# Logistical Networking: Developments and Deployment

**Micah Beck, *Assoc. Prof. & Director***

*Logistical Computing & Internetworking (LoCI) Lab*

Computer Science Department

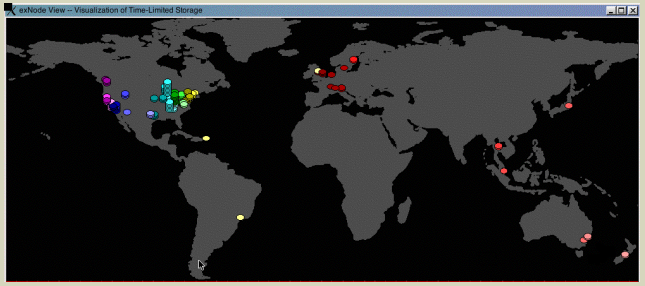
University of Tennessee

**mbeck@cs.utk.edu**

AMPATH Astronomy WG, Miami Jan 31, 2003

LoCI

LOGISTICAL COMPUTING AND INTERNETWORKING LAB




UNIVERSITY OF TENNESSEE

# Logistical Networking Research

- University of Tennessee
  - Micah Beck
  - James S. Plank
  - Jack Dongarra
- University of California, Santa Barbara
  - Rich Wolski

- Funding
  - Dept. of Energy SciDAC
  - National Science Foundation ANIR
  - UT Center for Info Technology Research

LoCI

# What is Logistical Networking

- A scalable mechanism for deploying shared storage resources throughout the network
- An general store-and-forward overlay networking infrastructure
- A way to break long transfers into segments and employ heterogeneous network technologies
- P2P storage and content delivery that doesn't using endpoint storage or bandwidth

LoCI

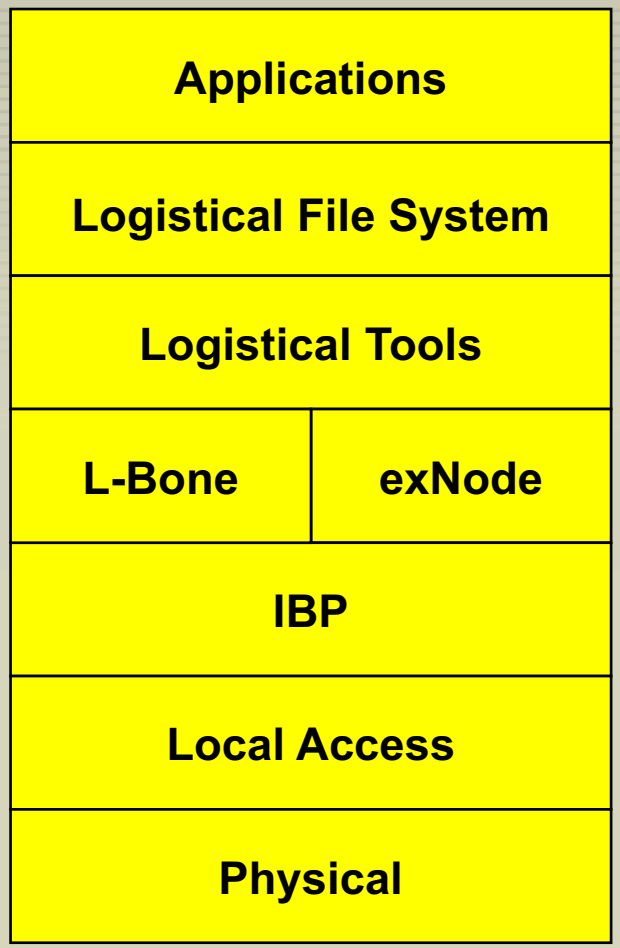
# Why "Logistical Networking"

- Analogy to logistics in distribution of industrial and military personnel & materiel
- Fast highways alone are not enough
  - Goods are also stored in warehouses for transfer or local distribution
- Fast networks alone are not enough
  - Data must be stored in buffers/files for transfer or local distribution
- Conventional vs logistical networking
  - Datagram routers make *spatial* choices
  - Storage depots enable *temporal* choices

LoCI

# The Network Storage Stack

- Our adaption of the network stack architecture for storage
- Like the IP Stack
- Each level encapsulates details from the lower levels, while still exposing details to higher levels




LoCI

# IBP: The Internet Backplane Protocol

- Storage provisioned on community “depots”
- Very primitive service (similar to block service, but more sharable)
  - Goal is to be a common platform (exposed)
  - Also part of end-to-end design
- Best effort service – no heroic measures
  - Availability, reliability, security, performance
- Allocations are time-limited!
  - Leases are respected, can be renewed
  - Permanent storage is to strong to share!

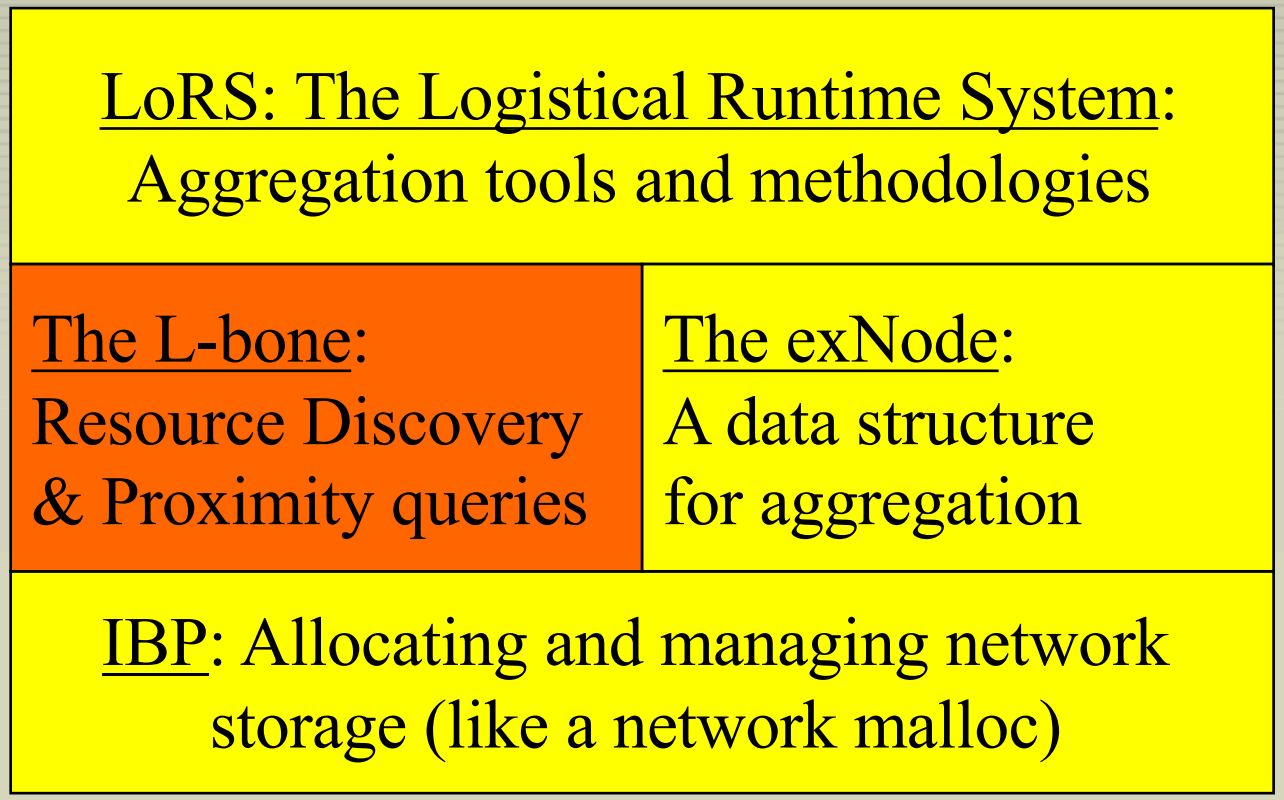
LoCI

# Models of Sharing: Logistical Networking

- Moderately valuable resources
  - Storage, server cycles
- Sharing enabled by relative plenty
- Internet-like policies
  - Loose access control
  - No per-use accounting
- Primary design goal: scalability
  - Application autonomy
  - Resource transparency
- Burdens of scalability
  - The End-to-End Principles
  - Weak operation semantics
  - Vulnerability to Denial of Service

LoCI

# The Network Storage Stack

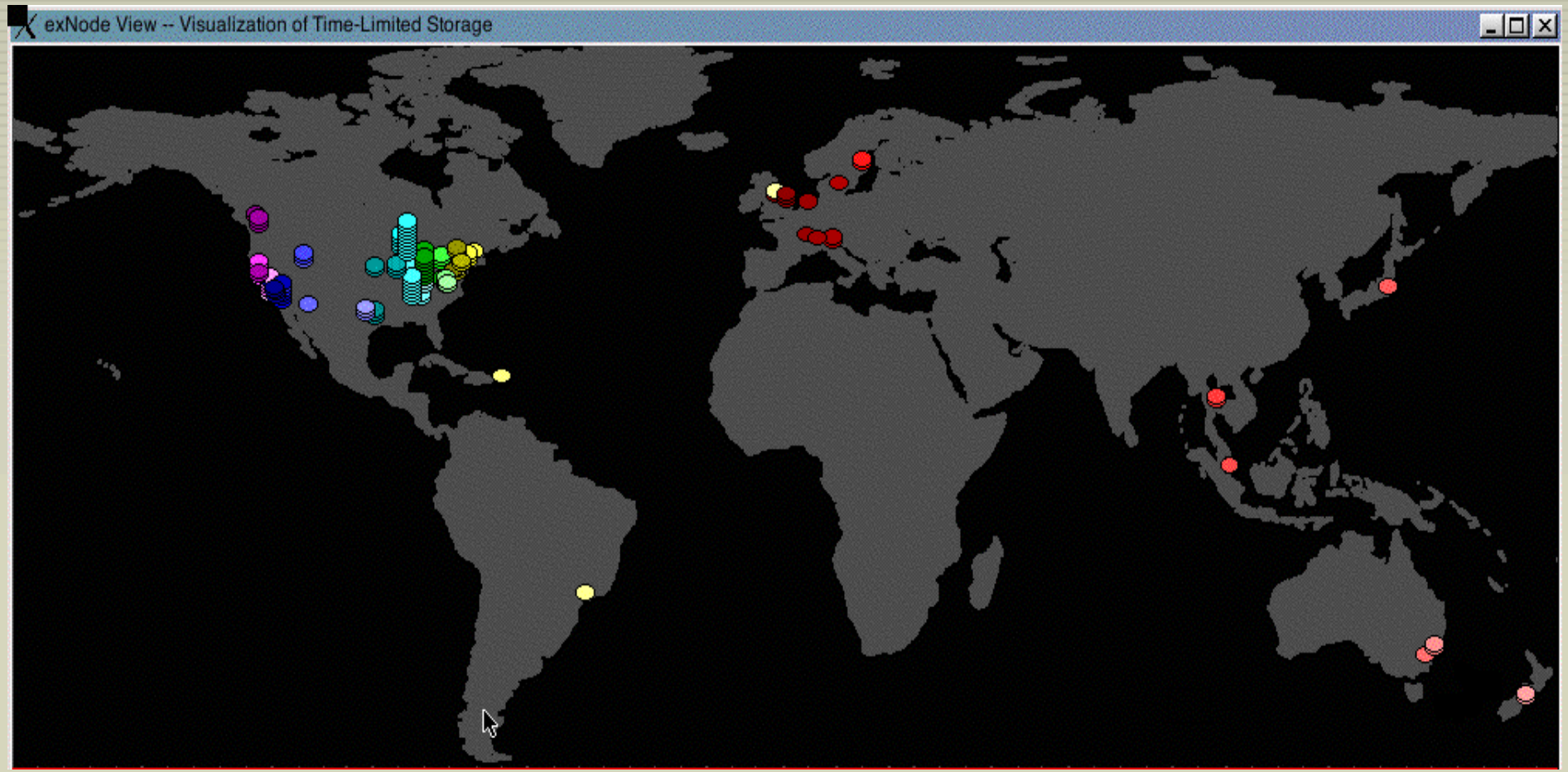


# The Logistical Backbone (L-Bone)

- LDAP-based storage resource discovery.
- Query by capacity, network proximity, geographical proximity, stability, etc.
- Periodic monitoring of depots.
- 10 Terabytes of shared storage. (with plans to scale to a petabyte...)

LoCI

# L-Bone: January 2003



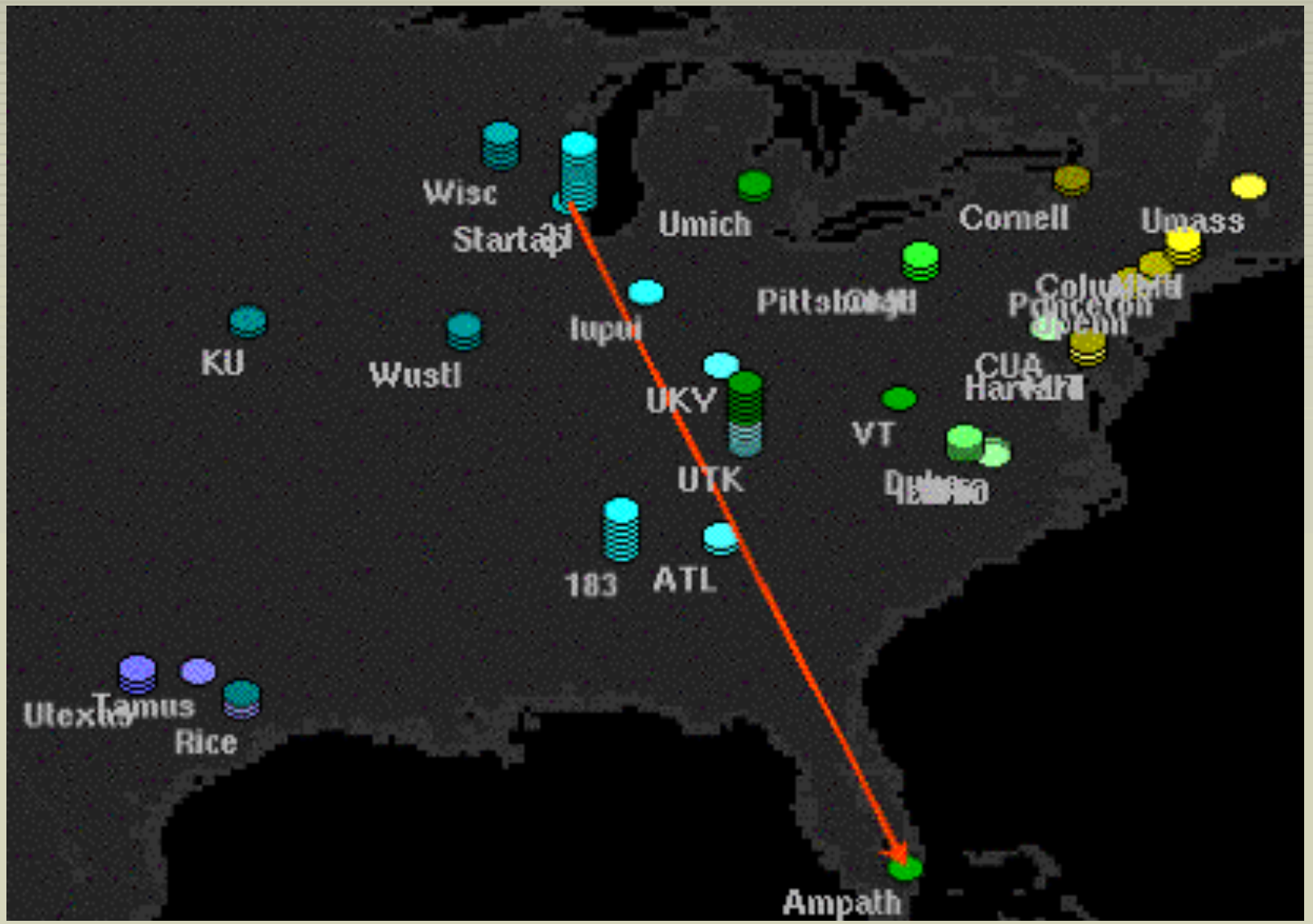

LoCI

# IBP Deployment

- Logistical Backbone
  - 147 depots in 15 countries
  - 10TB of shared storage
- Leverages Planet Lab nodes (Intel Research Labs)
- Depots/collaborations in AMPATH region
  - Puerto Rico (Guy Cormier, Univ. of Puerto Rico)
  - Brazil (Univ. of Sao Paolo)
- AMPATH Chicago/FIU 1 GB link test
  - 75 Mb/s to a depot attached at 100Mb/s

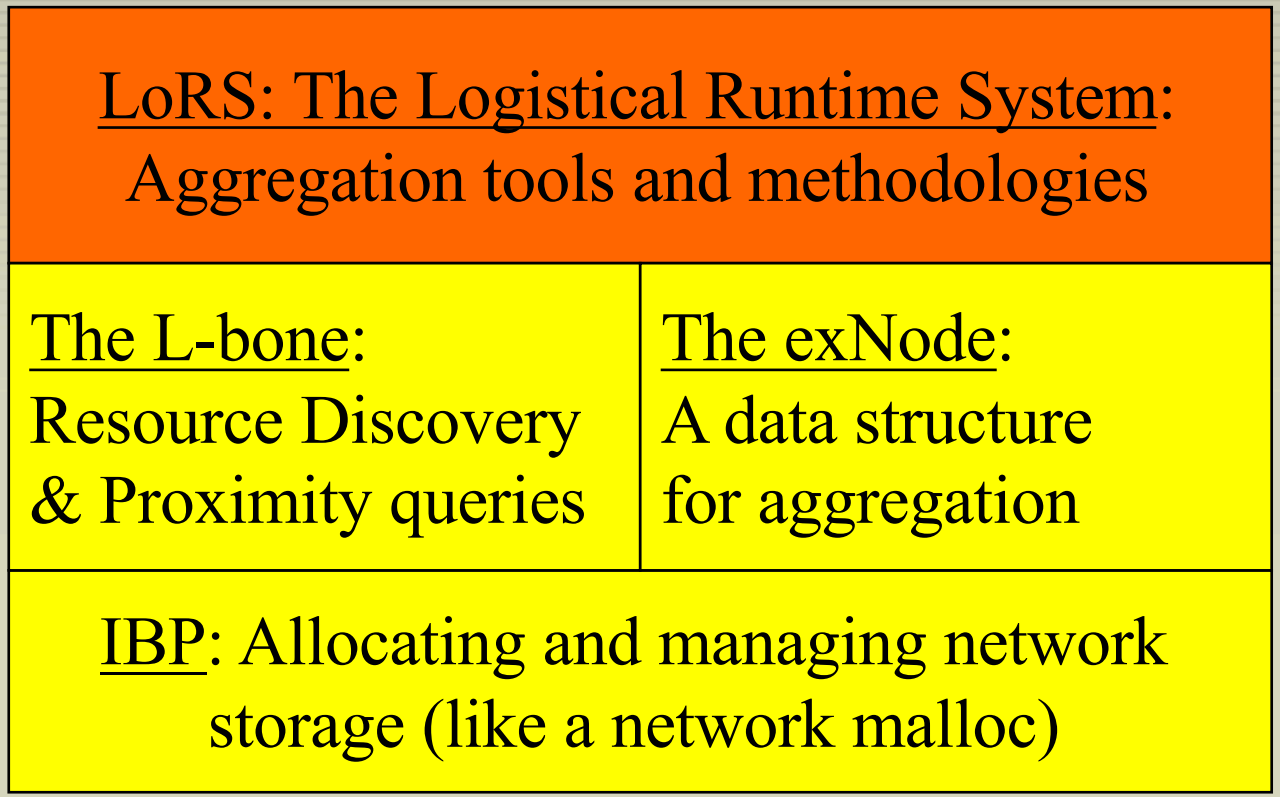
LoCI

# AMPATH-StarLight 1Gbps Link




LoCI

# The Network Storage Stack

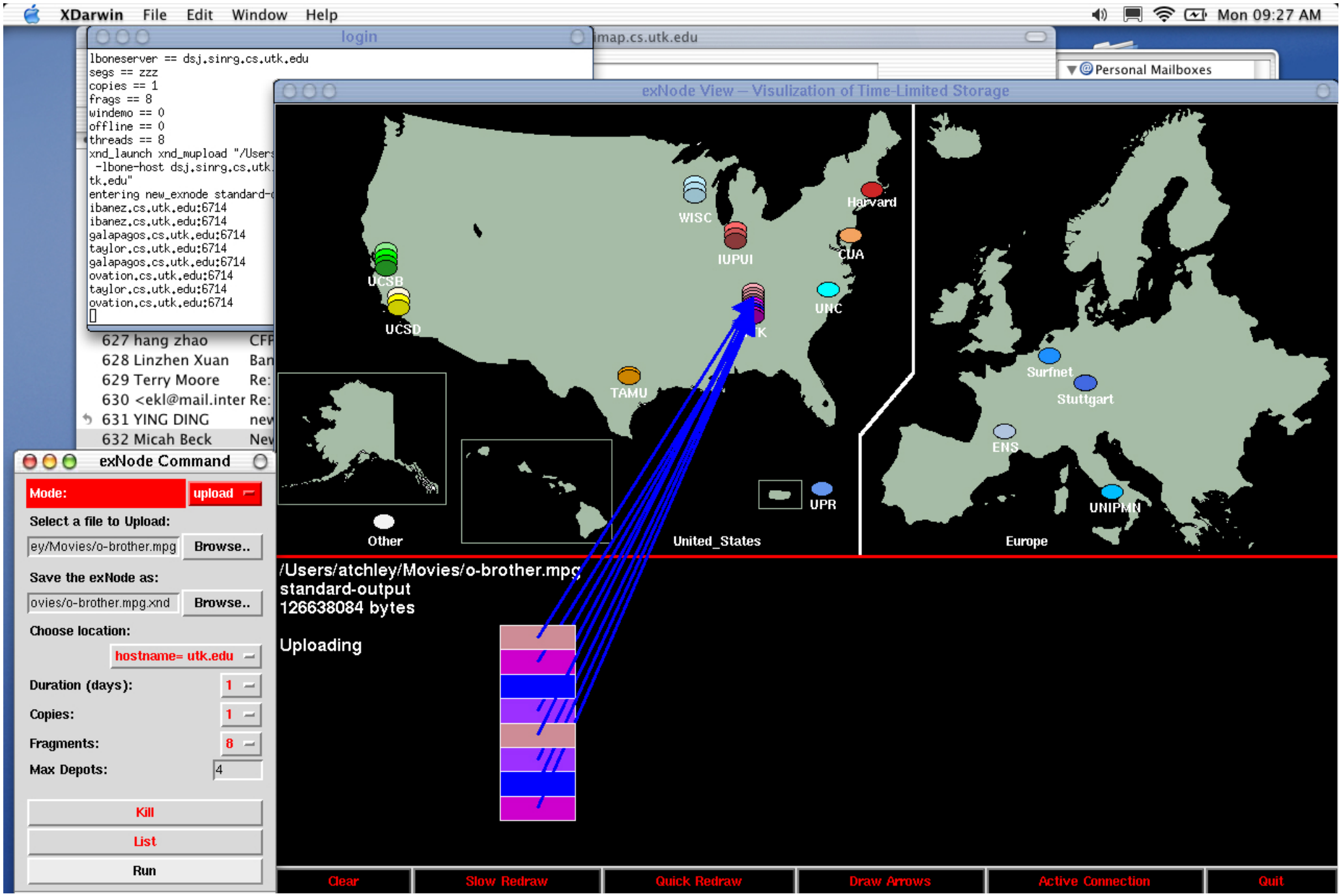


# Logistical Runtime System

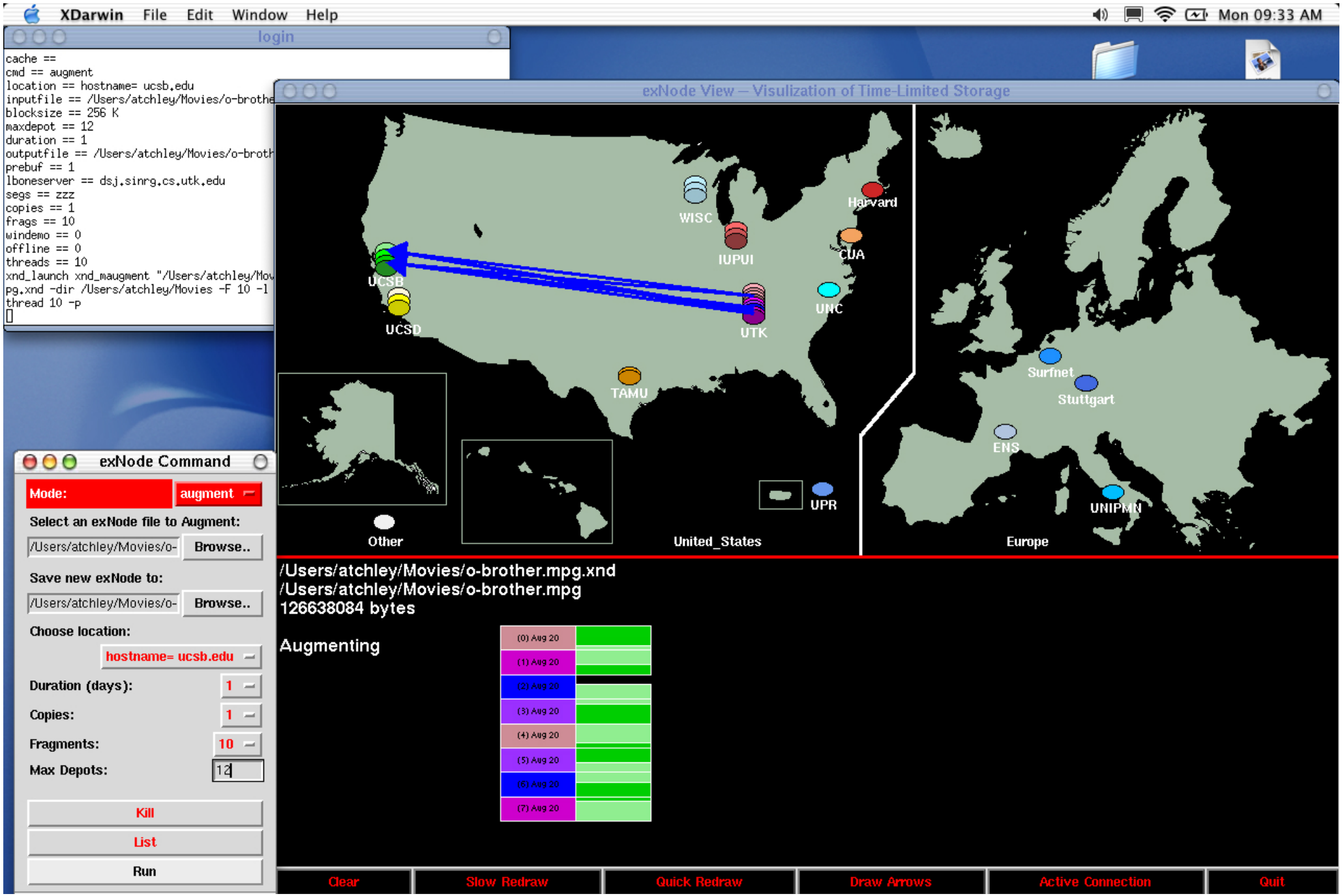
- Basic Primitives:
  - Upload, Download, Augment, Refresh
- End-to-end Services
  - Checksums, Encryption, Compression
- Other Things We Can Do
  - Routing through an intermediate depot to reduce IP RTT, speeding up TCP transfers
  - Overlay multicast using either multiple TCP streams or IP multicast at tree nodes

LoCI

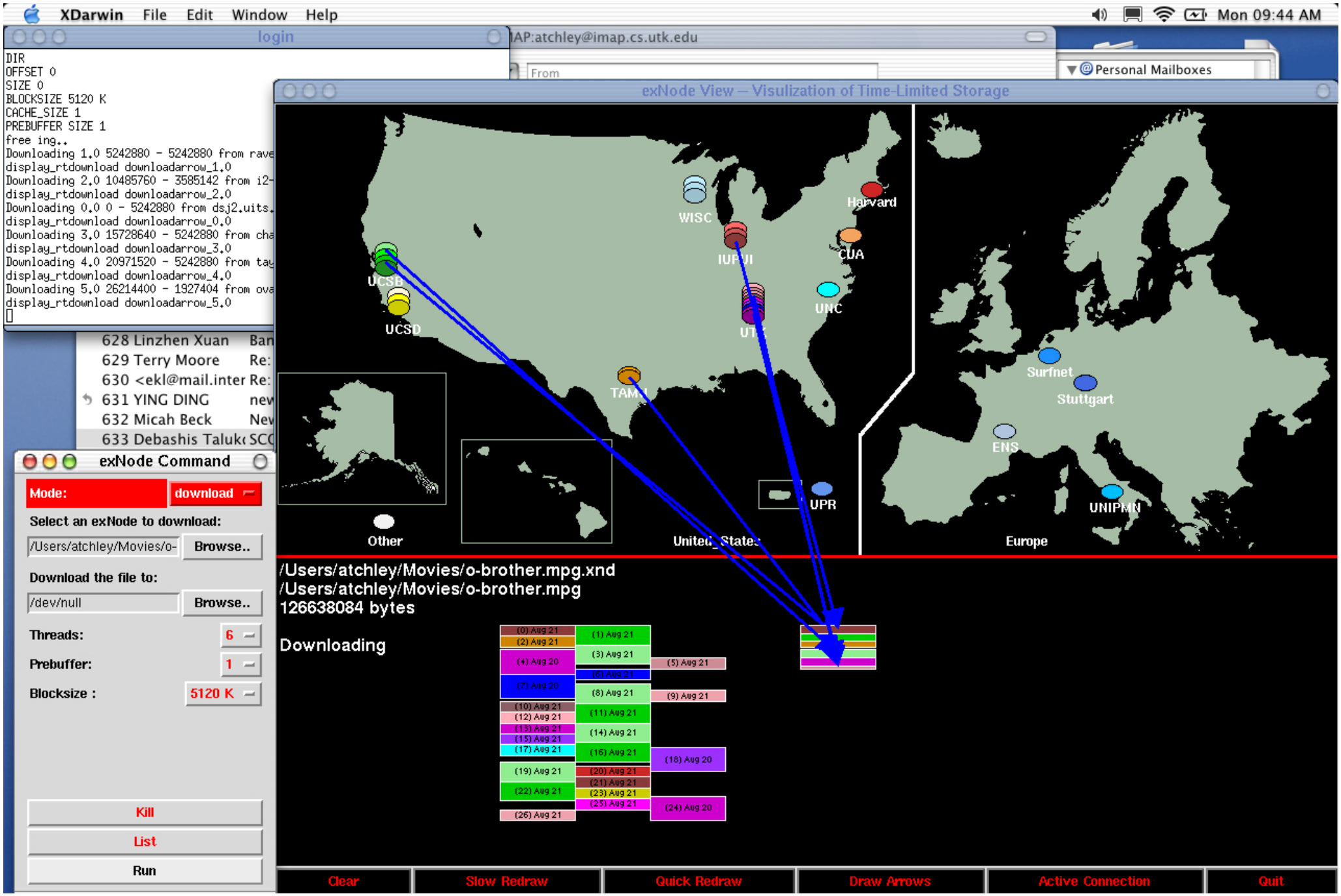
# Upload


XDarwin File Edit Window Help
Mon 09:27 AM
login
lboneserver == dsj.sinrg.cs.utk.edu
segs == zzz
copies == 1
frags == 8
windemo == 0
offline == 0
threads == 8
xnd_launch xnd_mupload "/Users
-lbone-host dsj.sinrg.cs.utk.
tk.edu"
entering new_exnode standard-
ibanez.cs.utk.edu:6714
ibanez.cs.utk.edu:6714
galapagos.cs.utk.edu:6714
taylor.cs.utk.edu:6714
galapagos.cs.utk.edu:6714
ovation.cs.utk.edu:6714
taylor.cs.utk.edu:6714
ovation.cs.utk.edu:6714
imap.cs.utk.edu
Personal Mailboxes
627 hang zhao
628 Linzhen Xuan
629 Terry Moore
630 <ekl@mail.inter
631 YING DING
632 Micah Beck
exNode View – Visulization of Time-Limited Storage
WISC
Harvard
IUPUI
CUA
UCSB
UCSD
UNC
TAMU
UPR
Other
United_States
Surfnet
Stuttgart
ENS
UNIPMN
Europe
exNode Command
Mode:
upload
Select a file to Upload:
ey/Movies/o-brother.mpg
Browse..
Save the exNode as:
ovies/o-brother.mpg.xnd
Browse..
Choose location:
hostname= utk.edu
Duration (days):
1
Copies:
1
Fragments:
8
Max Depots:
4
Kill
List
Run
/Users/atchley/Movies/o-brother.mpg
standard-output
126638084 bytes
Uploading
Clear
Slow Redraw
Quick Redraw
Draw Arrows
Active Connection
Quit

# Augment


XDarwin File Edit Window Help
Mon 09:33 AM
login
cache ==
cmd == augment
location == hostname= ucsb.edu
inputfile == /Users/atchley/Movies/o-brothe
blocksize == 256 K
maxdepot == 12
duration == 1
outputfile == /Users/atchley/Movies/o-broth
prebuf == 1
lboneserver == dsj.sinrg.cs.utk.edu
segs == zzz
copies == 1
frags == 10
windemo == 0
offline == 0
threads == 10
xnd_launch xnd_maugment "/Users/atchley/Mov
pg.xnd -dir /Users/atchley/Movies -F 10 -l
thread 10 -p
exNode View – Visulization of Time-Limited Storage
WISC
Harvard
IUPUI
CUA
UCSB
UNC
UCSD
UTK
TAMU
Surfnet
Stuttgart
ENS
UNIPMN
UPR
Other
United_States
Europe
exNode Command
Mode:
augment
Select an exNode file to Augment:
/Users/atchley/Movies/o-
Browse..
Save new exNode to:
/Users/atchley/Movies/o-
Browse..
Choose location:
hostname= ucsb.edu
Duration (days):
1
Copies:
1
Fragments:
10
Max Depots:
12
Kill
List
Run
/Users/atchley/Movies/o-brother.mpg.xnd
/Users/atchley/Movies/o-brother.mpg
126638084 bytes
Augmenting
(0) Aug 20
(1) Aug 20
(2) Aug 20
(3) Aug 20
(4) Aug 20
(5) Aug 20
(6) Aug 20
(7) Aug 20
Clear
Slow Redraw
Quick Redraw
Draw Arrows
Active Connection
Quit

# Download


XDarwin File Edit Window Help
Mon 09:44 AM
login
DIR
OFFSET 0
SIZE 0
BLOCKSIZE 5120 K
CACHE_SIZE 1
PREBUFFER SIZE 1
free ing..
Downloading 1.0 5242880 - 5242880 from rave
display_rtdownload downloadarrow_1.0
Downloading 2.0 10485760 - 3585142 from i2-
display_rtdownload downloadarrow_2.0
Downloading 0.0 0 - 5242880 from dsj2.uits.
display_rtdownload downloadarrow_0.0
Downloading 3.0 15728640 - 5242880 from cha
display_rtdownload downloadarrow_3.0
Downloading 4.0 20971520 - 5242880 from tay
display_rtdownload downloadarrow_4.0
Downloading 5.0 26214400 - 1927404 from ova
display_rtdownload downloadarrow_5.0
IMAP:atchley@imap.cs.utk.edu
From
Personal Mailboxes
628 Linzhen Xuan
629 Terry Moore
630 <ekl@mail.inter
631 YING DING
632 Micah Beck
633 Debashis Taluk
exNode View – Visulization of Time-Limited Storage
WISC
Harvard
UCSB
UCSD
CUA
UNC
TAMU
UPR
Surfnet
Stuttgart
ENS
UNIPMN
Other
United States
Europe
exNode Command
Mode:
download
Select an exNode to download:
/Users/atchley/Movies/o-
Browse..
Download the file to:
/dev/null
Browse..
Threads:
6
Prebuffer:
1
Blocksize :
5120 K
Kill
List
Run
/Users/atchley/Movies/o-brother.mpg.xnd
/Users/atchley/Movies/o-brother.mpg
126638084 bytes
Downloading
Clear
Slow Redraw
Quick Redraw
Draw Arrows
Active Connection
Quit

# IBP Enables Data Intensive Collaboration

- Large files can be uploaded to nearby depots, then managed by movement between depots
  - End systems are not involved in long distance transfers
- Data can be moved near to distant collaborator without being downloaded into their end system
  - Direct access to collaborators private storage is not required
- Depot-to-depot transfers can take advantage of multithreading, UDP transfer, Net/Web 100, other high-performance optimizations

LoCI

# Example Application: IBPvo

- Web interface allows television shows to be recorded in U.S., uploaded to IBP depots
- Resulting AVI files are O(1GB) in size
- ExNode is delivered to user by mail
- Multithreaded transfer to APAN region depots
- Users watch programs by downloading to their own workstations, viewing locally
- A reciprocal service would allow users in U.S. direct access to AMPATH region television
- **`http://promise.sinrg.cs.utk.ed/votest`**

LoCI

# Other Areas of Application

- Management of massive data sets
  - Produced by simulation
  - Captured from experimentation
  - Generated by sensors and instruments
- Caching and staging of of data in high-performance wide are (e.g. Grid) computation
- Content Distribution of highly popular content
- Overlay routing, multicast
- Digital Libraries
- Checkpoints and backups
- Wide area file systems

LoCI

# LoCI Lab Online

## http://loci.cs.utk.edu

- IBP server and clients for Unix/Linux/OS X
  - Additional clients for Java, Win32
- Logistical Runtime System libraries and tools
  - Run under Unix/Linux/OS X natively
  - Ported to Windows under Cygwin
  - Includes visualization (Tcl/tk)
  - Web interface
- Logistical Backbone resource discovery server
  - Unix/Linux/OS X only
- Publications, documentation, L-Bone status

LoCI