



RDMA Stacks

MOFED, OFED & Linux Kernel

Fernando Garcia (fg@gentwo.org)
and
Christoph Lameter (cl@gentwo.org)

RDMA libraries and stacks



Problem: Numerous RDMA / IB stacks from multiple sources with various features

What does one use for a production environment? We have a custom built kernel on top of Ubuntu 10.04.

One particular issue were the kernel modules and therefore Linux kernel dependencies. With that came fragility and lack of integration into Linux distributions.

What we need:

1. Extreme low latency
2. Multicast issues
3. Ethernet support

RDMA stacks



<i>Stack</i>	<i>Versions</i>	<i>Characteristic</i>
<i>Linux Kernel</i> kernel.org	<i>2.6.32-3.15</i>	<i>No external module build</i>
<i>OFED</i> openfabrics.org	<i>1.1.x-1.5.x</i> <i>3.2/3.5/3.12</i>	<i>OpenFabrics Releases</i>
<i>MLNX_OFED</i> mellanox.com	<i>1.4.X/1.5.X</i> <i>2.0/2.1</i>	<i>Mellanox enhanced OFED releases</i>
<i>OFED-MIC</i> openfabrics.org	<i>3.5</i>	<i>Intel Xeon Phi Ofed stack</i>
<i>OFED-VMA</i>	<i>1.5.X?</i>	<i>Early flow steering implementation</i>

Non Linux OFED stacks



- Windows “WinOFED”
 - AIX 7.1 OFED
 - Solaris/Illumos OFED
 - FreeBSD OFED
-

Linux Distributions supported



RHEL 6.X	Needs OFED/MOFED
RHEL 7 beta	Works out of box
Ubuntu 10.04	No OFED support, custom hacks
Ubuntu 12.04	OFED/MOFED available
Ubuntu 14.04	Works out of box
SLES11/SLES10	OFED/MOFED required
OLE	Not dealt with it
Debian	Usually requires customization work

History with RDMA Stacks



- 2008 SDR/DDR OFED 1.2/1.3 with VMA
- 2009 DDR with IPoIB OFED-1.4
- 2010 QDR OFED-1.5.X -> RDMA apps
- 2014 QDR Linux Kernel 3.14 IB stack

OFED API breakage



Binaries built against OFED 1.X break with strange errors when run with OFED 3.X or the Linux IB stack

Mismatch in data structures. Checks on symbol versioning of the linker do not trigger.

Other issues with header changes but those are to be expected when new features are introduced:

Ethernet support changes of MOFED/OFED vs Linux IB.

Flow steering APIs vs. earlier hacks

How to deploy an upstream RDMA stack



3.12 kernel with extra patches or vanilla 3.14

<https://www.kernel.org/pub/scm/libs/infiniband/>

<https://git.kernel.org/cgit/libs/infiniband/>

libibverbs verbs extensions patches

Missing send flags

Device controlled flow steering

`log_num_mgm_entry_size=-1`



Questions?
Suggestions?



OPENFABRICS
ALLIANCE
