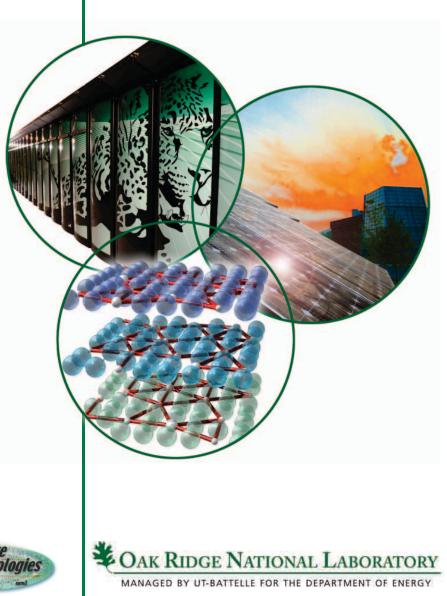
Toward Performance Prediction of Tree-Based Overlay Networks on the Cray XT

Philip C. Roth

Computer Science and Mathematics Division Oak Ridge National Laboratory

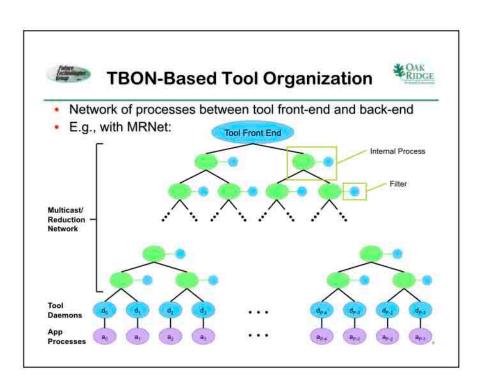






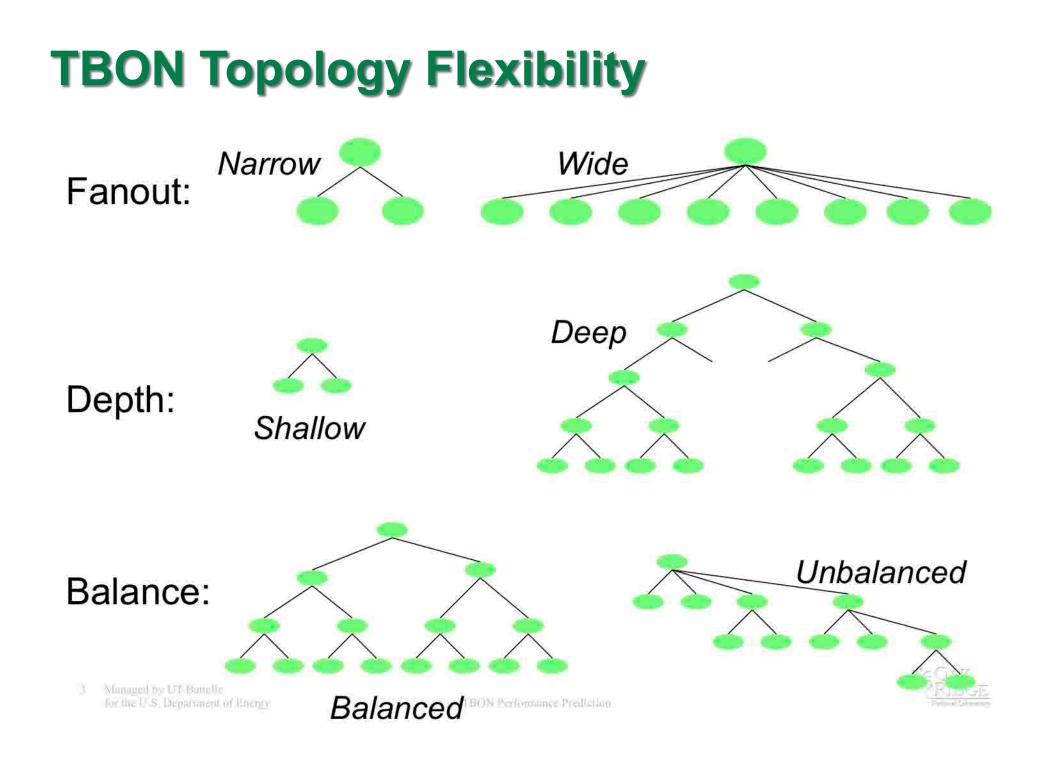
TBONs and MRNet

- A Tree-Based Overlay Network (TBON) like MRNet provides
 scalable infrastructure for tools and applications
- MRNet's process topology and placement support is extremely flexible (on most platforms)
 - Any tree topology
 - Internal processes on same nodes as application processes, or on distinct nodes









The Problem With Flexibility

- Flexibility leads to questions identifying "best" process topology and placement
- Interaction of several factors determine "best"
 - Performance (tool and application)
 - System hardware and software
 - Purpose
 - Even economics (e.g., can I afford to request "extra" nodes for MRNet processes given my allocation budget?)
- Decision process often not rigorous using "rule of thumb"



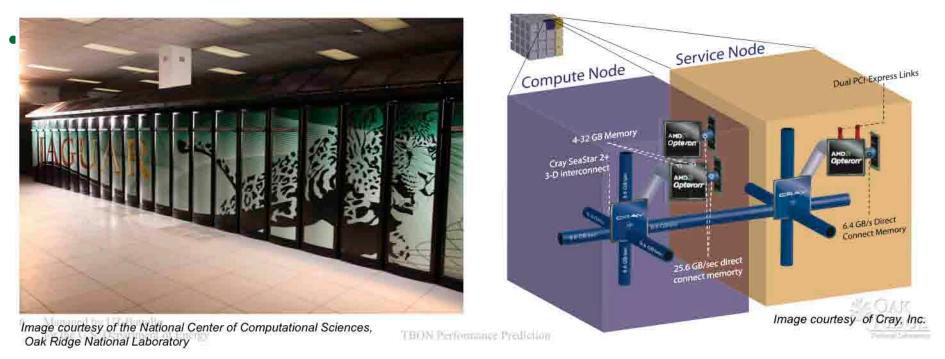
TBON Performance Prediction

- Goal: Given a node allocation on a leadership class system, to be able to identify "best" MRNet process placement and topoogy
- Several constraints:
 - Tool multicast and reduction requirements
 - Behavior of application under study
 - Other activity on the system
 - System software and hardware



Target Platform

- Cray XT is target platform
 - Jaguar XT4 and XT5 systems at Oak Ridge National Laboratory (ORNL)
 - Hopper XT5 at NERSC
 - Kraken XT5 at ORNL
- Opteron-based nodes arranged in 3D mesh with possibility of torus links

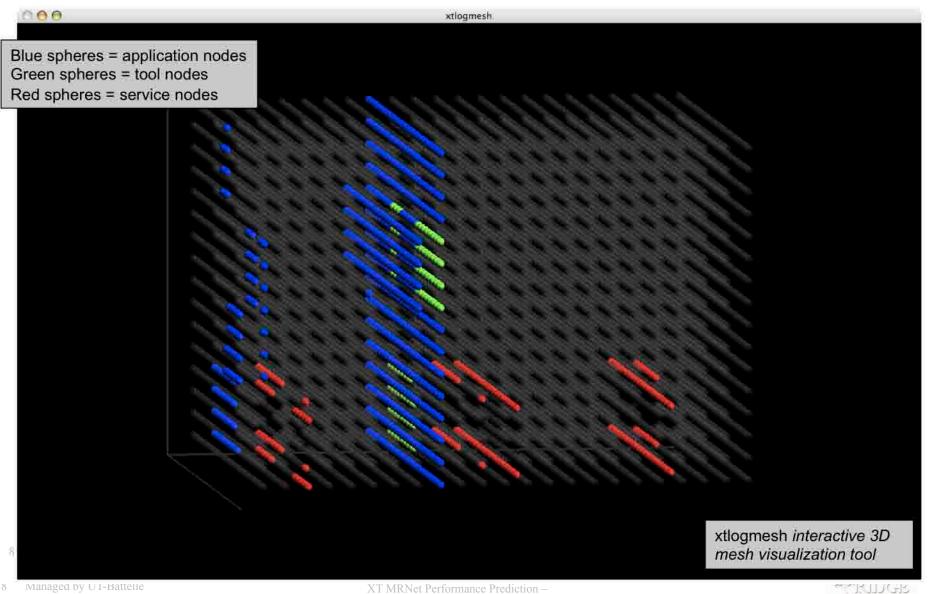


Cray XT TBON Process Placement Tests

- Goal: understand Cray XT allocation characteristics & their impact on MRNet-based tool process placement
- Used simple MPI/Portals program to collect node number and position within the XT mesh
 - Earlier generation ORNL Jaguar with dual-core Opterons
- Batch job launched two independent instances of the program:
 - 512 application nodes (1024 processes)
 - 72 tool nodes (enough for balanced 8-way TBON topology assuming front-end is on batch script service node)



Jaguar Placement Trial Example

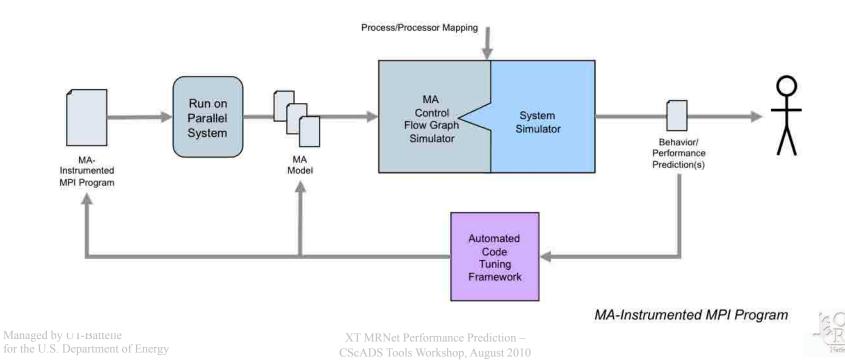


Managed by UI-Battelle 8 for the U.S. Department of Energy

Our Approach

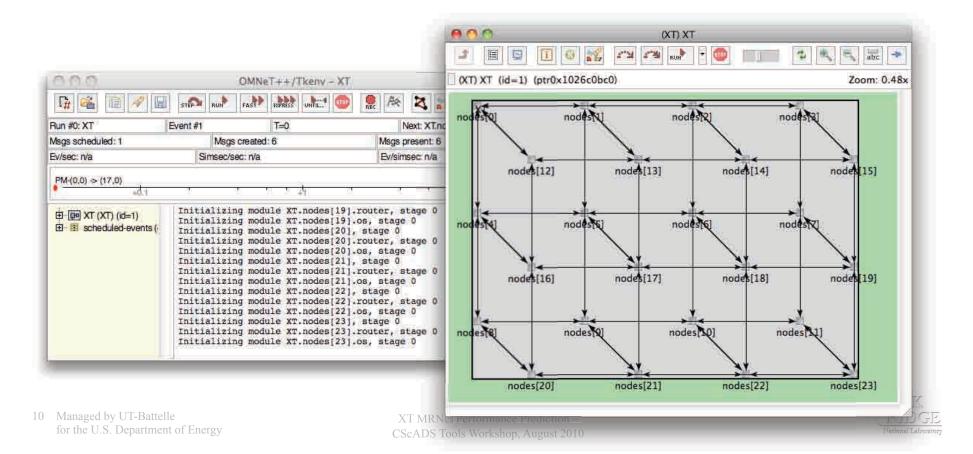
9

- Discrete event simulation of XT system nodes running application and MRNet processes
- Component of MAST framework: Modeling Assertions, Simulation, and Tuning

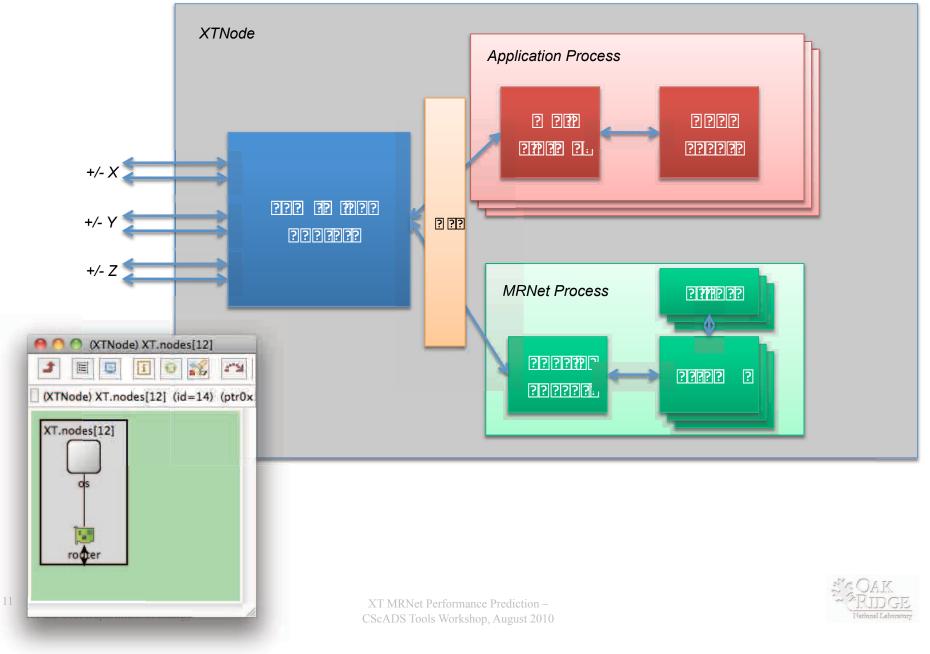


System Model

- Node modules connected in 3D torus
- Implemented using OMNeT++ (http://www.omnetpp.org)



Node Model

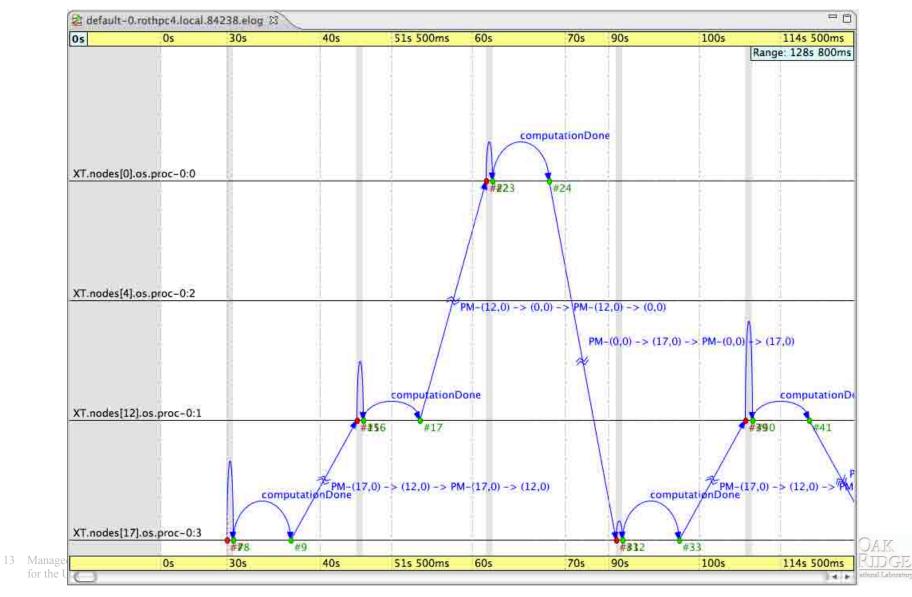


Simulated MPI Hot Potato Activity

Os 20s 28s 30s 340ms 50s 51s 50ms 55s 62s 50ms S XTSim Includes XT.nodes[9].router Range: 64s Range: 64s 100ms S src XT.nodes[10].router XT.nodes[10].router XT.nodes[10].router XT.nodes[10].router 100ms		💩 🛛 📽 🛛 🕸 - 🤤 - 🖘 🗢			28	N 1997) 54			
Image: Signal		default-0.rothpc4.loca	al.84238.elog	X				H	
A 1. Additis (). Additional (). Add	E 😫 🎽	Os 20s	28s	30s 340ms 00ms	40s	50s	51s 500ms 55s		
Simoladies XT.nodes(9).0s XT.nodes(9).0s Winderlos XT.nodes(10).0s XT.nodes(10).0s Winderlos XT.nodes(11).0s XT.nodes(11).0s Wi	xTSim	XT.nodes[9].router						Range: 64s 100ms	
Properties 70 Standes120.couter March20.00-x-11.1.3 Windes120.couter Windes120.couter March20.00-x-11.1.3 Windes121.couter Windes121.couter Windes121.couter Windes121.couter Windes121.couter SSM-c11.1.1.> < <0.0.1.5									
Poperty Yale Property Yale		1000 1 12 01		and the states	- 10 M	N 17 2 3 4			
Produktions Marketin Properties XI Prodesilizione Prodesilizione Prodesilizione			27W-<0,0,02	-><_		1	1		
Implementations Implementations	101 A Guerra (1993) (101							PM-(12.0) -> (0.0)	
Properties (2) Outline (1) Properties (2) Outline (the second se		1						
Bit Bit Lo. Drobest. Kord Windest Dialowski with the second withe second with the second with the second with t			1			8 40 10			
B ometpp.in Im ometpp.in Im of the converte B ometpp.in Im ometpp.in Im ometpp.in B ometpp.in Im ode in information Site of information Makefile Im ode information Site of information M node information Im ode information Site of information M node information Im ode information Im ode information M node information Im ode information Im ode information M node information Im ode information Im ode information M node information Im ode info	A REAL PROPERTY OF A READ REAL PROPERTY OF A REAL P		1			A 13-(1)	7;0) -> (12;0 #19		
Prometyp-common.in/ B package.net Windesi14.losure Windesi15.losure Windesi15.losure Windesi15.losure Windesi15.losure Windesi15.losure Windesi15.losure Windesi16.losure Windesi16.losure Windesi16.losure Windesi18.losure Windesi18.losure Windesi18.losure Windesi18.losure Windesi18.losure Windesi18.losure Windesi20.losure Windesi20.losure Windesi20.losure Windesi20.losure Windesi20.losure Windesi20.losure Windesi20.losure Windesi20.los Windesi20.los Windesi21.losure Windesi22.losure Windesi21.losure Windesi21.losur	And the second						#18	$(0,0) \qquad PM(0,0) \rangle > (1)$	
B onnet(p)=-Onnet(p)=-Onnet(p)=-Sign=-Construction W test If nodes[13]-router XT: nodes[21]-router XT: nodes[22]-router XT: nodes[21]-router XT: nodes[22]-router XT: nodes[21]-router XT: nodes[22]-router XT: nodes[21]-router XT: nodes[22]-router XT: nodes[21]-router XT: nodes[21]-router XT: nodes[21]-router XT: nodes[21]-router XT: nodes[21]-router XT: nodes[21]-router	the full sector state - the sector state at a	and so that the second s	1			/			
B package.ned XT nodes[15]:souter Windes[15]:souter XT nodes[15]:souter Makefile XT nodes[16]:souter Windes[16]:souter XT nodes[16]:souter XT nodes[16]:souter XT nodes[17]:souter XT nodes[17]:souter XT nodes[17]:souter XT nodes[17]:souter YT nodes[17]:souter XT nodes[19]:souter Figure 1 XT nodes[19]:souter Figure 1 XT nodes[19]:souter YT nodes[19]:souter XT nodes[19]:souter XT nodes[10]:souter XT nodes[19]:souter XT nodes[10]:souter XT nodes[10]:souter XT nodes[20]:souter XT nodes[20]:souter XT nodes[20]:souter XT nodes[21]:souter XT nodes[21]:souter XT nodes[21]:so:souter XT nodes[21]:souter XT nodes[22]:souter XT nodes[21]:souter XT nodes	the set of		1.		/				
It est IX1 nodes[15]:souter Makefile XT.nodes[15]:souter XT.nodes[16]:souter XT.nodes[16]:souter XT.nodes[16]:souter XT.nodes[16]:souter XT.nodes[16]:souter SSM-<1,1,1>-><0,0,1> XT.nodes[18]:souter III: 0:0 XT.nodes[18]:souter III: 0:0 XT.nodes[19]:souter III: 0:0 XT.nodes[19]:souter III: 0:0 XT.nodes[19]:souter III: 0:0 XT.nodes[19]:souter III: 0:0 XT.nodes[10]:souter III: 0:0 XT.nodes[10]:souter III: 0:0 XT.nodes[21]:souter III: 0:0 XT.nodes[21]:sos: III: 0:0<			1		/s	SM-<1.1.1> -	> <0,0,1>		
iiii make.out XT.nodesi151.os Makefile XT.nodesi151.os Vindesi161.os XT.nodesi16.os XT.nodesi17.nos Vindesi17.nos XT.nodesi17.nos Vindesi17.nos XT.nodesi18.os Vindesi17.nos XT.nodesi18.os Vindesi17.nos XT.nodesi18.os Vindesi17.nos XT.nodesi18.os Vindesi17.nos XT.nodesi18.os Vindesi17.nos XT.nodesi18.os Vindesi18.os XT.nodesi19.os Vindesi20.os XT.nodesi20.os Vindesi20.os XT.nodesi20.os Vindesi20.os XT.nodesi20.os.proc-0.0 Vindesi20.os XT.nodesi20.os.proc-0.0 Vindesi20.os XT.nodesi21.los.proc-0.0 Vindesi21.los Vindesi21.los.proc-0.0 Vindesi21.los<			1		1				
Makefie XT.nodes[16].outer XT.nodes[17].router SSM-<1,1,1>-><0,0,0]> XT.nodes[17].router SSM-<1,1,1>-><0,0,0]> XT.nodes[18].router 6 XT.nodes[18].router 6 XT.nodes[19].router PM-(17,0)-x(12,0) XT.nodes[19].router PM-(17,0)-x(12,0) XT.nodes[19].router PM-(17,0)-x(12,0) XT.nodes[19].router PM-(17,0)-x(12,0) XT.nodes[10].router PM-(17,0)-x(12,0) XT.nodes[10].router PM-(17,0)-x(12,0) XT.nodes[10].router PM-(12,0)-x(12,0) XT.nodes[20].os PM-(12,0)-x(12,0) XT.nodes[21].router XT.nodes[21].router XT.nodes[21].router XT.nodes[21].router XT.nodes[22].os IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	11 200 041237 0000				/				
NT.nodes[1]7.router PM-(0,0) SSM-(1,1,1)-><0,0,1> NT.nodes[1]7.router PM-(0,0) SSM-(1,1,1)-><0,0,1> NT.nodes[1]7.router PM-(17,0)-><(12,0)	Makefile				1				
XT:nodesi[17].router SSM-<<1,1,1>-><0,0,3> XT:nodesi[18].router #6 XT:nodesi[18].router #7 XT:nodesi[20].router XT:nodesi[20].router XT:nodesi[21].router XT:nodesi[21].router XT:nodesi[21].router XT:nodesi[21].router XT:nodesi[21].router XT:nodesi[22].router XT:nodesi[21].router XT:nodesi[22].router XT:nodesi[22].router XT:nodesi[23].router XT:nodesi[21].router XT:nodesi[23].router XT:nodesi[1		#12				
XT:nodes[12].os.proc-0:1 Yale Yale Yale				S	SM-<1,1,1>-:	> <0.0.1>			
XT.nodes[18].router #6 #210 PM (17.0) -> (12.0) Wroperties X Contine PM (17.0) -> (12.0) PM (17.0) -> (12.0) YT.nodes[19].router XT.nodes[19].router XT.nodes[19].router XT.nodes[20].router YT.nodes[21].router XT.nodes[21].router PM (12.0) -> (12.0) PM (12.0) -> (0.0) XT.nodes[21].router XT.nodes[22].router XT.nodes[22].router XT.nodes[23].router XT.nodes[23].router XT.nodes[23].router XT.nodes[23].router XT.nodes[23].router XT.nodes[23].router XT.nodes[23].router XT.nodes[24].router XT.nodes[26].router XT.nodes[23].router XT.nodes[24].router XT.nodes[26].router XT.nodes[26].router			PM-	$(0,0) \longrightarrow (17,0)$ #11					
Minodes[18].os Minodes[18].os Winodes[18].os Minodes[18].os Winodes[18].os Minodes[18].os Winodes[18].os Minodes[18].os Winodes[18].os Minodes[18].os Winodes[18].os Minodes[12].os Winodes[12].os Minodes[20].outer Winodes[21].os Minodes[22].os Walue Minodes[22].os Winodes[22].os Minodes[22].os Winodes[22].os Minodes[22].os Winodes[22].os Minodes[22].os Winodes[22].os Minodes[22].os Winodes[22].os.proc-0:0 Minodes[22].os Winodes[22].os.proc-0:0 Minodes[22].os Winodes[12].os.proc-0:1 computationDone Winodes[12].os.proc-0:3 Minodes[12].os.proc-0:3 Winodes[12].os.proc-0:3 Minodes[12].os.proc-0:3 Winodes[12].os.proc-0:3 Minodes[12].os.proc-0:3 Winodes[12].os.proc-0:3 Minodes[12].os.proc-0:3 Winodes[12].os.proc-0:3 Minodes[12].os Winodes[12].os.proc-0:3 Minodes[12].os Winodes[12].os.proc-0:3 Minodes[12].os Winodes[12].os.proc-0:3 Minodes[12].os		Service Providence and the service of the service of the	#6	#10	-> (12,0)				
NT.nodes[19].router NT.nodes[20].router Windes[19].os PM-(17,0)-> (12,0) Property Value Property Value NT.nodes[21].os PM-(12,0)-> (12,0) NT.nodes[21].os PM-(12,0)-> (0,0) NT.nodes[21].os Image: ComputationDone NT.nodes[21].os Image: ComputationDone NT.nodes[21].os.proc-0:0 Image: ComputationDone NT.nodes[21].os					100				
Properties IX Image: Image						PM-(17,0) -> (12,0)		
Properties IX B Outline IIII Property Value Value PM-(12,0) -> (0,0) Property Value Value PM-(12,0) -> (0,0) In odes[2].os IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII))416								
XT. nodes[20].os PM-(12,0) -> (12,0) PM-(12,0) -> (0,0) YT. nodes[21].router XT. nodes[21].router PM-(12,0) -> (0,0) XT. nodes[21].router XT. nodes[22].router PM-(12,0) -> (12,0) XT. nodes[22].router XT. nodes[23].router PM-(12,0) -> (12,0) XT. nodes[23].router XT. nodes[23].router PM-(12,0) -> (12,0) XT. nodes[2].router XT. nodes[2].router YT. nodes[2].router XT. nodes[1].router YT. nodes[2].router YT. nodes[2].router XT. nodes[1].router YT. nodes[1].router PM-(12,0) -> (12,0) YT. nodes[2].router YT. nodes[2].router YT. nodes[2].router YT. nodes[2].router YT. nodes[2].router YT. nodes[2].router YT. nodes[1].rou		STATE PERSONAL PROPERTY AND ADDRESS OF ADDRE			- 10 M				
Wate XT.nodes[2].router XT.nodes[2].router XT.nodes[2].os XT.nodes[2].router XT.nodes[2].os.proc-0:0 XT.nodes[4].os.proc-0:1 XT.nodes[1].os.proc-0:1 computationDone XT.nodes[1].os.proc-0:3 XT.nodes[1].router Bs 20s 28s 20s 28s 30s 340ms [00ms 40s 50s 51s 500ms 55s Begin calling initialize(0) in module (cCompoundModule) XT Image: State (CompoundModule) XT.nodes[0]	stroperties to the outline			(0.0) - (1.7 m) (1.7 0)	2 11 2 10		PM-/12 0) -> //	A OV	
Property Value XT.nodes[21].os XT.nodes[22].os XT.nodes[23].router XT.nodes[23].router XT.nodes[23].router XT.nodes[23].os computationDone XT.nodes[23].os computationDone XT.nodes[10].os.proc-0:0 XT.nodes[10].os.proc-0:1 XT.nodes[11].os.proc-0:1 computationDone XT.nodes[12].os.proc-0:1 computationDone XT.nodes[17].os.proc-0:3 #78 #8 20s 28s 30s 340ms [00ms 40s 50s 51s 500ms 55s 62s 500ms 65s #78 #9 Simulation started with runid default-0-20100803-14:01:54-84238 #* @* @* @* @* @* @* @* @* @* @* @* @* @*	1 월 화 등 주		PN	-(0,0) -> 1(2(8)(-(12,0)) -	-> UZ,W		1191-(12,0) -> (0	597	
XT.nodes[22].router XT.nodes[23].os XT.nodes[23].os XT.nodes[2].os XT.nodes[2].os XT.nodes[2].os XT.nodes[2].os XT.nodes[2].os XT.nodes[2].os XT.nodes[2].os.proc-0:0 XT.nodes[1].os.proc-0:1 computationDone XT.nodes[12].os.proc-0:1 xT.nodes[12].os.proc-0:1 xT.nodes[12].os.proc-0:1 xT.nodes[12].os.proc-0:3 XT.nodes[12].os.proc-0:3 XT.nodes[17].os.proc-0:3 XT.nodes[17].os.proc-0:3 XT.nodes[17].os.proc-0:1 ComputationDone XT.nodes[17].os.proc-0:3									
XT.nodes[22].os xT.nodes[23].outer XT.nodes[23].os xT.nodes[23].os XT.nodes[23].os.proc-0:0 computationDone XT.nodes[12].os.proc-0:1 computationDone XT.nodes[12].os.proc-0:1 computationDone XT.nodes[12].os.proc-0:3 computationDone XT.nodes[17].os.proc-0:3 computationDone XT.nodes[17].os.proc-0:3 computationDone XT.nodes[17].os.proc-0:3 rest Bas 20s 28s 30s 30s 30s Vent # Time Details #0 0 Simulation started with runid default-0-20100803-14:01:54-84238 ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** **	- 57-91				- 10 M	N 77 27 2 4			
XT.nodes[23].router XT.nodes[23].os XT.nodes[23].os XT.nodes[23].os XT.nodes[23].os XT.nodes[23].os XT.nodes[23].os XT.nodes[23].os XT.nodes[23].os XT.nodes[12].os.proc-0:0 XT.nodes[12].os.proc-0:1 ComputationDone XT.nodes[17].os.proc-0:3 #9 8 20s 28 30s 340ms [00ms 40 0s Event # Time Details #0 0 Simulation started with runid default-0-20100803-14:01:54-84238 * Simulation started with runid default-0-20100803-14:01:54-84238 * * * Simulation started with runid default-0-20100803-14:01:54-84238 * * * * * * * * * * * * * * * * * * * * * * * <td< td=""><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></td<>									
XT.nodes[2].os computationDone XT.nodes[0].os.proc-0:0 computationDone XT.nodes[1].os.proc-0:1 computationDone XT.nodes[1].os.proc-0:1 computationDone XT.nodes[1].os.proc-0:3 #9 8s 20s 20s 28s 30s 30s #78 #9 8s 20s 20s 28s 30s 30s #10 #9 8s 20s 20s 28s 30s 300ms 40 0s 5 Simulation started with runid default-0-20100803-14:01:54-84238 8 Simulation started with runid default-0-20100803-14:01:54-84238 9 Simulation started with runid default-0-20100803-14:01:54-84238 9 Begin calling initialize(0) in module (cCompoundModule) XT.nodes[0] 9 Begin calling initialize(0) in module (cCompoundModule) XT.nodes[0] 9 Begin calling initialize(0) in module (scaStar) XT.nodes[0].to i=misc/square s									
XT.nodes[0].os.proc-0:0 XT.nodes[4].os.proc-0:1 computationDone #223 #24 XT.nodes[12].os.proc-0:1 computationDone #24 #24 XT.nodes[17].os.proc-0:3 #### #9 8s 20s 28s 30s 340ms 00ms 40s 50s 51s 500ms 55s 62s 500ms 65s ### #9 ### #9 ### ### ### ### ### ### ### #### #### #### ####################################									
XT.nodes[4].os.proc-0:2 computationDone #223 #24 XT.nodes[12].os.proc-0:1 computationDone #24 XT.nodes[17].os.proc-0:3 #78 #9 8s 20s 28s 30s 340ms 00ms 40s 50s 51s 500ms 55s 62s 500ms 65s #78 #9		and the second				······································		computationDone	
XT.nodes[12].os.proc-0:1 computationDone XT.nodes[17].os.proc-0:3 #9 8s 20s 28s 30s 340ms 00ms 40s 50s 51s 500ms 55s 62s 500ms 65s Image: Stress of the						1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	Summer was an Block	#223 #24	
XT.nodes[17].os.proc-0:3 #78 8s 20s 28s 30s 340ms 00ms 40s 50s 51s 50oms 62s 50oms 65s 62s 50oms 65s 62s 50oms 65s 62s 62s 50oms 65s 62s 62s <td></td> <td></td> <td>1</td> <td>Contraction and a strategy of the</td> <td></td> <td></td> <td>computationDone</td> <td>TRACT (</td> <td></td>			1	Contraction and a strategy of the			computationDone	TRACT (
8s 20s 28s 30s 340ms 00ms 40s 50s 51s 500ms 55s 62s 500ms 65s 1 Problems Image: Module Hierarchy Image: NED Parameters Image: NED Inheritance Image: Console Image: Event Log Ima				computationDone		#19	6 #17		
8s 20s 28s 30s 340ms 00ms 40s 50s 51s 500ms 55s 62s 500ms 65s 4		A Lindestri hosipide o		#8		# #.A	CM		
Image: Problems Module Hierarchy NED Parameters NED Inheritance Console Event Log Image: Problems Image: Problem		0c 20c			402	FOr	Fle FOOme FFe	62c 500mc 65c	
Problems Module Hierarchy NED Parameters NED Inheritance Console Event Log ************************************		05 205	205	SUS SAOMS DOMS	405	505	315 300ms 335		-
Event # Time Details #0 0s Event in module (cCompoundModule) XT #0 0s Simulation started with runid default-0-20100803-14:01:54-84238 #0 0s Pi Begin calling initialize(0) in module (cCompoundModule) XT.nodes[0] #0 0s Pi Begin calling initialize(0) in module (cCompoundModule) XT.nodes[0] #0 0s Pi Begin calling initialize(0) in module (SeaStar) XT.nodes[0].router #0 0s Pi Begin calling initialize(0) in module (SeaStar) XT.nodes[0].router #0 Display string changed for module (cCompoundModule) XT.nodes[0] to i=misc/square s		la Terraria de la companya de la company	many fine			<u> </u>	Contract Sec.	and the second se	3
#0 0s Event in module (cCompoundModule) XT Image: Simulation started with runid default-0-20100803-14:01:54-84238 Image: Simulation		2 Problems 3 Module	lierarchy	NED Parameters 👔 NED	Inheritance 🔄	Console 苯	Event Log 🕄	第・41・1 ⑧・旦・1 🔗 🧮	4
Simulation started with runid default-0-20100803-14:01:54-84238 Begin calling initialize(0) in module (cCompoundModule) XT.nodes[0] In calling module Begin calling initialize(0) in module (SeaStar) XT.nodes[0].router Display string changed for module (cCompoundModule) XT.nodes[0] to i=misc/square s		Event # Time	Details						n l
Image: Control of the second secon		#0 Os	Eve	nt in module (cCompoun	dModule) XT				
ゆ G ゆ ゆ End calling module ゆ G ゆ ゆ Begin calling initialize(0) in module (SeaStar) XT.nodes[0].router ● Display string changed for module (cCompoundModule) XT.nodes[0] to i=misc/square s		10 Os	S	imulation started with ru	nid default-0-	20100803-14	1:01:54-84238		
ゆ G ゆ ゆ End calling module ゆ G ゆ ゆ Begin calling initialize(0) in module (SeaStar) XT.nodes[0].router ● Display string changed for module (cCompoundModule) XT.nodes[0] to i=misc/square s									
신 Display string changed for module (SeaStar) XT.nodes[0].router							CALCULATION AND AND AND AND AND AND AND AND AND AN		
Display string changed for module (cCompoundModule) XT.nodes[0] to i=misc/square s			the second se	a service of the serv	n module (Sea	Star) XT.node	s[0].router		6
		40 04		33 (5) Hours			and the second se	isc/square_s	
		C		staping string changed) moune (C	compound with	sels/ A langaes(0) to (=))	A P	- 2

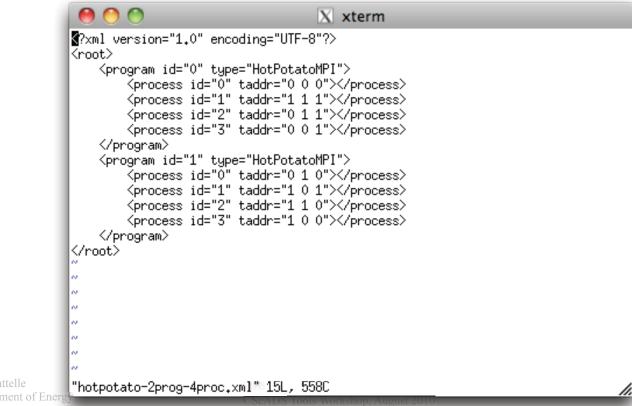
12 Manag

Simulated MPI Hot Potato Activity, Filtered



Workload Specification

- XML file
- Multiple parallel programs per file, including type and associated attributes like "input"
- Mapping of processes to system nodes

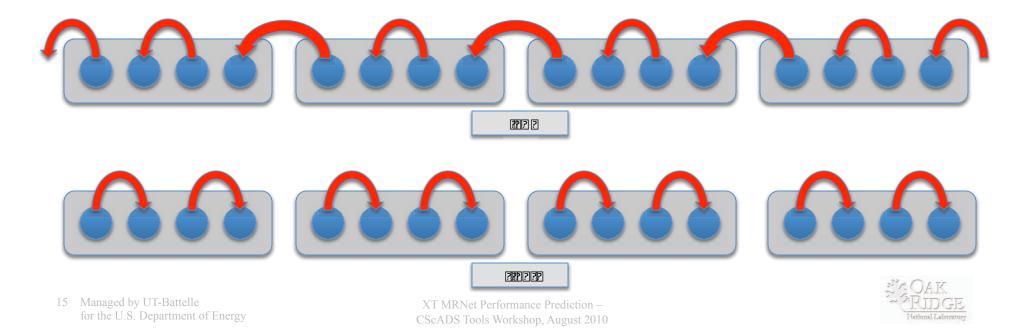




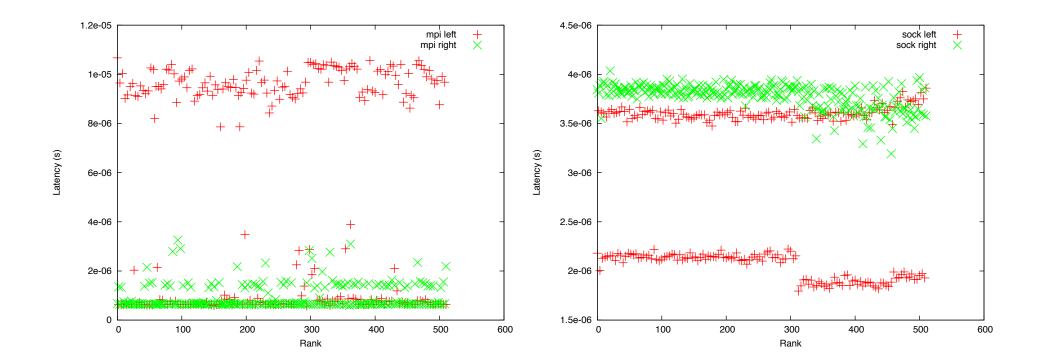
14 Managed by UT-Battelle for the U.S. Department of Energy

Model Parameterization

- Measuring process-to-process latency and bandwidth
 - MPI, Sockets
 - Fully populated nodes, one process per node
- Pairs of processes
 - Even ranks first pair left, then right



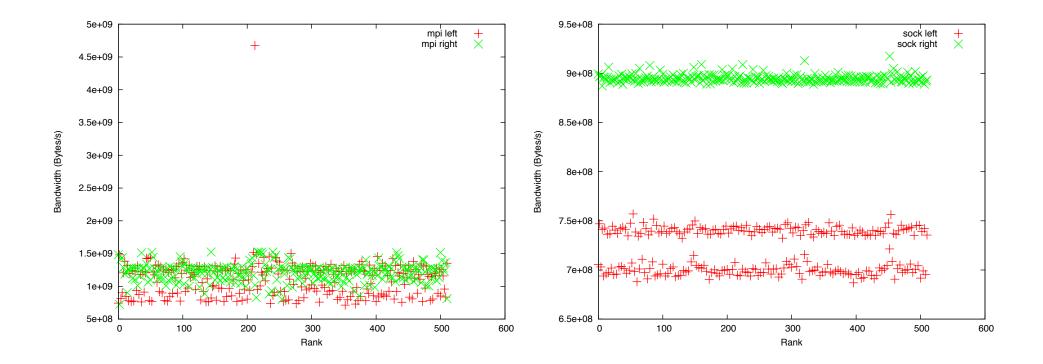
Jaguar XT4 MPI and Socket Latency, Fully Populated Nodes



16 Managed by UT-Battelle for the U.S. Department of Energy



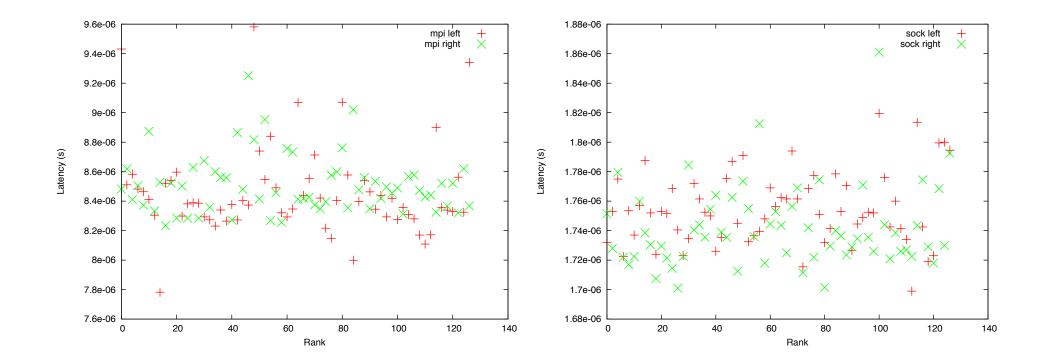
Jaguar XT4 MPI and Socket Bandwidth, Fully Populated Nodes



17 Managed by UT-Battelle for the U.S. Department of Energy



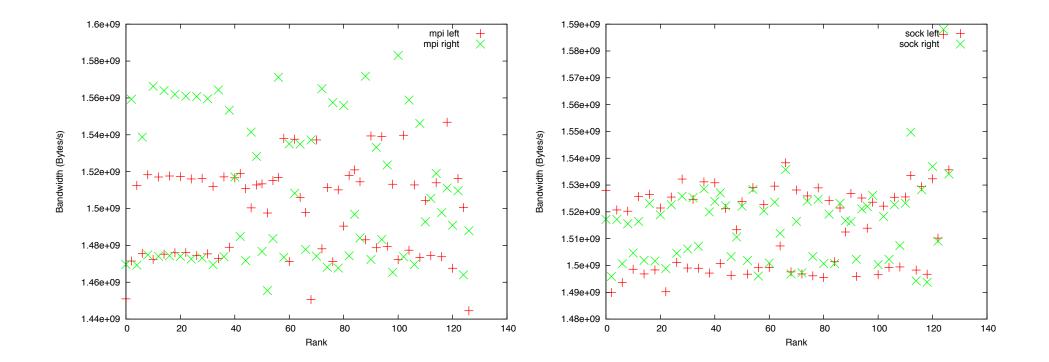
Jaguar XT4 MPI and Socket Latency, One Process Per Node



18 Managed by UT-Battelle for the U.S. Department of Energy



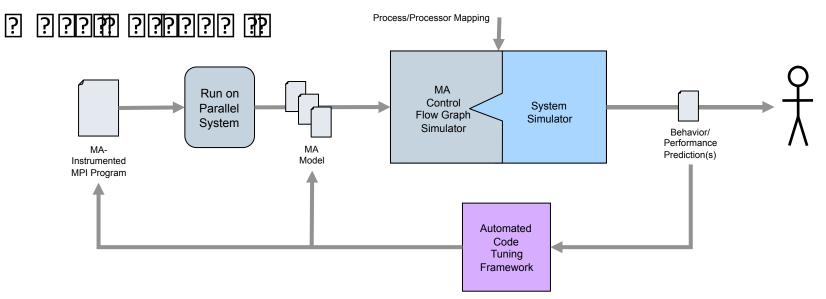
Jaguar XT4 MPI and Socket Bandwidth, One Process Per Node

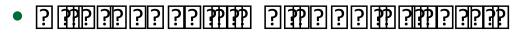


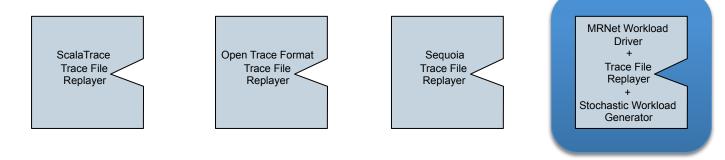
19 Managed by UT-Battelle for the U.S. Department of Energy



Simulation Flexibility







RIDGE National Laboratory

20 Managed by UT-Battelle for the U.S. Department of Energy



- Basic XTNode with SeaStar router is implemented
 - Parameterization still in progress as described earlier
- Support for simple MPI-based workloads
 - Hardcoded behaviors (hot potato, 1D exchange)
 - OTF and Sequioa trace readers implemented for previous version, must be resurrected
- Support for TBON processes designed and partially implemented
- Recently adapted model from OMNeT++ 3.2 to 4.1 (changes in simulation time)

21 Managed by UT-Battelle for the U.S. Department of Energy



Acknowledgements

- This research is sponsored by the Office of Advanced Scientific Computing Research; U.S. Department of Energy. The work was performed at the Oak Ridge National Laboratory which is managed by UT Battelle, LLC under Contract No. De-AC05-00OR22725.
- This research used resources of the Center for Computational Sciences at Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. De-AC05-00OR22725.





- Predicting TBON performance on Cray XT is highly desirable
 - Matching TBON process topology and placement to tool needs subject to application and system constraints
 - May support online reconfiguration of TBON topology
- Developing simulation-based TBON prediction capability
 - Expect predictions of realistic scenarios soon
 - Easily adaptable to expected future architectures (e.g., GPU-enabled nodes, Infiniband clusters)
 - Embeddable (in theory)





For more information

RIDGE

http://ft.ornl.gov rothpc@ornl.gov http://www.paradyn.org/mrnet