



OPENFABRICS
ALLIANCE

14th ANNUAL WORKSHOP 2018

OPENFABRICS VERBS MULTI-VENDOR SUPPORT

Brian Hausauer, Intel
Saqib Jang, Chelsio
Nishant Lodha, Cavium

April 12, 2018



NOTICES AND DISCLAIMERS

INFORMATION IN THIS DOCUMENT IS PROVIDED IN CONNECTION WITH INTEL® PRODUCTS. NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT. EXCEPT AS PROVIDED IN INTEL'S TERMS AND CONDITIONS OF SALE FOR SUCH PRODUCTS, INTEL ASSUMES NO LIABILITY WHATSOEVER, AND INTEL DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO SALE AND/OR USE OF INTEL® PRODUCTS INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT, COPYRIGHT OR OTHER INTELLECTUAL PROPERTY RIGHT. INTEL PRODUCTS ARE NOT INTENDED FOR USE IN MEDICAL, LIFE SAVING, OR LIFE SUSTAINING APPLICATIONS.

Intel may make changes to specifications and product descriptions at any time, without notice.

All products, dates, and figures specified are preliminary based on current expectations, and are subject to change without notice.

Intel processors, chipsets, and desktop boards may contain design defects or errors known as errata, which may cause the product to deviate from published specifications. Current characterized errata are available on request.

Intel product plans in this presentation do not constitute Intel plan of record product roadmaps. Please contact your Intel representative to obtain Intel's current plan of record product roadmaps.

Intel, Xeon the Intel logo are trademarks of Intel Corporation in the United States and other countries.

*Other names and brands may be claimed as the property of others.

Copyright ©2018 Intel Corporation.

Information in this document does not constitute plan of record for Intel, Cavium, or Chelsio products. Please contact your Intel, Cavium, or Chelsio representative to obtain current plan of record information.

OVERVIEW

- **Over the years OpenFabrics Verbs (OFV) has grown beyond a “classic” RDMA API**
 - Now supports non-RDMA networking features
 - And some vendor-specific features

- **This session promotes *multi-vendor OFV*, defined as the subset of Verbs supported by a majority of surveyed NIC vendors**
 - Survey currently covers a subset of Ethernet NIC vendors supporting iWARP and/or RoCE

- **Goals**
 - Make it transparent and easy for RDMA developers to know which verbs can be used across a wide set of NICs
 - Drive an industry conversation on how/when/if to expand *multi-vendor OFV* definition

- **Non-goal**
 - Eliminate NIC innovation or differentiation

DETAILS ON SURVEYED NICs



- Intel® Ethernet Connection X722 with iWARP RDMA



- 41000 Series 10/25GbE NICs with Universal RDMA (iWARP and RoCE/v2)
- 45000 Series 40/50/100GbE NICs with Universal RDMA (iWARP and RoCE/v2)



- Terminator 6 (T6) 1/10/25/40/50/100GbE Offload Adapters with iWARP
- Terminator 5 (T5) 1/10/40/GbE Offload Adapters with iWARP

DETAIL ON TABLES

- The following tables describe the OFV function calls included in *multi-vendor OFV*
- **Important:** Each NIC vendor supports *more* verbs than shown in the tables! Tables show only the subset of verbs supported by a majority of surveyed NIC vendors.

Table Column Descriptions

These parameters are excluded from the multi-vendor verb. This is loosely defined, possibly at the sub-struct or feature level.

Exceptions from full support across all surveyed NICs. Full support is defined to exclude any Verb parameters that are not multi-vendor

Multi-vendor Verb name

Verb	excluded parameters	NIC Support Exceptions
ibv_resize_cq		x722,4x000 : future sw release, t5t6 : currently unsupported
ibv_post_send	atomics, xrc, tso	

Examples

UVERBS: DEVICE, PD

Verb	excluded parameters	NIC Support Exceptions
ibv_get_device_list		
ibv_free_device_list		
ibv_open_device		
ibv_close_device		
ibv_get_device_name		
ibv_get_device_guid		
ibv_query_device		
ibv_query_port		
ibv_query_gid		
ibv_query_pkey		
rdma_get_src_port		
rdma_get_dst_port		
rdma_get_local_addr		
rdma_get_peer_addr		
rdma_get_devices		
rdma_free_devices		
rdma_getaddrinfo		
rdma_freeaddrinfo		
ibv_alloc_pd		
ibv_dealloc_pd		

UVERBS: COMPLETION, EVENT

Verb	excluded parameters	NIC Support Exceptions
ibv_get_async_event		
ibv_ack_async_event		
ibv_create_comp_channel		
ibv_destroy_comp_channel		
ibv_create_cq		
ibv_resize_cq		x722,4x000 : future sw release, t5,t6 : currently unsupported
ibv_destroy_cq		
ibv_get_cq_event		
ibv_ack_cq_events		
ibv_req_notify_cq		
ibv_poll_cq		
rdma_create_event_channel		
rdma_destroy_event_channel		
rdma_get_cm_event		
rdma_ack_cm_event		

UVERBS: QP, MEMORY

Verb	excluded parameters	NIC Support Exceptions
ibv_create_qp		
ibv_destroy_qp		
ibv_modify_qp		
ibv_query_qp		
ibv_post_send	atomics, xrc, tso	
ibv_post_recv		
rdma_create_qp		
rdma_destroy_qp		
ibv_reg_mr		
ibv_dereg_mr		
ibv_alloc_mw		x722,4x000,t5,t6 : future sw release
ibv_dealloc_mw		x722,4x000,t5,t6 : future sw release
ibv_bind_mw		x722,4x000,t5,t6 : future sw release
ibv_inc_rkey		t5,t6 : future sw release

UVERBS: SRQ, CONNECTION MANAGEMENT

Verb	excluded parameters	NIC Support Exceptions
ibv_create_srq		x722,t5 : currently unsupported
ibv_modify_srq		x722,t5 : currently unsupported
ibv_query_srq		x722,t5 : currently unsupported
ibv_get_srq_num		x722,t5 : currently unsupported
ibv_destroy_srq		x722,t5 : currently unsupported
ibv_post_srq_recv		x722,t5 : currently unsupported
rdma_create_id		
rdma_destroy_id		
rdma_migrate_id		
rdma_set_option		
rdma_bind_addr		
rdma_resolve_addr		
rdma_resolve_route		
rdma_connect		
rdma_listen		
rdma_get_request		
rdma_accept		
rdma_reject		
rdma_disconnect		
rdma_event_str		

KVERBS: DEVICE

Verb	excluded parameters	NIC Support Exceptions
ib_query_port		
rdma_port_get_link_layer		
rdma_start_port		
rdma_end_port		
rdma_is_port_valid		
rdma_protocol_ib		
rdma_protocol_roce		
rdma_protocol_roce_udp_encap		
rdma_protocol_roce_eth_encap		
rdma_protocol_iwarp		
rdma_ib_or_roce		
rdma_protocol_raw_packet		
rdma_protocol_usnic		
rdma_cap_ib_mad		
rdma_cap_opa_mad		
rdma_cap_ib_smi		
rdma_cap_ib_cm		
rdma_cap_iw_cm		
rdma_cap_ib_sa		
rdma_cap_ib_multicast		

KVERBS: DEVICE, PD

Verb	excluded parameters	NIC Support Exceptions
rdma_cap_af_ib		
rdma_cap_eth_ah		
rdma_cap_opa_ah		
rdma_max_mad_size		
rdma_cap_roce_gid_table		
rdma_cap_read_inv		
ib_query_pkey		x722,4x000,t5,t6 : na for iwarp
ib_find_pkey		x722,4x000,t5,t6 : na for iwarp
ib_get_rdma_header_version		
ib_get_vector_affinity		
ib_query_gid		
ib_find_gid		
rdma_get_gids_from_rdma_hdr		
rdma_cap_ib_switch		
ib_alloc_pd		
ib_dealloc_pd		

KVERBS: COMPLETION, EVENT, QP

Verb	excluded parameters	NIC Support Exceptions
ib_register_event_handler		
ib_unregister_event_handler		
ib_dispatch_event		
ib_alloc_cq		
ib_create_cq		
ib_resize_cq		x722 : future sw release, t5,t6 : currently unsupported
ib_destroy_cq		
ib_poll_cq		
ib_req_notify_cq		
ib_create_qp		
ib_modify_qp_is_ok		
ib_modify_qp		
ib_query_qp		
ib_destroy_qp		

KVERBS: QP

Verb	excluded parameters	NIC Support Exceptions
ib_post_send	atomics, lso, sig_mr	
ib_post_recv		
ib_drain_rq		
ib_drain_sq		
ib_drain_qp		
rdma_create_qp		
rdma_destroy_qp		
rdma_rw_ctx_init		
rdma_rw_ctx_destroy		
rdma_rw_ctx_signature_init		
rdma_rw_ctx_destroy_signature		
rdma_rw_ctx_wrs		
rdma_rw_mr_factor		
rdma_rw_init_qp		
rdma_rw_init_mrs		
rdma_rw_cleanup_mrs		
ib_mr_pool_get		
ib_mr_pool_init		

KVERBS: MEMORY, SRQ

Verb	excluded parameters	NIC Support Exceptions
ib_alloc_mr	signature, gaps MRs	
ib_dereg_mr		
ib_update_fast_reg_key		
ib_inc_rkey		
ib_check_mr_access		
ib_map_mr_sg		
ib_map_mr_sg_zbva		
ib_sg_to_pages		
ib_create_srq		x722,t5 : currently unsupported
ib_modify_srq		x722,t5 : currently unsupported
ib_query_srq		t6 : future sw release, x722,t5 : currently unsupported
ib_destroy_srq		x722,t5 : currently unsupported
ib_post_srq_recv		x722,t5 : currently unsupported

KVERBS: CONNECTION MANAGEMENT

Verb	excluded parameters	NIC Support Exceptions
rdma_destroy_id		
rdma_bind_addr		
rdma_resolve_addr		
rdma_resolve_route		
rdma_init_qp_attr		
rdma_connect		
rdma_listen		
rdma_accept		
rdma_reject		
rdma_disconnect		
rdma_set_service_type		
rdma_set_reuseaddr		
rdma_set_afonly		
rdma_get_service_id		
rdma_reject_msg		
rdma_is_consumer_reject		
rdma_consumer_reject_data		

CHARACTERISTICS OF MULTI-VENDOR OFV

What it includes: The set of *essential verbs* widely used across many types of mainstream RDMA applications, and supported by a majority of surveyed NIC vendors

What it excludes: Verbs and features used only by specific types of RDMA applications, or supported by a minority of surveyed NIC vendors. **Some examples:**

- Verbs with “_ex” suffix
- UD and XRC transports
- Atomic and TSO ops
- Flow verbs
- Application-tailored CQE formats
- Per CQ event moderation
- Signature and Gaps MRs and T10 DIF, see link below
https://www.openfabrics.org/images/eventpresos/workshops2014/DevWorkshop/presos/Tuesday/pdf/11_Signature_Verbs.pdf
- Old-style FMR (does not include post_send with opcode reg_mr, see link below)
<https://www.openfabrics.org/images/eventpresos/2016presentations/204KernelVerbs.pdf>

QUESTIONS FOR THE AUDIENCE

Developers: Is the multi-vendor OFV definition valuable in development or maintenance OFV apps?

RDMA end users: Is there value in knowing which OFV applications conform to multi-vendor OFV?

All: Is this concept valuable enough to be owned, maintained, and expanded by a neutral consortium of NIC vendors, Developers, and RDMA end users (possible OFA activity?)

NEXT STEPS?

- **Ask other NIC vendors to participate**
- **Develop a process for changing (expanding) the definition**
- **Survey existing OFV apps to see which conform to multi-vendor OFV**
- **Treat “rdma_” verbs consistently. Currently, some are defined as multi-vendor and others are defined “out-of-scope” since they are simple wrapper functions**



OPENFABRICS
ALLIANCE

14th ANNUAL WORKSHOP 2018

THANK YOU

Brian Hausauer, Intel
Saqib Jang, Chelsio
Nishant Lodha, Cavium

