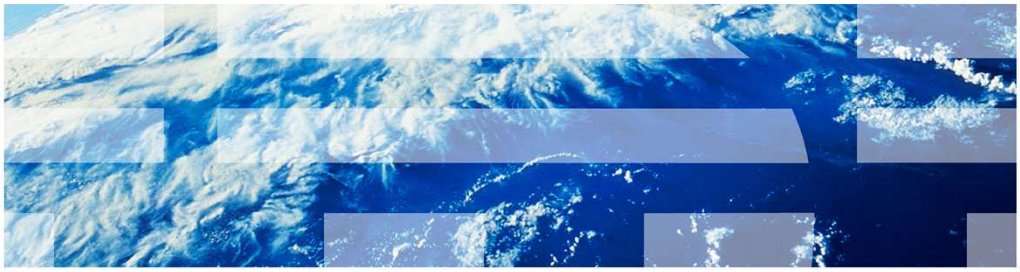
z/OS V2R1 Communications Server

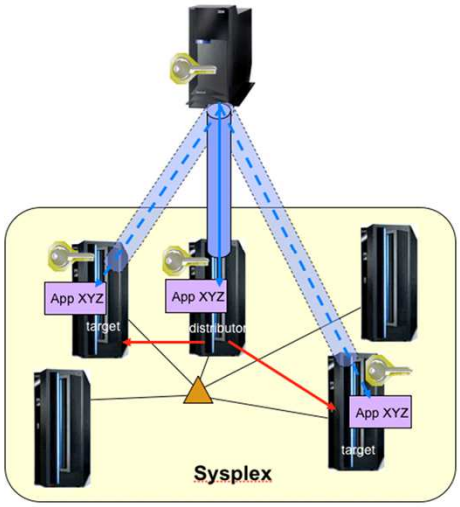
Availability



This presentation provides an overview of the enhancements in z/OS® V2R1 Communications Server for availability.

Background - Sysplex-Wide Security Associations

- Sysplex-Wide Security Associations allow IPSec-protected traffic to be distributed through a sysplex
- Security associations (SA) and characteristics are moved and distributed through the sysplex to the target stacks
  - VIPA Takeover
    - Ability of an SA to follow a DVIPA when it moves from one stack to another
  - Critical information stored in the Coupling Facility



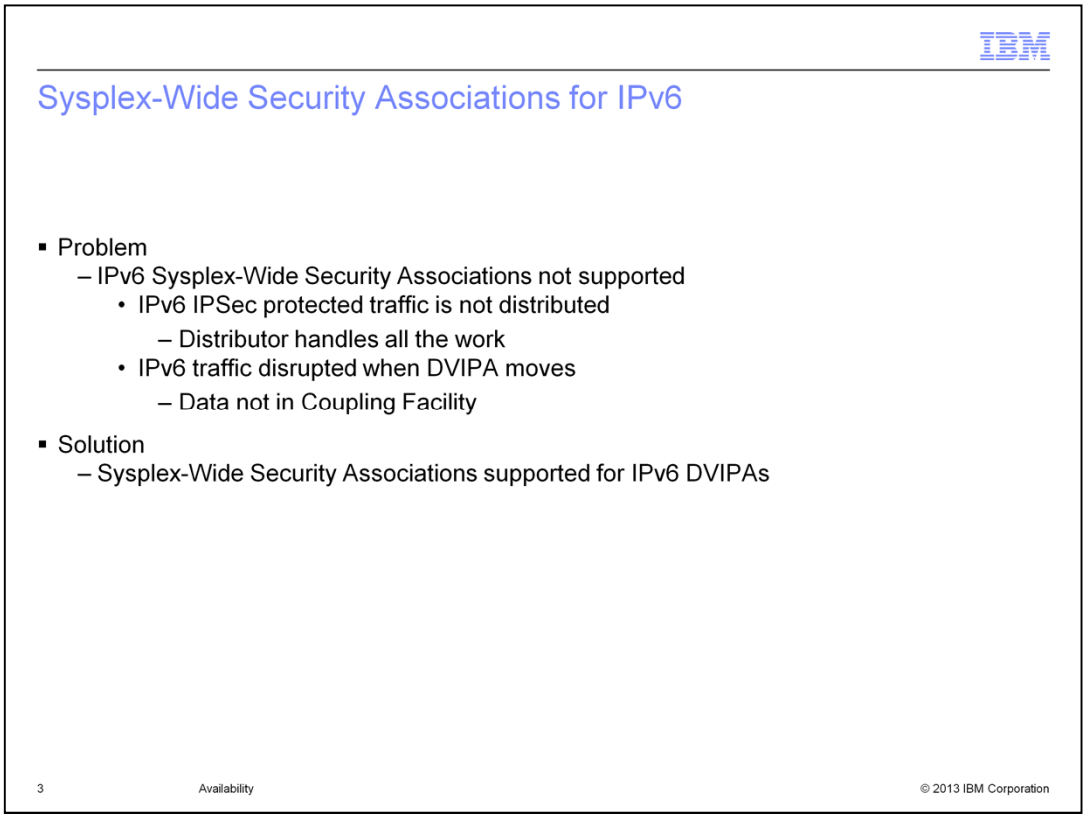
2 Availability

Sysplex-Wide Security Associations represent the intersection of two z/OS Communications Server functions. The first, IPSec, protects network data using Security Associations (SAs). The second, Sysplex Distributor, distributes application workload and provides backup and recovery mechanisms for Dynamic Virtual IP addresses (DVIPAs).

In an IPv4 network, Sysplex-Wide Security Associations allow you to exploit both functions together. With Sysplex-Wide Security Associations, you can encrypt a Sysplex Distributor workload. The distributor is responsible for negotiating an SA with a remote host. Copies of the SA, known as shadow SAs, are sent to any target stacks that can potentially receive workload for the DVIPA. In addition to distribution, you can recover SAs associated with DVIPAs that are migrated to an alternate TCP/IP stack (a DVIPA takeover). Any stacks that back up the DVIPA do not receive SA data from the distributor directly, but have access to SA data in the Coupling Facility needed for SA recovery if a DVIPA moves.

IKED must run on any stack that potentially negotiates SAs, including the distributor and backup stacks. Target stacks are not required to run IKED.

Critical SA data that is shared among Sysplex members is stored in the Coupling Facility for SAs that are distributed, or are candidates for takeover. This data includes phase one identities of the IKE peers. It also includes the IPSec protocol of the phase two SA and sequence numbers for the SA. Also, the protocol and TCP/UDP ports for the SA are included.

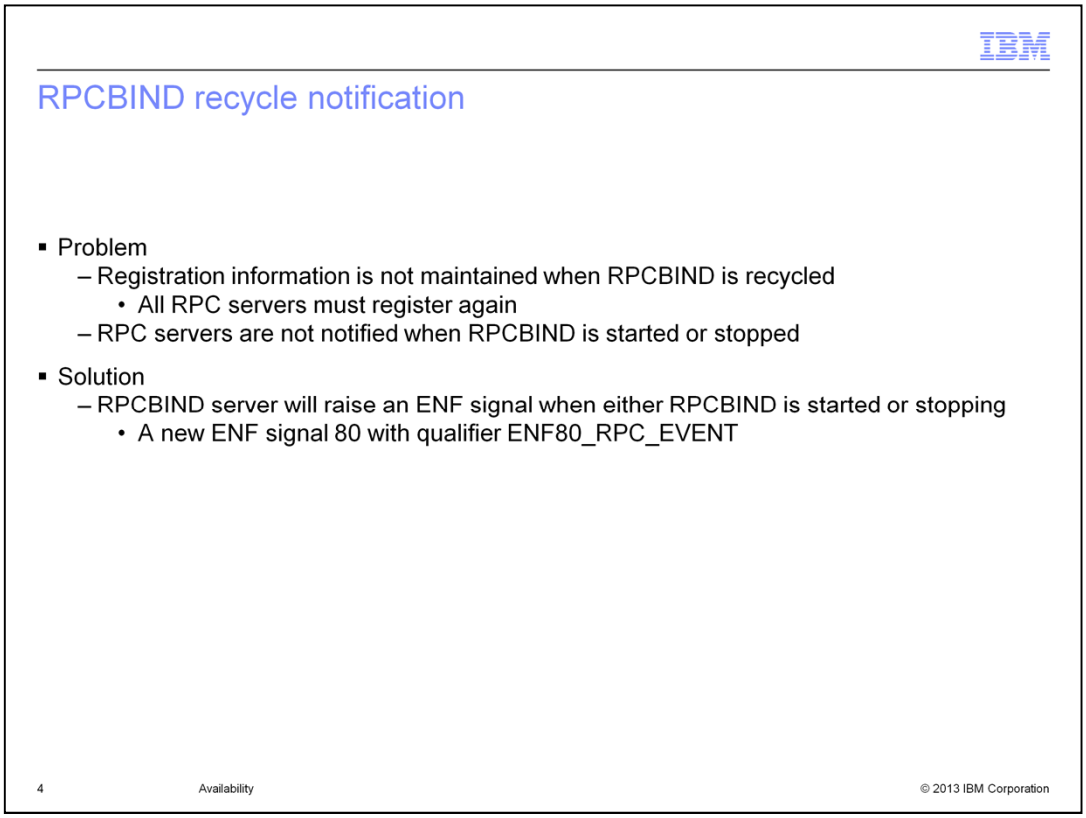


Before V2R1, IPSec tunnels for IPv6 DVIPAs were ineligible for sysplex distribution and takeover. A distributor that owned an IPv6 DVIPA can negotiate an SA with a remote peer. However, shadow SAs were not created or distributed for tunnels whose endpoint was an IPv6 DVIPA. Consequently, all workloads whose endpoint was an IPv6 DVIPA were forced to be serviced at the distributing stack only.

Furthermore, no data for IPv6 tunnels were stored in the Coupling Facility. Therefore, in the event that the IPv6 DVIPA moved because of a planned or unplanned takeover, the backup stack had no way to recover the lost SA. Any IPSec traffic to that IPv6 DVIPA was disrupted.

Starting in z/OS V2R1, IPSec tunnels for IPv6 DVIPAs can be used for sysplex distribution and takeover. When a distributor negotiates an SA for an IPv6 DVIPA, shadow SAs are distributed to all eligible targets.

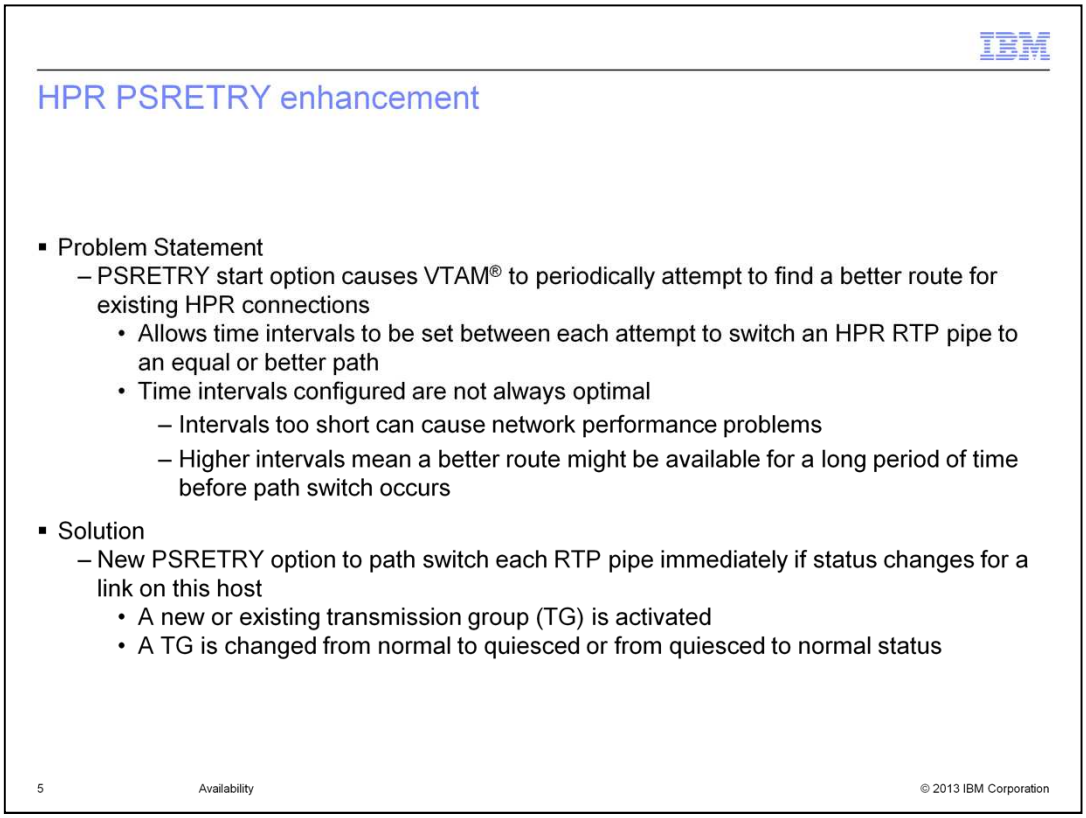
An eligible target must be at release level V2R1 or higher. Connections are also distributed among eligible targets. A target for the connection must be at release level V2R1 or higher to be eligible for an IPv6 connection protected by IPSec. Note that there are additional conditions that are checked to determine a target's eligibility, such as having a listener open on the target.



RPCBIND allows RPC services to register with it and respond to RPC client requests asking where an RPC service is registered. When the RPCBIND server is recycled, all registration information for the RPC services is lost. The RPC service must register again with RPCBIND.

Starting in V2R1, the RPCBIND server will now raise an ENF signal when either RPCBIND is started or is stopping. The RPCBIND server will send an ENF signal when it has started and is prepared to accept registrations from RPC applications. RPC applications can monitor this ENF signal and register again with the RPCBIND server if it is stopped and restarted for any reason.

The RPCBIND server will send an ENF signal when it is stopped or cancelled. RPC applications can monitor this ENF signal and take action when the RPCBIND server is not available to RPC clients. RPCBIND implements a new ENF signal for event code 80, with qualifier ENF80_RPC_EVENT indicating when RPCBIND has completed initialization or is terminating.



PSRETRY is an existing VTAM start option that allows you to set a time interval from 1 minute to 24 hours between each automatic attempt to switch an HPR rapid transport protocol (RTP) pipe to another path. That path might be equal to or better than the current path. Whether the path switches to an equal weight route or switches only if the other path is better is controlled by the PSWEIGHT start option. A different PSRETRY interval can be specified for each of the four defined transmission priorities: low, medium, high, and network. By default, PSRETRY is set to (0,0,0,0), which means that VTAM will not make periodic attempts to find a better path for each RTP pipe.

To avoid network performance problems, especially in configurations that use a large number of RTP pipes, the time intervals specified on PSRETRY should not be set too short. However, setting the intervals to longer periods of time means it might take a relatively long time after a better route becomes available before RTP finds it and switches to the new route. This means that RTP pipes can remain on worse-performing routes for a considerable time, even though a better route can be used.

A path switch always occurs when a TG is inactivated or INOPs. Starting in z/OS V2R1, when the status of a local link (TG) changes, PSRETRY will path switch to an equal or better route. It does this for each RTP pipe immediately rather than waiting for the PSRETRY timer to expire. This provides quicker path switches to better-performing routes than the current timer-only mechanism. These TG status changes include:

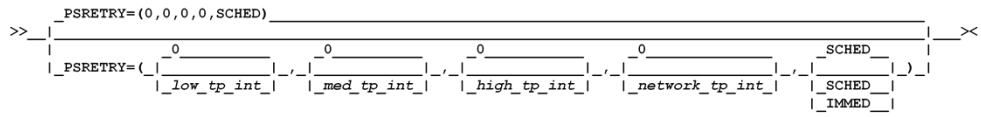
A new or existing TG is activated,

The weight of a TG changes or

A TG is changed from normal to quiesced or from quiesced to normal status.

Usage and invocation - HPR PSRETRY enhancement

- To invoke this function, configure IMMED on the PSRETRY start option



  - PSRETRY=(0,0,0,0,IMMED)
    - Disables scheduled path switch retry attempts to reduce the performance impact
    - Still triggers path switch retry attempts when local links are activated or change status
    - Will not trigger path switch retry attempts if an intermediate link is activated or changes status
- PSRETRY start option, including the SCHED or IMMED value, is modifiable with MODIFY VTAMOPTS command

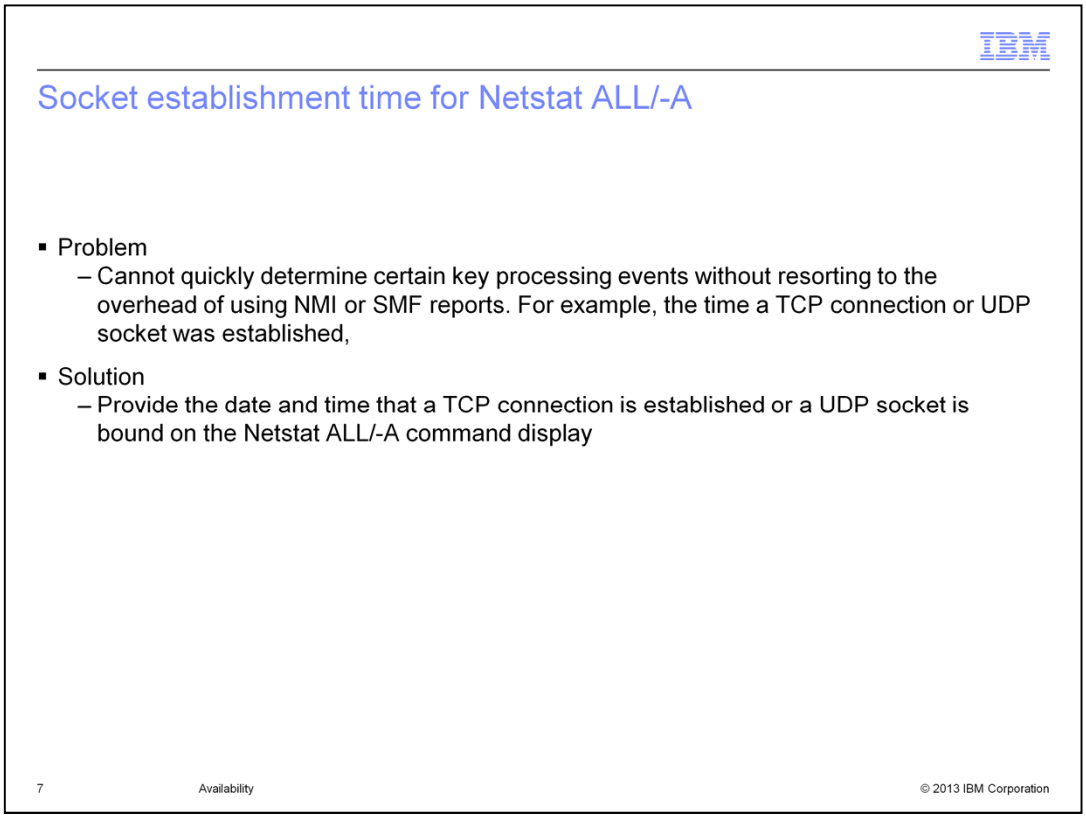
The new PSRETRY value specifies whether an immediate attempt to switch RTP sessions to another path is triggered when a link on this host is activated or changes status, regardless of the specified time intervals. The value can be set to IMMED or SCHED. In both cases, the path switch retry function also continues to follow the scheduled transmission priority time intervals, as it has in prior releases.

IMMED specifies that a link activation or status change triggers an attempt to switch RTP sessions to another path. With the IMMED value specified, the result includes both the PSRETRY behavior provided in prior releases and the new behavior introduced with the IMMED function.

SCHED specifies that a link activation or status change does not trigger an attempt to switch RTP sessions to another path. With the SCHED value specified or allowed to default, the path switch retry function represents the same behavior as PSRETRY provided in previous releases.

This slide includes the syntax of the PSRETRY start option with the new value of SCHED or IMMED. The default value is SCHED.

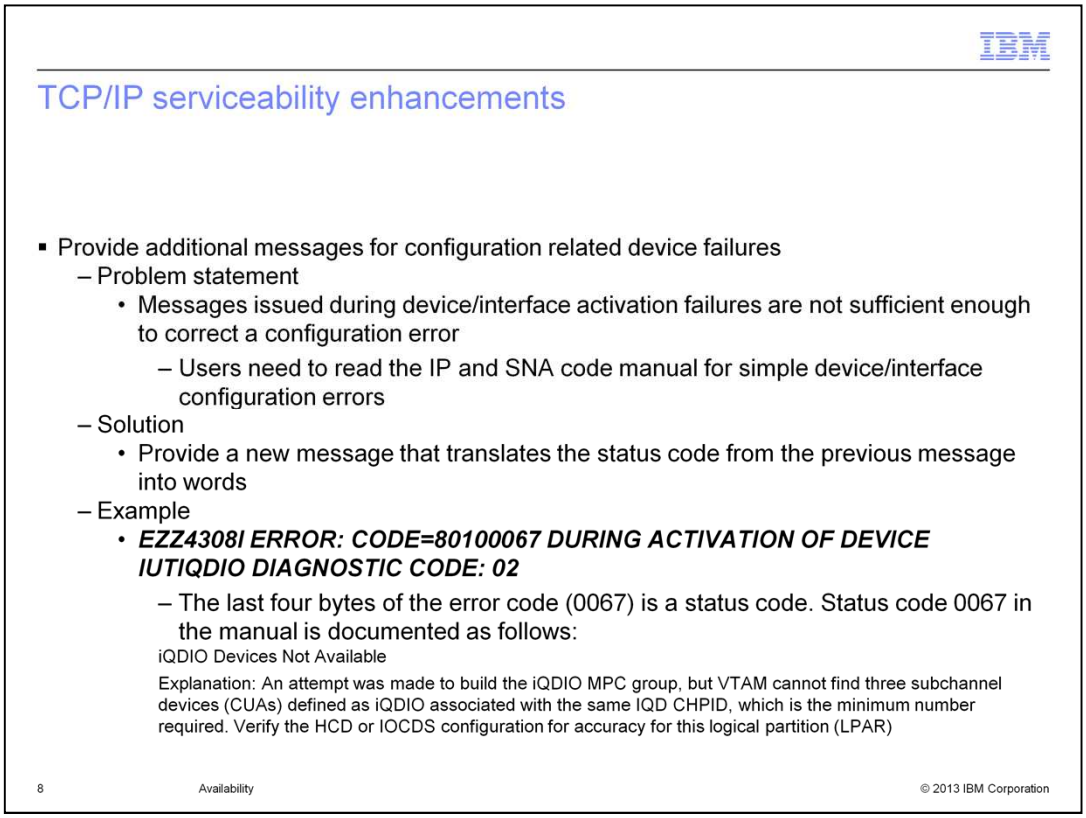
You can set the PSRETRY values to (0,0,0,0,IMMED) to disable regularly scheduled path switch retry attempts to reduce the performance impact, but still trigger the path switch immediate retry function. However, keep in mind that the immediate retry function is triggered only when a local link is activated or changes status. If you have RTP pipes that traverse intermediate links, status changes or activation of intermediate TGs will not trigger an immediate path switch attempt. To detect changes in the middle of a route, you must still perform periodic PSRETRY processing by setting non-zero time intervals. You can also modify the start option, including the SCHED or IMMED value, with the MODIFY VTAMOPTS command.



Problem determination can be related to the time a TCP connection or UDP socket is established. Connection distribution based on load balancing can depend on loads at the time of connection establishment. Delays in connection establishment occur because of system conditions at the time.

System Management Facility records and Network Management Interface information allow TCP connection start time and UDP socket start time to be gathered through their respective interfaces. However, before V2R1, there was no way to quickly and easily determine the time a TCP connection or UDP socket was established without resorting to the overhead of NMI or SMF reports.

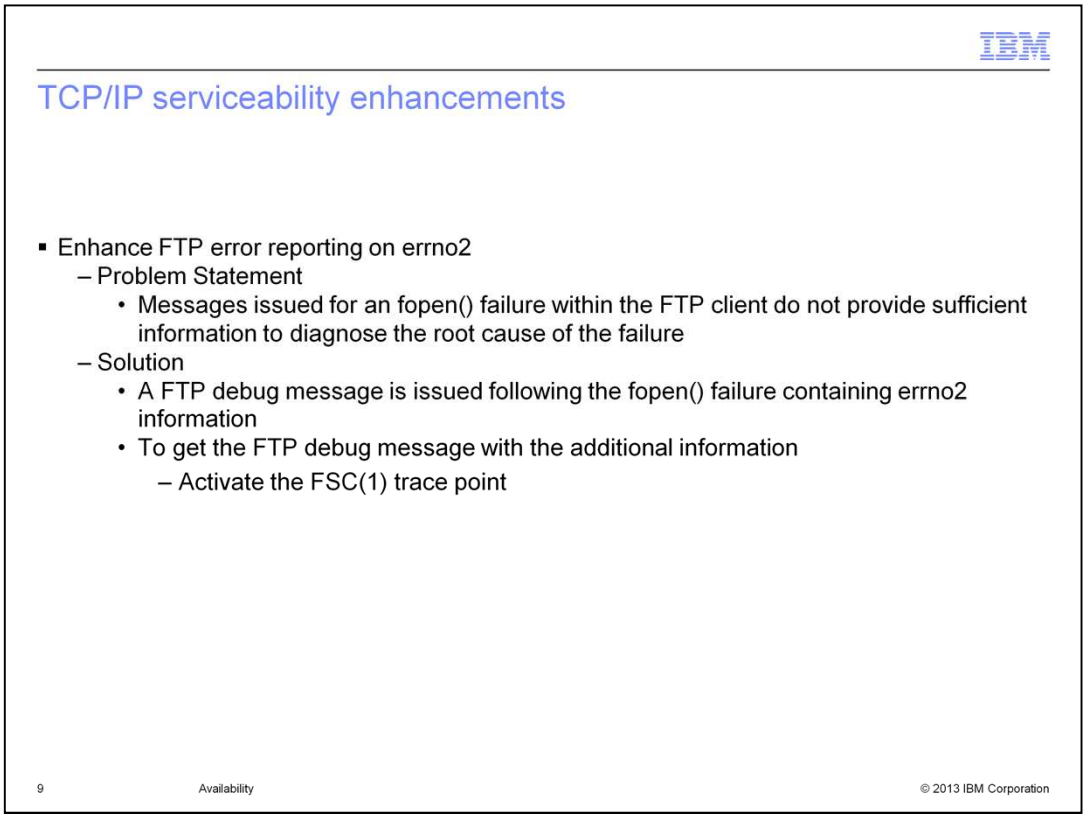
The Netstat ALL / -A report contains detailed information about TCP connections and UDP sockets. The purpose of this report is to aid in debugging problems with TCP connections and UDP sockets. Starting in V2R1, the output of the Netstat ALL / -A command will include the date and time that a TCP connection is established or a UDPsocket is bound.



When a device or interface activation fails, the messages reporting the failure contain error codes that identifies the problem. The error codes are documented in the IP and SNA codes manual.

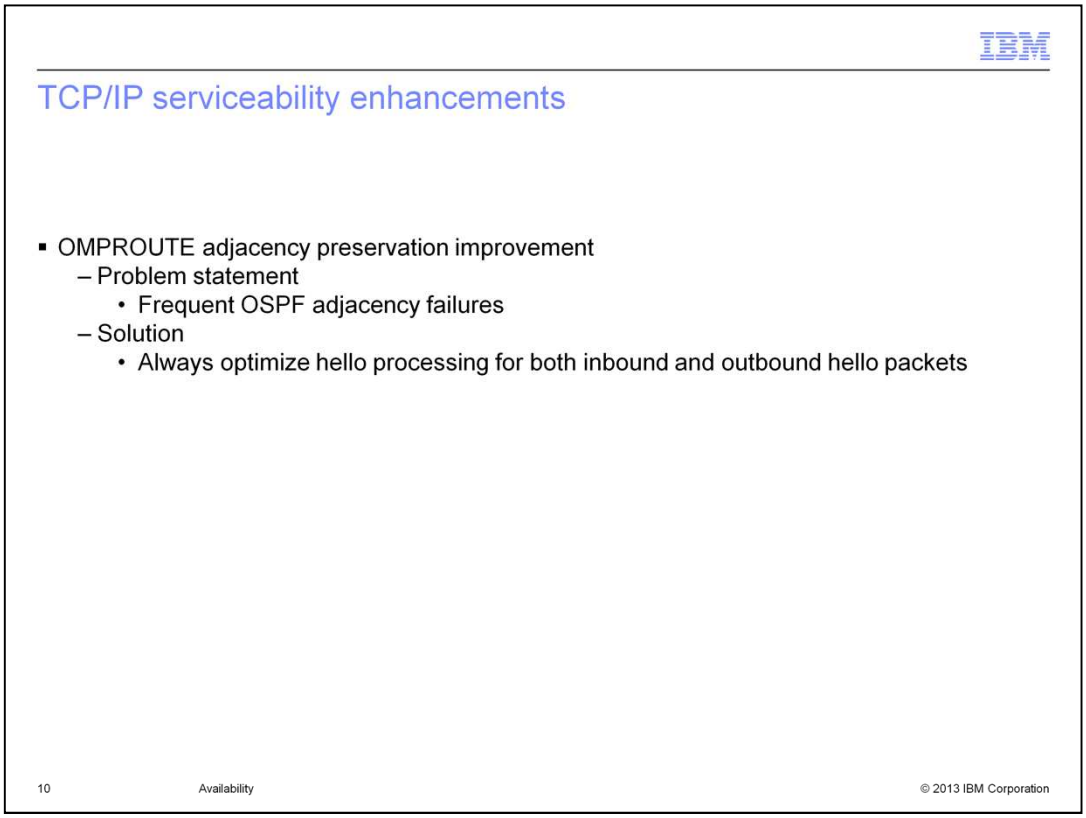
The current messages issued during device or interface activation failures are not sufficient to correct a configuration error. You need to read the IP and SNA code manual for simple device or interface configuration errors.

Starting in z/OS V2R1 Communications Server, a new message, EZD2028I, is issued after the existing error message. The new message translates the status code from the previous message into words. This enhancement reduces the need to read the IP and SNA codes manual for problem determination for simple configuration errors.



Currently in FTP there are two messages that are displayed in the event of a data set open or allocation failure: EZA2564W and EZA1735I. EZA2564W provides the name of the data set involved in the error, and EZA1735I provides standard error and return codes about the nature of the failure. However, the codes that are provided on EZA1735I are vague, and can apply to a large number of root causes.

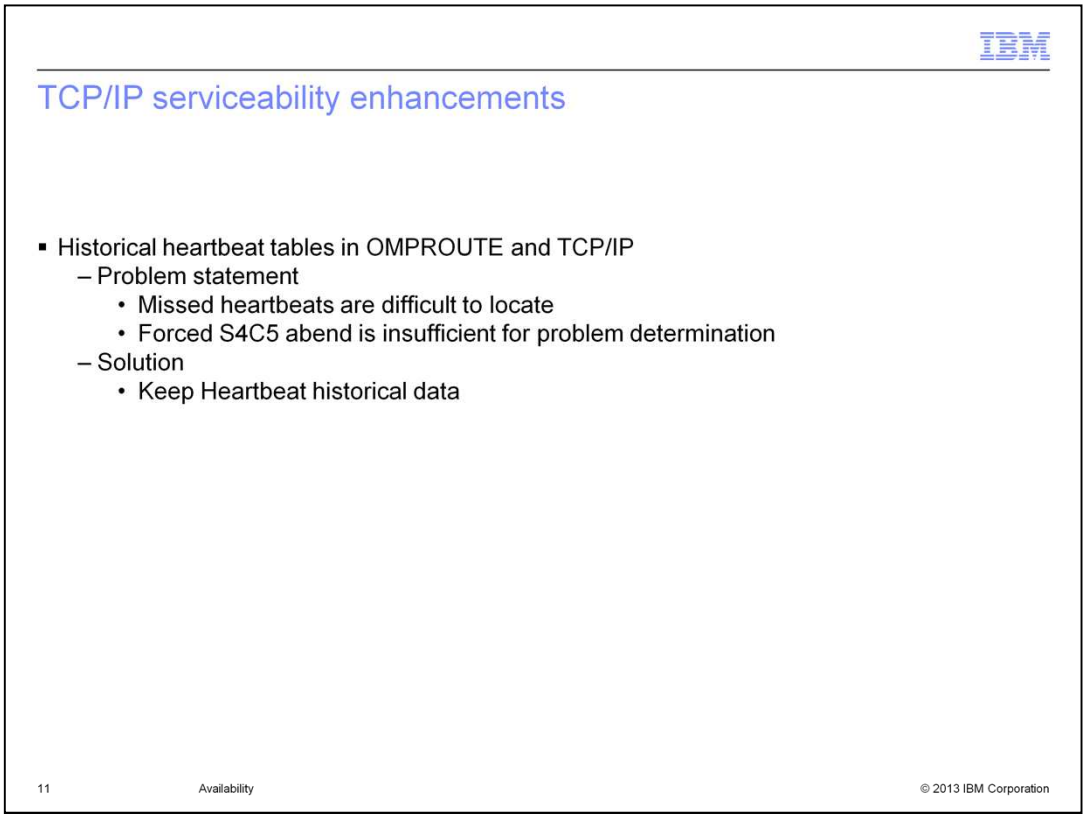
To improve the information that is made available at the time of this type of error, a new trace message is added to the FTP client in z/OS V2R1. The message is included within the FSC (or FSC(1)) trace point, so it is visible only when this point has been activated either on the command line or in the FTP client's FTP.DATA file.



The OMPROUTE_OPTIONS environment variable with the Hello_Hi option, was introduced in OS/390® Communications Server V2R7 as a new function PTF. The new function causes the inbound and outbound hello packets to be processed at a higher priority so that the potential adjacency failures can be minimized. While OMPROUTE can get flooded with the protocol packets in the received order sequence, the Hello_Hi option forces OMPROUTE to prioritize the processing of these packets to maintain the adjacencies.

The Hello_Hi option is disabled by default. If the Hello_Hi option is not specified, the inbound and outbound hello packets might not be processed in a timely manner, resulting in adjacency failures.

Starting in z/OS V2R1, the processing of the inbound and outbound OSPF hello packets is enabled by default at a high priority for ideal optimization and to minimize adjacency failures from the missed hellos. This optimization feature cannot be disabled. The OMPROUTE_OPTIONS environment variable is deprecated by having OMPROUTE ignore it when it is coded and issue a warning message. The message indicates that the variable is ignored and retired in a subsequent release.

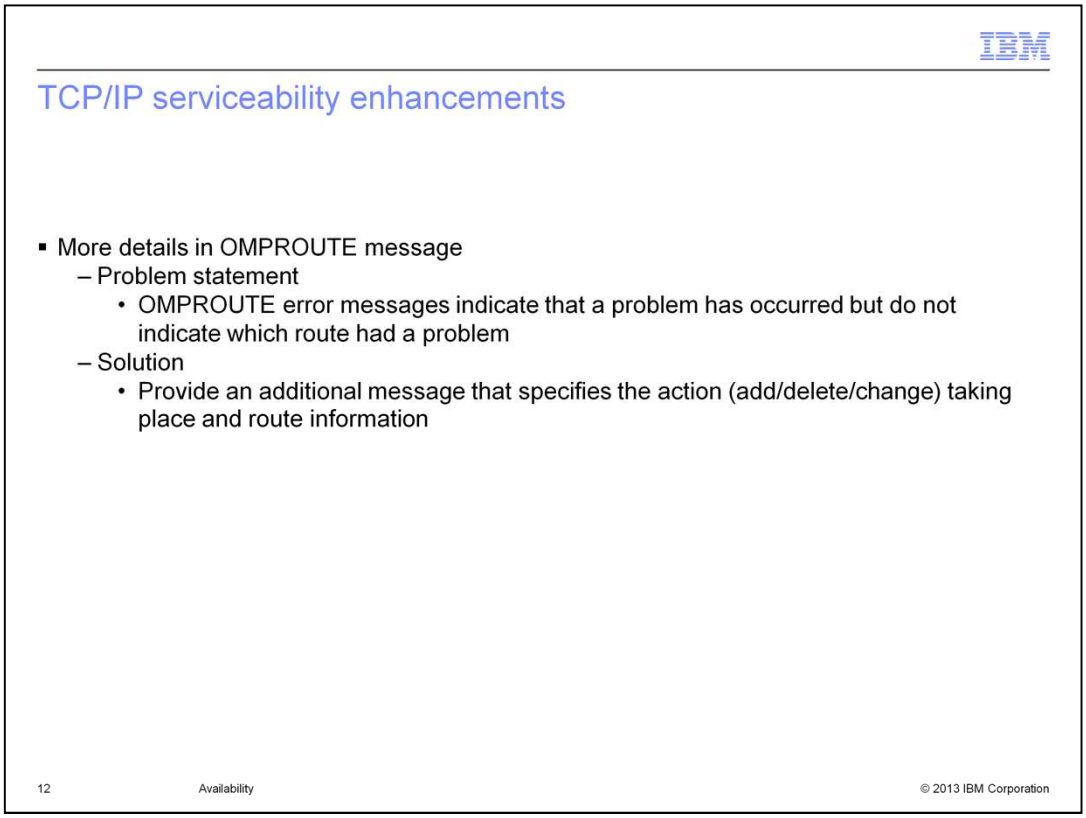


When a TCP/IP stack is configured with sysplex monitoring function enabled (GLOBALCONFIG SYSPLEXMONITOR TIMERSECS) for problem detection, it listens for the heartbeats sent from OMPROUTE. To send a heartbeat, OMPROUTE issues a SIOCSOMPACTIVE ioctl socket call with the active status to the TCP/IP stack. When OMPROUTE is terminating, it uses this ioctl socket call with the termination status to the TCP/IP stack so that the sysplex recovery actions can be taken when necessary.

The TIMERSECS value defaults to 60 seconds and ranges from 10 to 3600 seconds and is used to determine how quickly the sysplex monitor reacts to problems with the needed sysplex resources. Based on the TIMERSECS value, the TCP/IP stack issues warning messages and performs recovery actions as necessary if the RECOVERY option was specified in SYSPLEXMONITOR. If the TCP/IP stack has not received a heartbeat from OMPROUTE for the duration of half of TIMERSECS, then EZZ9672E message is issued to the system console. If no heartbeats were received for the full duration of TIMERSECS, then EZZ9678E message is issued to the system console. The RECOVERY option is used to determine whether to have the TCP/IP stack leave the sysplex group and to force abend dumps of OMPROUTE and TCP/IP address spaces for problem determination. The last recorded time stamp of a received heartbeat in the TCP/IP stack does not help because there is no recorded time stamp when OMPROUTE has sent the correlated heartbeat.

Unfortunately, the dump is typically insufficient because it was not taken near the time of the problem. There are several factors that can contribute to the missed heartbeats and they are difficult to locate. OMPROUTE might not be sending the heartbeats in time or the TCP/IP stack might not be processing the received heartbeats in time.

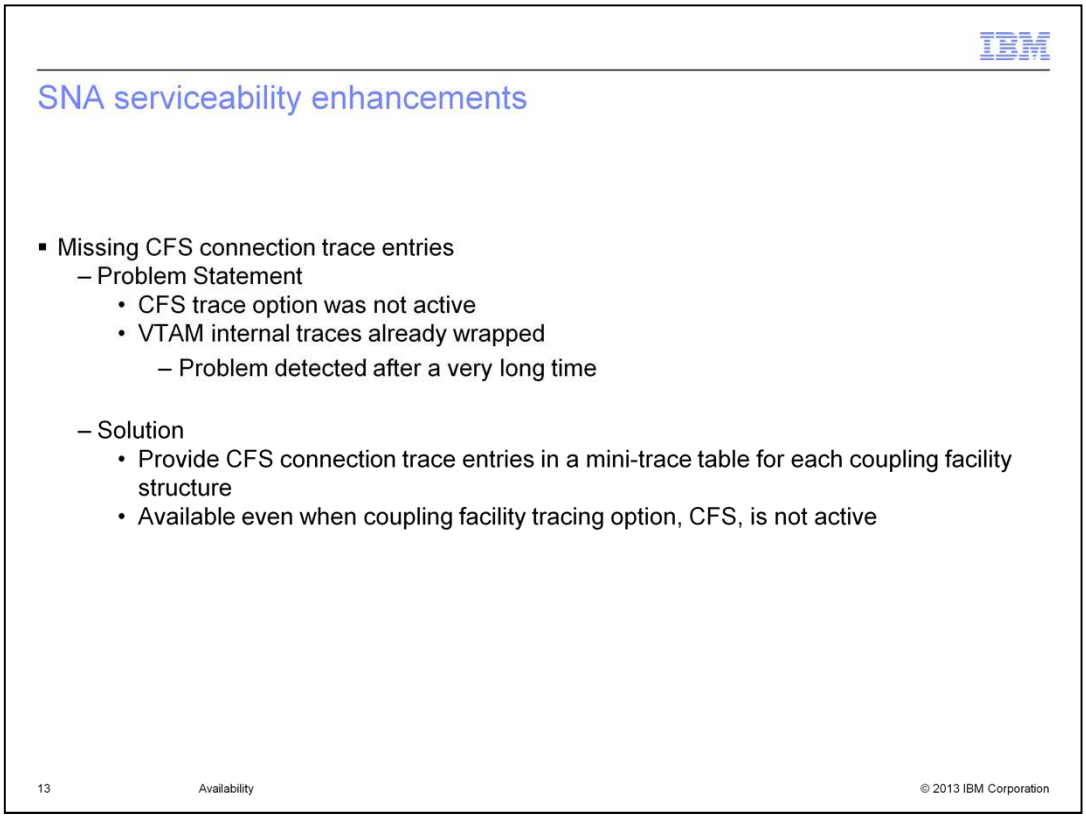
In z/OS V2R1, historical heartbeat tables are added to OMPROUTE and the TCP/IP stack to aid in diagnosis. OMPROUTE includes the time stamp of when it sent the heartbeat in information passed on the ioctl socket call. When the TCP/IP stack receives the ioctl, it saves this time stamp in its heartbeat table and it also saves the time stamp of when the heartbeat was received. Now when a forced abend occurs, the information in the OMPROUTE heartbeat table and the information in the TCP/IP stack heartbeat table, will aid in debugging the problem.



OMPROUTE issues error messages EZZ7828I, EZZ7885I, EZZ7810I, EZZ7829I when a problem is encountered when trying to update the stack's routing table. However, these messages do not indicate which route had a problem. Even though there are some existing info/debug level messages that indicate the route in question, they are issued every time the ioctl is issued, not just in error cases.

Customers typically only run at error level in production to avoid high message volume.

In z/OS V2R1, an additional message, EZZ8174I, is issued that specifies the action (add/delete/change) that is taking place and the route information. This message is issued only when the ioctl socket call to modify the stack's routing table fails.

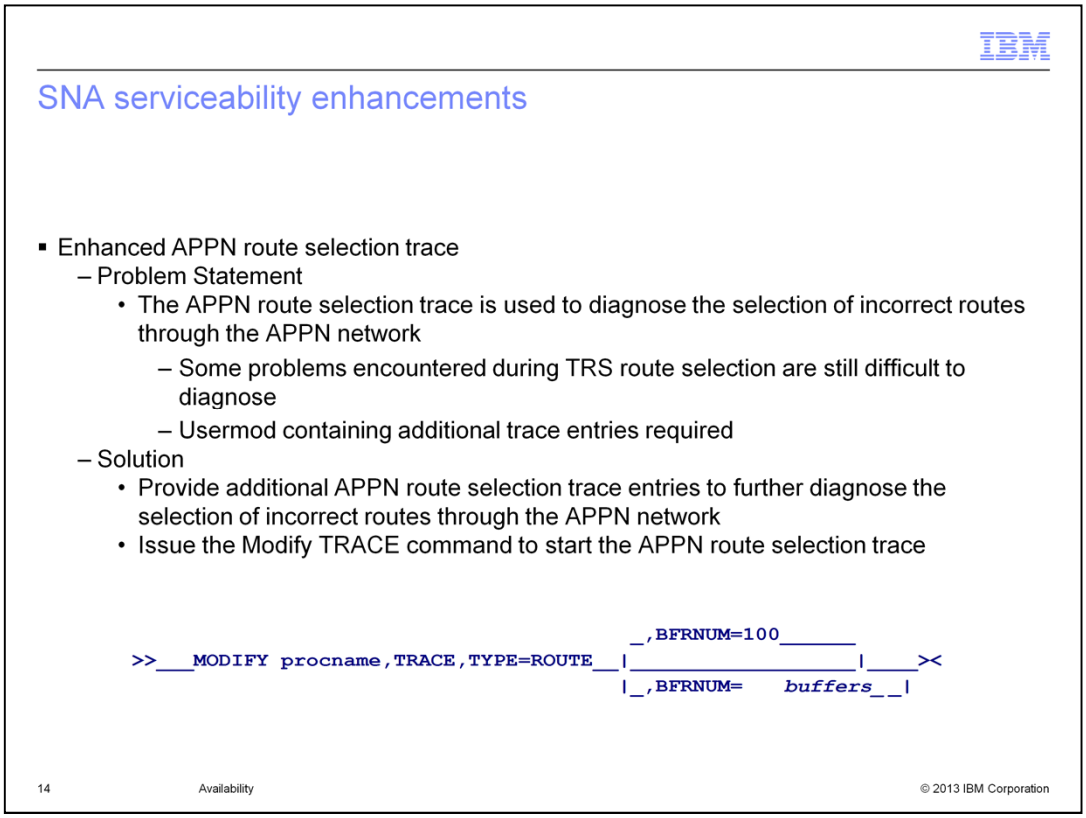


The coupling facility services (CFS) trace option in VTAM internal trace (VIT) traces coupling facility related events. Sometimes a customer is not aware of a coupling facility connection problem until hours later. More than likely, at that point, the VIT has wrapped and does not have any CFS connection related traces. Customers are then requested to re-create the problem by using the external VIT to capture the error information. The external VIT might not have the error information if the problem is not re-creatable or the problem occurred days earlier. If the problem still occurs, then a usermod is needed to obtain the traces at the time of the error.

Starting in z/OS V2R1, VTAM stores CFS connection related traces in a mini-trace table. There is a table for each coupling facility structure used by Communications Server with the exception of the Multi Node Persistent Sessions (MNPS) structure. These tables are available even if the CFS trace option is not enabled.

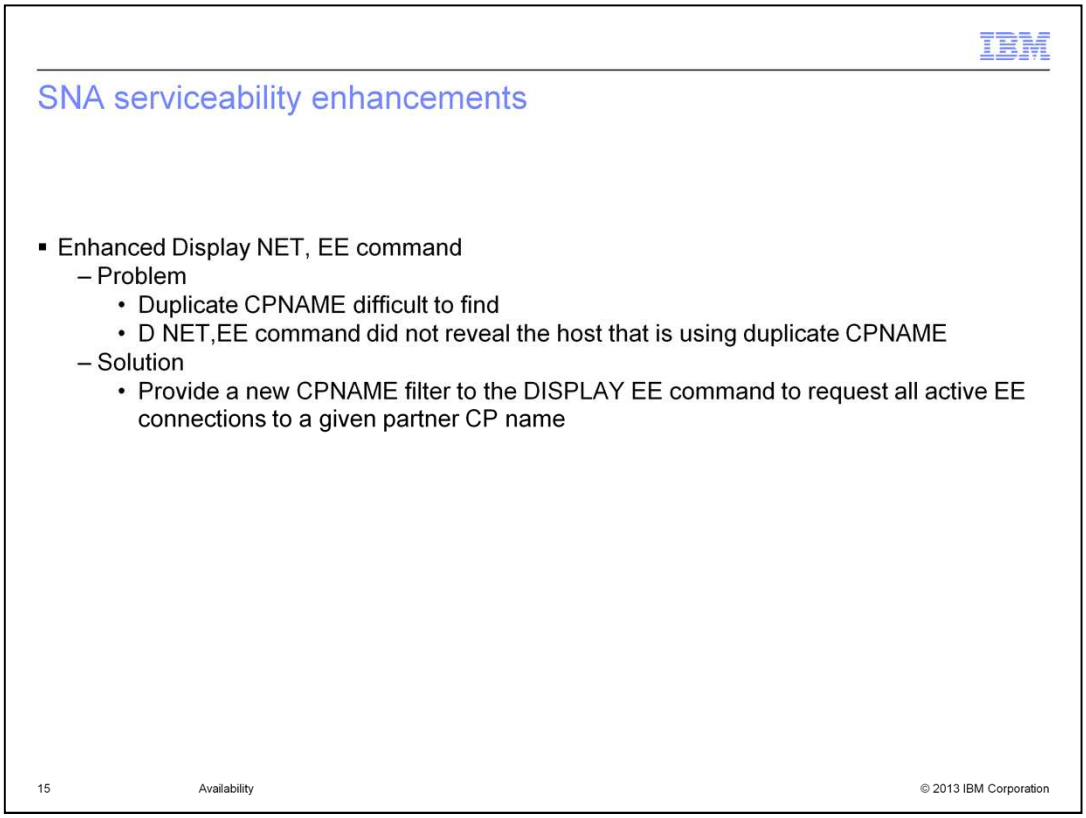
VTAM will always trace coupling facility connection related activities and other important information in the mini-trace table. Each Coupling Facility Structure will have its own mini-trace table. The information provided in the mini-trace tables is similar to the information provided in the CFS VIT trace entries.

Having the information available all the time ensures quicker determination of the problem and minimizes the need for a re-create.



The APPN route selection trace was implemented to diagnose the selection of incorrect routes by TRS. Because all route calculation is performed in network nodes, this trace is valid only on a network node. The route selection trace table is not part of the VIT, but is a separate internal trace table. Even with the existing APPN route selection trace, some problems were resolved only by adding trace entries with USERMODs. However, this process is not satisfactory because you must apply the USERMOD, then wait for the problem to occur again.

In z/OS Communications Server V2R1, additional trace entries are now included in the APPN route selection trace. These trace entries have been used in USERMODs provided to customers and were instrumental in debugging the problem. The MODIFY, TRACE, TYPE=ROUTE command should be issued to start the APPN route selection trace. This allows for further diagnosis of APPN route selection problems without re-creates.

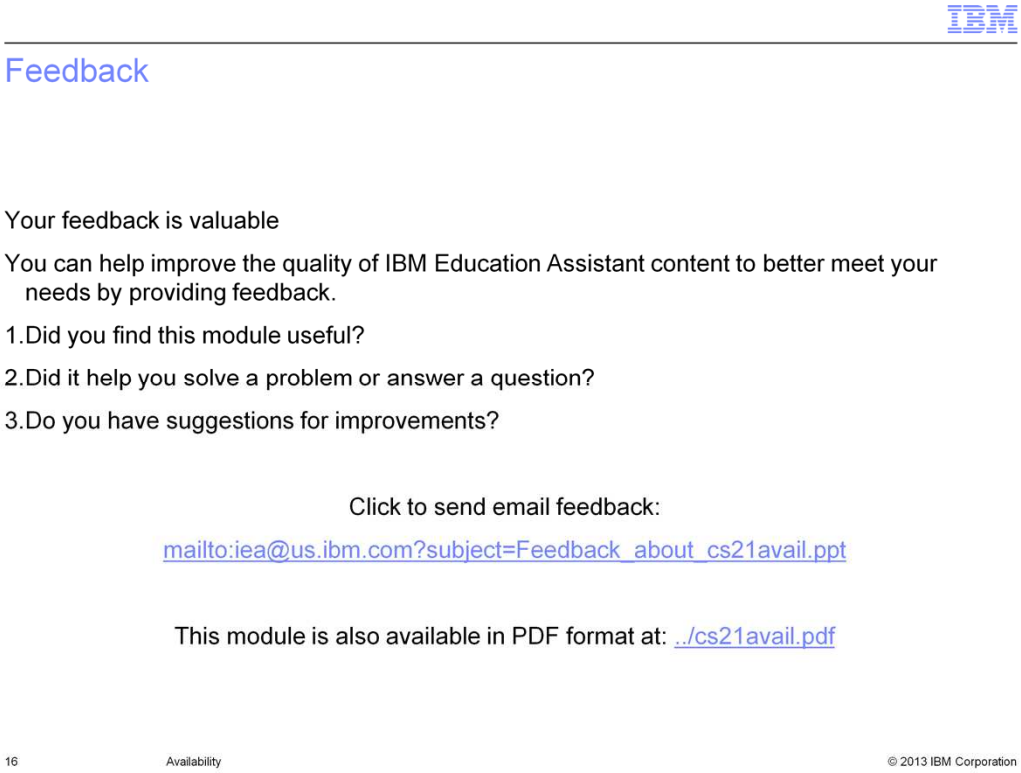


EE connectivity issues exist when two hosts are using the same CPNAME. It can be difficult to find the duplicate CPNAME, even though a failure message indicates that a TG is already active to a CPNAME when a second TG is activated to that CPNAME.

To display EE information, VTAM has a D NET,EE command. Filter operands are ID, IPADDR, and HOSTNAME. Issuing this command did not help with determining what connection was using the duplicate CPNAME.

Starting in z/OS V2R1, a new CPNAME filter is added to the DISPLAY EE command. The filter specifies the name of the CP that is at the other end of the EE connection. The name can be network qualified. If a network identifier is omitted, the host network identifier is assumed.

The output from the command is similar to the output when a remote host name or remote IP address is specified on the command.



You can help improve the quality of IBM Education Assistant content by providing feedback.

Trademarks, disclaimer, and copyright information