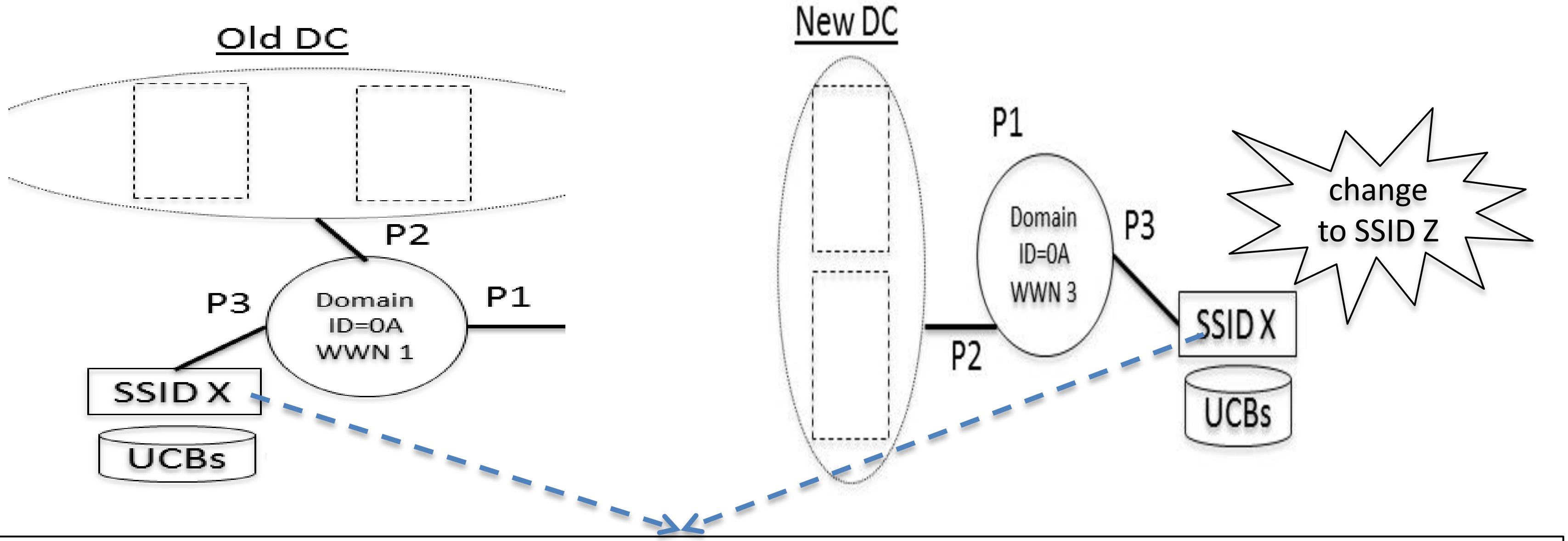
After connecting and unblocking ISLs to new DC FICON directors, the FABRIC binding wasn't successful and ISLs didn't work

```
2012 May 12 02:17:41 FD2112 %PORT-SECURITY-3-BINDING_VIOLATION: %$VSAN 2: 2012 Fri May 11 23:33:42.69998%$ <Fabric Binding:: sWWN: 20:02:54:7f:ee:02:5d:81>
2012 May 12 02:17:46 FD2112 %PORT-SECURITY-3-BINDING_VIOLATION: %$VSAN 2: 2012 Fri May 11 23:33:47.70008%$ <Fabric Binding:: sWWN: 20:02:54:7f:ee:02:5d:81>
2012 May 12 02:17:56 FD2112 %PORT-SECURITY-3-BINDING_VIOLATION: %$VSAN 2: 2012 Fri May 11 23:33:57.70016%$ <Fabric Binding:: sWWN: 20:02:54:7f:ee:02:5d:81>
```

To go ahead with the move, in the middle of the night.., we had to **upgrade NX-OS from 4.2.1. to 4.2.7b** There was a bug in the code and the new WWN wasn't considered a CISCO one... (SAN INTEGRITY issue)

Subsystem IDs (SSIDs)

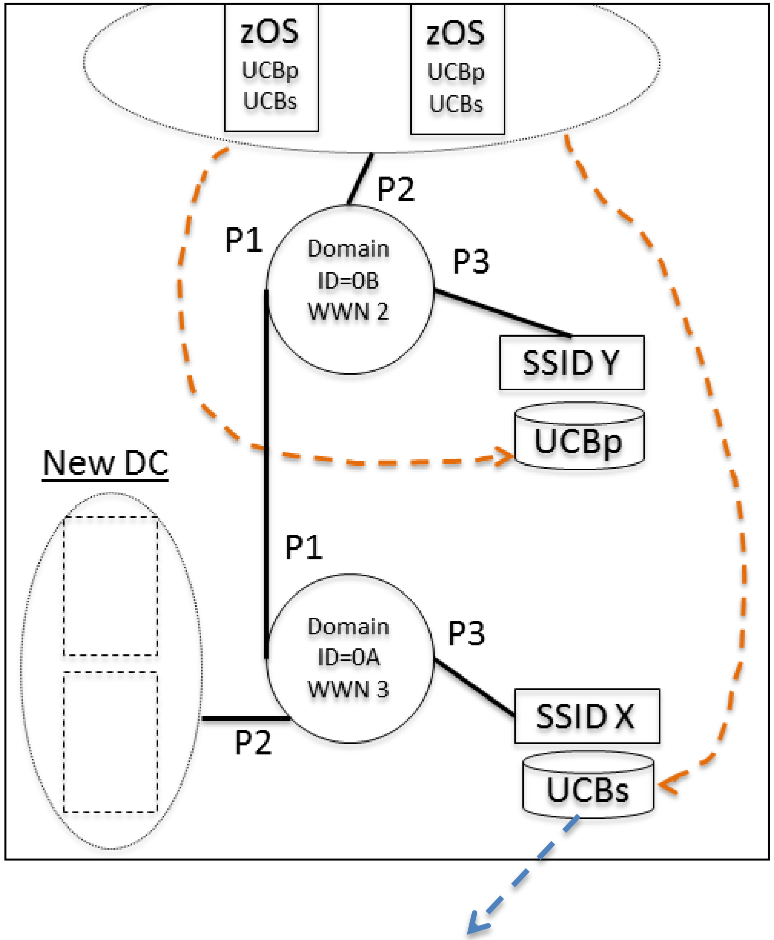
Don’t duplicate Disk Subsystem IDs in new DC!



Configuration in new DC can’t be an exact clone of the one in old DC
SSIDs for disk subsystems must be different. If not, when you try to bring a new UCBs online:
IEC334I DUPLICATE SUBSYSTEM X'*ssid***' CCA X'***cca***', DEVICE** *addr* **NOT BROUGHT ONLINE**

```
IEE421I V 950C,ONLINE 215
IEE103I UNIT 950C NOT BROUGHT ONLINE
IEE763I NAME= IECDINIT CODE= 0000000001000884
IEC334I DUPLICATE SUBSYSTEM X'2155', CCA X'0C', SERIAL=XX55-10943
IEE764I END OF IEE103I    RELATED MESSAGES
```

HPAV



HPAV binding issues after changing to new secondary

When bringing new DC disks online (CDS volumes, for instance), we got

IOS291I IOS1291I
CONFIGURATION DATA COULD NOT BE READ ON PATH (*devn*, *xx*) RC=*rc textline1* [*textline2*]

```
18:07:15.50 V 9481,ONLINE
  IOS291I  CONFIGURATION DATA COULD NOT BE READ ON PATH(9488,F5) RC=21
    TOKEN NED MISMATCH HAS BEEN DETECTED
  IOS291I  CONFIGURATION DATA COULD NOT BE READ ON PATH(9489,F5) RC=21
    TOKEN NED MISMATCH HAS BEEN DETECTED
<all HPAV addresses>
```

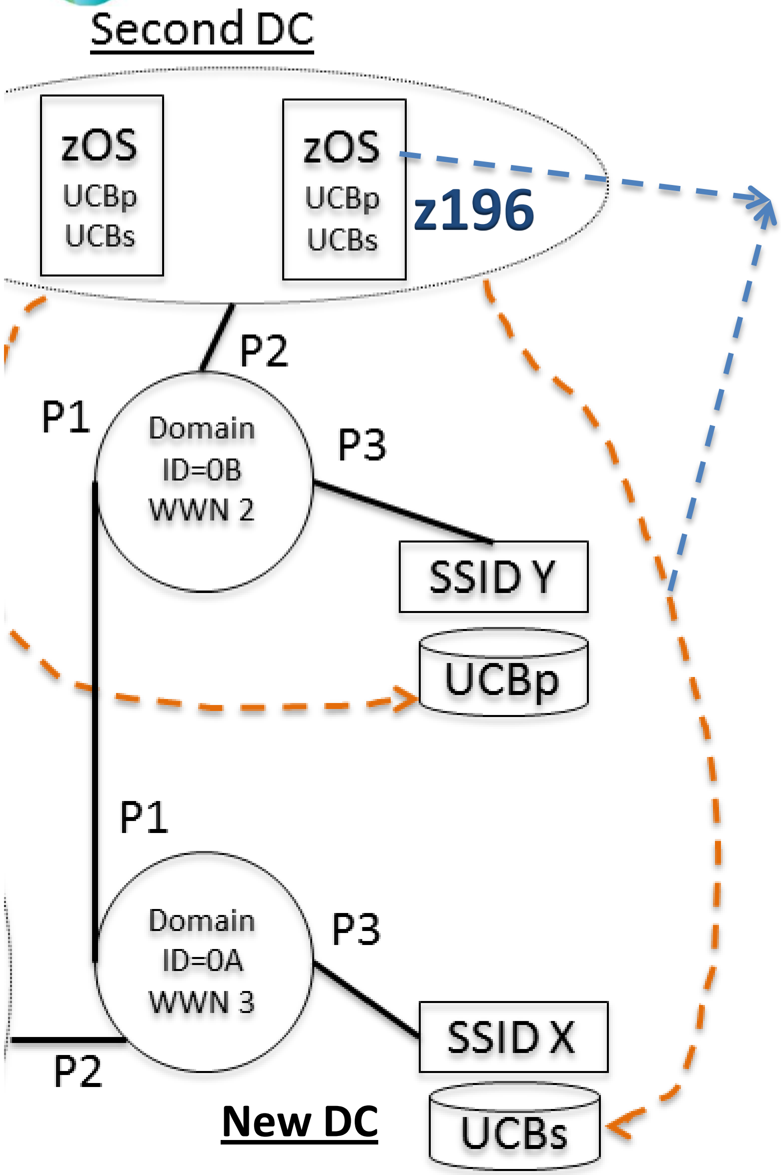
for HPAV devices and they didn't work (didn't bind to any base device)

New APAR **OA38759**:

IOS291I ISSUED FOR ALIAS DEVICES AFTER MAKING HARDWARE CONFIGURATION CHANGES

```
If a configuration change is made that does not require a
Dynamic ACTIVATE (for example, a Push/Pull of a control unit
who's device configuration remains the same), residual Self
Description data may be left in IOS's Configuration Data Table
for alias devices causing Self Description processing to fail
for the alias device with an IOS291I message when the devices
are initialized via a VARY device online for its base device
or Hyperswap Configuration Load processing.
```

New Fabric Recognition by z196



For z10 it’s enough to V PATH offline from old UCBs, change Fabric to new UCBs and V PATH / V DEVICE online new DC UCBs

Something has changed with z196 and you need to CONFIGURE CHANNEL OFFLINE /ONLINE to bring new UCBs online. Otherwise **path** is **PHYSICALLY UNAVAILABLE**

```
V PATH(2200-2201,4A),ONLINE
IEE386I PATH(2200,4A) NOT BROUGHT ONLINE 022
IEE763I NAME= IECVIOPM CODE= 0000000400000000
IOS552I PATH NOT PHYSICALLY AVAILABLE
IEE764I END OF IEE386I    RELATED MESSAGES
IEE386I PATH(2201,4A) NOT BROUGHT ONLINE 023
IEE763I NAME= IECVIOPM CODE= 0000000400000000
IOS552I PATH NOT PHYSICALLY AVAILABLE
IEE764I END OF IEE386I    RELATED MESSAGES
```

NETWORK & COMS SERVER

New Network Infrastructure

In the new DC we have important differences regarding the network infrastructure

- Nexus for the backbone instead of Catalyst
- 10 Gb in the backbone (in old DCs is 1 Gb)
- Fewer building blocks (more simple)

→ What we took for granted in the old DCs, changed in the new one

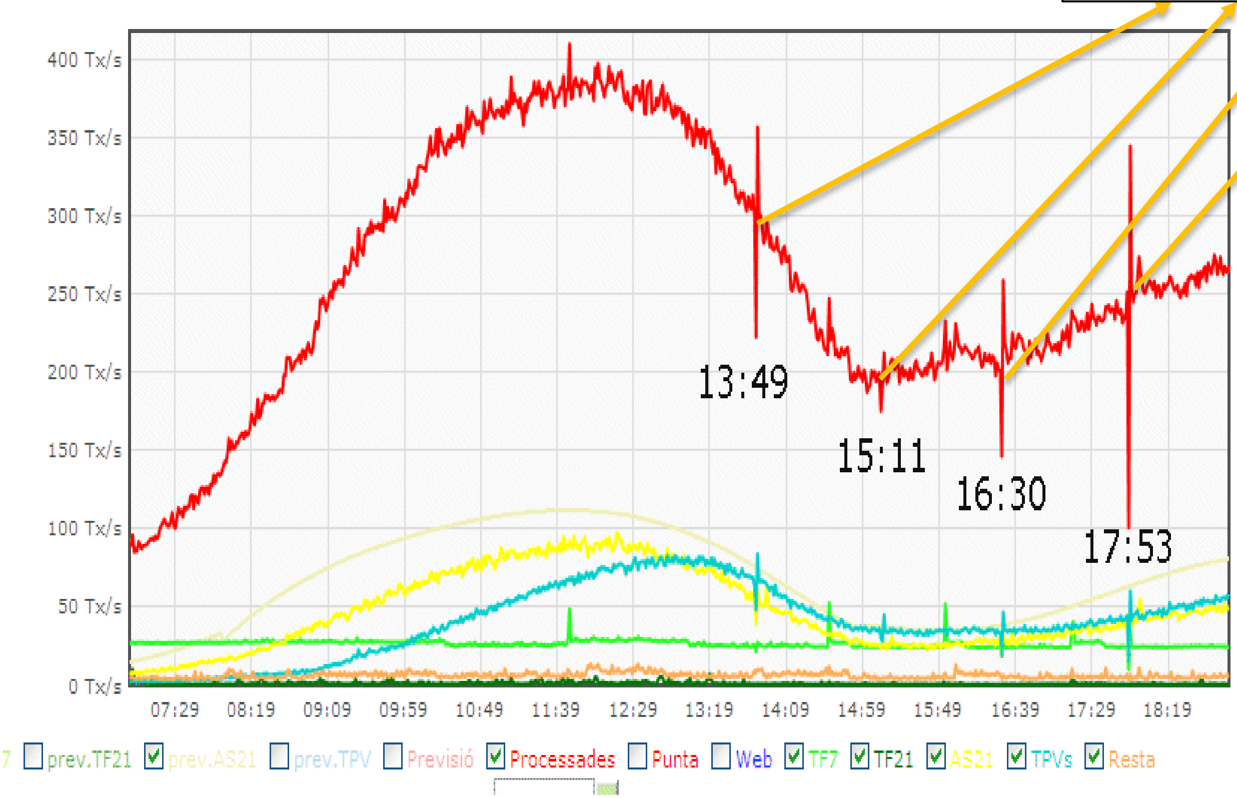
→ We got some issues and some of them are being analyzed by IBM & CISCO

→ All the problems are related to XCF being used as network adapter in addition to the regular network. Basically, these problems arise when we have a multisite sysplex configuration

All Through One (1)

As we try new DC infrastructure by moving there one system (from ten) for a while (multisite testing), every time we stop it, there is a drop in IMS processing.

IMS processing drops when stopping the system in new DC



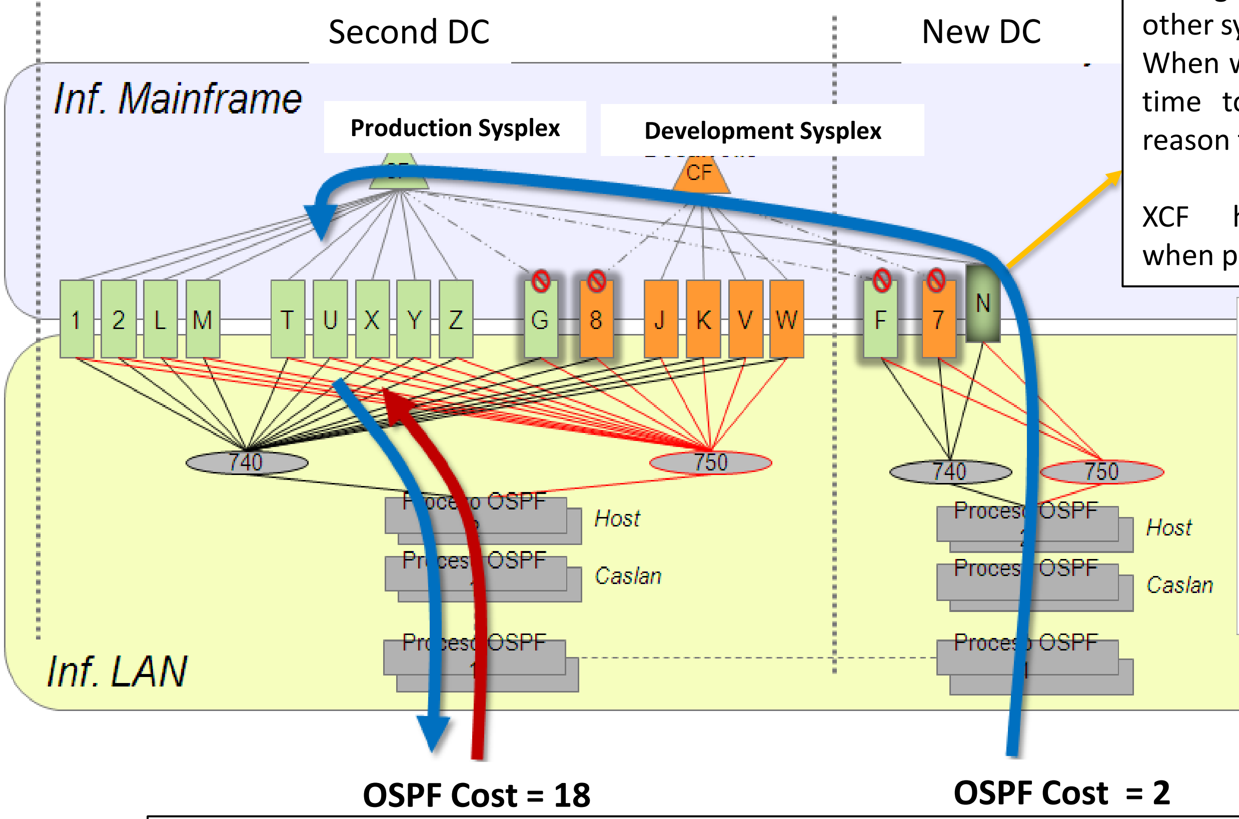
All Through One (2)

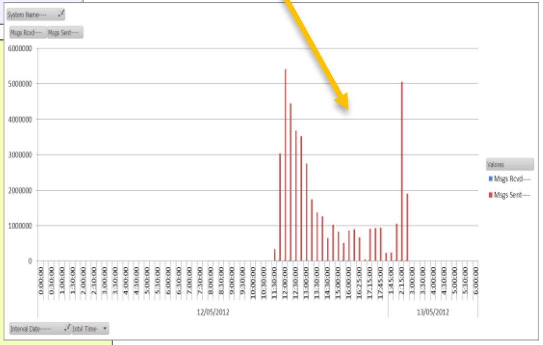
The problem is that when we move one Sysplex System to the new DC, the network sends all the incoming traffic to it (for the whole sysplex) because the route through XCF has a lower OSPF cost!

All traffic for production goes through SYSN and by XCF reaches other systems in the sysplex.

When we stop SYSN the network needs some time to reevaluate the routes. That's the reason for IMS processing drop.

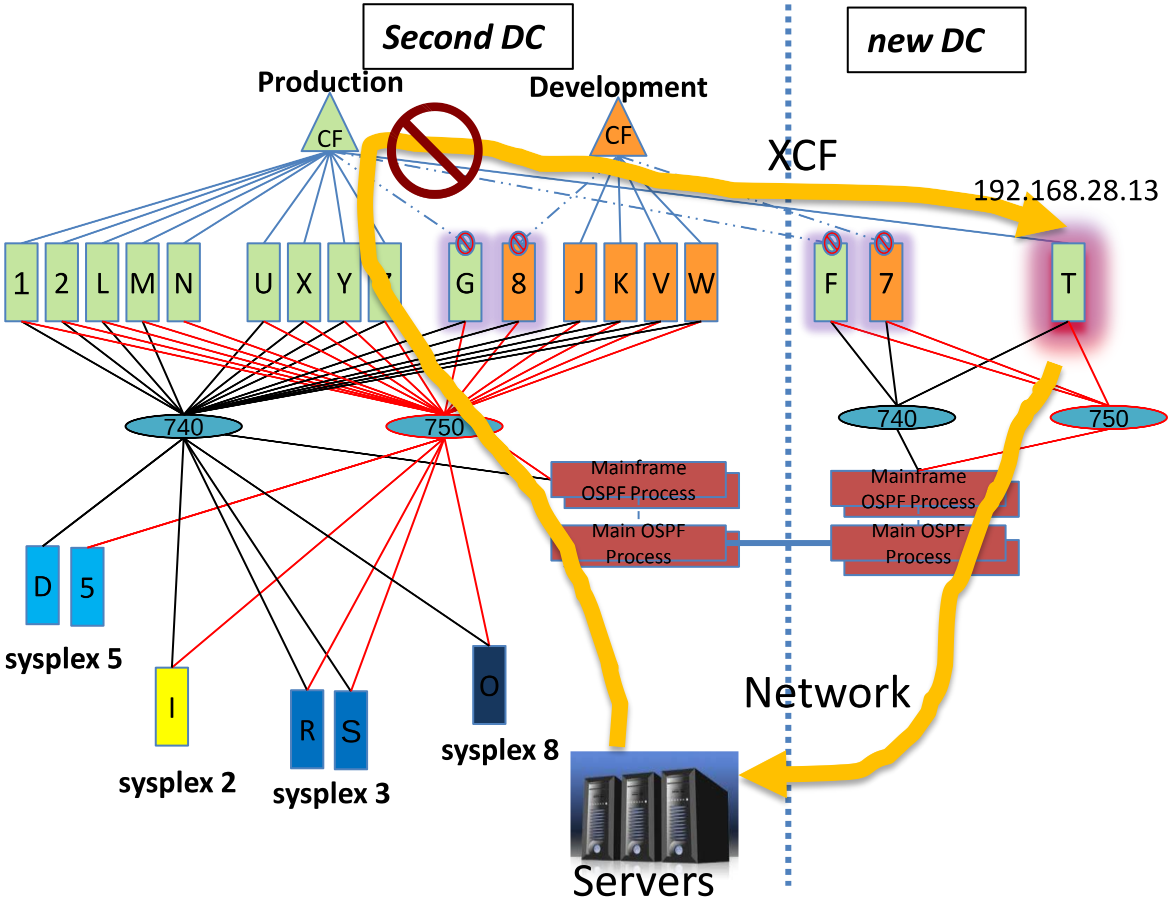
XCF heavily used by Coms Server when performing multisite





The solution is to artificially increase OSPF costs in new DC network infrastructure to be the same as in second DC

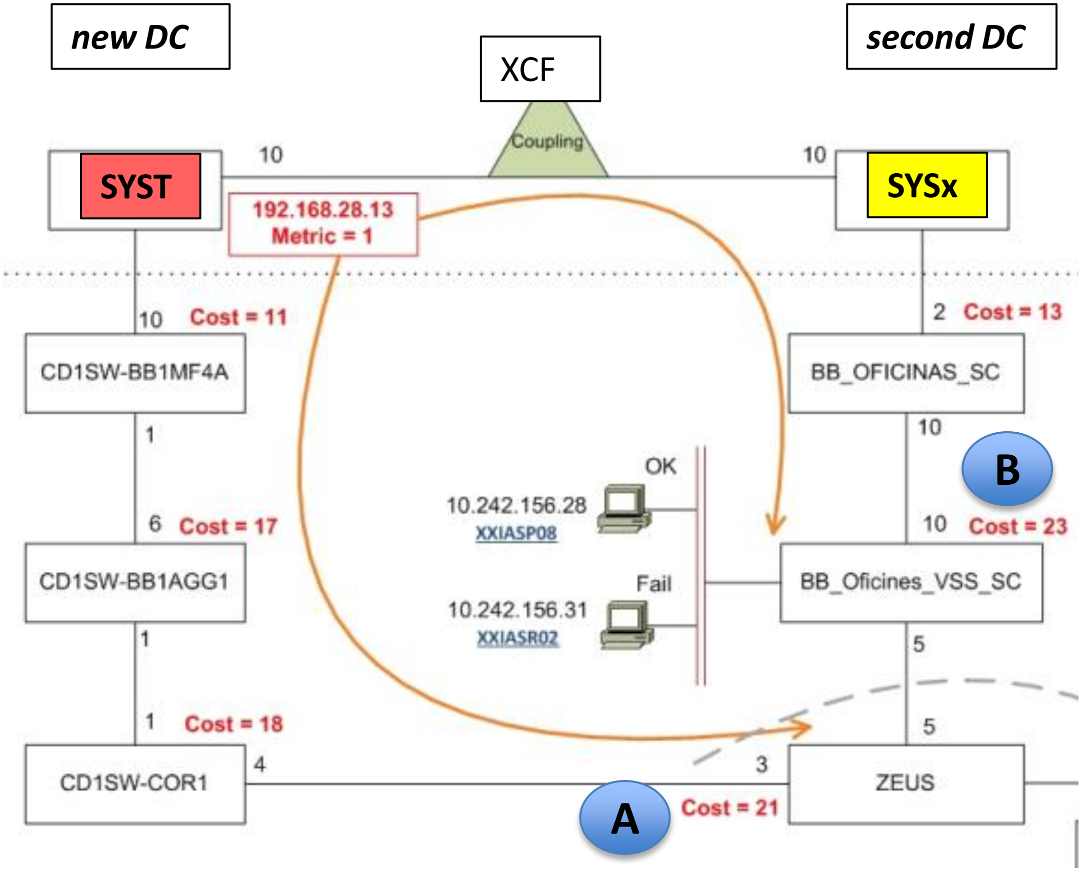
Multisite IP Connectivity Issues (1)



When moving SYST (and only SYST) to new DC it loses IP connectivity to some servers in second DC. Some servers in the same VLAN have visibility to SYST and some others don't.

By performing network troubleshooting, it seems that communication is lost between the server and SYST through XCF (using another system in second DC)

Multisite IP Connectivity Issues (2)



Normal network should be used (A). However, servers in VSS vlans use XCF to send back answers and here is where we experienced the problem.

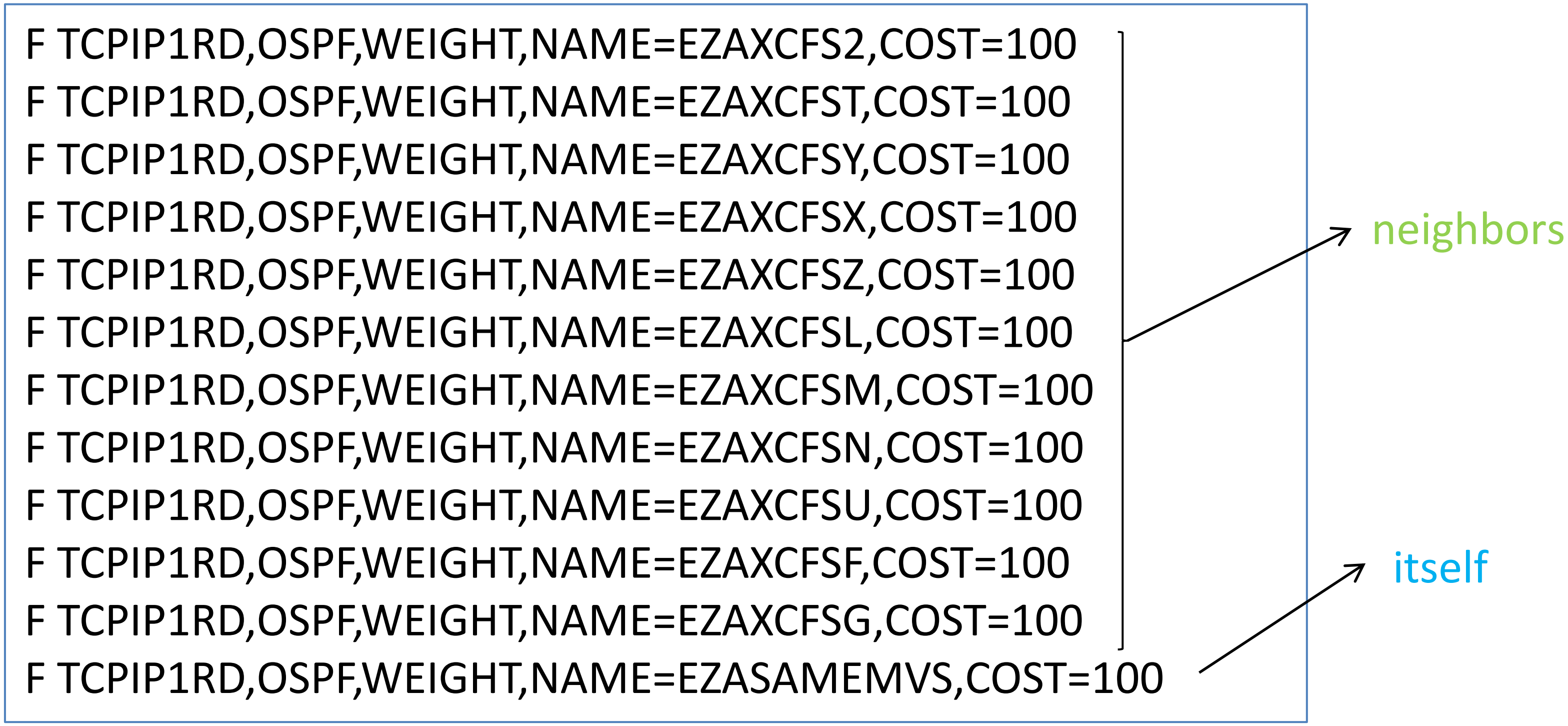
Sometimes this doesn't work

Solution: Increase XCF metric to 100 to avoid it completely !!! (sysplex distributor continues working well)

Multisite IP Connectivity Issues (3)

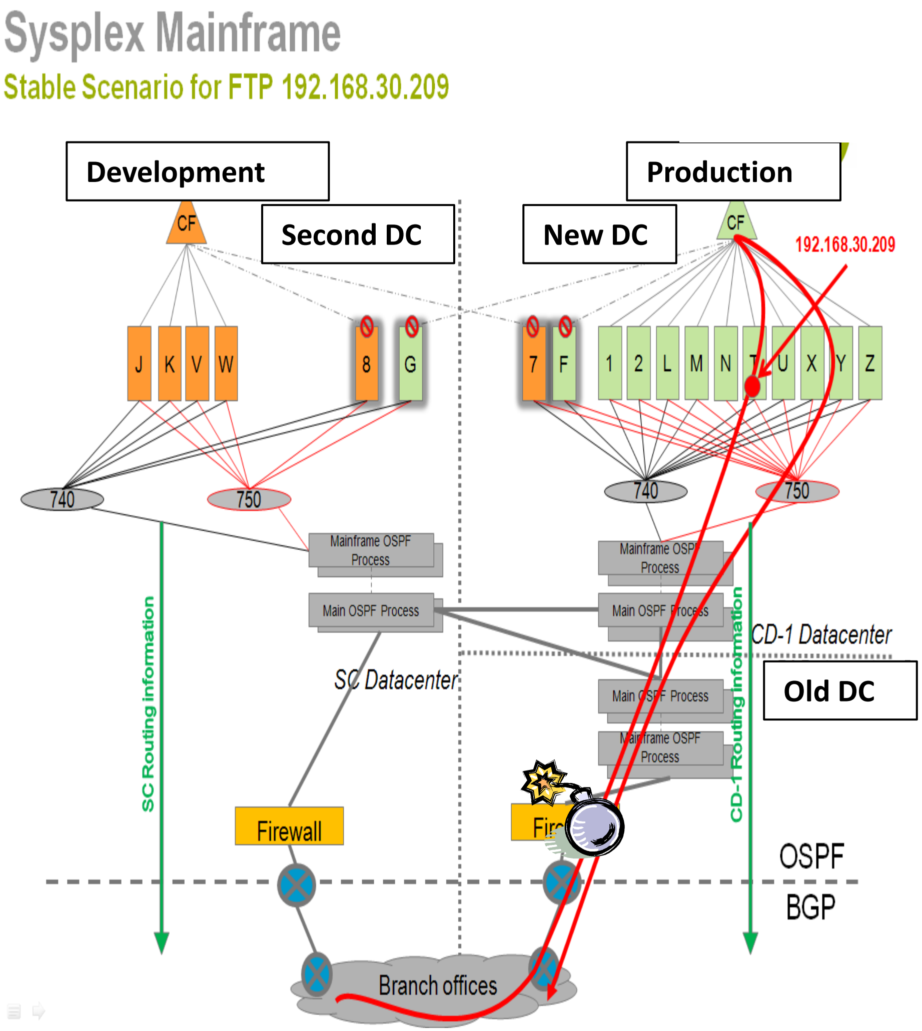
You can increase XCF costs dynamically

On each system in the sysplex we increased the cost of all OSPF XCF interfaces (for the neighbors and for itself)



```
F TCPIP1RD,OSPF,WEIGHT,NAME=EZAXCFS2,COST=100
F TCPIP1RD,OSPF,WEIGHT,NAME=EZAXCFST,COST=100
F TCPIP1RD,OSPF,WEIGHT,NAME=EZAXCFSY,COST=100
F TCPIP1RD,OSPF,WEIGHT,NAME=EZAXCFSX,COST=100
F TCPIP1RD,OSPF,WEIGHT,NAME=EZAXCFSZ,COST=100
F TCPIP1RD,OSPF,WEIGHT,NAME=EZAXCFSL,COST=100
F TCPIP1RD,OSPF,WEIGHT,NAME=EZAXCFSM,COST=100
F TCPIP1RD,OSPF,WEIGHT,NAME=EZAXCFSN,COST=100
F TCPIP1RD,OSPF,WEIGHT,NAME=EZAXCFSU,COST=100
F TCPIP1RD,OSPF,WEIGHT,NAME=EZAXCFSF,COST=100
F TCPIP1RD,OSPF,WEIGHT,NAME=EZAXCFSG,COST=100
F TCPIP1RD,OSPF,WEIGHT,NAME=EZASAMEMVS,COST=100
```

Multisite with Firewalls State-full (1)



FTP distributed VIPA works well when there is no multisite sysplex. Each DC announces only its own routes

When a system is in Multisite, even from a different sysplex (they belong to same OSPF area), routing information is sent back to second DC through XCF, which advertises routes from the two DCs and leads into a firewall state-full issue

Second DC shouldn't announce routes from the two DCs. CISCO is looking into it.

Bypass:

Don't use XCF for sending routing information (Sysplex Distributor continues working well)

Multisite with Firewalls State-full (2)

Every time we IPL a system in the “other DC” (Multisite IPL) we modify OMPROUTE profile to avoid using XCF as OSPF interfaces. That is:

~~OSPF_Interface~~
~~IP_Address=192.168.49.*~~
~~Subnet_Mask=255.255.255.248~~
~~Attaches_To_Area=0.0.38.12~~
~~MTU=32768~~
~~Cost0=100~~
~~Router_Priority = 0;~~

XCF

```
OSPF_Interface
   IP_Address=10.242.70.62
   NAME=LOSA0860
   ROUTER_PRIORITY = 0
   SUBNET_MASK=255.255.255.0
   ATTACHES_TO_AREA=0.0.38.12
   MTU=1500;

OSPF_Interface
   IP_Address=10.242.80.70
   NAME=LOSA0890
   ROUTER_PRIORITY = 0
   SUBNET_MASK=255.255.255.0
   ATTACHES_TO_AREA=0.0.38.12
   MTU=1500;
```

OSA adapters

Thanks

Any Questions?