OpenSM and InfiniBand Management Update

OPENFABRICS ALLIANCE

Sasha Khapyorsky

Agenda

- General
- New
  - OFED 1.3
  - The current
- Plans
  - OFED 1.4
  - Long terms
- Wishes
- Thanks

General

- Git tree
  - git://git.openfabrics.org/~sashak/management
- Downloads
  - http://www.openfabrics.org/downloads/management/
- OFED-1.3:
  - libibcommon-1.0.8, libibumad-1.1.7, libibmad-1.1.6, OpenSM-3.1.10, infiniband-diags-1.3.6, ibsim-0.4+
- The current (master):
  - libibcommon-1.1.0, libibumad-1.2.0, libibmad-1.2.0, OpenSM-3.2.1, infiniband-diags-1.4.0

OFED-1.3: core libraries

- libibumad-1.1.6
  - Partition support
    - Pkey index support in user MAD kernel module
  - Multiple open()s
  - Thread safe now
  - Many fixes
- libibmad-1.1.7
  - GRH encoding support
  - IB_DEVICE_MGMT_CLASS support

OFED-1.3: OpenSM-3.1.10

- QoS manager (experimental)
  - Based on IBTA annex
  - Highly configurable QoS policy definition
  - SA PR/MPR queries
  - Simplified QoS per ULPs configuration
  - Port parameters (SL2VL, VLArb) setup is needed
- Performance manager (experimental)
  - Able to work when SM is not in a master state
  - Event plugin interface

OFED-1.3: OpenSM-3.1.10

- Routing
  - Dimension Order Routing (DOR) – contributed by SGI – useful with hypercube and mesh topologies
  - Performance improvements for existing routing engines: Min hops, Up/Down, LASH
  - FatTree routing algorithm can be used with not pure FatTree topologies
- Multicast re-routing speedup
  - Process all pended multicast requests
- Support for non-local IPoIB scopes

OFED-1.3: OpenSM-3.1.10

- Per subnet prefix routing – contributed by Obsidian
  - Router per subnet prefix configuration file
  - SA PathRecord query returns path to a router for non-local DGIDs
- Incremental port (pkey, sl2vl, vlarb) and switch (LFTs) tables update
  - Fetch and update only when needed and only modified blocks
- Fast port and switch reset detection

OFED-1.3: OpenSM-3.1.10

- Force links speed option
  - Setup ports' enabled link speed to defined value
- Duplicate GUIDs/port moving detection
- Handle “babbling” ports
  - Suppress trap storms
- Node name map
  - Internal node description redefinition for logging
  - Using defined node names in configurations (currently QoS manager only)

OFED-1.3: OpenSM-3.1.10

- Native daemon mode (--daemon option)
- More console commands
- Improved build and packaging
- Other improvements and many fixes

OFED-1.3: infiniband-diags-1.3.6

- Saquery – more queries
- Ibnetdiscover – link list (--ports option)
- Most utils work with any CA/port
- Node name map support for additional diags
- set_nodedesc.sh instead of set_mthca_nodedesc.sh

OFED-1.3: ibsim-0.4

- New package – not part of the management
  - git://git.openfabrics.org/~sashak/ibsim.git
  - http://www.openfabrics.org/downloads/management
- Infiniband user MAD layer simulator
  - Redefines system calls open(), read(), write(), etc. (using LD_PRELOAD trick)
  - Libibumad based tools (OpenSM, infiniband-diags, etc.) work unmodified
  - Uses ibnetdiscover output as fabric description
  - Has simple command console interface

The Current: OpenSM-3.2.1

- IPv6 SNM (Solocited Node Multicast) MGIDs consolidation
- Paths balancing for LMC > 0
- Node IDs configuration for Up/Down routing
- Many code cleanups

Plans

- OpenSM configuration unification
- MGIDs compression
  - An option to map many MGIDs to single MLID
- Routing engines chain
  - Opensm -R ftree -R updn -R minhops
- Failover/Handover improvements
  - Tests and fixes
  - Query standby SMs (sminfo) during light sweep
  - Trap 144 on SM priority change

Plans:

- Incremental routing
- Shadow SA DB
  - Keep SA DB copy in order
  - More granular SM/SA DB update locking
- APM disjoin
- MKey management
- QoS manager integrations
  - Share port group definition with Partition Manager
  - Low level QoS parameters: SL2VL, VLArb, etc

Plans:

- Secure console
- IBA 1.2.1 additions
  - unpath/repath support
  - SL2VL setup optimization
  - Traps changes
- Ibnetdiscover library
  - Should speedup other diag tools dramatically
- OpenSM diagnostic capabilities

Plans: Long Term

- SM/SA DB replication
- Scalability
  - SA scalability
  - Distributed SA
- IB-to-IB routers support (as will be released by IBTA)
- Congestion management

Wishes

- More collaboration is needed
  - OpenSM usability is increased
  - More feedback, testing, development are desired
- Thoughts, Ideas?
  - general@lists.openfabrics.org

Thanks

- git log ofed_1_2..master | grep Author | sort -u

Albert L. Chu
Arne Redlich
Bernd Schubert
Dale Purdy
David A. McMillen
Dotan Barak
Doug Ledford
Eitan Zahavi
Erez Strauss
Hal Rosenstock
Ira K. Weiny
Jeff Becker
Jeremy Brown
Michael S. Tsirkin
Philippe Gregoire
Roland Dreier
Rolf Manderscheid
Sasha Khapyorsky
Sean Hefty
Timothy A. Meier
Todd Rimmer
Yevgeny Kliteynik

Thank You