z/TPF V1.1

# TPF Users Group - 2010

# Communications and Security Update

Name: Mark Gambino
Venue: Communications Subcommittee

**AIM Enterprise Platform Software**
**IBM z/Transaction Processing Facility Enterprise Edition 1.1.0**

Any reference to future plans are for planning purposes only. IBM reserves the right to change those plans at its discretion. Any reliance on such a disclosure is solely at your own risk. IBM makes no commitment to provide additional information in the future.

# Online IP Message Counters

- **Output of ZSTAT command includes the number of TCP/IP input messages and output messages since the last IPL**
- **Message counters were 4 bytes**
  - Counter would wrap after 4G worth of messages
- **APAR PJ37312 expands the counters to 8 bytes**

# Improving TCP/IP Stack Performance

- **Multiple I-streams attempting to obtain the same lock in memory results in lock contention**
- **APAR PJ37312 reduces the overhead associated with contention on the socket block (TCP/IP) lock**
  - Most applicable for z/TPF systems running multiple I-streams on modern processors

# Slow Socket Sweeper Support

- **Socket sweeper detects and cleans up sockets that have not been used by any application for a certain amount of time**
  - Based on the SOCKSWP value in CTK2
- **The TPF_NOSWEEP option on the *ioctl* function allows a socket to remain active, even if no application has used it for a long period of time**
- **APAR PJ37586 introduces a slow socket sweeper option**
  - New SLOWSWP parameter in CTK2 defines how many hours to wait before cleaning up a socket that has not been used even though it is marked as TPF_NOSWEEP
  - If SLOWSWP=0 (which is the default value) or SOCKSWP=0, the slow socket sweeper is disabled

# Local Sockets Support

- **Allows a TCP/IP client application to communicate with a server application that resides on the same z/TPF system without needing/using a real network**
- **When an output packet is built, if the destination IP address is one of the IP addresses of this z/TPF system:**
  - Packet is placed in a 4K frame and placed on the input list
  - When the input list is processed, the packet is passed to the TCP/IP stack making it look like a packet was received from the network

# Local Socket Support Enhancements

- **APAR PJ37312 improves local sockets behavior**
- **Output packets are now passed in the IP message table (IPMT) rather than 4K frames**
  - Eliminates spikes in 4K frame usage by local sockets
- **Output packets are now placed on the OSA input message queue rather than on the input list**
  - Allows local sockets input messages to be processed by any I-stream
  - Eliminates CPU utilization spike on one I-stream that generates a large amount of local sockets messages

# OSA-Express Communications

- **z/TPF communicates to OSA-Express using the Queued Direct I/O (QDIO) protocol**
- **QDIO already has interrupt reduction algorithms built in to make the protocol very efficient**
  - If the host has sent output packets to OSA that OSA is still processing and the host now has more output packets to send, no need to signal OSA again
  - If OSA has presented input packets to the host that the host is still processing and OSA now has more input packets to deliver, no need to interrupt the host again

# Optimized Latency Mode (OLM)

- **New option on OSA-Express3 adapters**
- **Reduces latency even more**
  - After OSA has processed all pending output packets from the host, OSA automatically looks for more output packets from this host for a short period of time
    - If the host generates more output packets in this period of time, host no longer needs to signal OSA
  - When the host has processed all input packets from OSA and new packet(s) arrive from the network, OSA sends the interrupt earlier to wake up the host

# OSA-Express Sharing and OLM

- **One OSA-Express adapter can be shared by many hosts (LPARs and z/VM guests)**
  - For example, dozens of z/TPF test systems running as z/VM guests can share the same OSA-Express adapter
- **To provide the best quality of service, the number of host connections is limited when using OLM**
  - Currently on OSA-Express3, a maximum of 4 host connections per network port is allowed if any of those connections have OLM enabled

# OLM Support for z/TPF

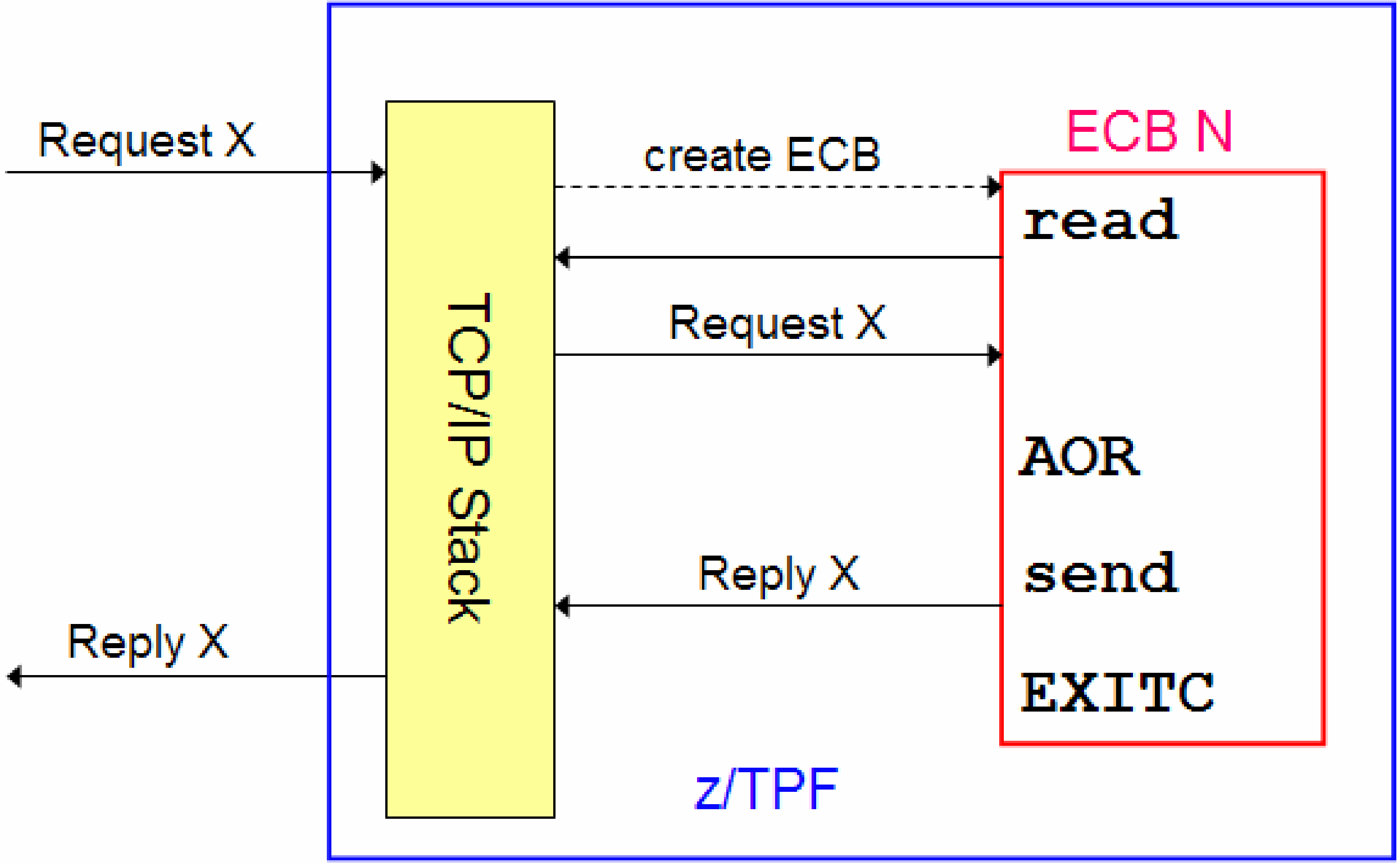
- **APAR PJ37341 provides OLM support for z/TPF**
- **New OLM parameter on ZOASE DEFINE and MODIFY commands**
  - OLM=YES
    - Activate the connection using OLM if OLM is support by the adapter
      - If adapter does not support OLM, the connection is activated without using OLM
  - OLM=NO (default)
    - Activate the connection without using OLM

# activate_on_receipt (AOR) Processing

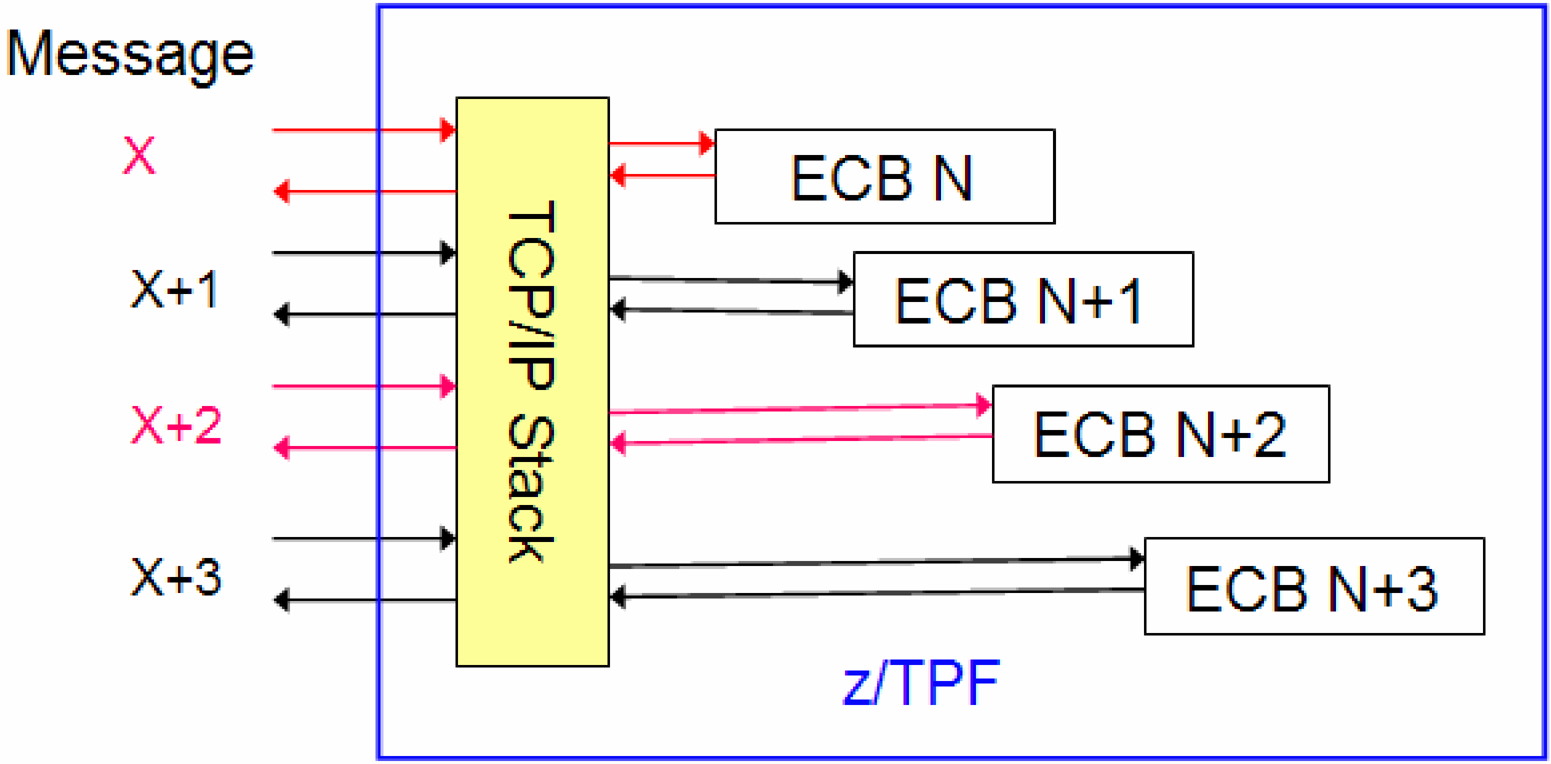
- **The AOR family of APIs allows a socket to remain active without any active application ECB when waiting for data to arrive from the remote TCP/IP application.**
  - When data arrives from the network, a new application ECB is created to read and process the data
- **Typical application design using AOR:**

1. ECB N created when data arrives on the socket
2. ECB N issues *read* API to read message X
3. ECB N issues *AOR* API that will create ECB N+1 when message X+1 arrives from the network
4. ECB N processes message X, then exits

# Request/Reply Model using AOR



# Request/Reply ECBs using AOR



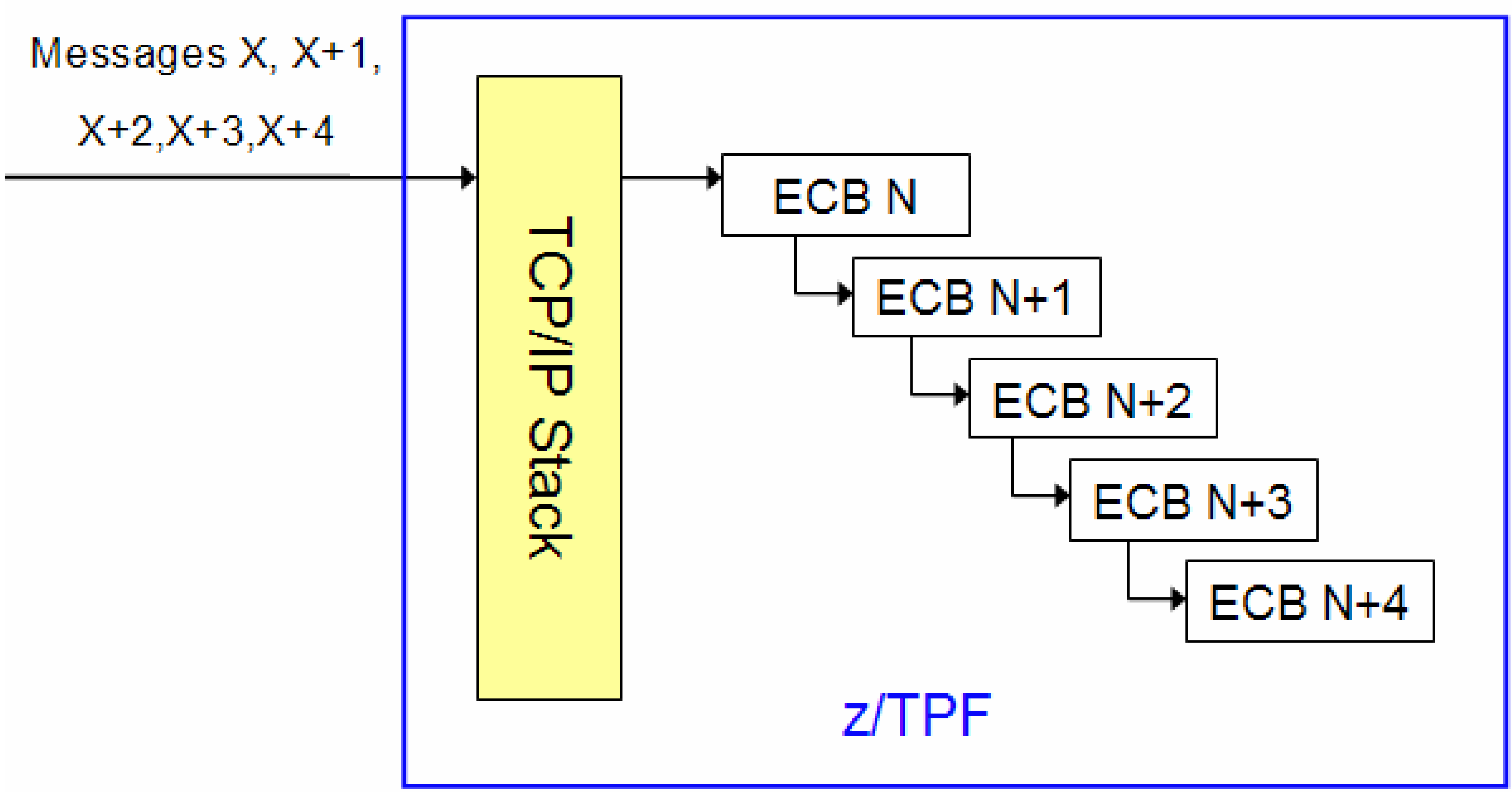
# TCP/IP Input Message Throttling

- **TCP/IP input messages are controlled by input list shutdown checks**
  - New input messages are not read from the network if z/TPF is in an input list shutdown condition
- **Sockets is a pull model architecture**
  - Application program must ask for new data (via read or AOR)
  - If new data arrives from the network before the application asks for it, that data is queued in the socket's input buffer

# Pipeline Sockets

- **Sockets is a full duplex architecture**
- **Besides request/reply model, there is a pipeline model where many input messages are sent over one socket at a time**
  - Application level responses might flow over the same socket or a different socket
- **Pipeline sockets are an easy way to increase throughput without having to increase the number of network/socket connections**
- **One batch of input messages might create many ECBs for servers using the AOR model**

# Pipeline Socket ECBs using AOR



# AOR Throttling Enhancements

- **APAR PJ38132 changes AOR API processing for pipeline sockets**
- **If an AOR API is issued and there is already enough data in the socket's input buffer to post the AOR**
  - The newly created ECB by AOR is now placed on the input list to ensure the next message for this socket does not begin application processing unless/until there are sufficient system resources available
- **Similar changes were made to the SSL_aor API**

# Dynamic Program Loads (E-type Loader)

- **Each ECB has an activation number that determines which version of an E-type (realtime) program will be used by that ECB if multiple versions of the program are active**
- **A new version of a program can be loaded via the E-type loader (ZOLDR command) while applications are active and processing messages**
  - Messages currently being processed by active ECBs continue to use the old version of the program
  - New messages in new ECBs will be processed using the new version of the program

# ZOLDR Operations and Long Running ECBs

- **Long running system ECBs, such as the Internet daemon (INETD) and servers managed by INETD, recycle themselves when a ZOLDR operation occurs**
  - Existing long running ECBs exit and new long running ECBs are created that can use the latest versions of programs
- **When new programs in a loadset have been tested and you want to make those programs part of the base, issue the ZOLDR ACCEPT command**
- **ACCEPT processing must wait for all existing ECBs to exit before the operation can be completed**
  - Existing ECBs might call a program that is in the loadset and; therefore, must use the old version of that program
- **Recycling long running server ECBs can be disruptive to network connections, message traffic, or both**

# To Recycle or to Not Recycle

- **APAR PJ37377 provides a new E-type loader recycle interface that long running ECBs can use to determine whether they need to recycle when ZOLDR operations occur**
- **When a long running server ECB detects that the system activation number has changed (such as a ZOLDR command was issued), it can now issue a new API passing the list of program names that are part of or used by this server**
  - If there are newer versions of any of the programs in the list passed on the API (or newer versions of any of the z/TPF standard libraries), the API return code tells the ECB that it needs to recycle
  - If there are no new versions of programs relevant to this server, the system will update the activation number in the ECB to be the current system activation number and the return code will indicate that no recycle is necessary
    - Long running ECB remains active without any disruption to traffic

# System Processes that Use the Recycle Interface

- **With APAR PJ37377, the following long running system processes (ECBs) use the new recycle interface**
  - INETD
  - Shared SSL daemons
    - Console message sent if shared SSL daemons need to recycle
  - SYSLOG daemon
  - Advanced HTTP client
  - OpenLDAP
  - MYSQL
- **APAR PJ37900 updates Apache 2.2 to use the new recycle interface**

# Recycle Interface for User Programs/Servers

- **Long running application ECBs can also use the recycle interface**
  - *tpf_etype_loader_recycle_interface* API
- **Option on API to update activation numbers in this ECB as well sibling and thread ECBs created by this ECB**
  - ECBs created using *pthread_create*, *fork*, or *tpf_fork*
- **User exit UERL allows you to specify user libraries that require ECBs to recycle if they are in a loadset**
- **User exit UERA is called to obtain the list of programs that make up a specific user server defined to INETD as  MODEL=DAEMON and the list of programs called by this server**
  - New CALLLIST option on maketpf build tool generates a list of programs that are called by that shared object

# 3215 Console Support

- **z/TPF customers with 3215 console support use:**
  - Channel-attached 3215 devices
  - OSA-ICC 3215 adapters (z9 or beyond)
- **At IPL time, if no 3215 addresses can be mounted and written to, the system goes into wait state until an interrupt is received from one of the 3215 devices**

# 3215 Console Fallback Enhancements

- **APAR PJ37340 creates an option to automatically fallback to and use the hardware management console (HMC or SE) operating system messages interface**
- **New INCLUDEHMC statement on CRASTB statement in SIP**
  - INCLUDEHMC=YES means try and use HMC when:
    - No 3215 devices are available at IPL time
    - z/TPF is up and running, then the primary and alternate 3215 devices fail
  - INCLUDEHMC=NO means do not try and use HMC (default)
    - Manual fallback (ZACRS FBK) to HMC is still possible
- **Automatic fallback to the HMC is only attempted when z/TPF is running native**
  - INCLUDEHMC=YES is ignored when z/TPF is running as a z/VM guest

# z/TPF Keystore Backup and Restore

- **z/TPF keystore contains secure symmetric keys and RSA public/private key pairs**
- **For disaster recovering, a backup copy of the keystore can be created in a file that is then sent (FTP'd) to a remote system for safe keeping**
  - Can restore the z/TPF keystore from a backup copy
- **Backup copy of the keystore can optionally be password protected**
  - Must specify the correct password to restore the keystore from this backup copy
  - Prevents a backup copy of one z/TPF system's keystore from being (mis)used by another z/TPF system

# New Keystore Protection

- **APAR PJ37831 give you the ability to prevent z/TPF test systems copied from a production system from using secure keys to decrypt production data**
- **New ZKEYS PASSWORD ASSIGN command allows you to assign a password to the online copy of the z/TPF keystore**
  - This is a different password than the one used to protect the backup copy of the z/TPF keystore

# How the New Keystore Protection Works

- **If a password has been assigned to the online keystore:**
  - Keystore is disabled at IPL time
  - ZKEYS PASSWORD ENABLE command must be issued during system restart specifying the correct password to enable the keystore
    - The password itself is not echoed to the console
    - You can configure the TPF Operations Server (TOS) to mask the password in the TOS command and message areas
  - ZKEYS BYPASS ENABLE command allows system restart to continue leaving the keystore as disabled
- **When the keystore is disabled**
  - Any API call attempting to use a secure symmetric key or and RSA private key will be rejected

# Secure Symmetric Key Names

- **A z/TPF secure symmetric key is referenced by key name from application programs**
- **There are two key names for each key:**
  - *Encryption key name*
    - Key name to be used to encrypt data.
    - There can be many keys with the same encryption key name, but only one is active at a time
      - Allows you to change keys without having to make any changes to application programs
  - *Decryption key name*
    - Key name to be used to decrypt data.
    - Each key has a unique decryption key name.

# Get Properties of a Secure Symmetric Key

- **Many applications do not need to know anything about the properties of a given key other than the name of the key**
- **Some applications/middleware do need to know some specifics about a key**
  - APAR PJ37375 gives application programs the ability to obtain the information about a specific secure symmetric key via the new *tpf_get_symmetric_key_info* function

# *tpf_get_symmetric_key_info* Function

- **Input**
  - Either an encryption key name or a decryption key name
- **Output**
  - Encryption key name
  - Decryption key name
  - Cipher name
  - Key size
  - Block size
  - Date and time when the key was last activated (0 if never active)
  - Indicator whether the key is active

# Summary

- **Expanded IP message counters (PJ37312)**
- **Improved TCP/IP stack performance (PJ37312)**
- **Slow socket sweeper socket (PJ37586)**
- **Local sockets enhancements (PJ37312)**
- **Optimized Latency Mode (OLM) support (PJ37341)**
- **activate_on_receipt (AOR) throttling enhancements (PJ38132)**
- **E-type loader recycle interface (PJ37377)**
- **3215 Console Fallback Enhancements (PJ37340)**
- **Additional keystore password protection (PJ37831)**
- **Get secure symmetric key properties API (PJ37375)**

# Trademarks

- IBM is a trademark 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.