Name Space, Migration, and Replication in NFSv4

* Jiaying Zhang
Center for Information Technology Integration
University of Michigan
Ann Arbor

Motivation

- Improve wide-area interoperation & access to distributed storage
- Facilitate file system administration

Goals

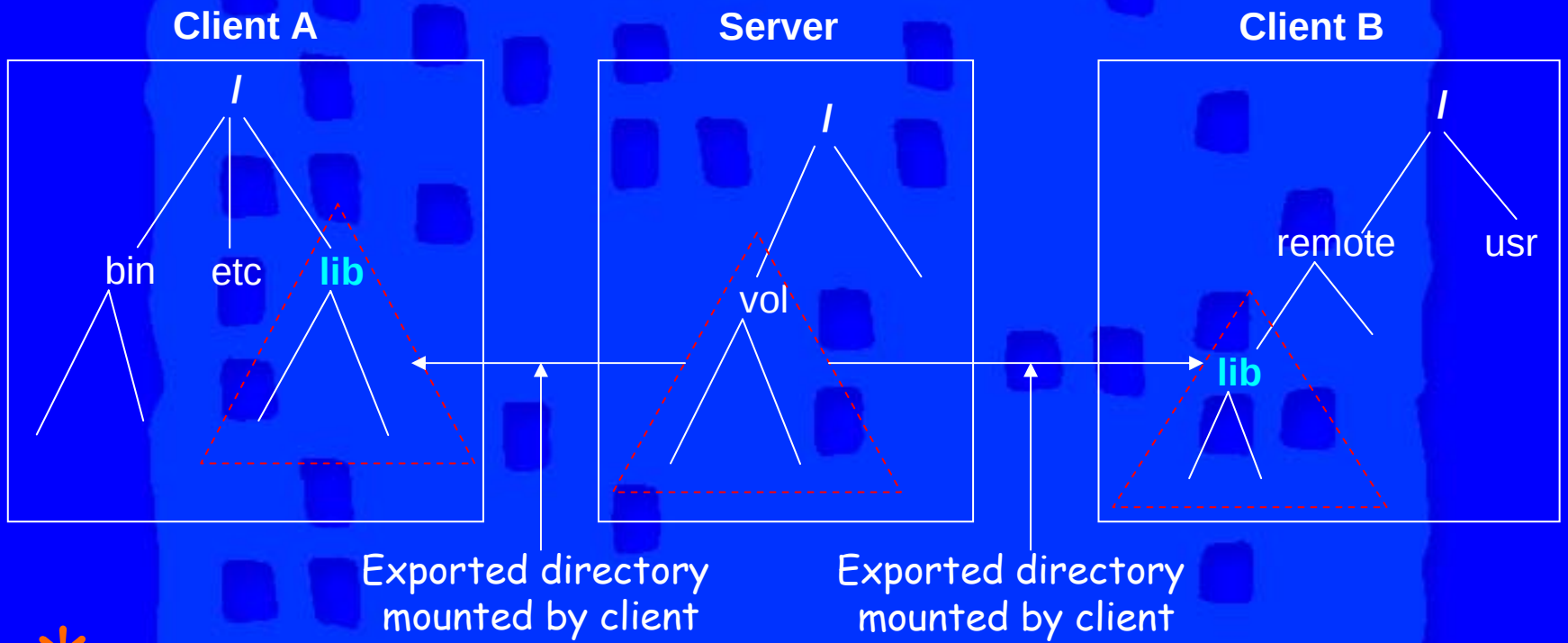
- Global name space
  - Provides a common frame of reference
- Migration
  - Simplify file system administration
  - Load balancing
- Replication
  - Multiple copies improve performance & availability
- Mutable replication
  - Users should be able to modify data as needed

Mechanisms

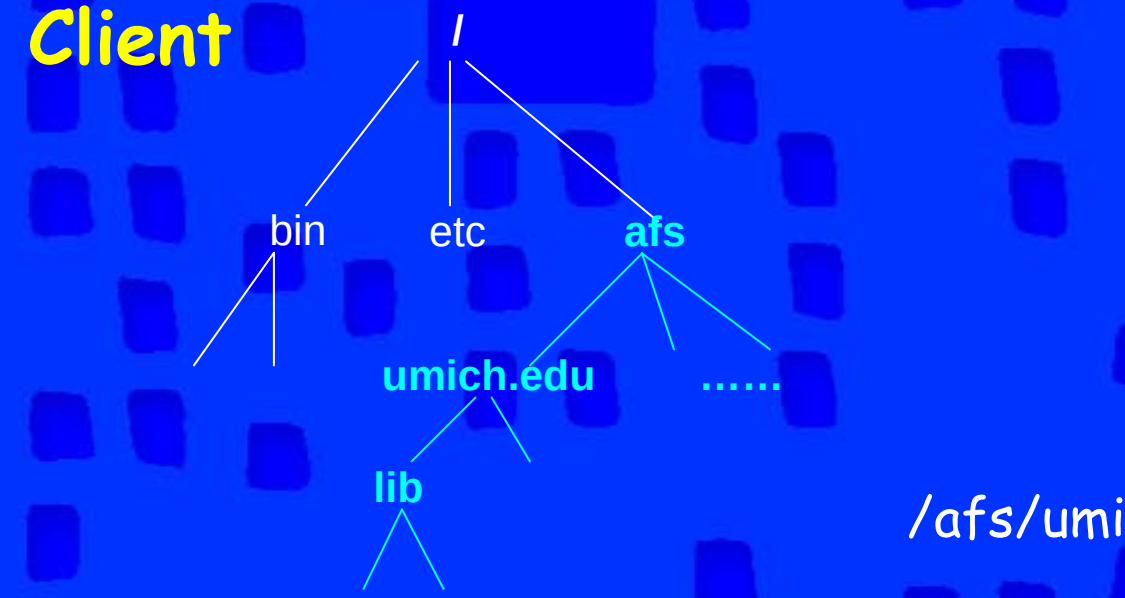
- Global name space, file system migration and replication
  - DNS resolution
- Directory migration and replication
  - FS_LOCATIONS attribute
- Mutable Replication
  - Server redirection

Global Name Space – NFSv3

- No global name space support



Global Name Space — AFS



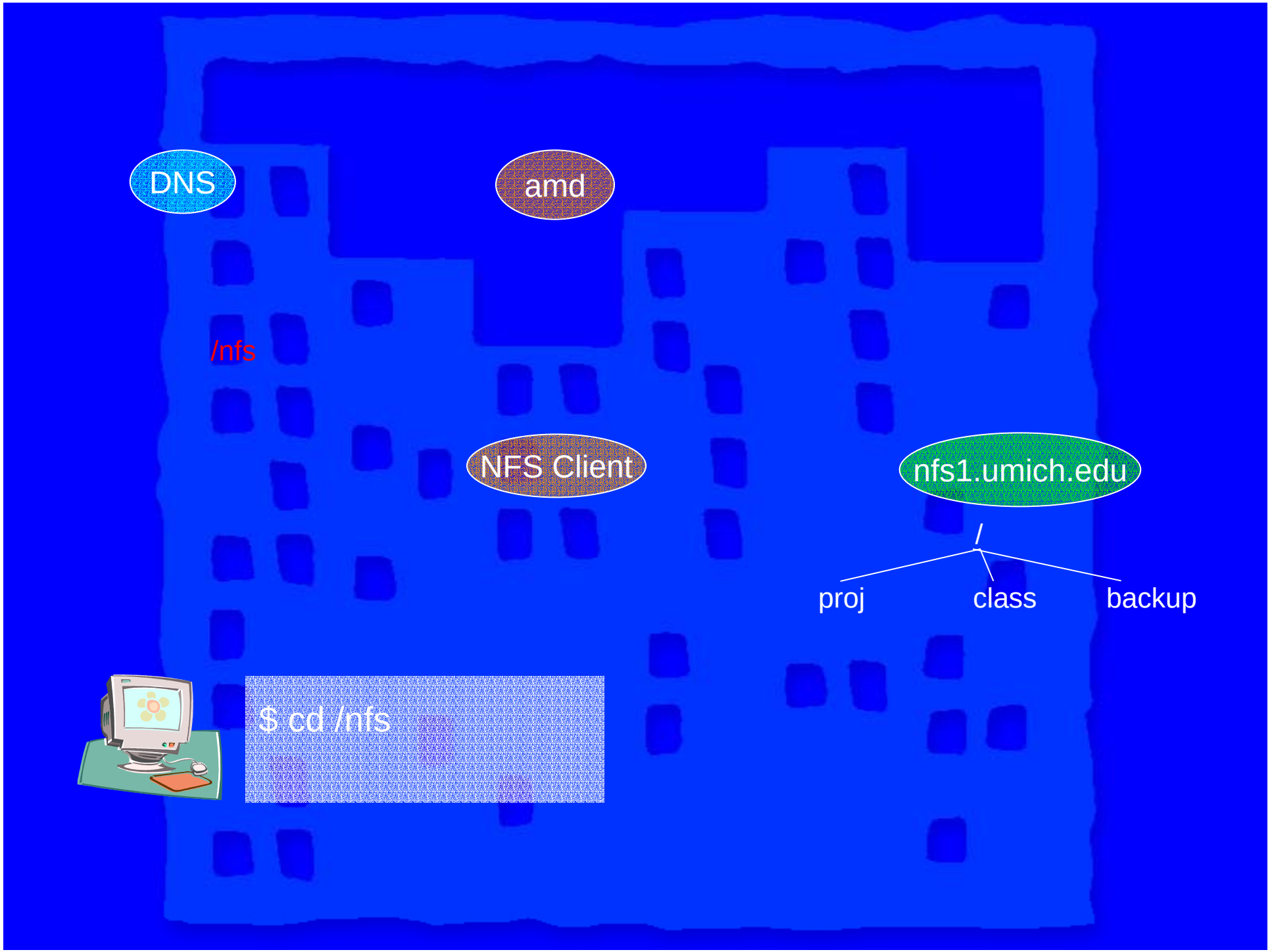
- We want to provide NFS users a similar view

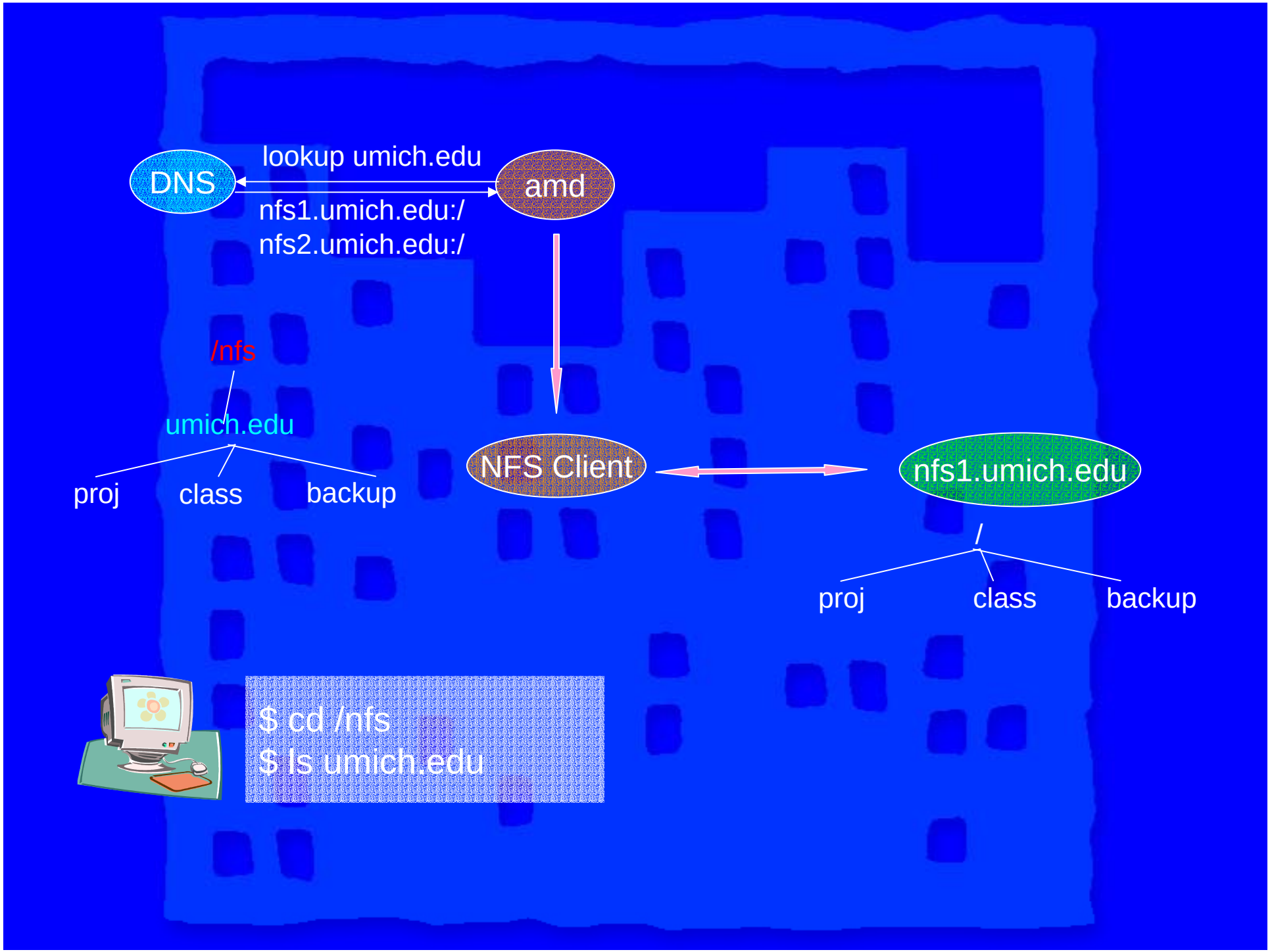
Global Name Space — NFSv4

- /nfs
  - Global root of all NFS file systems
  - Holds recently accessed NFS file systems
- Entries under /nfs
  - Mounted on demand
  - Referred following DNS conventions:
    /nfs/umich.edu/lib/file1

Mechanisms

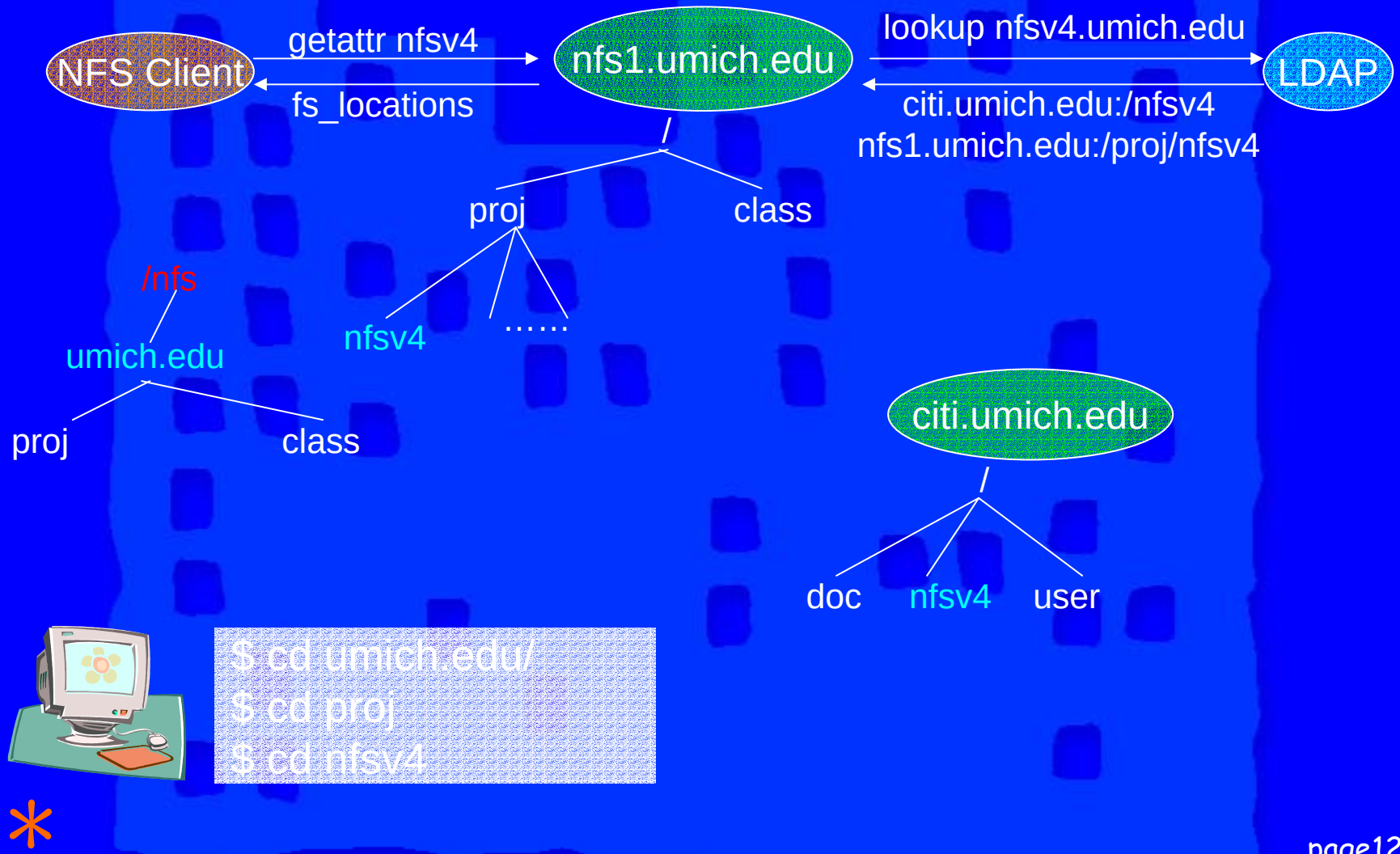
- DNS server maps logical name to NFS server locations
  - TXT RR or SRV RR
- Extend AMD to support DNS query
- Provides file system level migration & (read-only) replication

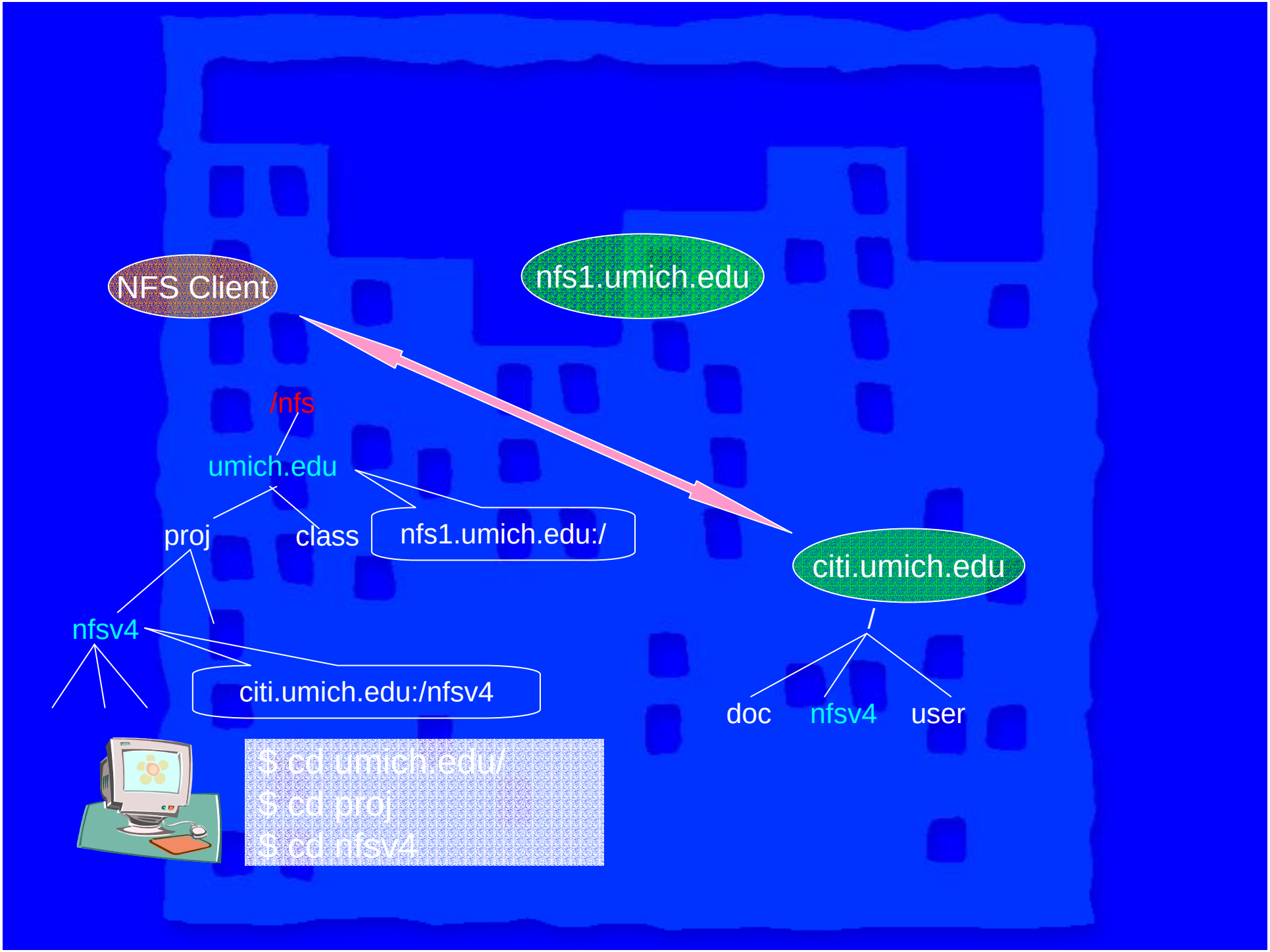

DNS
amd
/nfs
NFS Client
nfs1.umich.edu
/
proj
class
backup
$ cd /nfs


DNS
lookup umich.edu
amd
nfs1.umich.edu:/
nfs2.umich.edu:/
/nfs
umich.edu
proj
class
backup
NFS Client
nfs1.umich.edu
/
proj
class
backup
$ cd /nfs
$ ls umich.edu

Directory Migration and Replication

- Attach a reference string to each migrated/replicated directory
- Support multiple lookup methods
  - LDAP, DNS, FILE or SERVER REDIRECTION
- Server sends client replica locations through fs_locations attr
- Client selects a replica and mounts it


NFS Client
getattr nfsv4
fs_locations
nfs1.umich.edu
lookup nfsv4.umich.edu
LDAP
citi.umich.edu:/nfsv4
nfs1.umich.edu:/proj/nfsv4
/
proj
class
nfsv4
.....
/nfs
umich.edu
proj
class
citi.umich.edu
/
doc
nfsv4
user
$ cd umich.edu/
$ cd proj
$ cd nfsv4


NFS Client
nfs1.umich.edu
/nfs
umich.edu
proj
class
nfs1.umich.edu:/
citi.umich.edu
nfsv4
citi.umich.edu:/nfsv4
doc
nfsv4
user
$ cd umich.edu/
$ cd proj
$ cd nfsv4

Migration and Active Client

- Server sends client NFSERR_MOVED
- Client gets new location through FS_LOCATIONS attribute
- Client remounts the specified server
- Client side recovery – similar to server reboot recovery

Mutable Replication

- Make common accesses fast
  - **Exclusive read**: most often
  - **Shared read**: common
  - **Exclusive write**: less common
  - **Write with concurrent access**: infrequent
  - **Server failure and network partition**: rare

Mutable Replication

- Absent writes: clients read nearby servers
- Client write-opens a file: the connected server
  - Disables replications on other replicas
  - Becomes the temporary primary for the file
- Concurrent writes: direct all accesses to the primary server
  - FS_LOCATIONS attribute

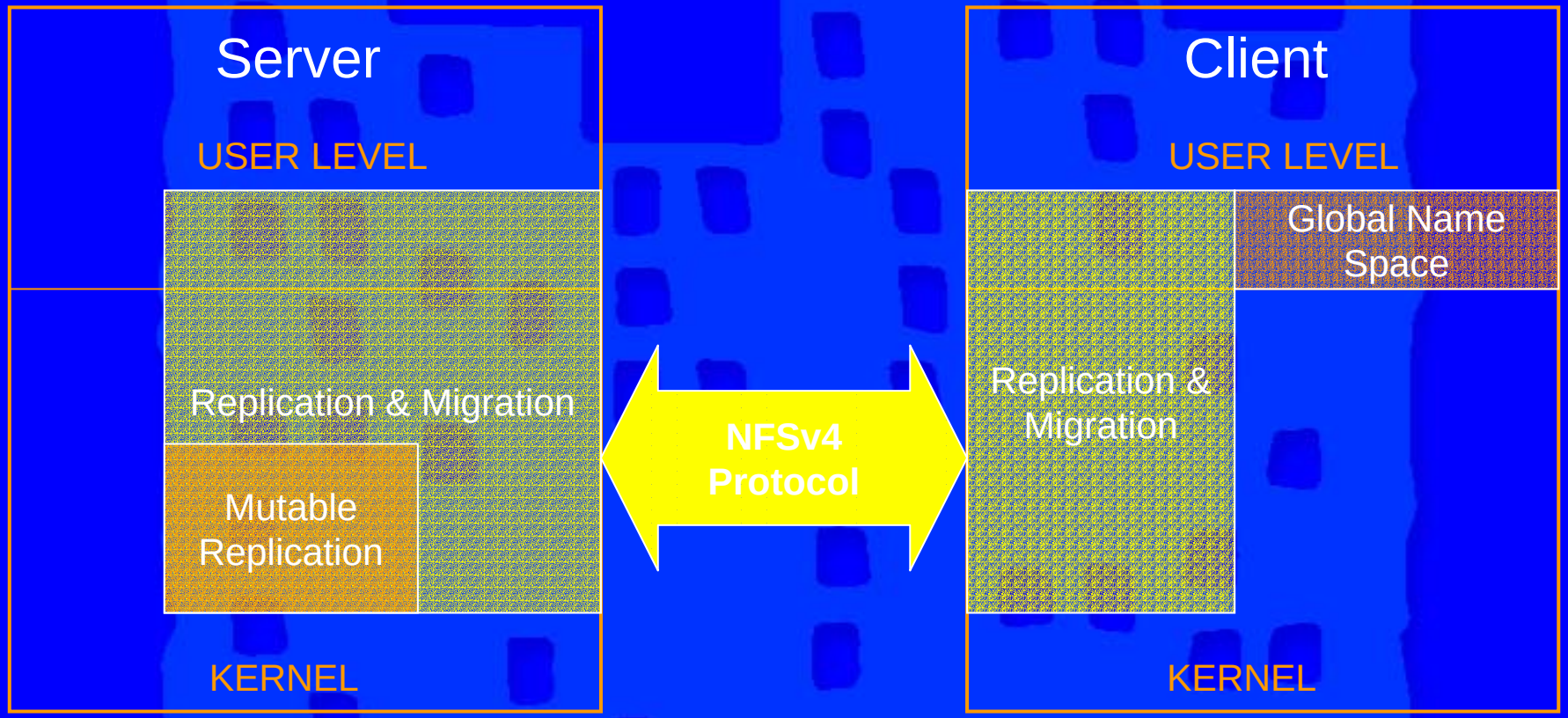
Consistency Guarantees

- Assure sequential consistency with view-based mechanism
- Can support strict consistency by
  - Disabling writes before failure is recovered
- No overhead when free of failure
- Simple client recovery procedure
- Details and correctness proof:
  http://www.citi.umich.edu/techreports/reports/citi-tr-04-1.pdf

Mutable Replication

- Allow users to modify data when they need to
- Read (usually) comes at **no additional cost**
- Write performance is (generally) fabulous
- Can support strong consistency

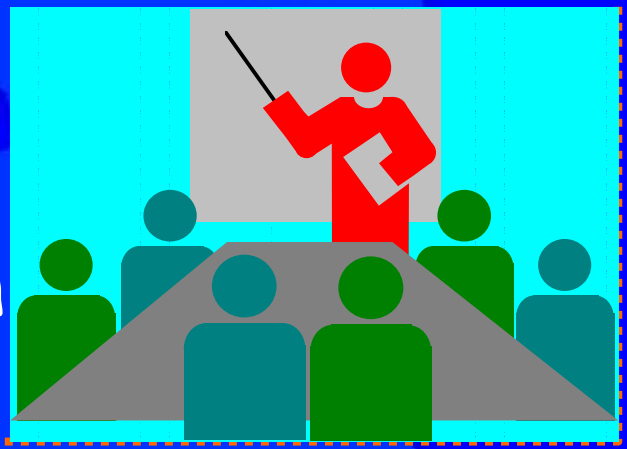
Putting It All Together



We Are Looking For …

- Your feedback
- Use cases to help us evaluate our design
- Application workloads to help us evaluate performance

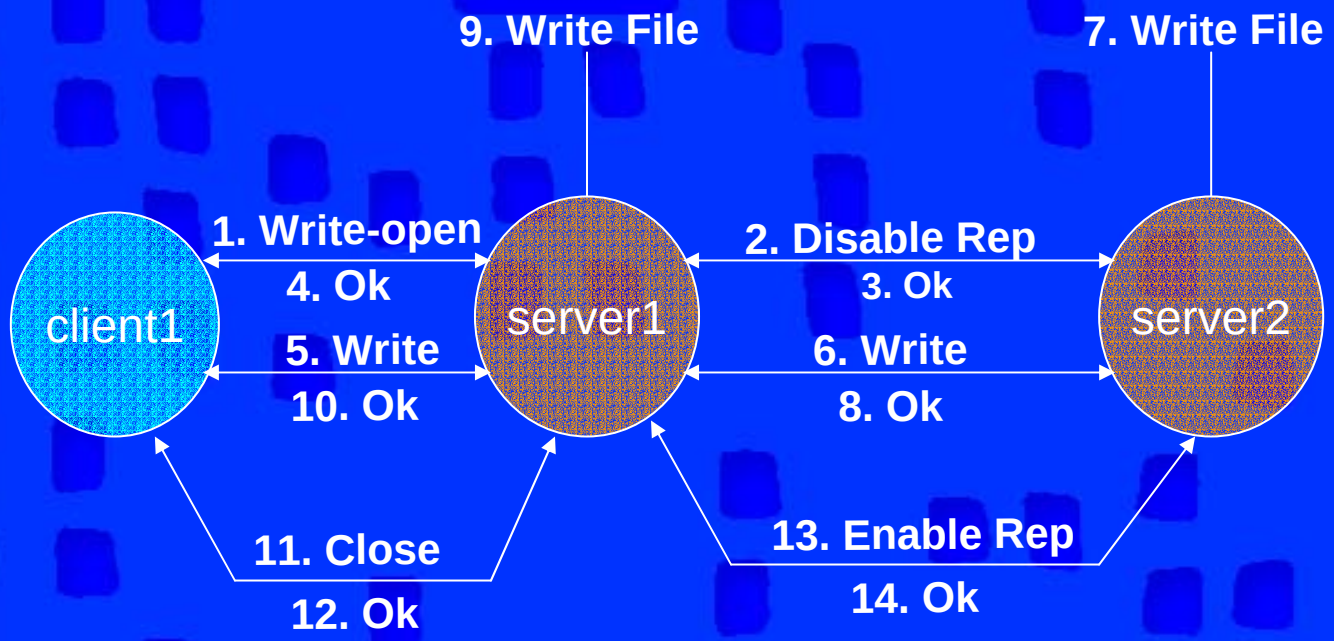
- Thank you for your attention



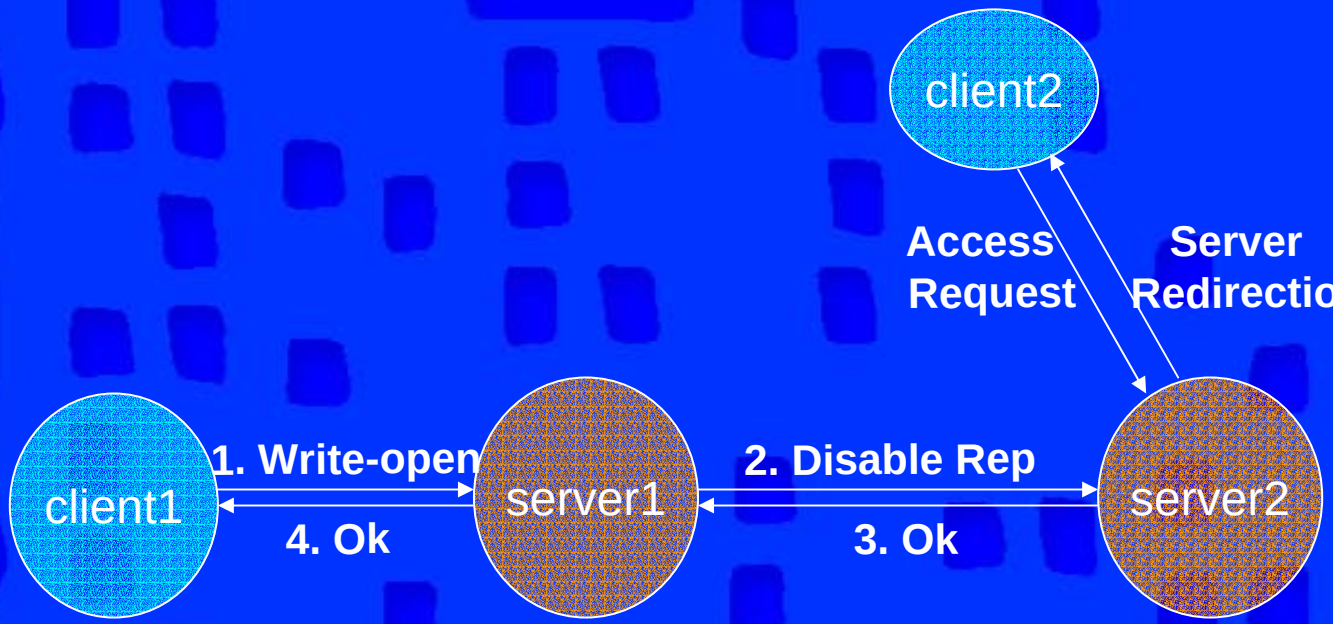
Reference String

- LDAP:
  - ldap://ldapserver/lookup-key [-b searchbase] [-p ldapport]
- DNS
  - dns://lookup-name
- FILE
  - file://pathname/lookup-key
- SERVER REDIRECT
  - server://hostname:/path [mount-options]

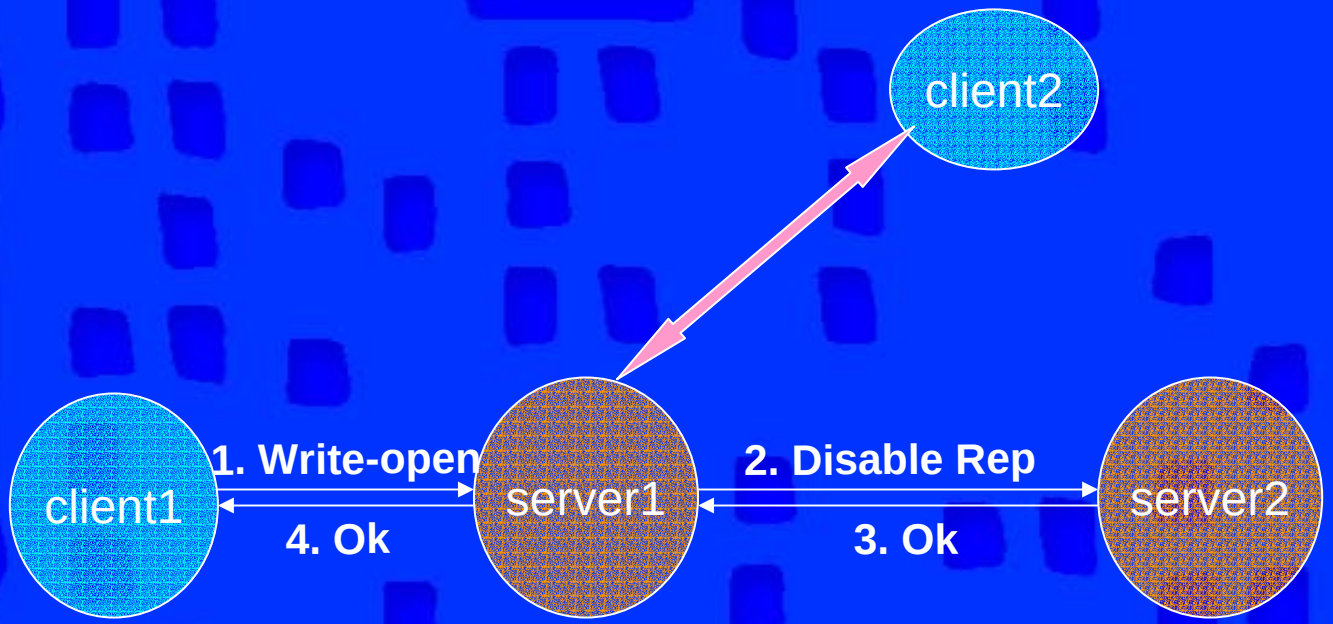
File Modification



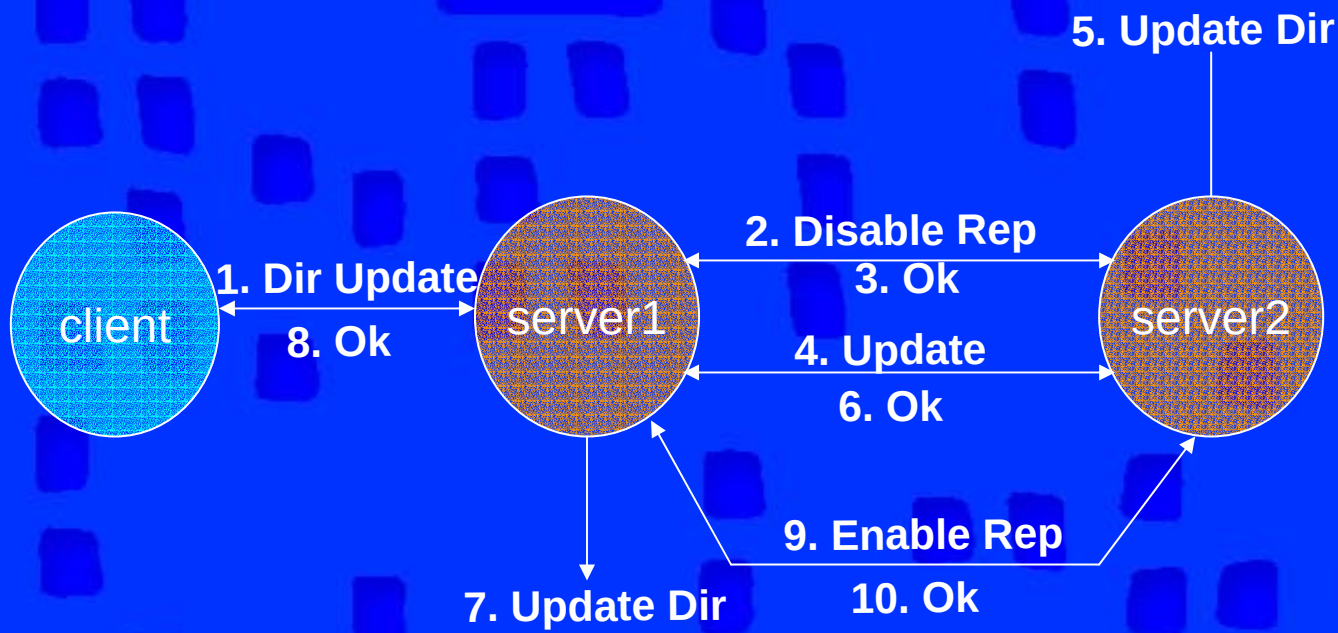
File Modification



File Modification



Directory Modification



Directory Modification