IMPI

- IMPI – Interoperable Message Passing Interface
- Developed and Proposed by NIST
- Standard for inter-operation of multiple
  - Implementations (IMPI, IMPI-2)
  - Architectures
  - Networks
  - Operating Systems
Client, Host Concept

- MPI processes spread across multiple clients
- Clients represent MPI processes belonging to a single implementation
- Hosts represent gateways for processes of Clients
- IMPI Application may have two or more clients
- Client may have one or more hosts
- Hosts serve as gateways for one or more MPI processes
Typical Scenario – Multi-vendor MPI

- 3 Clients (Each cluster make one client)
  - Client 1
    - 2 hosts, 2 MPI processes
  - Client 2
    - 1 host, 3 MPI processes
  - Client 3
    - 2 hosts, 6 MPI processes
MPI/Pro 1.7.0

- MPI/Pro 1.7.0 provides first complete implementation of IMPI
- Enables Interoperation between
  - Windows, Linux and Mac OSX operating systems
  - 32-bit and 64-bit architectures
  - TCP, GM and VAPI Networks
  - Any combination of all the above
Extensions

- IMPI does not address issues such as
  - Private IP Addresses
  - Owner domains
  - Faults
  - Interacting Dynamic Groups
- Above issues play vital role in Grid
- Verari proposed and implemented a novel method to address issue of Private IP Addresses
Case Study

Private IP Enabled IMPI
Typical Cluster Setup

• Compute Nodes have private IP addresses
• External communication through single head node or gateway
• Unsuitable for multi-cluster multi-site communication
Network Address Translation (NAT)

- Firewalls block incoming connections
- NAT used to serve requests generated within the cluster
NAT-based IMPI

- Use NAT to generate dynamic mappings between head node and compute nodes
- Dissipate dynamic mapping info through IMPI server
- Release mapped ports on head node on completion of application
Performance

<table>
<thead>
<tr>
<th>Configuration</th>
<th>Latency (us)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MPI/Pro without IMPI</td>
<td>142.45</td>
</tr>
<tr>
<td>MPI/Pro with IMPI</td>
<td>147.35</td>
</tr>
</tbody>
</table>
Performance

IMPI using NAT - Bandwidth

- Bandwidth
- Streaming Bandwidth

MB/s vs Message Length (KBytes)
Proposed Extensions

- IMPI extensions for MPI-2
- Open protocol-based initialization such as SOAP
- Adaptation to the Grid
- Reduce user involvement
- Optimize for performance
References