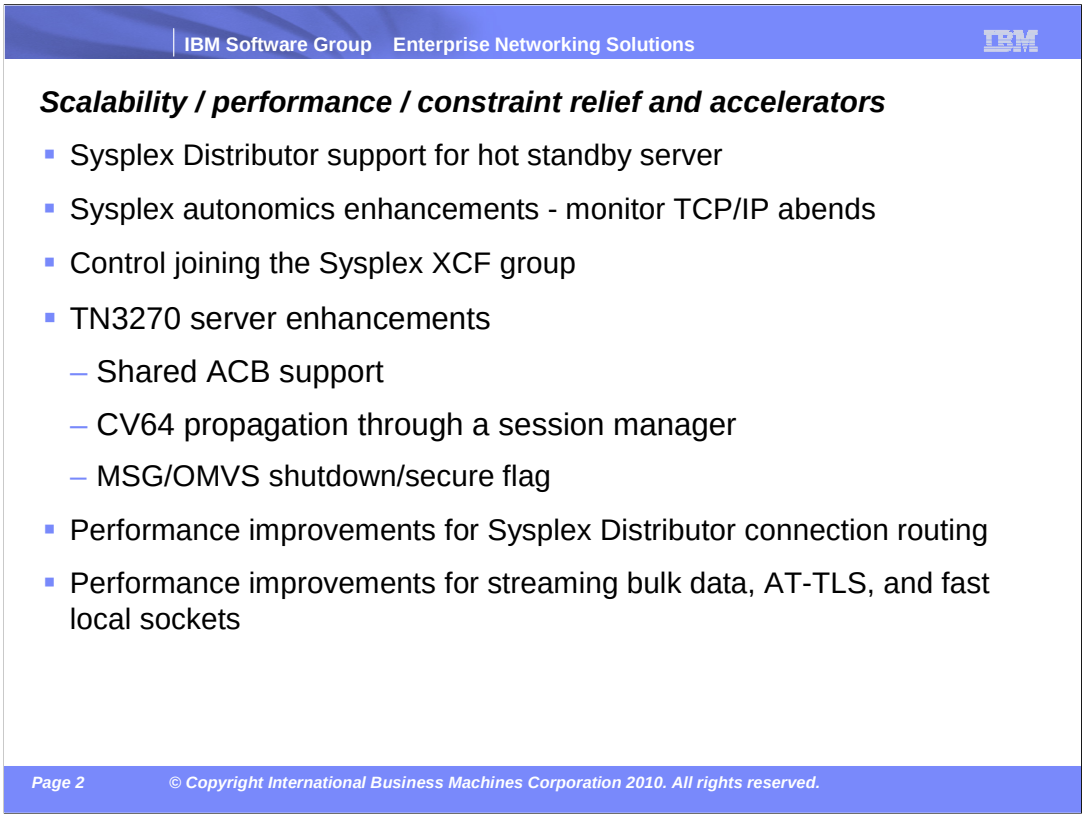
This presentation provides an overview of the new functions in z/OS® V1R12 Communications Server for the Scalability / Performance / Constraint Relief and Accelerators theme.



There are quite a few enhancements in the scalability, performance, constraint relief and accelerators theme.

Sysplex Distributor is enhanced to support a hot standby server.

Sysplex autonomics is improved by monitoring for TCP/IP abnormal ends.

And you can now control whether a given TCP/IP stack joins the Sysplex XCF group.

The TN3270 server now has the ability to share an ACB among multiple LUs, reducing both ECSA and telnet private storage consumption.

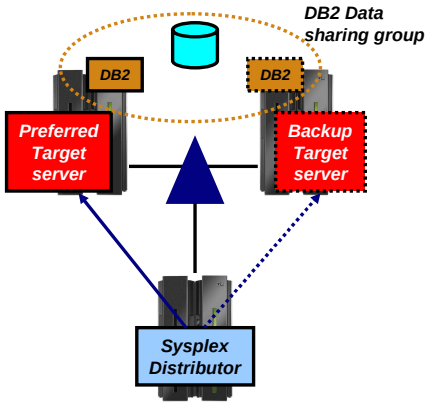
The CV64 control vector that is received by the TN3270 server from a TN3270 client can now be propagated through a session manager.

There are several other miscellaneous telnet enhancements.

Performance for both AT-TLS and fast local sockets is also improved.

Sysplex distributor hot standby support

- A single target server receives all new connection requests
  - Automatically route traffic to a backup target server when the active target server is not available
- Enable using new HOTSTANDBY distribution method
  - One preferred target
    - AUTOSWITCHBACK option - switch to the preferred target if it becomes available
  - And one or more backup targets ranked in order of preference
  - A target is not available when:
    - Not ready OR
    - Route to target is inactive OR
    - If HEALTHSWITCH option configured and target is not healthy

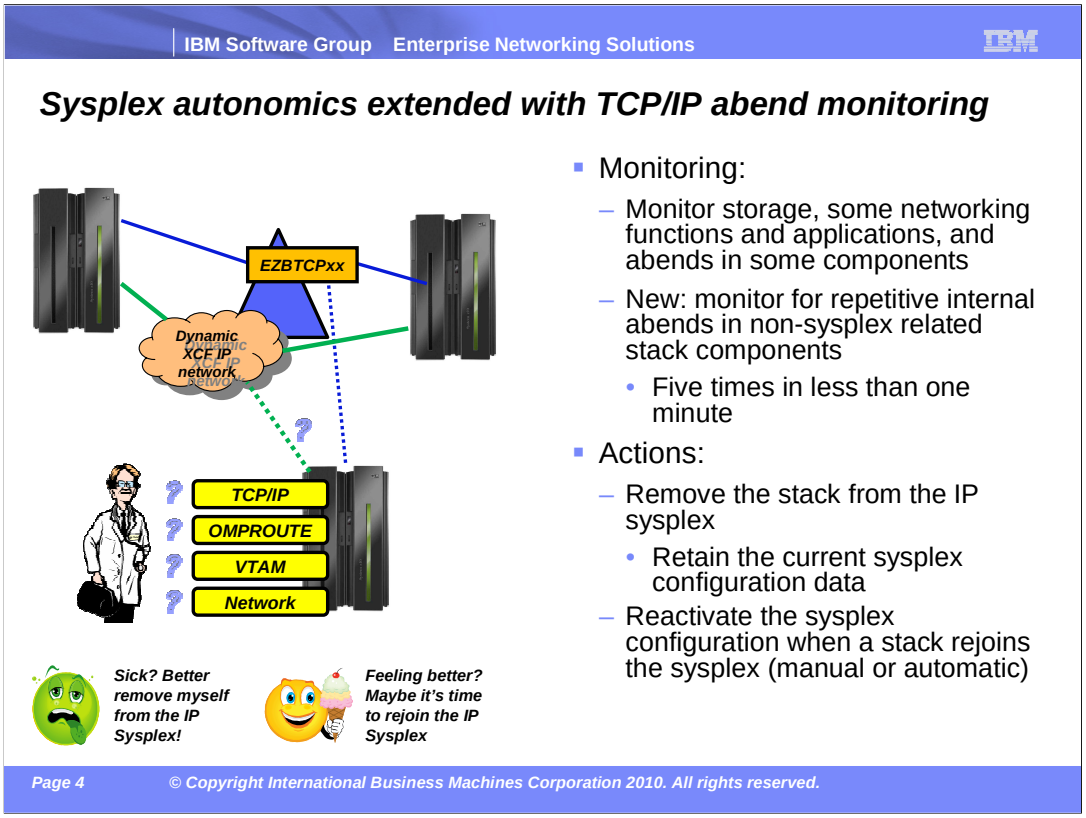



```
VIPADEFINE DVIPA1
VIPADISTRIBUTE DISTMETHOD HOTSTANDBY
  AUTOSWITCHBACK HEALTHSWITCH
  DVIPA1 PORT nnnn
  DESTIP  XCF1 PREFERRED
  DESTIP  XCF2 BACKUP 50
  DESTIP  XCF3 BACKUP 100
```


A new distribution method, HotStandby, is supported.

A single target server receives all new connection requests while other target servers are active but not receiving any new connection requests. Sysplex distributor automatically routes traffic to a backup target server when the active target server is not available.

One target can be designated as a preferred target. If AUTOSWITCHBACK is configured, the distributor will switch to the preferred target when it is available unless the switch from the preferred target occurs because of health problems. Designate one or more targets as backup targets. Use a backup target if the preferred target is not available. A target is not available if it is not ready, the route from distributor to target is inactive, or if HEALTHSWITCH is configured and the target is not healthy. A target is not healthy if the TSR is 0%, the abnormal termination rate is 1000 (out of 1000 transactions), or the server reported health is 0%.



Before V1R12, several indicators are monitored. TCP/IP monitors Communications Server health indicators and makes corrections if storage usage is critical (>90%) for more than TIMERSECS seconds. This includes CSM, TCP/IP private and ECSA storage.

Sysplex autonomics monitors dependent networking functions including OMPROUTE availability, VTAM® availability, and XCF links availability.

Sysplex autonomics monitors for abends in sysplex-related stack components (selected internal components that are vital to sysplex processing). Sysplex autonomics also monitors selected network interface availability and routing.

In z/OS V1R12, sysplex autonomics monitors stacks for multiple abends in a short time. If a stack suffers five or more abends within one minute, these three actions are taken.

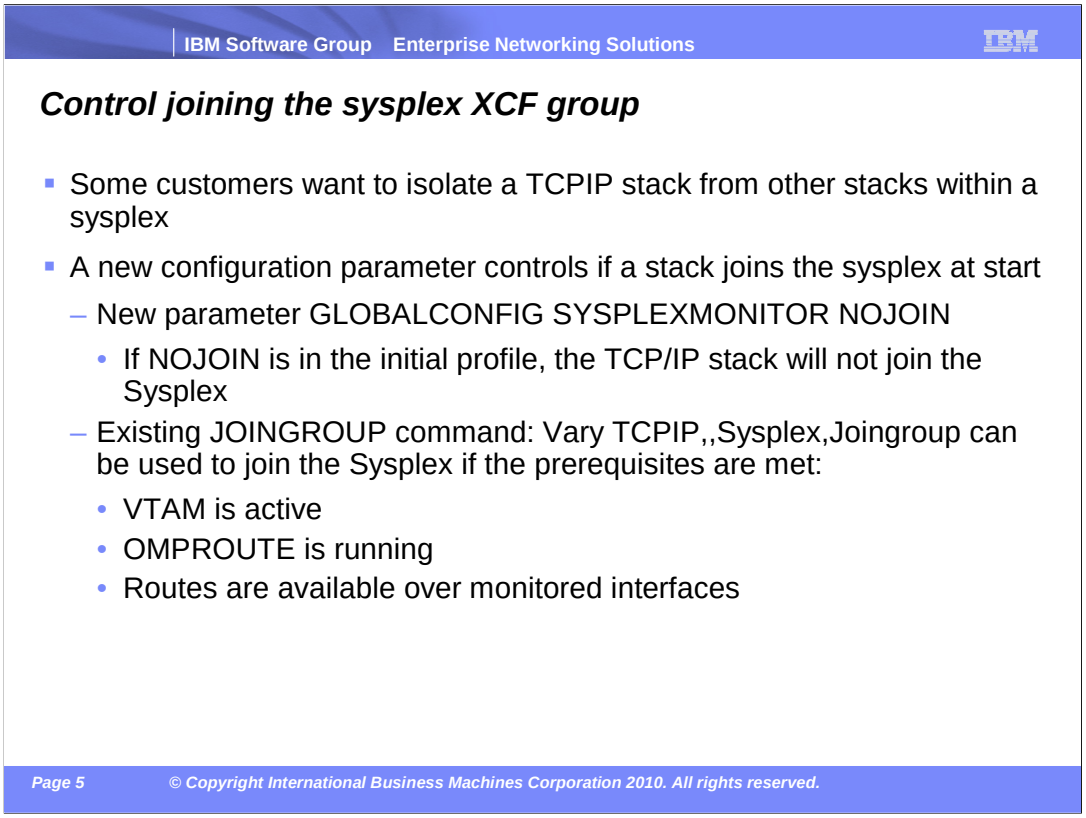
(1) The message EZD1973E MULTIPLE tcpip NONRECOVERABLE ERRORS ARE ADVERSELY AFFECTING SYSPLEX PROCESSING is issued.

(2) If GLOBALCONFIG SYSPLEXMONITOR RECOVERY is enabled, the stack is removed from the sysplex.

(3) The stack's load balancing agent will stop reporting this stack to the load balancing advisor, thus preventing its use by an external load balancer.

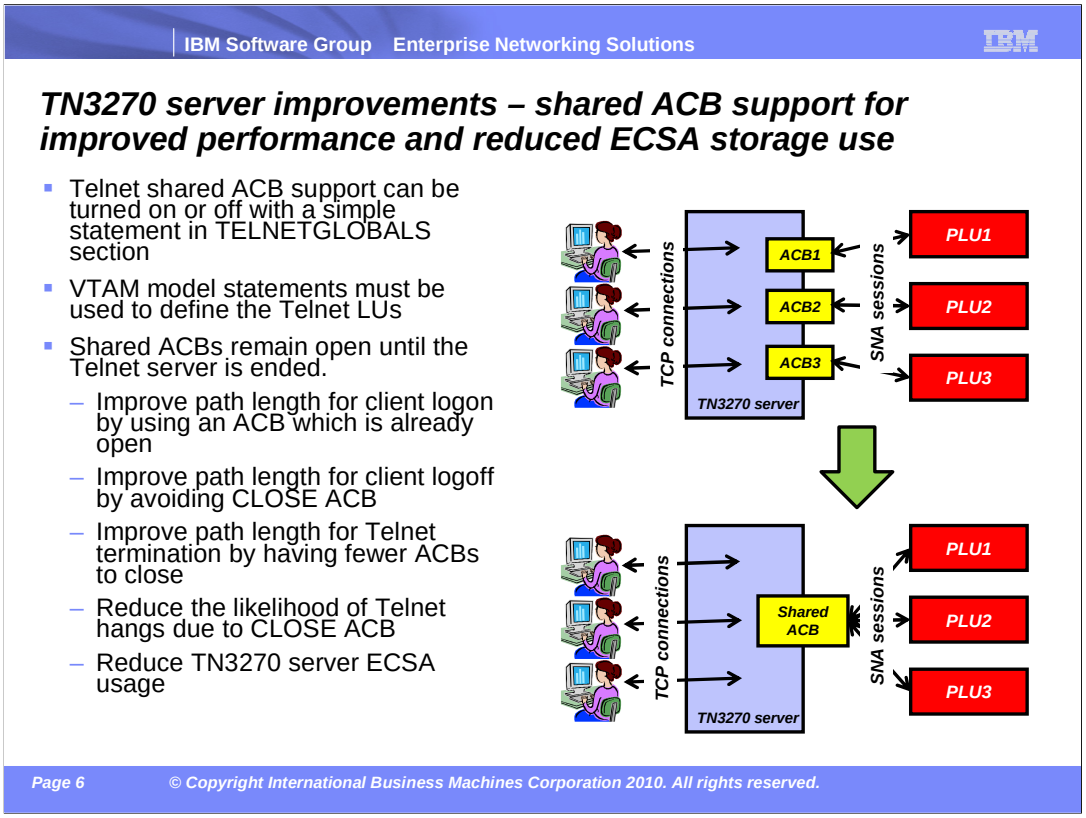
There is nothing new that needs to be configured for this new monitoring.

Sysplex autonomics continues to perform the same actions. It removes the stack from the IP sysplex. It retains the current sysplex configuration data in an inactive state when a stack leaves the sysplex. And it reactivates the currently inactive sysplex configuration when a stack rejoins the sysplex (manually or automatically).



When a sysplex-enabled TCP/IP stack is started, it always joins the sysplex group unless sysplex autonomics detects a problem and delays the join. This happens when VTAM is not active or if there are no routes over monitored network interfaces. The delayed join also happens when GLOBALCONFIG SYSPLEXMONITOR DELAYJOIN is configured and OMPROUTE is not active.

You might want to isolate a TCP/IP stack from other stacks in a sysplex. A new configuration parameter GLOBALCONFIG SYSLEXMONITOR NOJOIN will control if a stack joins the sysplex at startup. If this parameter is in the initial profile, the stack will not join the sysplex. The existing command, Vary TCPIP,,Sysplex,Joingroup can override this parameter. The command will not override existing sysplex autonomics problem detection functions and parameters. The stack will not join the sysplex group until VTAM is active. Remaining SYSPLEXMONITOR parameters are activated as the command is issued. If DELAYJOIN is configured, the stack will not join the sysplex group until OMPROUTE is active.



A telnet logical unit (LU) must be represented by a corresponding VTAM definition such as a static APPL statement or a VTAM model statement.

To establish a session with a target host application, telnet first issues an OPEN ACB for each assigned LU name, followed by a SETLOGON RPL and REQSESS RPL. ECSA based control blocks are allocated by VTAM for every OPEN ACB issued by telnet. The ACB resides in telnet private storage.

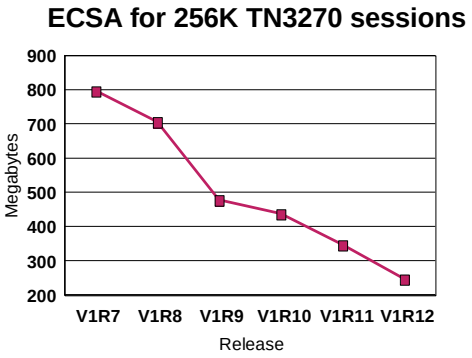
To reduce the amount of ECSA used by VTAM and the amount of telnet private storage used by telnet, telnet will now use a single ACB for multiple telnet client sessions. Four shared ACBs are initially opened for each telnet port. Each ACB can be assigned 300 clients. New shared ACBs are opened as the client workload increases for a given port. The client workload per ACB increases from 300 to a maximum value of 1100 clients per ACB.

Using shared ACBs not only saves storage, it has an additional benefit. Path length is improved for logon, logoff, and termination processing because telnet does not have to issue OPEN ACB or CLOSE ACB for every session.

A new statement is provided to enable the function. The default is NOSHAREACB, which disables the new support. SHAREACB/NOSHAREACB statement can be coded only in the TELNETGLOBALS statement block.

TN3270 server ECSA usage improvement up to and including z/OS V1R12 Communications Server

Release	ECSA for 256K TN3270 sessions
V1R7	798M
V1R8	708M
V1R9	480M
V1R10	440M
V1R11	347M
V1R12 (1)	249M



The numbers are configuration dependent, but they should give you an idea of the magnitude of the savings achieved in the recent releases.

Note (1): The V1R12 number is a preliminary number based on use of shared ACBs - it may change before general availability of z/OS V1R12 Communications Server

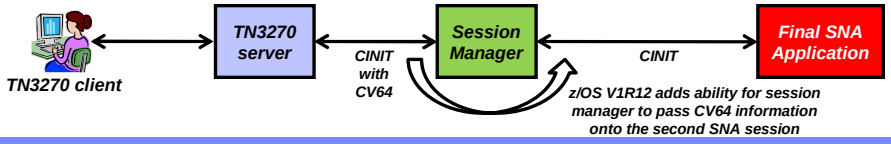
You can see that this new function reduced ECSA usage by almost 100M for a system with 256 thousand telnet sessions. There is also about 20M in telnet private savings for this configuration as well.

TN3270 server improvements – IP management information through a relay-mode session manager

- TN3270 server passes selected IP management information to the SNA side by way of a control vector known as a “CV64”
  - CV64 includes client IP address, port, and optionally host name
  - A VTAM display of the Telnet LU includes some IP information

```
IST1727I DNS NAME: CRUSET60P.RALEIGH.IBM.COM
IST1669I IPADDR..PORT 9.27.40.41..3907
```

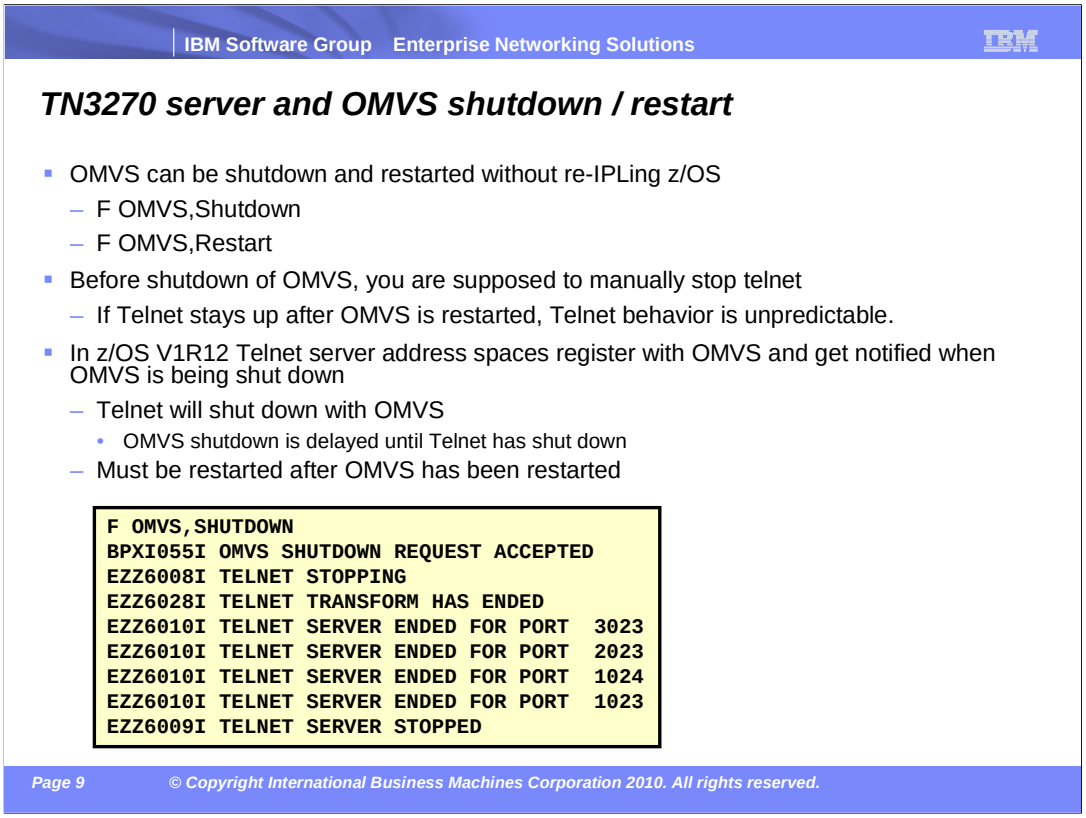
  - The CV64 is also passed to the SNA PLU by its logon exit
- When the SNA PLU is a session manager that relays the SNA session over another LU to the final SNA application PLU
  - The CV64 information is lost on that second session
  - The session manager has no SNA APIs available to propagate the CV64 information
- z/OS V1R12 adds such an API, allowing an enabled session manager to pass the CV64 information to the final SNA application



When you establish a session using telnet, telnet creates a secondary logical unit (SLU) or telnet LU name. Telnet uses the SLU to initiate a session with the target primary logical unit (PLU) by sending a CINIT to the PLU. Telnet builds a control vector, CV64, with IP and other information and appends it to the CINIT that is delivered to the PLU in its LOGON exit. The CV64 contains IP information that can be displayed by VTAM once the session is established.

If there is a session manager between telnet and the PLU, the session manager can have two sessions for the session from the telnet SLU to the target application PLU. One of the sessions is from telnet to the session manager and the other session is from the session manager to the target application. The session manager receives the CINIT with the CV64 in its LOGON exit, but it has no way to pass the CV64 on to the target application PLU. So, the PLU does not receive the CV64 that carries the IP information for the telnet client. Therefore any functionality that the PLU provided, because of the IP information, is lost. The session manager in this picture is IBM Session Manager (ISM). A session manager that does CLSDST PASS to establish the sessions does not have this problem.

VTAM has added API support that allows an application to provide a CV64 to associate with a session. During SETLOGON processing, the application provides the CV64 to VTAM. VTAM checks the CV64 to make sure it is built correctly. Then, VTAM stores the CV64 so that later it can append it to a CINIT for a session with the application. The CINIT being delivered to the target application PLU has a CV64 with it.



Because telnet uses UNIX® System Services, when OMVS is shut down, telnet can no longer access these services and can potentially abnormally terminate. When OMVS is restarted, the state of telnet is questionable. Operators should stop telnet before shutting down OMVS, but they don't always do so.

Telnet now registers with OMVS telling OMVS it must delay shutdown processing until telnet is stopped. The registration also includes a telnet exit routine that OMVS calls when the OMVS shutdown command is issued.

When OMVS shutdown is issued, OMVS calls the telnet routine that will shut down telnet. OMVS waits for telnet to stop. The new telnet exit drives the same process as if an operator issued a stop telnet console command. Once telnet is stopped, OMVS can continue its shutdown.

Various TN3270 server enhancements

- A new option is passed in the CV64 control vector to an SNA primary LU on the CINIT flow
  - The option informs the SNA application if the TN3270 connection is a secure connection or not
  - Can be used by the SNA application to determine requirements for additional security
- To prevent a change of TN3270 connection attributes during a takeover process, a new configuration option is added to the takeover definitions:
  - TKOGENLURECON and TKOSPECLURECON – SAMECONNTYPE
- TN3270 server messages will now indicate the name of the TN3270 server address space instead of just saying 'TELNET'

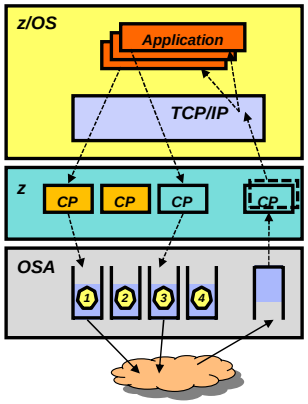
Several customers have asked for a real-time indicator that can be used to determine the security of the TCP/IP connection while the session is being established. Telnet provides the security level of the TCP/IP connection using SMF records and appldata on the TCP connection termination SMF record. However, this information is available too late to be useful as a real-time security check. A flag that indicates whether the TCP/IP connection is secure has been added to the CV64.

If a telnet connection is taken over with the reconnect option by another connection, the type of connection can be identical or at a higher level of security. Today it is possible for a secure telnet connection to take over a basic connection. It is also possible for a connection which uses client authentication to take over a connection which does not use it. However, the application will continue to have security information relating to the original connection and not the new connection. To prevent this, a new option, SAMECONNTYPE, can be specified so that the taker connection must be the same CONNTYPE as the target connection.

When telnet was first introduced, it ran as part of the TCP/IP stack. It was important to identify messages as telnet messages, but not the job name as that was the name of the TCP/IP stack. Since z/OS V1R9, telnet must run in its own address space. If you use multiple telnet servers, you need a way to determine which telnet server issued the message. The word TELNET in messages is replaced with the job name. LUNS/LUNR messages had all begun with TELNET and contained the telnet server name. These messages have also been modified. TELNET has been removed and the messages now begin with the telnet server name.

Pre V1R12 OSA inbound processing overview

- QDIO uses multiple write queues for traffic separation
- QDIO uses only one read queue
  - Multiple CPs are used only when data is accumulating on the queue
  - Single process for initial interrupt and read buffer packaging
    - TCP/IP stack performs inbound data separation for sysplex distributor traffic, bulk data, EE traffic, and so on
  - z/OS Communications Server is becoming the bottleneck as OSA nears 10GbE line speed
    - Inject latency
    - Increase processor utilization
    - Impede scalability



Outbound traffic separation (that is, assignment to a specific priority queue) on the multiple write queues can be accomplished by using the policy agent. You configure a policy with the SetSubnetPrioTosMask statement. Beginning in z/OS V1R11 Communications Server, outbound traffic separation can also be accomplished by using the WLMPRIORITYQ parameter on the GLOBALCONFIG statement. Each priority queue is processed independently of the others. For example, one processor can be building writes on priority queue one while another processor is building writes on priority queue four.

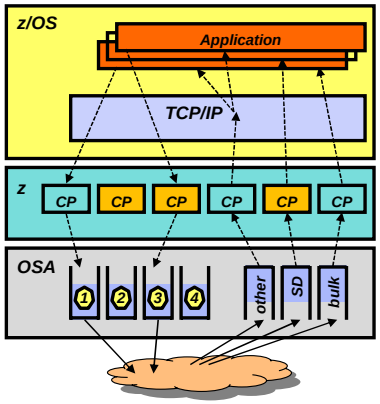
All inbound traffic is received on the single read queue. This includes both batch and interactive traffic, and both traffic destined for this TCP/IP stack and traffic to be forwarded by this TCP/IP stack. Multiple processes only run for inbound traffic when data is accumulating on the read queue. This typically happens during burst periods when z/OS Communications Server is not keeping up with the OSA. A single process is used to package the data, queue it, and schedule the TCP/IP stack to process it. This same process also performs acceleration functions, such as sysplex distributor connection routing acceleration. The TCP/IP stack separates the traffic types to be forwarded to the appropriate stack component that will process them.

For these reasons, z/OS Communications Server is becoming a bottleneck as OSA-Express3 10GbE nears line speed. z/OS Communications Server is injecting latency and increasing processor utilization.

An example of a performance problem observed for bulk inbound traffic is that multiple processes run when data accumulates on the read queue. Then the inbound data for a single TCP connection can arrive at the TCP layer out of order. Then TCP transmits a duplicate ACK every time it sees out of order data. Then the sending side enters fast retransmit recovery.

OSA multiple inbound queue support

- Allow inbound QDIO traffic separation by supporting multiple read queues
  - "Register" with OSA which traffic goes to which queue
  - OSA-Express Data Router function routes to the correct queue
- Each input queue can be serviced by a separate process
  - Primary input queue for general traffic
  - One or more ancillary input queues (AIQs) for specific traffic types
- Supported traffic types are bulk data and sysplex distributor
  - All other traffic not backed up behind bulk data or SD traffic
- Dynamic LAN idle timer updated per queue



z/OS V1R12 Communications Server separates inbound traffic using multiple read queues. TCP/IP will register with OSA which traffic is received on each read queue. The OSA-Express Data Router function routes traffic to the correct queue. Each read queue can be serviced by a separate process. The primary input queue is used for general traffic.

One or more ancillary input queues (AIQs) are used for specific traffic types.

The supported traffic types are streaming bulk data and sysplex distributor.

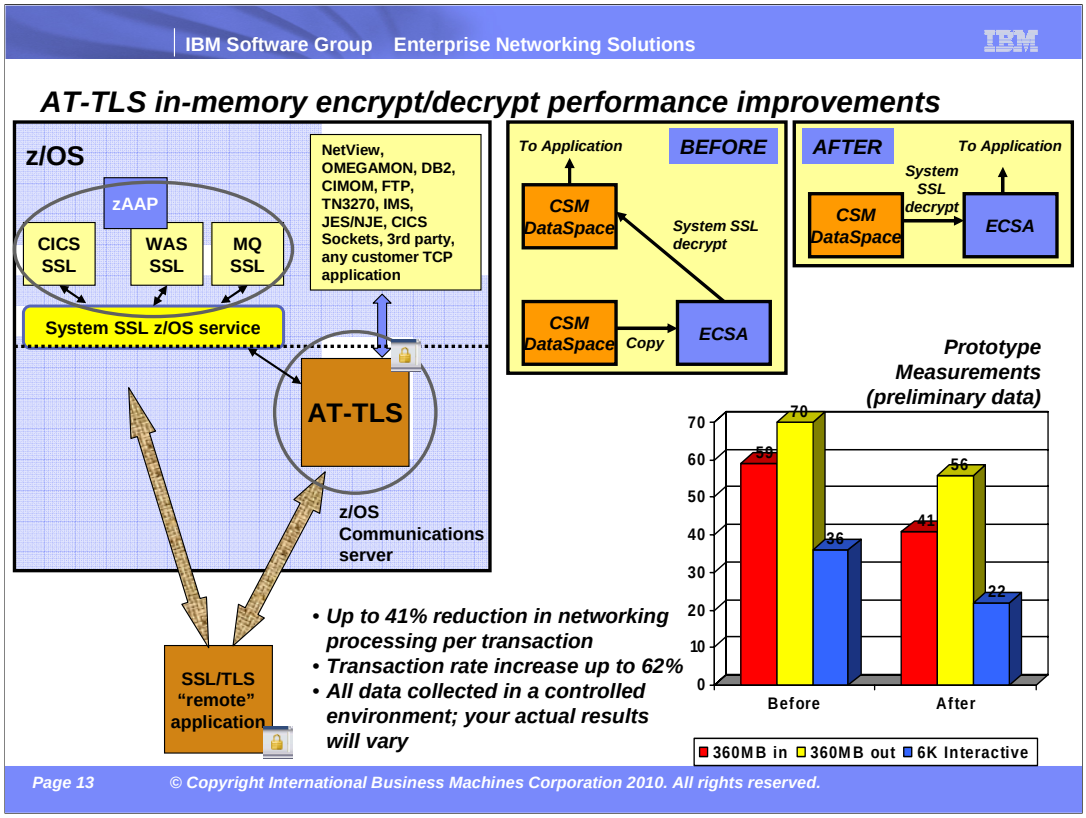
Examples of bulk data traffic are FTP, TSM, NFS, and TDMF®.

Both IP versions are supported for all types of traffic.

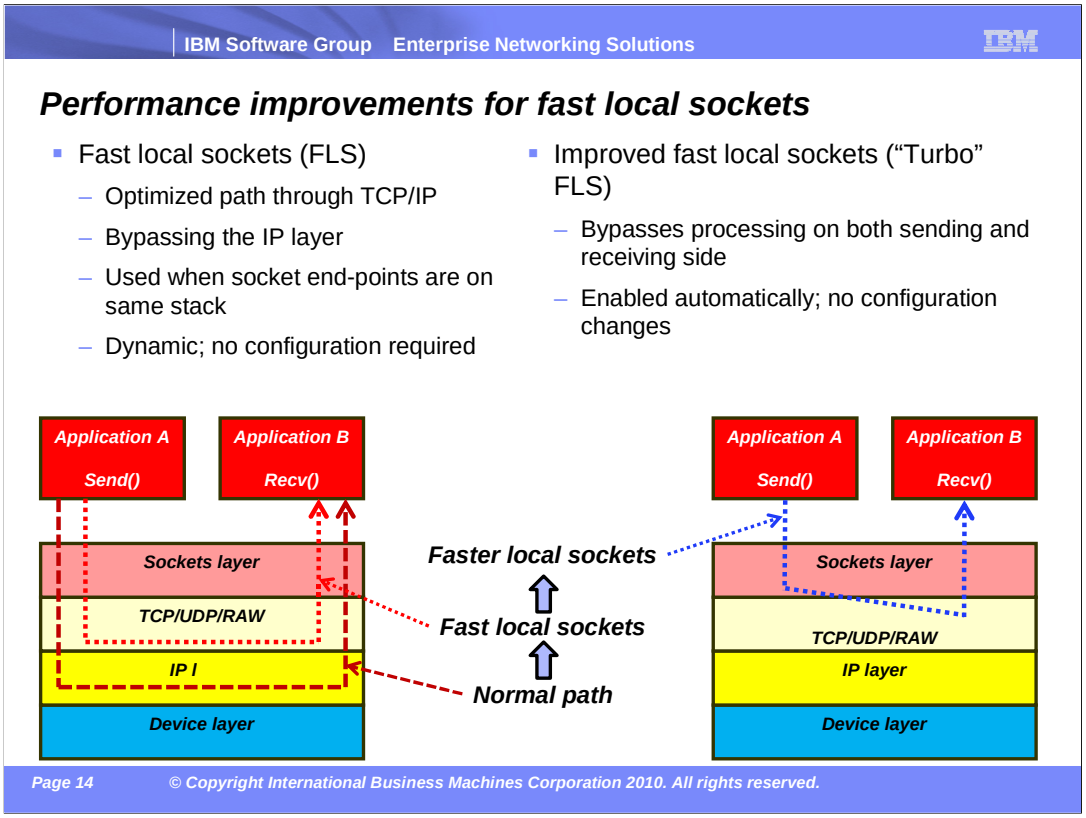
With bulk data traffic separated onto its own read queue, TCP/IP services the bulk data queue from a single processor. This solves the out of order delivery issue – there are no more race conditions. With sysplex distributor traffic separated onto its own read queue, it can be efficiently accelerated or presented to the target application. All other traffic is processed simultaneously with the bulk data and sysplex distributor traffic.

The dynamic LAN idle timer is updated independently for each read queue. This ensures the most efficient processing of inbound traffic based on the traffic type.

TCP/IP defines and assigns traffic to queues dynamically based on local IP address and port. For bulk traffic, the application sets a send or receive buffer to at least 180K. TCP/IP registers each connection (five-tuple). Sysplex distributor traffic data queuing is based on active VIPADISTRIBUTE definitions and TCP/IP registering each DVIPA address.



In z/OS V1R12, AT-TLS performance enhancements have been achieved in encrypting and decrypting data. Early performance measurements show significant processing reduction for request/response and streaming workloads. For short-lived connections (CRR), the cost of AT-TLS connection establishment mitigates the encrypting and decrypting savings.



The performance of fast local sockets has improved by reducing some of the TCP layer processing.

Before z/OS V1R12, fast local sockets bypassed the IP layer. Data was placed on the TCP send queue, then moved to the TCP receive queue. The ACKs were built and sent from the receive side.

In z/OS V1R12, data that is being sent is no longer placed on the sender’s queue and then moved to the receive queue.

The data is now moved directly from the sender onto the receiver’s queue. This bypasses much of the TCP inbound processing.

Acknowledgements of the data are no longer built and sent by the receiver.

This function is enabled automatically and requires no configuration changes. TCP/IP automatically reverts to fast local sockets if packet trace in general or AT-TLS for this specific connection is enabled. There is no impact for data trace.

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 email feedback:

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

This module is also available in PDF format at: ../wnperf.pdf

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

Trademarks, copyrights, and disclaimers

IBM, the IBM logo, ibm.com, IBM, TDMF, VTAM, and z/OS are trademarks or registered trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of other IBM trademarks is available on the web at "Copyright and trademark information" at http://www.ibm.com/legal/copytrade.shtml

THE INFORMATION CONTAINED IN THIS PRESENTATION IS PROVIDED FOR INFORMATIONAL PURPOSES ONLY. WHILE EFFORTS WERE MADE TO VERIFY THE COMPLETENESS AND ACCURACY OF THE INFORMATION CONTAINED IN THIS PRESENTATION, IT IS PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED. IN ADDITION, THIS INFORMATION IS BASED ON IBM'S CURRENT PRODUCT PLANS AND STRATEGY, WHICH ARE SUBJECT TO CHANGE BY IBM WITHOUT NOTICE. IBM SHALL NOT BE RESPONSIBLE FOR ANY DAMAGES ARISING OUT OF THE USE OF, OR OTHERWISE RELATED TO, THIS PRESENTATION OR ANY OTHER DOCUMENTATION. NOTHING CONTAINED IN THIS PRESENTATION IS INTENDED TO, NOR SHALL HAVE THE EFFECT OF, CREATING ANY WARRANTIES OR REPRESENTATIONS FROM IBM (OR ITS SUPPLIERS OR LICENSORS), OR ALTERING THE TERMS AND CONDITIONS OF ANY AGREEMENT OR LICENSE GOVERNING THE USE OF IBM PRODUCTS OR SOFTWARE.

