



# INTEROPERABILITY WORKING GROUP IWG AND CANDI UPDATE

- **Bob Noseworthy, UNH-IOL Chief Engineer** Presenting on behalf of IWG co-chairs Paul Bowden, Intel and Gilad Margalit, Mellanox
- **Paul Grun, OFA Vice Chair & Cray**

# OVERVIEW

## OFA-IWG Overview

Presented by Bob Noseworthy

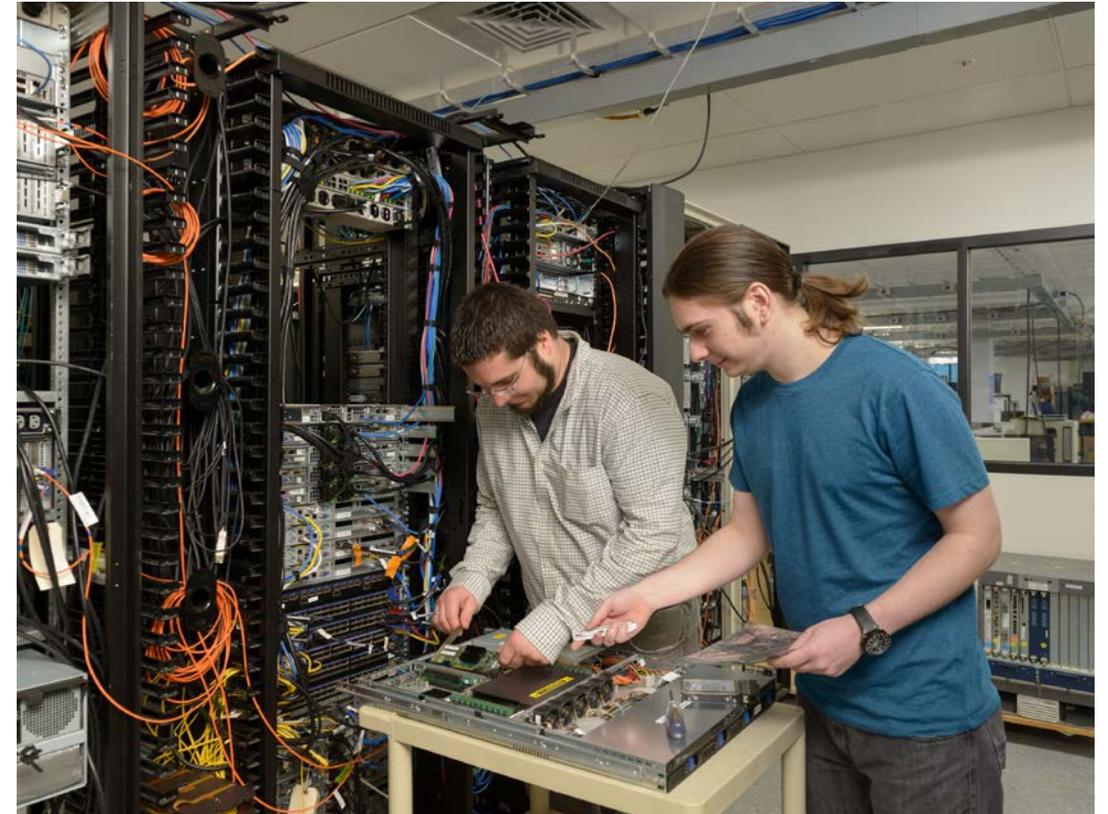
### ▪ OFA Interoperability Logo Program Update

- OS DISTRO / ON-DEMAND TESTING
- Test Plan expansion
  - Co-Existence
  - RoCEv2
- Opportunities with NVMe Over Fabrics

## Compliance and Interoperability (CANDI)

Presented by Paul Grun

### ▪ CANDI Update



# INTEROPERABILITY WORKING GROUP (OFA-IWG)

## Original Purpose

To deliver on the promise of end-user readiness, the OFA software running on servers or hosts needs to support interoperability in many ways – between different supported transports that run the same OFA software, and with switches, gateways, servers, and storage targets that contain external (e.g., OEM provided) software elements that work in conjunction with OFA software to provide fabric wide functionality.

The OFA-IWG was formed to address the interoperability requirements above.



- **Enable confidential 3<sup>rd</sup>-party testing in a multi-vendor environment**
- **Identify interoperability issues**
- **Conduct a neutral Interoperability Logo Program (OFILP)**

**Add value to the networking community through testing and validation**

# OFA-IWG

Open to participation by all

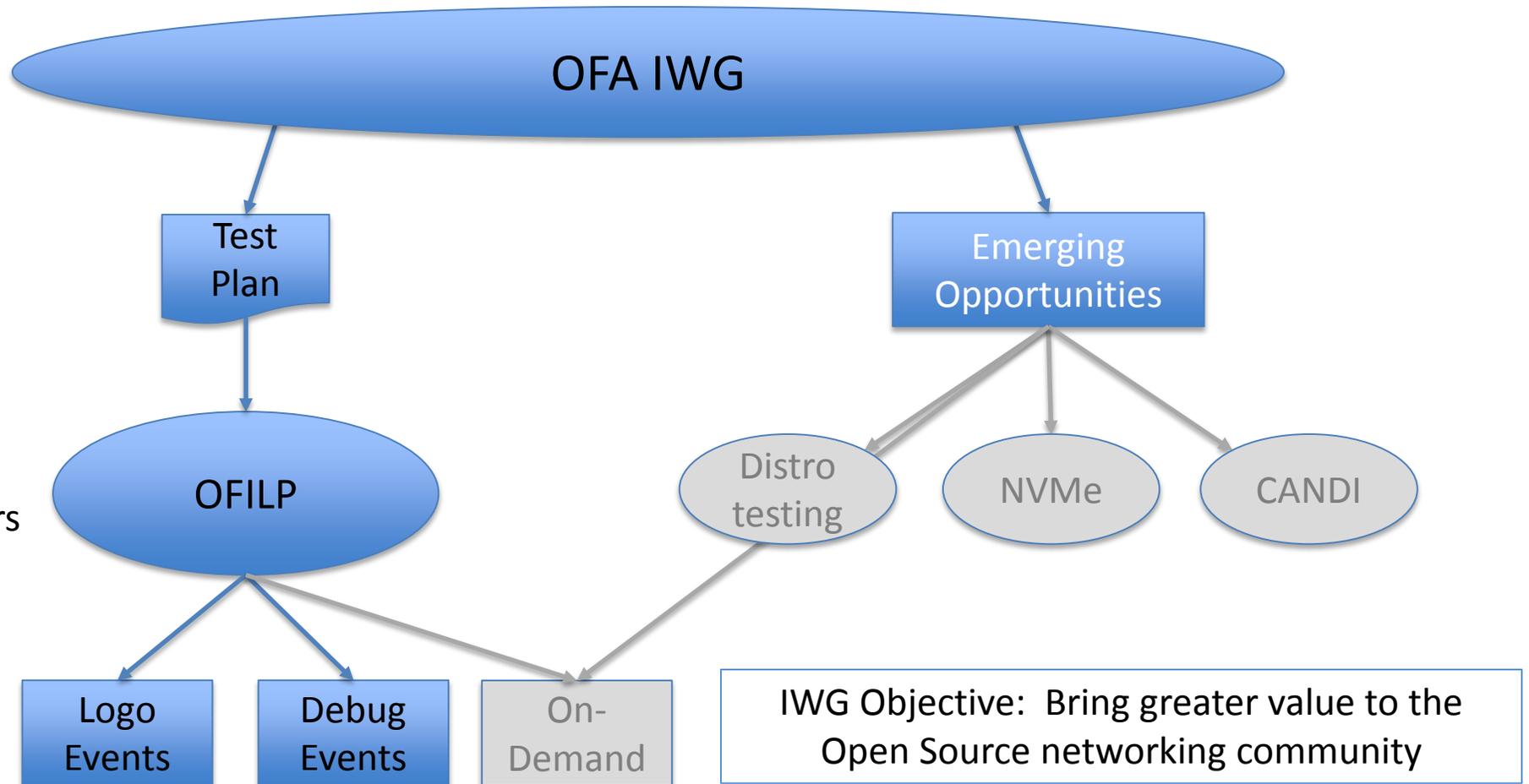
IWG oversees the Interop program via the Test Plan

OFA Interop Logo Program

- Funded by OFILP members
- Confidential

Programs include

- biannual Debug Events
- Logo events



# IMPROVING THE LOGO PROGRAM

Driven by members to address current and emerging needs

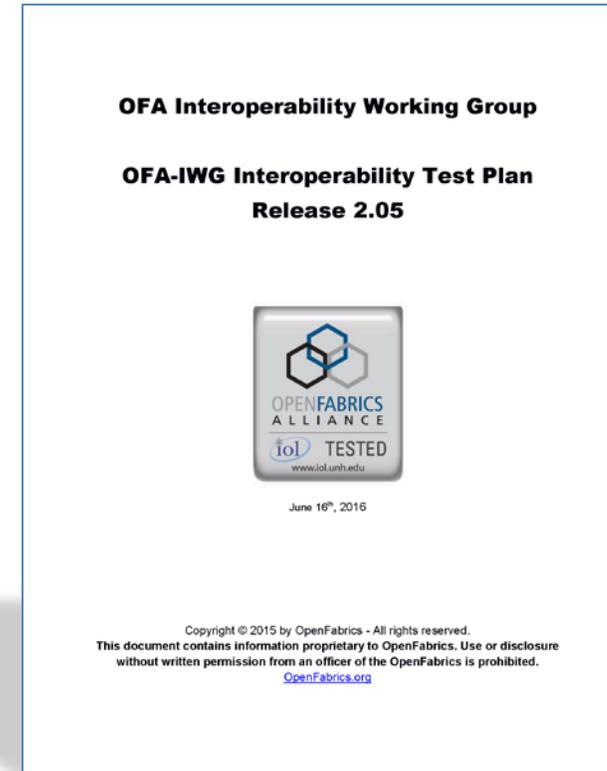
## ■ Co-Existence Coverage

- Co-Existence of multiple adapters within the same system
- Call to action: Is this Logo Program expansion important to you?

## ■ RoCE v2 Coverage

- RoCE v2 support needs member interest and definition for test coverage
- Call to action: Is this Logo Program expansion important to you?

**Are there other expansion areas important to you?**



**Interoperability Test plan** The OFA recognizes the need for the interoperability test effort to be industry-wide, where testing is conducted under the auspices of the appropriate networking organizations.

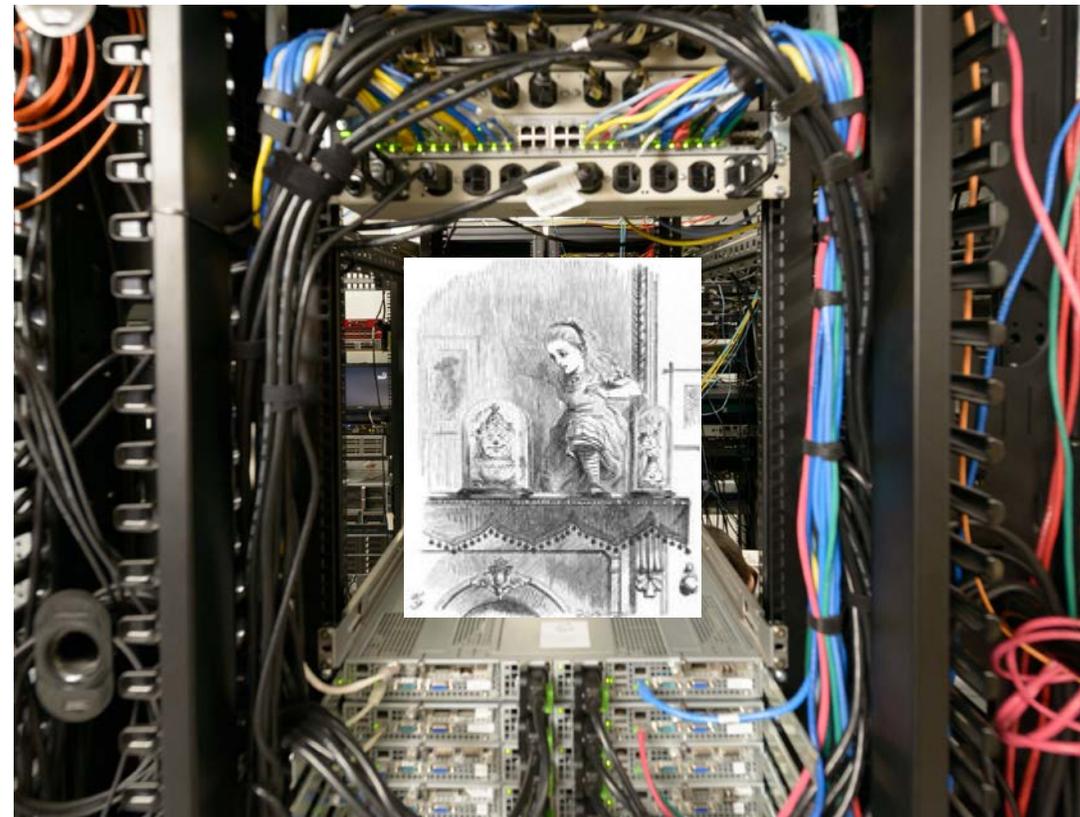
IBTA is responsible for InfiniBand compliance & interoperability testing on cables & devices. NVMe Org for NVMe storage devices. UNH-IOL provides DCB Switch testing. The OFA IWG is responsible for Interoperability testing on InfiniBand, iWARP and RoCE products based on OFA Software (OFED).

# PROPOSAL - OS DISTRO TESTING

Goal - improve upstream kernel and distro validation

## ■ OS DISTRO TESTING

- Working to address the needs of the Linux Distros
- Driven by requests from leading Distros, including SuSE and RedHat
- Apply the diverse multi-vendor OFA Cluster housed at UNH-IOL
- Enables inter-distro heterogeneous testing
  
- Address requests for more rapid testing
  - Increase automation of testing (eg: Jenkins)
  - Expand test robustness and speed reporting



*Delivering on the promise of end-user readiness*

# PROPOSAL - ON-DEMAND TESTING

Goal – Improve Responsiveness of the Testing Program

## ▪ ON-DEMAND TESTING

- OS Distro testing enables a pathway for On-Demand testing year-round utilizing the OFA Cluster
- Customized test cycles for members & expanded types of testing
- Avoid the need to wait for bi-annual interop/logo events
- Allows Distros to test alpha/beta or GA releases on their schedule
- Vendors can test their solutions when ready and still receive Logo
- Opportunity for greater upstream testing for ULPs and OFA software (RCs and GAs)

## ▪ Path forward still in discussion within IWG

- Focus is on meeting Linux Distro needs first and foremost

**Get involved with the IWG and help enable distro and upstream testing in the OFA Logo Program!**

# OFILG NEXT INTEROP/DEBUG EVENT

Play well with others!

- **OFA-IWG defined Test Plan**
  - Incremental updates always welcome
  - Contribution welcome from all OFA members in the IWG.
- **Target: OFED 4.8 RC1 (or latest)**

***Get involved !!  
Join the OFILG***

***Look to the OFILG Interoperability Logo List***

**Next Event:**

- **OFILG Interop/Debug Event #22**
- **May 1-5, 2017 at UNH-IOL**

The screenshot shows the website for the OFILG Interoperability Logo List. The page header includes the University of New Hampshire logo and the InterOperability Laboratory (IOL) logo. The main content area is titled "OFILG Interoperability Logo List - July 2016" and includes a breadcrumb trail: Home > Product Registries > OpenFabrics Interop Logo List > OFILG Interoperability Logo List - July 2016. Below the title, it identifies the "OpenFabrics Alliance Interoperability Working Group (OFA-IWG)" and states "OFED Version Tested: 3.18-1 and 3.18-2". An "Introduction" section explains that the logo list contains products that have passed mandatory testing. A "Contents" section lists categories such as Ethernet RNICs, RoCE Server System, SRP Targets, ISER Targets, and OFILG Cluster Contributors. An "Ethernet RNICs" section lists "Features Tested: Link Initialization, uDAPL, RDMA Basic Interop, RDMA Stress, MPI". At the bottom, a table header is visible with columns for Manufacturer, Model Number, Ports, Speed, HW, FW, and Report. On the right side of the page, there is a logo for the OpenFabrics Alliance with the text "IOL TESTED www.iol.unh.edu".

**Register for the next OFA Interop/Debug Event at: <https://www.iol.unh.edu/testing/hpc/ofa/grouptest>**

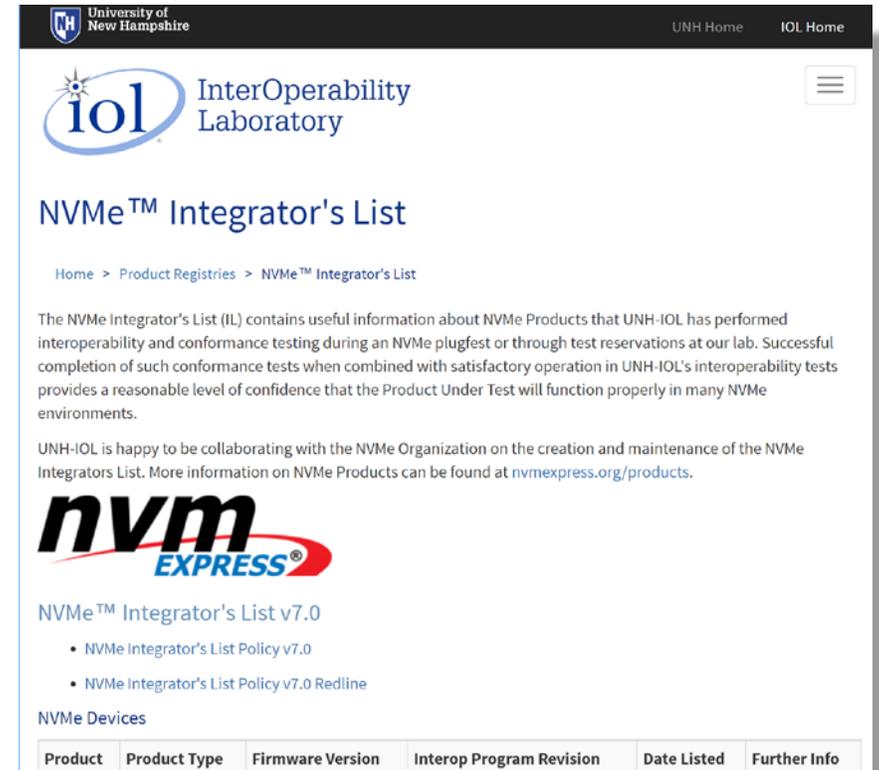
# OFA AND NVME OVER FABRICS

Opportunities ahead !

- **NVMe over Fabrics is an emerging technology**
- **UNH-IOL performs testing on behalf of NVMeexpress.org**
  - For OFILG members: UNH-IOL is providing access for RNIC, HCA and RCAs, to the next NVMe Plugfest, where initial exploration of testing for NVMe over Fabrics will occur.
  - Future events may be handled differently
  - Contact [kerry.munson@iol.unh.edu](mailto:kerry.munson@iol.unh.edu)

## Next Event:

- **NVMe Plugfest #7 with NVMe-MI + Fabrics**
- **May 22-25, 2017 at UNH-IOL**



The screenshot shows the UNH-IOL website. The header includes the University of New Hampshire logo and navigation links for UNH Home and IOL Home. The main content area features the IOL logo and the title "NVMe™ Integrator's List". Below the title is a breadcrumb trail: Home > Product Registries > NVMe™ Integrator's List. The text explains that the NVMe Integrator's List (IL) contains useful information about NVMe Products that UNH-IOL has performed interoperability and conformance testing during an NVMe plugfest or through test reservations at their lab. It also mentions that UNH-IOL is collaborating with the NVMe Organization on the creation and maintenance of the NVMe Integrators List. The NVMe EXPRESS logo is displayed. Below the logo, there is a section for "NVMe™ Integrator's List v7.0" with two bullet points: "NVMe Integrator's List Policy v7.0" and "NVMe Integrator's List Policy v7.0 Redline". At the bottom, there is a table header for "NVMe Devices" with columns: Product, Product Type, Firmware Version, Interop Program Revision, Date Listed, and Further Info.

Register for the next NVMe Plugfest at: <https://www.iol.unh.edu/event/2017/01/nvme-plugfest-7-nvme-mi-fabrics>



# CANDI UPDATE (COMPLIANCE AND INTEROPERABILITY)

Paul Bowden, Paul Grun, Sean Hefty, Bob Russell, Jim Ryan, Jerome Berryhill, Alexia Ingerson

# RE-CAP FROM LAST YEAR – A FIVE PART PROPOSAL

## (INFORMAL) PROPOSAL

**Augment the existing Interop Program with these elements:**

- **Training**
  - Key beneficiary is the consumer community
- **Provider Compliance Validation**
  - Key beneficiary is the provider vendors (with some benefit to the consumer)
- **Support for Consumer Code Development**
  - Assistance and advice in implementation and coding
  - Emphasize agile development techniques
- **Validation, Tuning, and Optimization**
  - Assistance in deploying an implementation
- **Support for testing at scale**
  - An unsolved problem, can a CoE add value here?

Create an OFA “Center of Excellence” in Networking

A small team began meeting in 2016 to assess the potential for creating more value add on top of the existing Interop program

Intended to meet a wide variety of needs, ranging from provider vendors to consumers and deployers

# RE-CAP FROM LAST YEAR – A FIVE PART PROPOSAL

## (INFORMAL) PROPOSAL

Augment the existing Interop Program with these elements:

- **Training**
  - Key beneficiary is the consumer community
- **Provider Compliance Validation**
  - Key beneficiary is the provider vendors (with some benefit to the consumer)
- **Support for Consumer Code Development**
  - Assistance and advice in implementation and coding
  - Emphasize agile development techniques
- **Validation, Tuning, and Optimization**
  - Assistance in deploying an implementation
- **Support for testing at scale**
  - An unsolved problem, can a CoE add value here?

Create an OFA “Center of Excellence” in Networking

15

OpenFabrics Alliance Workshop 2016

Starting with Provider  
Compliance Validation is  
logical

# RE-CAP FROM LAST YEAR – A FIVE PART PROPOSAL

## (INFORMAL) PROPOSAL

Augment the existing Interop Program with these elements:

- **Training**
  - Key beneficiary is the consumer community
- **Provider Compliance Validation**
  - Key beneficiary is the provider vendors (with some benefit to the consumer)
- **Support for Consumer Code Development**
  - Assistance and advice in implementation and coding
  - Emphasize agile development techniques
- **Validation, Tuning, and Optimization**
  - Assistance in deploying an implementation
- **Support for testing at scale**
  - An unsolved problem, can a CoE add value here?

Create an OFA “Center of Excellence” in Networking

15

OpenFabrics Alliance Workshop 2016

Developing a training program for OFI is also on the TODO list, but that task belongs to the XWG/BoD

# LAST YEAR WE IDENTIFIED A NEED FOR A COMPLIANCE PROGRAM

## COMPLIANCE

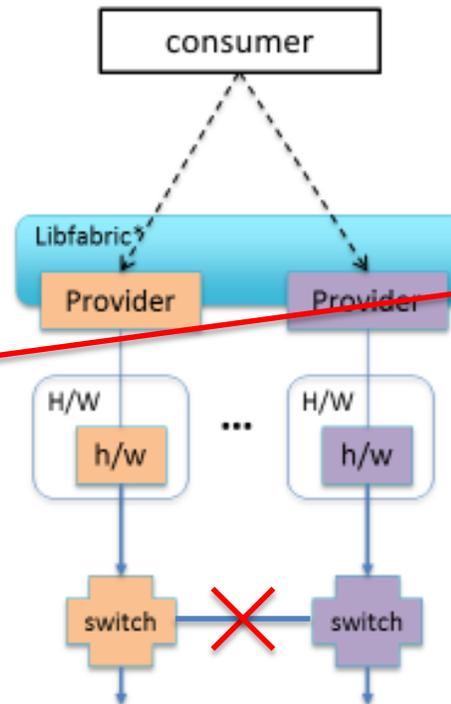
The goal now is to ensure that each vendor ('Provider') implements the API functions correctly, and that services are exported to the consumer correctly.

In other words, ensure that the design complies with the API as defined in the libfabric MAN pages.

Consumer code (middleware, application level) now becomes portable between providers.

\*libfabric – user mode library for distributed and parallel computing

10



wire protocols are different!

OpenFabrics Alliance Workshop 2016

Augments, but does not replace, the current Interop program

We concluded that 'compliance' and 'interoperability' are different things

OFI architecture doesn't require interoperability, (but also doesn't prohibit it)

# COMPLIANCE? INTEROPERABILITY?

## Interoperability:

- A device is interoperable if it works “correctly” with other devices – e.g. does this HCA work with other HCAs? Switches?
- Interoperability is established through a series of exhaustive tests, usually focusing on a matrix of components

## Compliance:

- An object is in compliance if it conforms to a set of requirements as measured by some objective criteria
- Typically, requirements are conveyed in an industry standard
- Compliance is determined by testing against each requirement

Today, devices are not tested for compliance  
Compliance testing is left to the standards bodies e.g. IBTA, IEEE...

Endpoint Types	bgq	gni	mxm	psm	psm2	sockets	udp	usnic	verbs
FI_EP_DGRAM	✗	✓	✗	✓	✓	✓	✓	✓	✗
FI_EP_MSG	✗	✓	✗	✗	✗	✓	✗	*	✓
FI_EP_RDM	✓	✓	*	✓	✓	✓	✗	*	*

Primary Capabilities	bgq	gni	mxm	psm	psm2	sockets	udp	usnic	verbs
FI_MSG	✓	✓	✗	✓	✓	✓	✓	✓	✓
FI_RMA	✓	✓	✗	✓	✓	✓	✗	✗	*
FI_TAGGED	✓	✓	✓	✓	✓	✓	✗	✗	*
FI_ATOMIC	✓	✓	✗	✓	✓	✓	✗	✗	*
FI_NAMED_RX_CTX	✓	✓	✗	✗	✗	✓	✗	✗	✗
FI_DIRECTED_RECV	✓	✓	✗	✗	✓	✓	✗	✗	✗
FI_READ	✓	✓	✗	✓	✓	✓	✗	✗	✓
FI_WRITE	✓	✓	✗	✓	✓	✓	✗	✗	✓
FI_SEND	✓	✓	✗	✓	✓	✓	✓	*	✓
FI_RECV	✓	✓	✗	✓	✓	✓	✓	*	✓
FI_REMOTE_READ	✓	✓	✗	✓	✓	✓	✗	✗	✓
FI_REMOTE_WRITE	✓	✓	✗	✓	✓	✓	✗	✗	✓
FI_MULTICAST	?	✗	?	?	?	?	?	?	?

- The OFI architecture is designed to support
- Multiple providers
  - Each delivering a unique feature set

Compliance Objective: if a provider supports a feature, is that feature implemented correctly?

# FABTESTS – A GOOD PLACE TO START

- **The OFIWG project has developed a set of FABTESTS**
  - Designed to help Provider developers ensure their code works with the libfabric API
  
- **Our objective is to leverage FABTESTS to validate the coverage of features and options by any given provider**
  
- **Naturally, several issues arose:**
  1. The existing FABTESTS may not be sufficient to test the complete set of endpoints
  2. Even if tests exist, there's no guarantee today that all options for that endpoint type are exercised
  3. Once the set of FABTESTS is complete, still need to validate a given provider works properly with its defined set

# HERE'S WHAT WE'RE DOING

- ✓ **Catalog all features and options described in the existing MAN pages – (done)**
- Augment the portfolio of Fabtests to cover any existing gaps**
  - Call to action – help is needed to round out the library of fabtests
- Statically analyze existing Fabtests code paths – which features of a libfabric provider could theoretically be exercised? (in process today)**
- Dynamically analyze Fabtests code paths – which features are actually testable using available command line options?**
  - Call to action – help is needed to develop tools and methods to accomplish this
- Decide how to certify any given Provider against the resulting set of FABTESTS**
  - Call to action – decide what level of certification is desired

man3	function/struct	field/param	value	Active MSG EP	Passive MSG EP	RDM EP	DGRAM EP	Active STREAM EP	Passive STREAM EP	SOCK DGRAM EP
<u>fi_getinfo</u>										
	<u>fi_getinfo()</u>									
		<u>flags</u>								
			FI_NUMERICHOST							
			FI_SOURCE							
			FI_PROV_ATTR_ONLY							
	<u>fi_freeinfo()</u>									
	<u>fi_allocinfo()</u>									
	<u>fi_dupinfo()</u>									
	<u>struct fi_info</u>									
		<u>next</u>								
		<u>caps</u>								
			<i>primary</i>							
			FI_MSG							
			FI_TAGGED							
			FI_RMA							
			FI_ATOMIC							
			FI_SEND							
			FI_RECV							
			FI_MULTICAST							
			FI_DIRECTED_RECV							
			FI_NAMED_RX_CTX							
			FI_READ							
			FI_REMOTE_READ							
			FI_WRITE							
			FI_REMOTE_WRITE							

A snippet of the canonical list of libfabric's functions and structures

As the program proceeds, the cells are filled in with a value to indicate if that combination is testable via one or more Fabtests

This becomes the basis for defining the set of FABTESTS to be run to verify compliance with the MAN pages

We plan to post this reference to github shortly

man3	function/struct	field/param	value	Active MSG EP	Passive MSG EP	RDM EP	DGRAM EP	Active STREAM EP	Passive STREAM EP	SOCK DGRAM EP
fi_msg										
	fi_recv()			Y	Y					
	fi_recvv()			0	0					
	fi_recvmmsg()			0	0					
		recv flags								
		FI_COMPLETION		0	0					
		FI_MORE		0	0					
		FI_MULTI_RECV		0	0					
	fi_send()			Y	Y					
	fi_sendv()			0	0					
	fi_sendmmsg()			Y	Y					
		send flags								
		FI_COMPLETION		0	0					
		FI_MORE		0	0					
		FI_FENCE		0	0					
		FI_REMOTE_CQ_DATA		0	0					
		FI_INJECT		Y	Y					
		FI_INJECT_COMPLETE		0	0					
		FI_TRANSMIT_COMPLETE		Y	Y					
		FI_DELIVERY_COMPLETE		0	0					
		FI_MULTICAST		0	0					
	fi_inject()			0	0					

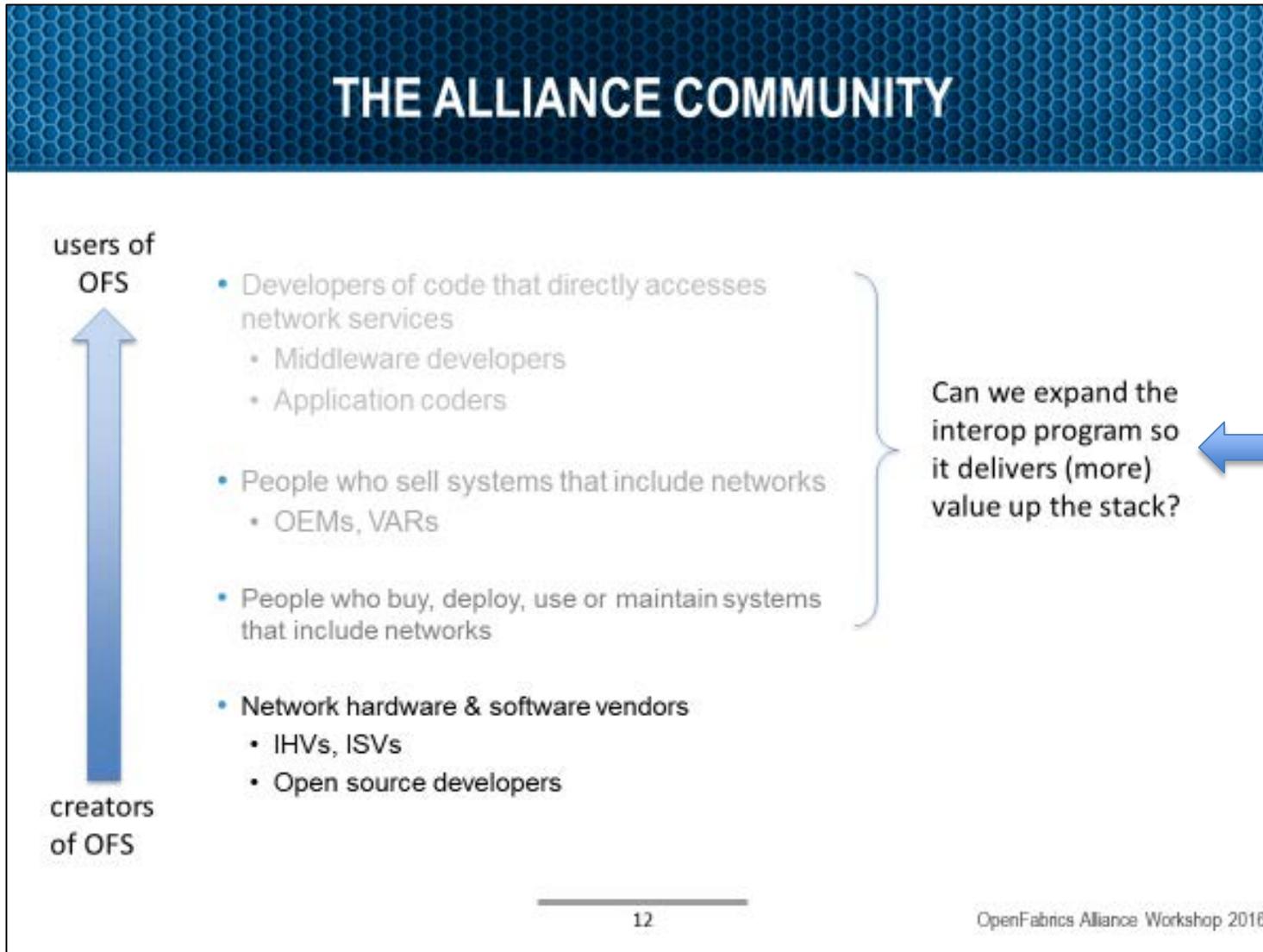
Results of the static path analysis for one FABTEST – this happens to be for a ping-pong test.

A ‘Y’ indicates that this FABTEST contains a code path that exercises that function.

Static analysis does not tell us if that path is actually exercised or not. That falls to the dynamic analysis

Note that none of the other endpoint types are exercised, because the ping-pong test only uses the Active MSG and Passive MSG EPs

# BEYOND COMPLIANCE TESTING



We need to hear from application developers and users – What do you want to get out of this emerging program?



OPENFABRICS  
ALLIANCE

13<sup>th</sup> ANNUAL WORKSHOP 2017

**THANK YOU**

Bob Noseworthy, UNH-IOL

Paul Grun, Cray