



IBM Software Group

z/OS® V1R9 Communications Server

OSA-Express network traffic analyzer and queued direct I/O diagnostic synchronization





@business on demand.

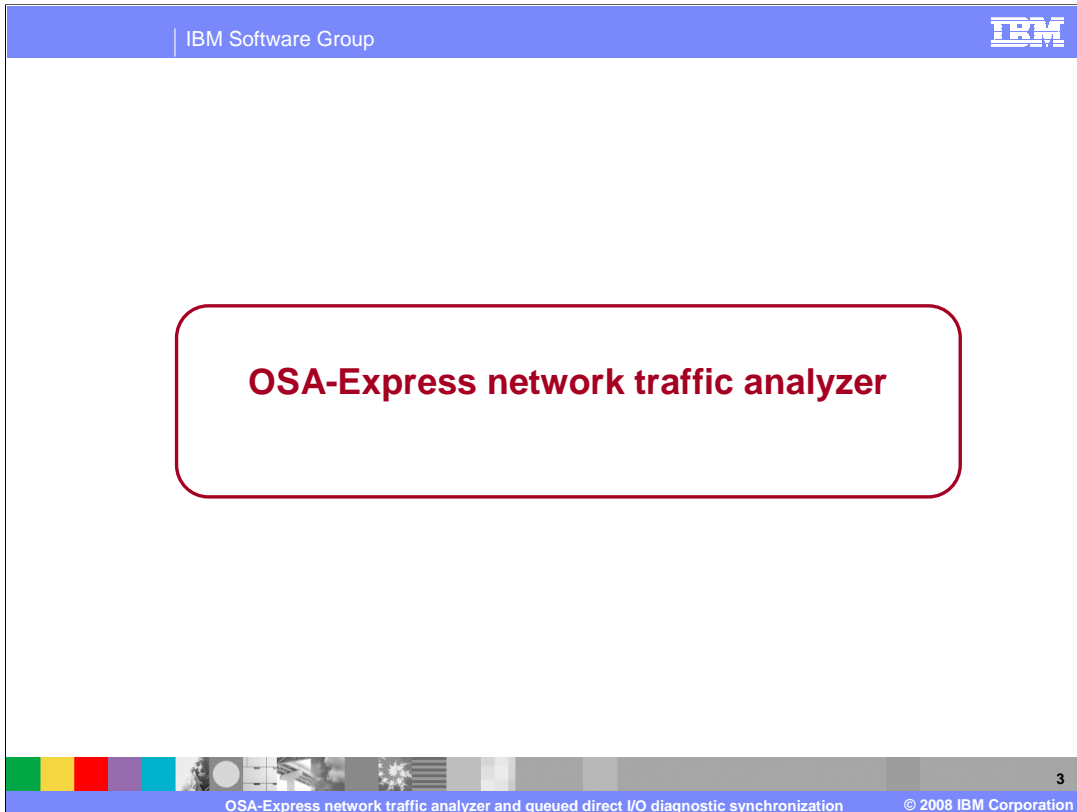
© 2008 IBM Corporation
Updated January 15, 2008

This presentation discusses the enhancements to OSA-Express Network Traffic Analyzer and Queued Direct I/O diagnostic synchronization for the z/OS V1R9 Communications Server.

Agenda

- 
- OSA-Express network traffic analyzer
 - Queued direct I/O diagnostic synchronization
- 

This presentation covers the enhancements for OSA-Express network traffic analyzer and queued direct I/O diagnostic synchronization.



This section describes the z/OS Communications Server implementation of the OSA-Express Network Traffic Analyzer.

Background - OSA-Express in QDIO mode

- Strategic network interface for Ethernet connectivity.
- Configuration in z/OS Communications Server
 - ▶ TRLE definition
 - ✓ The TRLE entry in a VTAM® TRL major node needs at least one DATAPATH address for each TCPIP stack on the LPAR
 - ▶ MPCIPA DEVICE statement and IPAQENET LINK statement for IPv4
 - ▶ IPAQENET6 INTERFACE statement for IPv6
 - ▶ IPv4 LINK and IPv6 Interface share a data device

```

TRLCS  VBUILD TYPE=TRL
*
OSAQ4  TRLE LNCTL=MPC,          *
        READ=(0E28),           *
        WRITE=(0E29),          *
        MPCLEVEL=QDIO,         *
        DATAPATH=(0E2A),       *
        PORTNAME=(QDIO4101,0)

```

```

DEVICE QDIO4101 MPCIPA PRIROUTER
LINK   QDIO4101L IPAQENET QDIO4101
INTERFACE QDIO41016 DEFINE IPAQENET6 PORTNAME QDIO4101

```

- Functions provided for an OSI Layer 3 application
 - ▶ ARP offload
 - ▶ VLAN
 - ▶ Checksum offload
 - ▶ TCP segmentation offload
- The OSA can be shared by multiple stacks, LPARs and CSSs

4

OSA-Express network traffic analyzer and queued direct I/O diagnostic synchronization

© 2008 IBM Corporation

The Open System Adapter (OSA) Express operating in Queued Direct Input/Output (QDIO) mode is the strategic network interface for Internet Protocol (IP) communications for the z/Series line of processors.

The z/OS Communications Server provides the interface to the OSA-Express for IP when a TRLE definition is configured for VTAM and the IPv4 Device and LINK statements or the IPv6 INTERFACE statement is configured for IP. The VTAM component of Communications Server provides the device driver interface between the OSA and IP component with a TRLE definition. The TRLE entry in a VTAM TRL major node needs at least one DATAPATH address for each TCPIP stack on the LPAR. The PORTNAME value on the TRLE statement is the name that is the same value used on the DEVICE and INTERFACE configuration statements.

The OSI Layer 3 functions of ARP offload, VLAN, Checksum offload and TCP segmentation offload can be used by IP. By moving these function to the OSA, the Communications Server reduces the processor load on the main processors.

In addition, multiple instances of the Communication Server in one LPAR or in multiple LPARs can share an OSA.

Problem - Diagnosing QDIO problems

- Diagnosing OSA-Express QDIO problems can be very difficult
 - ▶ TCP/IP stack (CTRACE or packet trace or both)
 - ▶ VTAM (VIT)
 - ▶ OSA (hardware trace) – SE initiated
 - ▶ Network (sniffer trace)
- Often it is not clear where the problem is and which traces to collect.
- Offloaded functions and shared OSAs can complicate the diagnosis.

5

OSA-Express network traffic analyzer and queued direct I/O diagnostic synchronization

© 2008 IBM Corporation

IP network problems can be complicated to resolve. Is the problem outside of the OSA? Is the network router not receiving packets from the OSA or not sending packets to the OSA? Is the problem inside the LPAR? Is the application or communications server not sending packets to the OSA or not receiving packets from the OSA? Tracking down where in the network path a packet is lost may require traces and logs from many different sources:

- Application logs showing that a network session is active and processing network data.
- NETSTAT displays of active connections and the routes that are active for those connections.
- Packet traces and system traces taken by the communications server.
- VTAM traces of the device driver processing the network data.
- OSA hardware traces and logs
- Sniffer traces taken from routers and switches along the network path.
- And the above traces at the other end of the network path.

In addition the offloaded functions and shared OSAs cause further complications. The OSA hardware trace requires IBM SE personnel to be on-site to initiate the trace function from the Hardware Maintenance Console (HMC).

Solution - OSA-Express network traffic analyzer

- Improve serviceability with an OSA-Express network traffic analyzer (OSAENTA) function.
- Supported on OSA-Express2 GA3 (in QDIO mode) on z9-109.
 - ▶ Refer to the 2094DEVICE Preventive Service Planning (PSP) and the 2096DEVICE Preventive Service Planning (PSP) buckets for the latest level of OSA-Express2 LIC.
- Minimizes the need to collect and coordinate multiple traces for diagnosis
- Minimizes the need for traces from the OSA Hardware Management Console (HMC)



The OSA-Express Network Traffic Analyzer (OSAENTA) function is designed to provide the serviceability function for OSA-Express by collecting packet traces between the z/Series processors and the LAN connected to the OSA.

OSAENTA is supported on OSA-Express 2 GA3 on the z9-109 class of processors. It will also require an upgrade of the OSA-Express LIC. To enable the OSA-Express network traffic analyzer, which may be referred to as NTA or OSAENTA, you must be running at least an IBM System z9[®] EC or z9 BC and OSA-Express2 in QDIO mode (CHPID type OSD). Refer to the 2094DEVICE Preventive Service Planning (PSP) and the 2096DEVICE Preventive Service Planning (PSP) buckets for further information.

By collecting the Ethernet data frames OSAENTA make it easier to collect trace data from multiple sources. The Hardware Management Console (HMC) will only be used to the set the security setting for collecting trace data.

Solution - OSA-Express network traffic analyzer

- Allows z/OS Comm Server to collect Ethernet data frames from OSA
 - ▶ Controlled by z/OS Comm Server
 - ✓ New OSAENTA command
 - Define trace filters and parameters
 - OSA sends trace records to the z/OS stack
 - ✓ Save and format the data using existing Ctrace facilities
 - ▶ Collected by OSA
 - ✓ Ability to see:
 - ARP packets
 - MAC headers (including VLAN tags)
 - Packets to/from other stacks shared by the OSA (which could be z/VM® or z/Linux)
 - SNA packets

7

OSA-Express network traffic analyzer and queued direct I/O diagnostic synchronization

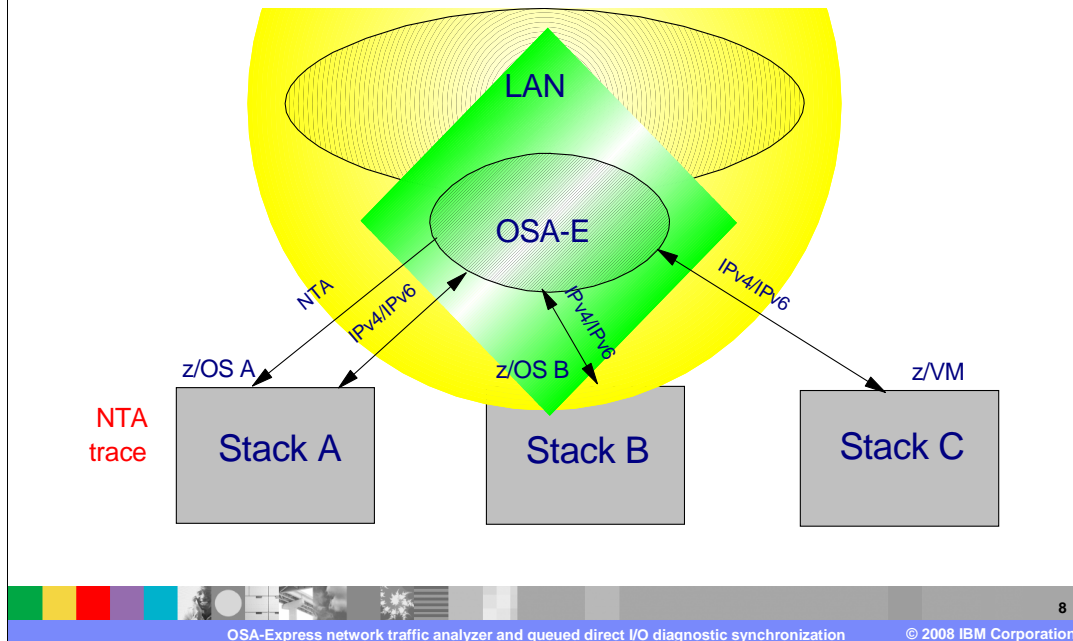
© 2008 IBM Corporation

An installation can now use the new OSAENTA command to define the trace filters and parameters, initiate the trace and terminate the trace. Using the currently available z/OS facilities (CTRACE and IPCS) the trace data can be written to z/OS data sets and formatted.

In addition the OSAENTA facility captures the Ethernet header which is not available with the current PKTTRACE command. The Ethernet header includes the MAC addresses, the VLAN tag, and the other 802.3 fields. Packets for other protocols not currently seen by the z/OS IP Communications Server such as ARP and SNA packets can be captured. Packets sent and received from other devices shared by the OSA can also be captured. These include IP stacks in the same LPAR, in other LPARs running z/VM, z/Linux and z/OS. This also includes other z/OS images with different releases of z/OS.

The OSA collects the data when it is sent across the PCI to the physical port (sometimes referred to as the NIC). The OSA also collects data for LPAR-LPAR packets which do not go onto the LAN. The SNA data collected is limited to Enterprise Extender data when OSA is configured in QDIO layer 3 mode and data to/from Communication Controller for Linux (CCL) on System z™ when OSA is configured in QDIO layer 2 mode. OSA supports only one stack sharing the OSA to perform NTA tracing. One stack can perform NTA tracing for multiple OSAs.

Separate data device for NTA



This diagram shows OSAENTA collecting data from an OSA-Express2 over a separate data path from a z/OS Communications Server LPAR. This z/OS A LPAR has a data path for the IPv4 and IPv6 network traffic. In addition the OSA is shared with another z/OS LPAR, z/OS B, and a z/VM LPAR. OSAENTA can collect packets flowing to and from each of these LPARs and to and from the LAN.

In this configuration Stack A has defined two DATAPATH addresses in its TRLE definition. One will be used for IP communications and the other for OSAENTA. If there are multiple IP stacks on z/OS A, then there must be one DATAPATH address for each IP stack and one for OSAENTA.

This example shows a stack which is using the OSA for IPv4/IPv6 data and for NTA tracing. This configuration requires a TRLE with at least two data devices. Stack A activates the OSA for IPv4 and IPv6. This causes z/OS Communications Server to allocate one data device which is shared for IPv4 and IPv6 data. Stack A also activates the OSA for NTA tracing. This causes z/OS CS to allocate another data device which is used exclusively for NTA.

Another alternative is to have a dedicated stack for NTA. In this configuration, Stack A could be a test system with a single DATAPATH address used exclusively by OSAENTA to capture packets from the other LPARs sharing the OSA, Stack B and Stack C. In this way Stack A will absorb the processor cycles needed to process and write the trace data and minimize the processor impact to the other LPARs. Of course, the impact to the OSA for the overhead of capturing and forwarding the trace data will remain the same in this configuration as in the previous configuration.

OSAENTA

- Control of the network traffic analyzer function
 - ▶ OSAENTA statement in TCPIP profile
 - ▶ VARY TCPIP,,OSAENTA command
- Common syntax for both the command and profile statement
 - ▶ Configure trace filters (which packets to collect)
 - ✓ Up to eight values per filter type
 - ✓ Filters are cumulative across multiple OSAENTA commands
 - ▶ Define trace limits (when to automatically stop the trace)
 - ▶ Specify tracing of discarded filters
 - ✓ Packets silently discarded are no longer silent. Packets that previously disappeared from view can now be traced by OSAENTA.
 - ✓ Discarded packets might be traced twice. Once when it is received during normal processing and once again when it is discarded.
 - ✓ Discarded packets are not matched against the other OSAENTA filters.
- Causes stack to interface with the OSA-Express
 - ▶ Start/stop the trace
 - ▶ Set/update trace filters and settings
- Network traffic analyzer trace interface
 - ▶ Created automatically on first OSAENTA command for a given PORTNAME
 - ✓ Appears as a TCP/IP interface
 - ✓ Only used for inbound trace data
 - ✓ No home IP address
 - ▶ Started with ON parameter of OSAENTA and stopped with OFF parameter of OSAENTA

The z/OS Communications Server OSAENTA command is used to control the trace process in an OSA. The OSAENTA statement can be in the profile data set or in a Vary OBEYFILE data set or issued as console command using VARY TCPIP,,OSAENTA. There are control parameters which start and stop the trace, tell the OSA how much data to collect out of each packet and when to automatically stop the trace. The filter parameters tell the OSA which packets to capture.

Both the OSAENTA statement in the TCPIP profile and the VARY TCPIP,,OSAENTA command provide equivalent function with common syntax. The only difference in syntax is that the command parameters are separated by commas while the profile statement parameters can be separated by blanks. This presentation uses the term "OSAENTA command" to refer to either the OSAENTA profile statement or the VARY TCPIP,,OSAENTA command.

There are seven filters available to define the packets to be captured. Refer to the z/OS Communications Server IP Configuration Reference for a description of each filter. There can be from zero to eight values per filter type. IP address is an exception: There can be up to eight IPv4 addresses and up to eight IPv6 addresses. However, the OSAENTA command will accept only one filter value for each type on a command. Therefore multiple OSAENTA commands are required to define multiple filter values.

A packet must pass all the rules for each filter type to be traced. The packet passes the rules if either there are no filters for the type or the value for the filter type matches one of the packet values. The IP address, IP port and Ethernet MAC addresses have two values in the packet. If one of the values matches the filter value, then the packet passes the filter. The IP address, IP protocol and IP port address are specific to IPv4 and IPv6 packet. If one of these filters is active and the packet is not for IPv4 or IPv6, then the packet does not pass these filters and is not captured. Packets that are being discarded are filtered only by the DISCARD parameter. These filters have no effect of the collection of discarded packets.

A discard reason code is associated with each discarded packet. A packet can be discarded for exceptional reasons or as part of the typical discard processing. The reason a packet is discarded is broken into two groups, exceptional reasons and the typical reasons. An exceptional reason can be no buffers available, the destination IP address is

Trace filter example

- These definitions

- OSAENTA PORTNAME=QDIO4101 IPADDR=9.67.1.1 PROTO=TCP PORTNUM=21
- OSAENTA PORTNAME=QDIO4101 IPADDR=9.67.2.0/24 PORTNUM=22 ON

- Produce these filters

- IPAddr: 9.67.1.1/32 9.67.2.0/24
- Protocol: TCP
- Portnum: 21 22

- These packets will be traced

- SrcIP = 9.67.1.1, Proto = TCP, DstPort = 22
- DstIP = 9.67.2.9, Proto = TCP, SrcPort = 21

- These packets will not be traced

- SrcIP = 9.67.1.1, Proto = UDP, DstPort = 22
- DstIP = 9.67.2.8, Proto = TCP, SrcPort = 23, DstPort = 24
- Ethtype = 80d5 (SNA)

This example shows the effects of using the different filters. Two OSAENTA commands specify selection based upon two IP addresses, the TCP protocol and two port numbers. If a packet matches all the criteria of the filters then it will be traced. Since in this example, the IPADDR, PORTOCOL and PORTNUM filters were used, then packets that are not ethernet type of IPv4 will not be traced.

Network traffic analyzer - additional information

- Security
 - ▶ V TCPIP,OSAENTA command authorization
 - ✓ Needs RACF® access to the MVS.VARY.TCPIP.OSAENTA resource in the OPERCMDS facility
 - ▶ OSA Hardware Management Console (HMC) authorization
 - ✓ Sets the OSA NTA trace authorization in the Support Element (SE)
 - ✓ These SE panels are password protected and require SE access administrator mode to enable which users can access the panels
- Restrictions
 - ▶ Only one network traffic analyzer per OSA
 - ▶ Need HMC authorization to see packets for other operating system images
 - ▶ No MAC headers for LPAR-LPAR traffic
 - ▶ The following are not traced by this function:
 - ✓ Devices not configured in QDIO mode
 - ✓ Data sent/received over the control devices
- z/OS Trace Command
 - ▶ SYSTCPOT - A new Ctrace component for collecting NTA trace data
 - ▶ CTINTA00 is the member in SYS1.PARMLIB
 - ✓ Specify the default buffer size
- NETSTAT DEVLINKS/-d
 - ▶ Enhanced to display Network Traffic Analyzer information
 - ▶ Can be filtered by using INTFName=**EZANTA**osaportname

Security is important since the data being traced in the OSA might contain confidential data. The first level of security is the protection provided by the z/OS security product to control access to the OSAENTA command. The OPERCMDS RACF class and the MVS.VARY.TCPIP.OSAENTA resource can be used to control access to the command.

The second level of security is provided by the Hardware Management Console (HMC). Each OSA can have one of three security levels. The default level, Logical Partition, where only packets associated with devices connected to the LPAR can be traced, CHPID is where packets from all devices sharing the OSA can be traced, and Disabled is where the tracing function is disabled. When the OSA security level is Disabled then the interface activation will fail.

There are some restrictions using the OSAENTA command. The OSA does not support multiple stacks activating the trace at the same time. Once the OSAENTA OFF command has been issued, another stack may start trace. The security authorization in the HMC is required to be set to CHPID to see packets from other operating system images. LPAR to LPAR traffic does not go over the LAN but is handled directly by the OSA. As such there are no MAC headers for these packets. This function applies only to OSA-Express2 Ethernet-type adapters configured in QDIO mode. Data to and from an OSA-Express2 adapter configured in Network Control Program (OSN) mode cannot be traced. SNA data tracing is currently limited to Enterprise Extender data when the OSA-Express adapter is configured in QDIO Layer 3 mode, and data to and from Communication Controller for Linux (CCL) on System z (TM) when the OSA-Express adapter is configured in QDIO Layer 2 mode. Data sent or received over the control devices are not traced. These include the IP assist commands and the OSA-Express SNMP subagent packets.

To provide Ctrace support which is used to write the OSAENTA trace data a new Ctrace component, SYSTCPOT, is defined. The parmlib member is CTINTA00. This member is used to define the size of the buffer space in the TCPIPDS1 data space reserved for OSAENTA Ctrace. The size can range from 1M to 624M with a default of 64M. The command TRACE CT,ON,COMP=SYSTCPOT,SUB=(tcpipprocname) starts the recording in the trace buffer into an external writer. The OPTIONS, JOBNAME and ASID keywords will be accepted by the TRACE command, but will be ignored by TCPIP. The TRACE

NETSTAT DEVLINKS/-d

```

OSA-Express Network Traffic Analyzer Information:
  OSA PortName: QDIO4101      OSA DevStatus:   Ready
  OSA IntfName: EZANTAQDIO4101 OSA IntfStatus: Ready
  OSA Speed:    1000         OSA Authorization: Logical
Partition
  OSAENTA Cumulative Trace Statistics:
    DataMegs:  0              Frames:           8
    DataBytes: 760            FramesDiscarded: 4
    FramesLost: 0
  OSAENTA Active Trace Statistics:
    DataMegs:  0              Frames:           8
    DataBytes: 760            FramesDiscarded: 4
    FramesLost: 0            TimeActive:       8
  OSAENTA Trace Settings:
    DataMegsLimit: 1024      Status: On
    Abbrev:        224        FramesLimit:    2147483647
    Discard:       ALL        TimeLimit:      10080
  OSAENTA Trace Filters:
    DeviceID: *              Nofilter: ALL
    Mac: *
    VLANid: *
    ETHType: *
    IPAddr: *
    Protocol: *
    PortNum: *

```

12

OSA-Express network traffic analyzer and queued direct I/O diagnostic synchronization

© 2008 IBM Corporation

This slide shows the partial output from a NETSTAT DEVLINKS/-d command relating to the Network Traffic Analyzer.

The display is divided into five sections. The first section shows the OSA portname, the interface name and status of the interface. In addition the last known security setting associated with the OSA is shown. If the interface has never been active, then UNKNOWN will be shown.

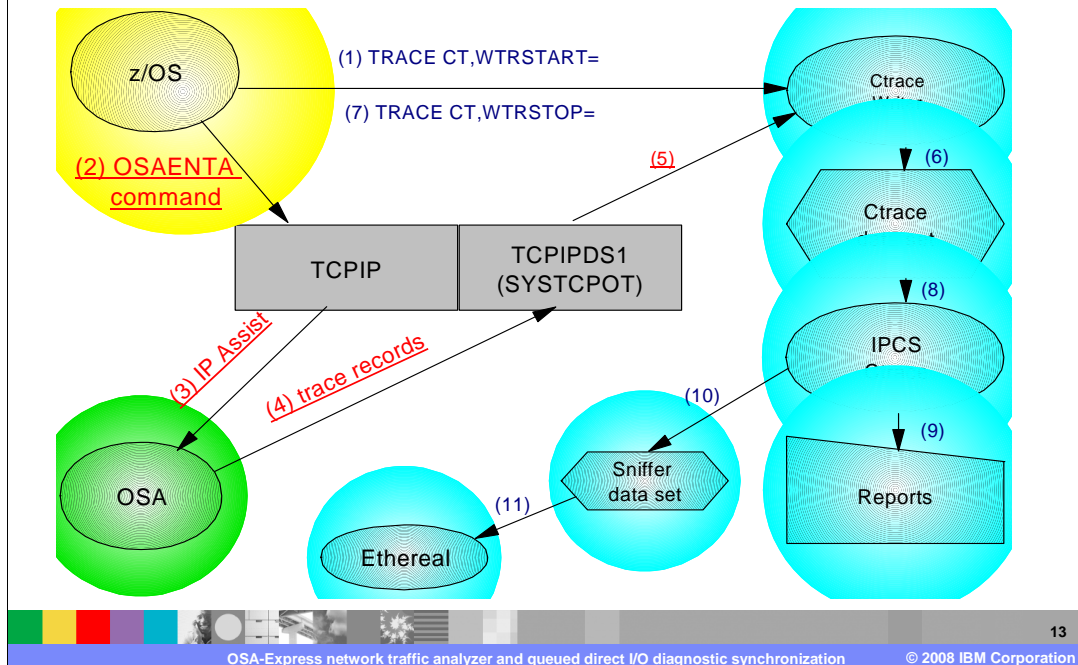
The second section is the statistics since the first time the interface was defined to TCPIP. Note that when the OSAENTA DELETE command is issued then these values are lost.

The third section is the statistics since the last time a OSAENTA ON command was issued.

The fourth section is the current trace settings for DATA, FRAMES, ABBREV, TIME and DISCARD.

The fifth section is the current filter values. If NOFILTER has been set to NONE, or left to default, and all the filters are set *, then the filter values are not displayed.

Overview of NTA function



This diagram shows the process of collecting trace data.

1. A MVS Ctrace external writer is started with the TRACE CT,WTRSTART=*wtrprocname*
2. The VARY OSAENTA,PORTNAME=*osaname*,ON command is issued to start the collection process
3. TCP/IP creates the required control blocks for the OSAENTA command. The interface name is EZANTA*osaname* . Then IP Assist orders are issued to create the data channel for the trace records and request the tracing commence using filters and parameters from the OSAENTA command (or commands).
4. The OSA collects the trace records into buffers for transmission to TCPIP. The same mechanism for regular data transfer is used to transfer the trace buffers.
5. TCPIP moves the trace data into the TCPIPDS1 data space buffers reserved for the SYSTCPOT Ctrace component. As the Ctrace buffers are filled they are written by the Ctrace external writer.
6. The Ctrace external writer copies the buffers into the Ctrace data sets allocated in the Ctrace writer procedure.
7. The operator disconnects the writer from TCPIP and stops the writer.
8. The IPCS CTRACE subcommand can now be used to select packets and format reports.
9. The same reports available for packet trace (SYSTCPDA) are available for OSA trace (SYSTCPOT).
10. The packets can also be copied to a sniffer formatted data set (OPTIONS ((SNIFFER))).
11. Once the sniffer data set has been downloaded to a PC, programs such as Ethereal can be used to further analyze the packet data.

D NET,TRL,TRLE=

- D NET,TRL,TRLE=trlename output

```

...
IST1716I PORTNAME = OSAQDIO4 LINKNUM = 0 OSA CODE LEVEL = 0630
IST1577I HEADER SIZE = 4096 DATA SIZE = 0 STORAGE = ***NA***
IST1221I WRITE DEV = 0E29 STATUS = ACTIVE STATE = ONLINE
IST1577I HEADER SIZE = 4092 DATA SIZE = 0 STORAGE = ***NA***
IST1221I READ DEV = 0E28 STATUS = ACTIVE STATE = ONLINE
IST1221I DATA DEV = 0E2A STATUS = ACTIVE STATE = N/A
IST1724I I/O TRACE = OFF TRACE LENGTH = *NA*
IST1717I ULPID = TCPCS1
IST1815I IQDIO ROUTING DISABLED
IST1918I READ STORAGE = 4.0M(64 SBALS)
IST1757I PRIORITY1: UNCONGESTED PRIORITY2: UNCONGESTED
IST1757I PRIORITY3: UNCONGESTED PRIORITY4: UNCONGESTED
IST2190I DEVICEID PARAMETER FOR OSAENTA TRACE COMMAND = 00-07-00-00
...
IST1221I TRACE DEV = 0E2B STATUS = ACTIVE STATE = N/A
IST1724I I/O TRACE = OFF TRACE LENGTH = *NA*
IST1717I ULPID = TCPCS1
IST1815I IQDIO ROUTING DISABLED
IST1918I READ STORAGE = 4.0M(64 SBALS)
IST1757I PRIORITY1: UNCONGESTED PRIORITY2: ***NA***
IST1757I PRIORITY3: ***NA*** PRIORITY4: ***NA***

```

14

This is the output of the VTAM display for an active OSA TRLE definition. Message IST1716I shows the OSA code level. Note that 0630 is **NOT** the code level that supports OSAENTA. Message IST2190I shows the DEVICEID value. Message IST1221I shows the device address and status of the trace DATAPATH device.

New messages

- New stack messages

```
EZD0015I osa_portname DOES NOT SUPPORT OSAENTA TRACE
EZD0016I OSAENTA TRACE ENABLED FOR osa_portname
EZD0017I OSAENTA TRACE MODIFIED FOR osa_portname
EZD0018I OSAENTA TRACE DISABLED FOR osa_portname
EZD0019I OSAENTA TRACE STOPPED FOR osa_portname - REASON: reason LIMIT REACHED
EZD0020I ERROR error_code ENABLING OSAENTA TRACE FOR osa_portname
EZD0021I ERROR error_code MODIFYING OSAENTA TRACE FOR osa_portname
EZZ0794I TOO MANY keyword VALUES DEFINED FOR OSAENTA osa_portname ON LINE n
```

15

OSA-Express network traffic analyzer and queued direct I/O diagnostic synchronization

© 2008 IBM Corporation

Message EZD0015I is issued if the OSAENTA ON command is issued for an OSA that does not support the Network Traffic Analyzer function.

Message EZD0016I is issued for a successful response to an initial OSAENTA ON command.

Message EZD0017I is issued for a successful response to a subsequent OSAENTA ON command.

Message EZD0018I is issued for a successful response to an OSAENTA OFF command.

Message EZD0019I is issued when one of the DATA, FRAMES or TIME limits is reached and the interface is deactivated.

Message EZD0020I is issued when an error occurs when a OSAENTA ON command is issued for the first time.

Message EZD0021I is issued when an error occurs when a OSAENTA command is issued after the interface is started.

Message EZZ0794I is issued when more than eight values have been specified for a given keyword across multiple OSAENTA commands.

Things to think about

- PTFs for the V1R8 APAR PK36947 are required
- Verify OSA microcode level (D NET,TRL,TRLE)
- Configure required HMC authorization setting
- Need an available data device (from the TRLE) for the NTA function
- LPAR-LPAR packets traced twice (once in each direction)
- Set appropriate filters to limit amount of data traced
 - ▶ Minimize chances of traces wrapping
 - ▶ Reduce performance impact

16

OSA-Express network traffic analyzer and queued direct I/O diagnostic synchronization

© 2008 IBM Corporation

For z/OS V1R8 Communications Server, install the PTFs for APAR PK36947. This APAR is required for correct functioning of OSAENTA on the z/OS V1R8 system. Since this a new function for V1R8 and V1R9 there is no migration from previous releases.

Before using this function you need to be aware of the correct OSA microcode level that needs to be installed. Refer to the 2094DEVICE Preventive Service Planning (PSP) and the 2096DEVICE Preventive Service Planning (PSP) buckets for the required microcode level.

Configure the HMC security settings for the OSAs. Either DISABLED, Logical Partition or CHPID should be configured.

Set the RACF permissions for the MVS.VARY.TCPIP.OSAENTA command.

Update the TRLE definition for the OSA to add an additional trace device address for OSAENTA.

Update CTINTA00 to set the Ctrace buffer size. Remember that this will use up auxiliary page space storage.

Remember to use filters to limit the trace records to prevent over consumption of the OSA processor resources, the LPAR processor resources, the TCPIPDS1 trace data space, memory, auxiliary page space and the IO subsystem writing trace data to disk.

For additional information you can refer to:

The Communications Server IP Configuration Guide for an overview of the function.


The Communications Server IP Configuration Reference for the syntax of the OSAENTA statement.

The Communications Server IP System Administrator's Commands for the syntax of the VARY TCPIP,,OSAENTA command.

The Communications Server IP Diagnosis Guide for packet trace formatting information.

The OSA-Express Customer's Guide and Reference for information about the OSA-Express support for the NTA function.

The System z9 Support Element Operations Guide for information about the OSA HMC authorization function.

IBM Software Group 

Queued direct I/O diagnostic synchronization

OSA-Express network traffic analyzer and queued direct I/O diagnostic synchronization © 2008 IBM Corporation 17

This solution is also informally referred to as 'OSA Trap'. This solution was part of z/OS **V1R8** Communications Server and subsequent releases. It is being presented here because OSA support is now available.

Background information

- Each OSA has its own trace table
 - ▶ Managed using the Hardware Management Console (HMC).
 - ▶ Trace table is snapshot using the HMC.
- Each host has its own trace table
 - ▶ VTAM has VTAM Internal trace, TCP/IP has CTrace



All references to OSA in this presentation implies OSA-Express2 in QDIO mode. Diagnostic information refers primarily to OSA and host trace tables. It does not preclude other diagnostic information such as counters, error logs, and so on. The OSA and the host maintain their own diagnostic information separately and each product's trace tables often are the most important piece of diagnostic information for that product.

Problem statement - Trace synchronization

- Difficult to synchronize the OSA and host trace tables.
- Difficult to stop the OSA trace table when a host dump is being taken.
 - ▶ Must be there when the problem occurs.
 - ▶ You must be physically quick (in some cases physically impossible).



This solution originated as a requirement from z/OS Communications Server System Verification Test (SVT). SVT wanted a way to automatically capture diagnostic information from multiple products simultaneously, hoping to minimize re-creates.

Solution - Automatic trace synchronization

- Exploit new OSA support which allows for automatic synchronization
- Managed using new control channel signals.
 - ▶ Arm (with optional OSA trace record filtering)
 - ✓ Arming the OSA puts it in a state where it will react to a Capture signal from the host or it detects abnormal loss of host connectivity. Arming an OSA will NOT adversely affect performance.
 - ✓ New TRACE TYPE **QDIOSYNC** is used to Arm the OSA
 - Specified on VTAM Modify Trace command or VTAM Trace start option
 - Granularity is on the TRLE level
 - ▶ Capture
 - ✓ The user can Capture based on the issuance of a specific message.
 - Requires the use of the **z/OS Message Processing Facility (MPF)** exit and the z/OS SLIP facility
 - ✓ The user can Capture based on the execution of a specific instruction.
 - Requires the use of a **z/OS Program Event Recording (PER)** SLIP
 - ✓ OSA will initiate Capture when it is Armed and detects abnormal loss of connectivity to the host
 - ▶ Disarm
 - ✓ Disarming the OSA causes it to ignore Capture requests. Also, the OSA will not snapshot its trace table when abnormal loss of host connectivity is detected.
 - ✓ New TRACE TYPE **QDIOSYNC** is used to Disarm the OSA
 - Specified on VTAM Modify NoTrace command or VTAM NoTrace start option
 - Granularity is on the TRLE level

This solution is effective only if supported by the OSA and enabled on z/OS Communications Server.

New control signals are used between z/OS Communications Server and OSA to facilitate implementation of this solution. z/OS Communications Server uses one of these signals to tell OSA to snapshot its trace table to the HMC hardfile. This new set of control signals is collectively known as SetDiagAsst.

Host initiated refers to z/OS Communications Server sending a trace synchronization command to the OSA. OSA initiated refers to action taken by the OSA without a specific command being received from the host.

Two synchronization states are defined for the OSA Armed and Disarmed. When Armed, if the OSA receives a capture request from the host or the OSA loses host connectivity, the OSA will snapshot its diagnostic information (trace table) to the HMC hardfile. When disarmed, the OSA will act no different than before this solution. Arming an OSA will NOT adversely affect performance because it causes the OSA to change an internal state and this internal state is not interrogated by the OSA during normal data transmission.

There are host initiated captures and an OSA initiated capture. For host initiated captures, z/OS Communications Server tells OSA to snapshot its trace table. For OSA initiated captures, OSA decides on its own to snapshot its trace table. There are 2 methods a user can use to initiate a Capture request from z/OS Communications Server. Note that a Capture is sent to all Armed OSAs. A user can Capture based on the issuance of a specific message. This requires the use of the z/OS Message Processing Facility (MPF) to drive the new V1R8 MPF exit (IUTLLCMP). The user will also need to use the z/OS SLIP facility on the same messages to initiate a host dump. The user can also Capture based on the execution of a specific instruction. This requires the use of a z/OS PER type SLIP specifying ACTION=(RECOVERY). In this case you can use the same PER SLIP to also get a host dump. Of the two host initiated capture mechanisms the MPF and SLIP mechanism is the most useful. The PER SLIP is the less useful and could significantly affect performance. The OSA will initiate a Capture when it is Armed and detects abnormal loss of connectivity to the host (includes any type of Halt subchannel (for

Trace management - Arm

- Use Modify TRACE to Arm an OSA (TRACE start option is similar).

```

>>_MODIFY procname,TRACE_,TYPE=QDIOSYNC |_, ID=*_____ |_____ >
|_, ID=_ * _____ |
|_trle_name_|
|_____ >

>_|_, OPTION=ALLINOUT |_, SYNCID=trle_name |_, SAVE=NO |_____ ><
|_, OPTION=_ ALLIN _____ | |_, SYNCID=identifier | |_, SAVE=_ NO _____ |
|_ ALLINOUT _____ |
|_ ALLOUT _____ |
|_ IN _____ |
|_ INOUT _____ |
|_ OUT _____ |

```

21

OSA-Express network traffic analyzer and queued direct I/O diagnostic synchronization

© 2008 IBM Corporation

New options and values are highlighted in red. There are new values for the OPTION parameter. ALLINOUT should be used unless directed to do otherwise. ALLIN directs OSA to collect only inbound diagnostic information for all devices. ALLOUT directs OSA to collect only outbound diagnostic information for all devices. ALLINOUT directs OSA to collect inbound and outbound diagnostic information for all devices. IN directs OSA to collect only inbound diagnostic information for devices defined to this VTAM. OUT directs OSA to collect only outbound diagnostic information for devices defined to this VTAM. INOUT directs OSA to collect inbound and outbound diagnostic information for devices defined to this VTAM. Note that OSA currently does not support IN, OUT and INOUT. They can be specified, but OSA will convert them to ALLIN, ALLOUT, and ALLINOUT.

SyncID is a (user chosen) EBCDIC correlator value passed to OSA on an Arm request (OSA will convert to ASCII and use it in its diagnostic information).

The asterisk value is now accepted for the ID keyword when TYPE=QDIOSYNC and means all TRLEs. The asterisk is NOT a wildcard variable and if used must be the only character specified.

You can issue Modify Trace even if the OSA is already Armed, which effectively updates the values.

Note that SAVE=NO is the default for the modify trace command where SAVE=YES is the default for the trace start option.

Trace management - Disarm

- Use Modify NOTRACE to Disarm an OSA (NOTRACE start option is similar).

```

>>__MODIFY                                _, ID=*_____
procname,NOTRACE__,TYPE=QDIOSYNC_|_____><
|_,ID=_*_____ -|
|_trle_name_|

```

New options and values are highlighted in red.

The asterisk value is now accepted for the ID keyword when TYPE=QDIOSYNC and means all TRLEs. The asterisk is NOT a wildcard variable and if used must be the only character specified.

The Vary TCPIP,tcpprocname,STOP command results in an automatic Disarm if the OSA is Armed.

Trace management - Display trace

- Use Display TRACES (TYPE=NODES or TYPE=ALL).

```

d net,traces,type=nodes,id=*
IST097I DISPLAY ACCEPTED
IST350I DISPLAY TYPE = TRACES,TYPE=NODES 506
IST075I NAME = A50CDRMC, TYPE = CDRM SEGMENT
IST1041I C01N                CDRM
IST1042I  BUF                = ON    - AMOUNT = PARTIAL  - SAVED = NO
IST924I -----
IST075I NAME = A0362ZC, TYPE = PU T4/5
IST1041I A03S16              LINE
IST1042I  LINE                = TRACT
IST924I -----
IST075I NAME = TRLHYDRA, TYPE = TRL MAJOR NODE
IST1041I TRLHYDRA            TRL MAJOR NODE
IST1042I  IO                  = ON    - AMOUNT = **NA**  - SAVED = NO
IST1041I NSQDIO11            TRLE
IST1042I  IO                  = ON    - AMOUNT = **NA**  - SAVED = NO
IST2183I  QDIOSYNC = ALLINOUT - SYNCID = NSQDIO11 - SAVED = YES
IST314I END
  
```

A new message is added to the Display response and is highlighted in red. The message is only issued if the TRLE is Armed (or will be armed when activated).

In this case the OPTION specified on (or defaulted for) the Trace command/start option is ALLINOUT. ALLINOUT means OSA is to collect diagnostic data pertaining to ALL LPARs and in both directions. The SYNCID is NSQDIO11, which is being retained by the OSA in case a capture occurs (the SYNCID value will be apparent in the OSA diagnostic data). The QDIOSYNC trace is also saved meaning it is effected when the TRLE or any of its devices are activated.

Trace management - Display TRLE

- Use Display TRL (Display ID=trlename is similar).

```
d net,trl,trle=of8geth
IST097I DISPLAY ACCEPTED
IST075I NAME = OF8GETH, TYPE = TRLE
IST1954I TRL MAJOR NODE = TRLHYDRA
IST486I STATUS= ACTIV, DESIRED STATE= ACTIV
IST087I TYPE = LEASED , CONTROL = MPC , HPDT = YES
IST1715I MPCLEVEL = QDIO MPCUSAGE = SHARE
IST1716I PORTNAME = OF8GETHP LINKNUM = 0 OSA CODE LEVEL = 0314
IST2184I QDIOSYNC = ALLINOUT - SYNCID = OF8GETH - SAVED = NO
IST1577I HEADER SIZE = 4096 DATA SIZE = 0 STORAGE = ***NA***
IST1221I WRITE DEV = 2E81 STATUS = ACTIVE STATE = ONLINE
IST1577I HEADER SIZE = 4092 DATA SIZE = 0 STORAGE = ***NA***
IST1221I READ DEV = 2E80 STATUS = ACTIVE STATE = ONLINE
IST1221I DATA DEV = 2E82 STATUS = ACTIVE STATE = N/A
```

A new message is added to the Display response and is highlighted in red. The message is only issued if the TRLE is Armed (or will be armed when activated).

Trace management - Changed messages

- Changes to existing messages.

IST1077I OPTION **SYNCID** AFTER **QDIOSYNC** NOTRACE IS NOT VALID

IST1515I **QDIOSYNC** TRACE ACTIVE

IST225I **QDIOSYNC** FOR ID = NSQDIO11 FAILED - **NOT SUPPORTED**

IST1137I **QDIOSYNC** FAILED, NSQDIO11 - **NOT SUPPORTED**

IST225I **QDIOSYNC** FOR ID = NSQDIO11 FAILED - **ARM REJECTED**

IST1137I **QDIOSYNC** FAILED, NSQDIO11 - **ARM REJECTED**

IST176I TRACE FAILED - TYPE AND **SYNCID** ARE CONFLICTING OPTIONS

This is a summary of the changes to the messages issued by the trace facility with the new variables shown in red.

In general, these are error messages issued when an incorrect command is entered or the command is valid but for some reason the OSA cannot be armed.

Sample - Using MPF to initiate capture

- Sample MPF ParmLib member
 - Restriction - Message must be first in group or ungrouped.

```
* This MPFLSTxx identifies the messages which lead to capture of
* armed OSA devices. If any of the following message are issued,
* IUTLLCMP (VTAM provided MPF exit) gains control and schedules
* the capture of all armed OSA devices.
*
* EZZ4343I ERROR xxxxx REGISTERING IP ADDRESS<IP_Addr> FOR ...
* EZZ4339I INTERFACE interface_name FAILED - ADAPTER SIGNAL ...
* EZZ4327I ERROR XXXX REGISTERING IP ADDRESS
* EZZ4328I ERROR XXXX SETTING ROUTING FOR DEVICE
EZZ4343I,SUP(NO),USEREXIT(IUTLLCMP)
EZZ4339I,SUP(NO),USEREXIT(IUTLLCMP)
EZZ4327I,SUP(NO),USEREXIT(IUTLLCMP)
EZZ4328I,SUP(NO),USEREXIT(IUTLLCMP)
```

- When using the MPF exit, use a SLIP for each message in the ParmLib member to get a synchronized host dump (need 4 of these for the MPF ParmLib sample)
 - Note: This is a sample, check the job and dataspace names and modify if necessary.

```
SL DEL, ID=MEZx, END
SL SET, ID=MEZx, MSGID=EZZ43xxI, A=(STOPGTF, SVCD), MATCHLIM=1,
JOBLIST=(TCP*, NET*),
DSPNAME=('TCP*'.*, 01.CSM*, 'NET*'.IST*),
SDATA=(RGN, ALLNUC, CSA, LSQA, PSA, SQA, SUM, SWA, TRT, LPA),
END
```

Using the MPF exit will tend to have an insignificant effect on performance.

Additional information on the MPFLSTxx ParmLib member can be found in the z/OS MVS publications. Search on 'MPFLSTxx'.

The messages in this list are indicative of OSA errors commonly seen by the System Verification Test (SVT) group.

When z/OS Communications Server issues any message in this ParmLib, z/OS generates a call to IUTLLCMP which locates all Armed OSAs and builds and sends each of them a Capture request.

In order to get 'matching' host documentation for the OSA trace table, create a message SLIP for each of the 4 messages. This set should be repeated 4 times using MEZ1, MEZ2, MEZ3, and MEZ4 with the corresponding message number change. Order of the SLIPs and order of the ParmLib messages need not match, as long as there's a SLIP for each message in the ParmLib.

Sample - Using PER SLIP to initiate capture

- Sample PER SLIP trap.
- Specifying A=(RECOVERY) initiates capture on all armed OSA devices.
- Note: This is a sample, check the job and dataspace names and modify if necessary.

```
SL DEL, ID=MEZ2, END
SL SET, IF, ID=MEZ2, RA=(address), A=(STOPGTF, RECOVERY, SVCD),
MATCHLIM=1, JOBLIST=(TCP*, NET*),
DSPNAME=( 'TCP*' . *, 01.CSM*, 'NET*' . IST*),
SDATA=(RGN, ALLNUC, CSA, LSQA, PSA, SQA, SUM, SWA, TRT, LPA),
END
```

Unlike message SLIPs, only 1 PER SLIP can be active at any time (which restricts its usefulness). Also, using a PER SLIP trap can have a significant adverse effect on performance.

This is not intended to be used with the MPF ParmLib member but instead by itself.

Automatic trace synchronization - Additional information

- Arming/Disarming is done on a TRLE basis so either all data devices in the group are armed or all are not.
- If an Arm failure occurs attempting to Arm one of the devices in the TRLE, any other devices in the same TRLE that were previously Armed are forced Disarmed.
- Subsequent Arm attempts (commands) are rejected once it's discovered the OSA does not support SetDiagAsst or an Arm failure occurred. Only a TRLE recycle will allow a subsequent Arm attempt.
- The host TOD value is sent to the OSA in every QDIOSYNC signal (Arm, Capture, and Disarm). This intended to be exposed by the OSA and used to correlate the host and OSA diagnostics.
- In order to use the QDIOSYNC function, you must be running at a minimum with an IBM System z9 EC or z9 BC and OSA-Express2 in QDIO mode (CHPID type OSD). Refer to the 2094DEVICE Preventive Service Planning (PSP) and the 2096DEVICE Preventive Service Planning (PSP) buckets for further information.

This slide contains some additional information that will be useful when using this function.

Feedback

Your feedback is valuable

You can help improve the quality of IBM Education Assistant content to better meet your needs by providing feedback.

- Did you find this module useful?
- Did it help you solve a problem or answer a question?
- Do you have suggestions for improvements?

Click to send e-mail feedback:

mailto:iea@us.ibm.com?subject=Feedback_about_OSA_Exp_QDIO_diag.ppt

This module is also available in PDF format at: [../OSA_Exp_QDIO_diag.pdf](http://OSA_Exp_QDIO_diag.pdf)



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

Trademarks, copyrights, and disclaimers

The following terms are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both:

IBM RACF System z System z9 VTAM z/OS z/VM

Product data has been reviewed for accuracy as of the date of initial publication. Product data is subject to change without notice. This document could include technical inaccuracies or typographical errors. IBM may make improvements or changes in the products or programs described herein at any time without notice. Any statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only. References in this document to IBM products, programs, or services does not imply that IBM intends to make such products, programs or services available in all countries in which IBM operates or does business. Any reference to an IBM Program Product in this document is not intended to state or imply that only that program product may be used. Any functionally equivalent program, that does not infringe IBM's intellectual property rights, may be used instead.

Information is provided "AS IS" without warranty of any kind. THE INFORMATION PROVIDED IN THIS DOCUMENT IS DISTRIBUTED "AS IS" WITHOUT ANY WARRANTY, EITHER EXPRESS OR IMPLIED. IBM EXPRESSLY DISCLAIMS ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE OR NON-INFRINGEMENT. IBM shall have no responsibility to update this information. IBM products are warranted, if at all, according to the terms and conditions of the agreements (for example, IBM Customer Agreement, Statement of Limited Warranty, International Program License Agreement, etc.) under which they are provided. Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products in connection with this publication and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products.

IBM makes no representations or warranties, express or implied, regarding non-IBM products and services.

The provision of the information contained herein is not intended to, and does not, grant any right or license under any IBM patents or copyrights. Inquiries regarding patent or copyright licenses should be made, in writing, to:

IBM Director of Licensing
IBM Corporation
North Castle Drive
Armonk, NY 10504-1785
U.S.A.

Performance is based on measurements and projections using standard IBM benchmarks in a controlled environment. All customer examples described are presented as illustrations of how those customers have used IBM products and the results they may have achieved. The actual throughput or performance that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput or performance improvements equivalent to the ratios stated here.

© Copyright International Business Machines Corporation 2008. All rights reserved.

Note to U.S. Government Users - Documentation related to restricted rights-Use, duplication or disclosure is subject to restrictions set forth in GSA ADP Schedule Contract and IBM Corp.

