Mobitopolo: A Portable Infrastructure to Facilitate Flexible Deployment and Migration of Distributed Applications with Virtual Topologies

Richard POTTER
NICT
Akihiro NAKAO
University of Tokyo
NICT

Virtual Infrastructure for Testbeds

- Increasing availability of hosting environments:
  - PlanetLab, CoreLab, EmuLab, Amazon EC2
- New challenges:
  - Consistent execution environment across heterogeneous hosts
  - Live migration between hosts
  - Maintain connections between components during migration

1. Consistent Execution Environment

- (VM) User-Mode Linux (UML)
  - Runs inside of virtual environments provided by PlanetLab (VServer), CoreLab (KVM), Amazon EC2(XEN)
  - Full Linux kernel functionality
- (NETWORKING) Added Ethernet/UDP tunnels
  - modified UML's TUN/TAP device driver to connect to UDP socket, instead of **`/dev/net/tun`**.
  - no root privileges needed
  - supports any protocol on top of Ethernet

2. Live Migration

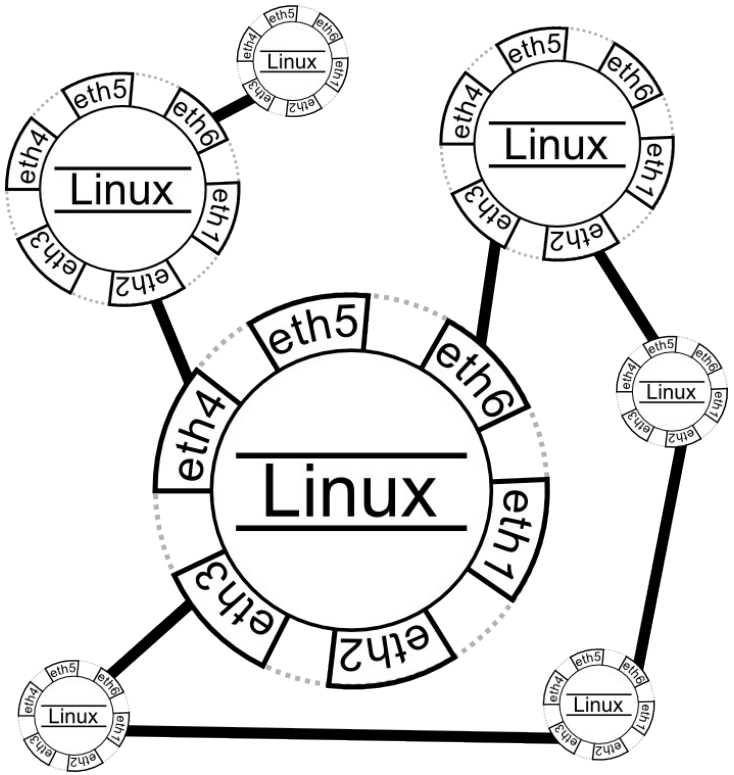
- Scrapbook for User-Mode Linux (SBUML)
  - Provides VM Snapshots to UML (since 2003)
  - Automatic HTTP download with demand fetching
- Added Live Migration over WAN
  - Iterative copy while VM is still running
    - Copy both RAM and DISK
  - Each pass copies smaller delta
  - Final copy with VM frozen
- Implemented with modified `tar`
  - Downtime can be less than 1 second
  - Depends on Internet bandwidth and machine activity

3. Maintaining Connections

- Central Control Software
  - Automatic Deployment
    - VMs initialized from snapshots
    - Tunnels automatically configured
  - Automatic reconnection of UDP-tunnel connections after migration

Result: Mobitopolo

- User-Mode Linux + Ethernet/UDP Tunnels + SBUML + Live Migration+ Central Control
- Distributed Applications see Linux OS connected by Ethernet
- Physical Host Differences are hidden

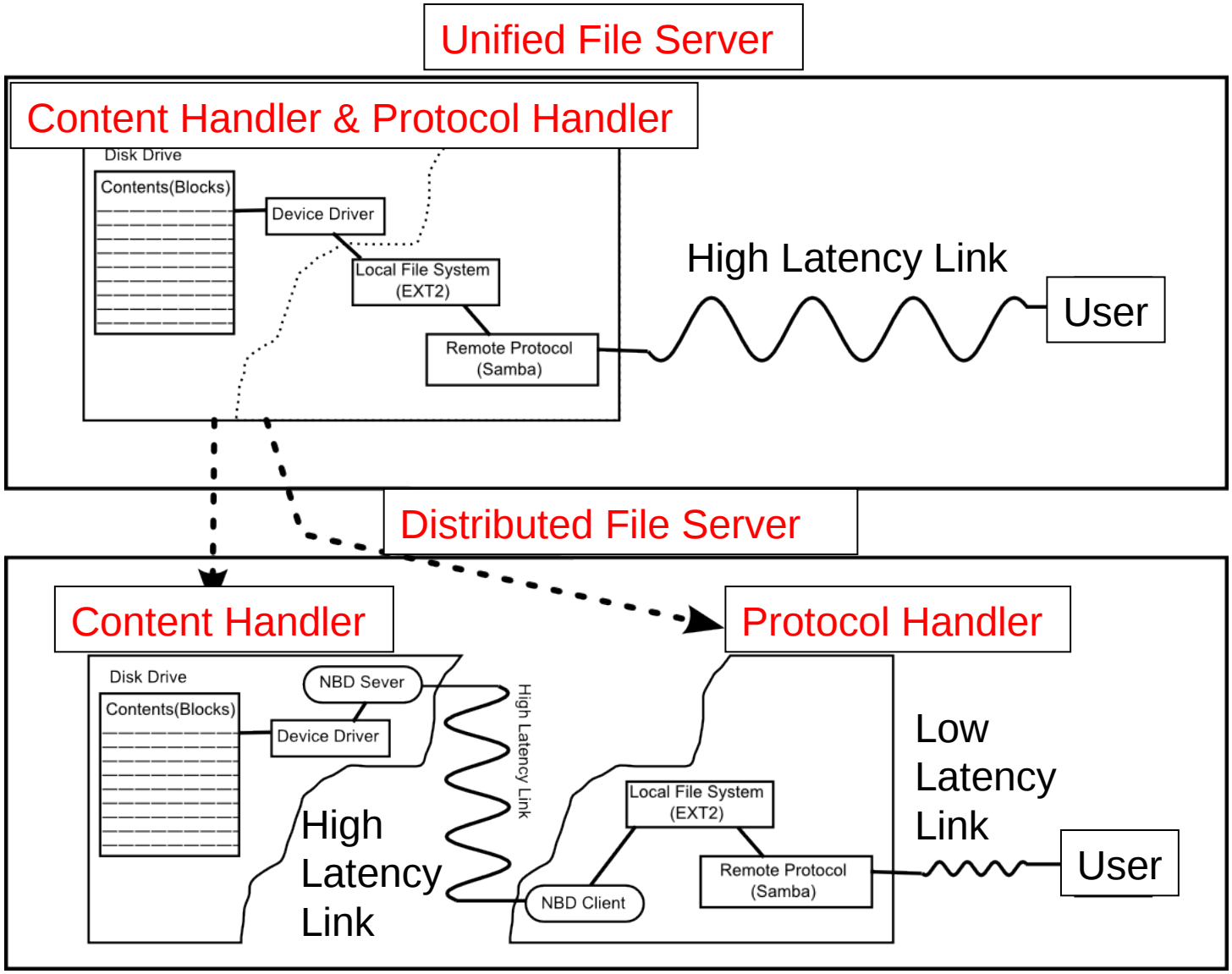


More General Benefits:
Consistent execution environment.....plus...

- Distributed application's physical host dependencies are minimized
  - Design, implementation, configuration,....
  - ...and runtime state!
  - Internal IMPLEMENTATION becomes independent of physical DEPLOYMENT
- *Preconfigured* distributed snapshots!
- Flexible, fast, automatic deployment
- Simplified application development
- Replication for experiments

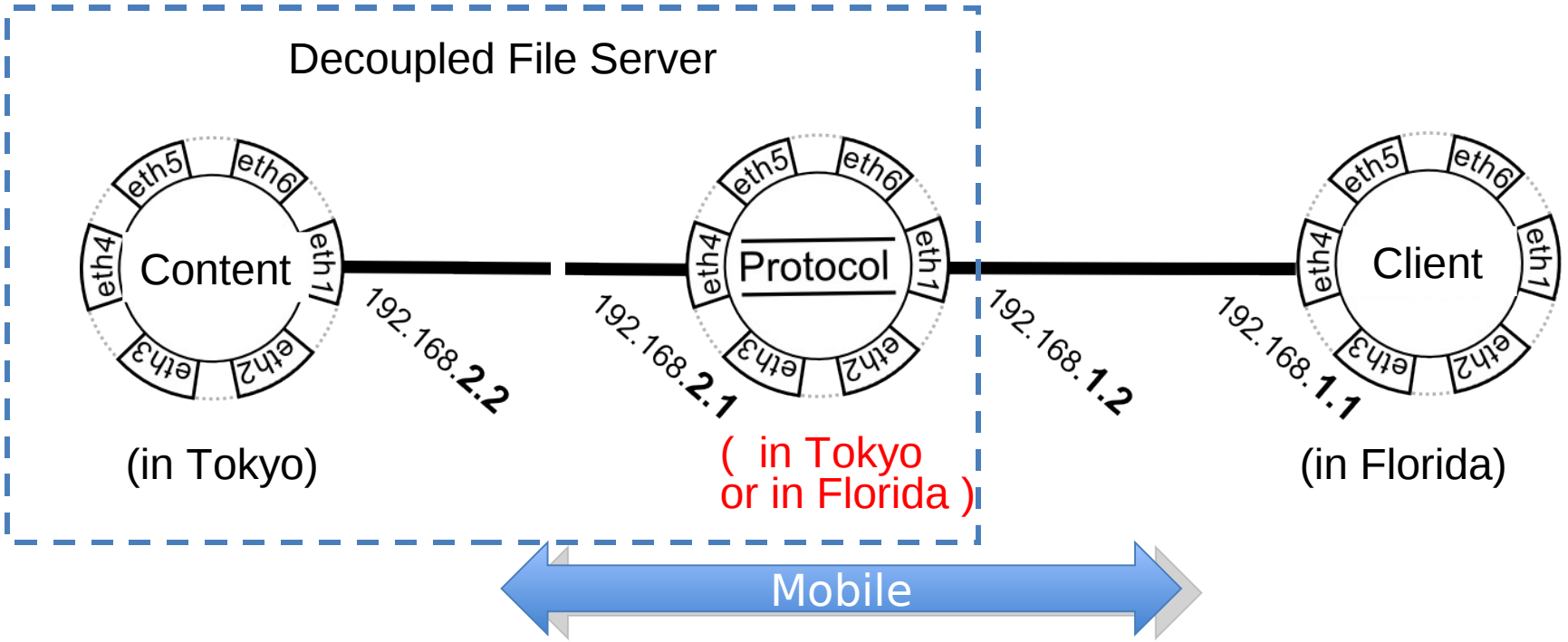
What would be a good small (3 node?) distributed application for illustrating these? ...plus generate some performance data

Example Distributed Application



Now “Protocol Handler” can follow you !!

Experiment with Mobitopolo



If protocol VM is in Tokyo, file copy BW = 120Kbps

If protocol VM is in Florida, file copy BW = 790Kbps

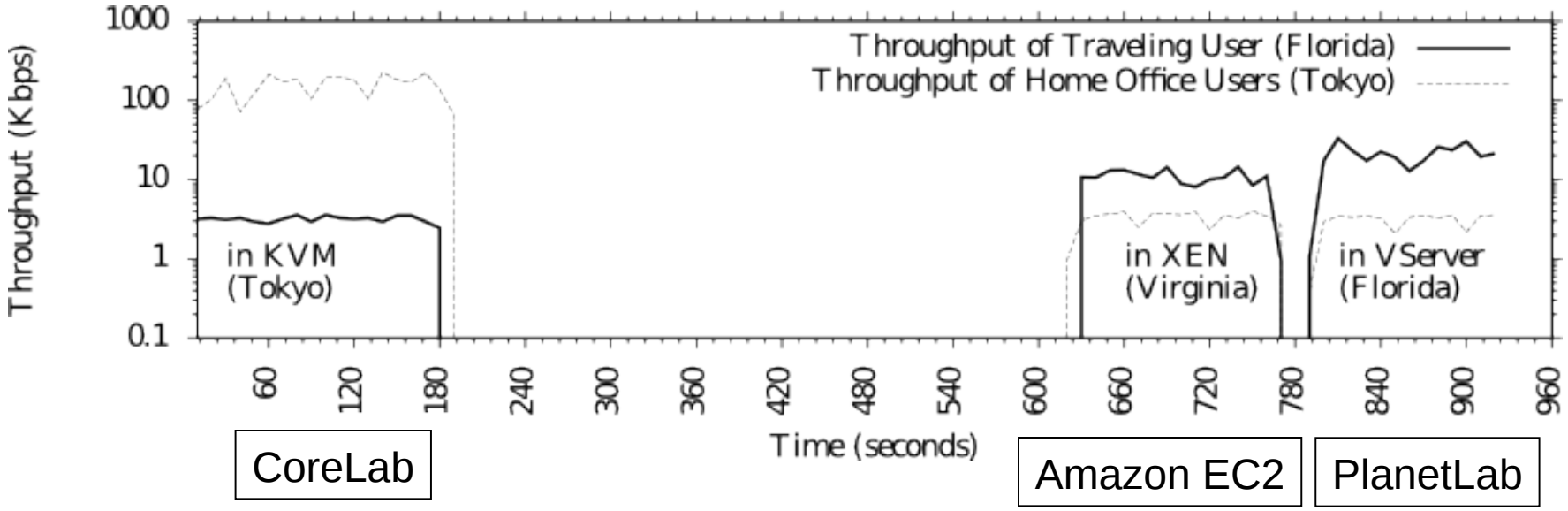
Can replicate EXACT experiment many times

Replicated 24 times

connection to contents in Tokyo				Host of Protocol VM	connection to client in Florida					
H-Th (Mbps)	VM-Th (Mbps)	H-La (ms)	VM-La (ms)		H-Th (Mbps)	VM-Th (Mbps)	H-La (ms)	VM-La (ms)	File-Th (Mbps)	(cached) File-Th (Mbps)
				collocated with contents:						
5791	117.96	.06	.71	node1-net0.koganei.corelab.jp	2.29	2.13	203.00	204.00	.12	.12
				collocated with client:						
1.93	1.18	203.00	204.00	planetlab2.acis.ufl.edu	6367	117.17	.02	.32	.79	19.00
				in Eastern North America:						
2.71	1.91	177.00	178.00	planetlab5.csres.utexas.edu	8.55	7.90	28.70	29.00	.49	.91
6.23	2.32	185.00	180.00	ec2-67...amazonaws.com	8.03	7.82	28.90	29.35	.43	.83
3.12	2.44	155.00	156.00	planetlab2.utdallas.edu	8.06	7.37	43.70	44.95	.36	.60
2.47	2.15	188.00	188.00	planetlab2.isi.jhu.edu	8.78	7.74	48.20	48.55	.32	.54
2.50	1.96	188.50	188.50	planetlab4.cnds.jhu.edu	8.55	7.26	48.40	49.10	.32	.53
2.38	1.63	188.00	189.00	planetlab3.cnds.jhu.edu	8.71	7.48	49.30	49.55	.31	.53
2.41	2.24	186.00	193.00	planet1.pitts...intel-research.net	1.12	0.61	54.90	55.30	.30	.46
				in Western North America:						
4.21	3.81	116.00	185.50	planlab2.cs.caltech.edu	8.64	7.28	56.40	56.70	.33	.48
4.08	2.83	120.00	121.00	planetlab-2.calpoly-netlab.net	7.34	6.21	60.30	60.65	.32	.45
3.40	3.16	140.00	141.00	planetlab7.flux.utah.edu	7.00	6.73	64.65	65.00	.31	.42
3.16	2.95	140.00	141.00	planetlab6.flux.utah.edu	6.67	5.99	64.70	65.05	.27	.37
0.79	0.67	243.00	244.00	planetlab4.postel.org	5.48	5.12	83.40	83.80	.21	.30
				in Europe:						
1.62	1.14	284.00	292.00	planetlab4.lublin.rd.tp.pl	3.07	2.55	155.00	156.00	.14	.16
1.63	1.03	289.00	290.00	plebt2.essex.ac.uk	2.90	2.76	163.00	163.00	.14	.16
1.55	1.34	294.00	296.00	planetlab-node1.it-sudparis.eu	2.70	2.65	171.00	171.00	.13	.15
1.50	1.41	297.00	298.00	node1pl.p...telecom-lille1.eu	2.73	2.59	173.50	174.00	.13	.15
1.50	0.49	294.00	295.00	plane-lab-pb2.uni-paderborn.de	2.66	2.34	176.00	177.00	.13	.15
1.40	1.31	309.00	310.00	planetlab2.it.uc3m.es	2.45	0.99	191.50	192.00	.12	.13
1.36	1.01	323.00	324.00	planetlab3.upc.es	2.27	2.15	205.00	206.00	.11	.12
0.28	0.27	304.00	301.00	planetlab1.mwrl.net	0.24	0.29	135.00	136.50	.07	.08
				elsewhere:						
10.20	7.95	.12	1.07	planetlab1.koganei.wide.ad.jp	2.09	2.15	203.00	203.00	.12	.13
0.57	0.34	333.00	334.00	planetlab1.tau.ac.il	1.97	1.89	216.00	216.00	.10	.11

WAN Migration

(first draft implementation)



- 7.5 min downtime from Tokyo to Virginia over 6.2Mbps link
  - Difficult migration due to high VM load in Tokyo undermining pre-copy effectiveness
- 28 sec downtime from Virginia to Florida over 8.0Mbps link
- Most WAN migration is tested on 100Mbps or 1Gbps links

Related Work
(User-Mode Networking)

- Bavier, Feamester, Huang, Peterson, & Rexford: In VINI Veritas: Realistic and Controlled Network Experimentation
- Jiang, & Xu: Violin: Virtual Internetworking on Overlay Infrastructure.
  - Both used UML
  - 2nd used custom UDP tunnels
  - Neither had snapshots or migration

Conclusion

- Standard Linux functionality and network interfaces
- Portable user-mode implementation
- Live migration across WAN
- Deployment of preconfigured VM snapshots and network topologies
- Physical deployment details transparent to distributed system