IBM Software Group

z10: z/OS V1R10 Communications Server Large Send Performance Summary

Dan Patel
csperf@us.ibm.com

September 2009

Trademarks and Notices

The following are trademarks of the International Business Machines Corporation in the United States and/or other countries.

BatchPipes*
CICS*
DB2*
DB2 Connect
DB2 Universal Database
Domino*
e-business logo
ESCON*
FICON
HiperSockets
IBM*
IBM logo*
Lotus*
Multiprise
NetView*
Notes*
OS/390*
Parallel Sysplex*
PR/SM
QMF
RMF
S/390*
S/390 Parallel Enterprise Server
Tivoli
VTAM
xSeries
WebSphere
z/Architecture
z/OS
z/VM
zSeries

* Registered trademarks of IBM Corporation

The following are trademarks or registered trademarks of other companies.

Cisco is a registered trademark of Cisco Systems, Inc.
UNIX is a registered trademark of The Open Group in the United States and other countries.
Microsoft, Windows and Windows NT are registered trademarks of Microsoft Corporation.

* All other products may be trademarks or registered trademarks of their respective companies.

Notes:

Performance is in Internal Throughput Rate (ITR) ratio based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual performance that any user will experience will vary significantly 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, and there is no guarantee, that an individual user will achieve performance or throughput improvements equivalent to the performance ratios stated here. Users of this document should verify the applicable data for their specific environment.

IBM hardware products are manufactured from new parts, or new and serviceable used parts. Regardless, our warranty terms apply.

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.

Performance Disclaimer

The performance measurements discussed in this presentation were collected using a dedicated system environment. The results obtained in other configurations or operating system environments may vary significantly depending upon environments used. Therefore, no assurance can be given, and there is no guarantee to achieve performance or throughput improvements equivalent. Users of this document should verify the applicable data for specific environment.

Performance Presentation

- Measurement Techniques
- z/OS V1R10 Large Send
- Hardware/Software Configurations
- z10: z/OS V1R10 CS Large Send Measurements
  - Large Send vs. No Large Send (RR,CRR, Stream workloads) OSA Exp3 1 Gb
  - Large Send vs. No Large Send (FTP workload) OSA Exp3 1Gb
  - Large Send vs. No Large Send (RR,CRR, Stream workloads) OSA Exp2 1 Gb
  - Large Send vs. No Large Send (FTP workload) OSA Exp2 1Gb
  - Large Send vs. No Large Send (RR,CRR, Stream workloads) OSA Exp3 10 Gb
- Summary

Measurement Techniques

Performance benchmarks in this presentation were obtained using the IBM Application Workload Modeler (AWM) for z/OS (V1R1)

- ***"IBM Application Workload Modeler for z/OS Release 1 provides the ability to model, measure, and analyze the performance of networks and applications in a client/server, multiprotocol, multiplatform environment. With Application Workload Modeler R1, you can more accurately plan for the roll-out of additional software or function, and determine where upgrades may be required in your network and systems."***

- **For more information, visit the Application Workload Modeler web site:**
  http://www.ibm.com/software/network/awm/index.html

For comparison of measurements, transaction rate and CPU cost differences within +/- 3% was considered statistically insignificant.

z/OS Comm Server Large Send

Large Send:

TCP Segmentation Offload (also called Large Send) transfers the overhead of segmenting outbound data into individual TCP packets to the QDIO (Queued Direct I/O) attached OSA-Express3 or OSA Express2 device(s). Offloading segmentation of streaming type workloads reduces CPU utilization

z/OS V1R10 Large Send Feature

- **Z/OS V1R10 Large Send / Segmentation Offload feature provides :**
  - **OSA Hardware**
    - Support for OSA-Express2 1000BaseT (#3366)
    - Support for OSA-Express2 Gigabit SX (#3365)
    - Support for OSA-Express2 Gigabit LX (#3364)
    - Support for OSA-Express2 10Gigabit LR(#3368)
    - Support for OSA-Express3 10Gigabit LR(#3370)
    - Support for OSA-Express3 10 Gigabit SR(#3371)
    - Support for OSA-Express3 Gigabit LX (#3362)
    - Support for OSA-Express3 Gigabit SX 4 ports/card (#3363)
    - Support for OSA-Express3 Gigabit SX 2 ports/card (#3373)
    - Support for OSA-Express3 1000BaseT 4 ports/card (#3367)
    - Support for OSA-Express3 1000BaseT 2 ports/card (#3369)

  Note: 2 port/card OSA Express3 features are only available on the z10 BC (2098) machines, except for FC 3371 which has a maximum of 2 ports and is available on both 2097 and 2098.

  - **Configured and managed exactly like Gigabit ethernet**
  - **Offload most IPv4 TCP segmentation for Outbound processing to OSA-Express2 or OSA-Express3 in QDIO mode**
  - **Decreases host CPU utilization for outbound bulk data**
  - **Increases data transfer efficiency for IPv4 packets**

z/OS V1R10 Large Send Feature ....

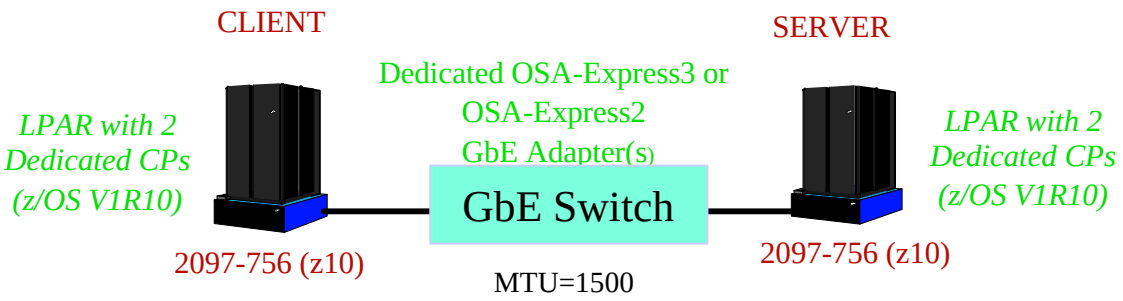
- **Z/OS V1R10 Large Send / Segmentation Offload feature provides :**
  - **Restrictions:**
    - Requires z10, z9, z990 or z890
    - IPv4 only
    - TCP transport only
    - Outbound packets only
    - Packets written to the LAN only (not to another stack sharing the OSA)
    - No IPSec packets
  - **Support enabled by adding the**
    - "GLOBALCONFIG SEGMENTATIONOFFLOAD " keyword in TCPIP profile
    - Checksum is also offloaded when segmentation is offloaded

z/OS V1R10/V1R9 Large Send PTFs

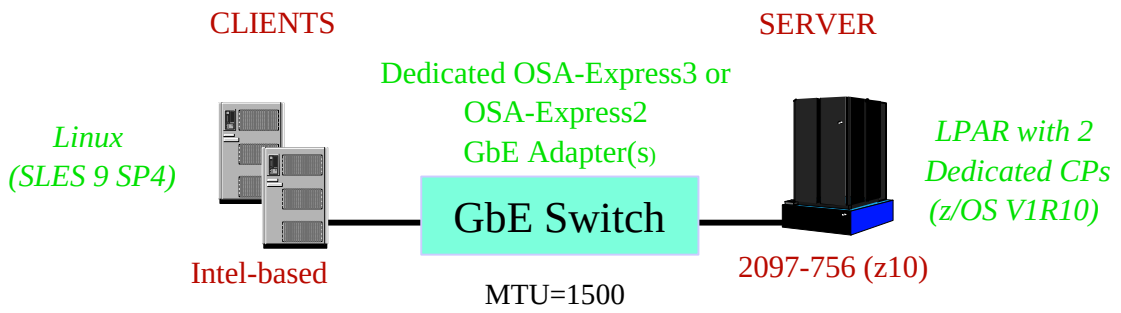
- **Minimum Communications Server Maintenance**
  - TCP/IP R1A0
    - APAR PK64756 - PTF UK37433
  - TCP/IP R190
    - APAR PK47376 - PTF UK26977
    - APAR PK56723 - PTF UK32713
    - APAR PK64756 - PTF UK37435
  - Detail for Other Releases and Minimum OSA Microcode Levels use the URL
    - http://www.ibm.com/support/docview.wss?rs=852&uid=swg21232599
- **Specific fixes Required for Large Send**
  - **Information for Fixes available in TDR H194022 or WSC FLASH10458**
    - z10 2097 and 2098,
      - Driver 76
      - OSA Express2 EC N10953 MCL001 (LIC Level 3.03)
    - z10 2097
      - Driver 73
      - OSA Express2 MCL not available, upgrade to D76
    - z9 2094 and 2096
      - OSA Express2 EC G40946 MCL007 (LIC Level 8.81)
    - z990 (2084 and 2086)
      - OSA Express2 EC J13476 MCL023 (LIC Level 0.2C)

Hardware/Software Configuration

➢ **AWM Client/Server Benchmarks (RR,CRR, STR)**



➢ **FTP Server**

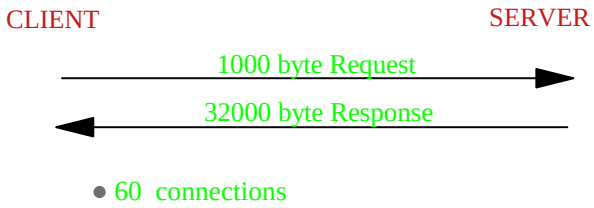


- All measurements were collected on a z10 (2097-756 2 CPs Client/Server LPARs) and z/OS V1R10
- For FTP measurements Linux machine was used as Client and z10 as LPAR as Server
- MTU size of 1500 bytes was used for all measurements

AWM Benchmark Description

RR Workload

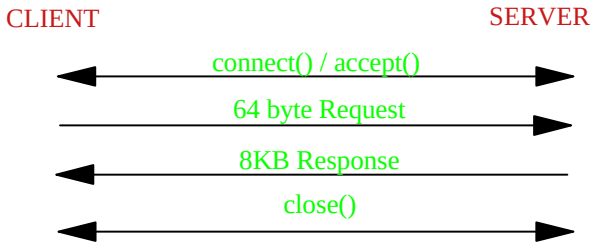
- Request-Response
  - Simulate TN3270
  - Interactive workloads
  - Bulk data used for Outbound



- 60 connections

CRR Workload

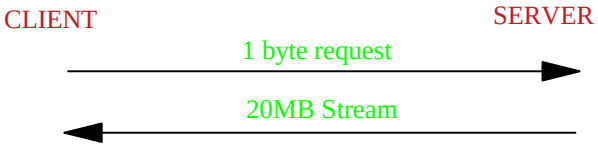
- Connect-Request-Response
  - Static Web Serving



- 9 connections

STR Workload

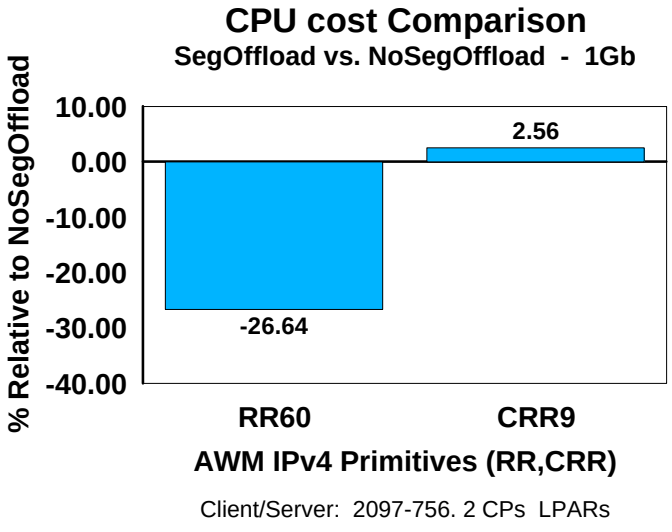
- Streaming
  - Simulate FTP
  - Memory-to-Memory
  - Workload used (1 byte IN/20,000,000 bytes OUT)



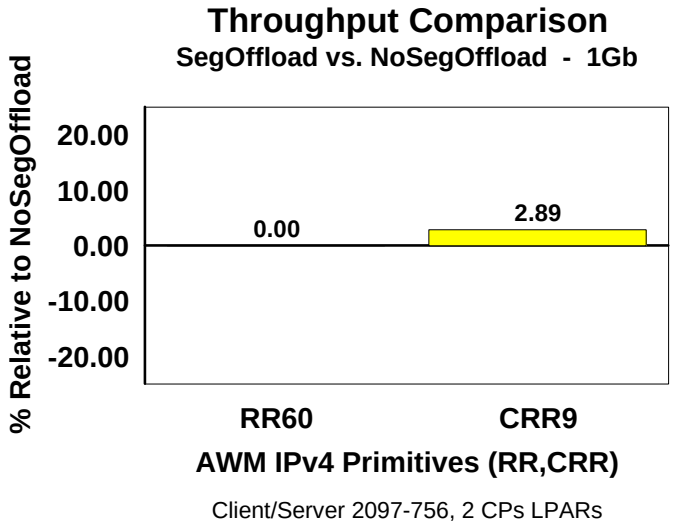
- 10 connections

V1R10 LSOffload OSA-Exp3 1 Gb Performance

- V1R10 SegOffload vs. NoSegOffload (RR, CRR) workload OSA-Exp3 1Gb
  - For RR60(1000/32000) workload CPU cost is 26.64% lower than NoSegOffload and Throughput for SegOffload and NoSegOffload is equivalent.
  - For CRR9(64/8192) workload CPU cost is 2.56% higher than NoSegOffload and Throughput for SegOffload is 2.89% higher compared to NoSegOffload.



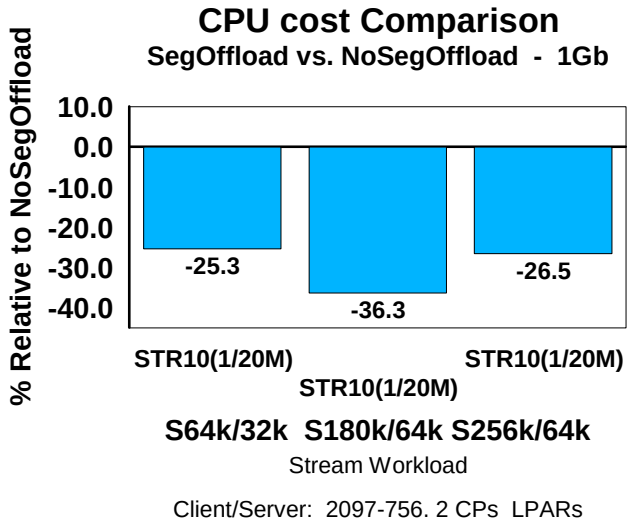
Client/Server: 2097-756. 2 CPs LPARs

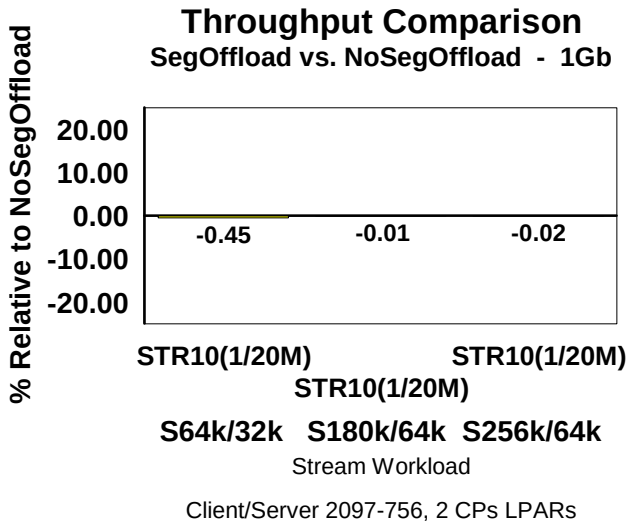


Client/Server 2097-756, 2 CPs LPARs

V1R10 LSOffload OSA-Exp3 1 Gb Performance ...

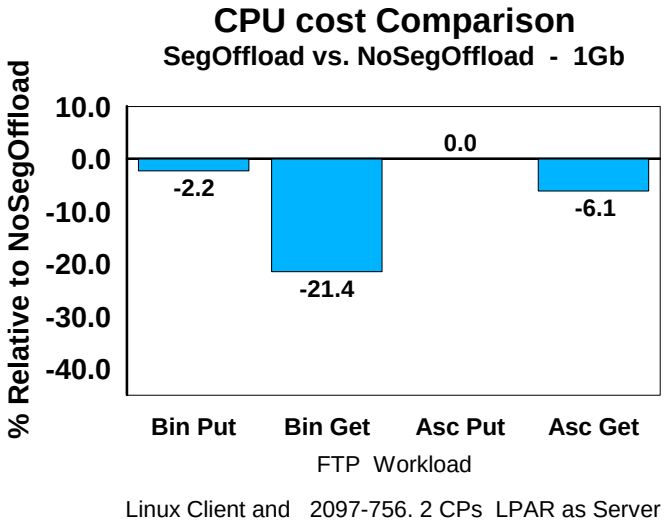
- V1R10 SegOffload vs. NoSegOffload (Stream) workload OSA-Exp3 1Gb
  - For Stream10(1/20M) workload CPU cost is 25.3-36.3)% lower than NoSegOffload and Throughput is within (-0.45 to 2.89)% range when compared to NoSegOffload.
  - S64k/32k, S180k/64k and S256k/64k represents SSOC=RSOC=64k or 180k or 256k and SNDR=RCVB = 32k or 64k.
    - SSOC -TCP/IP Send socket buffer size, RSOC-TCP/IP Receive socket buffer size
    - SNDR- TCP/IP Send socket size on Send() socket call and RCVB- TCP/IP Receive socket size on Receive() socket call



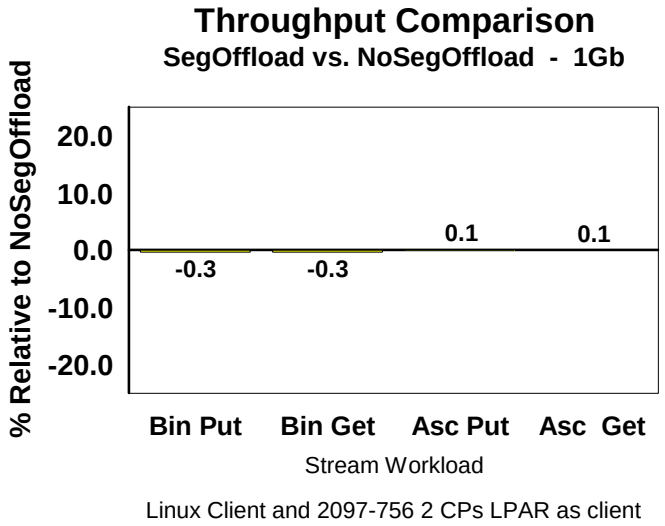


V1R10 LSOffload OSA-Exp3 1 Gb Performance FTP ...

- V1R10 SegOffload vs. NoSegOffload (FTP) workload OSA-Exp3 1Gb
  - For FTP workload CPU cost is 21.4% lower for Binary Get (Outbound) and 6.1% lower for ASCII Get with SegOffload compared to NoSegOffload.
  - SegOffload does not benefit to binary put or ASCII Put (Inbound).
  - Throughput difference between SegOffload and NoSegOffload is insignificant



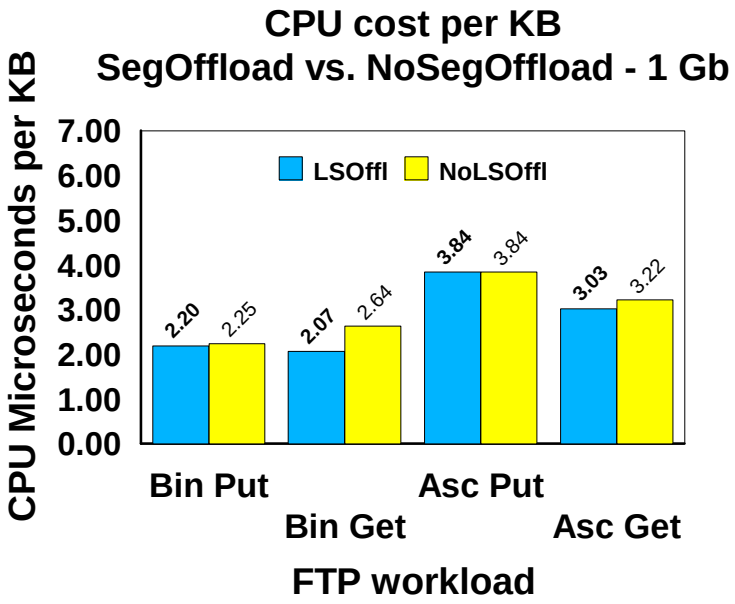
Linux Client and 2097-756. 2 CPs LPAR as Server



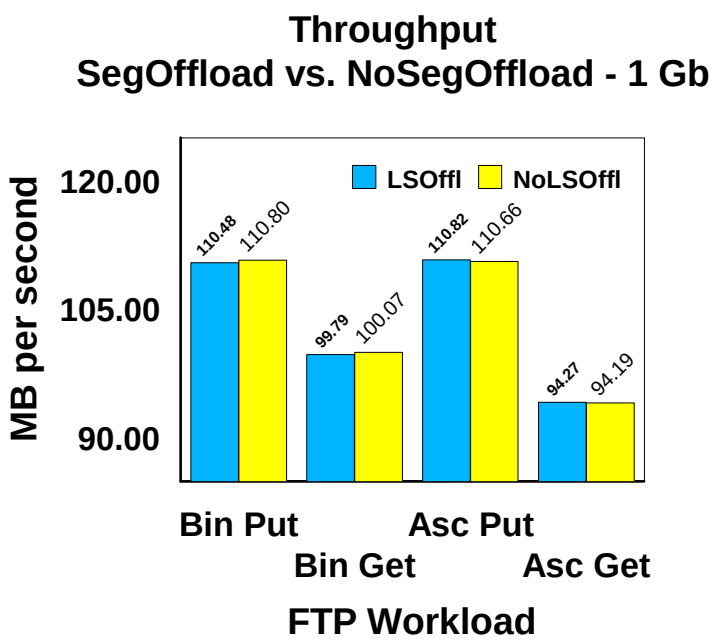
Linux Client and 2097-756 2 CPs LPAR as client

V1R10 LSOffload OSA-Exp3 1 Gb Performance FTP ...

- V1R10 SegOffload vs. NoSegOffload (FTP) workload OSA-Exp3 1Gb
  - FTP workload CPU and Throughput comparisons
  - SegOffload provides lower CPU cost for binary and ASCII GET



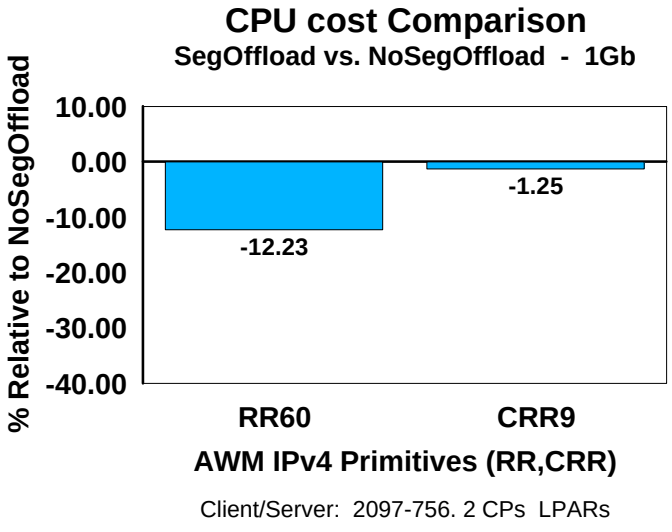
Linux Client and 2097-756 2 CPs LPAR as Server



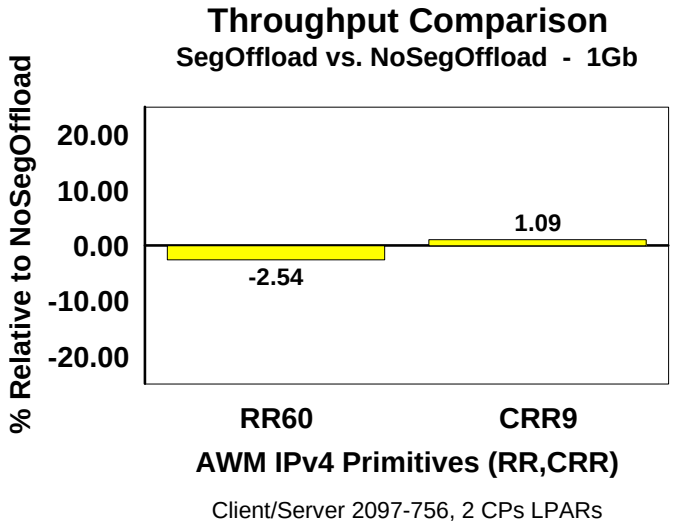
Linux Client and 2097-756 2 CPs LPAR as Server

V1R10 LSOffload OSA-Exp2 1 Gb Performance

- V1R10 SegOffload vs. NoSegOffload (RR, CRR) workload OSA-Exp2 1Gb
  - For RR60(1000/32000) workload CPU cost is 12.23% lower than NoSegOffload and Throughput is 2.54% lower compared to NoSegOffload.
  - For CRR9(64/8192) workload CPU cost is 1.25% lower than NoSegOffload and Throughput is 1.09% higher compared to NoSegOffload.



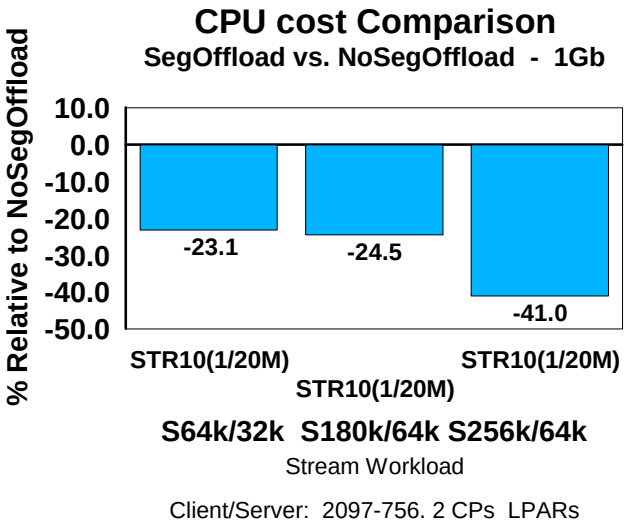
Client/Server: 2097-756. 2 CPs LPARs



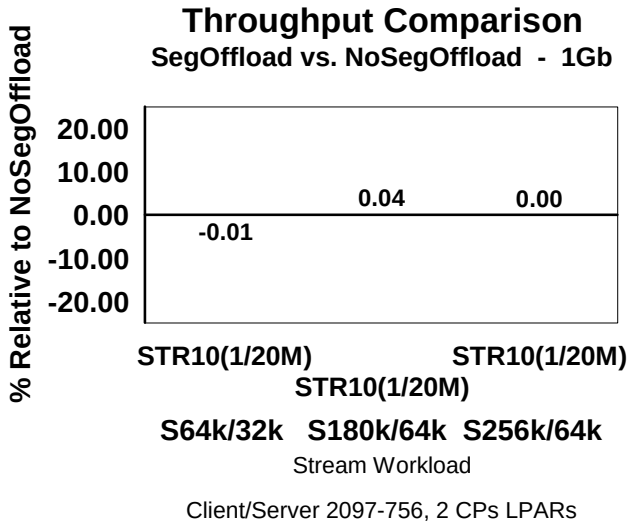
Client/Server 2097-756, 2 CPs LPARs

V1R10 LSOffload OSA-Exp2 1 Gb Performance ...

- V1R10 SegOffload vs. NoSegOffload (Stream) workload OSA-Exp2 1Gb
  - For Stream10(1/20M) workload CPU cost is (23.1-41.0)% lower than NoSegOffload and Throughput for SegOffload and NoSegOffload is equivalent in all cases.
  - S64k/32k, S180k/64k and S256k/64k represents SSOC=RSOC=64k or 180k or 256k and SNDR=RCVB = 32k or 64k.
    - SSOC -TCP/IP Send socket buffer size, RSOC-TCP/IP Receive socket buffer size
    - SNDR- TCP/IP Send socket size on Send() socket call and RCVB- TCP/IP Receive socket size on Receive() socket call



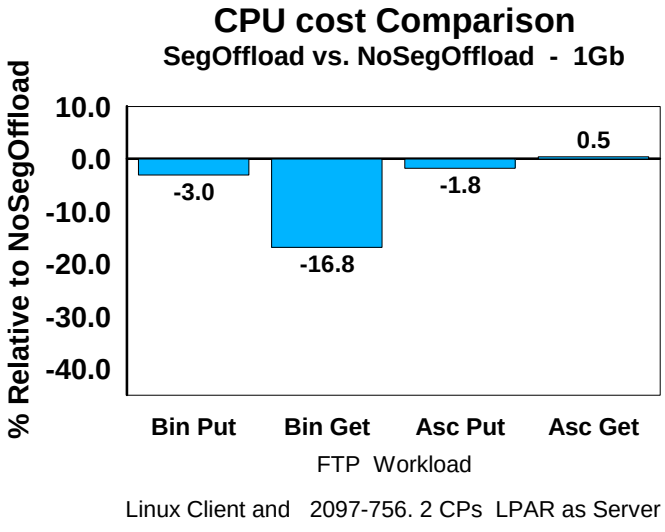
Client/Server: 2097-756. 2 CPs LPARs



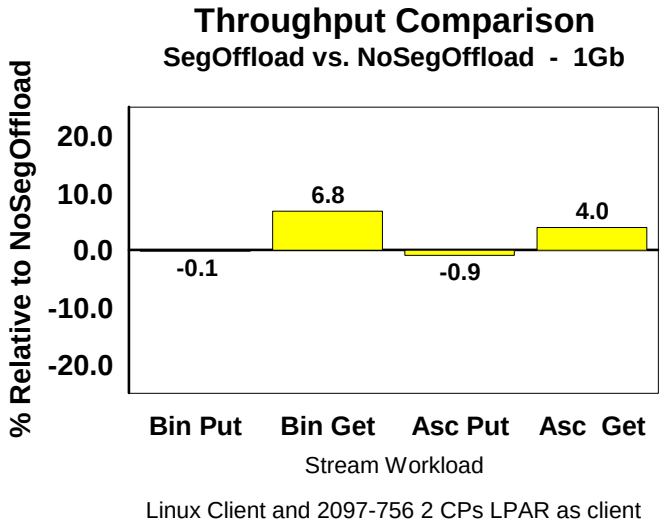
Client/Server 2097-756, 2 CPs LPARs

V1R10 LSOffload OSA-Exp2 1 Gb Performance FTP ...

- V1R10 SegOffload vs. NoSegOffload (FTP) workload OSA-Exp2 1Gb
  - For FTP workload CPU cost is 16.8% lower for Binary Get (Outbound) and 0.5% higher for for ASCII Get with SegOffload compared to NoSegOffload.
  - SegOffload does not benefit to binary put or ASCII Put (Inbound).
  - Throughput for binary Get and ASCII Get is (4-6.8)% higher compared to NoSegOffload.



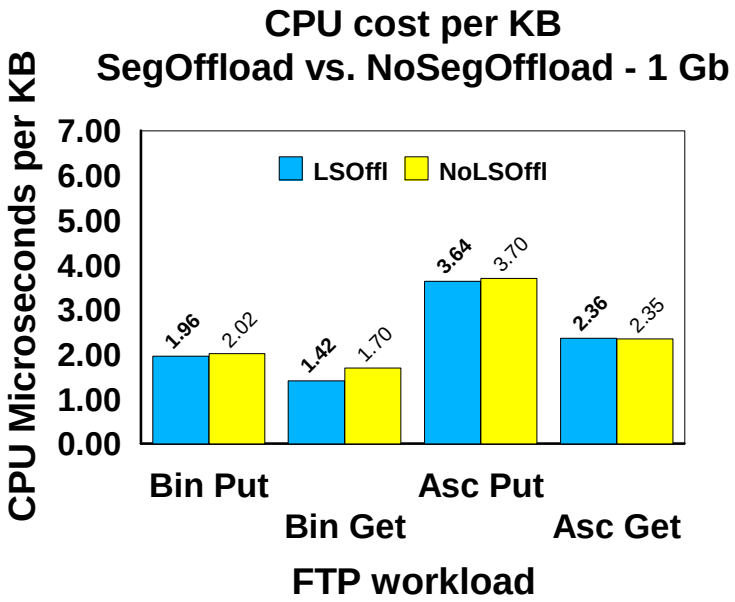
Linux Client and 2097-756. 2 CPs LPAR as Server



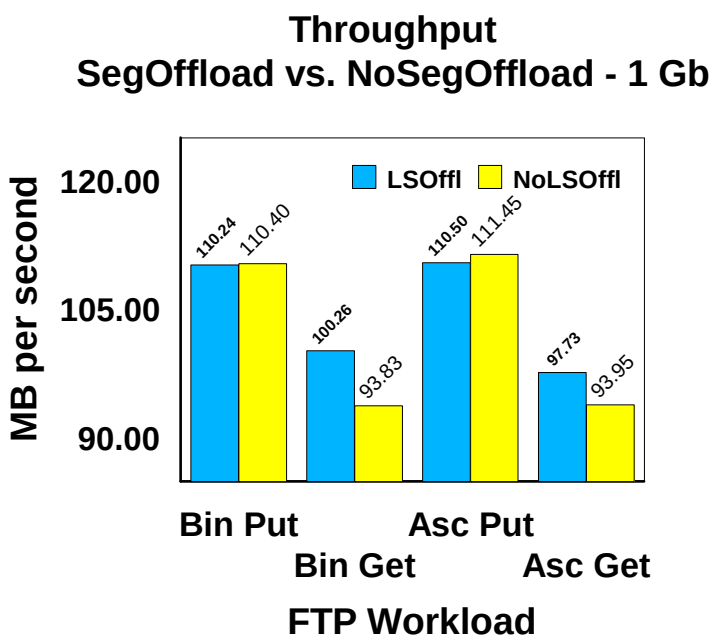
Linux Client and 2097-756 2 CPs LPAR as client

V1R10 LSOffload OSA-Exp2 1 Gb Performance FTP ...

- V1R10 SegOffload vs. NoSegOffload (FTP) workload OSA-Exp2 1Gb
  - FTP workload CPU and Throughput comparisons
  - SegOffload provides lower CPU cost for binary and ASCII GET



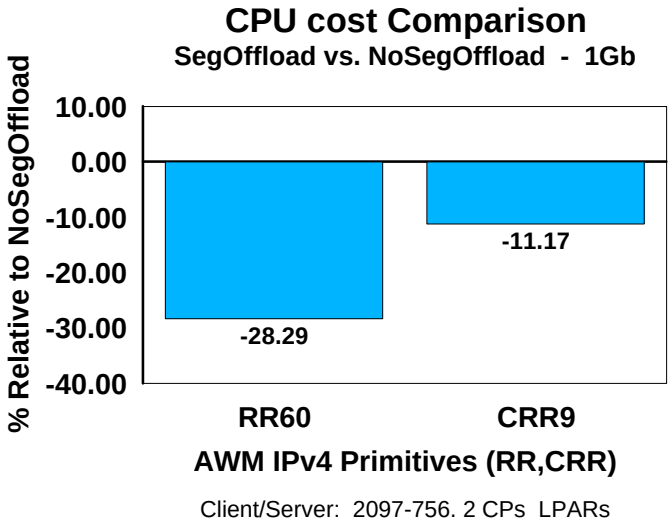
Linux Client and 2097-756 2 CPs LPAR as Server



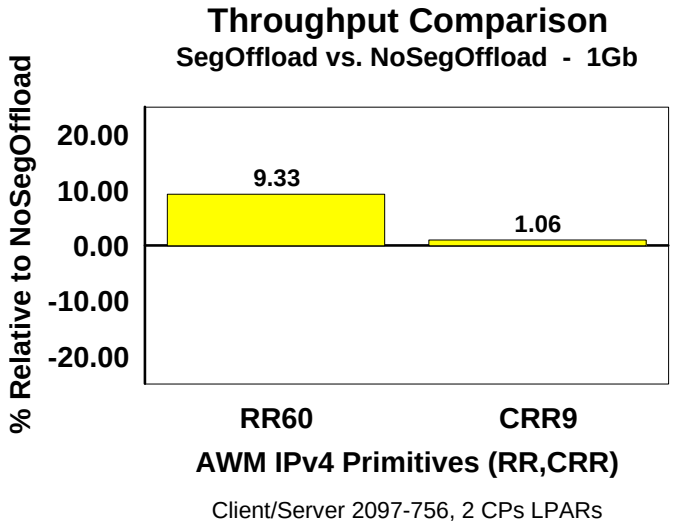
Linux Client and 2097-756 2 CPs LPAR as Server

V1R10 LSOffload OSA-Exp3 10 Gb Performance

- V1R10 SegOffload vs. NoSegOffload (RR,CRR) workload OSA-Exp3 10 Gb
  - For RR60(1000/32000) workload CPU cost is 28.29% lower than NoSegOffload and Throughput for SegOffload is 9.33% higher when compared to NoSegOffload..
  - For CRR9(64/8192) workload CPU cost is 11.17% lower than NoSegOffload and Throughput for SegOffload is 1.06% higher compared to NoSegOffload.



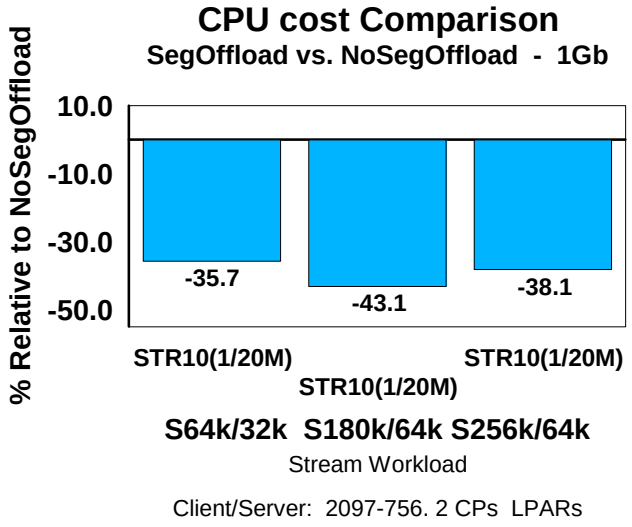
Client/Server: 2097-756. 2 CPs LPARs



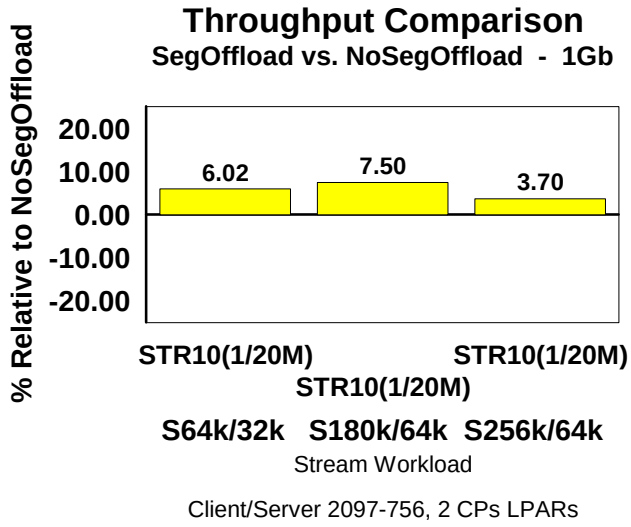
Client/Server 2097-756, 2 CPs LPARs

V1R10 LSOffload OSA-Exp3 10 Gb Performance ...

- V1R10 SegOffload vs. NoSegOffload (Stream) workload OSA-Exp3 10 Gb
  - For Stream10(1/20M) workload CPU cost is (35.7-43.1)% lower than NoSegOffload and Throughput is (3.70-7.50)% higher when compared to NoSegOffload.
  - S64k/32k, S180k/64k and S256k/64k represents SSOC=RSOC=64k or 180k or 256k and SNDR=RCVB = 32k or 64k.
    - SSOC -TCP/IP Send socket buffer size, RSOC-TCP/IP Receive socket buffer size
    - SNDR- TCP/IP Send socket size on Send() socket call and RCVB- TCP/IP Receive socket size on Receive() socket call



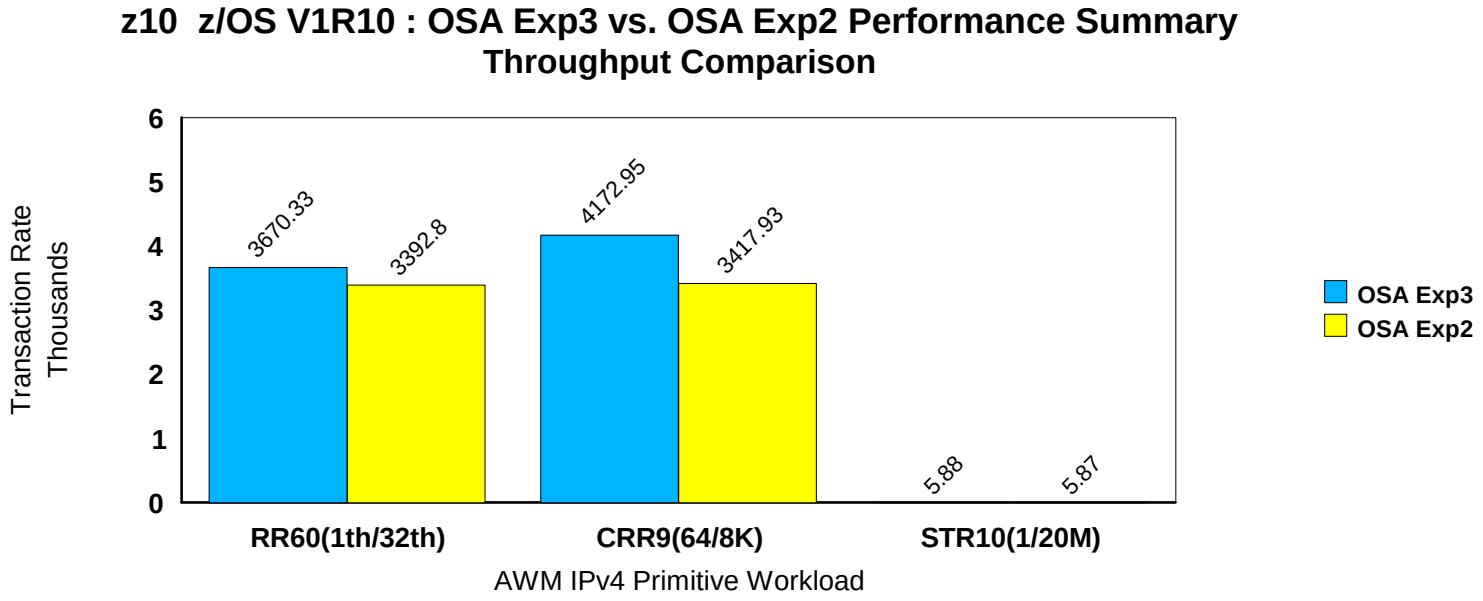
Client/Server: 2097-756. 2 CPs LPARs



Client/Server 2097-756, 2 CPs LPARs

V1R10 Throughput Comparison (OSA Exp3 and OSA Exp2 1Gb)

- **V1R10 Throughput Comparison for OSA Exp3 and OSA Exp2 1Gb)**
  - Throughput data used for comparison are for NoSegmentation Offload (Throughput difference is very minimum between Segmentation Offload and NoSegmentation Offload)
  - Workload:
    - For Client, Server used dedicated z10 LPARs each with 2 CPs
    - RR60(1th/32th) : 60 sessions, interactive Request Response workload with 1000 byte request, 32,000 byte reply
    - CRR9(64/8k): 9 sessions, Connect Request, Response workload with 64 byte request and 8192 byte reply
    - STR10(1/20M): 10 sessions, Streaming workload with 1 byte request and 20,000,000 byte reply



z/OS V1R10 Large Send Summary

- z/OS V1R10 Large Send helps in improving CPU cost for the outbound TCP (IPv4) Interactive, Streaming and FTP (binary and ASCII get) workloads

- OSA Exp3 1Gb interface:
  - For interactive workload - RR60(1000/32000), Large Send improves CPU cost per transaction by 26.64% compared to NoSegmentation Offload
  - For Streaming workload- STR(1/20M), Large Send improves CPU cost per transaction by (25.3-36.3)% compared to NoSegmentation Offload
  - For FTP workloads, CPU cost per transaction is improved by 21.4% for binary Get and 6.1% for ASCII Get with the use of Large Send.

- OSA Exp2 1Gb interface:
  - For Interactive workload - RR60(1000/32000), Large Send improves CPU cost per transaction by 12.23% compared to NoSegmentation Offload
  - For Streaming Workload - STR(1/20M), Large Send improves CPU cost per transaction by (23.1-41.0)% compared to NoSegmentation Offload
  - For FTP workloads, CPU cost per transaction is improved by 16.8% for Binary Get with the use of Large Send, No significant benefit to ASCII Get.

- OSA Exp3 10 Gb interface:
  - For interactive workload - RR60(1000/32000), Large Send improves CPU cost per transaction by 28.29% compared to NoSegmentation Offload
  - For Streaming workload- STR(1/20M), Large Send improves CPU cost per transaction by (35.7-43.1)% compared to NoSegmentation Offload

For More Information.

URL	Content
http://www.ibm.com/systems/z	IBM Enterprise Servers
http://www.ibm.com/systems/z/networking	zSeries Networking
http://www.ibm.com/software/network/commserver	IBM Communications Severs
http://www.ibm.com/software/network/commserver/zos	z/OS Communications Server
http://www.ibm.com/software/network/commserver/support	Communications Server Technical Support
http://www.ibm.com/software/network/commserver/library	Product publications, Red Books, White Papers, etc. for all IBM Comm Servers
http://www.ibm.com/software/network/commserver/zos/library	Product publications, Red Books, White Papers, etc. for z/OS CS
http://www.redbooks.ibm.com	ITSO Redbooks
http://www.ibm.com/software/network/commserver/zos/support/	z/OS Communications Server Technical Support
http://www.ibm.com/support/techdocs/	Technical Information Data Base (Flashes, Presentations, White Papers, etc.)
http://www.ibm.com/software/network/awm	IBM Application Workload Modeler (AWM)