Distributed Algorithms

CS 308 Compiler Techniques

160001026 - Niranjan Joshi

160001028 - Kanishkar J

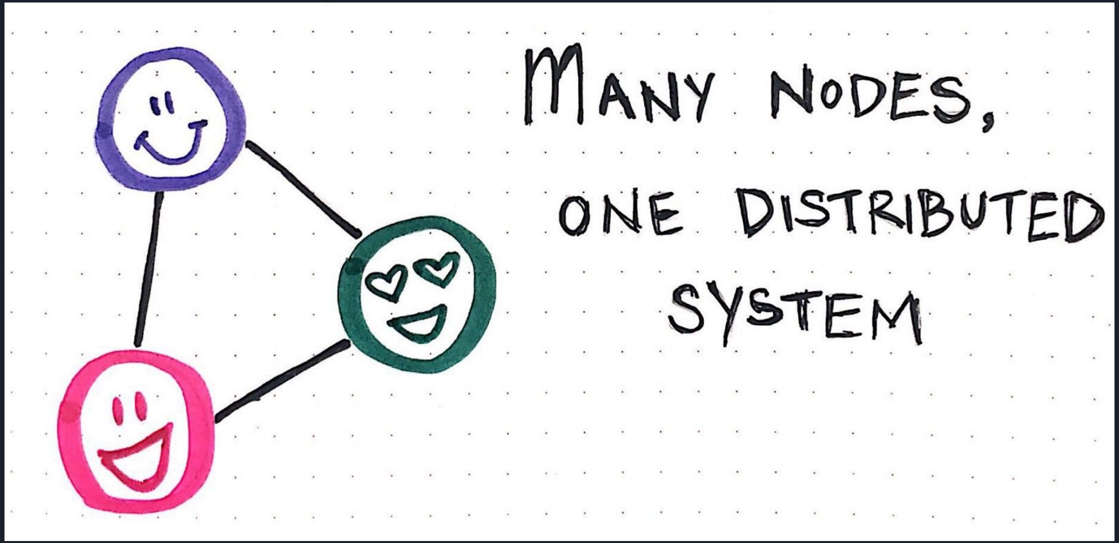
160001048 - Rishabh Kumar Verma

160001052 - Sahaj Khandelwal

Distributed Systems : Overview

- A model in which components located on networked computers communicate and coordinate their actions by passing messages.

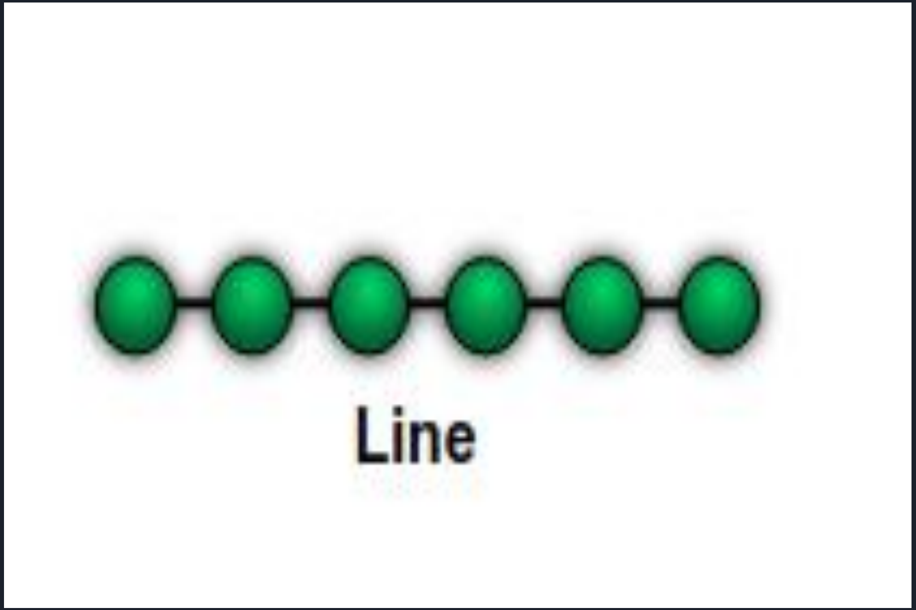
Distributed Systems : Overview



Source : medium.com

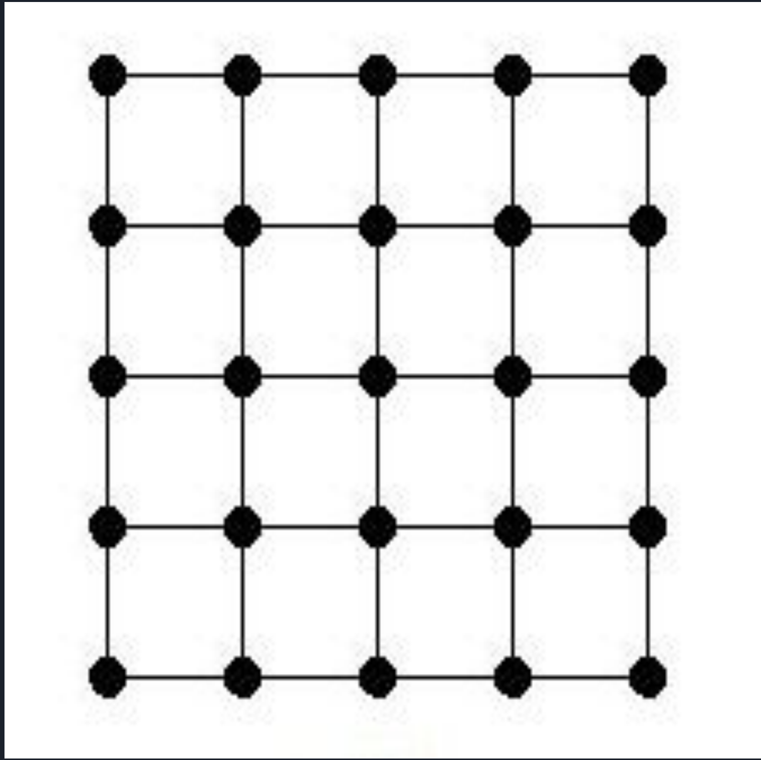
Network Topologies

Linear



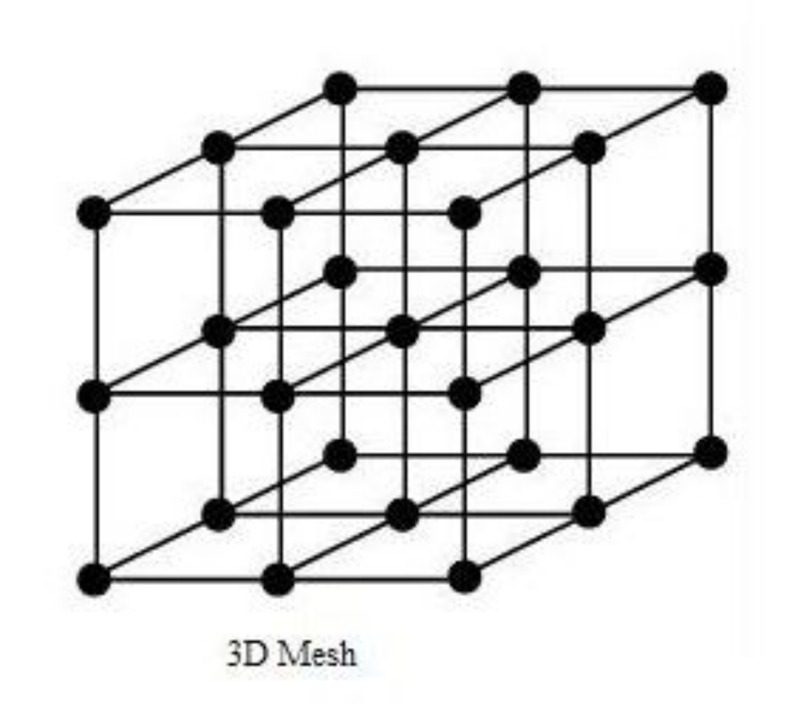
Source: Wikipedia

Grid



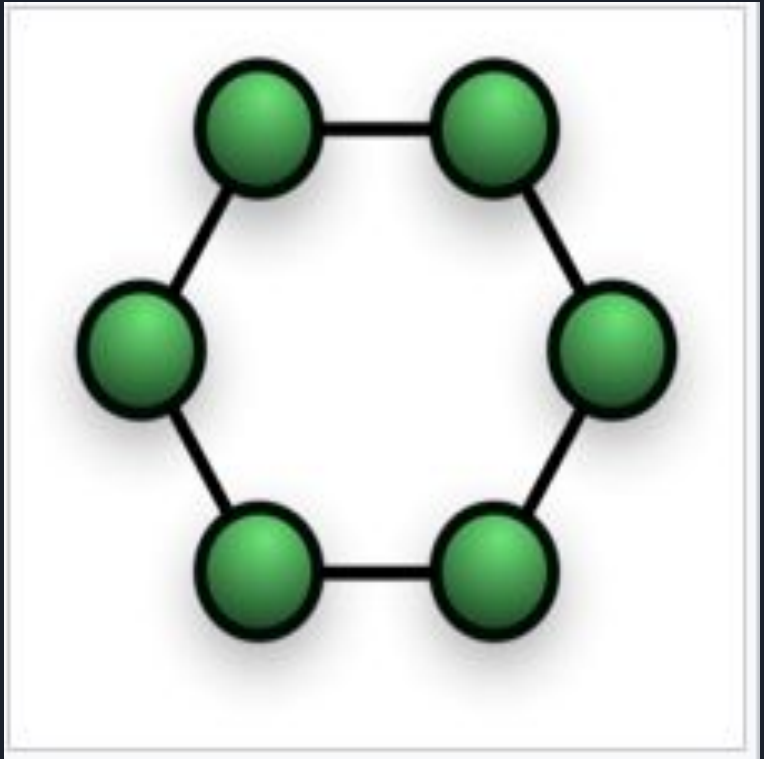
Source: http://wiki.expertiza.ncsu.edu

Mesh 3D



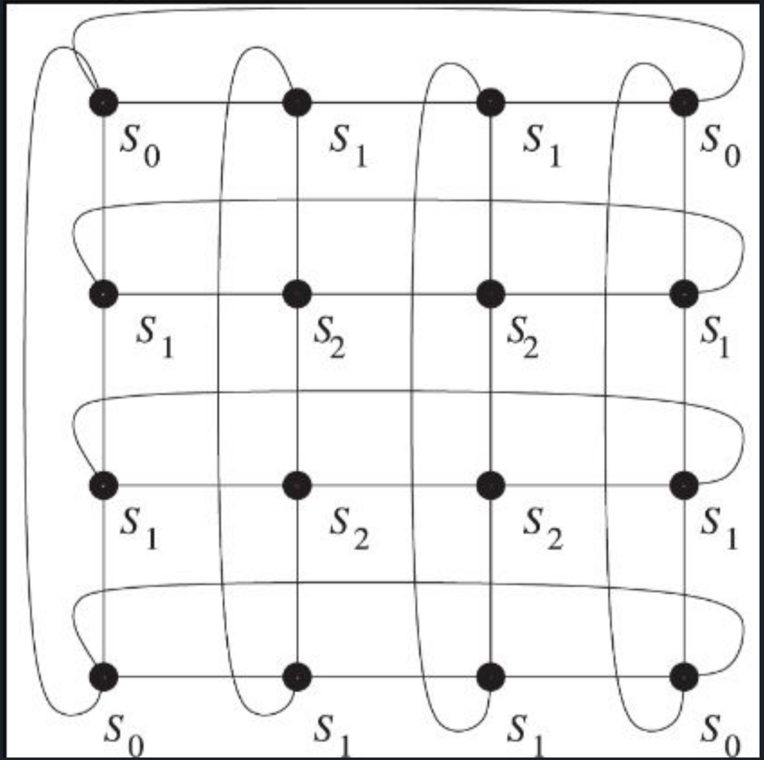
Source: http://wiki.expertiza.ncsu.edu

Ring



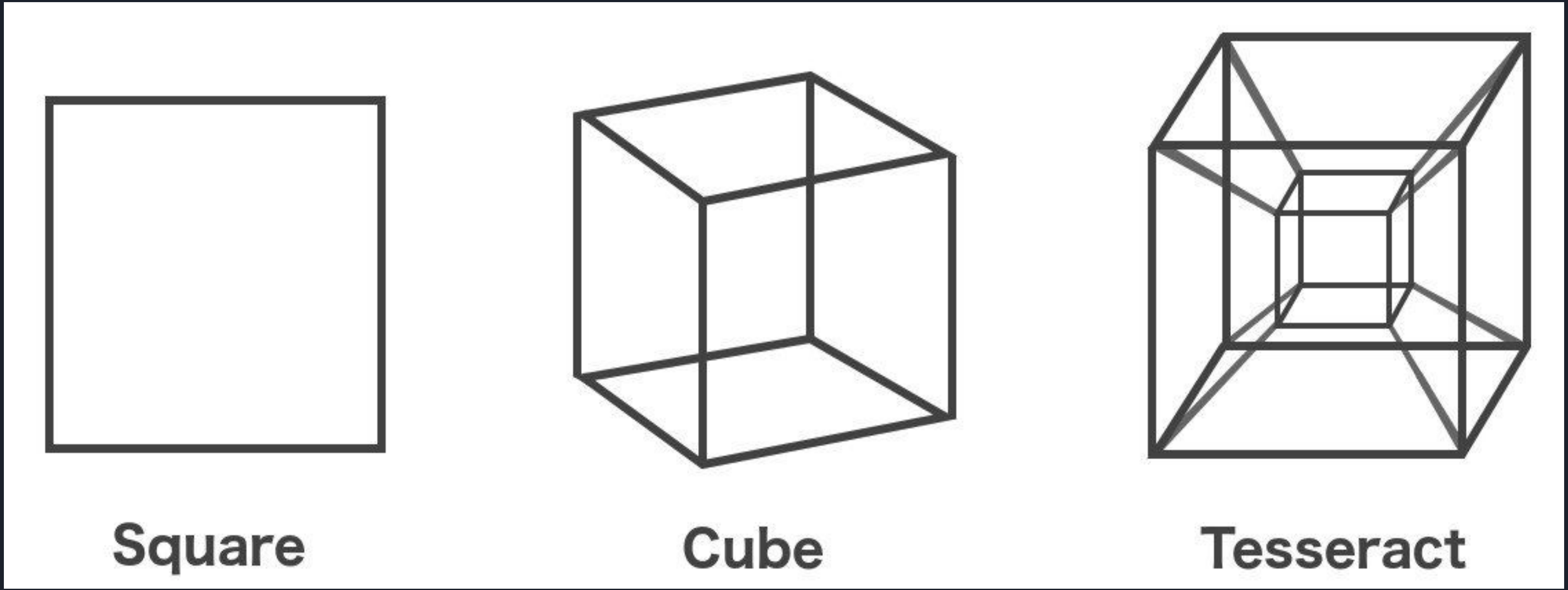
Source: Wikipedia

Torus 2D

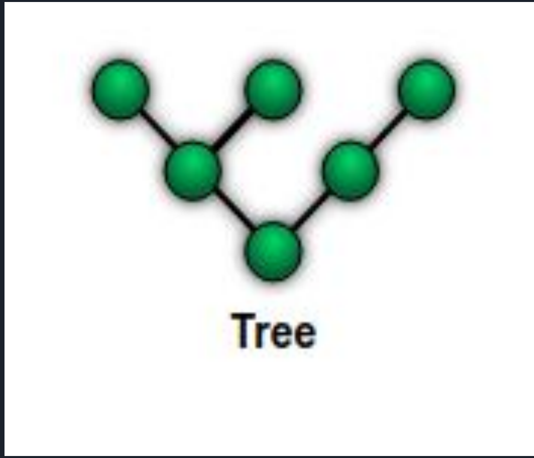


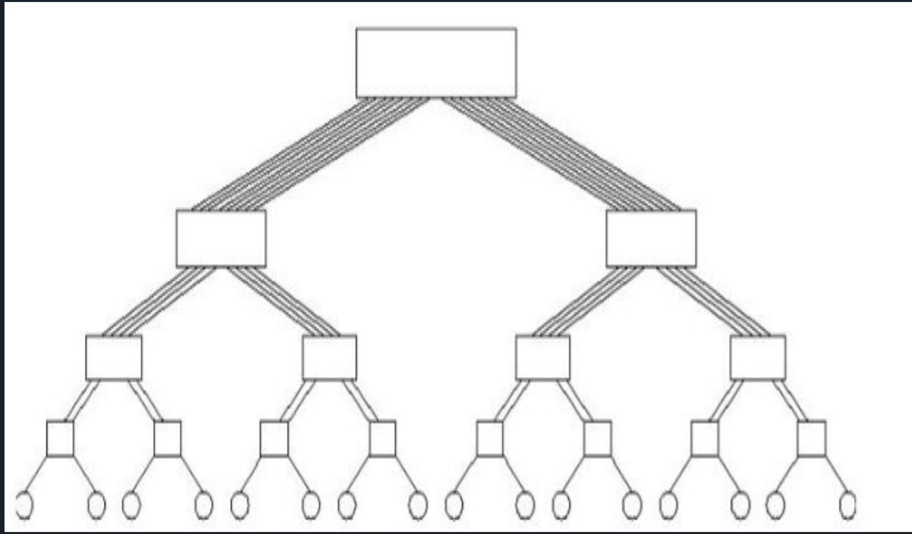
Source: www.computer.org

Hypercubes



Tree And Fat Tree

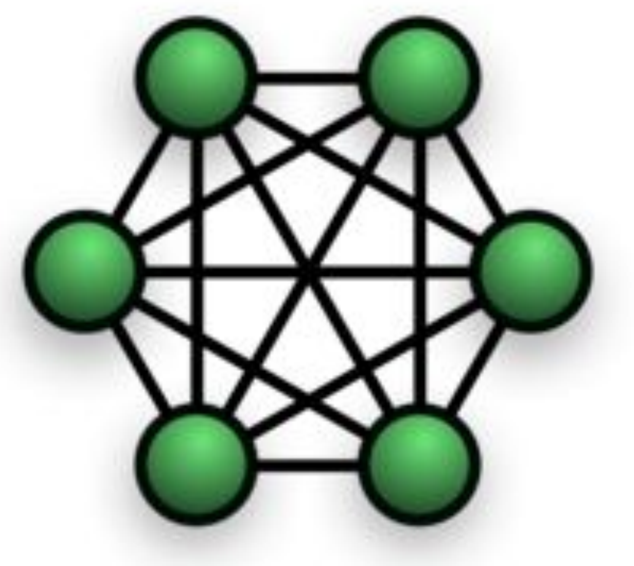




Fat Tree

Source: Ashley Dickerson

Fully Connected Graph



Fully connected mesh topology

Source: Wikipedia

Topology Parameters

- Number of Links
- Diameter
- Bisection Width
- Bisection Bandwidth
- Congestion

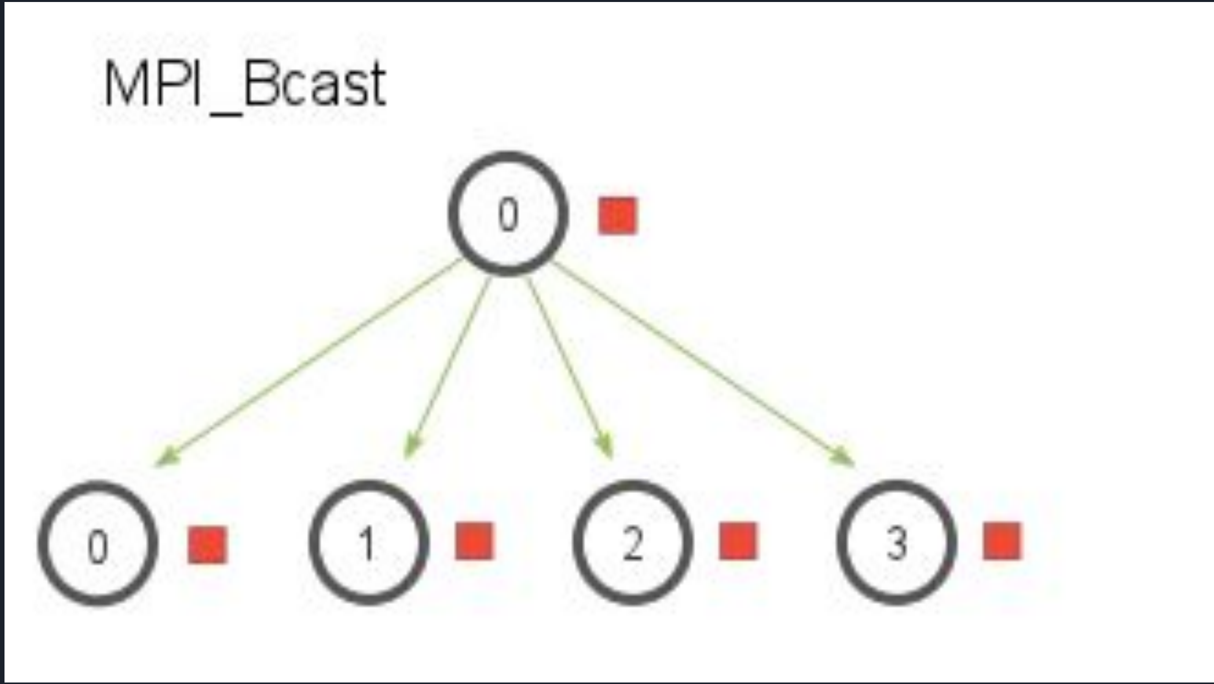
$$C = \Omega(B_L/B_p)$$

Topology Related Complexities

Topology	Metric		
	# Links	Diameter	Bisection Width
Linear	$p-1$	$p-1$	1
2D Mesh	$2p-2\sqrt{p}$	$2(\sqrt{p}-1)$	$\sqrt{p}$
kD Mesh	$kp-kp^{\left(\frac{k-1}{k}\right)}$	$kp^{\frac{1}{k}}$	$p^{\left(\frac{k-1}{k}\right)}$
Ring	p	$p/2$	2
2D Torus (Doughnut)	$2p$	$\sqrt{p}-1$	$2\sqrt{p}$
kD Torus	kp	$\frac{k}{2}(p^{\frac{1}{k}}-1)$	$2p^{\left(\frac{k-1}{k}\right)}$
$log(p)$ D Hypercube	$\frac{p\,log(p)}{2}$	$log(p)$	$p/2$
Binary Tree	$p-1$	$2log(p)$	1
Fully Connected	$\frac{p\,(p-1)}{2}$	1	$\frac{p^2}{4}$

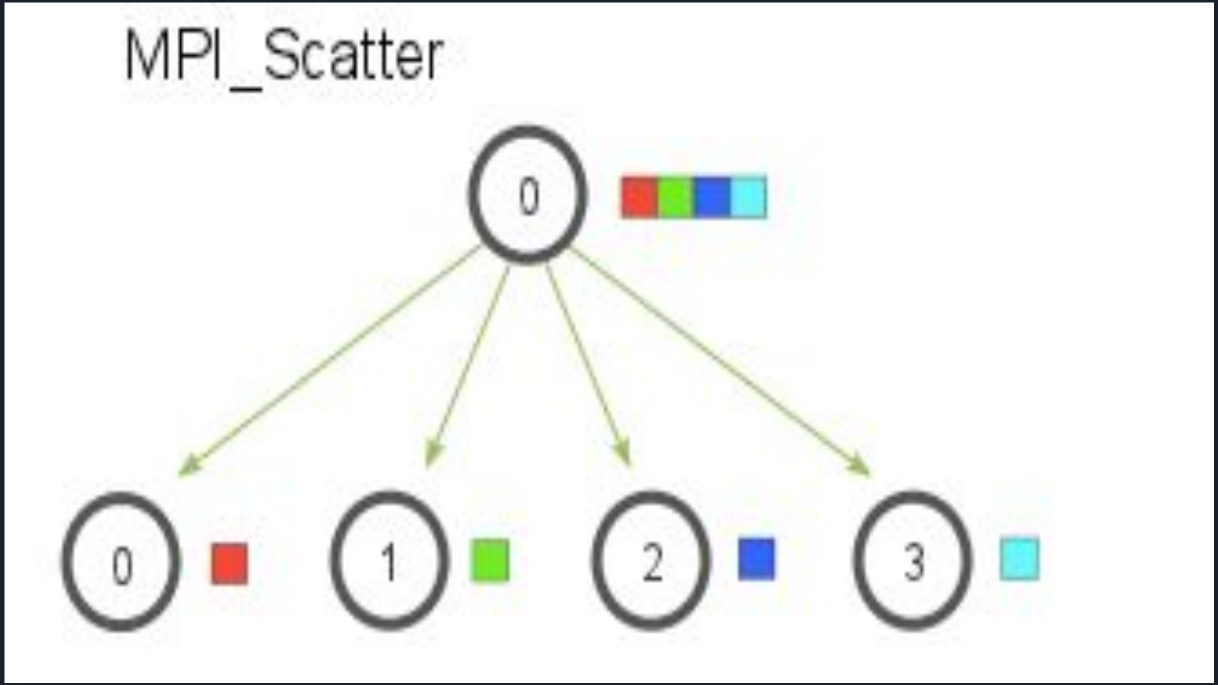
Distributed Algorithms

Broadcast Algorithm



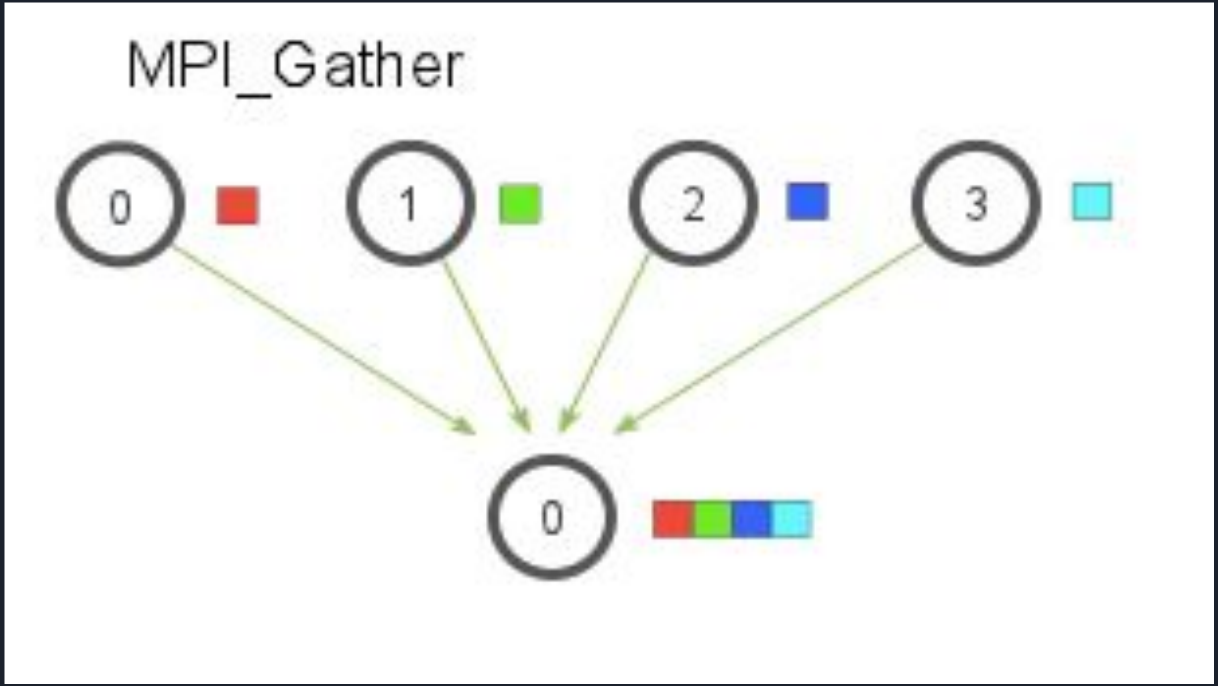
Source: mpitutorial.com

Scatter Algorithm



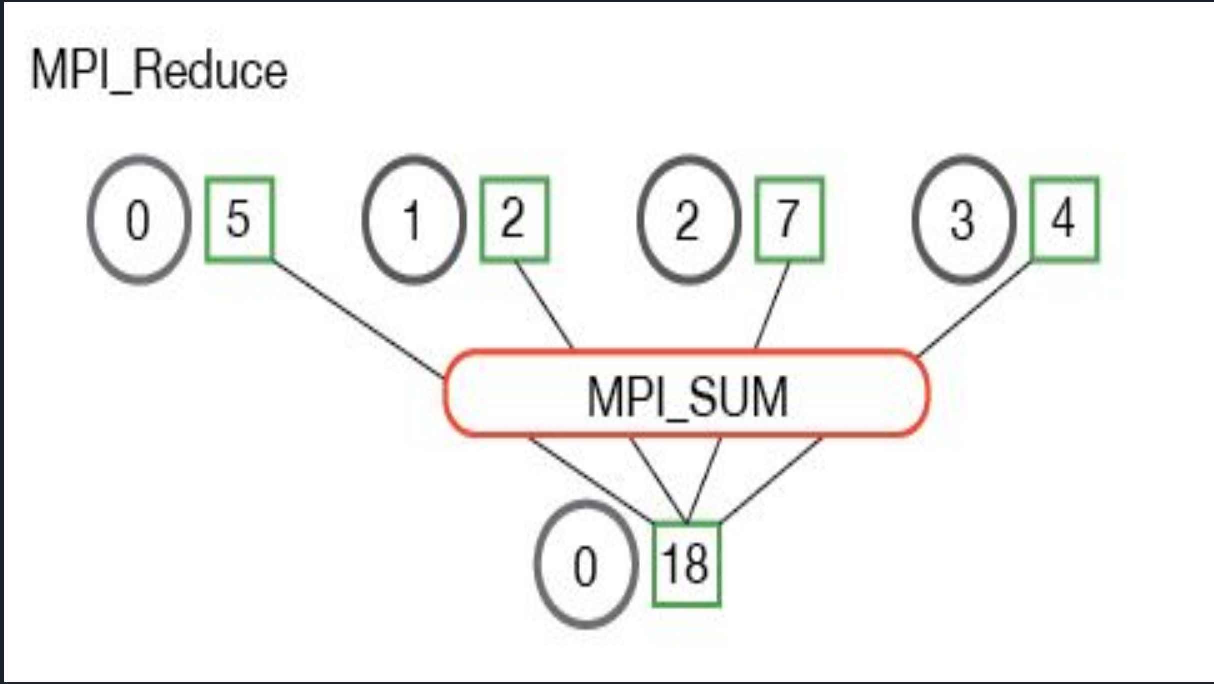
Source: mpitutorial.com

Gather Algorithm



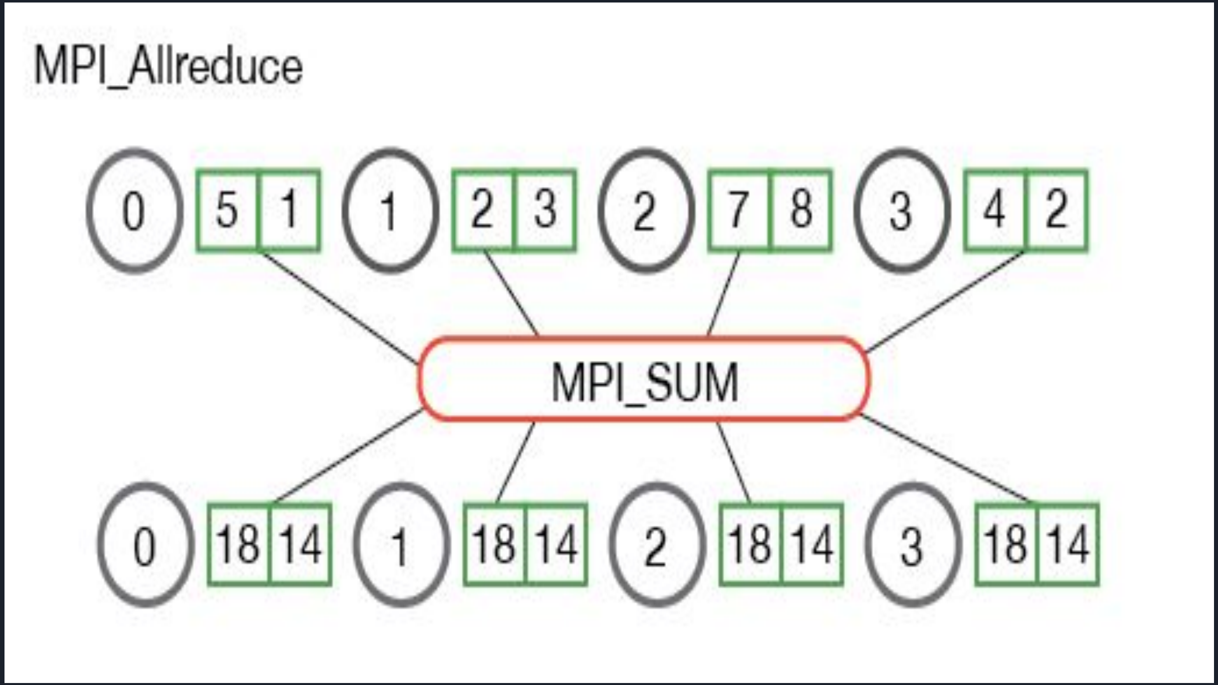
Source: mpitutorial.com

Reduce Algorithm



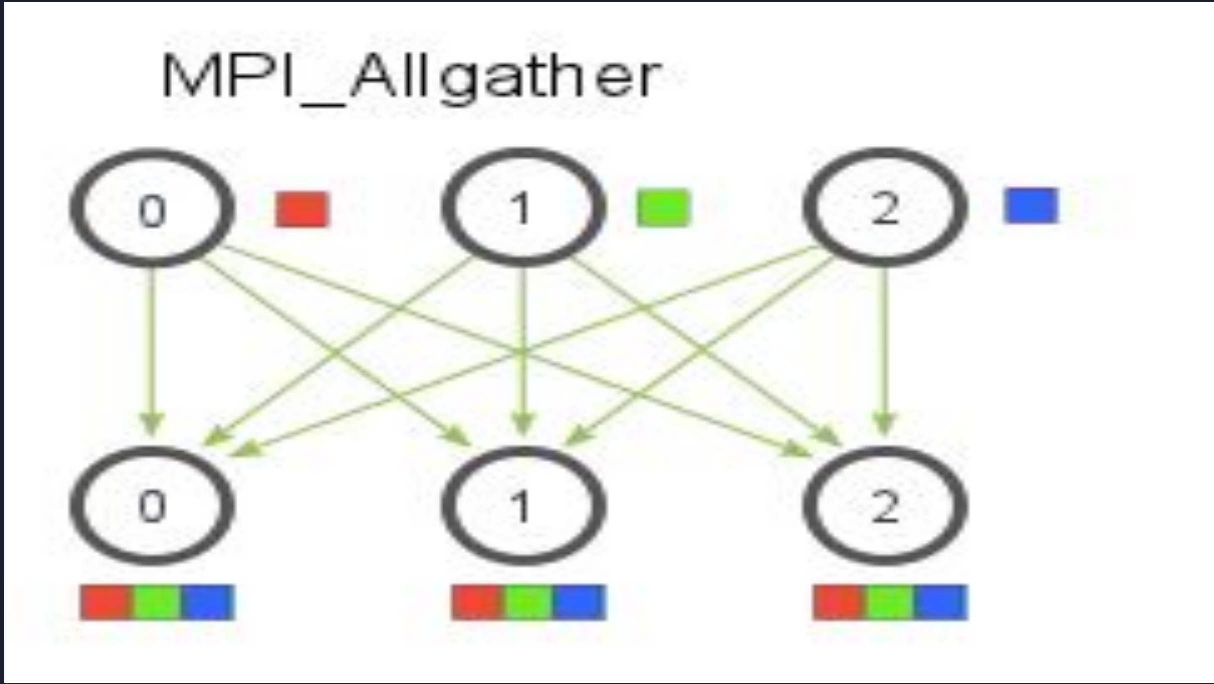
Source: mpitutorial.com

All Reduce Algorithm



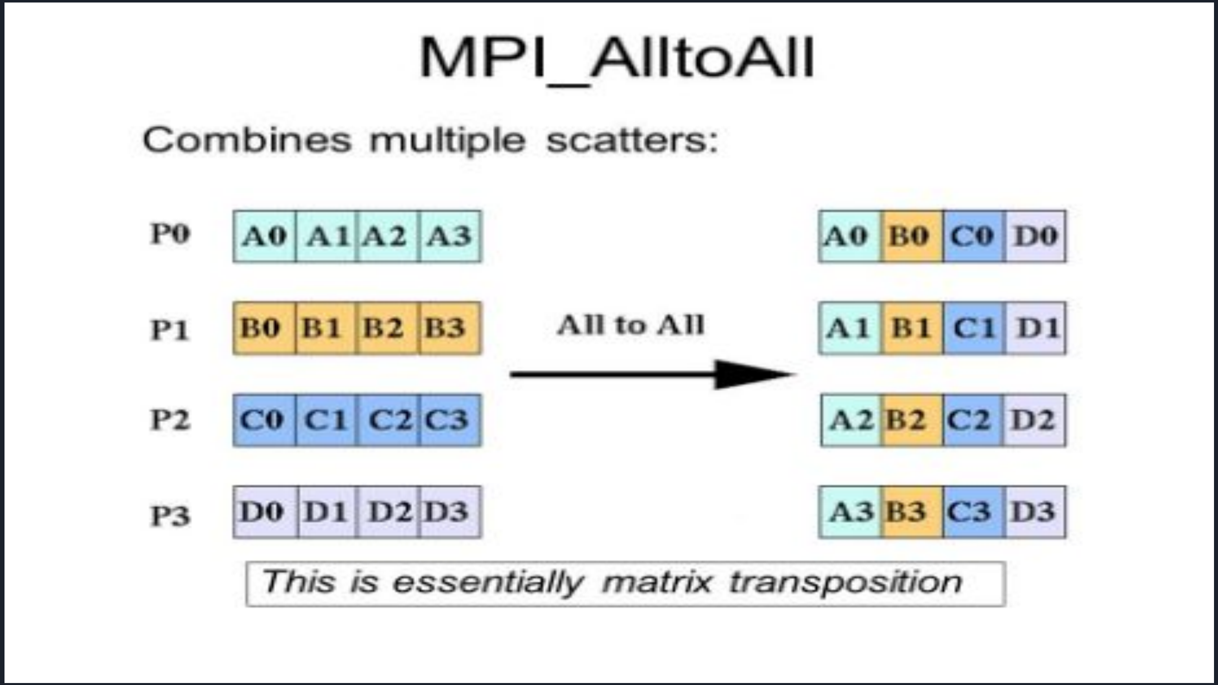
Source: mpitutorial.com

All Gather Algorithm



Source: mpitutorial.com

All-to-All Algorithm



Source: UNC-Charlotte

Algorithm Complexity Parameters

- α = latency (units of time)
- β = inverse bandwidth (units of time/word)
- Time to send a message of n-words between two nodes :

$$T(n) = \alpha + \beta n$$

Algorithm Complexity Parameters

- Time during K-way congestion :

$$T(n) = \alpha + \beta n k$$

Time Complexities

Algorithm	Topology	Time Complexity
One-to-all Broadcast All-to-one Reduction	*Linear*	$(\alpha + \beta m)(p)$
	Ring	$(\alpha + \beta m) \, log(p)$
	Mesh	$(\alpha + \beta m) \, log(p)$
	Torus (2-D)	$(\alpha + \beta m) \, log(p)$
	Hypercube	$(\alpha + \beta m) \, log(p)$
	Tree	$(\alpha + \beta m) \, p$
	FCG	$(\alpha + \beta m) \, log(p)$

Time Complexities

All-to-all Broadcast All-to-all Reduction	*Linear*	$(\alpha + \beta m)(p - 1)$
	Ring	$(\alpha + \beta m)(p - 1)$
	Mesh	$2\alpha(\sqrt{p} - 1) + \beta m(p - 1)$
	Torus (2-D)	$2\alpha(\sqrt{p} - 1) + \beta m(p - 1)$
	Hypercube	$\alpha log(p) + \beta m(p - 1)$
	Tree	$(\alpha + \beta m)(p - 1)$
	FCG	$(\alpha + \beta m)(p - 1)$

Time Complexities

All Reduce	*Linear*	$(\alpha + \beta m)\ log(p)$
	Ring	$(\alpha + \beta m)\ log(p)$
	Mesh	$(\alpha + \beta m)\ log(p)$
	Torus (2-D)	$(\alpha + \beta m)\ log(p)$
	Hypercube	$(\alpha + \beta m)\ log(p)$
	Tree	$(\alpha + \beta m)\ p$
	FCG	$(\alpha + \beta m)\ log(p)$

Time Complexities

Scatter Gather	*Linear*	$\alpha log(p) + \beta m(p-1)$
	Ring	$\alpha log(p) + \beta m(p-1)$
	Mesh	$\alpha log(p) + \beta m(p-1)$
	Torus (2-D)	$\alpha log(p) + \beta m(p-1)$
	Hypercube	$\alpha log(p) + \beta m(p-1)$
	Tree	$\alpha log(p) + \beta m(p-1)$
	FCG	$\alpha log(p) + \beta m(p-1)$

Time Complexities

Scatter-all Gather-all	*Linear*	$(\alpha + \beta m\frac{p}{2})(p-1)$
	Ring	$(\alpha + \beta m\frac{p}{2})(p-1)$
	Mesh	$(2\alpha + \beta mp)(\sqrt{p}-1)$
	Torus (2-D)	$(2\alpha + \beta mp)(\sqrt{p}-1)$
	Hypercube	$(\alpha + \beta m)(p-1)$
	Tree	$(\alpha + \beta m\frac{p}{2})(p-1)$
	FCG	$(\alpha + \beta m\frac{p}{2})(p-1)$

Simulations & Plots

MPI

- **Message Passing Interface (MPI)** is a standardized and portable message-passing standard designed by a group of researchers from academia and industry to function on a wide variety of parallel computing architectures.

SimGrid

SimGrid is a scientific instrument to study the behavior of large-scale distributed systems such as Grids, Clouds, HPC or P2P systems.

Simgrid and its features

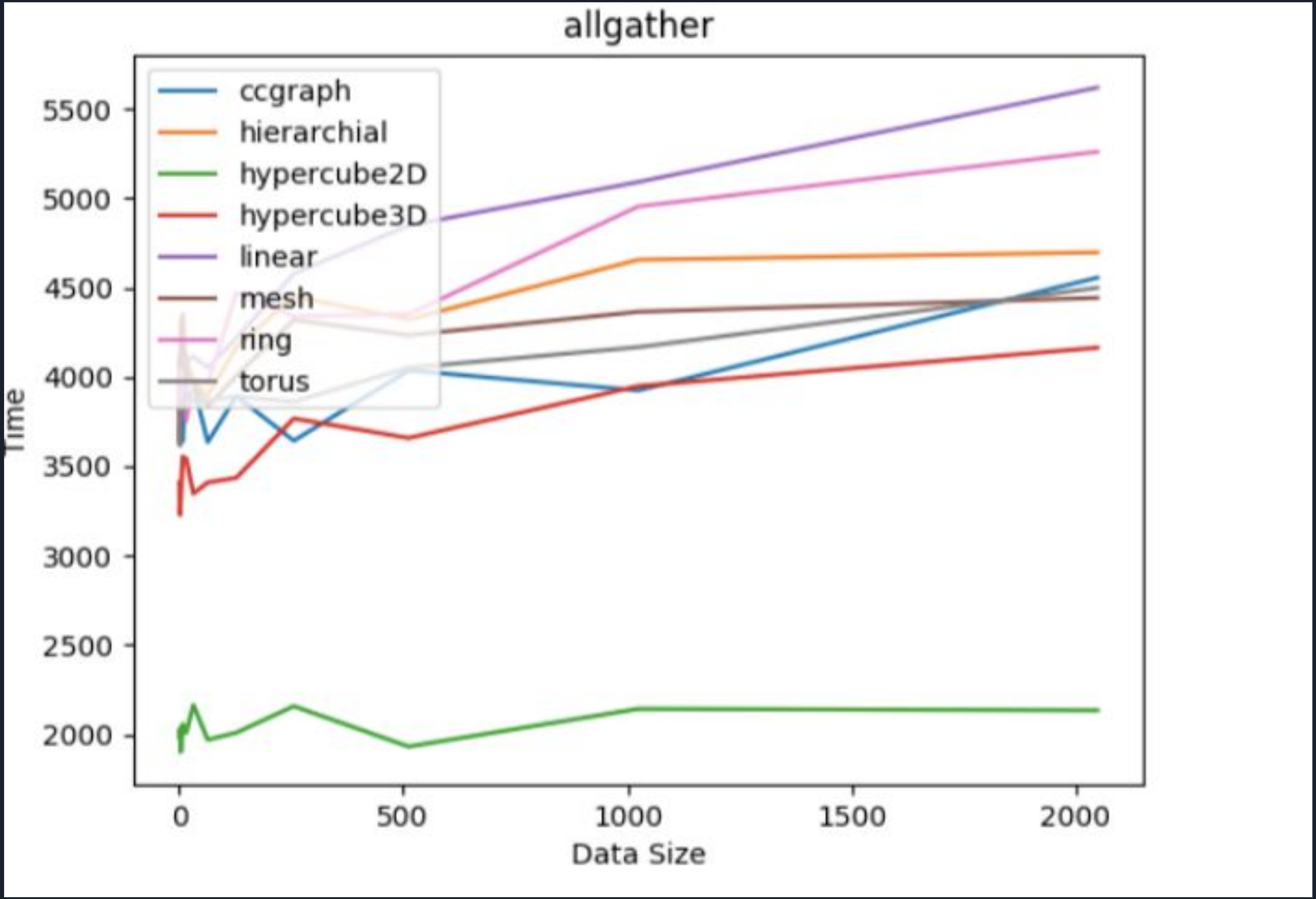
- Allows one to configure :
    - Per host computing power
    - Link latency and bandwidth
    - Simulate MPI programs without modifications
    - Simulate on different architectures

Simgrid and its features

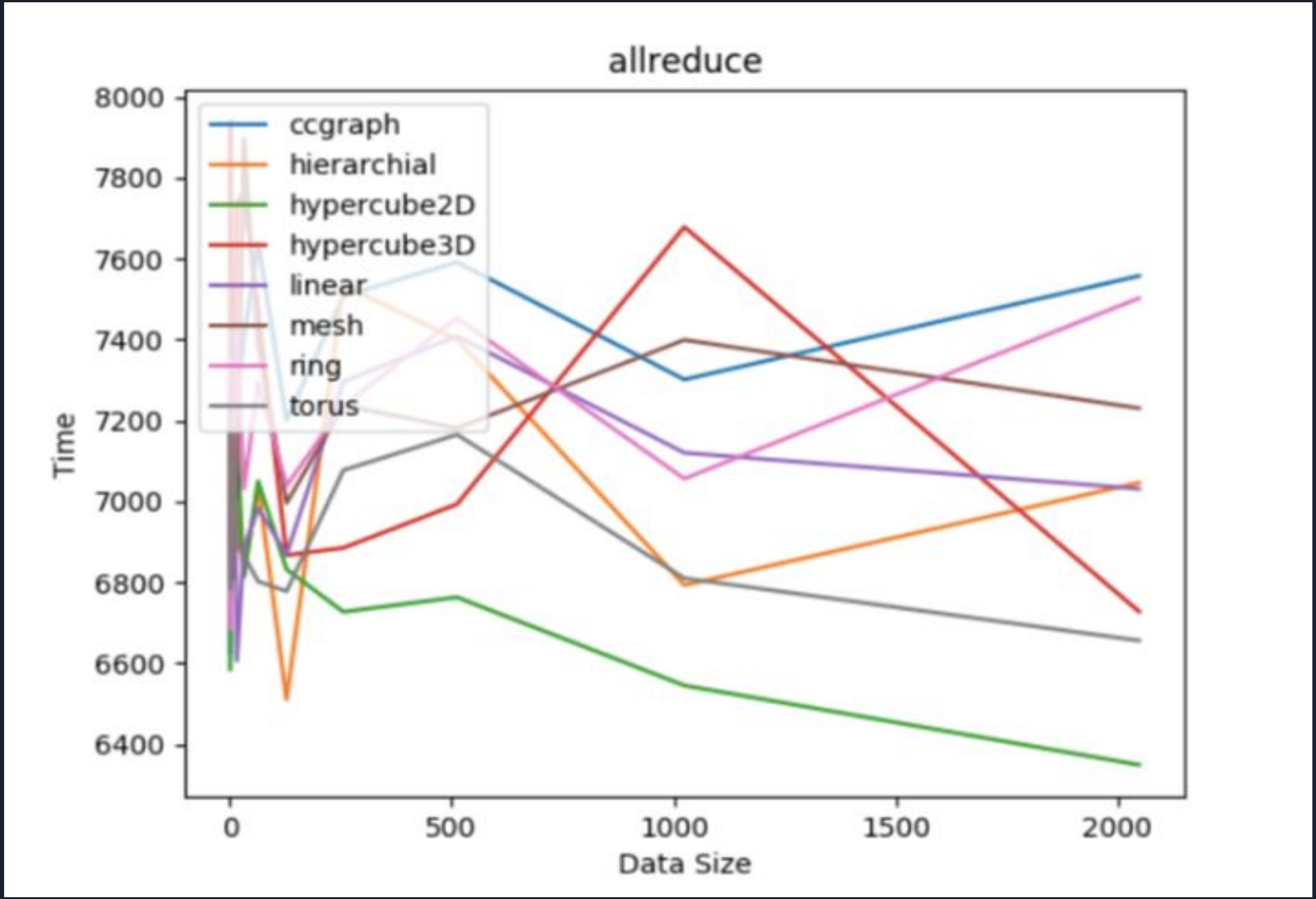
- Provides one with :
  - Link to link message count
  - Power consumption node wise
  - Execution time
- The trace information provided by Simgrid can be visualized using a tool pajeng.

Sample Trace Data after filtering

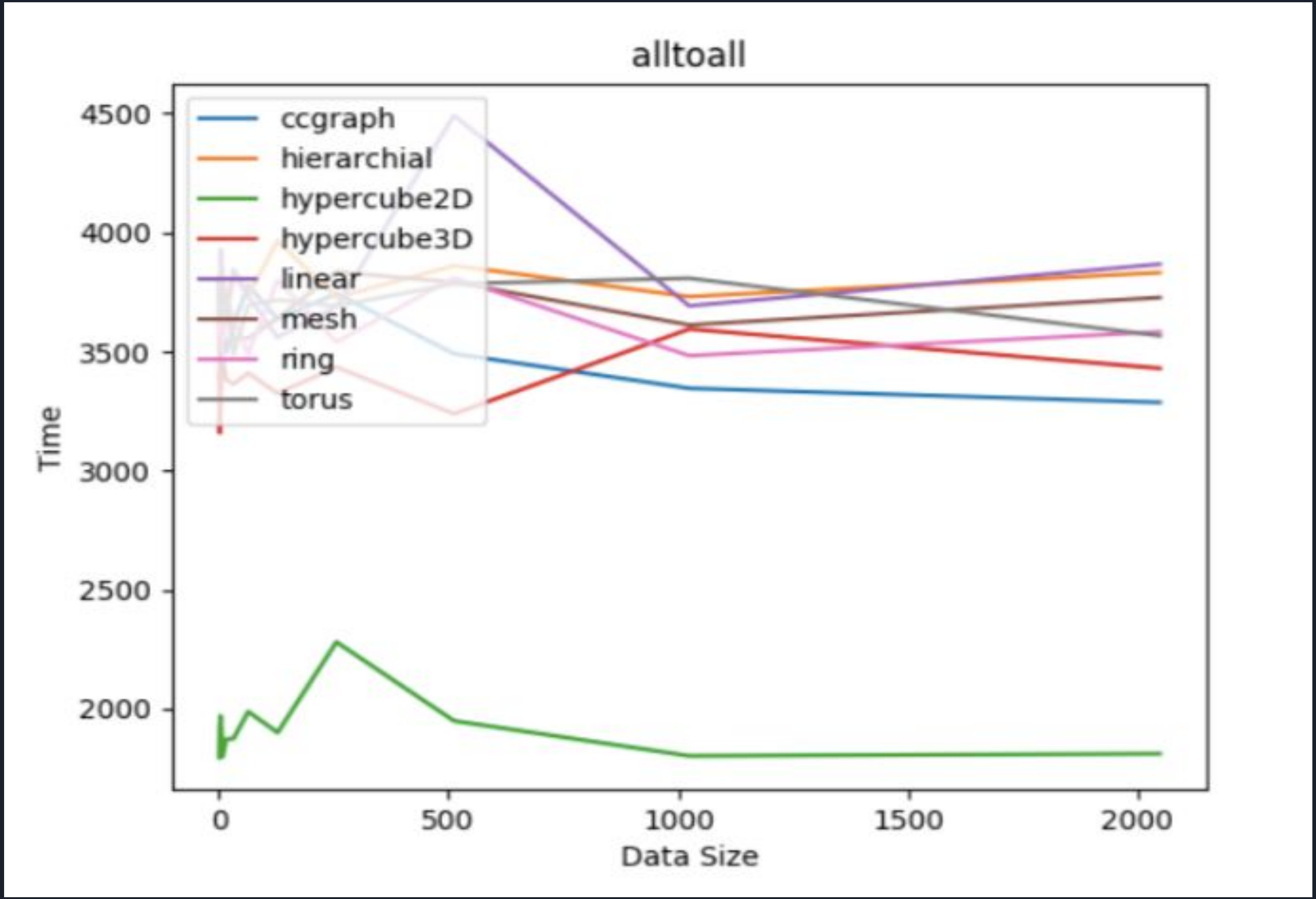
Source	Dest	Hop Count
1	2	4
1	9	1
1	3	3
3	4	5
1	5	2
5	6	7
5	7	6
7	8	8

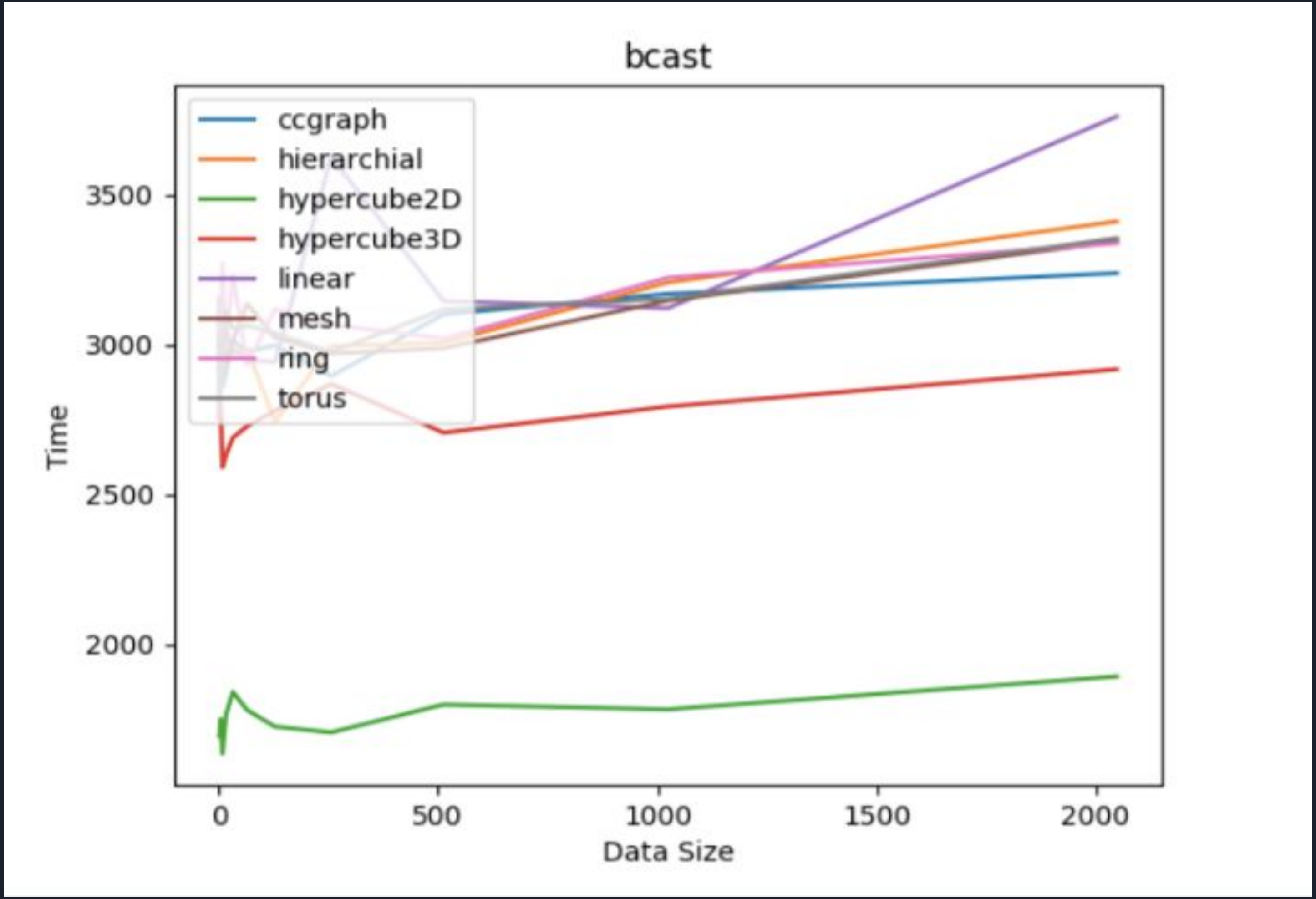


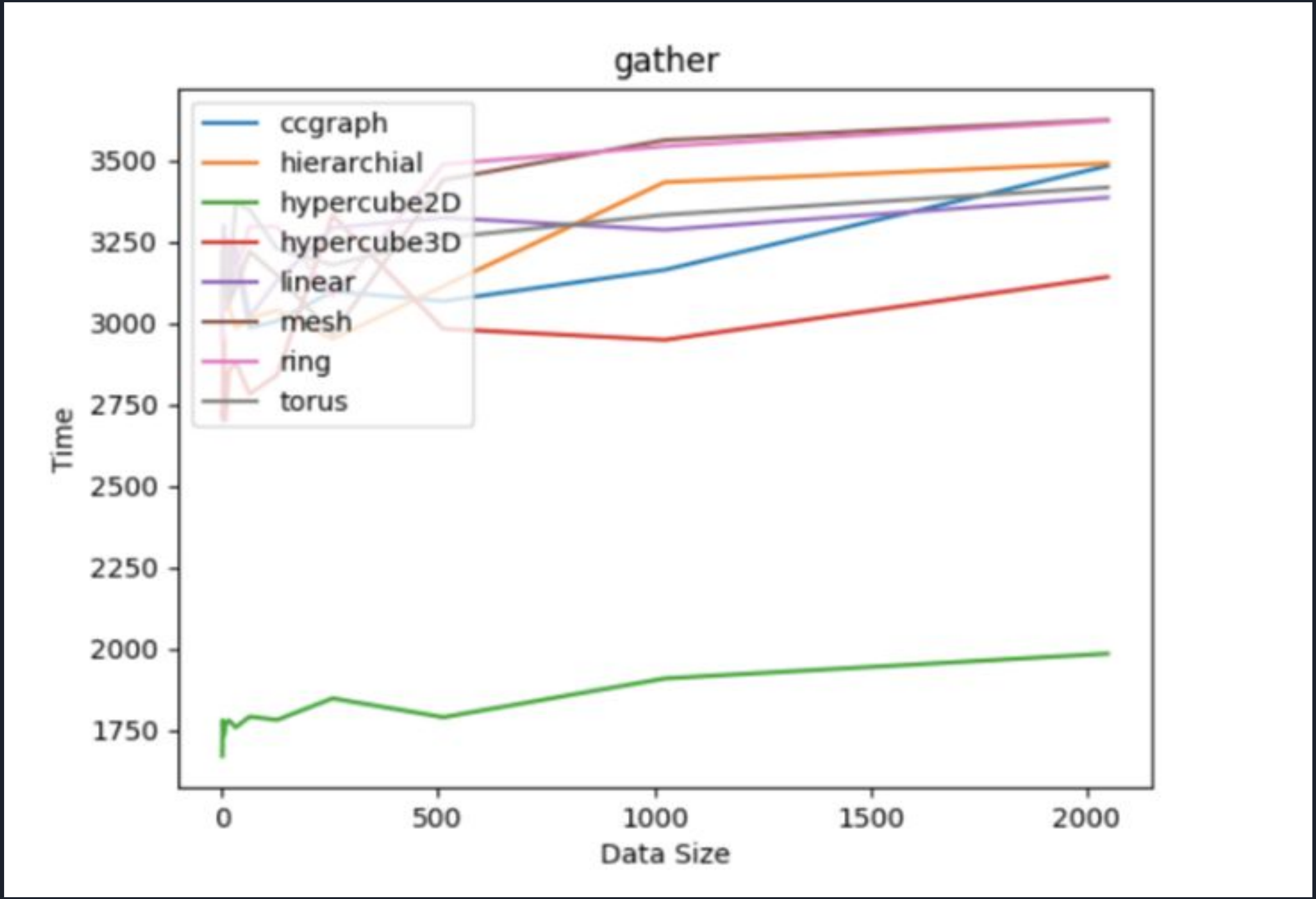
allgather
ccgraph
hierarchial
hypercube2D
hypercube3D
linear
mesh
ring
torus
Time
Data Size
5500
5000
4500
4000
3500
3000
2500
2000
0
500
1000
1500
2000

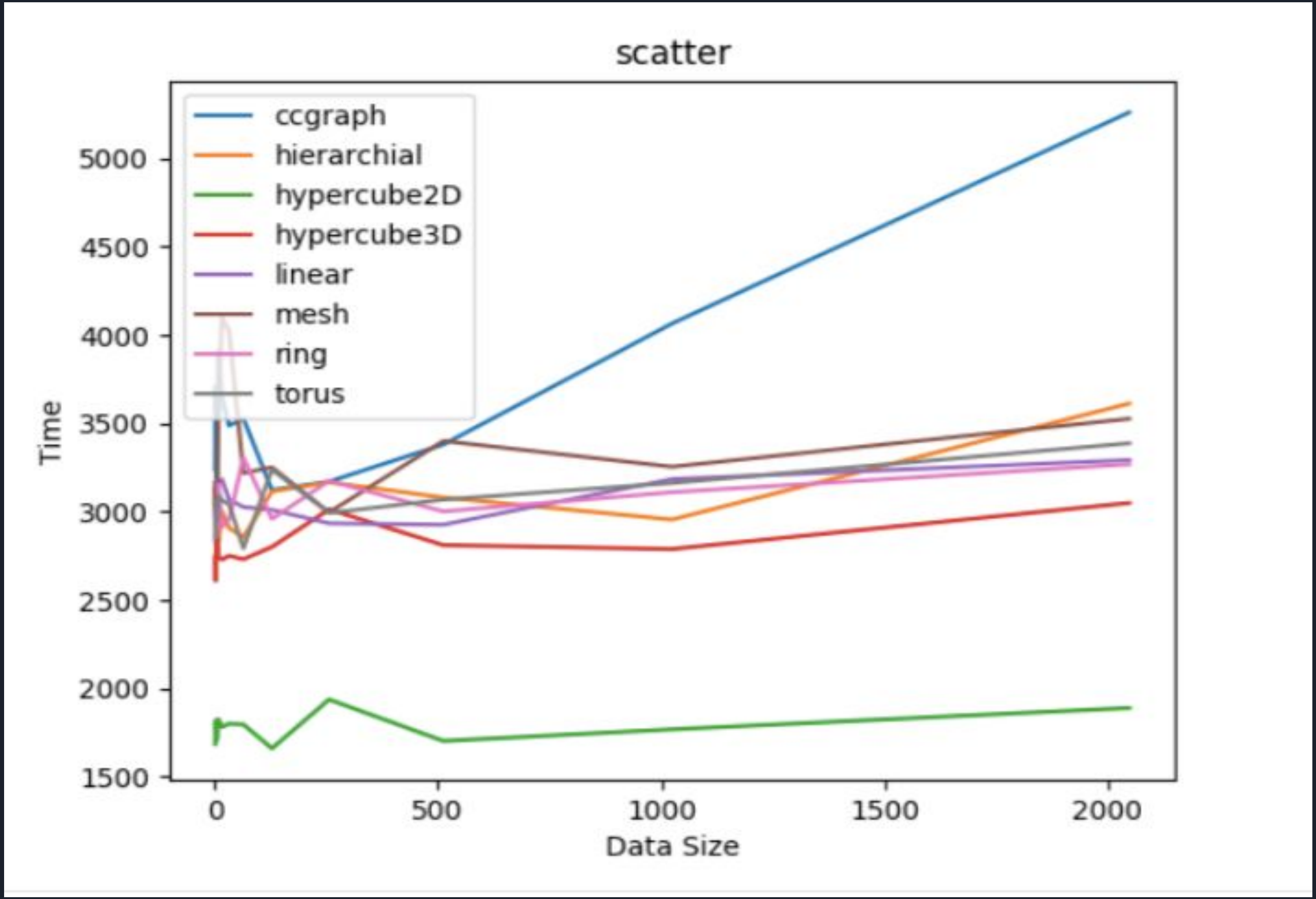


allreduce
8000
7800
7600
7400
7200
7000
6800
6600
6400
Time
0
500
1000
1500
2000
Data Size
ccgraph
hierarchial
hypercube2D
hypercube3D
linear
mesh
ring
torus


alltoall
ccgraph
hierarchial
hypercube2D
hypercube3D
linear
mesh
ring
torus
Time
Data Size
4500
4000
3500
3000
2500
2000
0
500
1000
1500
2000


bcast
ccgraph
hierarchial
hypercube2D
hypercube3D
linear
mesh
ring
torus
Time
Data Size
3500
3000
2500
2000
0
500
1000
1500
2000


gather
ccgraph
hierarchial
hypercube2D
hypercube3D
linear
mesh
ring
torus
Time
3500
3250
3000
2750
2500
2250
2000
1750
0
500
1000
1500
2000
Data Size


scatter
ccgraph
hierarchial
hypercube2D
hypercube3D
linear
mesh
ring
torus
Time
5000
4500
4000
3500
3000
2500
2000
1500
0
500
1000
1500
2000
Data Size

References

- https://www.cs.uky.edu/~jzhang/CS621/chapter5.pdf
- http://parallelcomp.uw.hu/ch04lev1sec1.html
- https://www8.cs.umu.se/kurser/5DV050/VT13/coll.pdf
- https://www8.cs.umu.se/kurser/5DV050/VT12/F1b.pdf
- https://s3.amazonaws.com/content.udacity-data.com/courses/gt-cse6220/Course+Notes/Lesson2-1+Basic+Model.pdf
- Udacity
- https://simgrid.org/tutorials/simgrid-smpi-101.pdf
- http://mpitutorial.com/tutorials/

Thank You