OPENFABRICS ALLIANCE

10TH ANNUAL INTERNATIONAL DEVELOPER WORKSHOP

iSER update

Eyal Salomon
Mellanox Technologies

2014 OFA Developer Workshop

Sunday, March 30 - Wednesday, April 2, 2014

Monterey CA

Agenda

- iSER Advantages and Status
- Support for new targets
  - LIO, SCST, TGT enhancements
- Initiator features and updates
  - Linux, VMware
  - Performance acceleration: iSER-BD and Block MultiQ
- iSER in OpenStack (Cinder)

iSER Advantages and Status

- Advantages
  - Enterprise Storage Solution (Security, HA, Discovery, ..)
  - OS tools and integration based on iSCSI
  - Ethernet (RoCE, iWarp) + InfiniBand
  - faster than SRP (on IOPs & Latency)
- Status and Adoption
  - In the kernel for few years, and constantly improving
  - Seamless integration into OpenStack (from Havana)
  - Adopted by many cloud providers
  - Will be available from multiple Tier1-2 storage vendors in the coming months

Storage Targets Support

	SRP	iSER	FCoE (Mlnx)
SCST (Kernel, external)	Yes	Yes	No
TGTD (User)	No	Yes	No
LIO (Kernel, inbox)	Yes	Yes	In-progress
GPL free, standalone reference (for storage OEMs and other OSes)	No, can base on SCST	Yes (can be obtained from Mellanox)	No, can base on LIO

LIO iSER

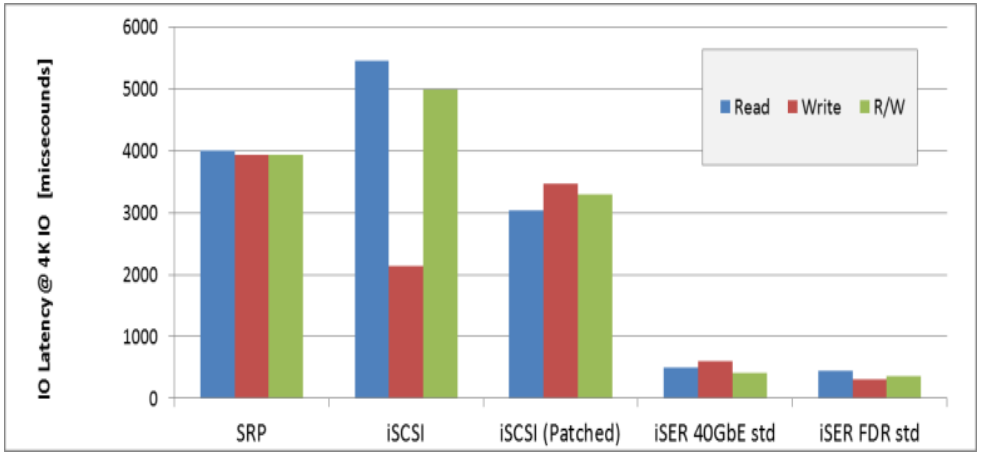
- iSER available from kernel 3.10
- T10-DIF
  - Implemented T10-DIF framework in target core
  - Implemented for iSER – fully offloading transport
    - Can do ADD/STRIP/PASS and verify of T10-DIF info
  - Implemented for File backing store
    - Can keep DIF interleaved in the data file (so one file has it all)
    - Can keep DIF in separate file (so data file stays clean with data only)
    - Good for development & debug (can inject errors…)
- working on performance optimizations

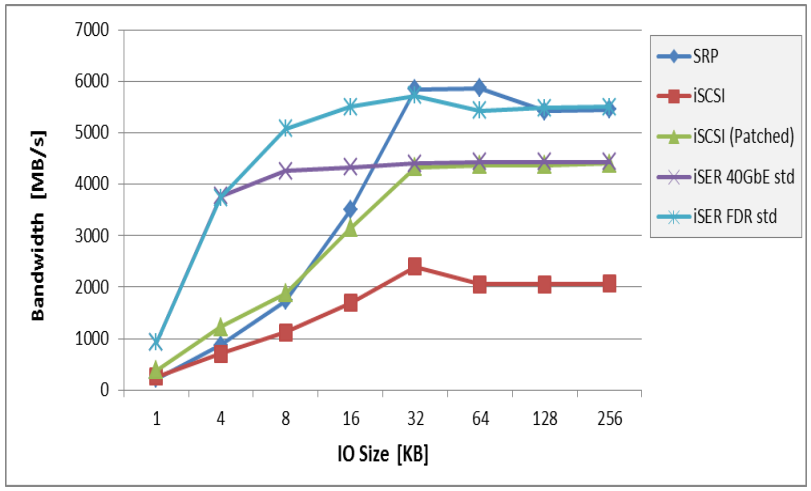
SCST

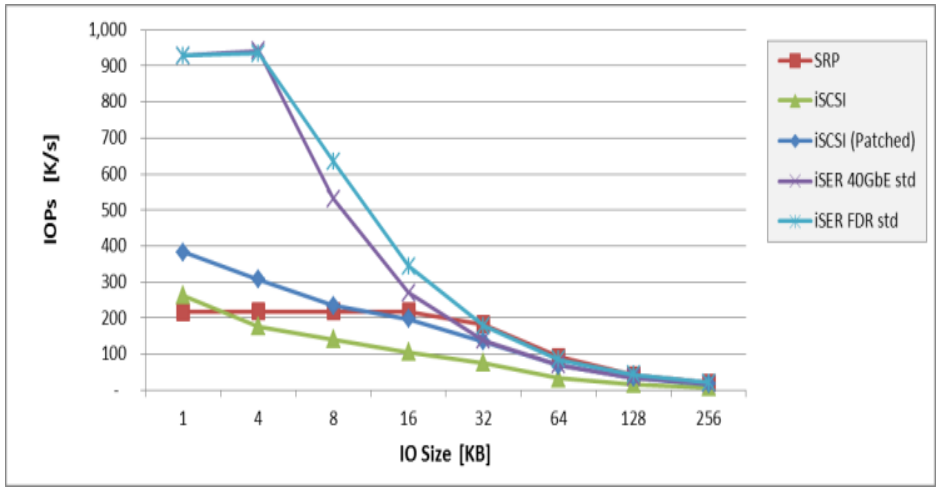
- SCST: A leading Linux target implementation
- iSCSI in SCST extended to support RDMA transport as well as TCP
- Transport API to abstract away from transport implementation
- High IOPs and bandwidth
- Supports Infiniband, RoCE and iWARP
- Available at http://scst.sourceforge.net/target_iser.html

Example Linux SCST Target Transport Comparison

Single Initiator to Single Target
Using HP DL380 with 2 x Intel 2650 CPUs
ConnectX-3 Adapter (40GbE or IB FDR)







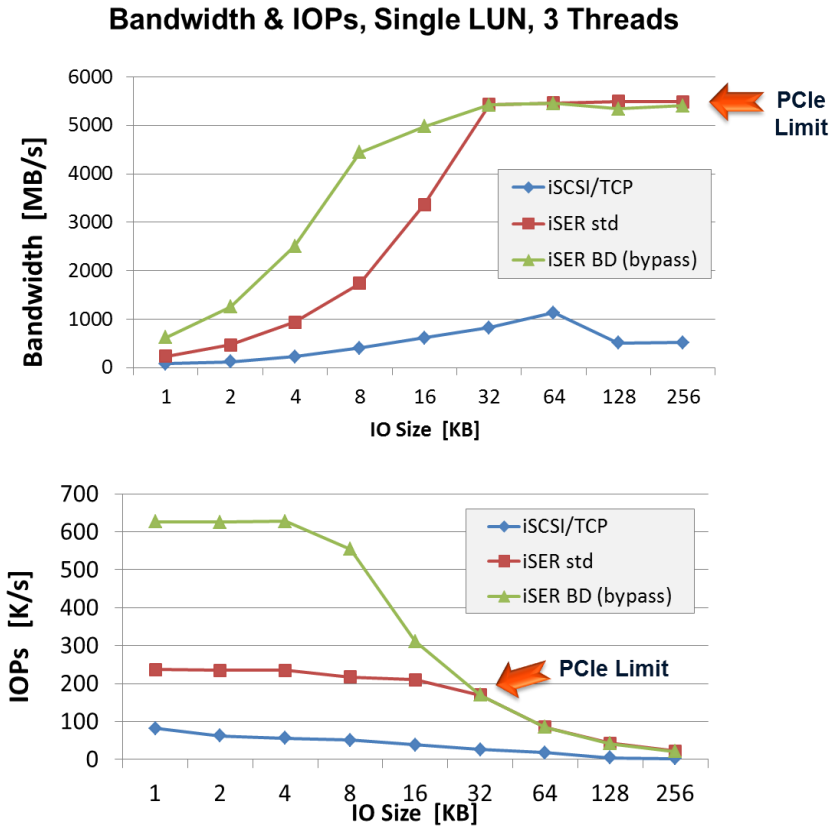
TGT iSER features

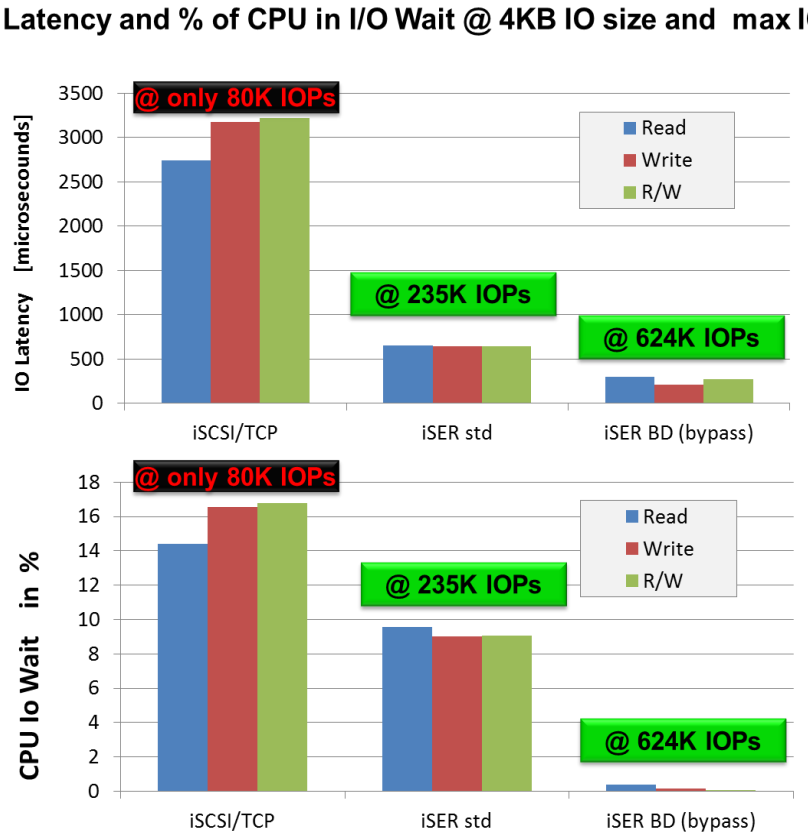
- sophisticated polling strategies to maximize IOPS
- support iSCSI Discovery over iser
- usage of Huge pages to minimize HCA cache misses
- support multiple instances for IOPS scalability
- support larger IO queue depth
- Configurable memory buffer
- Configurable CQ affinity

Initiator recent features

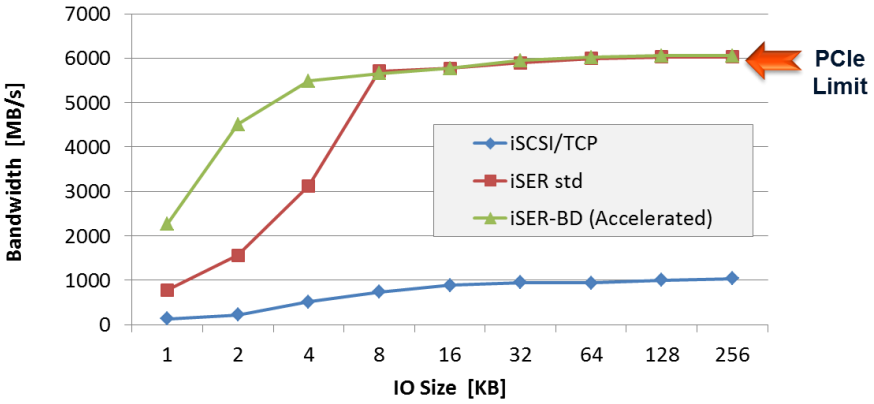
- Linux
  - ability to work in a VM over SRIOV VF
  - support fast registration (FRWR)
  - support iSCSI Discovery over iser
  - optimized fast-path locking scheme in libiscsi
  - support larger IO queue depth
  - support SCSI T10 signature/protection
  - misc stability fixes
- VMware ESX
  - iSER integrated into ESX 5.1/5.5 as an iSCSI HW offload
  - Work on native port to ESX 6 kernel

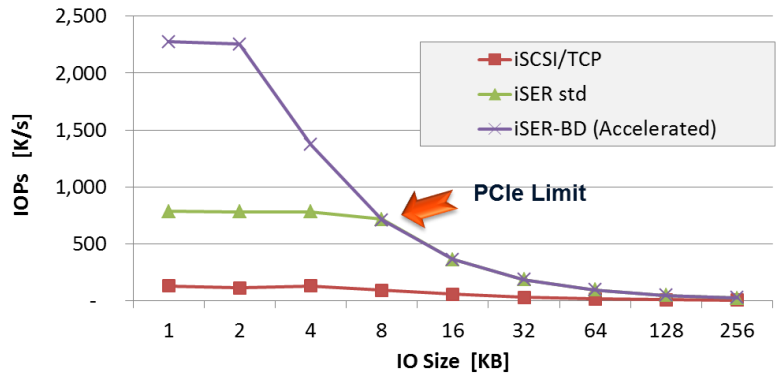
Accelerating IO Performance (Accessing a Single LUN)

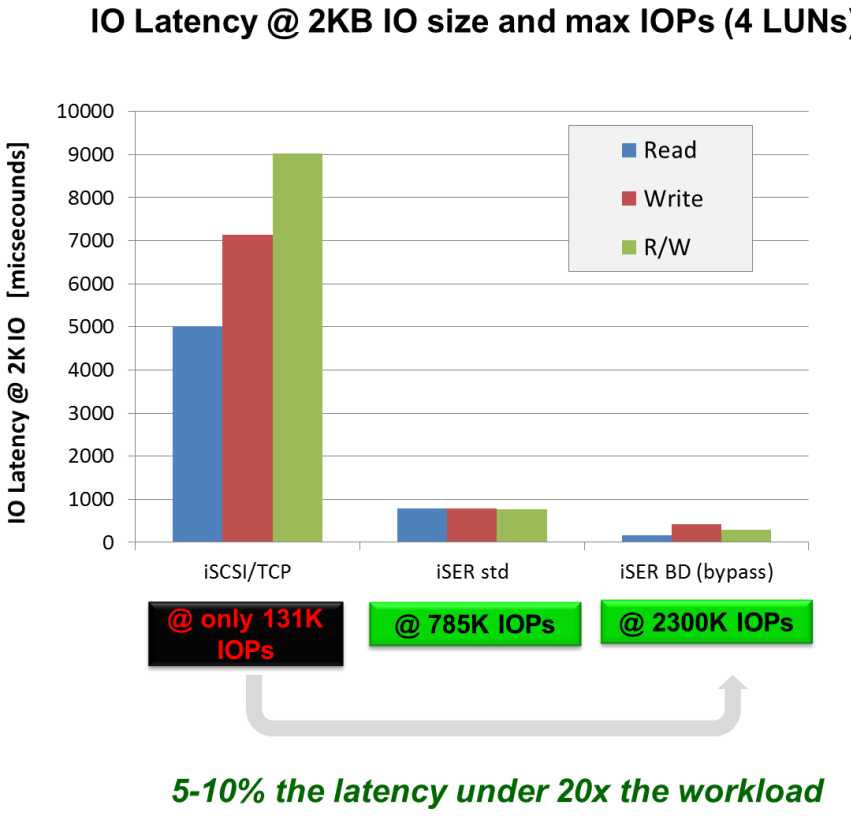




Accelerating IO Performance (Accessing 4 LUN in parallel)





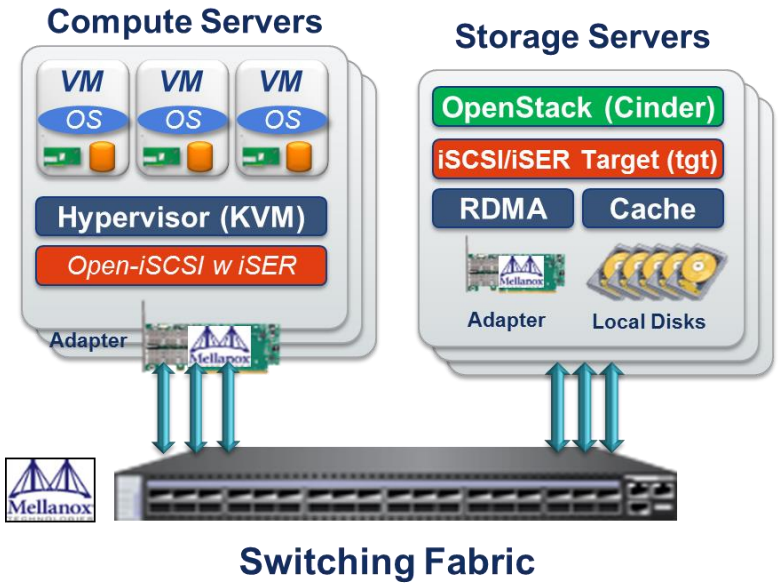


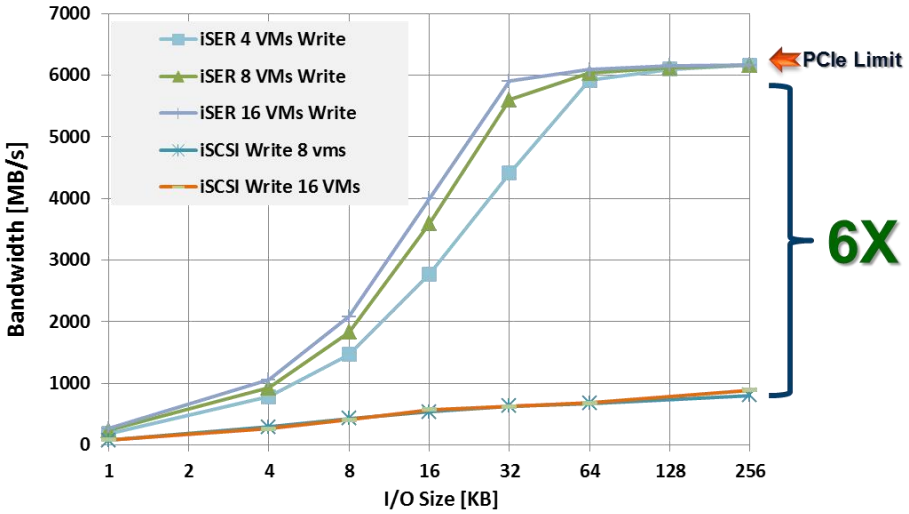
5-10% the latency under 20x the workload

MultiQ

- The iSER initiator exposes a SCSI Block device to the OS
- Both the Block and SCSI layers are performance bottlenecks e.g. for IOPS scalability
- in Linux 3.13 BLK-MQ was introduced to the block layer for generic block devices
- SCSI MQ is under the works by the community and iSER is planned to leverage on that
- The results in the previous slides were obtained by a prototype that emulated BLK/SCSI MQ

iSER in OpenStack





- Using OpenStack Built-in components and management (Open-iSCSI, tgt target, Cinder), no additional software is required, RDMA is already inbox and used by our OpenStack customers !
- Mellanox enable faster performance, with much lower CPU%
- Next step is to bypass Hypervisor layers, and add NAS & Object storage

Thank You

OPENFABRICS ALLIANCE

10TH ANNUAL INTERNATIONAL DEVELOPER WORKSHOP