Linux-iSCSI.org BoF

Current Status and Future of iSCSI on the Linux platform

Linux Plumbers Conference 2009

Nicholas A. Bellinger

nab@linux-iscsi.org

Overview

- ISCSI mini-overview
- ISCSI use cases today
- What is TCM/LIO..?
- Status update for TCM/LIO v3.2
- Future of iSCSI offload hardware
- Future of iSCSI on Linux
- Future of iSCSI on kernel.org
- Questions!? (please ask at any time)

What is iSCSI..?

- At it's core, the SCSI architecture model is a packet/message protocol designed to run on every possible physical link layer to every possible type of storage device.
- ISCSI is the method of layering the SCSI architecture and command set on top of TCP/IP.
- Originally used as an alternative to FC SANs, but increasingly being used as a commodity alternative for storage with Linux/x86 hosts
- Very popular on 1 Gb/sec ports, but 10 Gb/sec ethernet continues to push iSCSI in a number of ways

How is iSCSI being used..?

- Concentrating large amounts of raw storage into a powerful iSCSI Initiator machine
- Using iSCSI as a shared storage mechinism for different approaches to client side HA clustering aka 'iSCSI behind the cloud'
- Using iSCSI as a front-end protocol for different approaches to server side HA clustering aka 'iSCSI in front of the cloud'
- Using iSCSI as a storage mechinism for virtualization (On the VM host providing disk images on iSCSI LUNs or directly within the VM guest)

How is iSCSI being used..? Cont

- ISCSI for LAN replication
- iSCSI for WAN replication
- Using iSCSI for offsite backup
- iSCSI Boot using Pre Execution Environment (PXE)
- Watching commerical HD movies over the network

Current state of iSCSI on Linux

- All major distributions ship with Open-iSCSI initiator
- Many major distributions include STGT support for userspace iSCSI target mode, but is still unsupported in many commerical cases.
- Major distributors want to see proper in-kernel iSCSI target mode support, and the community offers out of tree modules for those interested
- Many developers, hardware vendors, and users want upstream kernel-level iSCSI target mode!

What is TCM/LIO..?

- TCM (target_core_mod) is the generic target infrastructure in lio-core-2.6.git on git.k.o
- LIO-Target (iscsi_target_mod) is the kernel-level iSCSI target fabric module that uses infrastructure provided by TCM for fabric indendent SCSI functionality and real-time control path using configfs
- LIO-Target provides exhaustive support for RFC-3720, including MC/S for bandwith trunking, and on-the-fly addition of new paths

TCM/LIO 3.2 update

- Since LPC 2008, there has been substainal work to turn the LIO-Target v2.9 codebase into the leading generic target engine on Linux.
- This includes previously un-heard of density for LUNs and iSCSI Target endpoints using upstream linux kernel infrastructure (configfs)
- Also includes SPC-4 cluster and multipath features never before available on OSS target mode Linux, and previously only available on the highest end of storage arrays and products

TCM/LIO v3.2 kernel update

- TCM v3.2 supports the complete subset of SPC-4 defined Persistent Reservation service actions and feature bits, including persistence across target power loss
- TCM v3.2 supports the complete subset of SPC-4 defined implict and explict Asymmetric Logical Unit Access logic for intelligent MPIO
- First open OR closed target mode to push 10 Gb/sec line rate to a single iSCSI Logical Unit using IOV capable hardware within Linux/KVM

TCM/LIO v3.2 kernel, cont

- Generic Target Engine (TCM) submitted for review during v2.6.32 merge window.
- Linux-iSCSI.org fabric module (LIO-Target) is being submitted seperately at a later date..
- Chicken and Egg problem: Can't merge a kernel-level iSCSI Target until the proper kernel-level SCSI target mode infrastructure exists upstream. So the iSCSI target code stays in lio-core-2.6.git, for now...

TCM/LIO v3.2 userspace

- ConfigFS interface between generic target core and Linux-iSCSI.org target now capable of creating, saving and restarting 10,000 unique HBA+LUNs with unique iSCSI Target Endpoints
- Python based CLI API (tcm_node.py and lio_node.py) in lio-utils.git for developers, integrators and advanced users.
- The CLI API exposes the complete set of functionality available from TCM/LIO v3.2 code, but a higher level interactive shell is next step..

Future of iSCSI offload hardware

- ISCSI + TOE design is dying, dying dead.
- Bugfixing Linux/Net is easier than bugfixing TCP/IP in silicon. (Potential security hole waiting to happen with TOE)
- Linux/Net folks (eg: DaveM) dislike TOE because it requires invasive changes to the Linux/Net stack.
- Stateless TCP/IP offloads with 10 Gb/sec hardware have proven to scale BETTER than TOE, and do not break the Linux/Net stack!

Future of iSCSI offload hardware

- Patent filed in July for hybrid iSCSI offload HW engine using stateless TCP offload by market leading 10 Gb/sec Ethernet adapter vendor.
- Allows for a per iSCSI PDU context offload depending on PDU layout in TCP segment
- Allows for Direct Data Placement on RX side for traditional iSCSI coming into TCM as pre-registered struct scatterlist for fast path
- Allows for existing iSCSI software to handle multiple PDUs within the same TCP segment.

Future of iSCSI offload hardware

- The hybrid iSCSI offload + TCP stateless offload design allows high bandwidth on the order of 1000s of MB/sec without breaking the Linux/Net or Linux/iSCSI software base!
- This design combined with IO virtualization will drive cost savings for IP storage on 10 Gb/sec Ethernet, with up to 256 virtual functions per adapter using Alternative Requester ID (ARI)
- Support for hybrid iSCSI offload in LIO-Target in v3.4 or v3.5 time frame (end of 2010)

Future of iSCSI on Linux

- Linux iSCSI target mode for upstream needs to lead and not follow
- Linux/iSCSI needs to support end-to-end data integrity using T10 DIF for iSCSI gateways to existing SAS and FC HBAs and drives.
- Linux/iSCSI needs to be ahead of the curve so once drive vendors support DIF on SATA drives it will be available out of the box for Linux <-> Linux iSCSI fabric setups.

Future of iSCSI on Linux, cont

- Provide upstream kernel level iSCSI code that can scale into 10,000s of LUNs+Endpoints and supports independent real-time configuration
- ISCSI Target independent cluster resource agent for Pacemaker work done by Florian Haas of Linbit
- Provide a interactive shell for day-to-day for the higher end critical use cases, but still make it easy enough for a every Linux admin for simple use cases.

The future of iSCSI (service) from boot.kernel.org

- boot.kernel.org announced on monday by John Hawley and the kernel.org team
- Currently offers boot services using gpxe for Linux via httpfs.
- Will also be offering the first public iSCSI targets for accessing Linux .isos over the internet using kernel.org infrastructure!
- This means that every Linux toaster (including you and your grandma's) will be using boot.kernel.org at some point.

Questions and Thank you!

- Linux/SCSI, Linux/iSCSI and open source target mode communities (Doug Gilbert, jejb, mnc, Tomo-san, Dr. Hannes Reincke, Boaz Harrosh, Ming Zhang, Richard Sharpe for presenting at LinuxCon, and many more)
- kernel.org team (hpa and warthog9)
- Upstream Linux Kernel team and Linux Foundation
- Linbit (Phillip, Lars and Florian)
- Neterion (Leonid Grossman)
- Rising Tide Systems (Marc Fleischmann)