

Z16

Server Time Protocol (STP)

Migration and Recovery Considerations

Noshir Dhondy (dhondy@us.ibm.com)

IBM
SYSTEM z9 AND zSERIES EXPO
October 9 - 13, 2006



Orlando, FL

May 16, 2005

© IBM Corporation 2006

Agenda

- STP Overview
 - ? Terminology
- Migration Considerations
 - ? Hardware installation planning
 - ? Connectivity planning
 - ? Software installation planning
- Recovery Considerations
 - ? STP Recovery concepts
 - ? Mixed-CTN recovery
 - ? STP-only CTN recovery
 - ? Site failure scenarios

What is STP?

- Provides capability for multiple servers to maintain time synchronization with each other and form a Coordinated Timing Network (CTN)
 - ? CTN: a collection of servers that are time synchronized to a time value called Coordinated Server Time (CST)
- Server-wide facility implemented in IBM System z9™ Enterprise Class (EC), z9 Business Class (BC), z990, z890 Licensed Internal Code (LIC)
 - ? Single view of “time” to PR/SM
 - ? PR/SM can virtualize this view of time to the individual logical partitions (LPARs)
 - ? (for example z/OS)
 - ? STP not available on z900, z800 or 9672 Gx servers
- Message based time synchronization protocol
 - ? Similar to Network Time Protocol (NTP) – an industry standard
 - ? Timekeeping information transmitted over Coupling Links
 - ? ISC-3 links (Peer mode), ICB3 and ICB4 links
 - ? **NOT standard NTP**

Terminology

- STP-capable server/CF
 - ? z9 EC, z9 BC, z990, z890 server/CF with STP LIC installed
- STP-enabled server/CF
 - ? STP-capable server/CF with STP FC 1021 installed
 - ? STP panels at the HMC/SE can now be used
- STP-configured server/CF
 - ? STP-enabled server/CF with a CTN ID assigned
 - ? STP message exchanges can take place
- CTN
 - ? Collection of servers that are time synchronized to a time value called Coordinated Server Time (CST)
- CTN ID
 - ? Servers / Coupling Facilities (CFs) that make up a CTN are all configured with a common identifier CTN ID

Terminology (continued)

- Two types of CTN configurations possible:
 - ? Mixed CTN
 - ? Allows servers/CFs that can only be synchronized to a Sysplex Timer (ETR network) to coexist with servers/CFs that can be synchronized with CST in the "same" timing network
 - ? Sysplex Timer provides timekeeping information
 - ? CTN ID format
 - STP network ID concatenated with ETR network ID
 - ? STP-only CTN
 - ? All servers/CFs synchronized with CST
 - ? Sysplex Timer is NOT required
 - ? CTN ID format
 - STP network ID only (ETR network ID field has to be null)

MIGRATION CONSIDERATIONS

Hardware installation planning

- 9037-002 concurrent LIC upgrade
 - ? If migrating from ETR network
 - ? 9037 code changes to support STP Mixed CTN
- z9 EC, z9 BC server/CF must be at EC Driver level 63J
- z990 and z890 server/CF must be at EC Driver level 55K
- Install all of the latest MCLs (concurrent install) for Driver 63J and/or Driver 55K on each z9, z990, z890 server/CF that will be configured in a Mixed or STP-only CTN
 - ? STP prerequisite MCLs (LIC) will be installed
 - ? Server becomes STP-capable
 - ? STP functions cannot be used until server/CF STP-enabled

7

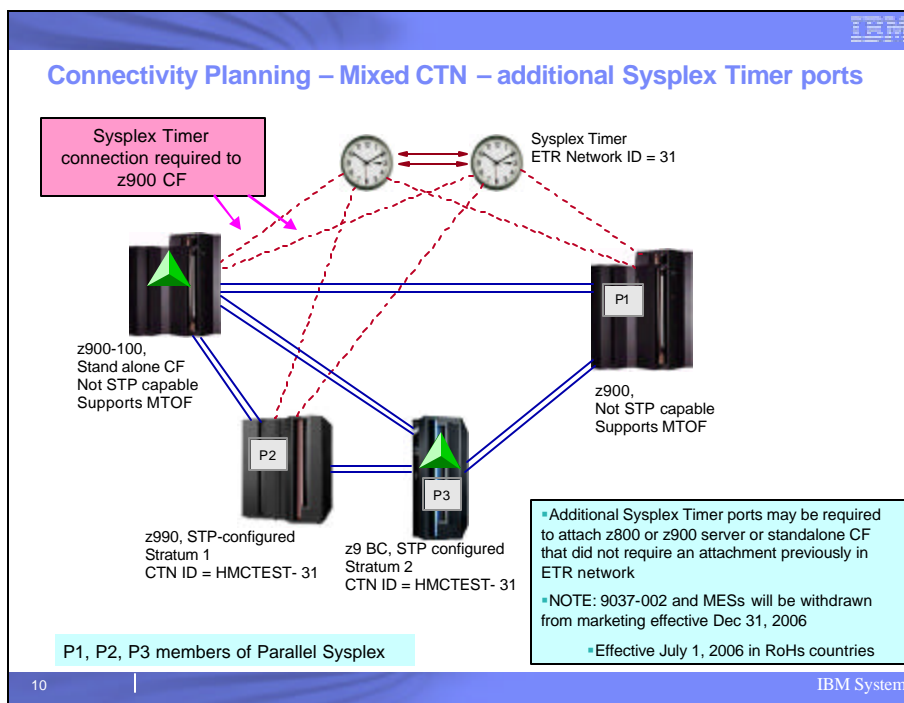
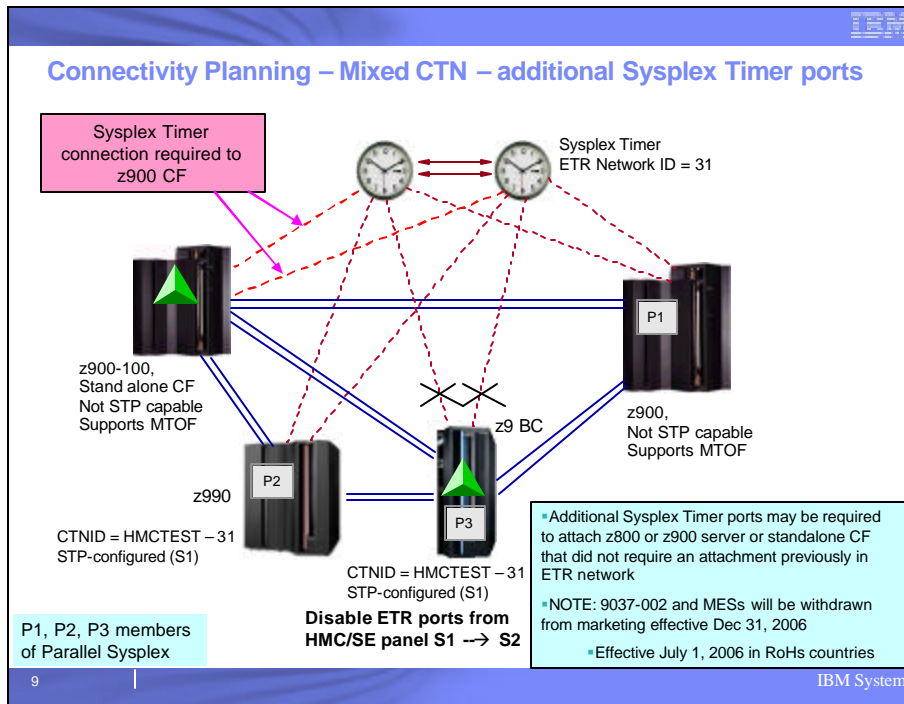
IBM Systems

Hardware installation planning (continued)

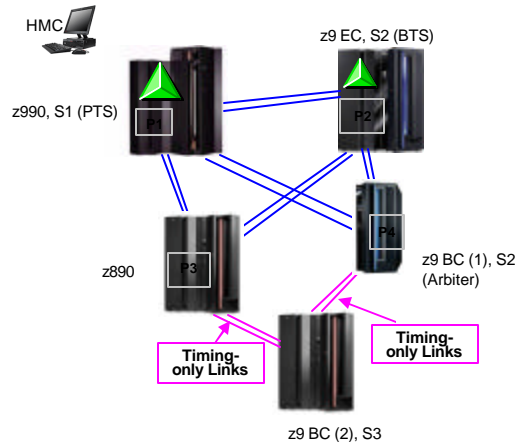
- HMC v2.9.1 (EC Driver level 64) or higher
 - ? Can upgrade z890/z990 HMC to new HMC code level
 - ? If existing HMC being upgraded, attached z990, z890, z900, z800 servers/ CFs need to be at specific driver levels and specific MCLs need to be installed.
- Install STP Enablement MCL (FC 1021) on each z9, z990, z890 server that will be configured in a Mixed or STP-only CTN
 - ? Concurrent install
 - ? Server/CF becomes STP-enabled
 - ? STP functions (eg HMC/SE panels) can now be used
 - ? Servers/ CFs can now be configured for STP
 - ? Chargeable feature

8

IBM Systems



Connectivity Planning – Timing-only links



- Coupling links that allow 2 servers to be synchronized when a CF does not exist at either end of link
 - ? Typically required when synchronization needed in non Parallel Sysplex configurations
 - ? For example: Base Sysplex; XRC
- HCD enhanced to define Timing-only links
- Can be defined in either Mixed CTN or STP-only CTN
- Timing-only links used to transmit STP messages only

P1, P2, P3, P4 members of Parallel Sysplex

11

IBM Systems

Connectivity Planning – multi-site sysplex

- If CTN across multiple sites:
 - ? Maximum fiber distance of ISC-3 Peer links without repeaters is 10 km
 - ? 12 km with RPQ
 - ? If fiber distance greater than 10 km, need to plan for Dense Wavelength Division Multiplexers (DWDM)
- IBM supports only those DWDM products qualified by IBM for use in multi-site sysplex applications, such as GDPS
- At STP GA several DWDM products are planned to be qualified for Sysplex Timer and ISC-3 links (transporting CF and STP messages)
- The latest list of qualified DWDM vendor products can be found on the Resource Link Web site at:
 - ? <https://www.ibm.com/servers/resourcelink>
 - ? They are listed under the Hardware Products for Servers on the Library page
- Plan for redundant, diverse fiber routes between sites to avoid a single point of failure of the fiber trunk

12

IBM Systems

Software Installation Planning

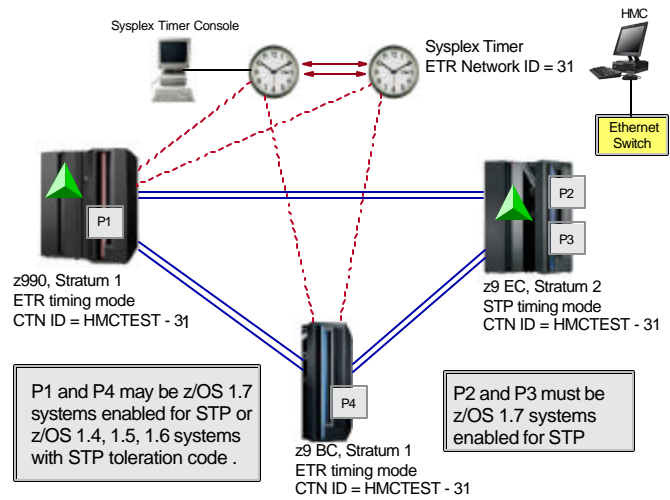
- Install or upgrade z/OS images to z/OS V1.7 or higher
- Install additional software maintenance required for z/OS V1.7, V1.8
 - ? Includes STP enablement APAR
 - ? Maintenance can be applied using “rolling IPL” process
- If you have z/OS V1.4 through z/OS V1.6 systems in a Mixed CTN:
 - ? Install toleration software maintenance required for z/OS V1.4, V1.5, V1.6
 - ? Mixed CTN can include pre-V1.7 systems
- If you have Timing-only links in either a Mixed or STP-only CTN:
 - ? Install software maintenance for HCD, HCM, IOCP to define these links
- Check Preventive Service Planning (PSP) buckets
 - ? Listed in the 2084DEVICE, 2086DEVICE, 2094DEVICE and 2096DEVICE PSP buckets for the z990, z890, z9 EC and z9 BC respectively
 - ? To simplify identification of PTFs for STP, functional PSP bucket created
 - ? Use the Enhanced Preventive Service Planning Tool (EPSPT)
 - ? <http://www14.software.ibm.com/webapp/set2/psp/srchBroker>

CLOCKxx

- Update the CLOCKxx member of SYS1.PARMLIB
- CLOCKxx statements
 - ? OPERATOR PROMPT|NOPROMPT
 - ? TIMEZONE W|E hh.mm.ss
 - ? ETRMODE YES|NO
 - ? ETRZONE YES|NO
 - ? SIMETRID nn (where nn = 0 – 31)
 - ? STPMODE YES|NO
 - ? Specifies whether z/OS is using STP timing mode
 - ? STPMODE YES default
 - ? STPZONE YES|NO
 - ? Specifies whether the system is to get the time zone constant from STP
 - ? ETRDELTA ss | TIMEDELTA ss (where ss = 0 – 99 seconds)

New statements for STP

Timing Modes in a Mixed CTN (example)

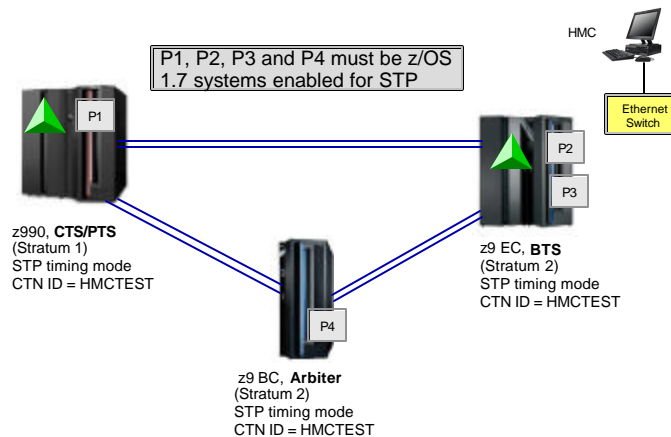


P1, P2, P3, P4 members of Parallel Sysplex

15

IBM Systems

Timing Mode in an STP-only CTN (example)



P1, P2, P3, P4 members of Parallel Sysplex

16

IBM Systems

RECOVERY CONSIDERATIONS

STP Recovery concepts

- Coordinated Server Time
 - ? Coordinated Server Time (CST) represents the time for the CTN and is the time at a Stratum 1 server
- Synchronization check threshold
 - ? Server/CF considered to be in synchronized state if TOD clock within synchronization check threshold of CST
 - ? STP synchronization check threshold 50 microseconds
 - ? If TOD clock differs from CST by more than +/- 50 microseconds, server/CF becomes unsynchronized
 - ? Can become a Stratum 0 server/CF
- Freewheel Interval
 - ? Amount of time a Stratum 2 or Stratum 3 server can remain synchronized without receiving messages from its clock source
 - ? Approximately 1 second (Mixed-CTN)
 - ? Approximately 10 seconds (STP-only CTN)

Mixed-CTN

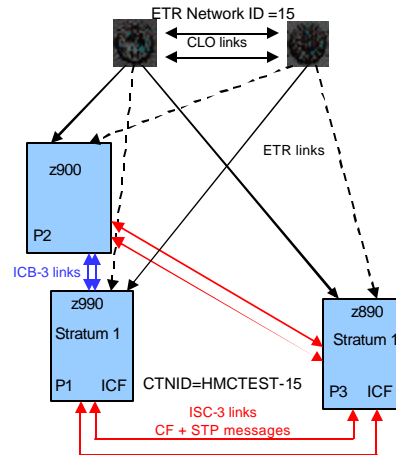
- ETR network recovery

- ? See Backup slides for details

- Improved RAS over ETR network for S1 servers

- ? If either z990 or z890 loses all timing signals from the 9037(s), they are capable of becoming Stratum 2 (S2) servers in a Mixed-CTN

P1, P2, P3 in Parallel Sysplex
 — Active ETR link
 - - - Alternate ETR link



19

IBM Systems

Mixed-CTN

- ETR network recovery

- ? See Backup slides for details

- Improved RAS over ETR network for S1 servers

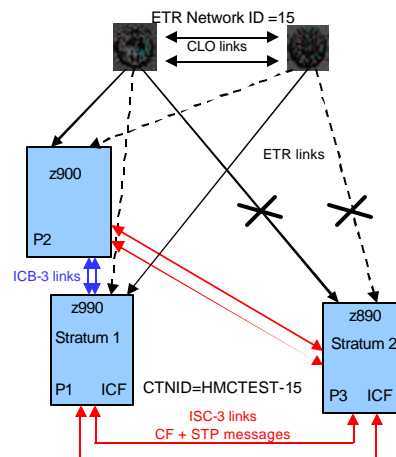
- ? If either z990 or z890 loses all timing signals from the 9037(s), they are capable of becoming Stratum 2 (S2) servers in a Mixed-CTN

- Example:

- ? z890 loses both ETR links

- ? Can synchronize to z990 and become a S2 server

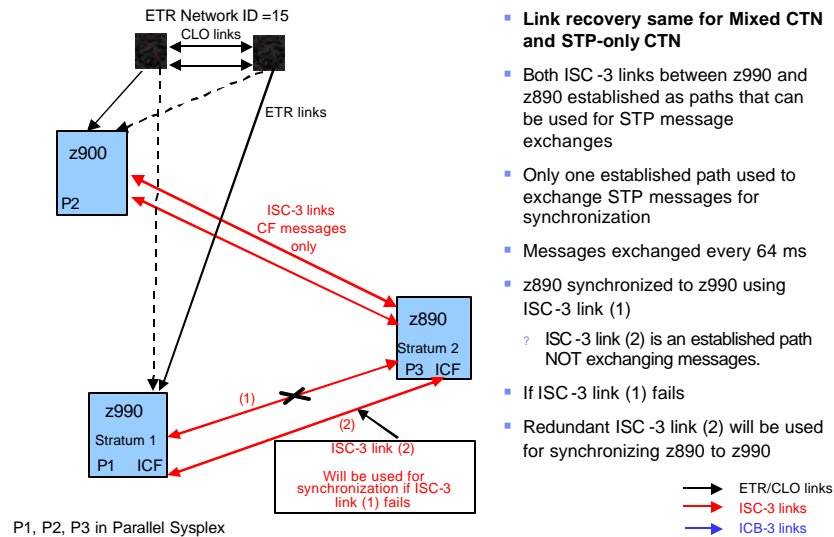
P1, P2, P3 in Parallel Sysplex
 — Active ETR link
 - - - Alternate ETR link



20

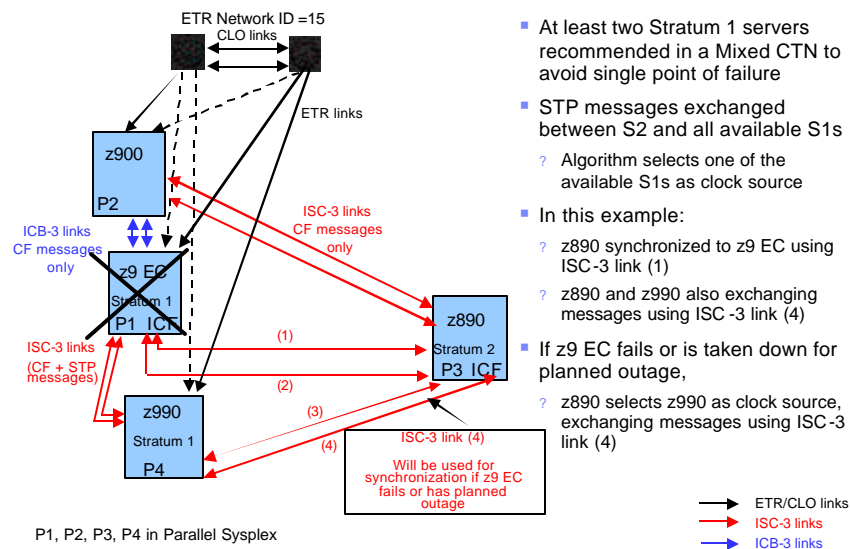
IBM Systems

Coupling link recovery – Mixed CTN example

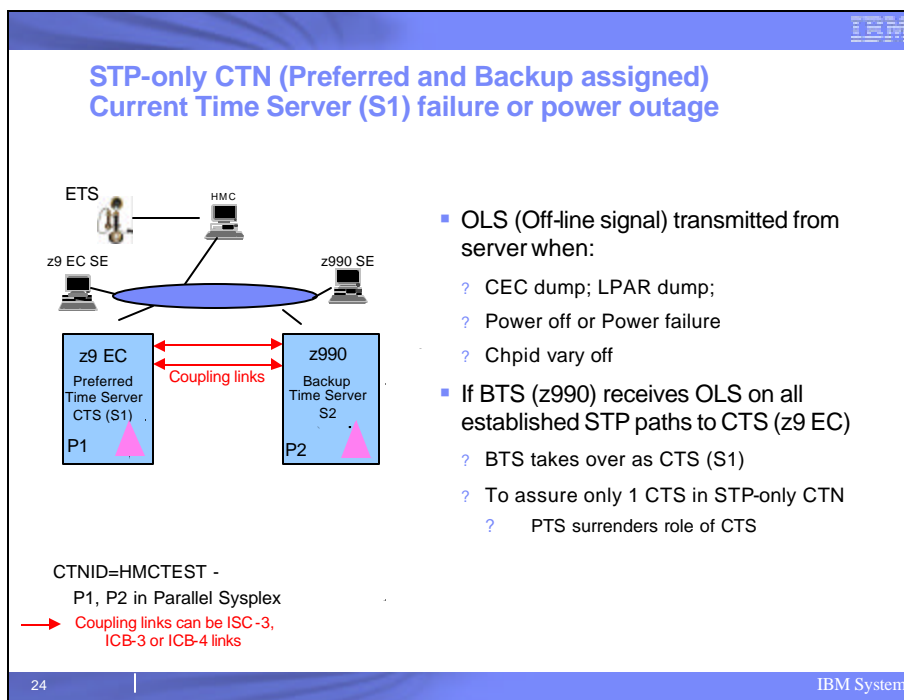
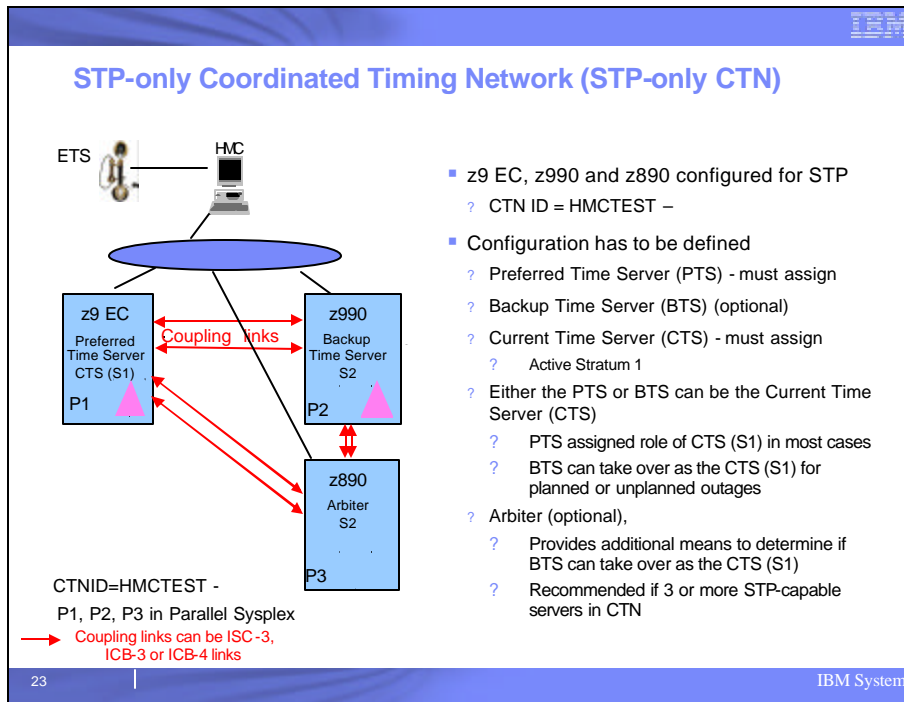


- Link recovery same for Mixed CTN and STP-only CTN
- Both ISC-3 links between z990 and z890 established as paths that can be used for STP message exchanges
- Only one established path used to exchange STP messages for synchronization
- Messages exchanged every 64 ms
- z890 synchronized to z990 using ISC-3 link (1)
 - ? ISC-3 link (2) is an established path NOT exchanging messages.
- If ISC-3 link (1) fails
- Redundant ISC-3 link (2) will be used for synchronizing z890 to z990

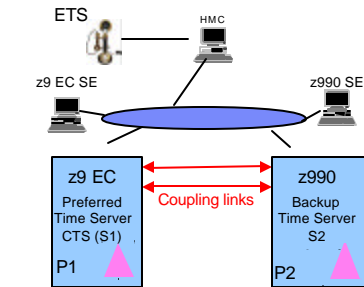
Mixed-CTN (Stratum 1 failure)



- At least two Stratum 1 servers recommended in a Mixed CTN to avoid single point of failure
- STP messages exchanged between S2 and all available S1s
 - ? Algorithm selects one of the available S1s as clock source
- In this example:
 - ? z890 synchronized to z9 EC using ISC-3 link (1)
 - ? z890 and z990 also exchanging messages using ISC-3 link (4)
- If z9 EC fails or is taken down for planned outage,
 - ? z890 selects z990 as clock source, exchanging messages using ISC-3 link (4)



STP-only CTN (Preferred and Backup assigned) Loss of communication



CTNID=HMCTEST -

P1, P2 in Parallel Sysplex

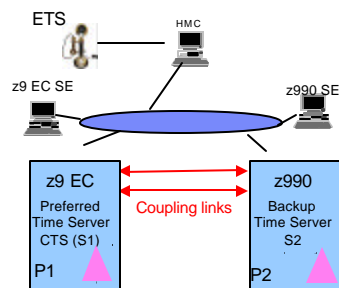
→ Coupling links can be ISC -3, ICB-3 or ICB-4 links

- Loss of communication on one Coupling link between BTS and CTS
 - ? BTS selects redundant link, if available
- OLS signal may not be transmitted for certain failures
 - ? Examples:
 - ? Channel subsystem fails
 - ? SAP recovery
 - ? All links between BTS and CTS fail
- If BTS **does not** receive OLS on all established STP paths to CTS
 - ? BTS "Freewheels"
 - ? BTS initiates "Console assisted recovery"
 - ? To determine if current CTS (PTS) has failed prior to taking over as the new CTS

25

IBM Systems

STP-only CTN (Preferred and Backup assigned) Console assisted recovery



CTNID=HMCTEST -

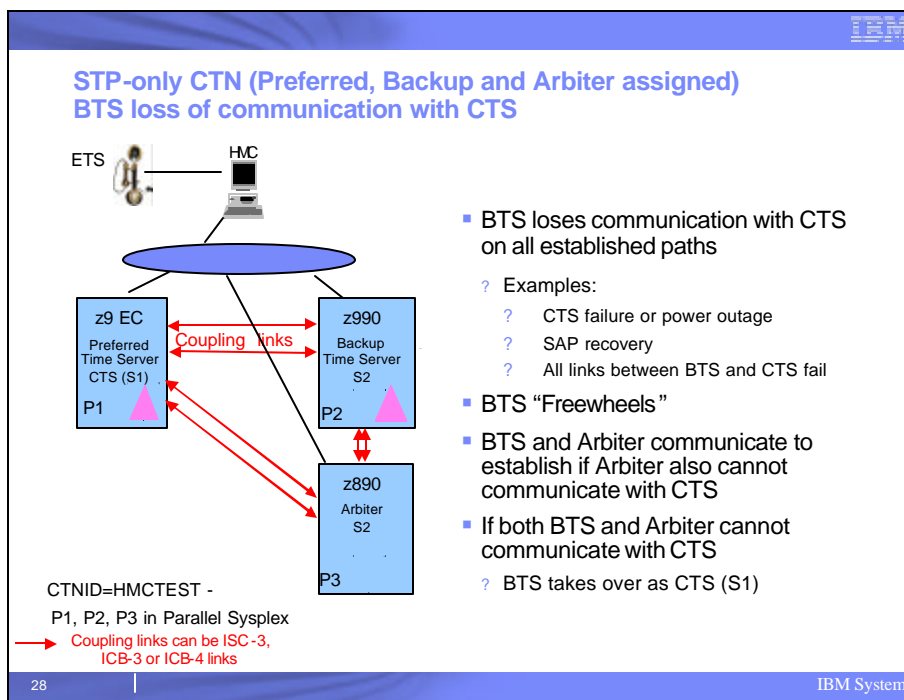
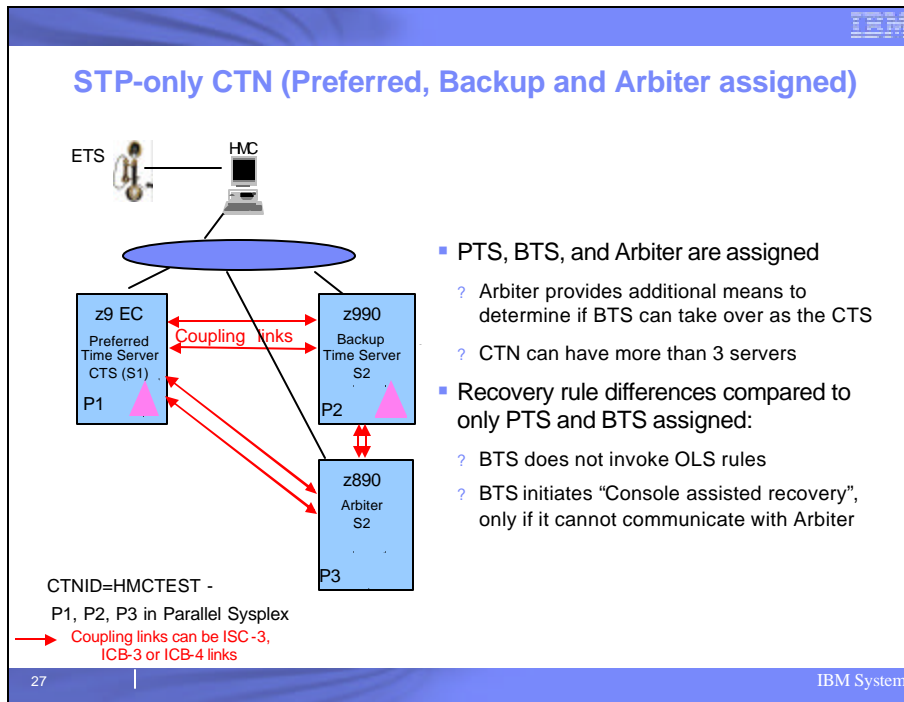
P1, P2 in Parallel Sysplex

→ Coupling links can be ISC -3, ICB-3 or ICB-4 links

- BTS (z990) sends command to its Service Element (SE) to poll the CTS (z9 EC)
- BTS (z990) SE attempts to determine state of CTS (z9 EC) by communicating via HMC with CTS (z9 EC) SE
- If CTS (z9 EC) determined to have "failed"
 - ? BTS takes over as CTS
- If CTS (z9 EC) state "good" or "indeterminate"
 - ? BTS CANNOT take over as S1
 - ? BTS eventually becomes unsynchronized at end of Freewheel Interval
 - ? z/OS systems (STPMODE YES) post WTOR

26

IBM Systems



STP-only CTN (Preferred, Backup and Arbiter assigned) BTS loss of communication with CTS (continued)

ETS HMC

z9 EC Preferred Time Server CTS (S1) P1

z990 Backup Time Server S2 P2

z890 Arbiter S2 P3

Coupling Links

CTNID=HMCTEST -
P1, P2, P3 in Parallel Sysplex

→ Coupling links can be ISC-3, ICB-3 or ICB-4 links

- BTS loses communication with CTS on all established paths
 - ? BTS "Freewheels"
- BTS and Arbiter communicate to establish if Arbiter also cannot communicate with CTS
- If Arbiter still has connectivity to CTS
 - ? BTS may synchronize to Arbiter and become S3
- If BTS cannot take over as CTS or become S3
 - ? BTS eventually becomes unsynchronized at end of Freewheel interval
 - ? All z/OS systems on BTS (STPMODE YES) post WTOR

29 IBM Systems

STP-only CTN (Preferred, Backup and Arbiter assigned) BTS loss of communication with CTS (continued)

ETS HMC

z9 EC Preferred Time Server CTS (S1) P1

z990 Backup Time Server S2 P2

z890 Arbiter S2 P3

Coupling Links

CTNID=HMCTEST -
P1, P2, P3 in Parallel Sysplex

→ Coupling links can be ISC-3, ICB-3 or ICB-4 links

- Can PTS continue as CTS when it loses communication on all links to both the BTS and Arbiter?
- Since only 1 CTS (S1) can exist,
 - ? PTS initially surrenders role of CTS
 - ? PTS initiates "Console Assisted Recovery" to determine if BTS failed or operational
 - ? If BTS determined to have failed
 - ? PTS retakes its role of CTS
 - ? If BTS state good (BTS is capable of taking over as CTS) or "indeterminate"
 - ? PTS either may become a S3 server (if connectivity to a S2 server exists) or
 - ? PTS goes unsynchronized (S0)
 - z/OS systems on PTS post WTOR

30 IBM Systems

RECOVERY CONSIDERATIONS

Site Failure Scenarios

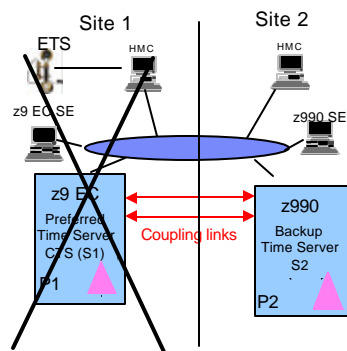
RECOMMENDATION: Provide redundant fiber trunks between sites

31

IBM Systems

STP-only CTN (Preferred and Backup assigned)

Site 1 Failure



- BTS (z990) loses all communication with CTS (z9 EC)
- ? BTS most probably does not receive OLS
- ? BTS initiates "Console assisted recovery"
- ? Results of "Console assisted recovery"
- ? CTS state most probably indeterminate
- ? **BTS eventually becomes unsynchronized at end of Freewheel Interval**
- ? z/OS systems (STPMODE YES) in site 2 post WTOR

CTNID=HMCTEST -

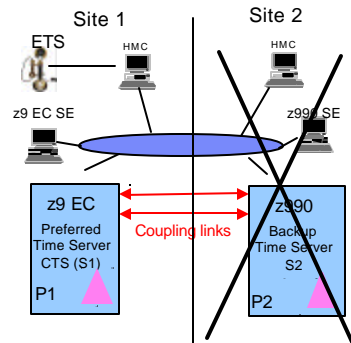
P1, P2 in Parallel Sysplex

→ Coupling links can be ISC-3, ICB-3 or ICB-4 links

32

IBM Systems

STP-only CTN (Preferred and Backup assigned) Site 2 failure



- PTS (z9 EC) continues role of CTS
- Servers in Site 1 stay synchronized to CTS (z9 EC)
- z/OS systems in Site 1 requiring STPMODE YES not affected

CTNID=HMCTEST -

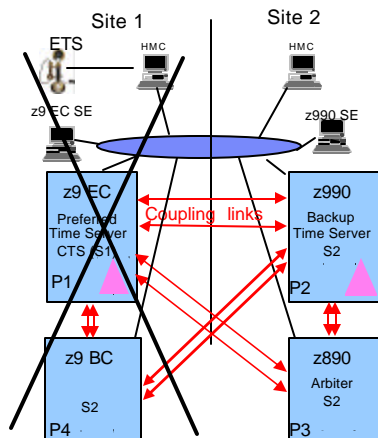
P1, P2 in Parallel Sysplex

→ Coupling links can be ISC-3, ICB-3 or ICB-4 links

33

IBM Systems

STP-only CTN (Preferred, Backup, and Arbiter assigned) Site 1 Failure – Arbiter in same site as BTS



- BTS (z990) loses all communication with CTS (z9 EC)
- BTS (z990) and Arbiter (z890) communicate to establish if Arbiter also cannot communicate with CTS
- Both cannot communicate
- BTS takes over as CTS (S1)
- z/OS systems in Site 2 requiring STPMODE YES not affected

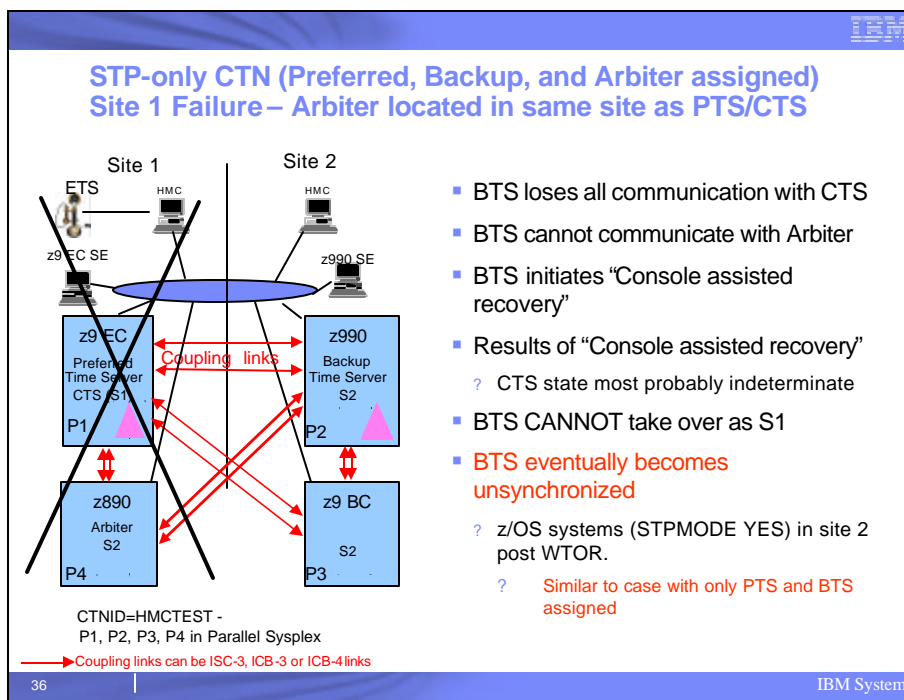
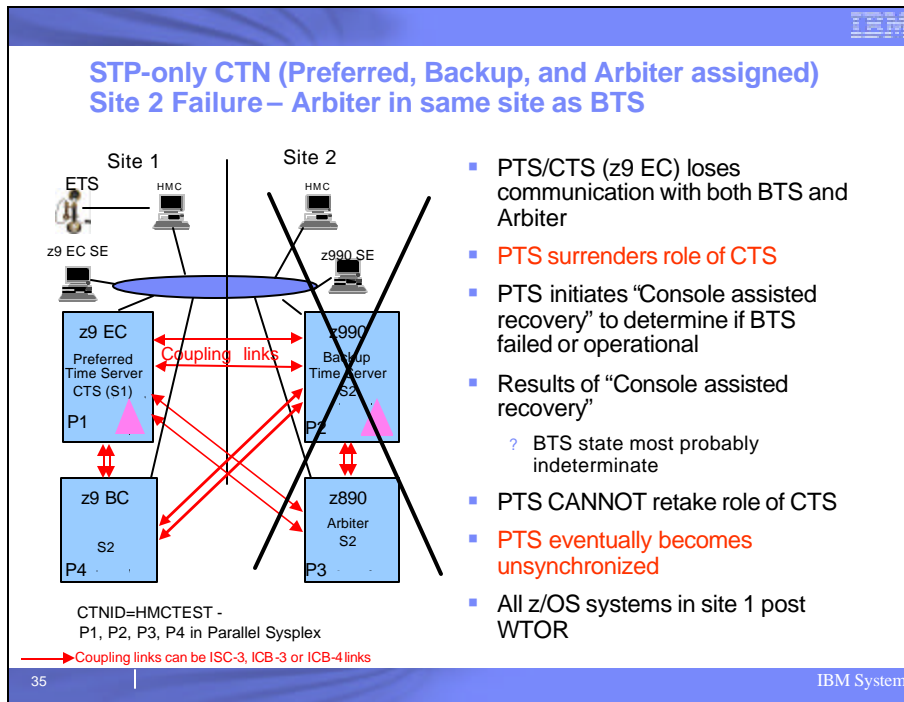
CTNID=HMCTEST -

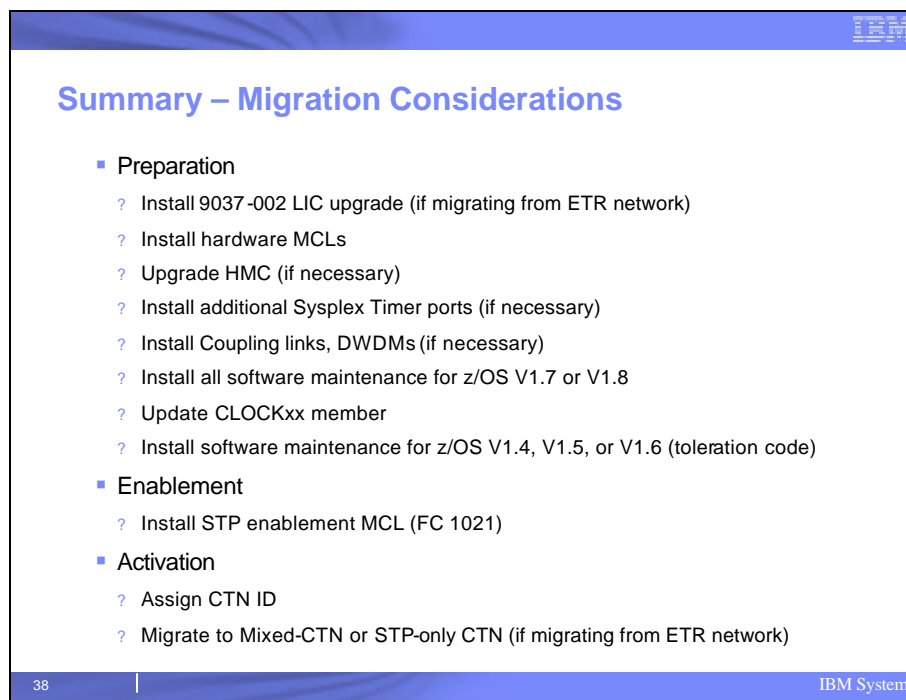
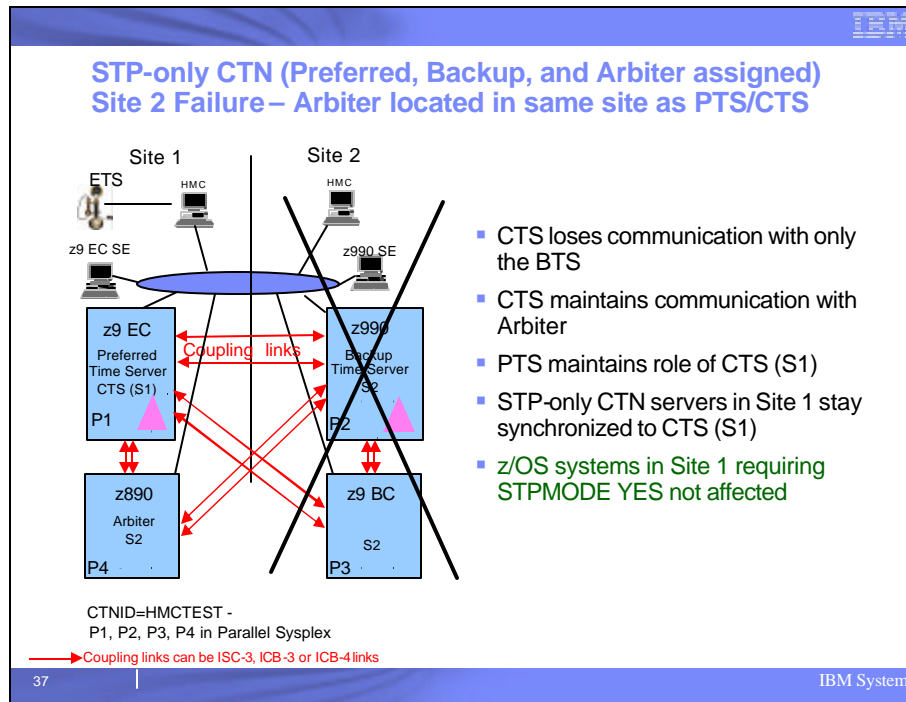
P1, P2, P3, P4 in Parallel Sysplex

→ Coupling links can be ISC-3, ICB-3 or ICB-4 links

34

IBM Systems





Summary – Recovery Considerations

■ Mixed-CTN

- ? Configure for link redundancy
- ? Attach (synchronize) at least 2 STP-configured servers to the Sysplex Timer
 - ? Multiple S1s allowed in Mixed-CTN

■ STP-only CTN

- ? Configure for link redundancy
- ? Initialize configuration with the PTS assigned as the Current Time Server
 - ? PTS, CTS must be assigned
- ? Assign at least a Backup Time Server (can take over as CTS - active S1)
- ? If 3 or more servers in CTN, assign BTS and Arbiter
- ? For configuration across 2 sites
 - ? Provide redundant routes for fiber links between sites
 - ? Locate the Arbiter in same site as PTS
 - ? Provides better recovery for scenarios when:
 - OLS may not be sent from CTS or
 - OLS may not be received by BTS

Additional Information

■ Redbooks

- ? Server Time Protocol Planning Guide SG247280
 - ? Available now
- ? Server Time Protocol Implementation Guide SG247281
 - ? Available at General Availability (GA)

■ Education

- ? Introduction to Server Time Protocol (STP)
 - ? Available on Resource Link
 - ? <https://www.ibm.com/servers/resourceLink>

■ STP website

- ? <http://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

Additional Information (BACKUP SLIDES)

41

IBM Systems

ETR Network Recovery

- **ETR link or CEC ETR port failure**

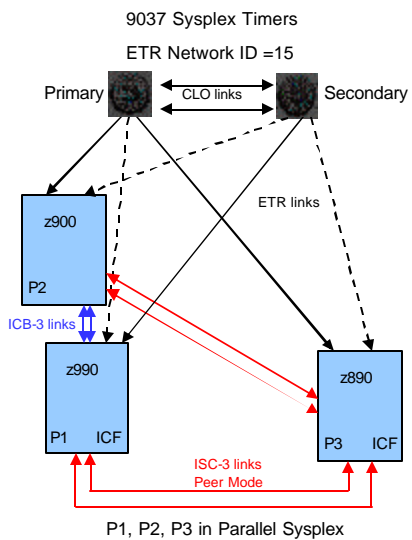
- ? When 9037 signals not received by active ETR port, the CEC switches to alternate ETR port

- **Single CLO link failure**

- ? 9037s stay in synch; continue to transmit

- **Both CLO links failure**

- ? Primary timer continues to transmit when loss of communication between Timers
- ? Secondary timer stops transmitting when loss of communication between Timers



42

IBM Systems

ETR Network Recovery (Continued)

- 9037 detects a failure/power outage

- ? 'Going away signal' Off Line Sequence (OLS) Symbol transmitted on CLO links

- OLS Recovery Rules

- ? Primary continues to transmit (if operational)
 - ? OLS received or not
 - ? If Secondary receives OLS
 - ? Secondary becomes primary
 - ? If Secondary does not receive OLS (for example when Site 1 fails)
 - ? Secondary discontinues transmission
 - ? z/OS systems in Site 2 (ETRMODE YES) post WTOR

— Active ETR link
 - - - Alternate ETR link

