Design and Implementation of a Multi-Fabric Message Passing Interface (MPI): Intel® MPI Library

Authors:

Dr. Michael D'Mello
Intel Americas Inc., Champaign, Illinois, USA

Correspondence:

Michael D'Mello
Intel FX-121
1906 Fox Drive
Champaign, IL 61820

Tel: +1-217-403-4211
Fax: +1-217-403-4399
Email: michael.d'mello@intel.com

ABSTRACT

**Design and Implementation of a Multi-Fabric Message Passing
Interface (MPI):
Intel® MPI Library
Version 1.0**

**Intel Corporation
Parallel & Distributed Solutions Division**

**Presented by:
Michael D'Mello
michael.d'mello@intel.com**

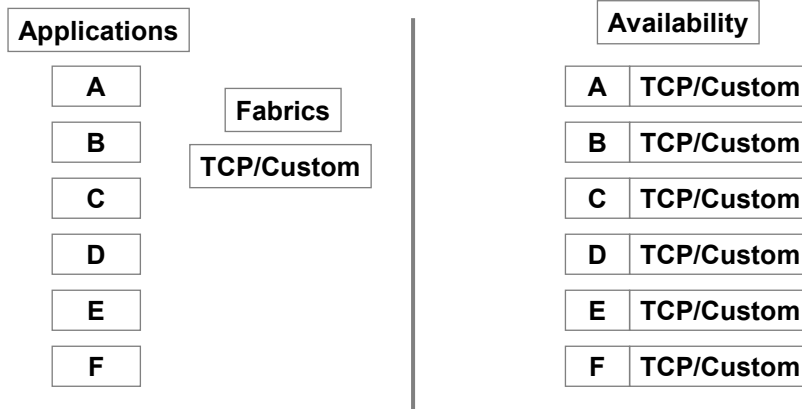**INTRODUCTION**

Contents

- Environment & Situation Analysis
- Program direction
- Intel® MPI Design Considerations
    - High Level Design
    - Stability
    - Functionality
    - Performance
    - Portability
    - Environment & Tools support
    - Extensibility
- Highlights of Intel® MPI Program

Current Environment – Single Fabric Clusters

Applications
A
B
C
D
E
F

Fabrics
TCP/Custom

Availability	
A	TCP/Custom
B	TCP/Custom
C	TCP/Custom
D	TCP/Custom
E	TCP/Custom
F	TCP/Custom

Full application coverage and simple deployment

intel.

3

Environment – Fragmentation effect

Applications
A
B
C
D
E
F

Fabrics
TCP
Myrinet
IB
Other

Availability			
A	T	IB	Other
B	T	Myrinet	Other
C	T	Myrinet	IB
D	T	IB	
E	T	Other	
F	T		
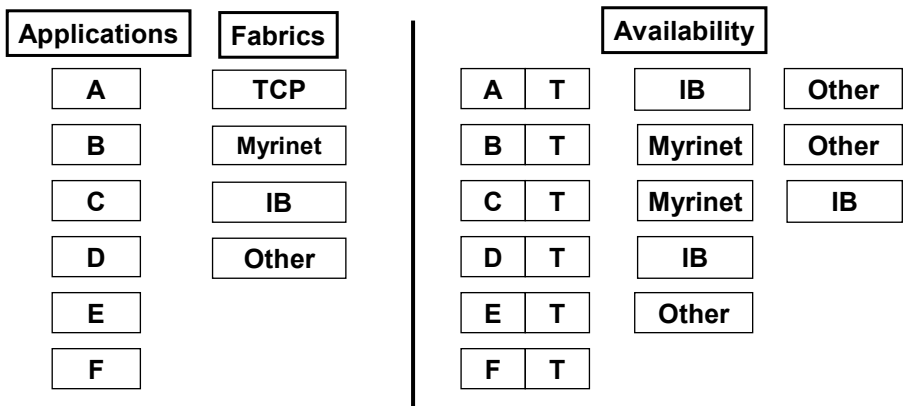
**Customer running apps A & F has to choose TCP*

intel.

4

A Simpler Solution

Applications

| A |
| B |
| C |
| D |
| E |
| F |

Intel® MPI with Virtual Fabric Technology

Fabrics

| TCP |
| Myrinet |
| IB |
| Shm |
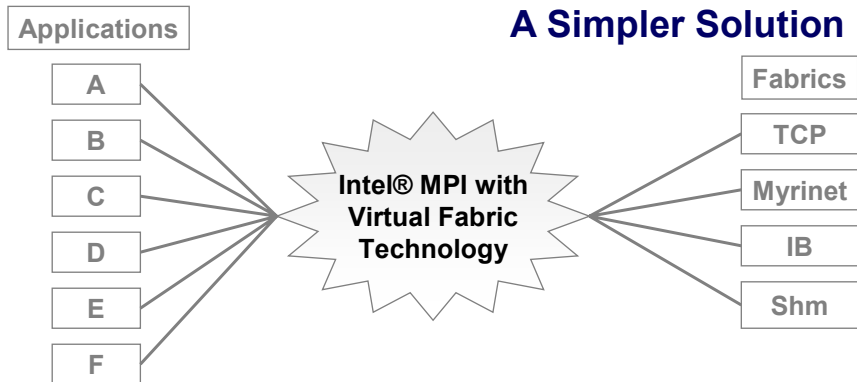
ISV's see & support a single interconnect fabric

End users enjoy simplified buying decision – enhanced value

intel.

5

Market Drivers for Intel® MPI Library

Target	Perceived Value	Reasoning
ISV's	Highest	Intel® MPI directly reduces SW production costs, increases adoption
IHV's & OEM's	Highest	Intel® MPI drives competitive advantage for their products
Commercial end users (GE, GM, etc.)	High	Cost reduction – more MPI apps will be enabled on a given fabric
National Labs	Medium	Often opt for single fabric solutions – will lean towards native implementations
Universities & Academics	Lowest	Have other free multi-fabric MPI options

intel.

6
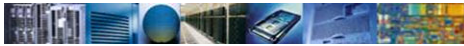
Intel® MPI Library Program Direction

- **Objective:** Provide a product quality, multi-fabric, high- performance, cost effective MPI library, with a high degree of portability, extensibility, and support
- **Key product focus areas:**
  - Stability
  - Functionality
  - Performance
  - Portability
  - Environment & Tools support
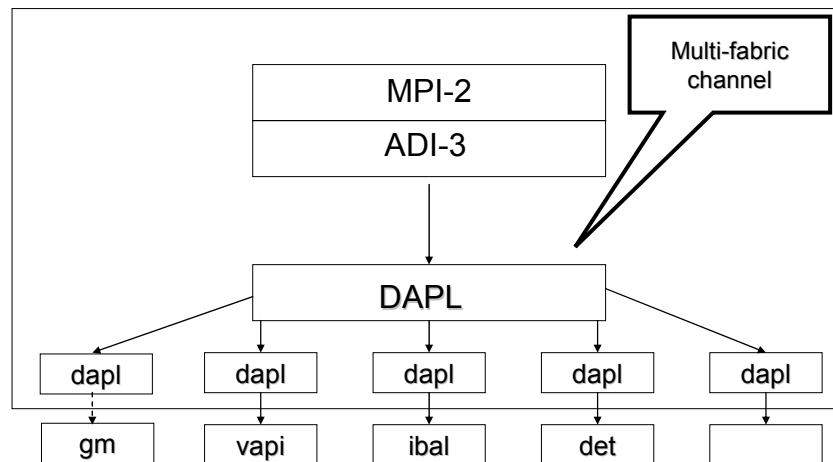 - Extensibility

intel.

7

Intel® MPI v1.0: High Level Schematic

intel.

8

Product Design Considerations

- Organized by focus areas:
    - Stability
    - Functionality
    - Performance
    - Portability
    - Environment & tools
    - Extensibility

Intel® MPI: Stability

- A product-quality implementation of MPI 2.0 based on MPICH-2
- Rigorous QA testing for base Functionality & Performance)
    - Several test suites including ANL's test suite & IMB
- Rigorous definition, design, and testing on the device layer
- Backup functionality mode

Intel® MPI: Functionality

- Based on ANL's MPICH-2 release of MPI-2 spec
- All MPI-2 enhancements targeted:
  - One sided communication
  - Dynamic process creation
  - MPI-2 I/O enhancement
- Daemon based job startup: *mpiexec*

intel.

11

Intel® MPI: Performance

- Performance addressed in two ways:
  - All drivers tuned for performance relative to best-in-class reported performance
  - Multi-fabric capability
- Demonstrated low latency overhead "penalty" with small messages
- Fabrics supported in v1.0:
  - Base: TCP/IP & Shared Memory
  - Infiniband
  - Myrinet
  - Direct Ethernet Transport (experimental)

intel.

12

Intel® MPI: Portability

- Based on ANL's MPICH-2 release of MPI-2 spec
- MPI-1 subset of MPI-2: allows for high degree of compatibility with previous versions of MPI
- Ease of I/O transformation of MPI-1 codes dependent on extent of use of ROMIO (1)

Intel® MPI: Extensibility

- Modular architecture
- Vendor drivers meeting specifications can easily be slotted into Intel® MPI framework
  - Specifications for device drivers in place
  - Component tests for device drivers in place
- Device selection via device list at runtime

Intel® MPI: Broad environment & tools support

- Intel® MPI supported on IA-32, Itanium processors, processors with Intel® EM64T
- Intel® MPI targeted to
  - RHEL 3
  - Intel® compilers 8.0 (Fortran, C, C++)
  - Support available for other versions of Linux and for gcc compilers
- Support for Intel® MPI tools:
  - Intel® Trace Analyzer
  - Intel® Trace Collector

intel.

15

Summary and Conclusions

Future of Intel® MPI Program

- Intel® MPI is only one part of a bigger picture of portability and interoperability driven by HPC market needs
- Interoperability program to simplify adoption by ISV's, IHV's, and OEM's
- Future opportunities to extend functionality to cover areas of high availability, security, and manageability

intel.

16

Q&A

17