TPF Users Group - 2011

Communications Update

Name: Mark Gambino
Venue: Communications Subcommittee

© 2011 IBM Corporation

It's Fun to Stay at the I-E-D-N

Some zEnterprise Terminology

- **z196**
  - Traditional System z central processing complex (CPC) with LPARs running operating systems like z/TPF, z/OS, z/VM, and Linux on System z

- **zEnterprise BladeCenter Extension (zBX) - optional**
  - Contains racks of accelerators and application server blades (ASBs) that help process System z workloads

- **zEnterprise Node**
  - One z196 along with its optional zBX

- **zEnterprise Ensemble**
  - 1 to 8 zEnterprise nodes that are centrally managed as a single system

- **Intraensemble data network (IEDN)**
  - Internal private data network used for communication between LPARs and blades within the ensemble

zEnterprise Node Data Networks

zEnterprise Node

z196   zBX

IEDN
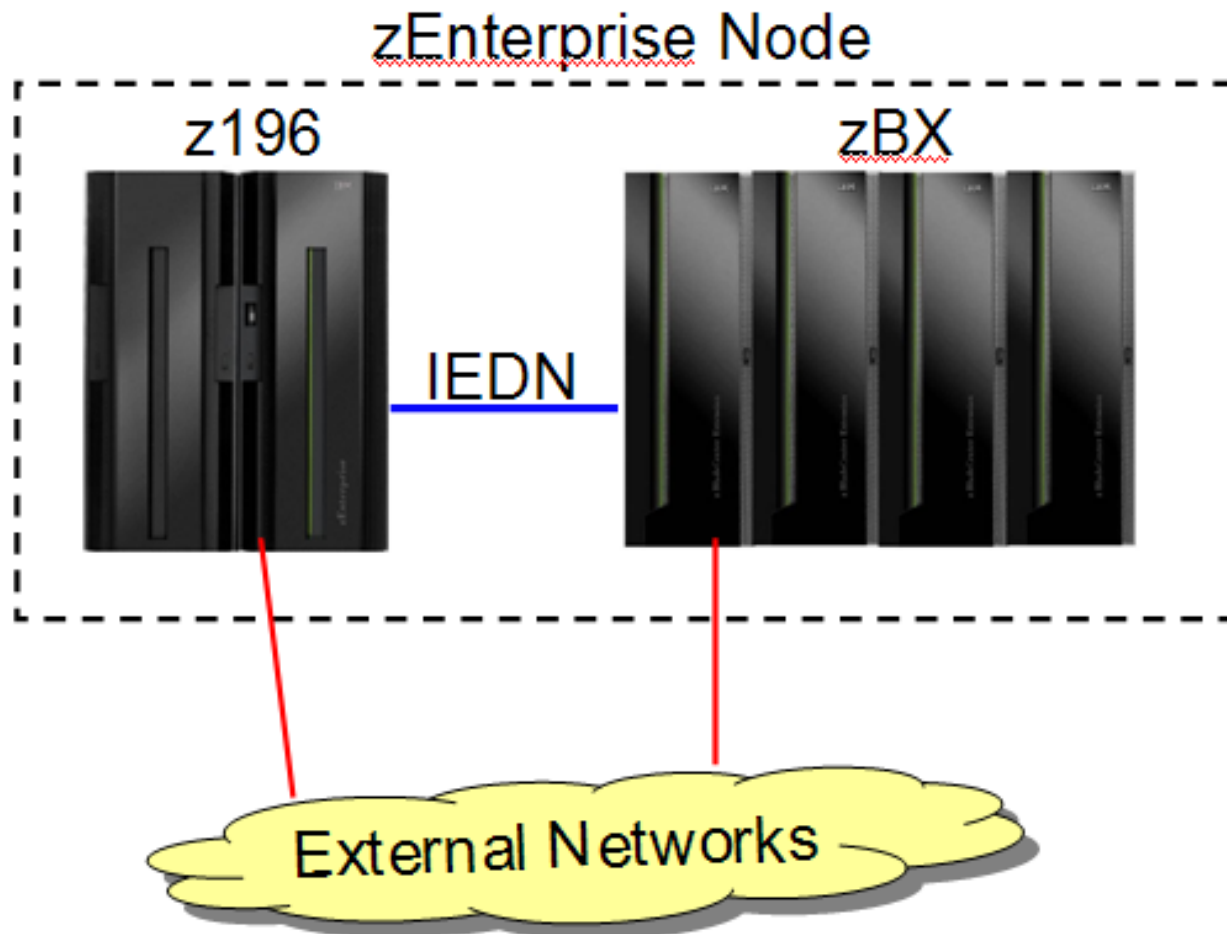
External Networks

zEnterprise Ensemble Data Networking with 3 Nodes
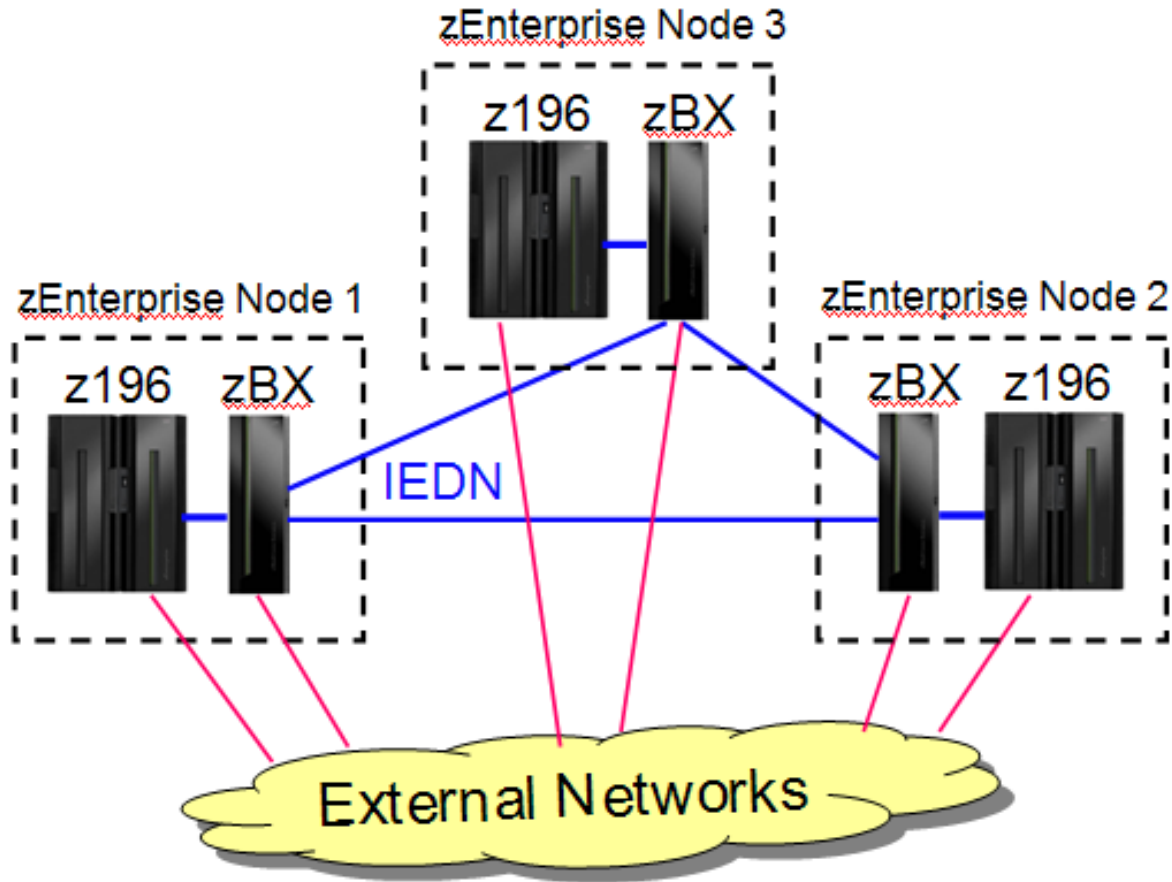
# IEDN Properties and Benefits

- **High speed internal network (10 GbE based)**

- **Reduce number of external switches, routers, firewalls**

- **Physically secure**
  - Dedicated and isolated internal network components

- **Logically secure**
  - Controlled access (which LPARs are allowed to communicate with which blades)

- **Data encryption/decryption may not be necessary based on your security policy**

IEDN – Can We Talk?

- **IEDN is one physical network that can be subdivided into multiple logical networks using virtual LAN (VLAN) technology**

- **From the zEnterprise Unified Resource Manager (zManager) you define each VLAN along with what LPARs and blades are in that VLAN**

  - A given LPAR or blade can connect to multiple VLANs

- **Operating systems running in LPARs or blades connected to the IEDN are *virtual servers***

- **Hypervisors and internal network components (like OSA-Express) enforce the VLAN definitions**

Sample IEDN Physically Connectivity
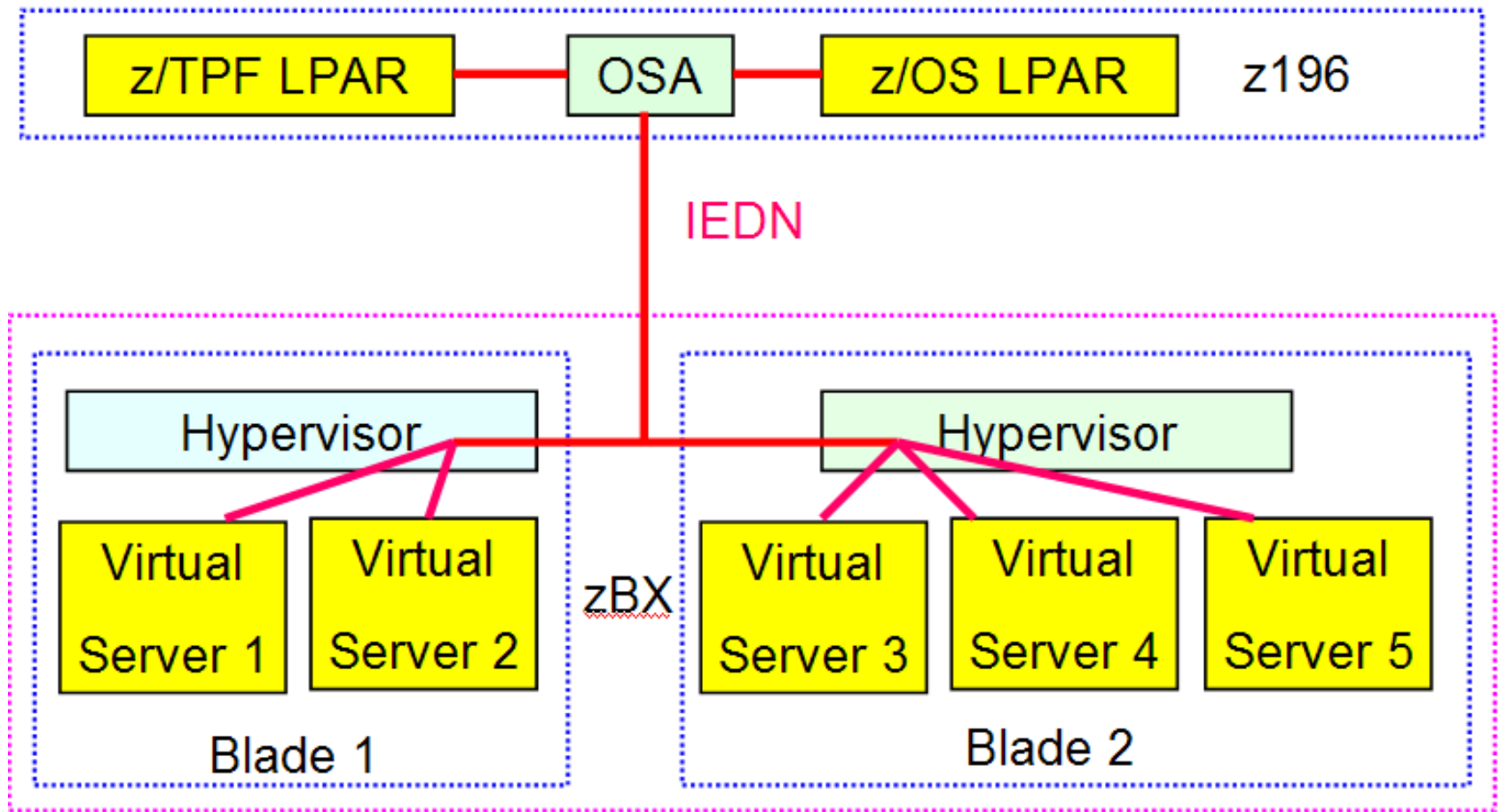
Sample IEDN – VLAN 100

Sample IEDN – VLAN 101

IBM z/Transaction Processing Facility Enterprise Edition 1.1.0
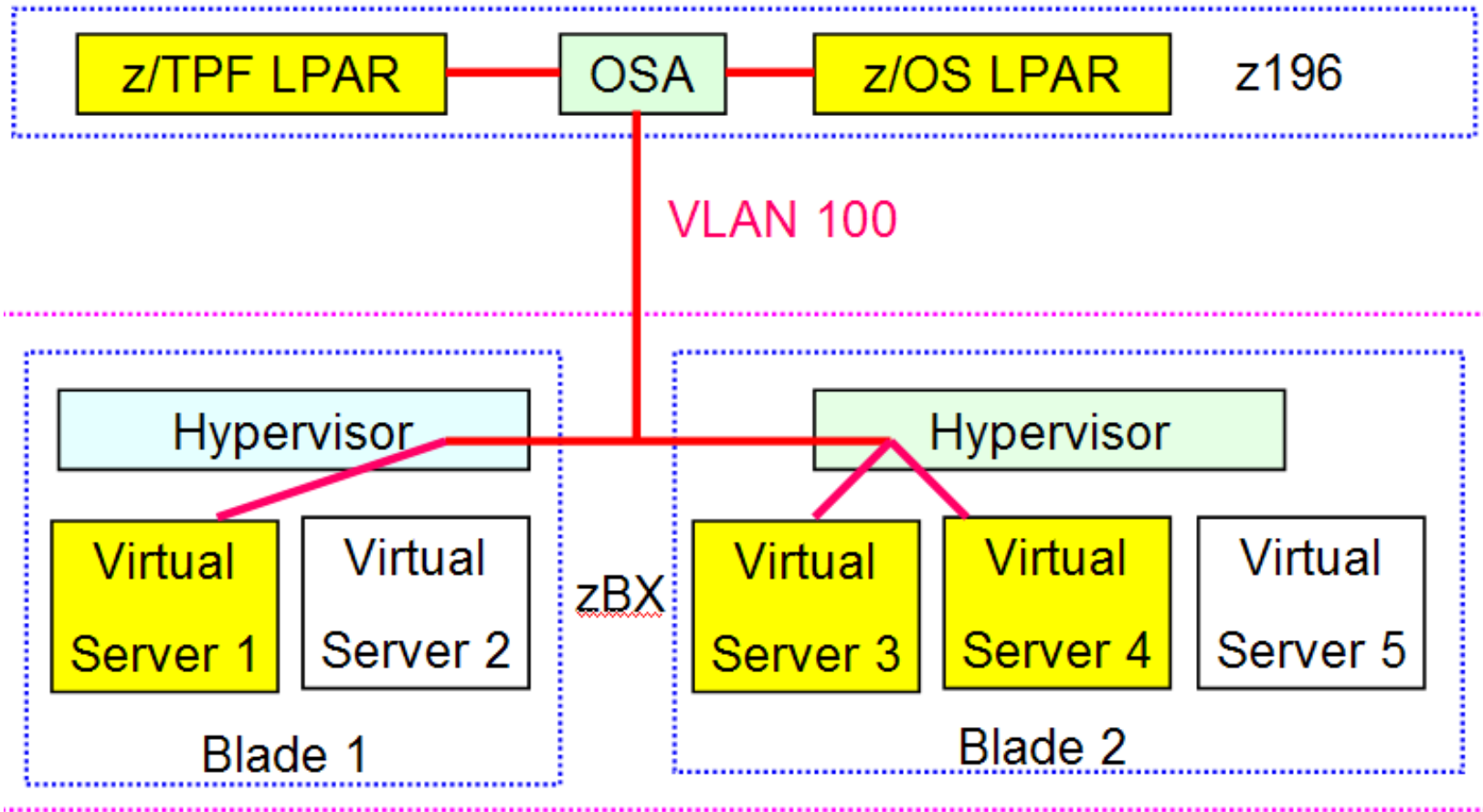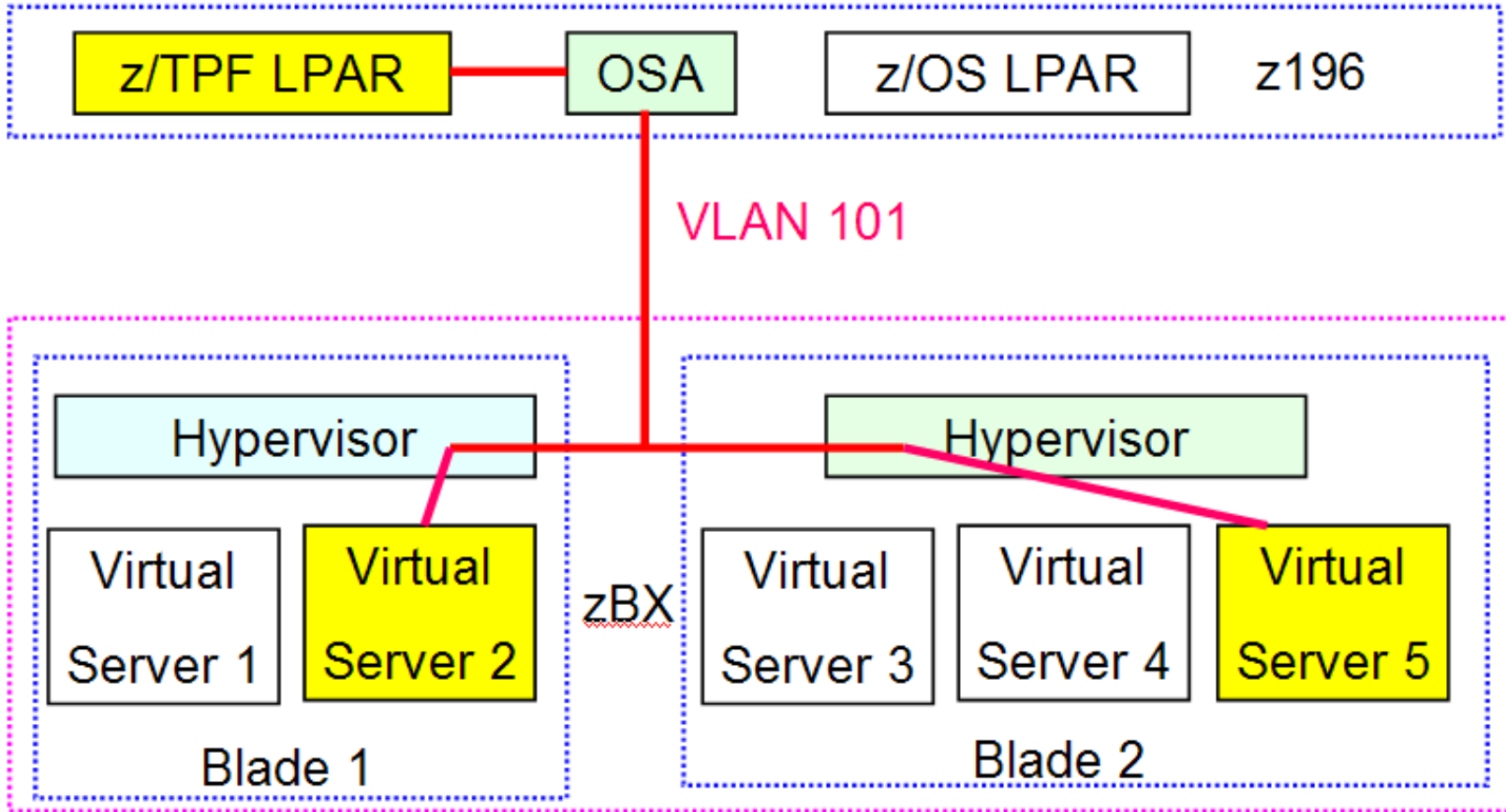
Defining OSA-Express Devices for z/TPF to the CEC

- **Each OSA-Express card is defined in the IOCP**

- **CHPID statement defines the channel path identifier on which this OSA-Express resides**

- **TYPE parameter on the CHPID statement defines the type of device:**

  - OSD – OSA-Express in QDIO mode connected to an external network

  - OSX – OSA-Express connected to the IEDN

Defining OSA-Express Connections to z/TPF

- **IEDN support for z/TPF is APAR PJ38383**

- **Use the ZOSAE DEFINE command**
  - NET parameter defines the type of network to which this OSA-Express connects:
    - FENET - Fast Ethernet
    - GENET – 1 Gigabit Ethernet
    - 10GBE – 10 Gigabit Ethernet
    - IEDN – Intraensemble Data Network
  - VLAN parameter defines the VLAN identifier
    - A VLAN ID must be defined when NET-IEDN is specified

Network Type Validation

- **When z/TPF activates an OSA-Express connection (ZTTCP START command), the handshake between z/TPF and OSA-Express verifies that the network type definition matches:**

  - If OSA-Express is connected to the IEDN (CHPID type is OSX), the connection must be defined to z/TPF as NET-IEDN

  - If OSA-Express is not connected to the IEDN (CHPID type is not OSX), the connection must not be defined to z/TPF as NET-IEDN

- **If there is a mismatch of definitions, the activation attempt will fail and produce a new console error message**

  - TTCP0164E or TTCP0165E

VLAN Support for External Networks

- **z/TPF now supports VLANs on OSA-Express connections to external networks**

- **VLAN parameter on ZOSAE DEFINE command is optional**
  - VLAN-0 means no VLAN for this connection
    - This is the default value if VLAN is not specified and is the value assigned to all connections defined prior to installing this support (APAR PJ38383)
  - VLAN ID is specified when defining a primary OSA-Express connection
    - Same VLAN ID value is set for the corresponding backup OSA-Express connection, if one exists

VLAN ID Included in Diagnostic Data

- **VLAN ID is included in the output of the ZOSAE DISPLAY command**

- **VLAN ID is included in the output of the ZSOCK DISPLAY FORMAT command**

- **VLAN ID is also included in IP trace for each packet sent or received**

  - Online system-wide IP trace display (ZIPTR FORMAT)

  - Individual IP trace trace (ZINIP FORMAT)

  - Offline IP trace (IPTPRT)

    - VLAN ID is also an optional input search criteria for IPTPRT
    - VLAN ID is included in packet capture (PCAP) format of data that open source tools can analyze

Show Me The ~~Money~~

Data

More Data Collected on a per Socket Basis

- **Date and time when the socket was created**

- **Number of bytes of user data sent**

- **Number of bytes of user data received**

- **For TCP listener sockets:**

  - Backlog value specified on the *listen*() API

  - Current backlog value

  - Number of connections accepted by this server

  - Number of connection requests rejected because the backlog limit has been reached

Additional Data Displayed for a Socket

- **The hostname of the remote partner**

  - Based on DNS lookup of the remote IP address

- **Token and program name passed as input on an *activate on receipt* (AOR) type API**

  - Values are displayed only if AOR is pending for this socket

- **VLAN ID**

- **What data is displayed depends on the type of socket:**

  - TCP listener socket

  - TCP connected socket

  - Other socket (UDP or RAW)

Partial ZSOCK DISPLAY FORMAT Sample Output for a TCP Connected Socket

```
SOCK0043I 10.45.54 TCP SOCKET CONTENTS FORMATTED

   LOCAL IP  -          9.117.236.131  LOCAL PORT  -          5005

   REMOTE IP -          9.117.241.140  REMOTE PORT -          1040

   PROTOCOL -                    TCP  SOCKET TYPE -       STREAM

   SOCKET DESCRIPTOR -      00C0001F  1052 STATE  -            Y

   FIRST HOP IP -       9.117.241.140  VLAN ID -               34

   SEND BUFF SIZE -            131072  SEND BUFF IN USE -     100

   RECV BUFF SIZE -            131072  RECV BUFF IN USE -       0

   BYTES SENT     -            134678  BYTES RECEIVED   -    7878

   DNS NAME - WWW.SECRET.TEST.SYSTEM.ORG

   AOR PENDING -                    Y

   AOR TOKEN   -      C8C3C2C6C9C3C2C1  AOR PROGRAM NAME -    QXYD

   SOCKET CREATED -  Tue Oct 5 16:18:55 2011
```

ZSOCK DISPLAY FORMAT Sample Output
for a TCP Listener Socket

```
SOCK0044I 10.45.54 TCP LISTENER SOCKET CONTENTS FORMATTED

    LOCAL IP  -          9.117.236.131 LOCAL PORT  -      5005

    PROTOCOL -                     TCP SOCKET TYPE -    STREAM

    SOCKET DESCRIPTOR -      00C0001E 1052 STATE   -        Y

    VLAN ID -                  1234

    STATE -                 LISTEN

    CONNECTIONS ACCEPTED - 1243832943

    BACKLOG EXCEEDED -            123

    BACKLOG LIMIT -               20

    BACKLOG CURRENT -              3

    SOCKET CREATED -  Tue Oct 5 16:18:55 2011

END OF DISPLAY
```

Display Socket Options

- **New SOCKOPT option on ZSOCK DISPLAY FORMAT command**

- **Displays the name and value of socket options that can be set for a given socket via the *setsockopt*() and *ioctl*() APIs**

- **Options that are applicable only for TCP sockets are only displayed if the socket is TCP**

ZSOCK DISPLAY SOCKOPT Sample Output for a TCP Socket

```
SOCK0047I 10.45.54 TCP SOCKET OPTIONS

  SOCKET DESCRIPTOR - 00C0001F      PROTOCOL -            TCP

  SO_INPUT_PRIORITY –          0

  SO_RCVBUF –             131072    SO_RCVLOWAT -         512

  SO_SNDBUF –             131072    SO_SNDLOWAT –         512

  SO_RCVTIMEO -              100     SO_SNDTIMEO –         100

  SO_TCPDELAY -               1      SO_TCPDUACK -           2

  SO_KEEPALIVE -              Y      SO_OOBINLINE -          N

  SO_REUSEADDR -              N      SO_LINGER -             Y

  AOR_BALANCE -               Y      FIONBIO -               Y

  TCP_PSH_LAST -              N      TPF_NOSLOWSTART –       N

  TPF_NOSWEEP -               Y      TPF_RST_IPL -           N

  TPF_SURVIVE_CYCLE -         N      TPF_RETURN_MESSAGES –  N

END OF DISPLAY
```

Socket Monitor

- **Monitors sockets for conditions that *may* indicate a problem or tuning is necessary**

- **New SOCKMON parameter in CTK2 (and ZNKEY command) to enable or disable the monitor**

  - Socket monitor is disabled by default

- **New user exit USMO is called when condition is detected allowing you to do things like:**

  - Log the incident

  - Notify the operator to investigate or take some action

  - Close the socket or stop the application

Socket Monitor User Exit (USMO) Interface

- **The following data is always passed to USMO:**

  - Socket descriptor

  - Local IP address and port

  - Remote IP address and port

  - Condition that has been detected

- **Additional data is passed to USMO based on which condition was detected**

Condition - Remote Partner not Reading Data

- **z/TPF output packets are waiting to be sent over a socket but the TCP window has been closed for more than 30 seconds**

- **Additional data passed to USMO:**

  - Send buffer size of the socket

  - Send buffer in use count

  - Number of bytes sent over the lifetime of this socket

Condition - z/TPF Application not Reading Data

- **Input messages have been queued to a socket for more than 10 seconds without the z/TPF application reading them**

- **Additional data passed to USMO:**

  - Number of bytes read over the lifetime of this socket

Condition - z/TPF Socket Receive Buffer is Full

- **z/TPF sent a TCP window size of 0 because the socket receive buffer is full**

- **Additional data passed to USMO:**

  - Receive buffer size of the socket

  - Number of times z/TPF has sent a TCP window size of 0 on this socket

- **USMO will be called at most once per minute per socket for this condition**

Condition - Connection Request Rejected due to Backlog Limit

- **z/TPF rejected a connection request from a remote TCP client because the server is at its backlog limit**

- **Additional data passed to USMO:**
  - Backlog limit
  - Number of connection requests rejected for this server because the backlog limit has been reached
  - Number of connection requests accepted by this server

- **USMO will be called at most once per minute per socket for this condition**

Bean

Counting

Core Block ECB Owners in z/TPF

- **z/TPF added the ability to assign ECB owner names to core blocks to show how many of each block type is used by a given application/owner.**

- **OWNER parameter on ZSTAT command shows block usage by ECB owner name**

- **Internet daemon (INETD) is started during the cycle up process and all servers (ECBs) started by INETD inherit the cycle up ECB's owner name "ISMP.ZCYCL"**

  - Makes it difficult to identify which resources are used by which server application

New ECB Owner Names for INETD and SSL

- **INETD monitor ECBs (includes NOLISTEN model):**

  - ECB owner name is "**INETD.MONITOR.subsystem**" where "subsystem" is the subsystem where the ECB is running

- **Other INETD server and application ECBs:**

  - ECB owner name is "**INETD.LISTENER.servername-subsystem**" where "servername" is the name of the server defined to INETD

- **Shared SSL daemon ECBs:**

  - ECB owner name is "**ISSL.SERVER.PROCNUMx-subsystem**" where "x" is the SSL daemon process number

Summary

- **IEDN support (APAR PJ38383)**

- **VLAN support (APAR PJ38383)**

- **Additional data collected per socket (APAR PJ38349)**

- **Enhanced socket (ZSOCK) displays including socket options (APAR PJ38349)**

- **Socket monitor (APAR PJ38349)**

- **ECB owner names for INETD and shared SSL (APAR PJ38385)**

Trademarks

- **IBM and zEnterprise are trademarks of International Business Machines Corporation in the United States, other countries, or both.**

- **Linux is a trademark of Linus Torvalds in the United States, other countries, or both.**

- **Other company, product, or service names may be trademarks or service marks of others.**

- **Notes**

- **Performance is in Internal Throughput Rate (ITR) ratio based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput 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 improvements equivalent to the performance ratios stated here.**

- **All customer examples cited or described in this presentation are presented as illustrations of the manner in which some customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics will vary depending on individual customer configurations and conditions.**

- **This publication was produced in the United States. IBM may not offer the products, services or features discussed in this document in other countries, and the information may be subject to change without notice. Consult your local IBM business contact for information on the product or services available in your area.**

- **All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.**

- **Information about non-IBM products is obtained from the manufacturers of those products or their published announcements. IBM has not tested those products and cannot confirm the performance, compatibility, or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.**

- **Prices subject to change without notice. Contact your IBM representative or Business Partner for the most current pricing in your geography.**

- **This presentation and the claims outlined in it were reviewed for compliance with US law. Adaptations of these claims for use in other geographies must be reviewed by the local country counsel for compliance with local laws.**