



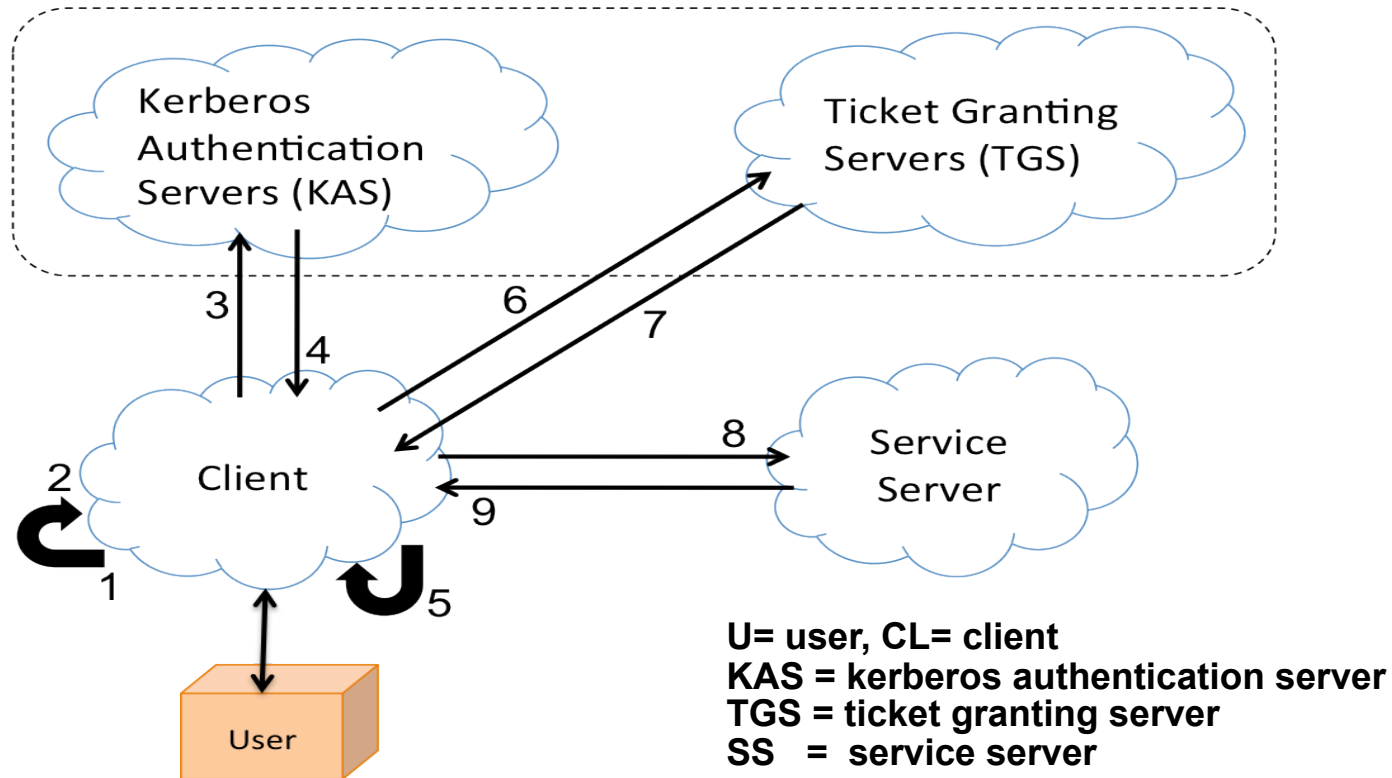
Kerberized Lustre 2.0 over the WAN

Josephine Palencia, Robert Budden, Kevin Sullivan
Pittsburgh Supercomputing Center

Content

- › Kerberos V5 primer
 - Authenticated lustre components**
- › Setup: Single Kerberos realms->Cross-realm
 - Partners: SDSC@TeraGrid, NRL, Uflorida**
- › Distributed OST/OST pools
 - Arranging user directories**
- › Lustre 2.0 tests
 - Available systems: Quota, ACL**
 - Kerberized scp/kftp/gridftp with konFUSEd**
 - Current constraints**
- › Integrating with kerberized NFS4
- › Summary
- › References

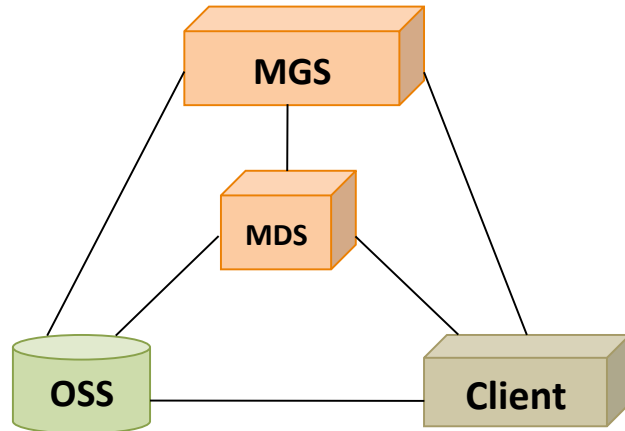
Kerberos V5 Primer



Client

- authenticates itself to KAS,
- demonstrates to TGS that it's authorized to receive a ticket for a service
- demonstrates to SS that it has been approved to receive service

Authenticated lustre components



```

lctl conf_param lwan.srpc.flavor.default=krb5p
lctl conf_param lwan.srpc.flavor.tcp0=krb5n
lctl conf_param lwan.srpc.flavor.tcp1=null
lctl conf_param lwan.srpc.flavor.default.cli2ost=krb5i
lctl conf_param lwan.srpc.flavor.default.mdt2mdt=null
lctl conf_param lwan.srpc.flavor.default.mdt2ost=krb5i
lctl conf_param _mgs.srpc.flavor.default=krb5p
  
```

| Flavor | Auth | RPC message protection | Bulk Data Prot |
|--------|----------|------------------------|----------------------|
| null | | | |
| plain | | null | checksum(adler32) |
| krb5n | GSS/krb5 | null | checksum(adler32) |
| krb5a | GSS/krb5 | partly integrity | checksum(adler32) |
| krb5i | GSS/krb5 | integrity | integrity(sha1) |
| krb5p | GSS/krb5 | privacy | privacy(sha1/aes128) |

```

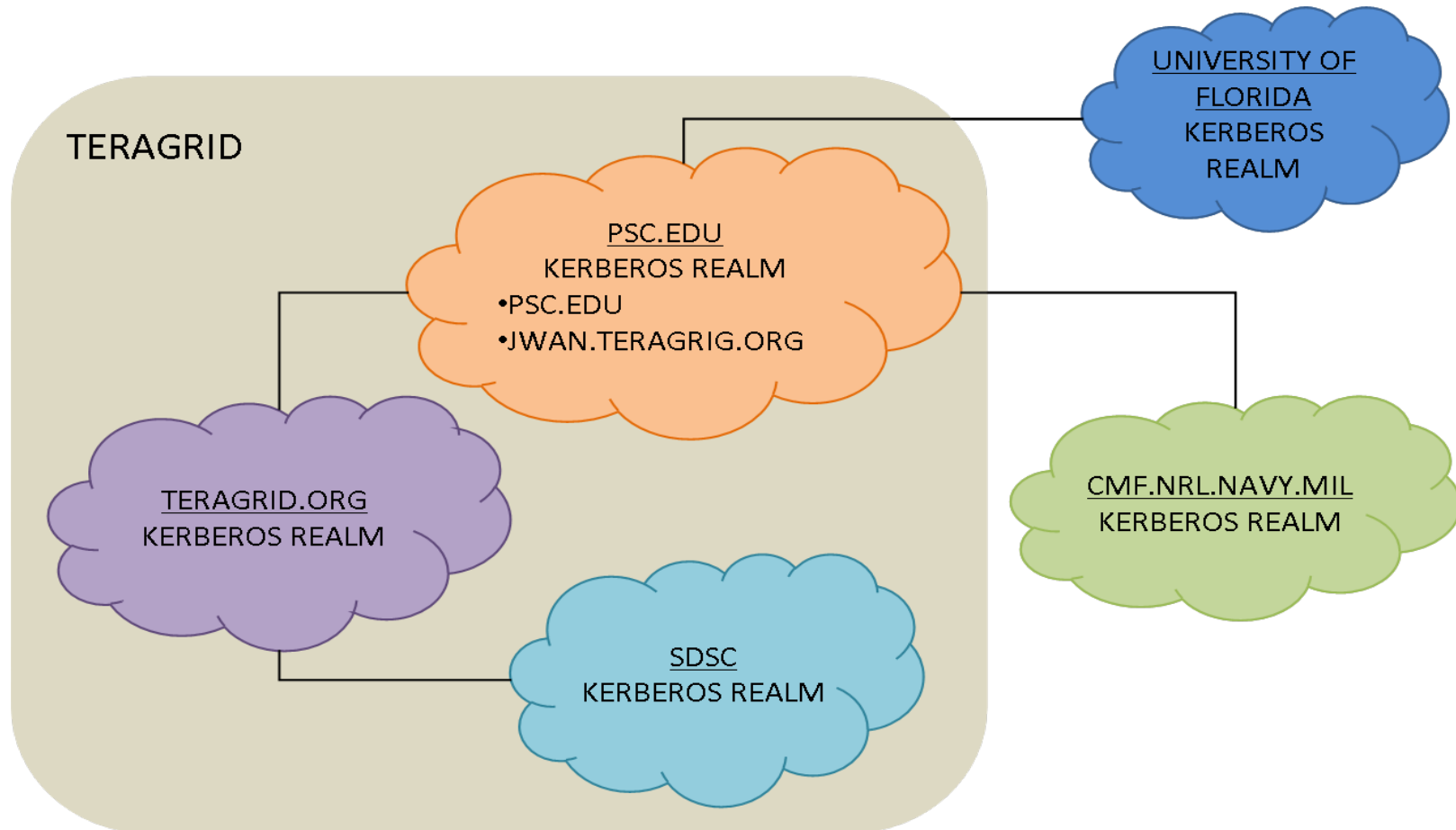
lustre_root/goldeneye.jwan.teragrid.org@TERAGRID.ORG (AES-256 CTS mode with 96-bit SHA-1 HMAC)
nfs/goldeneye.jwan.teragrid.org@TERAGRID.ORG (AES-256 CTS mode with 96-bit SHA-1 HMAC)
  
```

(lustre_mgs, lustre_mds, lustre_oss, lustre_root)

- Shift among kerberos flavors during mounts
- Preliminary IO benchmark comparing different flavors to look at performance overhead by kerberos

Setup

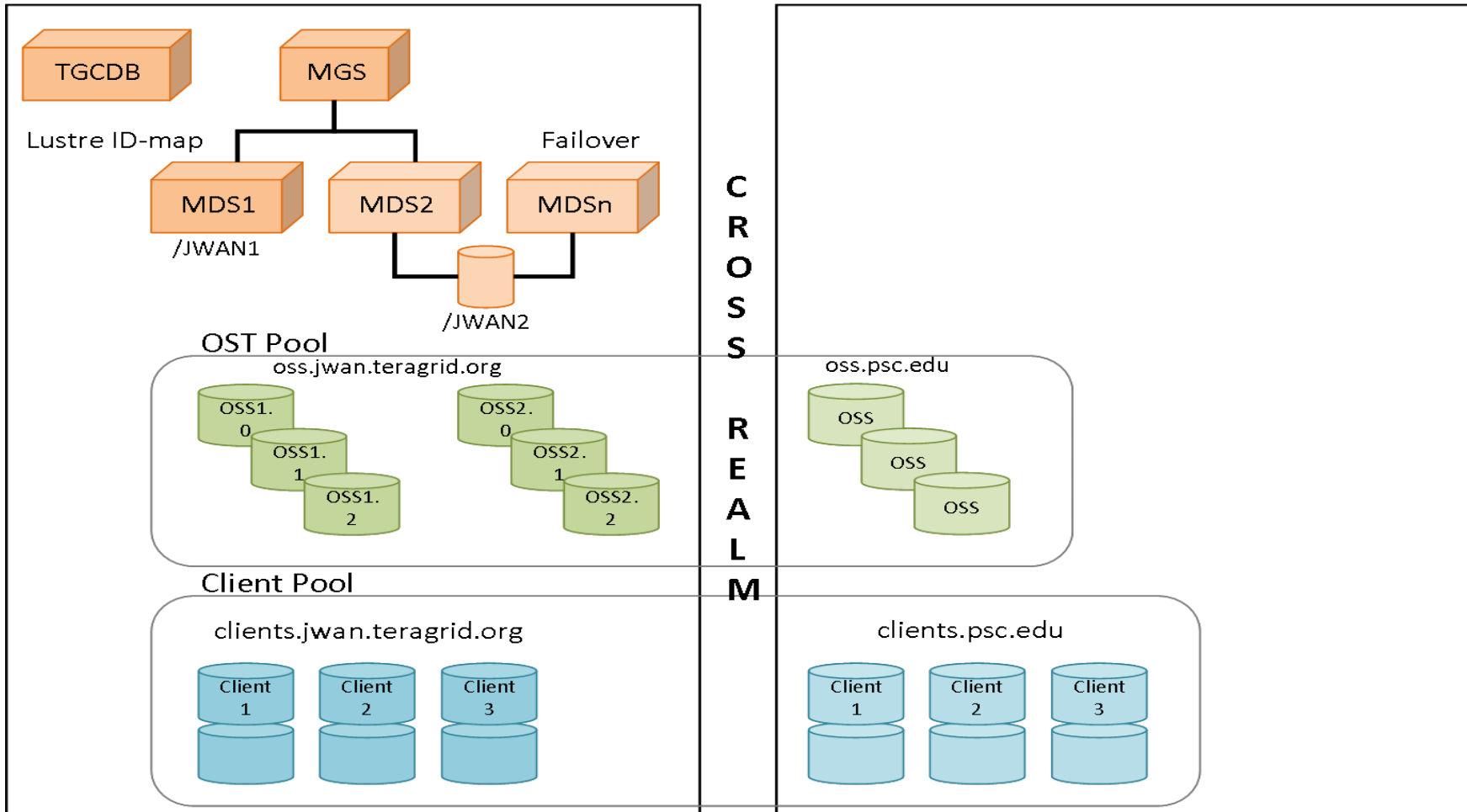
- Lustre 2.0 Alpha 5 (CentOS5)
- Release: 1.9.280
- Kernel: 2.6.18-128.7.1
- Kerberos auth: krb5p



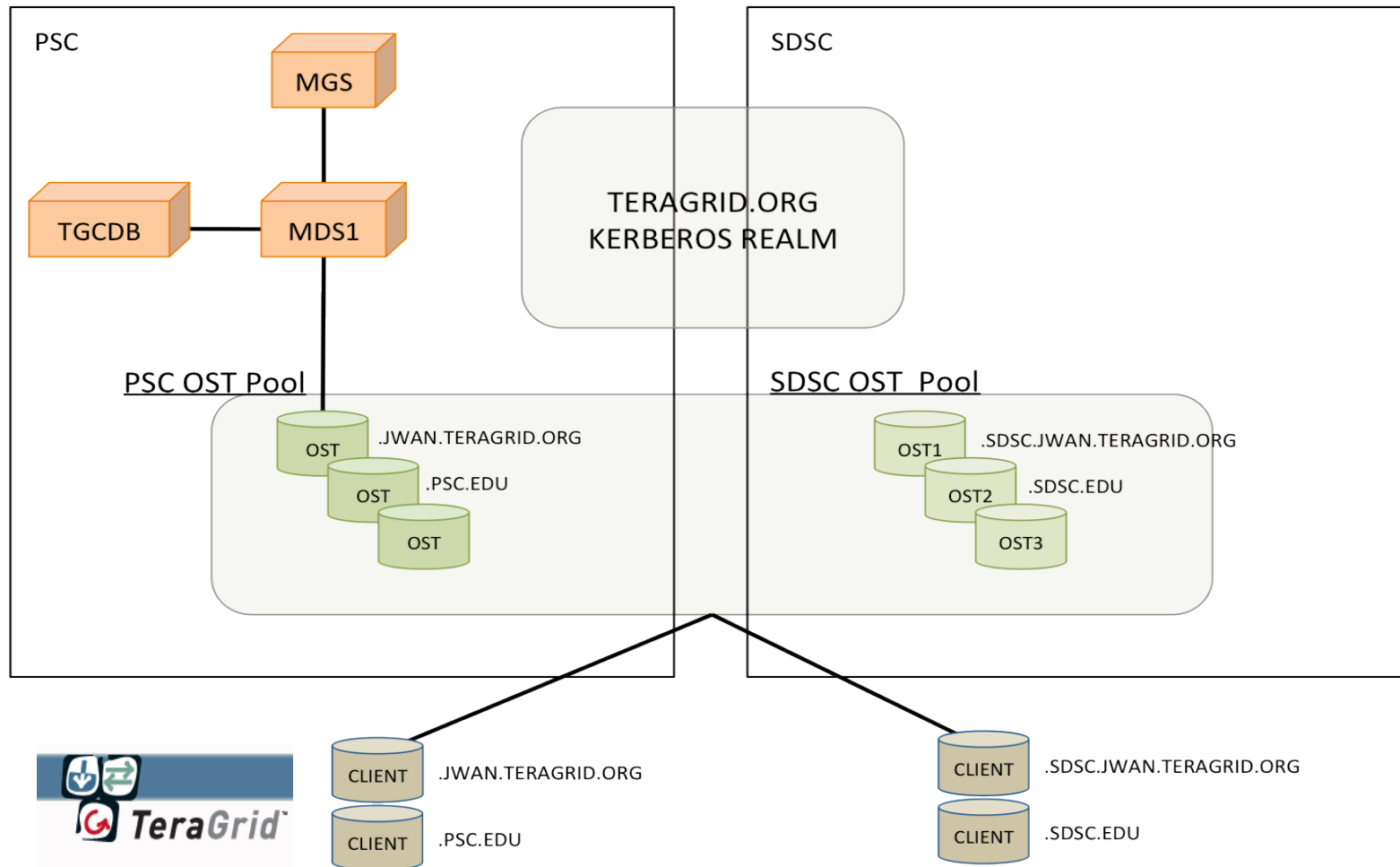
@PSC

TERAGRID.ORG KERBEROS REALM
DOMAIN: JWAN.TERAGRID.ORG

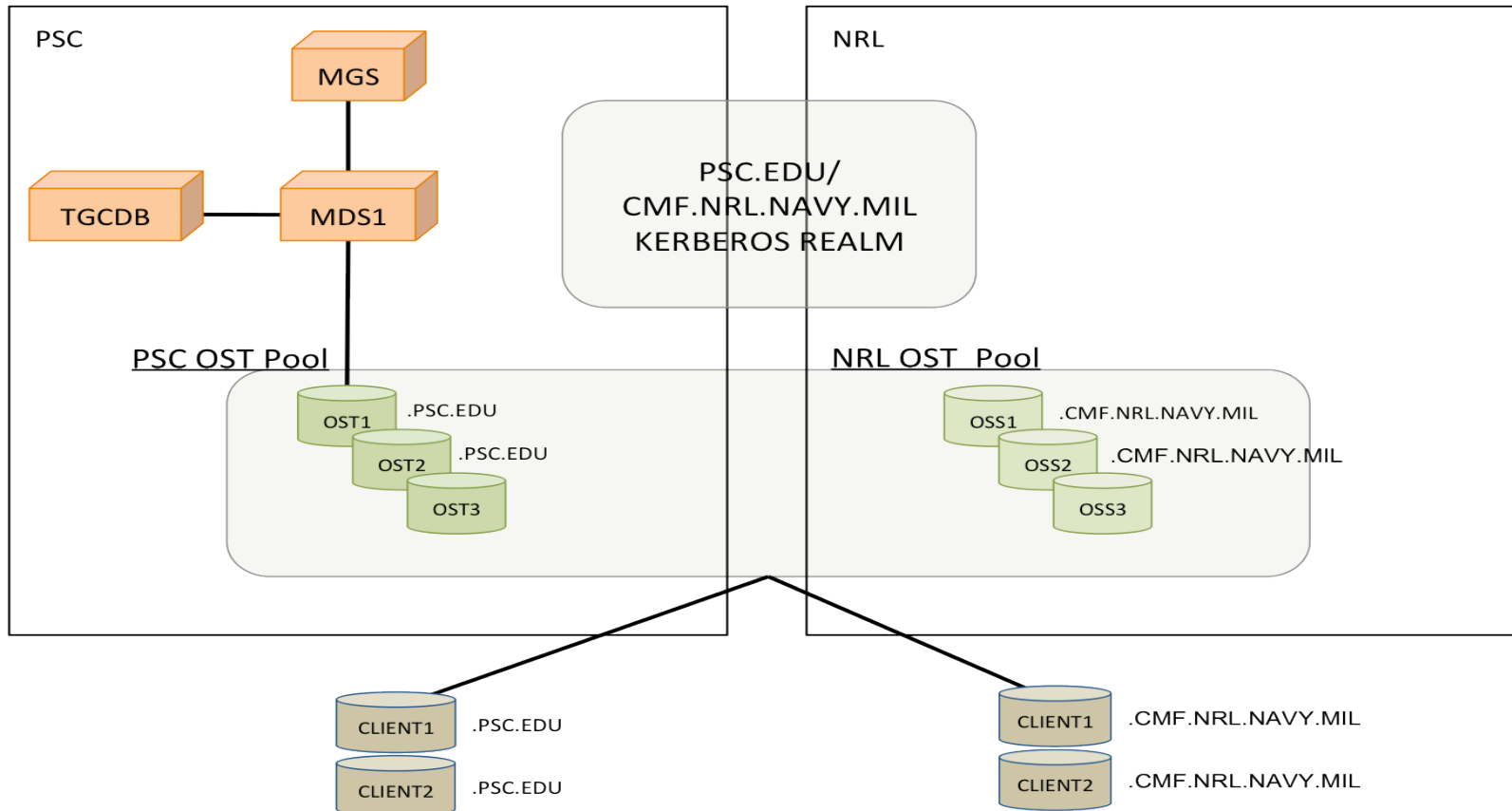
PSC.EDU KERBEROS REALM
DOMAIN: PSC.EDU



@PSC + SDSC

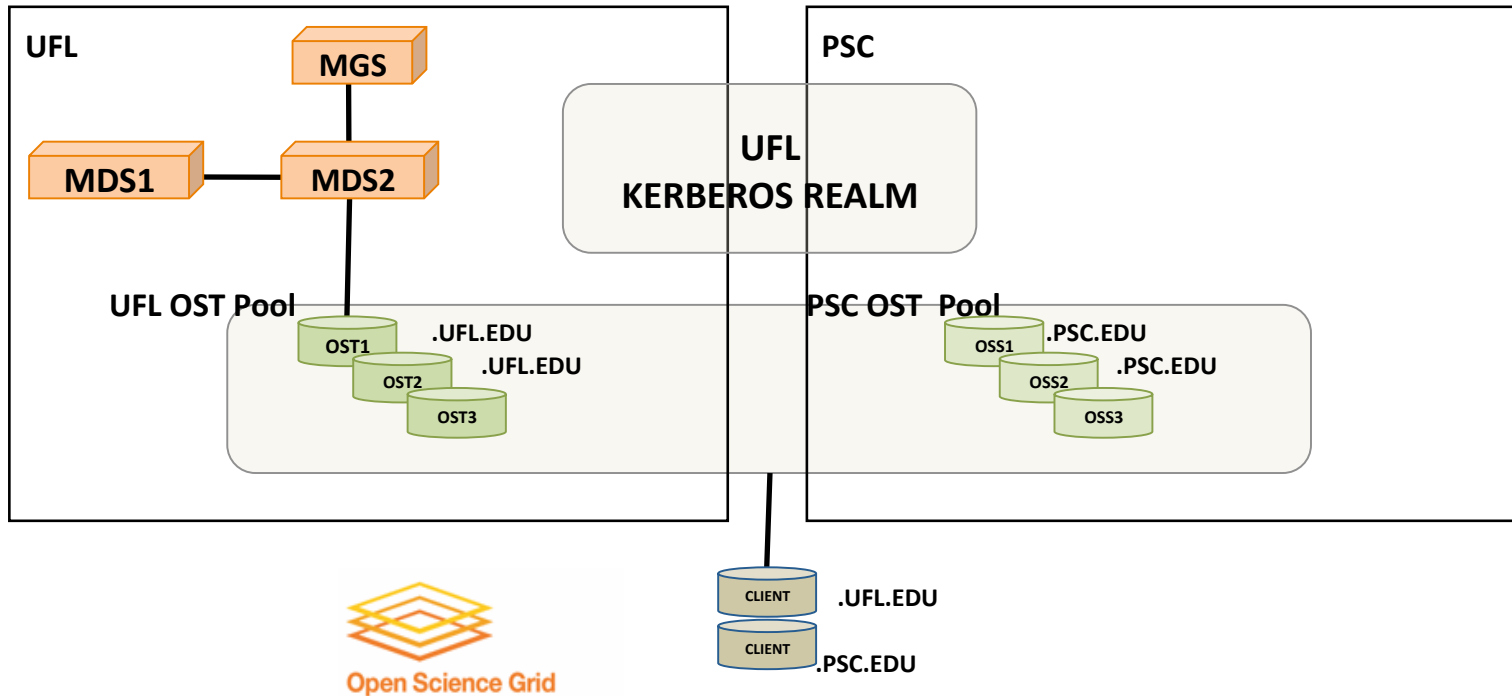


@PSC + NRL



Cross-Realm kerberos authentication work

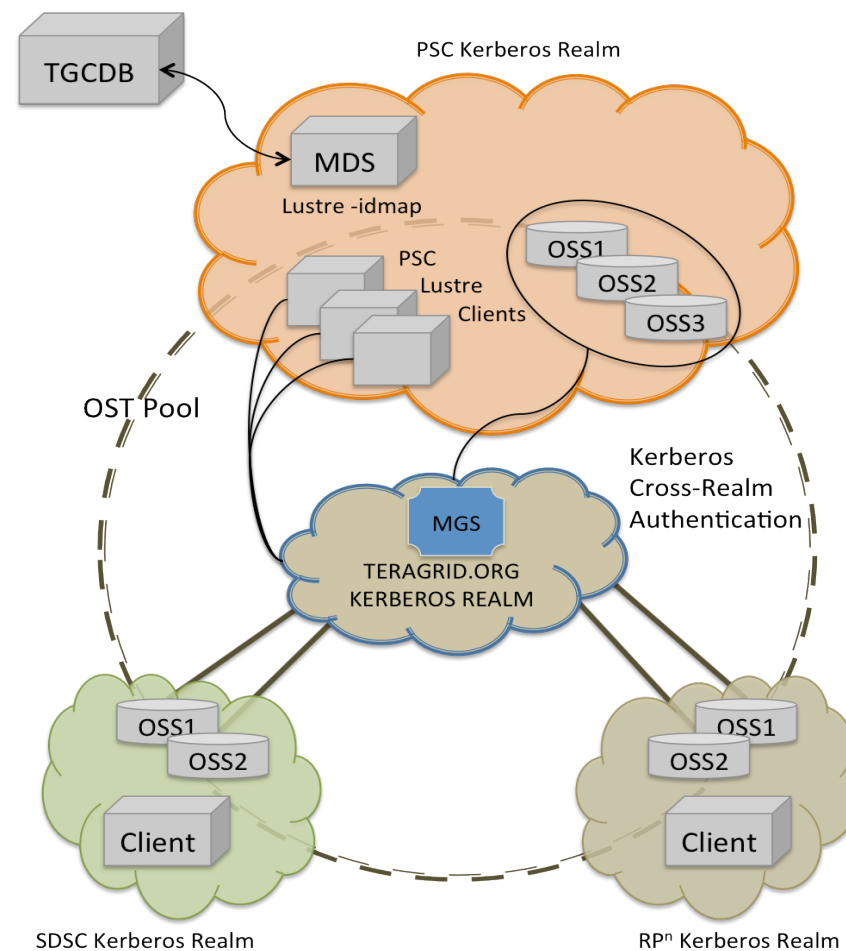
@PSC + OSG (UFlorida)



NSF funding to set up kerberized Lustre 2.0 for OSG
Secure Lustre 2.0 filesystem will be based at University of Florida
Project involves setting up a kerberos infrastructure and
choosing a kerberos realm.

Distributed OST/(optional) OST pool

- decentralize/distribute the secure Lustre WAN storage across several organizations/different sites
- sites write to local OST pools but the filesystem is visible on the WAN
- data striped
 - across PSC dist OSTs/pool
 - across SDSC dist OSTs/pool
 - across both sites OSTs



Distributed OST/OST pool

Best way to arrange TeraGrid user directories in a lustre WAN with distributed OST/OST pool?

1. Sysadmins: OSS-centric directory arrangement
Data is 'striped' to local OSS contributed by each site

```
so /jwan/$SITE/users
```

```
For the 6 sites and user joe (N-> 0-9),
```

```
    /jwan/psc/N/joe  
    /jwan/sdsc  
    /jwan/taccs  
    /jwan/iu  
    /jwan/nics  
    /jwan/ncsa...
```

2. Users: user-centric way of seeing their data in specific machines

```
so /jwan/$USER/$SITE/$MACHINE
```

```
    /jwan/joe/psc  
        /psc/pople  
    /jwan/joe/tacc  
        /tacc/ranger  
        /tacc/lonestar  
    /jwan/joe/nics  
        /nics/kraken...
```

Distributed OST/OST pool

Best way to arrange TeraGrid user directories in a lustre WAN with distributed OST pool?

3. Combining both and adopting something similar to AFS

```
so /jwan/users/$USER/$PSC/$machine symlink to /jwan/$SITE/N/$USER/$MACHINE
```

More specifically,

```
/jwan/users/joe/psc/pople --> /jwan/psc/users/N/joe/pople
```

where

```
/jwan/users/$USER and  
/jwan/users/$USER/$SITE level directories are both read-only and  
highly-available to users
```

Constraints: Part II

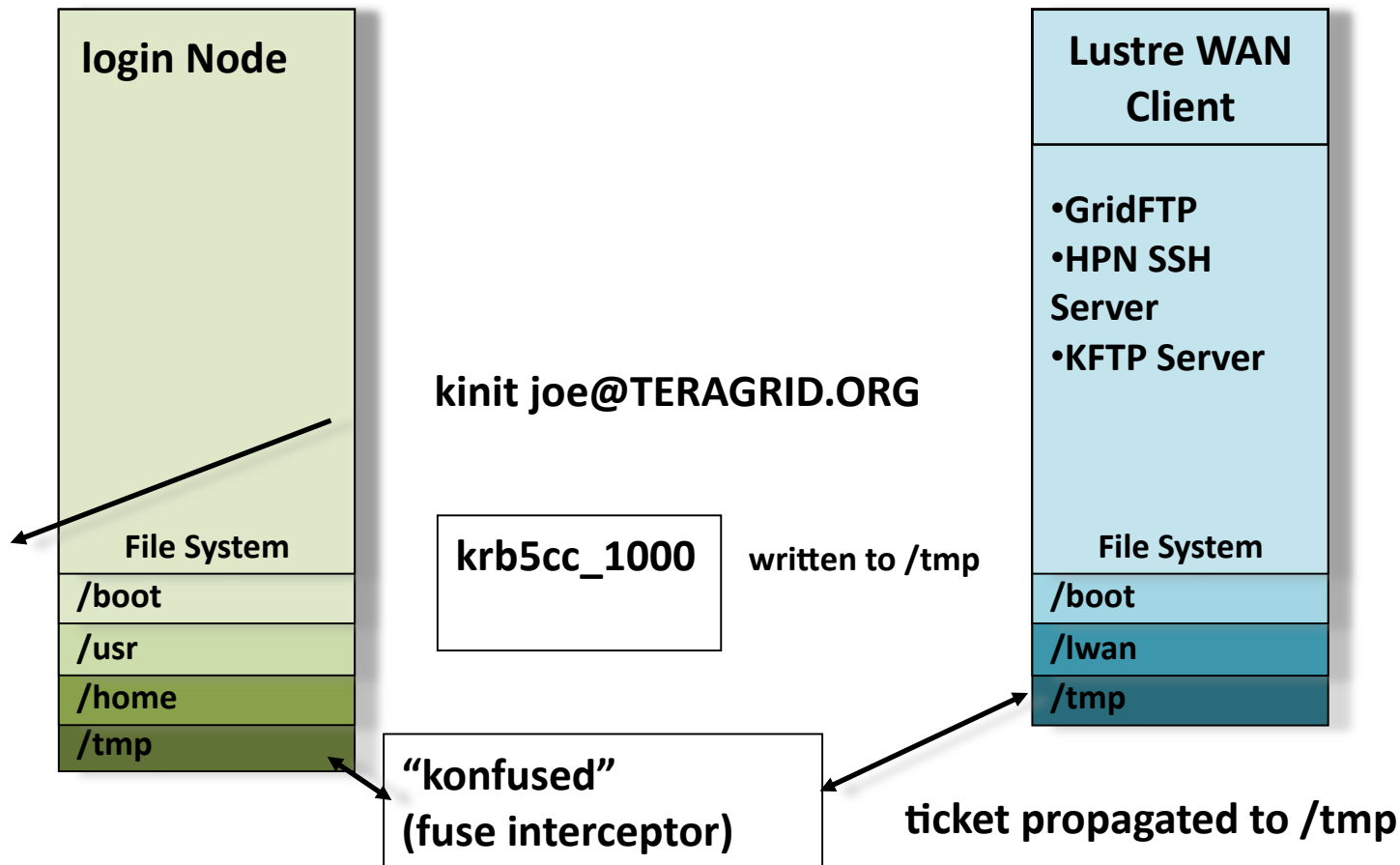
- **Lustre 2.0 – only interoperable with (non-kerberized) Lustre 1.8**
 - **Mount Lustre 1.8.2 on a kerberized Lustre 2.0 client without disabling kerberos on the Lustre 2.0 servers**
- **Native clients, OSTs- must be running lustre 2.0**
 - same release (kernel/lustre rpms- available)**
- **Other sites/partners don't have (will not have) the kerberos infrastructure**

How to make kerberized lustre 2.0 accessible and usable in the current framework?

Kerberized scp/kftp/gridftp: konFUSEd

pople.psc.edu

thunderball.jwan.teragrid.org



Kerberized scp/kftp/gridftp: konFUSEd

```
[joeuser@tg-login3.pople.psc.teragrid.org:~]$ scp file  
thunderball.jwan.teragrid.org:/jwan/users/joeuser/test
```

```
joeuser@thunderball.jwan.teragrid.org's password:  
scp: /jwan/users/joeuser/test: Operation not permitted
```

```
[joeuser@tg-login3.pople.psc.teragrid.org:~]$ module load konfuse
```

```
[joeuser@tg-login3.pople.psc.teragrid.org:~]$ env | grep KRB  
KRB5CCNAME=/konfusefs/krb5cc_20033
```

```
[joeuser@tg-login3.pople.psc.teragrid.org:~]$ kinit joeuser@TERAGRID.ORG  
joeuser@TERAGRID.ORG's Password:
```

```
[joeuser@tg-login3.pople.psc.teragrid.org:~]$ scp file  
thunderball.jwan.teragrid.org:/jwan/users/joeuser/test  
file 0% 0 0.0KB/s 0.0KB/s --:-- ETA file 100% 611KB 611.0KB/s 611.0KB/s 00:00  
Max throughput: 611.0KB/s
```

```
[joeuser@tg-login3.pople.psc.teragrid.org:~]$ exit
```

Available resources @ PSC




- **11,115 Teragrid Accounts auto-synced from TGCDB
(Teragrid Central Database)**
- **Any Teragrid user can login to VM kerberized lustre 2.0 client goldeneye.jwan.teragrid.org and**
 - **access the filesystem**
 - **run tests: Lustre 2.0 quota, ACLs**
 - **request writes to local PSC OSTs (pool), remote SDSC osts (pool), combination of local and remote OSTs (pool)**
- **Perform data transfer on Lustre 2.0 fs**
 - **Gridftp/Gsiscp/Kftp with another VM data target (databox) thunderball.jwan.teragrid.org**

Contribute resources

Active participants:

[Index of ftp://ftp.psc.edu/pub/jwan/Lustre-2.0-alpha/](ftp://ftp.psc.edu/pub/jwan/Lustre-2.0-alpha/)

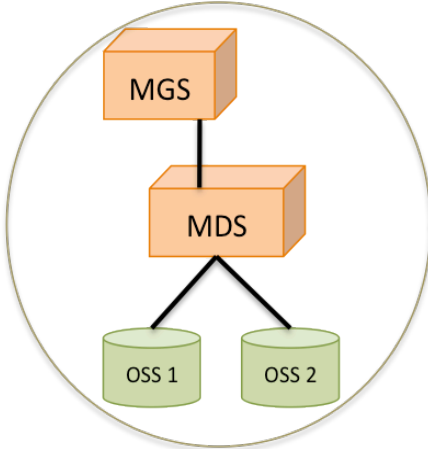
[↑ Up to higher level directory](#)

| Name | Size | Last Modified |
|---|------|------------------------|
|  1.9.210 | | 10/20/2009 04:19:00 AM |
|  1.9.280 | | 11/03/2009 10:07:00 PM |
|  1.9.50 | | 06/04/2009 12:00:00 AM |

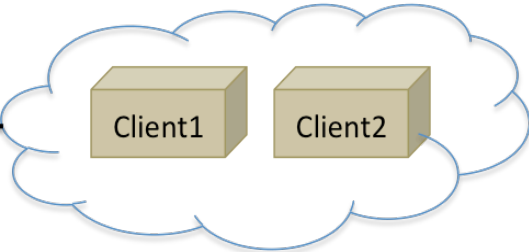
- **Kernel/Lustre rpms**
- **Kerberos infrastructure (with/without)**
 - **Sites** already in the TeraGrid
 - **Sites** not in TeraGrid (**untested**)

Kerberized NFS4

Kerberized Lustre 2.0 Servers



Kerberized Lustre 2.0 Clients



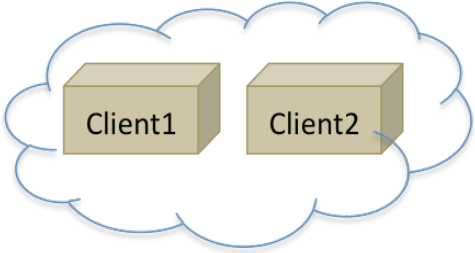
mount /lwan2.0
via Lustre with auth krb5p
lwan.psc.flavor.default = krb5p

Kerberized NFS4 Servers

Exports /lwan2.0
via NFS4 kerberos - enabled



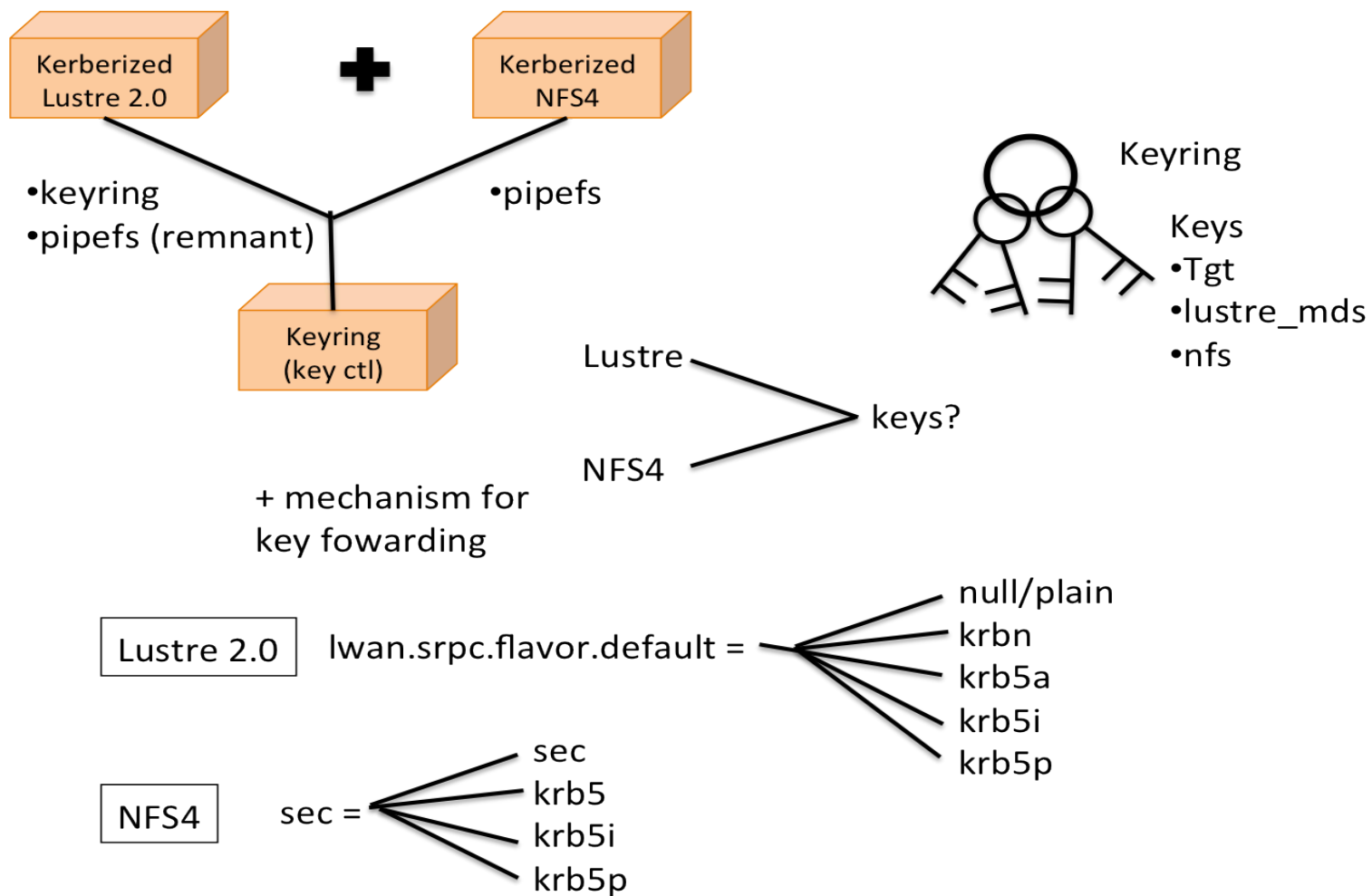
Kerberized NFS4 Clients



mount /lwan2.0-nfs4
via NFS4 (sec = krb5p)

| Lustre2.0 | NFS4 | User Access to /lwan-nfs4 |
|-------------------|------------------|---|
| null | sec, krb5 (i, p) | ✓ |
| krb5 (n, a, i, p) | - | ✗ /lwan-nfs4 remains mounted but inaccessible to users |

Kerberized NFS4



Constraint: Part II

- Lustre kerberos in Single Kerberos Realm
 - **necessary to sync all accounts on MDS and all clients**
 - **user-land: 3-fold check (uid, gid, kerberos principal)**
 - **lustre ID mapping turned off for single kerberos realm**
- Lustre kerberos- Cross-realm
 - **lustre ID mapping enabled**

Summary

- **Workable secure Lustre 2.0 system with constraints**
- **Working on kerberized NFS4 and Lustre 2.0**
- **Lustre kerberos realm: Single->Cross realm**
 - **Both user-land and systems**
 - **Key management**

References

TeraGrid Kerberized Lustre 2.0 Wiki

http://www.teragridforum.org/mediawiki/index.php?title=JWAN:_lustre-wan_advanced_features_testing

The screenshot shows a Mozilla Firefox browser window with the address bar containing the URL: http://www.teragridforum.org/mediawiki/index.php?title=Lustre-wan:_advanced_features_testing. The page title is "JWAN: lustre-wan advanced features testing" and it is noted as being redirected from "Lustre-wan: advanced features testing".

The page features a navigation sidebar on the left with sections for "navigation" (Main Page, Current events, Recent changes, Random page), "search" (with a search box and "Go" and "Search" buttons), and "toolbox" (What links here, Related changes, Upload file, Special pages, Printable version, Permanent link).

The main content area displays a "Contents" table of contents with the following items:

- 1 INTRODUCTION
- 2 BACKGROUND
- 3 GENERAL OBJECTIVES
- 4 SPECIFIC GOALS
- 5 PROJECTS
- 6 Lustre-WAN "2.0" Planning/Coordination
- 7 LINKS ARCHIVE
- 8 COLLABORATORS
- 9 HARDWARE
- 10 DOCUMENTATION
 - 10.1 Howto Outline
- 11 Instructions for lustre servers (OSS) and clients connecting to JWAN (Single Kerberos Realm)
- 12 Specific instructions for OST servers
- 13 Lustre Idmap
- 14 Multiple Kerberos Realms/Cross-Realm Authentication/Lustre ID-MAP
- 15 REFERENCES
- 16 KEY CONFIG FILES
- 17 TEST PLAN
 - 17.1 Phase I: Checking the Network
 - 17.2 Phase II: Defining the Kerberos Realms, Domains Setup for the WAN lustre filesystem
 - 17.2.1 Background
 - 17.2.2 Summary Test Table
 - 17.3 Phase III. Setting up Cross-Realm Authentication
 - 17.3.1 A. Checklist
 - 17.3.2 B. Debugging your kerberos setup
 - 17.4 Phase IV: Implementing OST Pools
 - 17.5 Phase V: Measuring the Kerberos Security Overhead
 - 17.6 Phase VI: Benchmark Proper
- 18 BENCHMARK ROLLS
- 19 BUGS & SECURITY
 - 19.1 Phase I:

References

PSC Kerblustre Wiki

<https://wiki.psc.edu/twiki/view/KerbLustre/WebHome>

PITTSBURGH SUPERCOMPUTING CENTER You are here: [TWiki](#) > [KerbLustre Web](#) > [WebHome](#)
r23 - 12 Apr 2010 - 14:51:56 - [JosephinAtPscEdu](#)

Web Home > [Webs](#) [Utilities](#) [Options](#)

Welcome to the KerbLustre web

Introduction

KerbLustre is a resource guide to the kerberos-enabled Lustre 2.0 filesystem we have established over the WAN. Lustre Kerberos authentication is enabled by default.

- Lustre version: 2.0.5 Alpha
- Lustre release: 1.9.280
- Kernel: 2.6.18_128.7.1
- Kerberos flavor: krb5p (default)
- Network eth: 112, 58 Subnets
- Infiniband: 1.5 (being worked on)
- Storage: 1.4TB (Single-realm) and 1.3TB (Cross-realm)

Systems in the TERAGRID.ORG Kerberos realm

PSC: Physical machines

| FS mounted | Hostname | Server | OS | Production Status | Lustre Purpose | Disk Space | Kerberos Realm | Network | Notes | Life Cycle DB |
|------------|------------------------------|------------------|----------------------------|-------------------|----------------|------------|----------------|---------|----------------------|-----------------------------|
| /jwan | mgs.jwan.teragrid.org | Physical Machine | CentOS ? 5 | Production | MGS | | TERAGRID.ORG | 112 | rcons4 | system info |
| | mds00w.psc.jwan.teragrid.org | Physical Machine | CentOS ? 5 | Production | MDS | | TERAGRID.ORG | 112 | rcons6 | system info |
| | oss01w.psc.jwan.teragrid.org | Physical Machine | CentOS ? 5 | Production | OSS | 1.4 TB | TERAGRID.ORG | 112 | rcons6 | system info |
| /jwan1 | mgs1.jwan.teragrid.org | Physical Machine | CentOS ? 5 | Production | MGS | | TERAGRID.ORG | 112 | | |
| | mds01w.psc.jwan.teragrid.org | Physical Machine | CentOS ? 5 | Production | MDS | | TERAGRID.ORG | 112 | rcons5 ; power reset | system info |
| | mds02w.psc.jwan.teragrid.org | Physical Machine | CentOS ? 5 | Production | MDS | | TERAGRID.ORG | 112 | rcons5 ; power reset | |
| | oss00w.psc.jwan.teragrid.org | Physical Machine | CentOS ? 5 | Production | OSS | 1.4 TB | TERAGRID.ORG | 112 | rcons6 | system info |
| | attractor.psc.edu | Physical Machine | CentOS ? 5 | Development | Other | | PSC.EDU | 58 | | system info |

[Add or Change a Machine](#)

PSC: Virtual machines:

| VM Name | Hostname | Server | OS | Production Status | Lustre Purpose | Disk Space | Kerberos Realm | Network | Notes | Life Cycle DB |
|---------|----------|--------|----|-------------------|----------------|------------|----------------|---------|-------|---------------|
|---------|----------|--------|----|-------------------|----------------|------------|----------------|---------|-------|---------------|

<https://wiki.psc.edu/twiki/edit/KerbLustre/WebHome?t=1271098317> wiki.psc.edu

Applications Places System 2:54 PM