IBM z16™ FICON Express32S Performance

May 2023

Brian Murphy
Yamil Rivera

Table of Contents

Notices, Disclaimer and Trademarks

Acknowledgements

The authors would like to thank the following colleagues for their expertise, insight and contributions:

- Shelly Black –IBM Systems, Poughkeepsie, NY
- Nora Borsare –IBM Systems, Poughkeepsie, NY
- Su KR Han – IBM Systems, Shanghai, China
- Andrea Harris –IBM Systems, Poughkeepsie, NY
- Fei Fei Li – IBM Systems, Shanghai, China
- Nicholas VanMater – IBM Systems, Poughkeepsie, NY

The authors would also like to thank all of the reviewers of this paper for their helpful comments.

Introduction

This whitepaper assumes a familiarity with the general concepts of Z Systems. Readers unfamiliar with these topics should consult the IBM Z Mainframe Capabilities website : https://www.ibm.com/it-infrastructure/z/capabilities/networking

In this paper, "FICON Express32S" and "FEx32S" are used interchangeably unless otherwise noted in the context of the text. When zHPF is used by itself, it refers to the zHPF Protocol. When FICON is used by itself, it refers to the FICON Protocol. When FCP is used by itself, it refers to the FCP Protocol.

The IBM z16™ supports the following FICON features:

- FICON Express32S (FEx32S) – new build
- FICON Express16SA (FEx16SA) – carry forward only
- FICON Express16S+ (FEx16S+) - carry forward only

The FICON features conform to the following architecture:

- Fibre Connections (FICON)
- High Performance FICON on Z (zHPF)
- Fibre Channel Protocol (FCP)

The FICON features provide connectivity between any combination of servers, directors, switches, and devices (control units, disks, tapes, and printers) in a SAN. The FEx32S feature is encryption capable in support of IBM Fibre Channel Endpoint Security (FC-ES). Each FICON feature occupies one I/O slot in the PCIe I/O drawer. Each feature has two ports, with one PCHID and one CHPID associated with each port.

Summary

The IBM z16™ FEx32S channel offers many benefits over previous generations of FICON channels, including support for a faster link data rate of 32Gbps.

Enabling IBM Fibre Channel Endpoint Security encrypts your data in flight, processing encryption directly on the FEx32S channel with less than 4% impact on maximum zHPF and FCP throughput.

The increased capability of the FEx32S channels are complemented by improved performance at many levels of the IBM z16™ I/O Subsystem.

The IBM z16™ PCIe fanout card, with two PCIe Interconnects or up to sixteen FEx32S channel cards, supports up to 38 GB/sec Read+Write.

Additional zHPF, FCP, and FICON product information is available on the IBM Z websites:

https://www.redbooks.ibm.com/redbooks/pdfs/sg245444.pdf

https://www.redbooks.ibm.com/redbooks/pdfs/sg248951.pdf

The High Performance FICON for IBM Z

High Performance FICON for IBM Z (zHPF) is implemented for improved throughput and latency, which it achieves by reducing the number of information units (IU) that are processed. Enhancements to the z/Architecture and the FICON protocol provide optimizations for online transaction processing (OLTP) workloads. Reflected in the bar charts in Figure 2 and Figure 3 below are the "best can do" capabilities of each of the generations of FICON Express channels supported on IBM z16™. For each of the generations of FICON Express channels, there are two bars, one which displays the maximum capability using the FICON protocol exclusively and another that shows the maximum capability using the zHPF protocol exclusively.

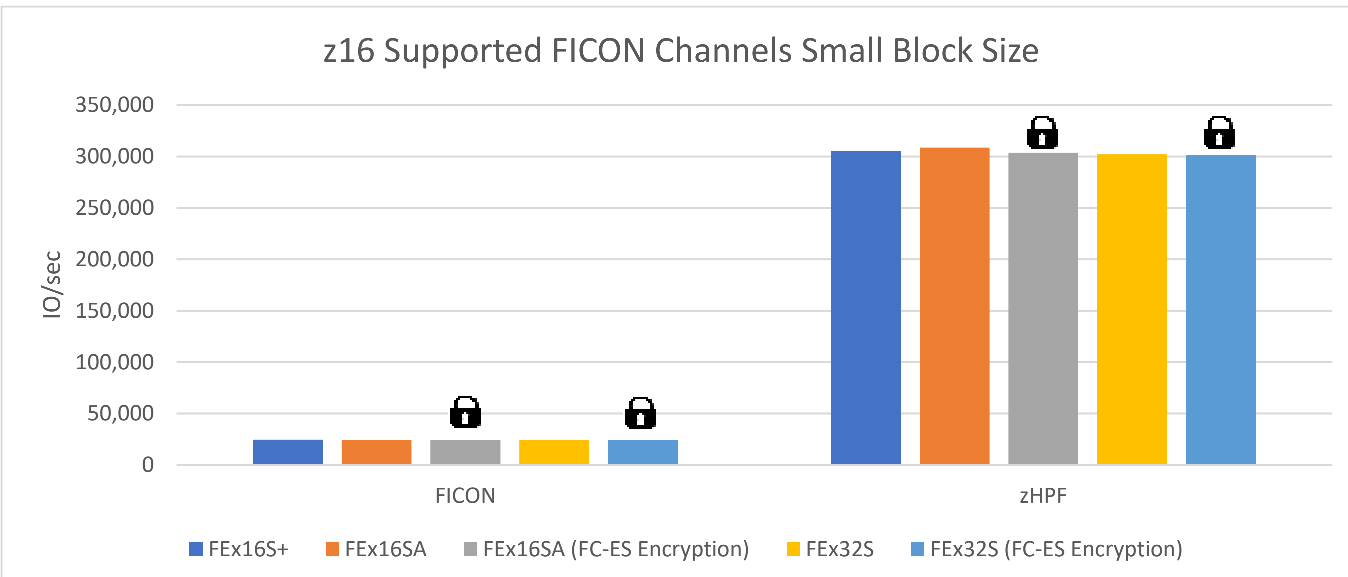


Figure 1 – FICON Express Channel Maximum IO/sec

The chart above displays the maximum IO/sec for each channel as measured at a point in time around the general availability (GA) dates of each product using an I/O driver benchmark program for 4K byte read hits. The size of most online database transaction processing (OLTP) workload I/O operations is 4K bytes. Laboratory measurements using FEx32S in an IBM z16™ with the zHPF protocol and small data transfer I/O operations achieved a maximum of 302K IO/sec.

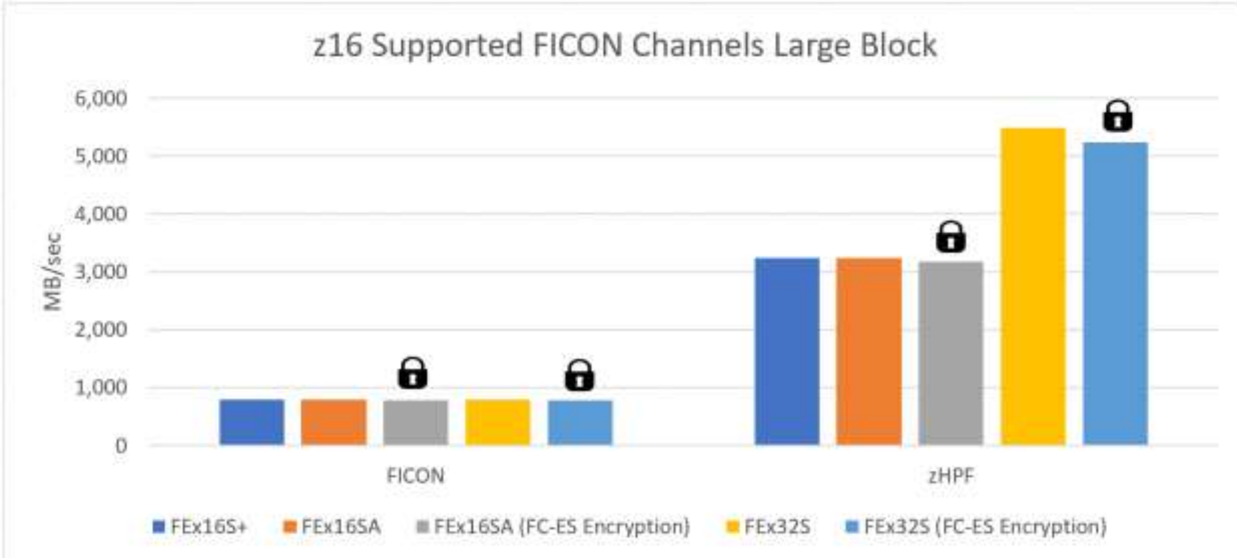


Figure 2 – FICON Express Channel Maximum MB/sec

The chart above displays the maximum READ+WRITE MB/sec for each channel. Laboratory measurements using FEx32S in an IBM z16™ with the zHPF protocol and a mix of large sequential read and write data transfer I/O operations achieved a maximum throughput of 5,400 READ+WRITE MB/sec.

The maximum zHPF 4K IO/sec measured on a FEx32S channel was approximately 12x greater than the maximum FICON protocol's capability. Response time curves for these results will be described later in this document. As shown in Figure 1 & Figure 2, with zHPF it is possible to achieve an improvement in both small block IO/sec processing for OLTP workloads and large sequential READ+WRITE I/O processing compared to previous FICON offerings.

High Performance FICON for System z (zHPF) Architecture Performance Features

zHPF is an extension to the FICON architecture designed to improve the execution of small block I/O requests. zHPF streamlines the FICON architecture and reduces the overhead on the channel processors, control unit ports, switch ports, and links by improving the way channel programs are written and processed. To understand how zHPF improves upon FICON, one needs to review the relevant characteristics of FICON channel processing.

A FICON channel program consists of a series of Channel Command Words (CCWs) which form a chain. The command code indicates whether the I/O operation is going to be a read or an update write from disk, and the count field specifies the number of bytes to transfer. When the channel finishes processing one CCW and either a command chaining or data chaining flag is turned on, it processes the next CCW, and the CCWs belonging to such a series are said to be chained. Each of these CCWs is packaged into a FICON channel Information Unit (IU) which requires separate processing on the FICON channel processor and separate commands to be sent across the link from the channel to the control unit. The zHPF architecture defines a single command block to replace a series of FICON CCWs as illustrated in Figure 3 below.

zHPF improves upon FICON by providing a Transport Control Word (TCW) that facilitates the processing of an I/O request by the channel and the control unit. The TCW has a capability that enables multiple channel commands to be sent to the control unit as a single entity instead of being sent as separate commands as is done with FICON CCWs. In addition, the channel is no longer expected to process and keep track of each individual channel command word. Instead, the channel forwards a chain of commands to the control unit to execute. The reduction of this overhead increases the maximum I/O rate possible on the channel and improves the utilization of the various sub-components along the path traversed by the I/O request.

zHPF provides a much simpler link protocol than FICON. Figure 3 below shows an example of a 4k read FICON channel program, where three IUs are sent from the channel to the control unit plus three IUs from the control unit to the channel. In this example, zHPF reduces the total number of IUs sent in half, using one IU from the channel to the control unit and two IUs from the control unit to the channel.

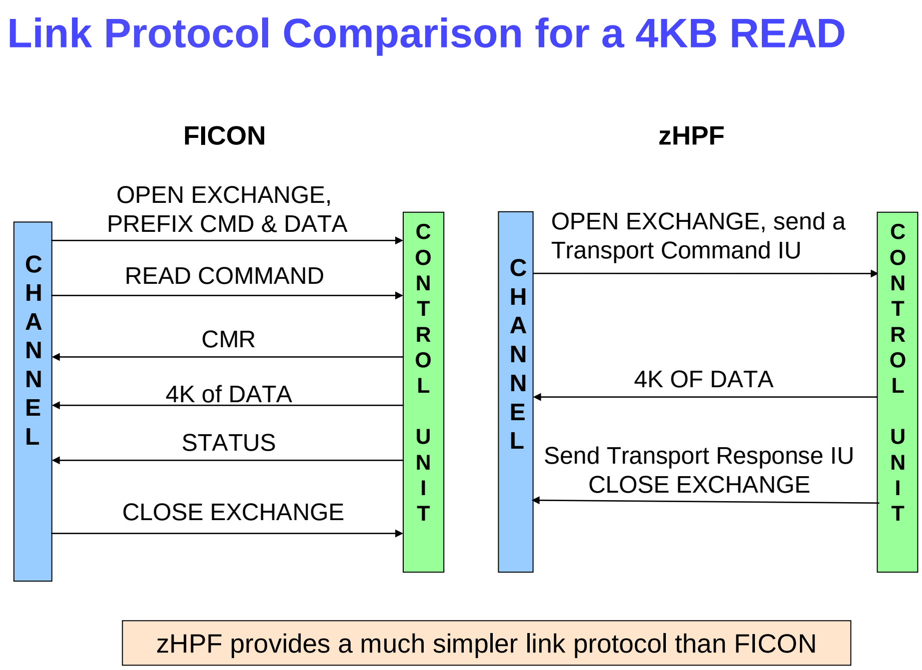


Figure 3 – Link Protocols

With zHPF, "well constructed" CCW strings are collapsed into a single new Control Word. Conceptually this is similar to the Modified Indirect Data Address Word (MIDAW) facility enhancement to FICON, which allowed a chain of data CCWs to be collapsed into one CCW. zHPF now allows the collapsing of both Command Chained as well as Data Chained CCW strings into one Control Word. zHPF-capable channels and devices support both FICON and zHPF protocols simultaneously.

The maximum number of open exchanges or the number of I/Os that can be simultaneously active on FEx32S channels is designed to be significantly higher with zHPF compared to FICON. An open exchange is an I/O that is active between the channel and the control unit, and it includes I/Os that are cache hits, which begin transferring data back to the channel immediately and those that are cache misses which might experience a delay of several milliseconds before the data can begin transferring back to the channel. Since higher I/O activity levels are possible now and expected to increase in the future with zHPF, the maximum number of open exchanges allowed per channel has been increased with zHPF. More information on how to find the actual average number of open exchanges used in a production workload environment is provided in the "zHPF fields on the RMF Channel Activity report" section of this document.

zHPF can be turned on or off. For z/OS exploitation, there is a parameter in the IECIOSxx member of SYS1.PARMLIB (ZHPF=YES/NO) and on the SETIOS command to control whether zHPF is enabled or disabled. The default is ZHPF=NO.

For zHPF exploitation, FEx32S (CHPID type FC) on an IBM z16™ requires at a minimum:

z/OS

- Version 2 Release 5 with PTFs
- Version 2 Release 4 with PTFs
- Version 2 Release 3 with PTFs
- Version 2 Release 2
    - IBM Software Support Services purchase

zHPF-capable channels and devices support both FICON and zHPF protocols simultaneously. The Media Manager component of DFSMS™ detects whether the device supports zHPF or not and builds the appropriate channel programs. Media Manager will build the zHPF Transport Mode channel programs for DB2, PDSE, VSAM, QSAM, BPAM, BSAM, zFS and Extended Format SAM that could benefit from the improved transfer technique.

FEx32S Channel Performance Using zHPF and FICON Protocols

With the introduction of the FEx32S channel, improvements can be seen in response times and maximum throughput for MB/sec for workloads using the zHPF protocol on FEx32S channels.

One primary performance improvement in a FEx32S channel with zHPF compared to FICON is the maximum number of small block IO/sec (4k bytes per I/O) that can be processed. As displayed in Figure 4 below, the maximum number of IO/sec that was measured on a FEx32S channel running an I/O driver benchmark with a 4k bytes per I/O workload exploiting zHPF is 302K IO/sec, which is approximately twelve times what was measured with FICON. In this case, two separate measurements were conducted: one where the channel was executing all zHPF channel programs and another where the channel was executing all FICON channel programs.

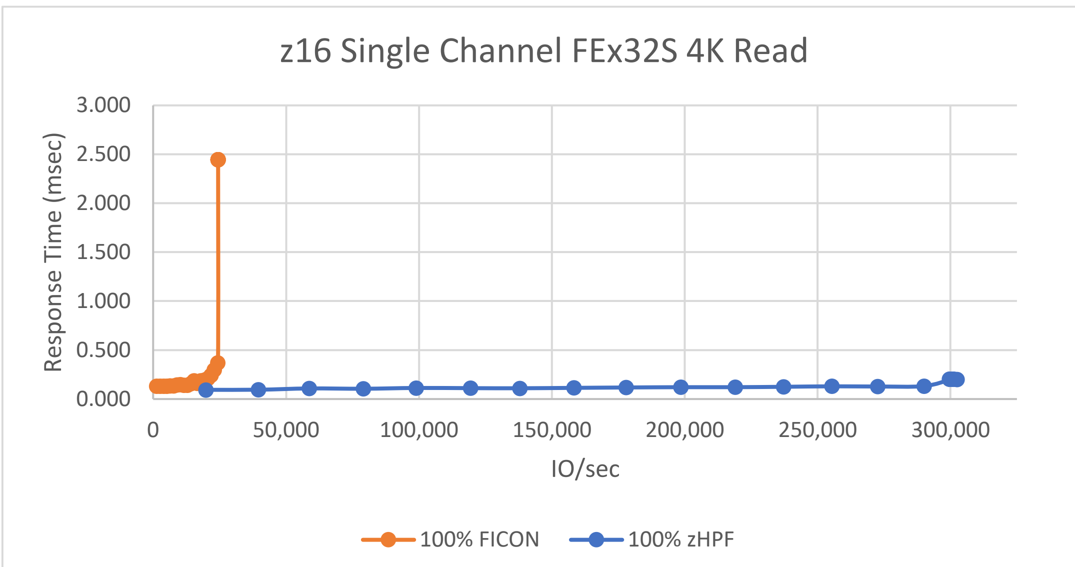


Figure 4 – FEx32S Maximum IO/sec

Figure 4 shows that the total response time for a 4k read I/O operation is approximately 100 microseconds better for zHPF compared to FICON channel programs at I/O activity levels less than 20,000 IO/sec per channel. For FICON, response times increase gradually to approximately 0.3ms for activity levels up to 18,000 IO/sec per channel and then increase sharply at a level of activity between 20,000 and 24,000 IO/sec. In contrast, response times for zHPF continue at a moderate rate of increase up to approximately 0.200ms at 300,000 IO/sec, reaching the maximum capability of the channel. Please note that these measurements were done with a single IBM z16™ channel connected to multiple host adapter ports, each on a different IBM System

Storage DS8900. Just as is the case with FICON, the maximum capability of any individual control unit (CU) port running zHPF channel programs is different than the maximum capability of the zHPF channel. Contact an IBM representative for more information on the performance of a DS8900.

Since the difference in maximum capability with 100% zHPF activity compared to 100% FICON activity is so great on the FEx32S channel, it is worthwhile to note how the maximum capability changes for I/O benchmarks that read 4K bytes of data using a mixture of zHPF and FICON protocols. As shown in Figure 5 below, with a mixture of 90% zHPF and 10% FICON, the maximum capability is 280K IO/sec. With a mix of 50% zHPF and 50% FICON, the maximum capability is 160K IO/sec on a FEx32S channel. Higher performance improvements can be experienced with higher percent zHPF activity.

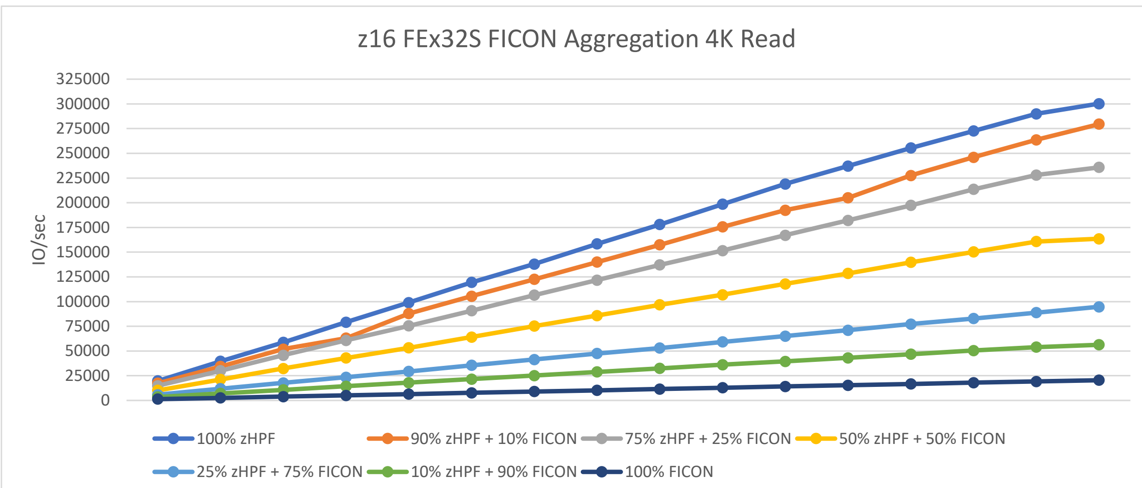


Figure 5 – FICON Aggregation

As shown in Figure 6 below, using zHPF on a FEx32S channel, a range from 1,200 MB/sec for a single 4k block to over 6,000 READ/WRITE (mix) MB/sec when 16 or more 4k blocks are transferred per I/O operation was achieved. In contrast, using the FICON protocol with the MIDAW facility, a range from 100 MB/sec for a single 4k block to over 700 READ/WRITE (mix) MB/sec when 16 or more 4k blocks are transferred per I/O operation was measured on a FEx32S channel. The maximum MB/sec possible with zHPF ranges from 7 times FICON for large data transfers (128 k bytes) to 12 times FICON for smaller data transfers (8k bytes per I/O). The graph below also demonstrates the performance benefits of the new FEx32S efficiencies of the zHPF protocol, which increases the average bytes per frame and reduces the number of frames per I/O that is transferred.

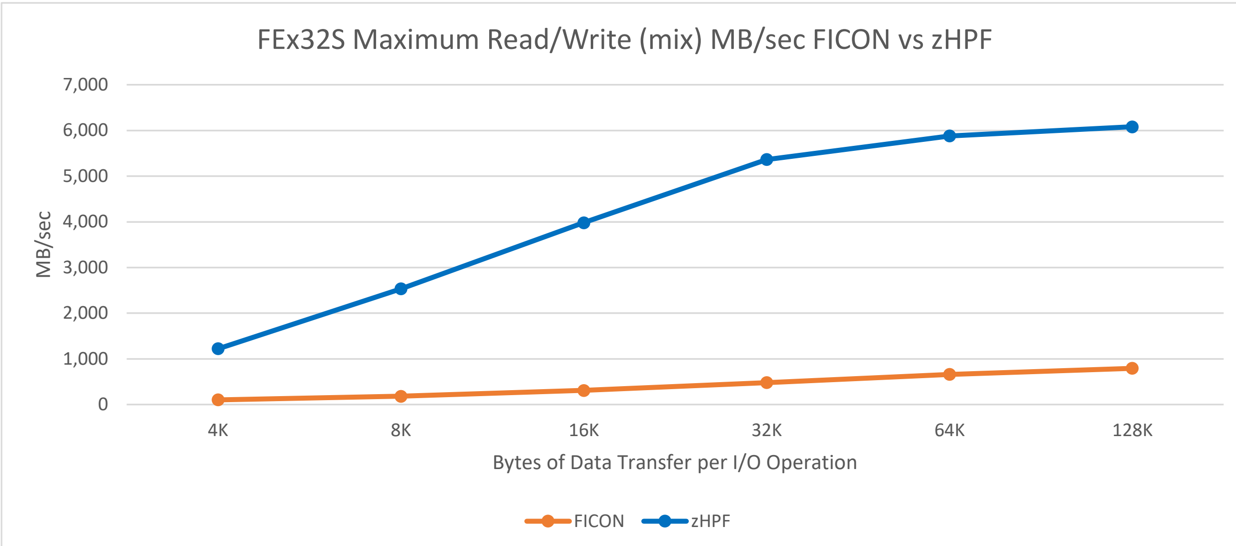


Figure 6 – FEx32S Maximum MB/sec

Examples of the response time benefits of both the new FEx32S channel and the efficiencies of the zHPF protocol are shown in Figure 7 below. Simple benchmark measurements were conducted using a FEx32S channel to READ or update WRITE 32 4K records or 128K bytes of data in total to a single device. PEND and CONN time response time components are displayed for zHPF compared to the FICON protocol. For I/O operations that READ 128K bytes of data, zHPF response times are 30% faster than FICON which is displayed in the CONN time component. For I/O operations that update WRITE 128K bytes of data, zHPF response times are 20% faster than FICON.

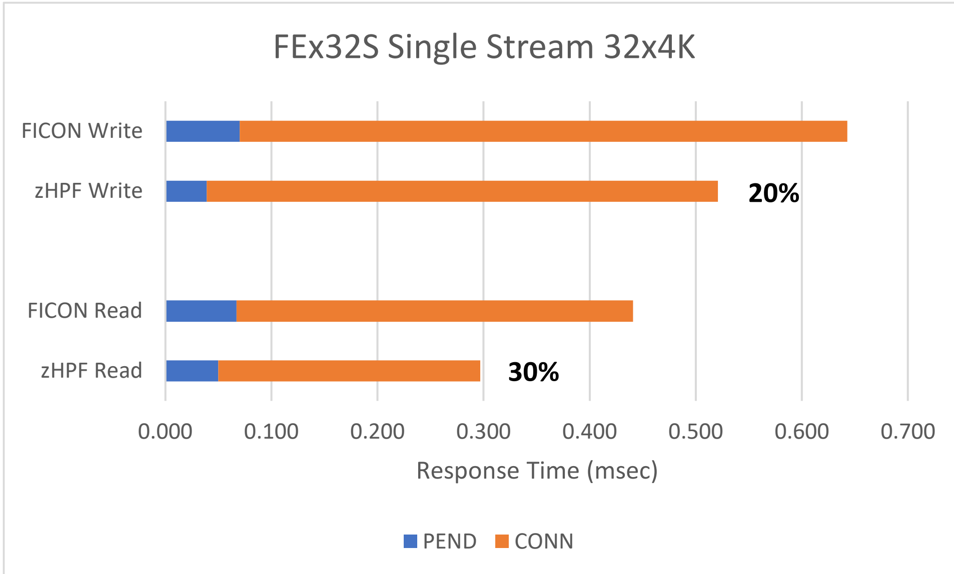


Figure 7 – FEx32S Response Time

Benefits of both the new FEx32S channel and the efficiencies of the zHPF protocol can also be observed in the channel processor utilization reported in the Resource Measurement Facility (RMF) Channel Activity Report. Figure 8 below shows the FEx32S channel processor utilization reported for both zHPF and FICON measurements for 32x4K READ I/O operations. Since the hardware data router is handling the movement of data for zHPF, the maximum FEx32S channel processor utilization reported for these large data transfer zHPF I/O operations is less than 12%. In contrast, the FEx32S channel microprocessor is involved in processing both the Channel Command Words (CCWs) and the transfer of each of the 8K byte data information units (IUs). The channel utilization reported for equivalent amounts of MB/sec processed is more than 38 times what is used for zHPF large data transfers. For large data transfer I/O operations using the zHPF protocol it is therefore more appropriate to observe either channel bus or link utilizations.

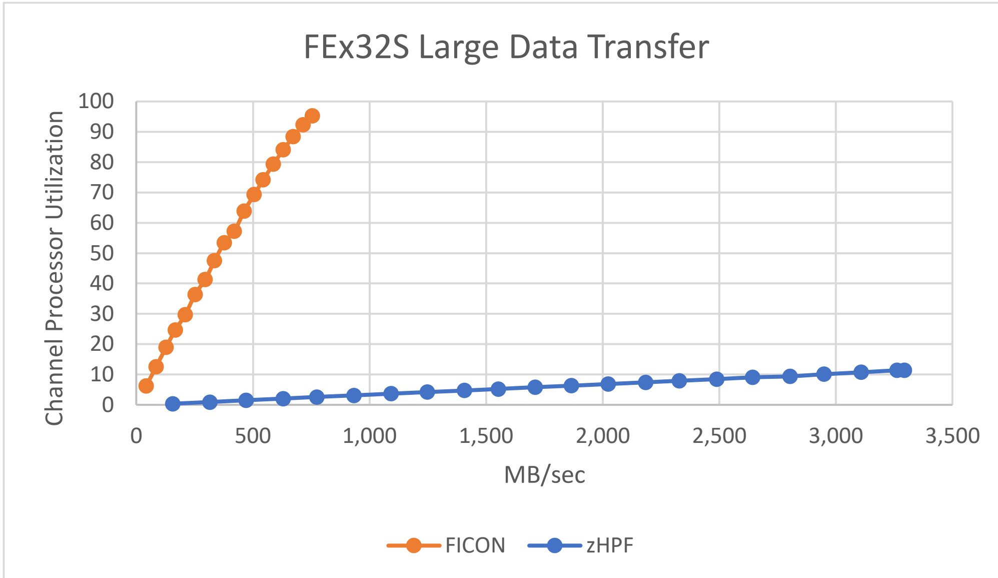
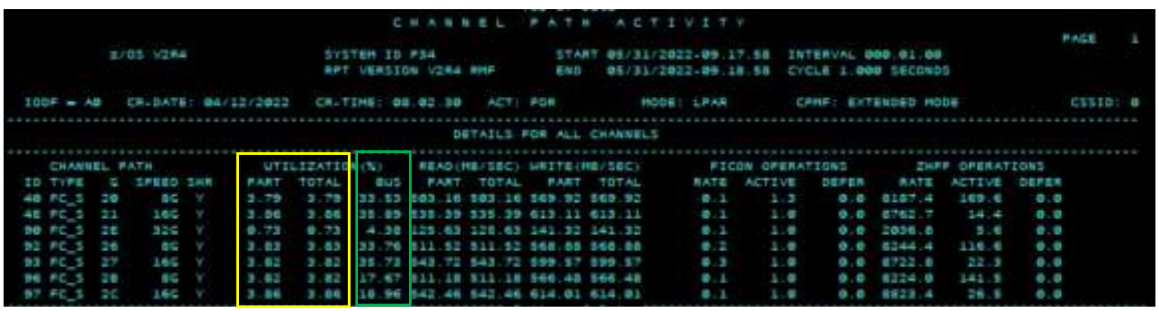


Figure 8 – FEx32S Channel Processor Utilization

In general, IBM recommends the following guidelines to achieve good response times:

- keep Channel Processor Utilizations less than 50%
- keep Channel BUS Utilizations less than 70%



C H A N N E L   P A T H   A C T I V I T Y

PAGE 1

z/OS V2R4 SYSTEM ID P34 START 05/31/2022-09.17.58 INTERVAL 000.01.00
RPT VERSION V2R4 RMF END 05/31/2022-09.18.58 CYCLE 1.000 SECONDS

IODF = A0 CR-DATE: 04/12/2022 CR-TIME: 08.02.38 ACT: POR MODE: LPAR CPMF: EXTENDED MODE CSSID: 0

DETAILS FOR ALL CHANNELS

CHANNEL PATH					UTILIZATION(%)			READ(MB/SEC)		WRITE(MB/SEC)		FICON OPERATIONS			ZHPF OPERATIONS		
ID	TYPE	G	SPEED	SHR	PART	TOTAL	BUS	PART	TOTAL	PART	TOTAL	RATE	ACTIVE	DEFER	RATE	ACTIVE	DEFER
40	FC_S	20	8G	Y	3.79	3.79	33.53	503.16	503.16	569.92	569.92	0.1	1.3	0.0	8187.4	169.6	0.0
4E	FC_S	21	16G	Y	3.06	3.06	35.85	535.39	535.39	613.11	613.11	0.1	1.0	0.0	8762.7	14.4	0.0
90	FC_S	2E	32G	Y	0.73	0.73	4.30	125.63	125.63	141.32	141.32	0.1	1.0	0.0	2036.0	5.6	0.0
92	FC_S	26	8G	Y	3.83	3.83	33.76	511.52	511.52	568.88	568.88	0.2	1.0	0.0	8244.4	116.6	0.0
93	FC_S	27	16G	Y	3.82	3.82	35.73	543.72	543.72	599.57	599.57	0.3	1.0	0.0	8722.8	22.3	0.0
96	FC_S	28	8G	Y	3.82	3.82	17.67	511.18	511.18	566.48	566.48	0.1	1.0	0.0	8224.0	141.5	0.0
97	FC_S	2C	16G	Y	3.86	3.86	18.96	542.46	542.46	614.01	614.01	0.1	1.0	0.0	8823.4	26.5	0.0

Figure 9 – FEx32S Channel Utilizations

When using either the FICON or the zHPF protocol, the FEx32S channel processor utilization is the most important factor to observe for small data transfers. The link or bus utilization is a more important factor to observe for large data transfers using the zHPF protocol. For large data transfers using the FICON protocol, both the channel processor and link utilization guidelines should be observed.

In summary, the FEx32S channel offers performance improvements in both response times and maximum throughput for IO/sec and MB/sec for workloads using the zHPF protocol.

FEx32S Channel Performance Using zHPF and FICON Protocols at Up to 100km of Distance

IBM recommends the use of Cascaded FICON directors with 32Gbps or greater Inter Switch Links (ISLs) between the two directors to support distances up to 100km with the use of DWDMs.

The FEx32S channel has 90 buffer credits for each individual channel port which is enough to support a 32 Gbps link speed at distances up to 5 km between each channel and its nearest neighbor (i.e. the longest such distance supported by a FEx32S channel), which could be a port on a director, or tape, or disk storage subsystem.

To evaluate FEx32S channel performance using zHPF and FICON protocols at 100 km of distance, using Dense Wavelength Division Multiplexer (DWDM), a series of performance measurements were conducted with multiple FEx32S channels on IBM z16™ connected to multiple storage subsystems using cascaded FICON directors with multiple 32 Gbps InterSwitch Link (ISL) between the two directors extended over a Dense Wavelength Division Multiplexer (DWDM) as illustrated in Figure 10 below.

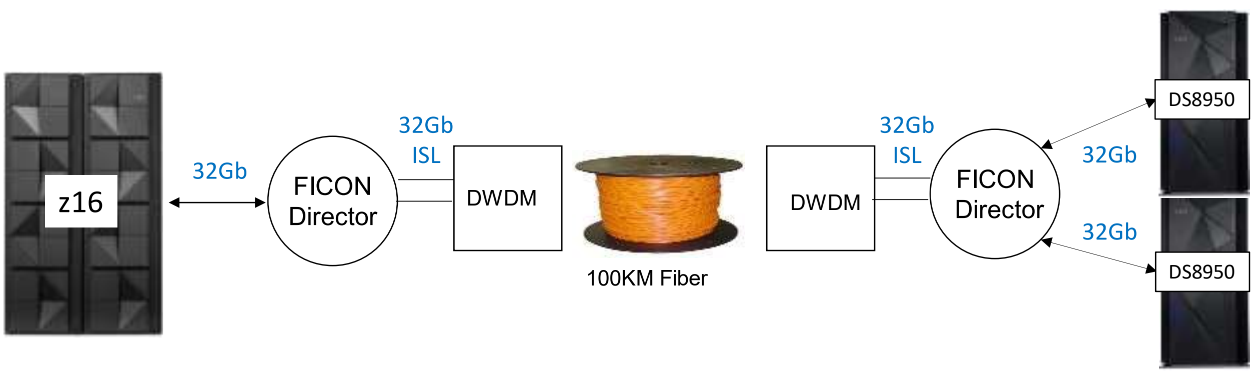


Figure 10 – Distance Diagram

The amount of throughput that can be achieved on an ISL at distances up to 100 km is dependent on a number of variables:

- Workload characteristics such as data transfer sizes and mix of READ and WRITE activity
- zHPF vs. FICON protocols
- Link speed and the bit encoding scheme used on the ISL
- Buffer credits available on the ISL ports on each of the FICON directors
- Actual distance between the two ISL ports on each of the FICON directors

Figure 11 below displays the maximum MB/sec achieved on 32 Gbps ISLs between two FICON directors separated by 100 km DWDM with both 720, and 1690 Buffer-to-Buffer (B2B) credits available on each of the ISL ports. Previously, the general recommendation was to use 0.5 buffers for each 1 km of distance and each 1 Gbps of link speed. This implies 50, 100, 200, 400 and 500 B2B credits for 1, 2, 4, 8 and 10 Gbps link speeds, respectively. These measurements used 1690 B2B credits to ensure the maximum was achieved.

For 128k byte READ or update WRITE I/O operations with 1690 B2B credits, a maximum of 3,200 MB/sec was achieved using zHPF protocols compared to 2,560 MB/sec with 720 B2B credits. For a 50/50 mix of 128k byte READ and WRITE I/O operations with 1690 B2B credits, a maximum of 6,020 MB/sec was achieved using zHPF protocols compared to 4,500 MB/sec with 720 B2B credits.

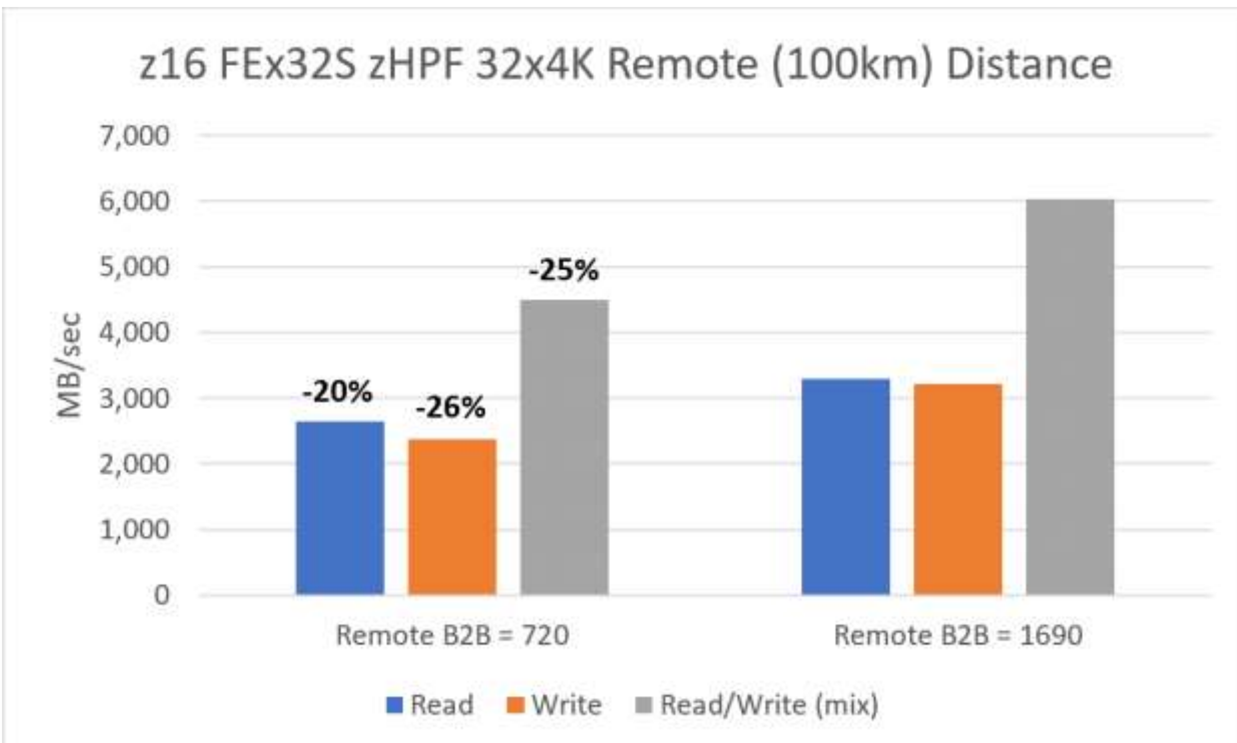


Figure 11 – Distance zHPF Large Data Transfer

Figure 12 below summarizes for READ or WRITE I/O operations, the average frame size shown is for the data frames flowing from the ISL port adjacent to the storage subsystem to the ISL port closest to the IBM z16™.

Data Transfer	zHPF Average Frame Size
4K Read or Write	1083
32x4K Read or Write	2023
128x4K Read or Write	2068

Figure 12 - Average Frame Sizes

From a response time perspective, most workloads will experience an additional 1 millisecond (ms) at 100 km of distance compared to local response times due to the speed of light through a fiber. Figure 13 below summarizes an additional 1ms to response times at reasonable levels of activity for READ I/O operations with small data transfers (4k bytes per I/O), large data transfers (32x4k or 128k bytes per I/O) and very large data transfers (128x4k or approximately 512k bytes per I/O) at 100 km compared to Local distance.

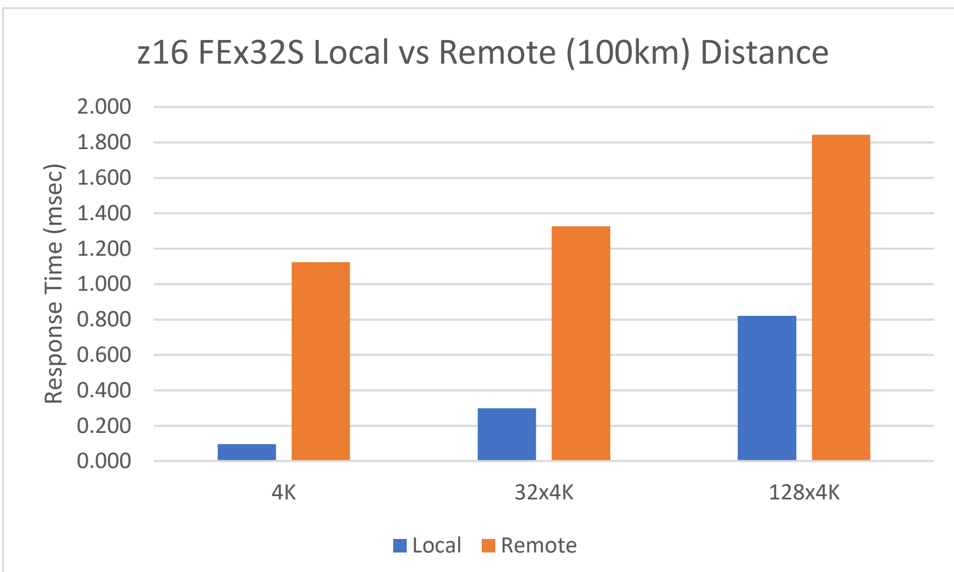


Figure 13 – Local vs Remote Distance

RMF Channel Activity Report

There are six fields in the RMF Channel Activity report that can be used to distinguish between FICON and zHPF traffic.

- The RATE field refers to the number of FICON or zHPF I/Os per second initiated at the total physical channel level (not by LPAR)
- The ACTIVE field refers to what we have previously called the "open exchanges", i.e. the number of I/Os that are simultaneously active within a channel.
- The DEFER field refers to the number of deferred FICON or zHPF I/O operations per second. This is the number of operations that could not be immediately initiated by the channel due to a temporary lack of resources.

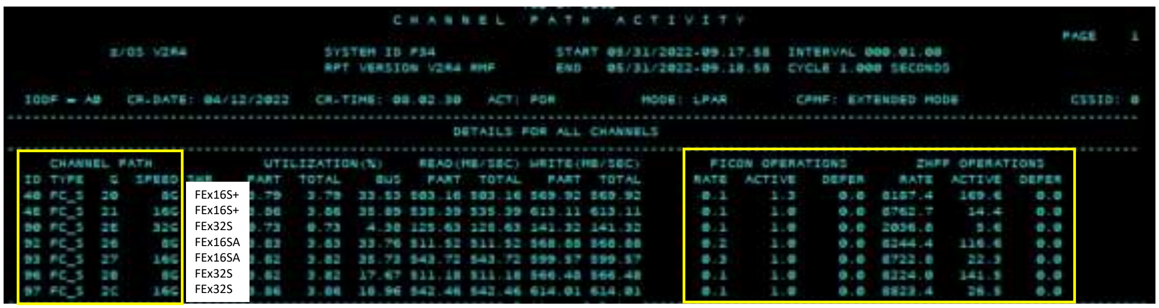


Figure 14 – RMF Diagram

Figure 14 above shows a mix of FICON and zHPF I/O operations on each channel as shown in the "G" and "SPEED" fields.

- CHPID 90 (FEx32S), the channel link is operating at 32 Gbps
- CHPID 97 (FEx32S), the channel link is operating at 16 Gbps
- CHPID 96 (FEx32S), the channel link is operating at 8 Gbps
- CHPID 93 (FEx16SA), the channel link is operating at 16 Gbps
- CHPID 92 (FEx16SA), the channel link is operating at 8 Gbps
- CHPID 4E (FEx16S+), the channel link is operating at 16Gbps
- CHPID 40 (FEx16S+), the channel link is operating at 8Gbps

The FEx32S channels support the following generation field ("G" column) of the RMF Channel Activity:

- "2B" the channel link has autonegotiated to a speed of 8 Gbps
- "2C" the channel link is operating at a speed of 16 Gbps
- "2D" the channel link is operating at a speed of 16Gbps with FEC (Forward Error Correction)
- "2E" the channel link is operating at a speed of 32 Gbps with FEC (Forward Error Correction)

The FEx16SA channels support the following generation field ("G" column) of the RMF Channel Activity:

- "26" the channel link has autonegotiated to a speed of 8 Gbps
- "27" the channel link is operating at a speed of 16 Gbps
- "28" the channel link is operating at a speed of 16Gbps with FEC (Forward Error Correction)

The FEx16S+ channels support the following generation field ("G" column) of the RMF Channel Activity:

- "1F" the channel link has autonegotiated to a speed of 4 Gbps
- "20" the channel link has autonegotiated to a speed of 8 Gbps
- "21" the channel link is operating at a speed of 16 Gbps
- "22" the channel link is operating at a speed of 16Gbps with FEC (Forward Error Correction)

For additional descriptions please refer to RMF User's Guide:
https://www-40.ibm.com/servers/resourcelink/svc00100.nsf/pages/zOSV2R4sc342664?OpenDocument

For additional information on FEC (Forward Error Correction) please refer to IBM z16 (3931) Technical Guide:
https://www.redbooks.ibm.com/redbooks/pdfs/sg248951.pdf

zHyperLink Express

zHyperLink Express is a feature introduced on the IBM z14™ system and the DS8880 storage family that will enable Synchronous I/O. zHyperLink Express provides a low latency direct connection between an IBM z16™ system and a DS8900. This is a point-to-point connection using PCIe Gen3 x8 physical and link layers. The optical MPO cables and transceivers are the same as those defined for sysplex coupling links, with a maximum cable length of 150m being supported. A new transport protocol is defined for reading ECKD data records.

IBM z16™ hardware support for this feature consists of PBU (PCI-Express Bridge Unit) and ZHB (zSystem Host Bridge) changes on the IBM z16™ processor and an adapter, named zHyperLink Express. The PBU changes consist of CRC generation and checking for individual data records and also some performance enhancements for dynamically monitoring each data transfer and determining when an operation completes. The zHyperLink Express adapter provides 2 zHyperLink Express ports and is managed using native z/PCI commands, with some minor enhancements for Synchronous I/O. A maximum of 8 zHyperLink Express adapters are supported in each IBM z16™ system.

The following key assumptions apply for IBM z16™:

- Only ECKD supported
- FICON/HPF paths required in addition to zHyperLink Express path for backup/recovery
- Only native LPAR supported
- All data transfers must be 16-byte aligned and with length being a multiple of 16 bytes
- ECKD Reads supported
  - Maximum Read Record length 4K

This low latency interface provides the opportunity for the operating system to read the data records synchronously, thus avoiding the scheduling and interrupt overhead associated with asynchronous operations. Consequently, a new Synchronous I/O command has been defined to allow the operating system to synchronously read one data record. Access to the zHyperLink Express adapter links are provided using z/PCI constructs such as a PCI Function Identifier. Up to 64 virtual functions are supported for each zHyperLink Express interface (2 per adapter). The zHyperLink Express Synchronous I/O Paradigm is illustrated in Figure 15 below.

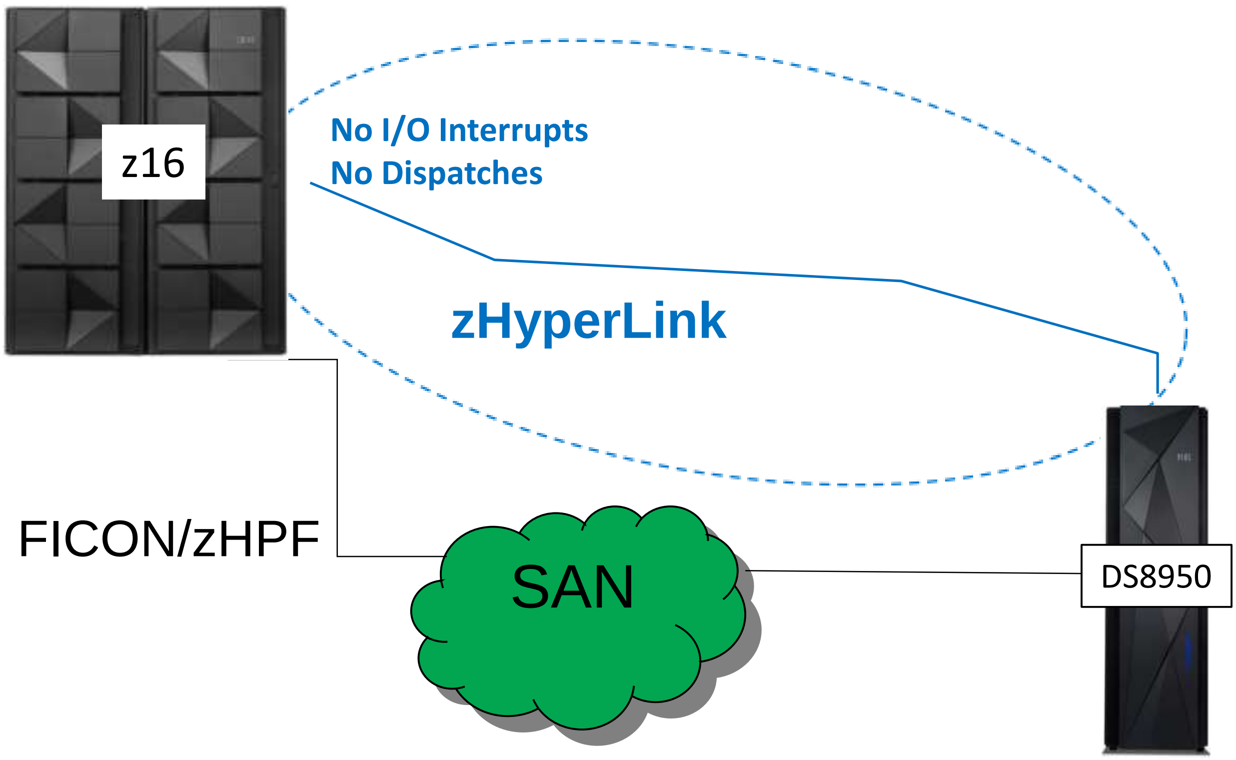


Figure 15 – zHyperLink Diagram

As illustrated in Figure 16, when using an IBM z16™ zHyperLink Express attached to DS8950 storage, Single-Stream 4K Read link latency is reduced by up to 4x as compared to IBM z16™ FEx32S attached to a DS8950 (direct connect) with a single 32Gb Host Adapter. Single-Stream 4K Write link latency is reduced by up to 3x as compared to IBM z16™ FEx32S attached to a DS8950 with a single 32Gb Host Adapter.

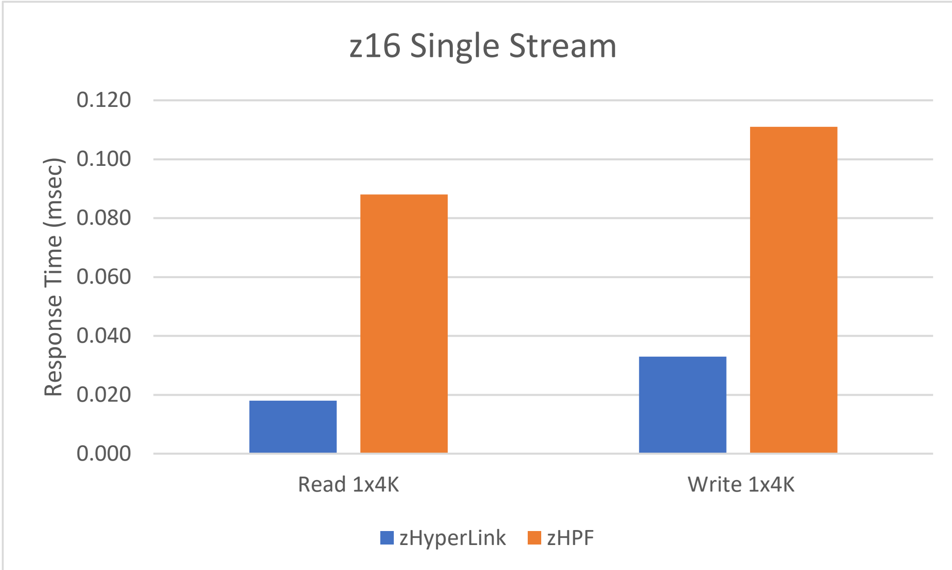


Figure 16 – zHyperlink Express vs FEx32S

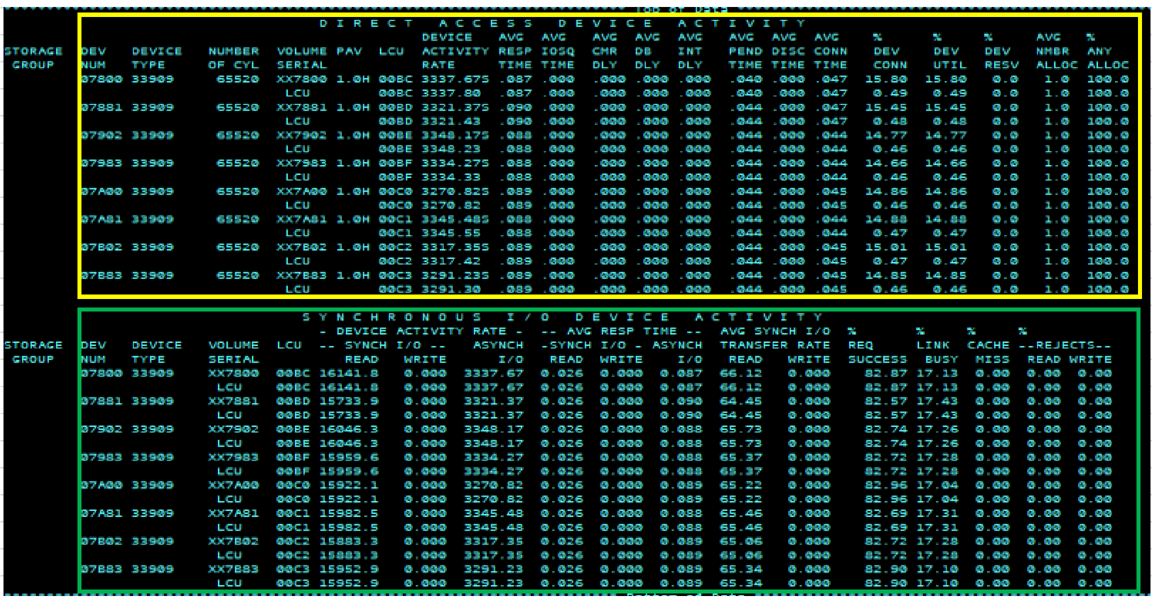
For additional information on zHyperLink Express please refer to the following guide: Getting Started with IBM zHyperLink Express for z/OS --> http://www.redbooks.ibm.com/Redbooks.nsf/RedbookAbstracts/redp5493.html?Open

RMF Synchronous I/O Device Activity

There is a section of the RMF Report for Synchronous I/O Device Activity that can be used to distinguish between synchronous and asynchronous device activity, Figure 17 below.

- DIRECT ACCESS DEVICE ACTIVITY REPORT
    - Device Activity Rate: Synchronous I/O Activity will be indicated by an "S".
- SYNCRONOUS I/O DEVICE ACTIVITY REPORT
    - Synchronous I/O Device Activity

For additional descriptions please refer to RMF User's Guide:
https://www-40.ibm.com/servers/resourcelink/svc00100.nsf/pages/zOSV2R4sc342664?OpenDocument



D I R E C T   A C C E S S   D E V I C E   A C T I V I T Y

STORAGE GROUP	DEV NUM	DEVICE TYPE	NUMBER OF CYL	VOLUME SERIAL	PAV	LCU	DEVICE ACTIVITY RATE	AVG RESP TIME	AVG IOSQ TIME	AVG CMR DLY	AVG DB DLY	AVG INT DLY	AVG PEND TIME	AVG DISC TIME	AVG CONN TIME	% DEV CONN	% DEV UTIL	% DEV RESV	AVG NMBR ALLOC	% ANY ALLOC
	07800	33909	65520	XX7800	1.0H	00BC	3337.675	.087	.000	.000	.000	.000	.040	.000	.047	15.80	15.80	0.0	1.0	100.0
				LCU		00BC	3337.80	.087	.000	.000	.000	.000	.040	.000	.047	0.49	0.49	0.0	1.0	100.0
	07881	33909	65520	XX7881	1.0H	00BD	3321.375	.090	.000	.000	.000	.000	.044	.000	.047	15.45	15.45	0.0	1.0	100.0
				LCU		00BD	3321.43	.090	.000	.000	.000	.000	.044	.000	.047	0.48	0.48	0.0	1.0	100.0
	07902	33909	65520	XX7902	1.0H	00BE	3348.175	.088	.000	.000	.000	.000	.044	.000	.044	14.77	14.77	0.0	1.0	100.0
				LCU		00BE	3348.23	.088	.000	.000	.000	.000	.044	.000	.044	0.46	0.46	0.0	1.0	100.0
	07983	33909	65520	XX7983	1.0H	00BF	3334.275	.088	.000	.000	.000	.000	.044	.000	.044	14.66	14.66	0.0	1.0	100.0
				LCU		00BF	3334.33	.088	.000	.000	.000	.000	.044	.000	.044	0.46	0.46	0.0	1.0	100.0
	07A00	33909	65520	XX7A00	1.0H	00C0	3270.825	.089	.000	.000	.000	.000	.044	.000	.045	14.86	14.86	0.0	1.0	100.0
				LCU		00C0	3270.82	.089	.000	.000	.000	.000	.044	.000	.045	0.46	0.46	0.0	1.0	100.0
	07A81	33909	65520	XX7A81	1.0H	00C1	3345.485	.088	.000	.000	.000	.000	.044	.000	.044	14.88	14.88	0.0	1.0	100.0
				LCU		00C1	3345.55	.088	.000	.000	.000	.000	.044	.000	.044	0.47	0.47	0.0	1.0	100.0
	07B02	33909	65520	XX7B02	1.0H	00C2	3317.355	.089	.000	.000	.000	.000	.044	.000	.045	15.01	15.01	0.0	1.0	100.0
				LCU		00C2	3317.42	.089	.000	.000	.000	.000	.044	.000	.045	0.47	0.47	0.0	1.0	100.0
	07B83	33909	65520	XX7B83	1.0H	00C3	3291.235	.089	.000	.000	.000	.000	.044	.000	.045	14.85	14.85	0.0	1.0	100.0
				LCU		00C3	3291.30	.089	.000	.000	.000	.000	.044	.000	.045	0.46	0.46	0.0	1.0	100.0

S Y N C H R O N O U S   I / O   D E V I C E   A C T I V I T Y

STORAGE GROUP	DEV NUM	DEVICE TYPE	VOLUME SERIAL	LCU	- DEVICE ACTIVITY RATE - -- SYNCH I/O -- READ	WRITE	ASYNCH I/O	-- AVG RESP TIME -- -SYNCH I/O - READ	WRITE	ASYNCH I/O	AVG SYNCH I/O TRANSFER RATE READ	WRITE	% REQ SUCCESS	% LINK BUSY	% CACHE MISS	% --REJECTS-- READ	WRITE
	07800	33909	XX7800	00BC	16141.8	0.000	3337.67	0.026	0.000	0.087	66.12	0.000	82.87	17.13	0.00	0.00	0.00
			LCU	00BC	16141.8	0.000	3337.67	0.026	0.000	0.087	66.12	0.000	82.87	17.13	0.00	0.00	0.00
	07881	33909	XX7881	00BD	15733.9	0.000	3321.37	0.026	0.000	0.090	64.45	0.000	82.57	17.43	0.00	0.00	0.00
			LCU	00BD	15733.9	0.000	3321.37	0.026	0.000	0.090	64.45	0.000	82.57	17.43	0.00	0.00	0.00
	07902	33909	XX7902	00BE	16046.3	0.000	3348.17	0.026	0.000	0.088	65.73	0.000	82.74	17.26	0.00	0.00	0.00
			LCU	00BE	16046.3	0.000	3348.17	0.026	0.000	0.088	65.73	0.000	82.74	17.26	0.00	0.00	0.00
	07983	33909	XX7983	00BF	15959.6	0.000	3334.27	0.026	0.000	0.088	65.37	0.000	82.72	17.28	0.00	0.00	0.00
			LCU	00BF	15959.6	0.000	3334.27	0.026	0.000	0.088	65.37	0.000	82.72	17.28	0.00	0.00	0.00
	07A00	33909	XX7A00	00C0	15922.1	0.000	3270.82	0.026	0.000	0.089	65.22	0.000	82.96	17.04	0.00	0.00	0.00
			LCU	00C0	15922.1	0.000	3270.82	0.026	0.000	0.089	65.22	0.000	82.96	17.04	0.00	0.00	0.00
	07A81	33909	XX7A81	00C1	15982.5	0.000	3345.48	0.026	0.000	0.088	65.46	0.000	82.69	17.31	0.00	0.00	0.00
			LCU	00C1	15982.5	0.000	3345.48	0.026	0.000	0.088	65.46	0.000	82.69	17.31	0.00	0.00	0.00
	07B02	33909	XX7B02	00C2	15883.3	0.000	3317.35	0.026	0.000	0.089	65.06	0.000	82.72	17.28	0.00	0.00	0.00
			LCU	00C2	15883.3	0.000	3317.35	0.026	0.000	0.089	65.06	0.000	82.72	17.28	0.00	0.00	0.00
	07B83	33909	XX7B83	00C3	15952.9	0.000	3291.23	0.026	0.000	0.089	65.34	0.000	82.90	17.10	0.00	0.00	0.00
			LCU	00C3	15952.9	0.000	3291.23	0.026	0.000	0.089	65.34	0.000	82.90	17.10	0.00	0.00	0.00

Figure 17 – RMF Synchronous I/O

IBM Fibre Channel Endpoint Security

IBM Fibre Channel Endpoint Security is an end-to-end solution which ensures all data flowing between the new encryption capable 32Gbps FICON Express32S channel and DS8900 storage is encrypted and protected. IBM Fibre Channel Endpoint Security is an additional data security technology that contributes to the IBM Z approach of encryption everywhere, further minimizing the risk of a security breach by extending the value of pervasive encryption.

IBM Fibre Channel Endpoint Security provides in-flight protection for all data over FICON and Fibre Channel (FCP) links, independent of the operating system, file system, or access method in use.

There are 3 key components of IBM Fibre Channel Endpoint Security:

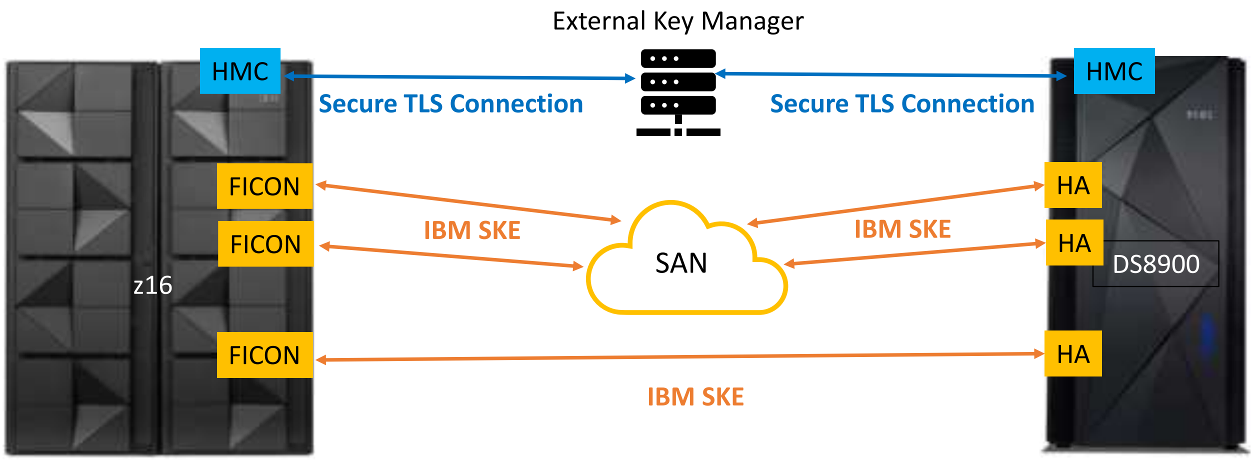


Figure 18 – IBM Fibre Channel Endpoint Security

1. IBM z16™ with FICON Express32S

   The IBM z16™ supports new encryption capable 32Gbps FICON cards for securing Channel-to-Storage Control Unit links.

2. IBM DS8900

   The new IBM DS8900 supports new 32Gbps encryption capable FICON adapter cards to support secured links between IBM Z and the storage control unit.

3. IBM Security Key Lifecycle Manager

ISKLM External Key Manager (EKM) appliance is required to serve a shared key to the server and storage that is used to secure messages between the endpoints in which further key material is derived/exchanged. Key Management Interoperability Protocol (KMIP) is the standard key management protocol used for key manager communication. The Channel and Storage System authenticate with the key server, and the Channel and Storage System CU query the server for shared keys.

Commercial Batch Workload Performance with IBM Fibre Channel Endpoint Security

To evaluate the effect of encryption enabled in the IBM Fibre Channel Endpoint solution we tested two different workload types. The first workload, known as 'Commercial Batch-Long' (CB-L), is a commercial batch job stream reflective of large batch jobs with fairly heavy CPU processing. The workload consists of a mix of Commercial Batch Jobs. The batch jobs include various combinations of C, COBOL, FORTRAN, and PL/I compile, link-edit, and execute steps. Sorting, DFSMS utilities (e.g. dump/restore and IEBCOPY), VSAM and DB2 utilities, SQL processing, GDDM® graphics, and FORTRAN engineering/scientific subroutine library processing are also included.

The results of running the CB-L workload is shown below in Figure 19. Execution of the CB-L workload with and without IBM Fibre Channel Endpoint Security encryption enabled demonstrated that IBM Fibre Channel Endpoint Security encryption had negligible impact on either the elapsed time, CPU consumption, or the I/O completion rate.

		I/O Interrupt	Elapsed
Fibre Channel Endpoint Security	CPU%	Rate	Time
Disabled	94.76	111,152.00	442.59
Enabled	93.86	109,883.12	444.09

Figure 19 – CB-L - IBM Fibre Channel Endpoint Security

Commercial Batch-Short (CB-S) is an I/O intensive batch workload which utilizes BSAM and QSAM access methods and was the second workload used to measure the effects of IBM Fibre Channel Endpoint Security encryption being enabled. CB-S consists of 18 jobs that repetitively copy data between input datasets and output datasets using various IBM standard utilities (e.g. IEBGENER, IEBDG, DFSORT, IFASMFDP, IDCAMS Repro and IEBCOMPR). For example, the DFSORT job splits one input dataset into three output datasets, and in the next step merges the three created datasets back into a single one.

		I/O Interrupt	Elapsed
Fibre Channel Endpoint Security	CPU%	Rate	Time
Disabled	5.18	12,458.55	621.7
Enabled	5.23	12,456.62	621.8

Figure 20 – CB-S - IBM Fibre Channel Endpoint Security

In both tests it was demonstrated that enabling IBM Fibre Channel Endpoint Security encryption had negligible impact on CPU consumption, elapsed job execution time and I/O completion rate.

There is a new field within the DISPLAY command for device (*M=DEV(xxxx),* where xxxx is the device address), indicating when IBM Fibre Channel Endpoint Security encryption is enabled or disabled. The Connection Security field signifies ENCR (encryption enabled) or NONE (encryption disabled).

```
D M=DEV(7800)
IEE174I 11.14.15 DISPLAY M 563
DEVICE 07800   STATUS=ONLINE
CHP                      50   51   52   53   54   55   56   57
ENTRY LINK ADDRESS       662B 6653 668B 66B3 66EB 6604 6664 66C4
DEST LINK ADDRESS        662C 66CF 666F 66B4 66DA 664A 66B2 66D5
PATH ONLINE              Y    Y    Y    Y    Y    Y    Y    Y
CHP PHYSICALLY ONLINE    Y    Y    Y    Y    Y    Y    Y    Y
PATH OPERATIONAL         Y    Y    Y    Y    Y    Y    Y    Y
MANAGED                  N    N    N    N    N    N    N    N
CU NUMBER                7800 7800 7800 7800 7800 7800 7800 7800
INTERFACE ID             0000 0032 0100 0132 0200 0232 0300 0332
CONNECTION SECURITY      Encr Encr Encr Encr Encr Encr Encr Encr
MAXIMUM MANAGED CHPID(S) ALLOWED:  0
DESTINATION CU LOGICAL ADDRESS = 00
SCP CU ND          = 002107.996.IBM.75.0000000KMX41.0000
SCP TOKEN NED      = 002107.900.IBM.75.0000000KMX41.0000
SCP DEVICE NED     = 002107.900.IBM.75.0000000KMX41.0000
WWNN               = 5005076309FFD439
HYPERPAV ALIASES CONFIGURED = 48
ZHYPERLINKS AVAILABLE = 16
FUNCTIONS ENABLED = MIDAW, ZHPF
```

Figure 21 – Device Display for IBM Fibre Channel Endpoint Security

The DISPLAY command for channel path (*D M=CHP(xx),PATHINFO*, where xx is the CHPID number), indicating when IBM Fibre Channel Endpoint Security encryption is enabled or disabled. The Connection Security Capability field identifies if the CHPID is eligible for encryption capability. The Conn Sec field identifies ENCR (encryption enabled) or NONE (encryption disabled).

```
D M=CHP(50),PATHINFO
IEE588I 11.16.44 DISPLAY M 565
CHPID 50:  TYPE=1B, DESC=FICON SWITCHED, ONLINE
Path Information for Channel Path 50
Connection Security Capability: Encryption
 Dest                                        Link-Speed   Conn
 Link Intf Node Descriptor                   Curr  Cap    Sec
 662C 0000 002107.996.IBM.75.0000000KMX41 .....  .....  Encr
```

Figure 22 – CHPID Display for IBM Fibre Channel Endpoint Security

FEx32S Channel Performance Using the FCP Protocol

FICON Express32S (FEx32S) supports the Fibre Channel Protocol (FCP) when defined as CHPID type FCP, allowing on IBM z16™ z/VM, z/VSE, and Linux on IBM Z LPARs to connect to industry-standard Small Computer System Interface (SCSI) storage controllers and devices . The following operating system versions are supported:

- SUSE SLES 15 SP3 or later
- SUSE SLES 12 SP5
- Red Hat RHEL 8.4 or later
- Red Hat RHEL 7.9
- Canonical Ubuntu 20.04.0x LTS
- zVM V7R2 with PTFs
- z/VM V7R1 with PTFs
- z/VSE V6.2
-

The latest generation IBM z16™ FCP channel offers significant performance improvements, delivering increased I/O throughput as well as response time benefits as a result of its 32Gb data rate link speed. This represents twice the link speed supported by the previous FICON Express 16SA (FEx16SA) on IBM z15™.

To evaluate the performance characteristics of FEx32S, measurements were performed on IBM z16™ by driving FCP operations using an internal Linux on IBM Z microbenchmark that initiates I/O operations. Both 4KB and 64KB data block sizes were tested. An LPAR running SuSE Linux Enterprise Server 12 was used, configured with 12 dedicated IFLs with SMT enabled and 16GB storage. The channel was tested under different load levels, varying the number of concurrent FCP operations incrementally: starting with a single I/O operation, up to hundreds of IOs in-flight. Two different configurations were tested:

1. *Driving FCP I/O operations using a single FEx32S port* (leaving the second port on the FEx32S adapter card unused),
2. *Driving FCP I/O operations on both ports available on the same FEx32S adapter*.

In all cases, FEx32S was connected to a 32Gb capable switch and IBM DS8950 storage with 32Gb FC ports. Both 4KB and 64KB data block sizes were used.

FCP Small Block Performance (IO/sec)

Based on a mixture of Read-only 4KB FCP operations, FEx32S delivers up to 600,000 IO/sec throughput per port; a 52% increase over FEx16SA. (A FEx32S adapter, with both ports executing the same mixture of FCP 4KB Read-only operations, supports up to 1.2 million IO/sec.) Figure 23 highlights these improvements in comparison to previous generations FCP channels going back to FICON Express4:

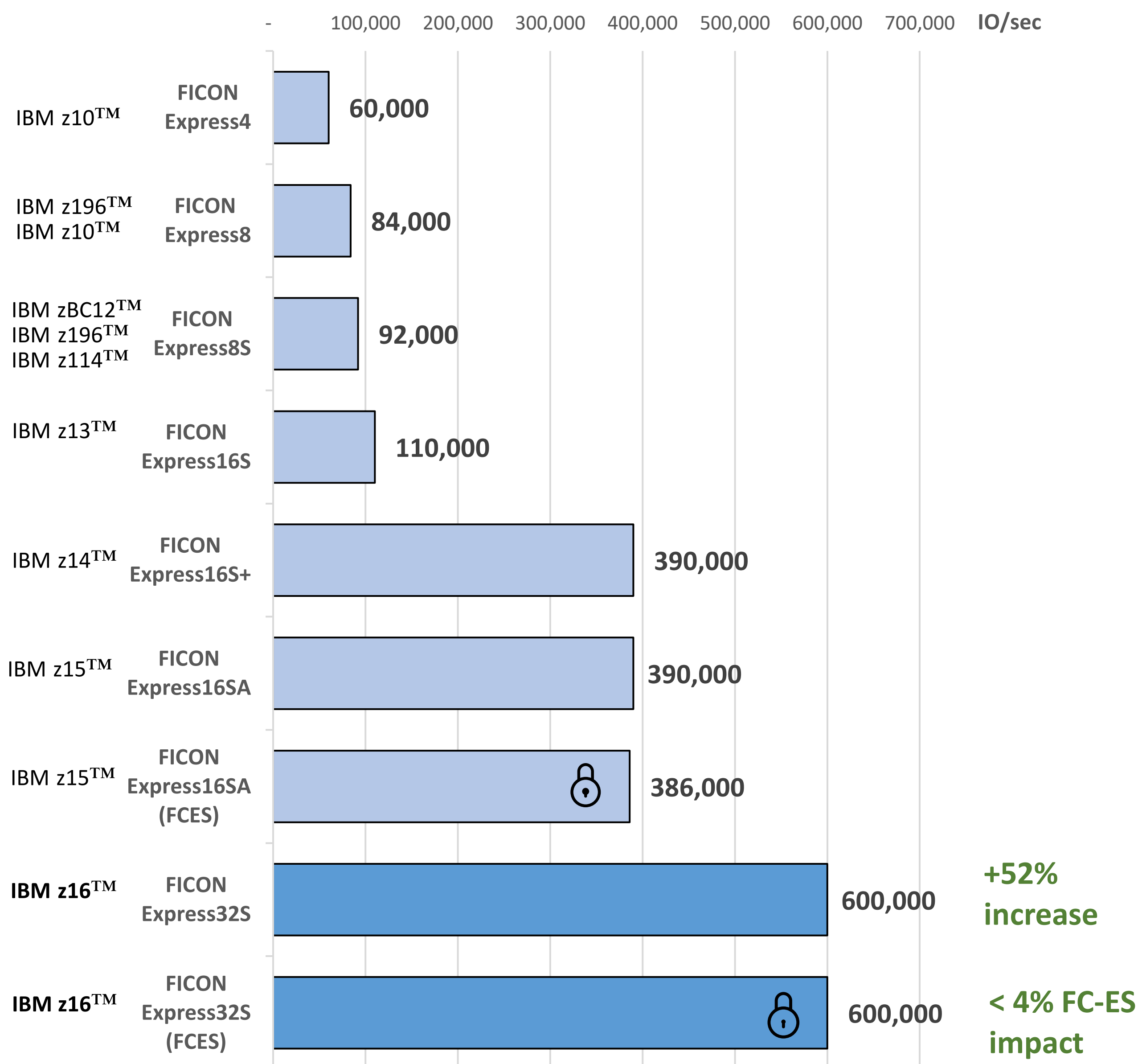


Figure 23 – Maximum FEx32S FCP Small Block Read IO/sec (Per Port)

In addition to the small block Read performance improvements highlighted above, FEx32S offers increased IO/sec capacity for Write and Read/Write (mix) operations. As displayed in Figure 24, a FEX32S FCP channel can deliver up to 460,000 Write IO/sec (+39%) and 590,000 Read/Write (mix) IO/sec (+5%) compared to FEx16SA.

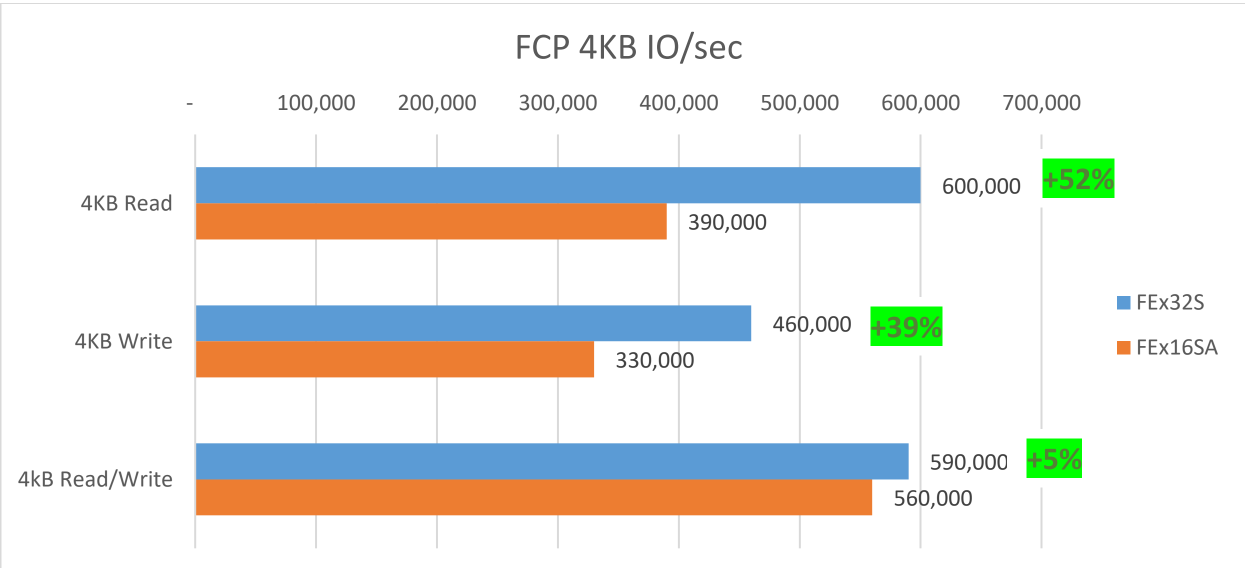


Figure 24 – FEx32S and FEx16SA FCP Small Block IO/sec Read, Write, Read/Write (mix)

Furthermore, FEx32S provides similar FCP I/O response time as FEx16SA, while avoiding the increase that occurs as throughput begins to saturate the 16Gb link. Based on measurements varying the number of concurrent FCP I/O operations, both channels deliver similar Read response time when processing up to 360,000 4KB Read-only IO/sec. However, their behavior starts to diverge as throughput on FEx16SA approaches 390,000 4KB Read IO/sec, the maximum supported by the 16Gb channel. (This throughput is roughly equivalent to 1,600 MB/sec, the unidirectional bandwidth limit of a 16Gb link. Therefore, the link is utilized at full capacity and additional I/O work pays a penalty in response time.)

In contrast, FEx32S FCP channels support a faster 32Gb link, overcoming the 390,000 4KB Read limit while providing a more gradual increase in I/O response time until approximately 550,000 IO/sec (see Figure 25). This may provide opportunities to consolidate FCP work from multiple FEx16SA channels into a single FEx32S depending on the workload and response time requirements. For example, consider two FEx16SA channels, each using a single port to process 200,000 4KB Read IO/sec at approximately 0.65ms response time. A single FEx32S port can process the same total amount of work (400,000 4KB Read IO/sec) while still delivering sub 0.70ms response time. This is roughly a 0.05ms increase versus splitting the work between individual FEx16SA channels.

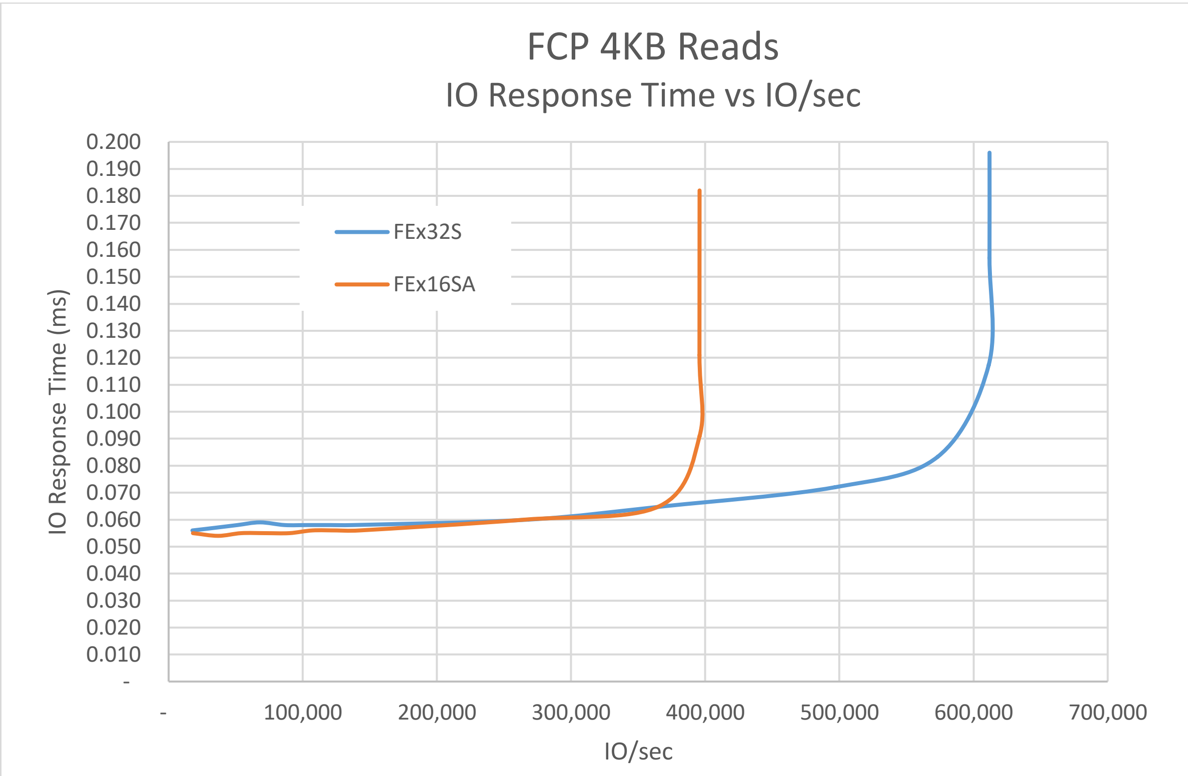


Figure 25 – FEx32S FCP Small Block Read I/O Response Time

FCP Large Block I/O Performance (MB/sec)

When handling a mixture of large 64KB Read/Write FCP I/O operations, a single FEx32S port can deliver up to 6,400 MB/sec throughput (Figure 26). Driving both ports available on the adapter card with the same mixture of Read/Write operations results in a total combined throughput of 9,700 MB/sec (Figure 27). This represents twice the full duplex throughput of a FEx16SA port, and a 54% increase over the total combined throughput of both FCP ports available on FEx16SA.

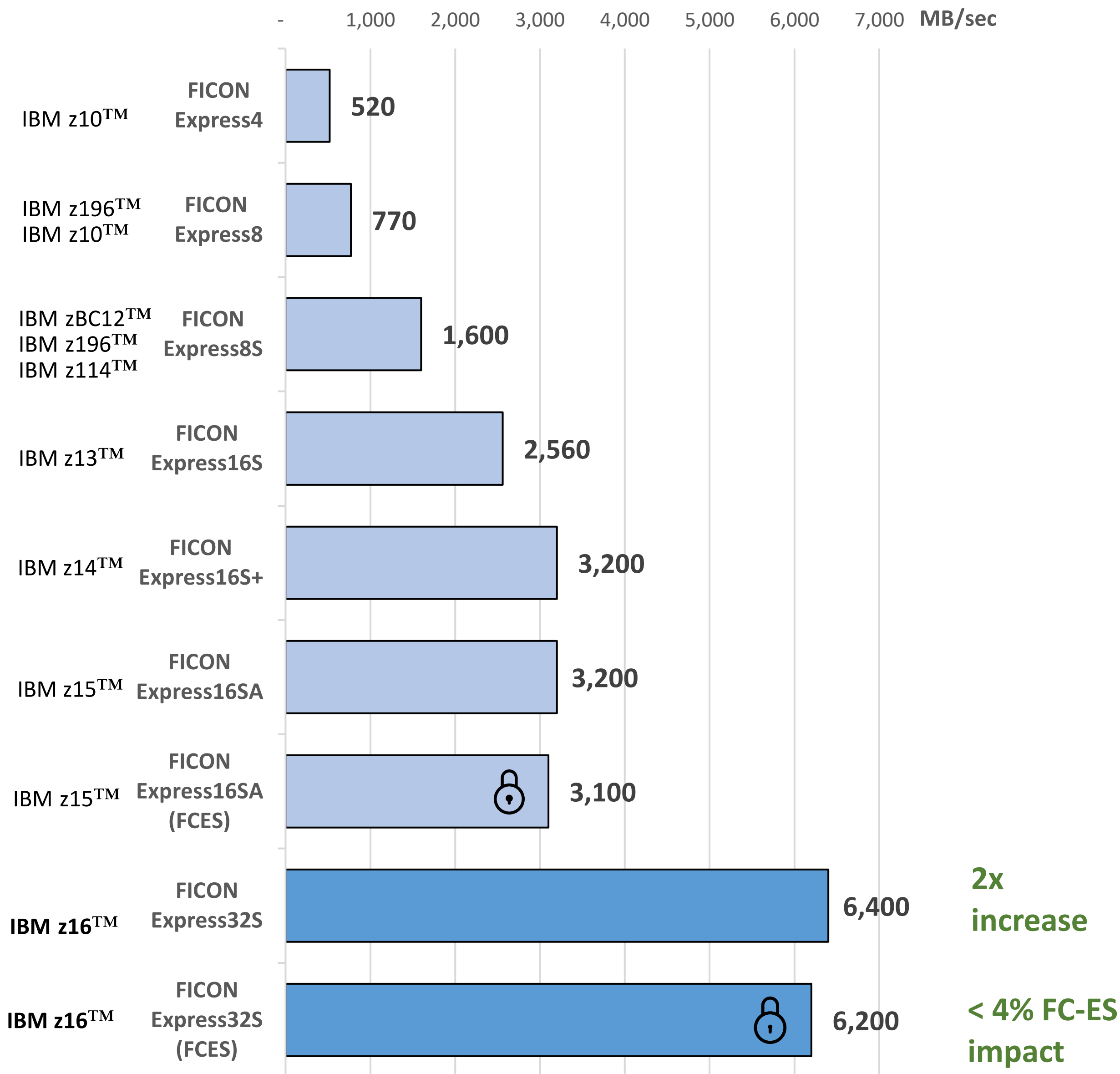


Figure 26 – Maximum FEx32S FCP Large Block Read/Write (mix) IO/sec (Per Port)

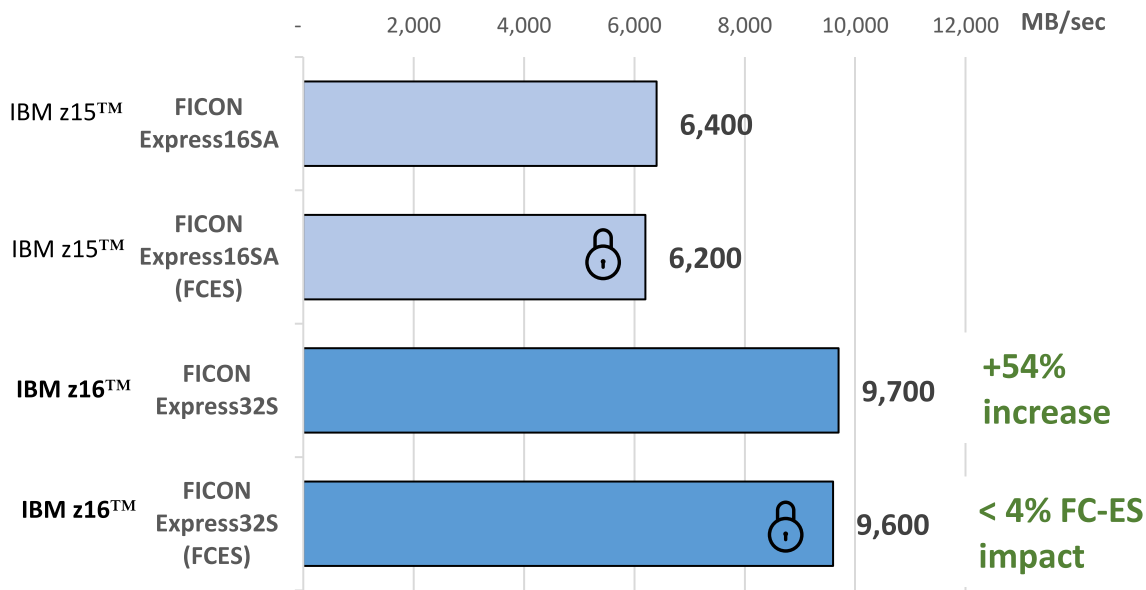


Figure 27 – Maximum FEx32S FCP Large Block Read/Write (mix) IO/sec (Combined Both Ports)

Bandwidth limits were also tested for Read-only and Write-only FCP operations, highlighting similar benefits over FEx16SA (Figure 28 & Figure 29). Overall, large block FCP I/O operations benefit the most from the increased 32Gb line rate offered by FEx32S, reaching full line speed when processing Read, Write or Read/Write (mix) operations on a single adapter card port, as well as when using both adapter ports with either Read-only or Write-only operations.

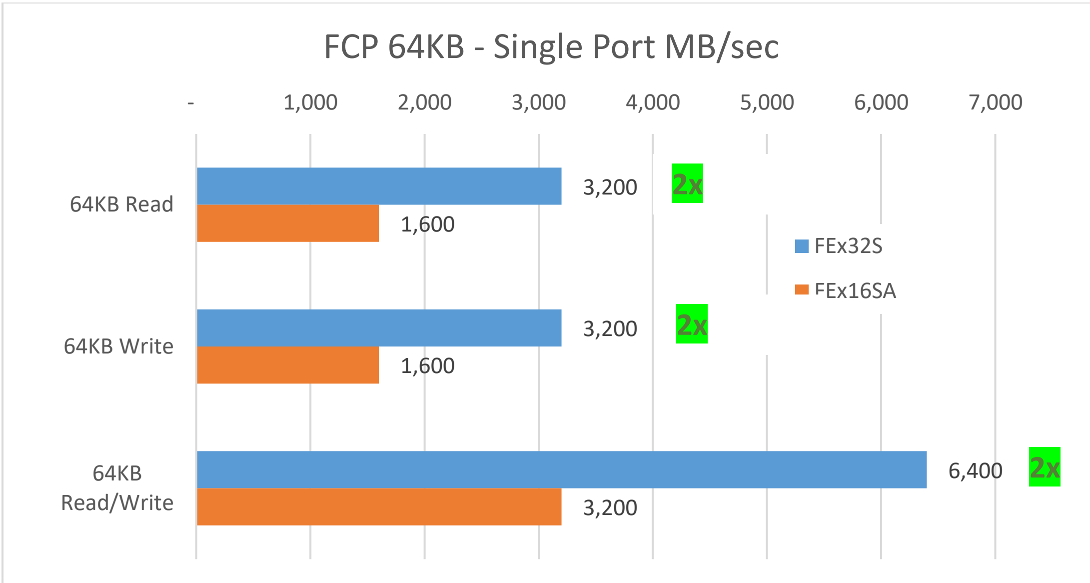


Figure 28 – FEx32S and FEx16SA FCP Large Block MB/sec (Single Port) Read, Write, Read/Write (mix)

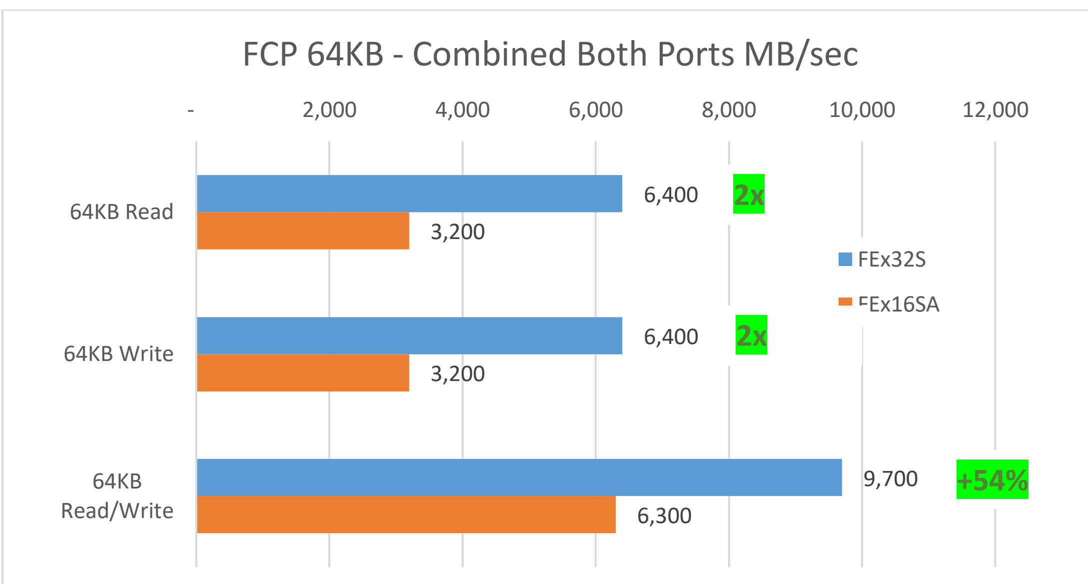


Figure 29 – FEx32S and FEx16SA FCP Large Block MB/sec (Combined Both Ports) Read, Write, Read/Write (mix)

Large block I/O response time is also improved compared to FEx16SA. Similar to the small block example provided on Figure 25, FEx32S can avoid the increased response time that FEx16SA channels experience when driving enough bandwidth to fill a 16G link (1,600 MB/sec in each direction). For example, as highlighted in Figure 30, FEx32S provides a gradual increase in response time for large block Read / Write (mix) operations until approximately 5,000 MB/sec :

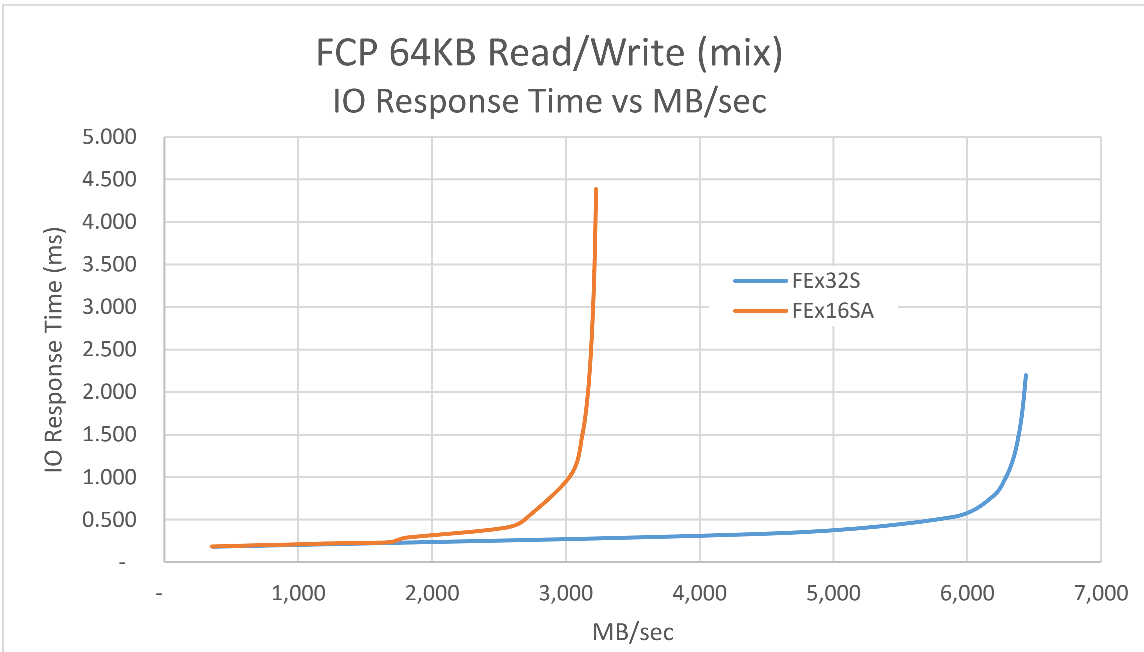


Figure 30 – FEx32S FCP Large Block Read/Write (mix) I/O Response Time

In summary, the latest FICON Express channel introduces significant performance improvements for FCP I/O on IBM z16™. Leveraging the faster link speed offered by FEx32S, workloads consisting of small block and large block I/O operations can benefit from both higher throughput and reduced response times versus previous generation FEx16SA.

FCP Performance with IBM Fibre Channel Endpoint Security

Based on laboratory measurements, IBM Fibre Channel Endpoint Security (FC-ES) enables encryption of data in-flight with minimal impact on performance. Running a suite of FCP Read-only, Write-only and Read/Write (mix) tests while increasing the number of concurrent operations, the FEx32S channel with FC-ES achieved a maximum throughput within 4% of the results produced without FC-ES. This was consistent for both 4KB and 64KB data block sizes tested. (Figure 31 & Figure 32.)

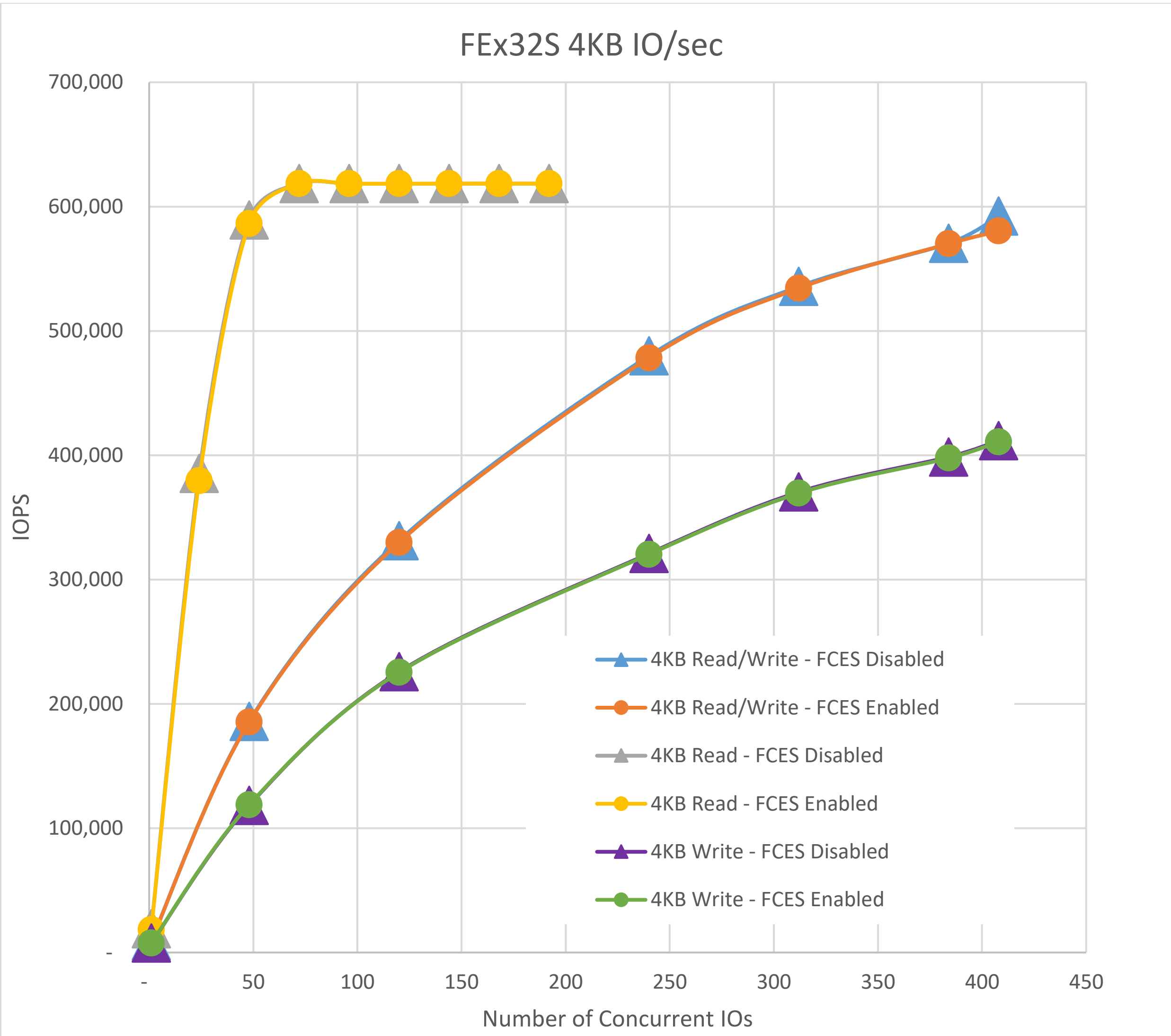


Figure 31– FCP Small Block Performance with FC-ES

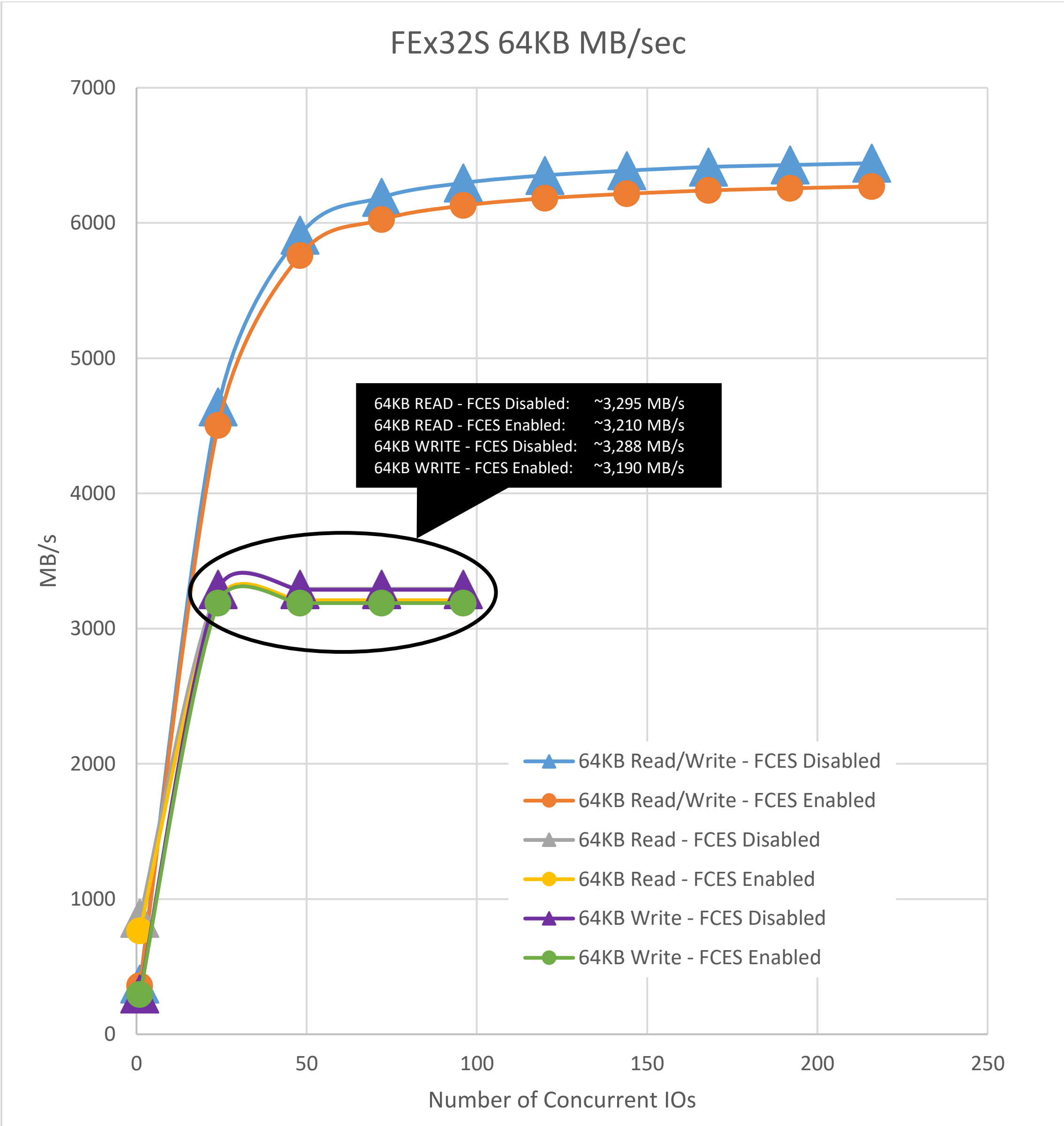


Figure 32 – FCP Large Block Performance with FC-ES

Furthermore, FC-ES processes encryption directly in the channel adapter HW, without using the LPAR CPU or IFL. Figure 33 highlights the average IFL utilization of a 12 IFL SLES 12 LPAR with SMT enabled, running a mixture of FCP 4KB Read/Write operations at up to 590,000 IO/sec throughput. Overall IFL busy was measured to be comparable when enabling and disabling FC-ES encryption, with less than 1% difference on average.

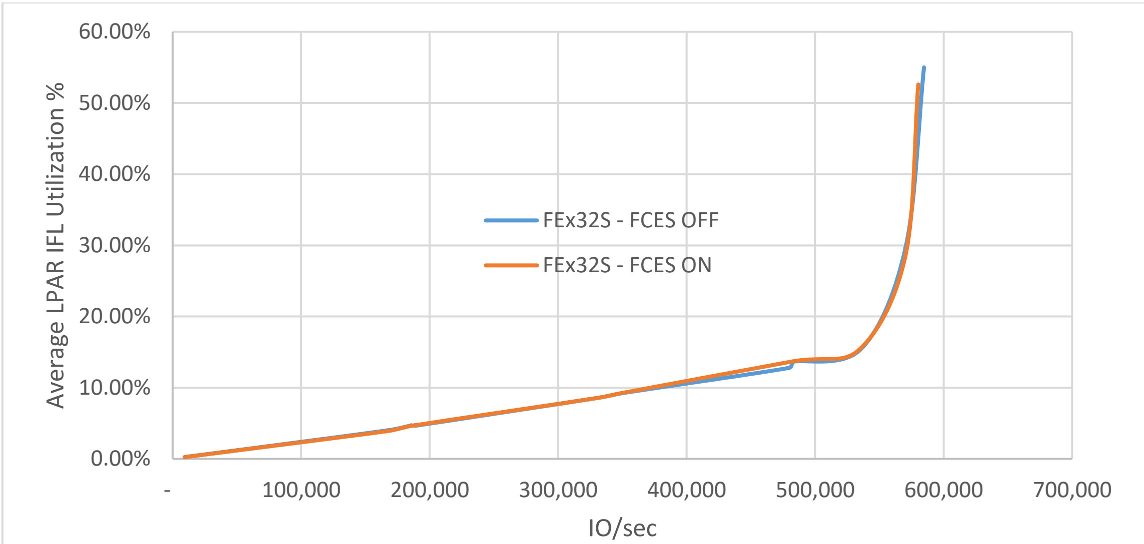


Figure 33 – FCP 4KB Read/Write
Average IFL Utilization (%) for a 12 Core LPAR vs IO/sec

Considering these results, we can conclude that FC-ES can offer additional data protection for FCP I/O transfers via encryption, with minimal impact on *both* FCP channel throughput and overall LPAR processor utilization.

FEx32S Card and I/O Subsystem (IOSS) Performance with zHPF

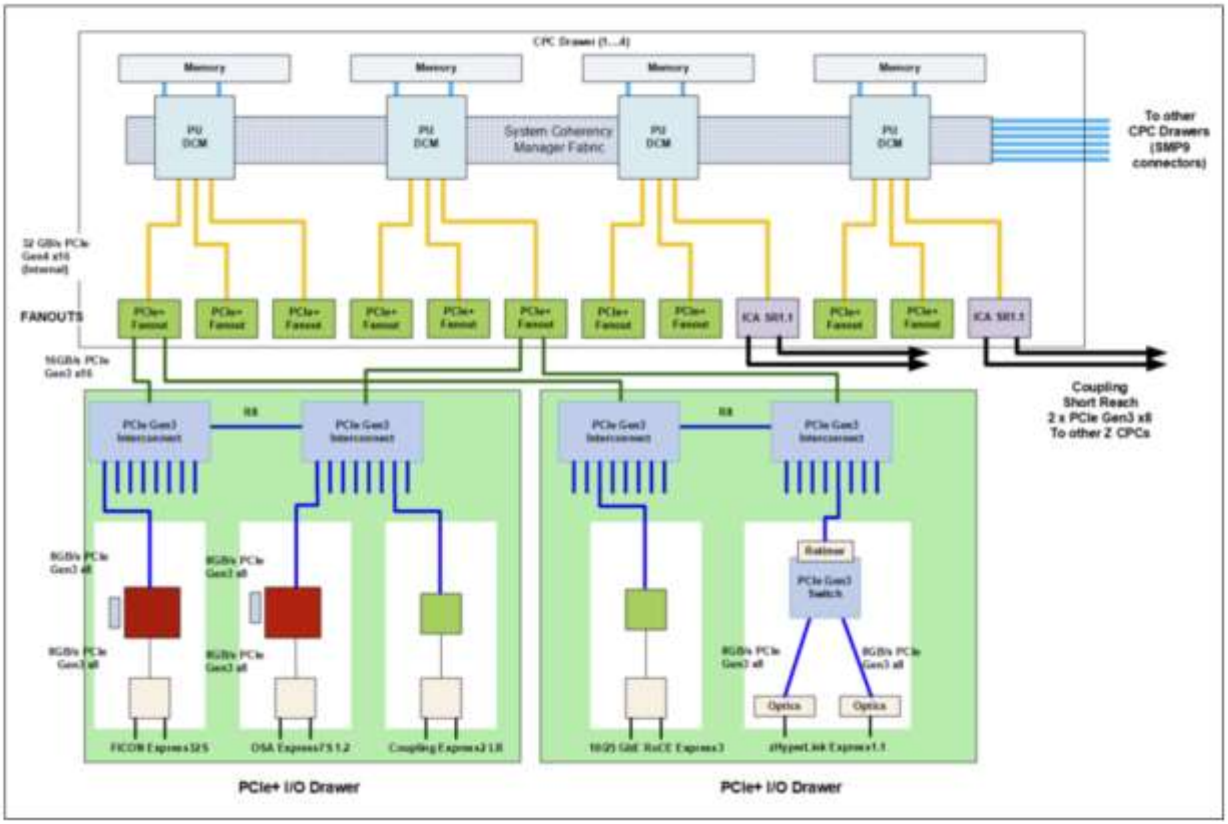


Figure 34 – IBM z16™ IOSS Diagram

The IBM z16™ supports the I/O drawer and form factor I/O cards which use Peripheral Component Interconnect Express Generation 3 (PCIe Gen3) links with increased capacity, granularity, and infrastructure bandwidth, as well as increased reliability, availability, and serviceability. The results of performance measurements done at each of the levels in this new I/O Infrastructure displayed in Figure 34 are summarized in this section.

At the card level, the new FEx32S channel card has two channels on the channel card.

Measurements were done with both FEx32S channels active Figure 35 on the same card to determine the maximum MB/sec capability of the FEx32S card with zHPF. When each I/O is transferring 128k bytes of data, the maximum MB/sec for two channels on a single card is approximately 6,300 MB/sec for READs, 5,480 MB/sec for update WRITEs and over 9,000 MB/sec for a mix or READ and WRITE I/O operations. These card measurement results are in the range of 1.6 to 1.9 times the maximum MB/sec of a single FEx32S channel using the zHPF protocol.

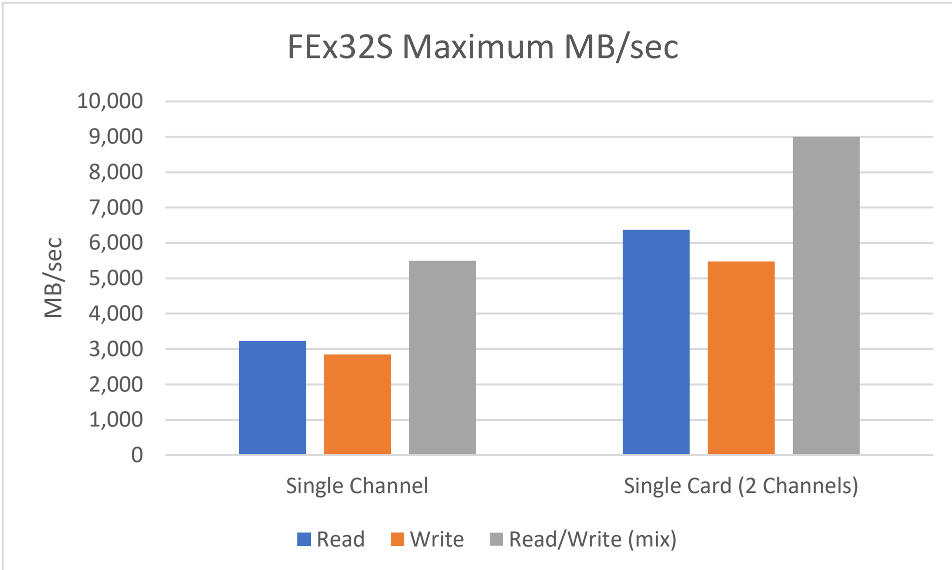


Figure 35 – FEx32S Channel vs Card Bandwidth

Figure 36 summarizes the expected I/O bandwidth available at each level of the new IBM z16™ I/O Infrastructure including a single FEx32S channel, a single FEx32S card with two FEx32S channels, single PCIe Interconnect domain with up to eight FEx32S channel cards and PCIe fanout card with two PCIe interconnects that allow connections to as many as 16 FEx32S channel cards.

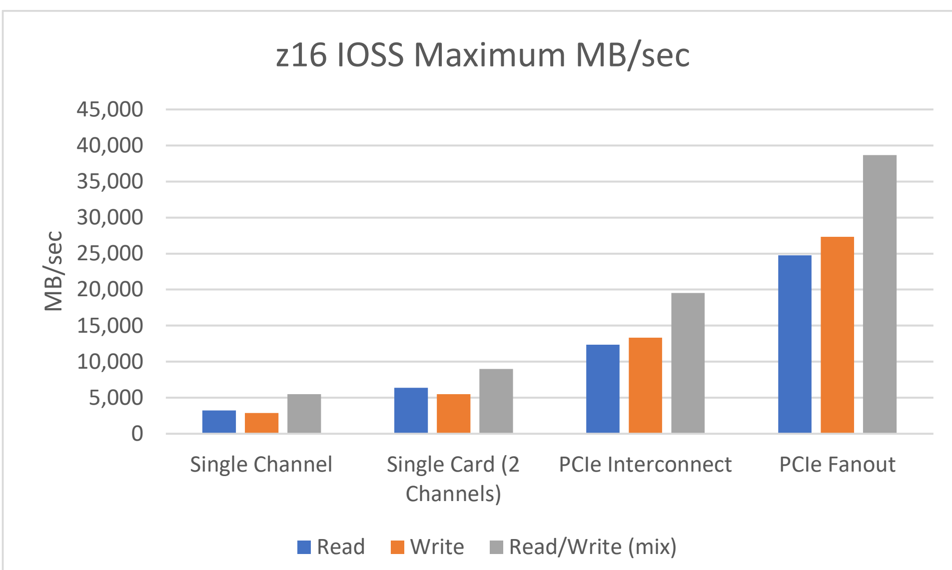


Figure 36 – IBM z16™ IOSS Bandwidth

The results achieved at the channel and card level of the I/O Infrastructure were each described in this document. For the PCIe interconnect level with up to eight FEx32S channel cards, approximately 12 GB/sec was measured for READs, 13 GB/sec for update WRITEs and over 19 GB/sec was measured for a mix or READ and WRITE I/O operations. For the PCIe fanout card with two PCIe Interconnects or up to sixteen FEx32S channel cards, approximately 24 GB/sec was measured for READs, 27 GB/sec for update WRITEs and over 38 GB/sec for a mix or READ and WRITE I/O operations.

The IBM z16™ supports the Redundant I/O Interconnect feature which is designed to help avoid unplanned outages by maintaining critical connections to I/O devices during path failures, or upgrades or repairs of a multi-drawer server. The IBM z16™ uses an alternate PCIe link which operates at the same speed as the primary PCIe link. The results achieved using an alternate path PCIe link into a single PCIe fanout card were similar to what was achieved using the primary PCIe links, namely 12 GB/sec READ or update WRITE.

This information about the maximum capabilities of various levels of the IBM z16™ IOSS is provided to help customers plan appropriately for high bandwidth demand workloads.

IBM Corporation
New Orchard Rd.
Armonk, NY 10504
U.S.A
Produced in the United States of America
5/2022