Executing Independent Parallel Applications

Using the SPEEDES Communications Library

Member of Technical Staff, RAM Labs

Terrence McGuinness

Objective

Performance Evaluation of a High Performance HLA RTI Communications Library

Acknowledgements

SPAWAR System Center: ***Dr. Larry Peterson, Program Manager***
RAM Labs: ***Dr. Jeffrey S. Steinman, Technical Lead***

Technical Staff

Chris Bailey
Joe Clark
Chris Landa
Randal Rogers
Terrence McGuinness

Maintain and modify the HPC shared memory communications

Communications Developers

Dr. Van Iwaarden
SPEEDES Communications
Shared Memory & TCP/IP

Dr. Roger Hillson
MPI interface to Speeds Comm

Overview

- **SPEEDES: (Simulation Engine)**

  Synchronous Parallel Environment for Emulation and Discrete-Event Simulation

- **RTI: (Interoperating Simulations)**

  Run Time Infrastructure

- **HLA: (DMSO standard)**

  High Level Architecture

High Performance HLA RTI

- RTI-NG 1.3, pRTI
  - Network Based

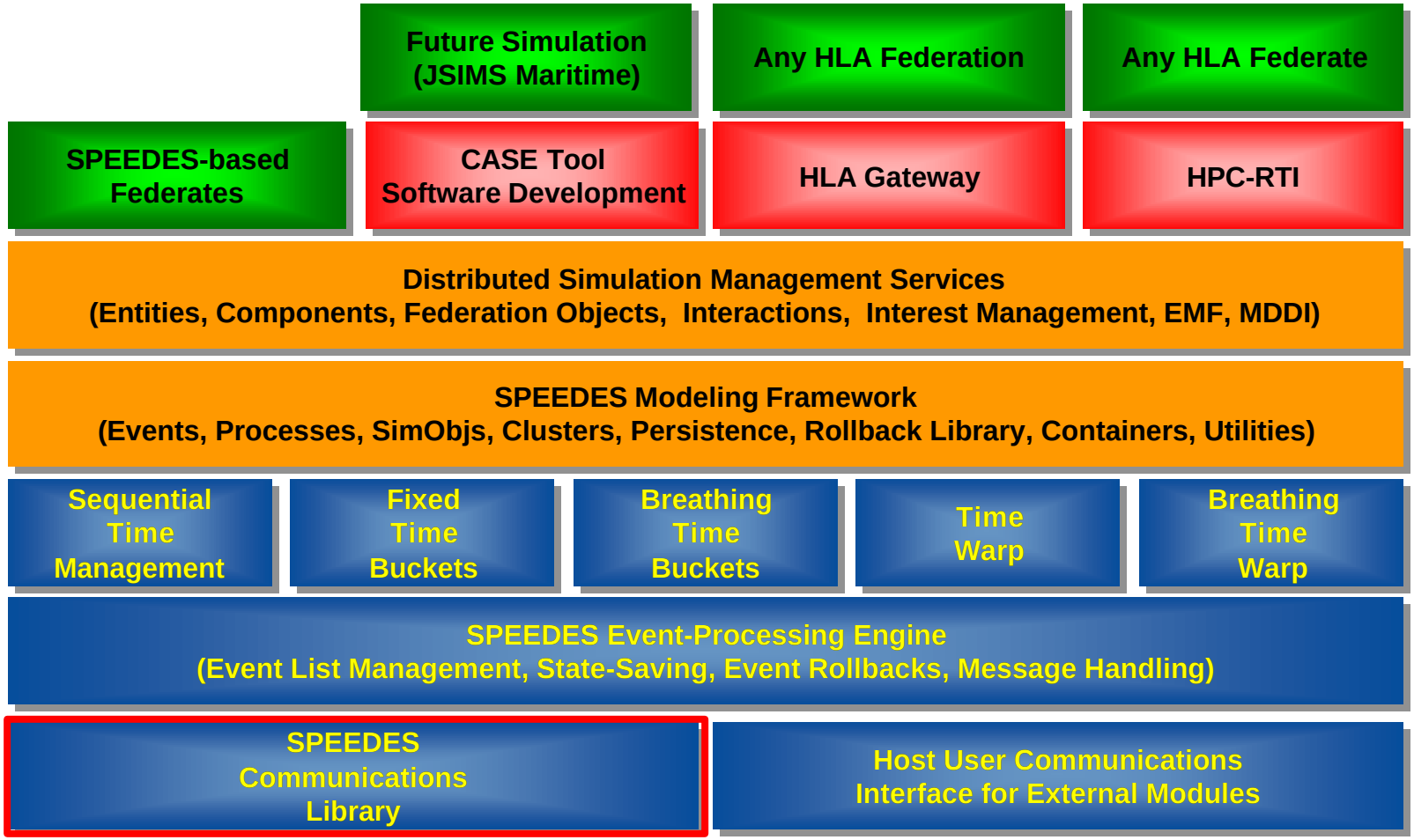
- High Performance Computing
  - Fast Communications
  - Large Data Sets
  - Computationally Bound
  - Automatic Parallelization

High Performance HLA RTI

- **SPEEDES**
  - Shared Memory and Network Communications
  - Declaration Management
  - Data Distribution Management
  - Ownership Management
  - Time Management

SPEEDES can inherently support HPC RTI

SPEEDES Layered Architecture



SPEEDES Communications

- **SPEEDES Communication Services are fine tuned to meet the specific needs of parallel event simulations.**

  - ***Buffered and asynchronous***
  - ***Immediate sending of short messages***
  - ***Non-Blocking Services***
  - ***Mixing of Network and Shared memory***

SPEEDES Communication Services

❑ Standard Communication Services

- **Initialization:** *SpComm_Startup*
  SpComm_GetNumNodes
  SpComm_GetNodeId
- **Asynchronous Buffered Point-to-point:** *SpSend/SpRecive*
- **Reduction:** *SpComm_GlobalMin(SpSimTime)*
- **Barrier Synchronization:** *SpComm_Barrier*

SPEEDES Communication Services

- **SPEEDES Communication Services unique to Parallel Event Simulations**

- **Non-blocking Barrier Synchronization:**
  - **SpComm_EnterFuzzyBarrier()**
  - **SpComm_ExitFuzzyBarrier()**

Once a process issues a fuzzy barrier subsequent calls to barrier poll rather than block

- **Buffered Asynchronous and Node-to-multiple-node:**

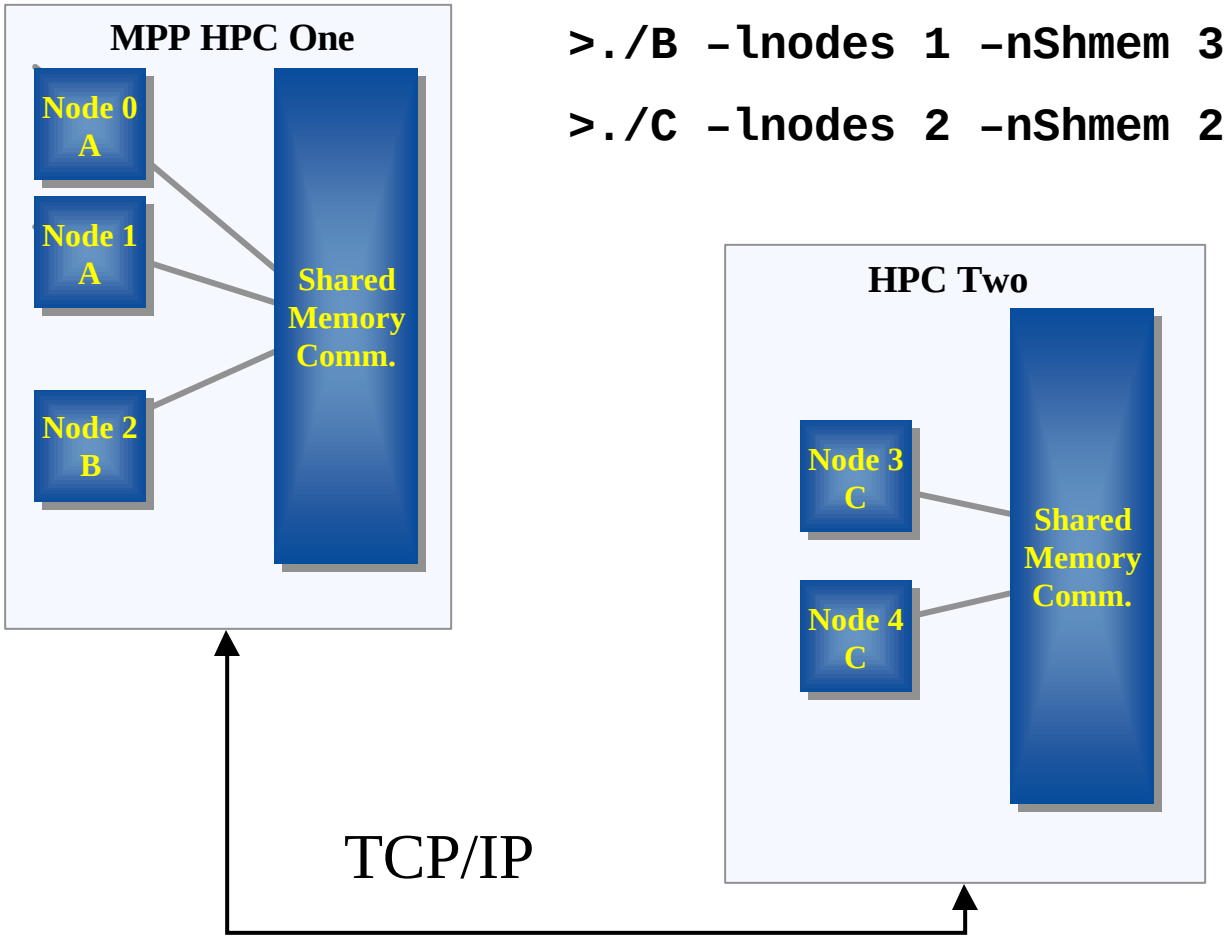
SpComm_Send(int dest, …)	***uni-cast***
SpComm_Send(SpDestination, …)	***multi-cast***
SpComm_Send()	***broadcast***

- **Coordinated Messaging:**
  - **SpComm_CoordinatedSend(SpDestination, …)**
  - **SpComm_CoordinatedReceive()**

Coordinated messages block all process in the designated set until all receives are completed

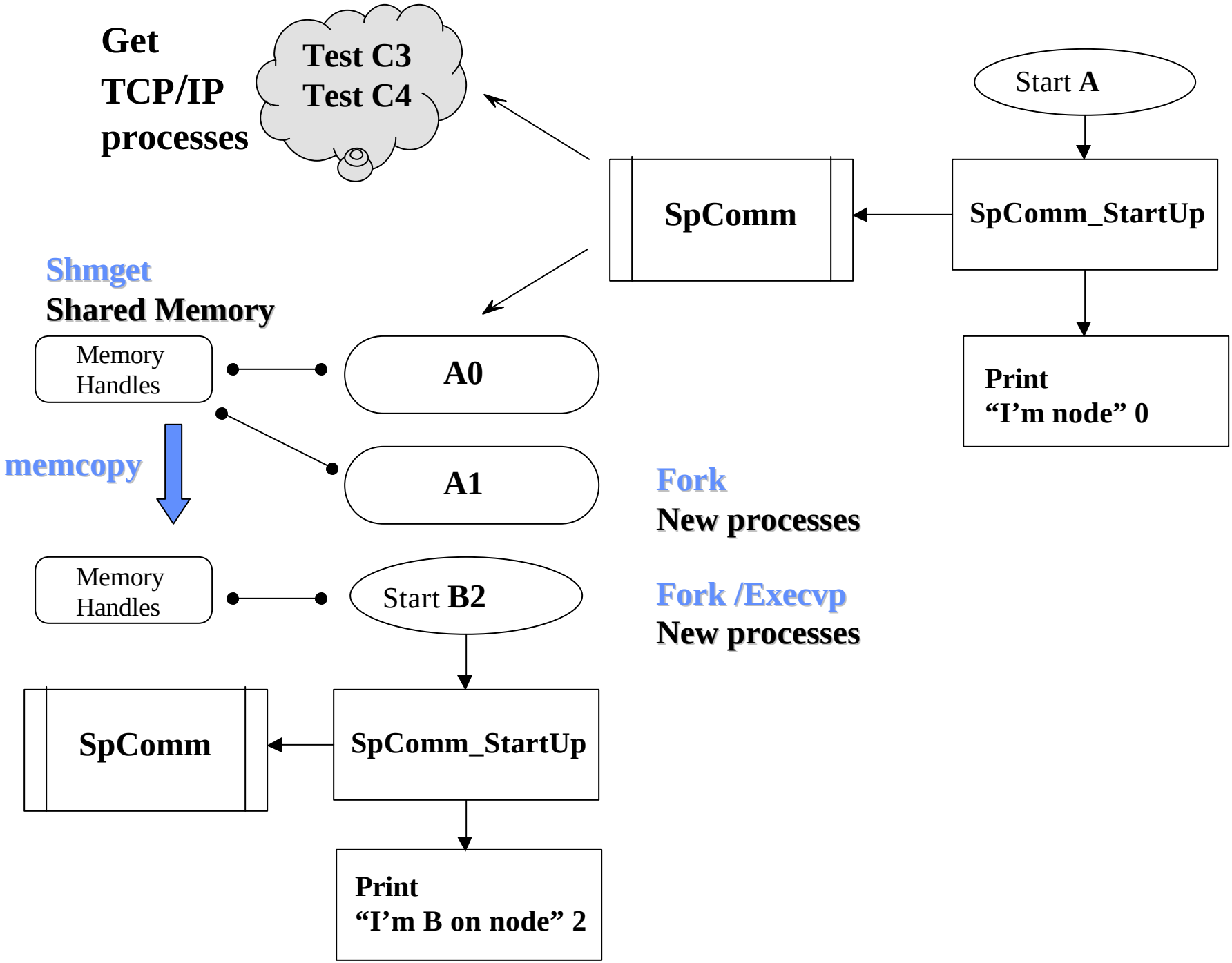
Communication Space

```
>./A -lnodes 2 -nShmem 3 -nnodes 5
>./B -lnodes 1 -nShmem 3 -nnodes 5
>./C -lnodes 2 -nShmem 2 -nnodes 5
```

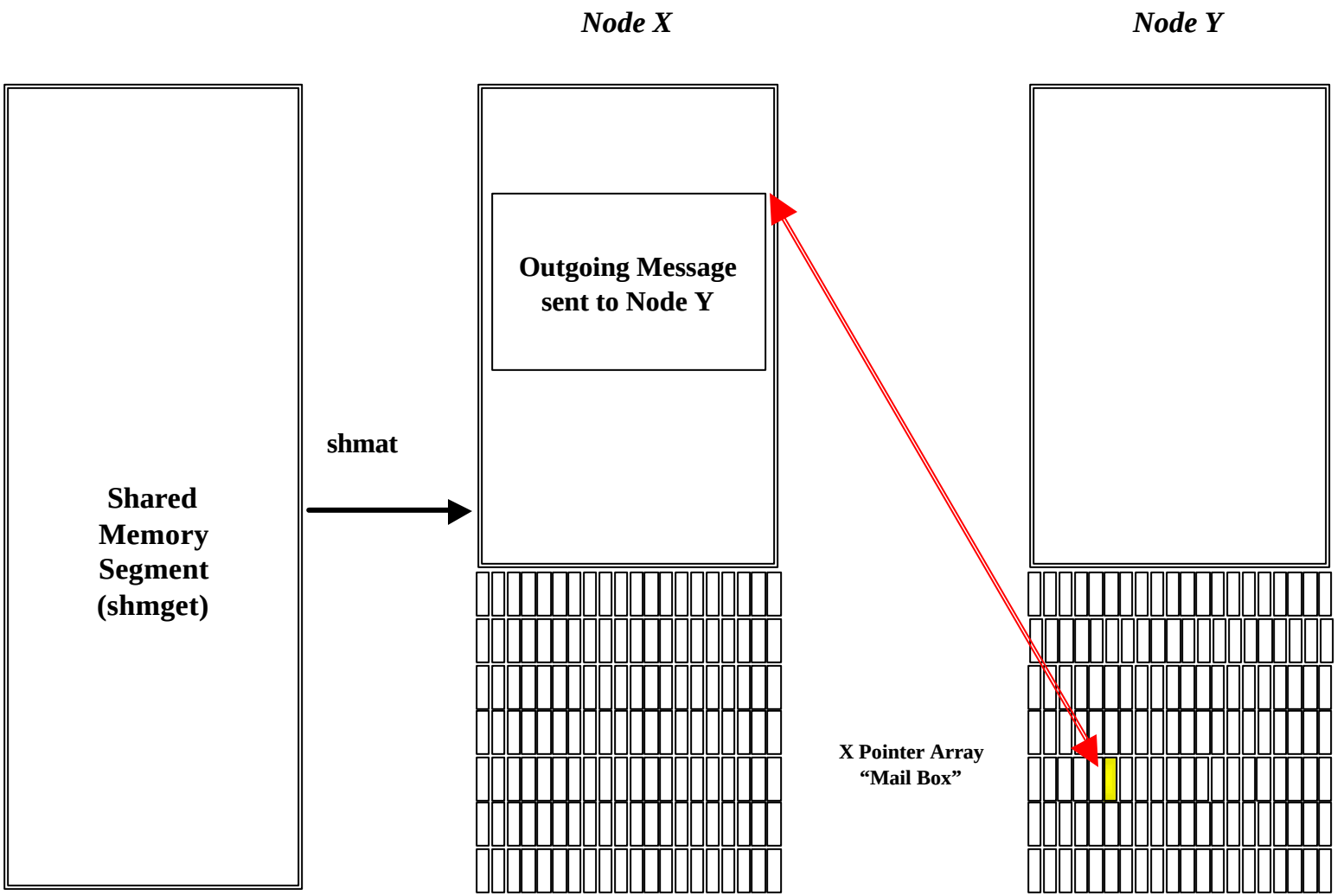


SPEEDES Communication Services Initialization

A quick overview of the SPEEDES Communications start up procedure to initialize the communication services …


Get
TCP/IP
processes
Test C3
Test C4
Start A
SpComm
SpComm_StartUp
Print
"I'm node" 0
Shmget
Shared Memory
Memory
Handles
A0
A1
memcopy
Fork
New processes
Memory
Handles
Start B2
Fork /Execvp
New processes
SpComm
SpComm_StartUp
Print
"I'm B on node" 2

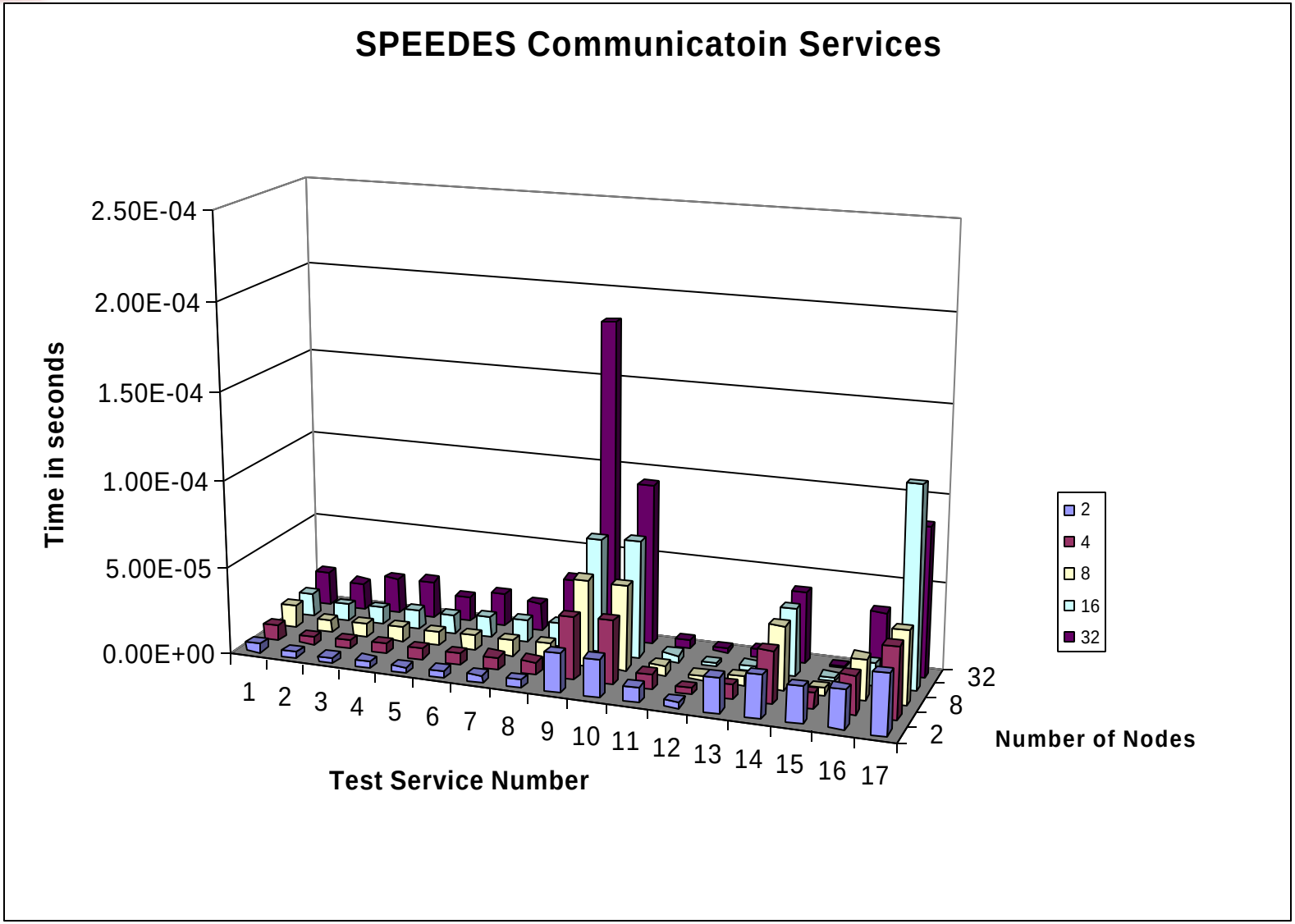
Shared Memory Communications



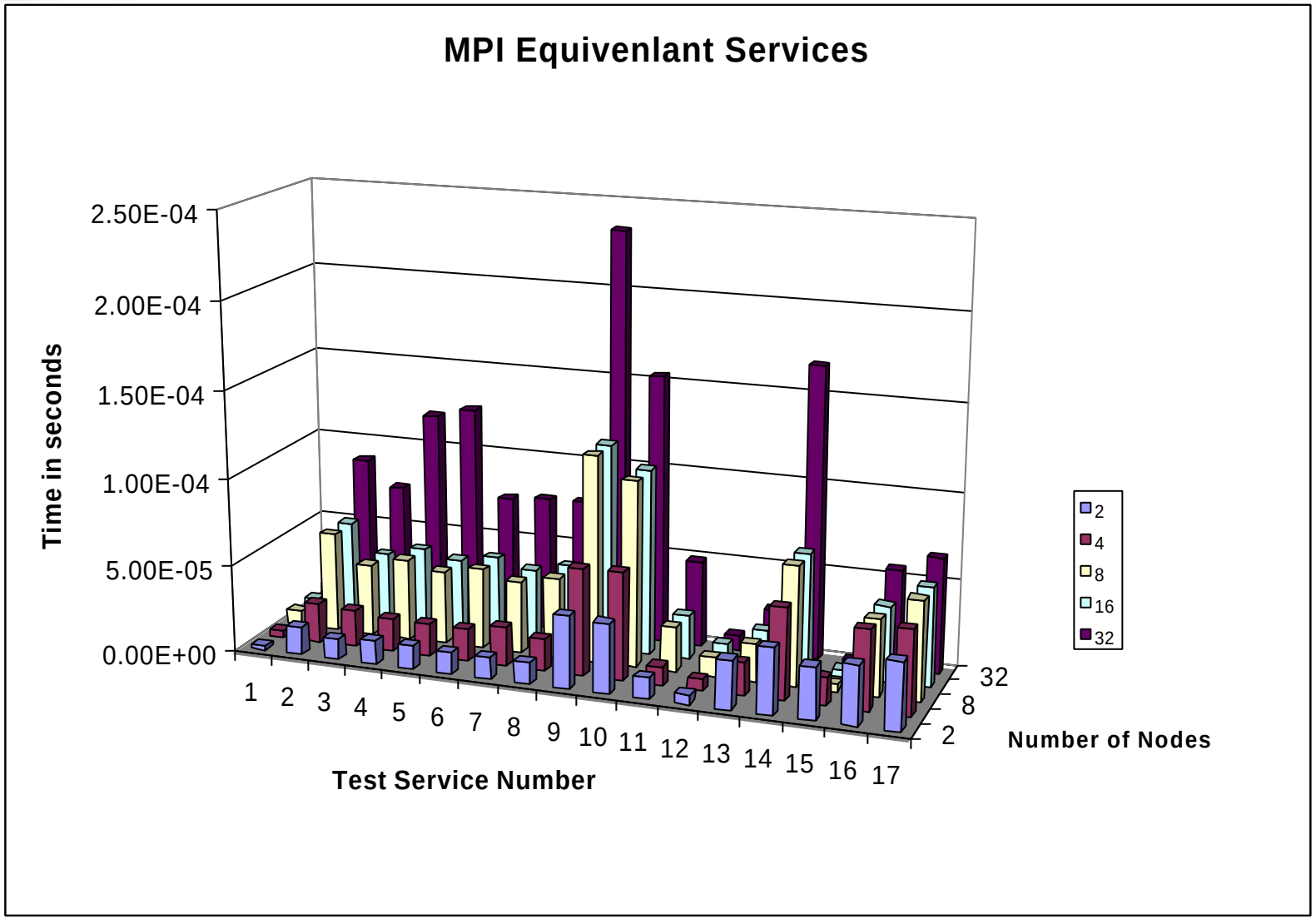
Tests and Timings

	SPEEDES					MPI			
Number of Nodes	2	4	8	16		2	4	8	16
1 SpComm_BarrierSync	5.70E-06	9.19E-06	1.35E-05	1.36E-05		2.40E-06	3.44E-06	8.12E-06	8.12E-06
2 SpComm_FuzzyBarrier	2.87E-06	4.87E-06	6.99E-06	9.58E-06		1.57E-05	2.26E-05	5.69E-05	5.69E-05
3 SpComm_GlobalSum(int)	3.17E-06	5.34E-06	7.42E-06	1.02E-05		1.21E-05	2.15E-05	4.06E-05	4.06E-05
4 SpComm_GlobalMin(int)	3.58E-06	6.34E-06	8.92E-06	1.16E-05		1.38E-05	1.89E-05	4.61E-05	4.61E-05
5 SpComm_GlobalMax(int)	3.36E-06	5.65E-06	7.82E-06	1.06E-05		1.32E-05	1.86E-05	4.20E-05	4.20E-05
6 SpComm_GlobalSum(double)	4.20E-06	5.98E-06	8.54E-06	1.26E-05		1.23E-05	1.87E-05	4.59E-05	4.59E-05
7 SpComm_GlobalMin(double)	4.26E-06	6.99E-06	8.54E-06	1.29E-05		1.23E-05	2.21E-05	4.08E-05	4.08E-05
8 SpComm_GlobalMax(double)	4.81E-06	6.99E-06	9.87E-06	1.35E-05		1.23E-05	1.81E-05	4.59E-05	4.59E-05
9 SpComm_GlobalMin(SpSimTime)	2.19E-05	3.55E-05	4.91E-05	6.53E-05		4.15E-05	6.12E-05	0.00012	0.00012
10 SpComm_GlobalMax(SpSimTime)	2.18E-05	3.57E-05	4.90E-05	6.67E-05		3.99E-05	6.17E-05	0.00011	0.00011
11 SpComm_send_receive_single_message()	9.06E-06	8.70E-06	5.76E-06	3.91E-06		1.25E-05	1.06E-05	2.53E-05	2.53E-05
12 SpComm_send_sync_message()	3.69E-06	3.54E-06	2.74E-06	2.20E-06		5.55E-06	6.21E-06	1.10E-05	1.10E-05
13 Multi-send/receive()	2.04E-05	8.16E-06	5.31E-06	3.43E-06		2.79E-05	1.86E-05	2.16E-05	2.16E-05
14 Time for TestBroadCast()	2.46E-05	2.95E-05	3.60E-05	3.83E-05		3.77E-05	5.20E-05	6.81E-05	6.81E-05
15 Time for Multi-send/receive()	2.12E-05	9.48E-06	4.42E-06	2.36E-06		2.99E-05	1.62E-05	4.25E-06	4.25E-06
16 Time for TestCoordinatedBroadCast()	2.24E-05	2.22E-05	2.35E-05	1.37E-05		3.38E-05	4.54E-05	4.34E-05	4.34E-05
17 Latency for a message of length 500	3.43E-05	4.06E-05	4.20E-05	0.00011		3.85E-05	4.85E-05	5.69E-05	5.69E-05

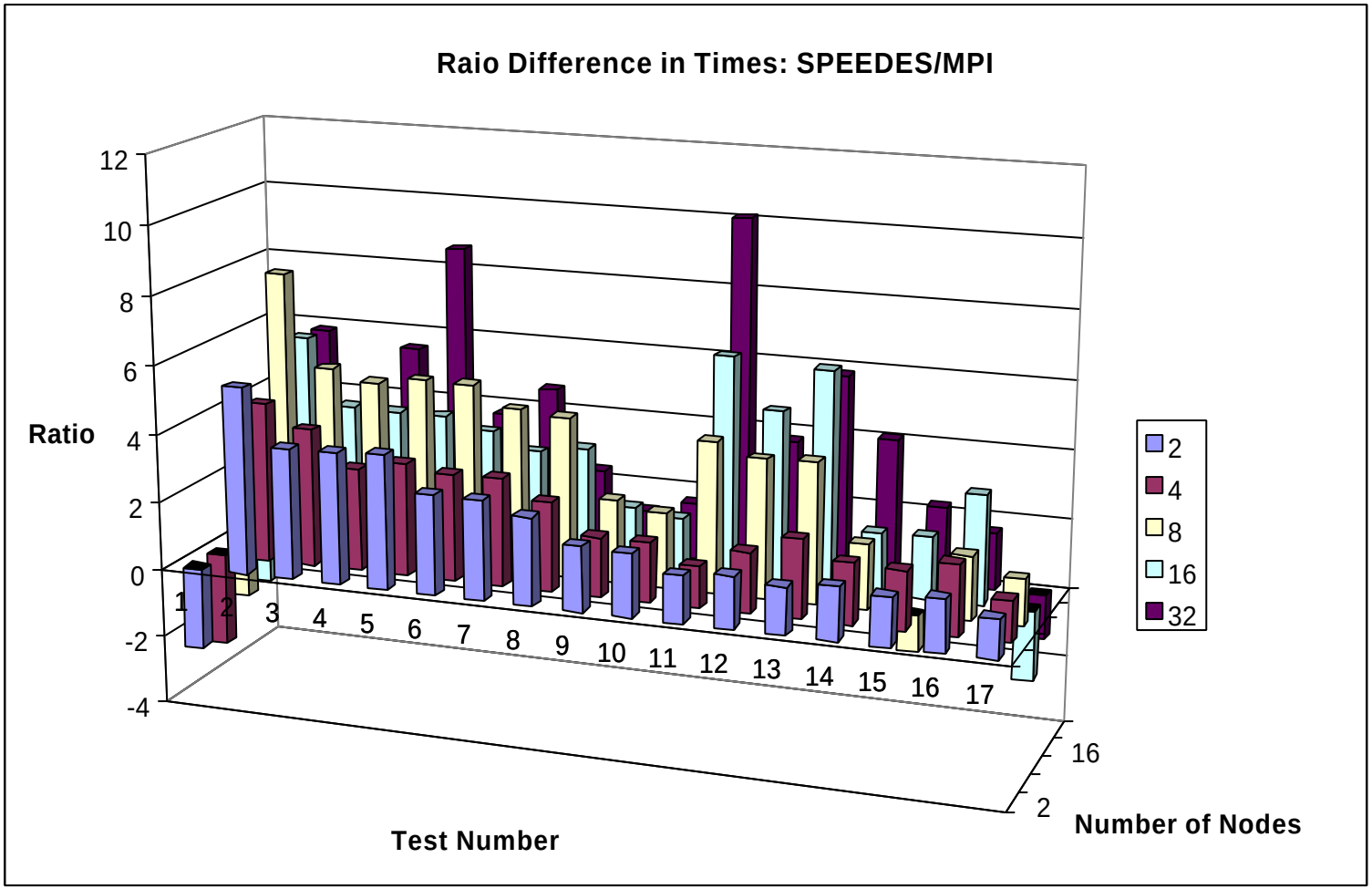
Verifications and Timings



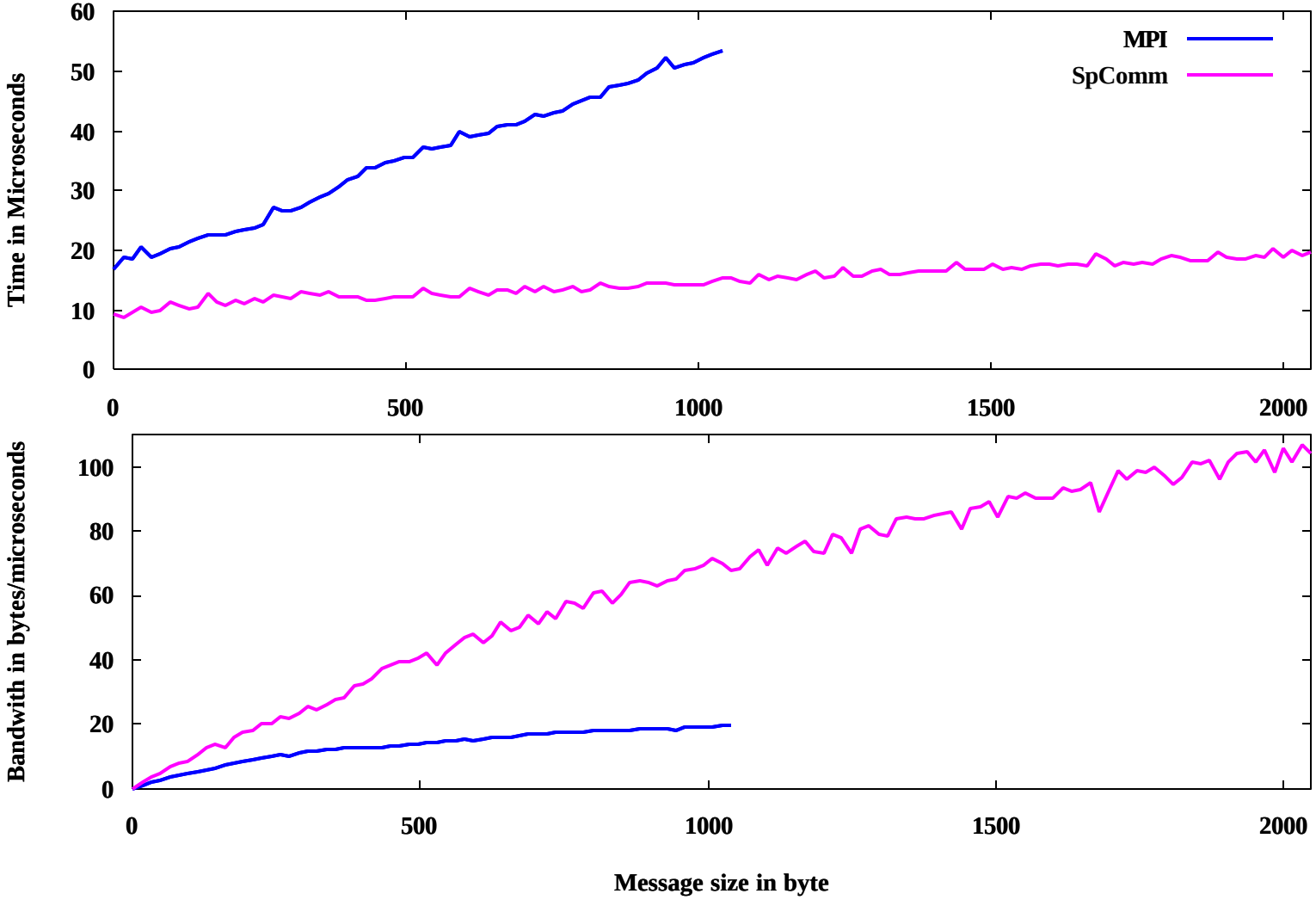
Verifications and Timings



Verifications and Timings



Point-to-Point / Bandwidth
Sun Ultra Enterprise 10000



Conclusion

- The SPEEDES Communications Services:
  - Are tailored to meet the specific needs of discrete event simulations
  - Sends point-to-point messages immediately and are: buffered asynchronous / multi-cast / coordinated
  - Have been updated to include multiple applications using shared memory communications
- SPEEDES can inherently support an HPC RTI
  - Object Management (Declaration, Ownership, Distribution)
  - Shared Memory and Network Communications
- Possibility of continued testing against MPI-2