Important Considerations for STP server role assignments

December 14, 2011

NOTICE:
See white papers WP102019 and WP102037 regarding relevant and updated information.
http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP102019
http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP102037

Gregory Hutchison
gkhutch@us.ibm.com

The recommendations in this White Paper apply only if you have configured a Server Time Protocol (STP) Coordinated Timing Network (CTN) with three or more servers, and have assigned the roles listed below. Recommendations for reassigning STP server roles when any of the assigned role servers has a planned or unplanned outage are documented.
NOTE: The recommendations, if not followed, MAY result in all the servers in the CTN becoming unsynchronized, a condition that results in a sysplex wide outage.

STP server roles in a CTN with three or more servers

In an STP-only CTN with three or more servers, it is recommended that the following roles be assigned to enable STP to recover from unplanned outages and provide near continuous availability:

1. Preferred Time Server (PTS), the server preferred to be the Stratum 1 (S1) server.
2. Backup Time Server (BTS), whose role is to take over as the S1 server when planned/unplanned outages affect the PTS.
3. Current Time Server (CTS), the server that is the Active Stratum 1 server. There can be only one Active S1, and only the PTS or the BTS can be assigned as the CTS. Typically the PTS is assigned as the CTS, and is therefore the Active S1. The BTS is typically the Inactive S1.
4. Arbiter, which provides a means to determine if the Inactive S1 should take over as the Active S1 when unplanned outages affect the CTN.

Figure 1, STP-only CTN with server roles, illustrates two examples of role assignments. The top example is the typical assignment of the PTS as the CTS (Active S1 server) and the bottom example is when the BTS is the CTS (Active S1 server).

Role assignment considerations and best practices are discussed in the following presentation: http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/PRS3897.

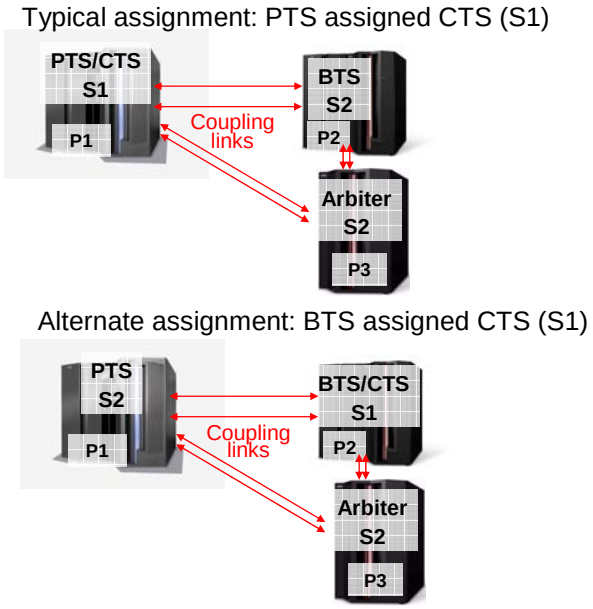

Figure 1: STP-only CTN with server roles

STP Recovery rules – PTS, BTS, Arbiter assigned

A fundamental aspect of an STP-only CTN is that the network design must make sure that it does *not* create an *island* condition, where two active Stratum 1 servers exist within the same CTN. This is because a Stratum 1 server does not synchronize to any other server's time in the CTN, but is the one all other servers are synchronized to. If two Stratum 1 servers were allowed to exist in the same CTN, the software and applications running on them would consider the two Stratum 1 servers to be synchronized, when in fact they most likely are not.

To ensure that recovery does not result in a condition of more than one Stratum 1 server, the STP recovery design in an STP-only CTN with 3 or more servers is based on the following rules, expressed in terms of Active and Inactive S1 first, and then clarified using the PTS, BTS terminology:

- If the Inactive S1 loses communication on all of its established paths to the Active S1, it attempts to determine the status of the Active S1 through the Arbiter
  - If both the Inactive S1 and the Arbiter cannot communicate with the Active S1, then the Inactive S1 becomes the new Active Stratum 1.
- If the Active S1 is still operational when it loses communication with both the Inactive S1 and the Arbiter, it can no longer be the Active S1 and becomes the Inactive S1. This is to ensure that the CTN does not have two Stratum 1 servers. If the new Inactive S1 does not have connectivity to another clock source, it will become an unsynchronized (Stratum 0 (S0)) server.

Assuming that the PTS is typically the CTS, and is therefore the S1 in the CTN, and that the main role of the BTS is to become a S1 server when the PTS is unavailable, either for planned or unplanned reasons, the above recovery rules can be rewritten in terms of the PTS and BTS terminology to state:

- If the BTS loses communication on all of its established paths to the PTS, it attempts to determine the status of the PTS through the Arbiter
  - If both the BTS and the Arbiter cannot communicate with the PTS, then the BTS becomes the Stratum 1.
- If the PTS is still operational when it loses communication with both the BTS and the Arbiter, it can no longer be the S1. This is to ensure that the CTN does not have two Stratum 1 servers. If the PTS does not have connectivity to another clock source, it will become unsynchronized (Stratum 0 (S0)).

Figure 2, STP-only CTN recovery after PTS failure, is an example of the BTS taking over as the CTS after a failure of the PTS.

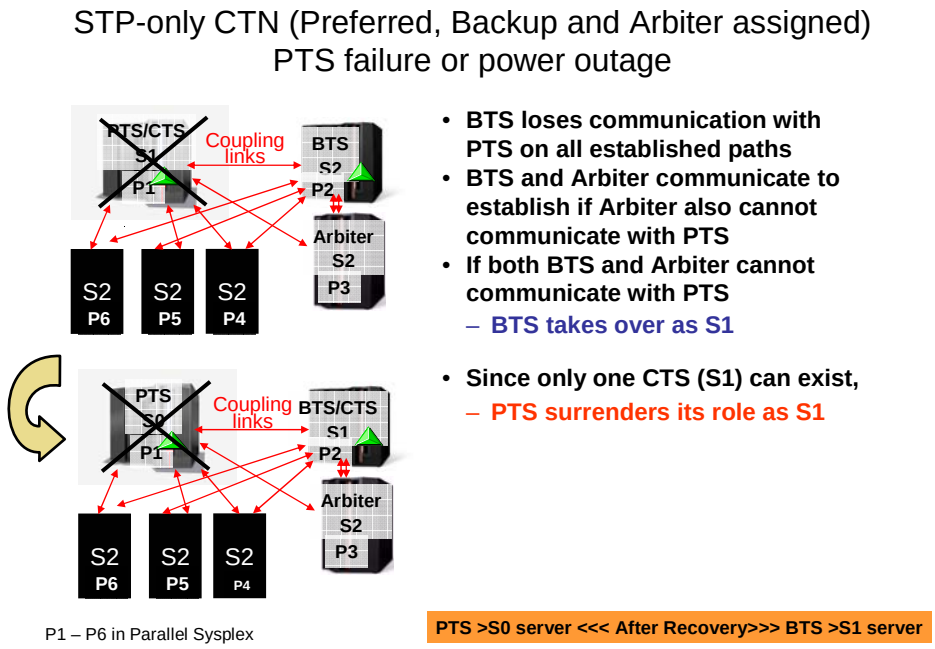

Figure 2: STP-only CTN recovery after PTS failure

Assumptions:

In order to keep the complex discussion that follows as understandable as possible, for the remainder of this document it will be assumed that the PTS has been assigned as the CTS, and the BTS is a Stratum 2 server capable of becoming the CTS. The terminology of Active S1 and Inactive S1 will not be used, but you should bear in mind that you can have only one CTS in the CTN and it can be either the PTS or BTS.

Recommended Server role reassignments – planned operations

STP has been designed to block a planned disruptive operation, if it is attempted on a PTS (assuming it is the CTS) with initialized coupling links to at least one Stratum 2 server in the CTN. Examples of disruptive operations are activate, deactivate, power off, power-on reset, disruptive switch to alternate SE, and so on.

STP code **does not block** planned disruptive operations on the BTS or Arbiter. Moreover, there is no message posted on the Hardware Management Console (HMC), warning you that the server has a role in an STP-only CTN.

Therefore, it is very important to modify your operational procedures to implement the following recommendations, which should be performed **prior** to performing planned disruptive operations affecting either the BTS or Arbiter.

1. Determine if the server that you are planning to perform the disruptive operation on is either the BTS or the Arbiter.
2. If the server is either a BTS or the Arbiter, the role should either be reassigned to another server in the CTN if possible or the role should be removed by reassigning it as "Not Configured". In other words,
    - Reassign the PTS, BTS, Arbiter roles if there are available servers and the connectivity exists to do so or
    - Reassign only the PTS and BTS roles if either there is no server available to assign an Arbiter or the connectivity does not exist to assign an Arbiter or
    - Reassign only the PTS as the CTS if no other server can perform the role of the BTS.

Note: After performing the planned disruptive actions, make sure the PTS, BTS, Arbiter roles are reassigned for the normal operational state of the CTN.

Role reassignment example – planned disruptive action on BTS

Figure 3, STP-only CTN example – configuration before planned disruptive action, shows an STP-only CTN during normal operation with six servers, and the roles of PTS, BTS, and Arbiter assigned as shown. P1, P2, P3, P4, P5, and P6 are z/OS systems in a Parallel Sysplex running on the indicated servers.

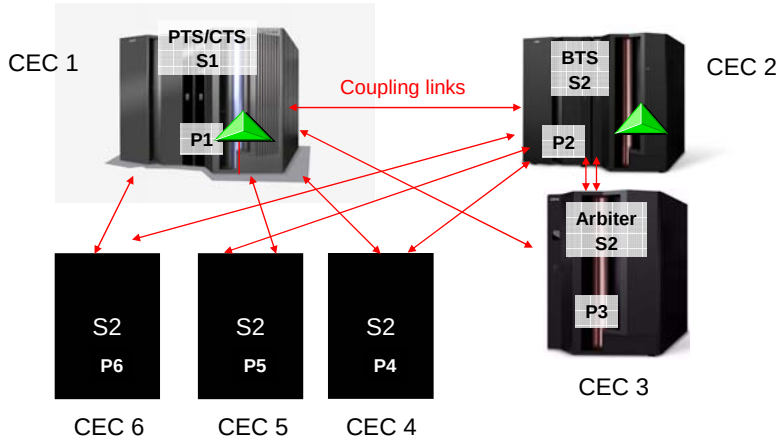

Figure 3: STP-only CTN example – configuration before planned disruptive action

A disruptive action (POR) has been planned on the BTS. Prior to performing the POR on the BTS, the BTS role was reassigned from CEC 2 to CEC 3. The Arbiter role could not be reassigned because even though there were sufficient servers available to assign the role, the connectivity was not available between the reassigned BTS and one of the other S2 servers to reassign the Arbiter. Note also that the BTS assignment does not provide any fault protection for the remaining CECs in the CTN, because of lack of connectivity from the new BTS to the other CECs. If the PTS were to fail during the planned outage of CEC 2, even though CEC 3 would become the new S1, CECs 4, 5, and 6 would become unsynchronized.

See *Figure 4: STP-only CTN example – reassignment before planned disruptive action.*

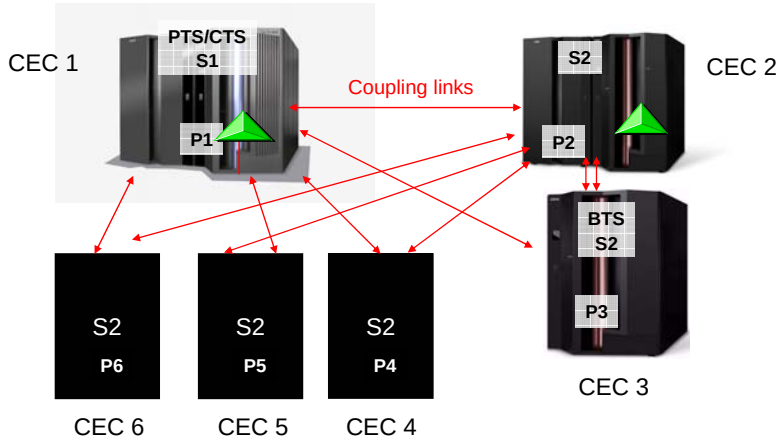

Figure 4: STP-only CTN example – reassignment before planned disruptive action

CEC 2 can be PORed after the reassignment has been completed.

Planned disruptive action example – roles not reassigned

Consider the following example which illustrates the consequences of **not following the above recommendations prior to performing planned disruptive actions** on the BTS and Arbiter.

Again using Figure 3 as the STP-only CTN example, a planned disruptive action (POR) was performed on the BTS **without following the above recommendation to reassign the BTS role or remove it.** This resulted in the PTS losing communication with the BTS. While the BTS was still non-operational, the Arbiter was PORed for a different planned disruptive action – **also without either reassigning or removing the role**. This resulted in the PTS also losing communication with the Arbiter.

Following the above recovery rules, since the PTS lost communication with both the BTS and Arbiter, it could no longer continue as a S1 server. Not only did the PTS become an unsynchronized (S0) server, but the 3 other servers (CECs 4, 5, and 6) hosting z/OS images P4, P5, P6 lost their clock source and therefore also became S0 servers. This resulted in a **sysplex wide outage** with z/OS systems P1, P4, P5, P6 posting WTOR IEA394A.

Figure 5, Unsynchronized STP-only CTN after planned disruptive actions, illustrates the result.

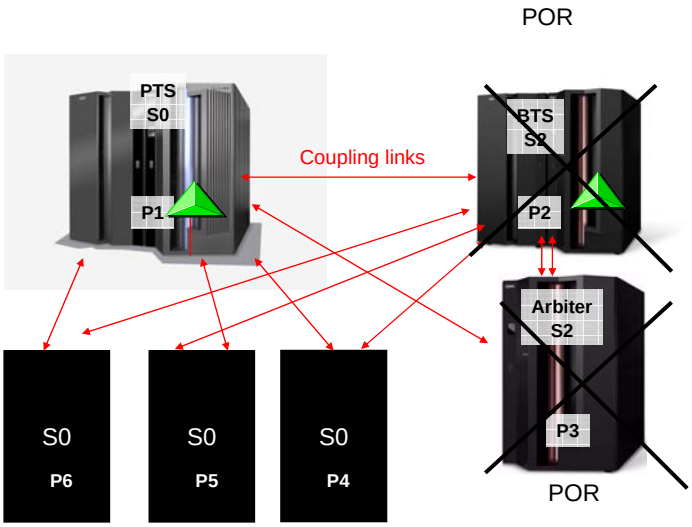

Figure 5: Unsynchronized STP-only CTN after planned disruptive actions

Recommended Server role reassignments – unplanned failure

If a failure condition has resulted in either the PTS or BTS or Arbiter to be no longer an operational synchronized server in the CTN, perform one of the following actions **after** the failure to maintain a resilient STP-only CTN:

- Reassign the PTS, BTS, Arbiter roles if the connectivity exists to do so or
- Reassign only the PTS and BTS roles if either there is no server available to assign an Arbiter or the connectivity does not exist to assign an Arbiter or
- Reassign only the PTS as the CTS if no other server can perform the role of the BTS.

Notes:
A failure of the PTS most likely will result in the BTS becoming the new S1. After the recovery action, reassignments should be performed using the remaining servers in the CTN, as soon as possible after the failure. See the section “Role reassignment example – after PTS failure for an example.
A failure of the BTS or Arbiter will not result in changing which server is the S1. Reassignments should still be performed using the remaining servers in the CTN as soon as possible after the failure.
After the failure condition is repaired, make sure the PTS, BTS, and Arbiter roles are reassigned for the normal operational state of the CTN.

Role reassignment example – after PTS failure

In the STP-only CTN example shown in *Figure 6, STP-only CTN after BTS takeover as S1*, a checkstop (failure) of the PTS (CEC 1) has resulted in an STP recovery action, and CEC 2, the BTS has taken over the role of S1.

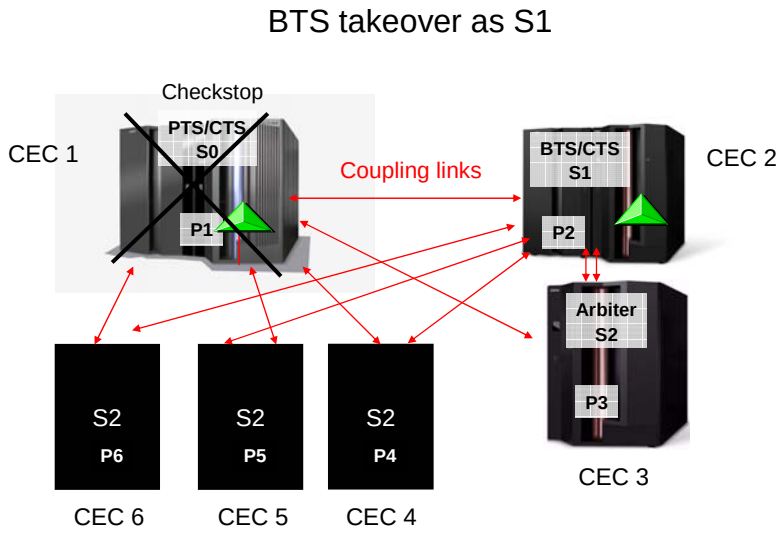

Figure 6: STP-only CTN after BTS takeover as S1

A reassignment is recommended after this recovery action. The following reassignment was performed:

- Reassign CEC2 the PTS/CTS, since it is the only CEC that has connectivity to CECs 3-6.
- Reassign either CEC 3, 4, 5, or 6 as the BTS. Note that the BTS assignment does not provide any fault protection for the remaining CECs in the CTN, when the BTS takes over as the S1 because of lack of connectivity from the BTS to the other CECs.
- An Arbiter cannot be reassigned, because an Arbiter needs connectivity to the BTS and PTS and the required connectivity does not exist.

Figure 7, Reassignment after BTS takeover as S1, illustrates the result.

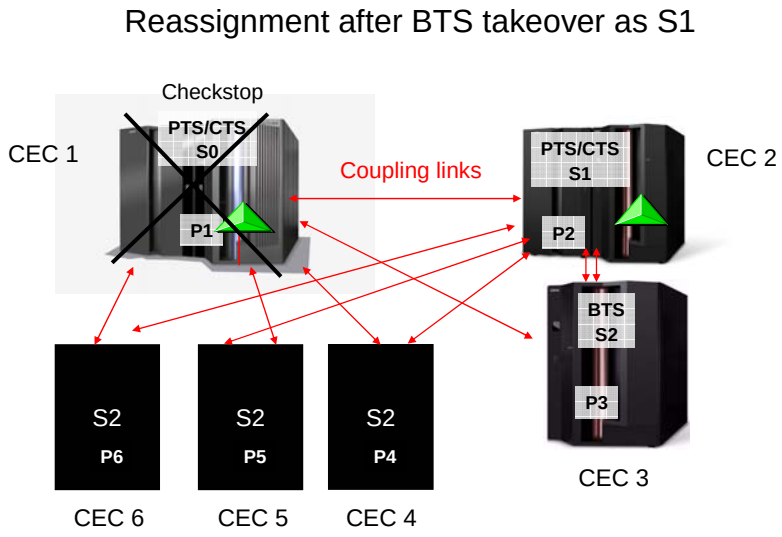

Figure 7, Reassignment after BTS takeover as S1

Unplanned failure example – roles not reassigned

Consider again the same example of a failure in an STP-only CTN shown in *Figure 6: STP-only CTN after BTS takeover as S1*. A failure of the server with the PTS role resulted in STP recovery and the BTS took over as the new S1 server. Note that after the takeover the BTS does not have connectivity to the PTS, since it has failed.

The roles were not reassigned as recommended above. If subsequently, the Arbiter requires a planned disruptive operation or it fails, the BTS will also lose connectivity to the Arbiter.

Following the STP design recovery rules, since the BTS (the new Active S1 (CTS)) lost communication with both the PTS and the Arbiter, it could no longer continue as a Stratum 1 server. Not only did the BTS become an unsynchronized (S0) server, but the 3 other servers (CECs 4, 5, and 6) hosting z/OS images P4, P5, P6 lost their clock source and therefore also became S0 servers. This results in a sysplex wide outage with z/OS systems P2, P4, P5, P6 posting WTOR IEA394A. *Figure 8, Unsynchronized CTN after BTS loses connectivity to Arbiter*, illustrates the result.

Note that a sysplex outage would have resulted if the BTS had failed, since there was no other server defined to take over as the S1.

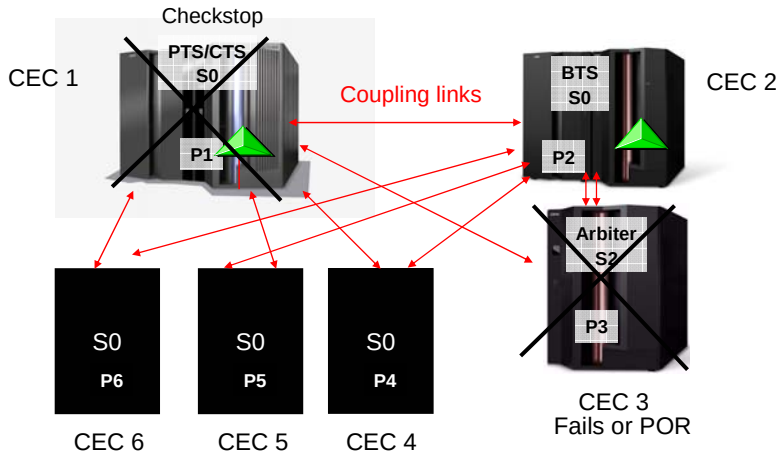

Figure 8: Unsynchronized CTN after BTS loses connectivity to Arbiter

Summary

An STP-only CTN can have only one Stratum 1 server to maintain data integrity. To ensure that recovery does not result in a condition of more than one Stratum 1 server, the STP recovery design in an STP-only CTN with 3 or more servers requires that the Active Stratum 1 server surrender its role as the Current Time Server, when it loses connectivity to the Inactive S1 and the Arbiter.

It is therefore strongly recommended that STP roles of PTS, BTS, and Arbiter be reassigned or removed prior to planned disruptive operations or after a failure affecting one of the role servers.