Deploying the TeraGrid PKI

Grid Forum Korea Winter Workshop
December 1, 2003

Jim Basney
Senior Research Scientist
National Center for Supercomputing Applications
University of Illinois
jbasney@ncsa.uiuc.edu

NCSA

Grid-building Challenges

- Many challenges in deploying Grids
  - – software compatibility
  - – resource discovery (information services)
  - – resource allocation
  - – accounting (charging for resource usage)
  - – performance optimization
  - – monitoring / support / helpdesk
  - – …

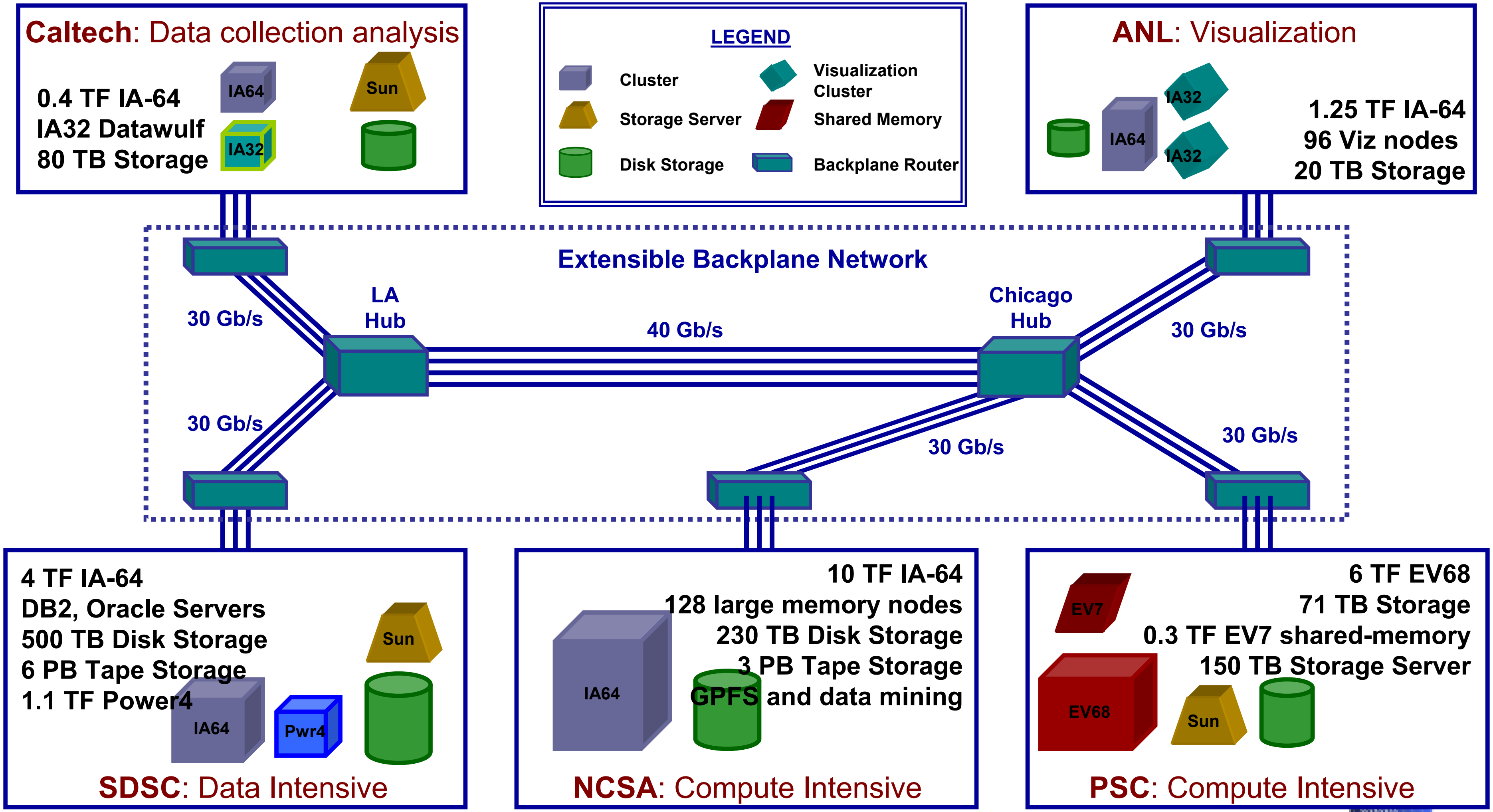
Managing Trust for Grid Single Sign-on

- A major Grid deployment challenge
- What CAs are trusted?
  - Can a CA gain universal acceptance for single sign-on?
  - What CA practices are acceptable?
  - Use hierarchical CAs or cross-certification?
- How do users obtain and manage credentials?
  - user enrollment, certificate renewal, private key security, …
- How are users authorized to use resources?
  - How are ACLs and authorization services managed?
- Consider the TeraGrid as a Case Study

Outline

- TeraGrid Overview
- Globus Security Infrastructure
  - Authentication and Authorization
  - Proxy Credentials
- TeraGrid Online CAs
- TeraGrid Single Sign-on
- Grid-Mapfile Management
- Credential Management

TeraGrid



Caltech: Data collection analysis

0.4 TF IA-64
IA32 Datawulf
80 TB Storage

LEGEND

- Cluster
- Visualization Cluster
- Storage Server
- Shared Memory
- Disk Storage
- Backplane Router

ANL: Visualization

1.25 TF IA-64
96 Viz nodes
20 TB Storage

Extensible Backplane Network

4 TF IA-64
DB2, Oracle Servers
500 TB Disk Storage
6 PB Tape Storage
1.1 TF Power4

SDSC: Data Intensive

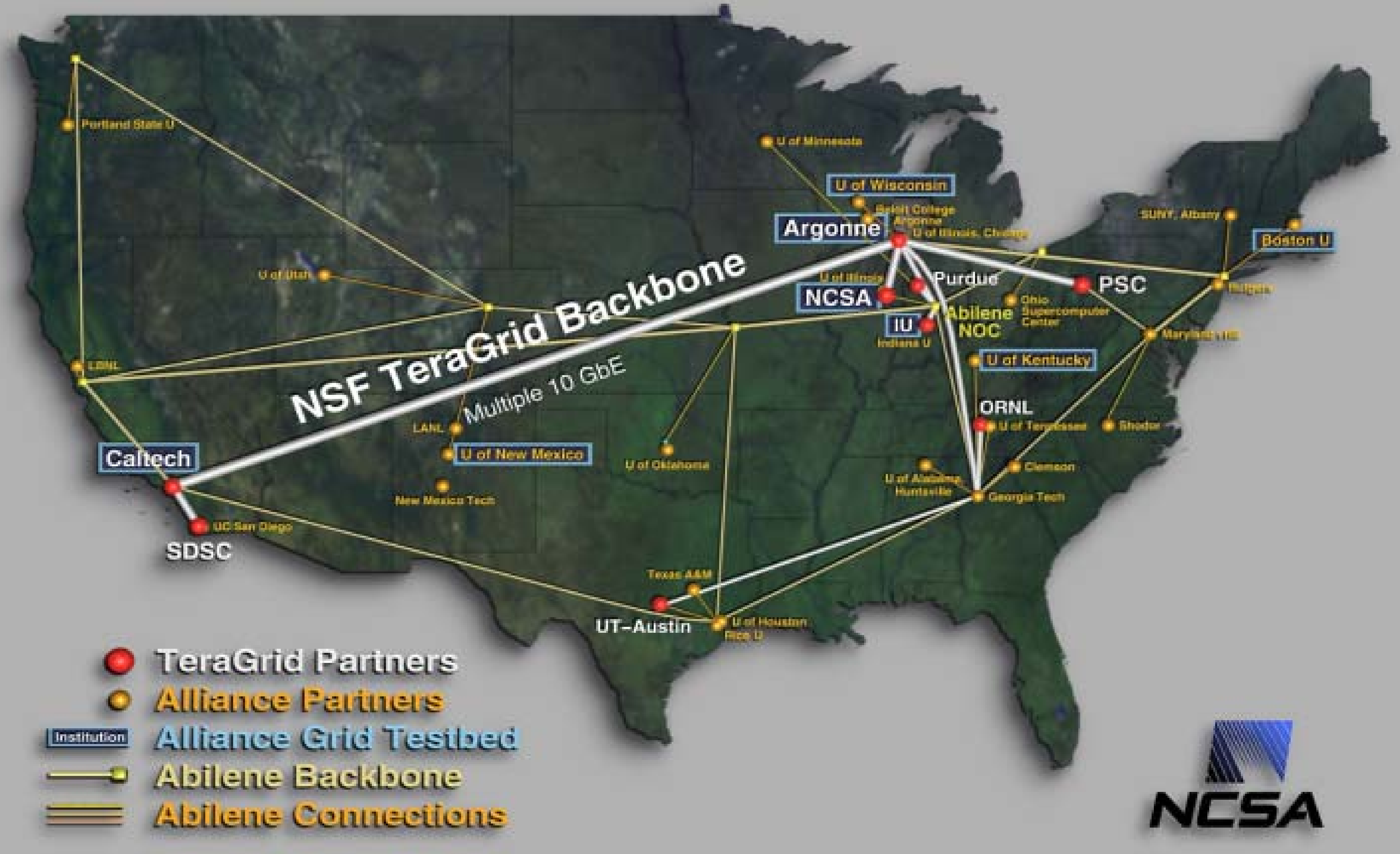
10 TF IA-64
128 large memory nodes
230 TB Disk Storage
3 PB Tape Storage
GPFS and data mining

NCSA: Compute Intensive

6 TF EV68
71 TB Storage
0.3 TF EV7 shared-memory
150 TB Storage Server

PSC: Compute Intensive

Additional TeraGrid Sites

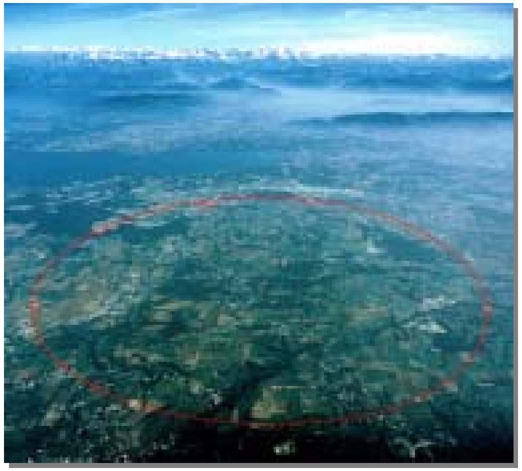


Building Something New

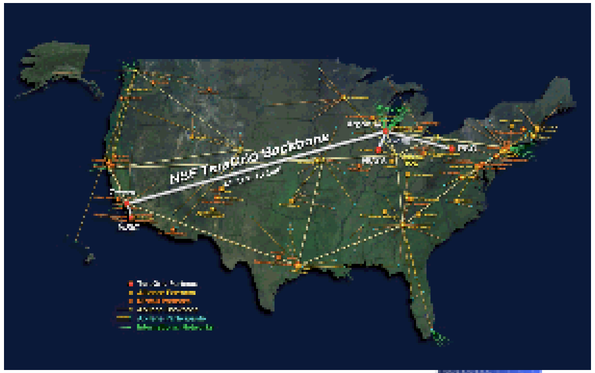
One Organization (merge institutions)	**The TeraGrid** **(A Grid hosting environment)**	Very Loose Collaboration (current situation)
- One sysadmin team - One management team - Distributed machine room, centralized control - e.g. Google data centers	- Single development environment - Single software stack to learn - Develop here, run there - Run here, store there	- Different MPIs - Hit-and-miss grid software: - Globus version? - Condor-G? - MPICH-G2? - Unique development environment
Not a Grid	**Applications are developed for the Grid because the barriers are low and the return large**	**Not a Grid, but with significant user investment, Grid applications can be developed**

TeraGrid and CMS

- Data and software testing challenge
  - test and validate analysis software
    - 100,000,000 events
- Testing approach
  - particle-detector interaction simulator (CMSIM)
    - energy deposition in the detector
  - ORCA (Object Reconstruction for CMS Analysis)
    - reconstruct QCD background sample
  - tracks and reconstructed particles, ready for analysis
- *Computing, storage and networking*
  - *1.1M SUs on the TeraGrid now*
    - *400 processors through April 2005*
  - 1M SUs on NCSA Platinum Pentium III cluster
  - 1.5M SUs on NCSA Tungsten Xeon cluster
  - 1 TB for production TeraGrid simulations
    - 400 GB for data collection on IA-32 cluster

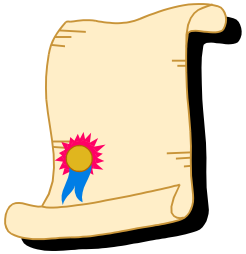


http://cmsinfo.cern.ch/



Globus Security Infrastructure

- Credentials
  - asymmetric public/private key pair
  - X.509 certificate, signed by Certificate Authority, binds distinguished name to key pair

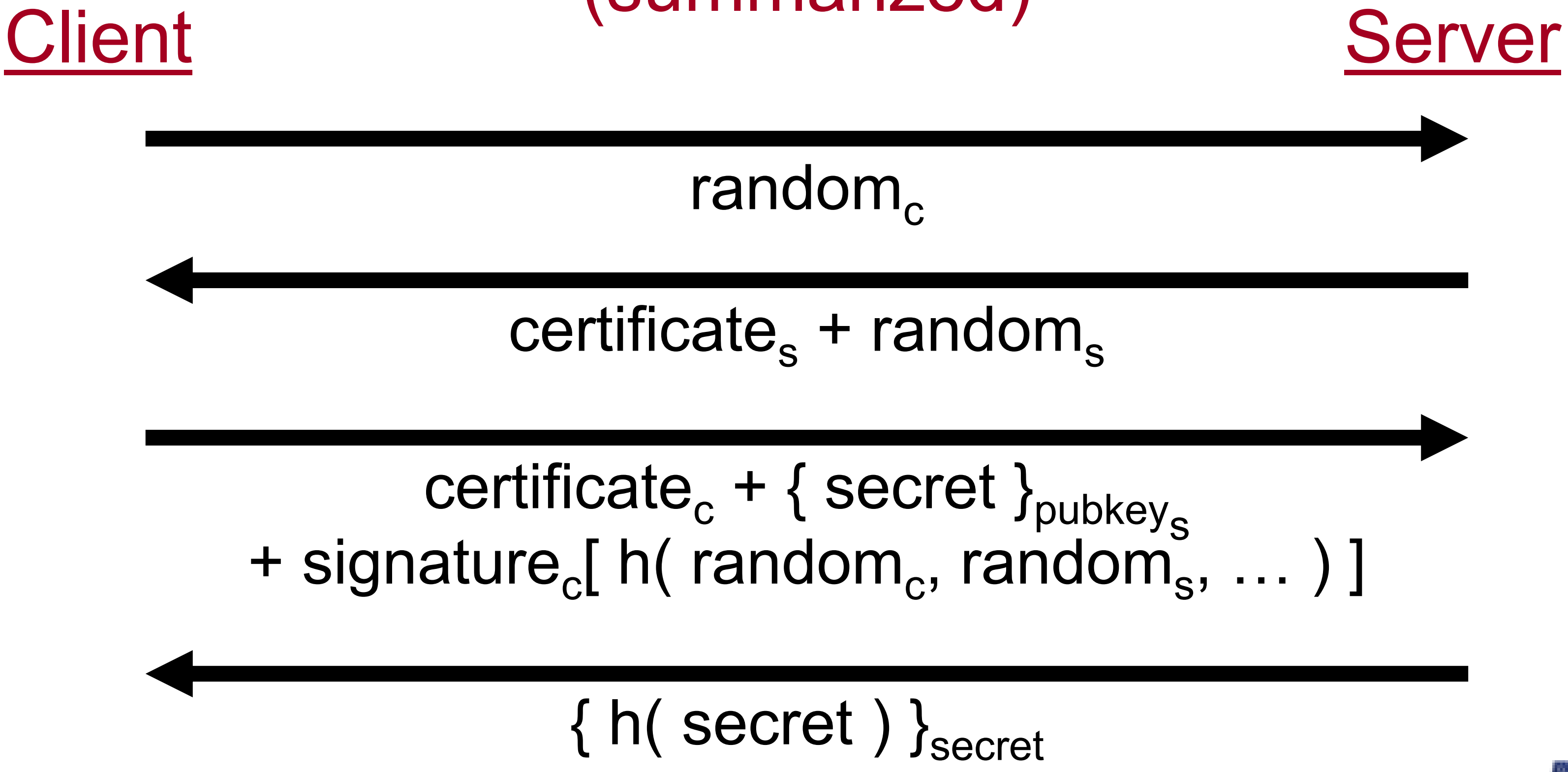


- Authentication (Who are you?)
  - proof of possession of private key
  - verify CA signature on X.509 certificate

- Authorization (What can you do?)
  - based on distinguished name in certificate
  - typically mapped to local account

GSI Mutual Authentication

Standard SSL/TLS Protocol (summarized)



Client

Server

random_c

$\text{certificate}_s + \text{random}_s$

$\text{certificate}_c + \{ \text{ secret } \}_{\text{pubkey}_s}$
$+ \text{ signature}_c[\text{ h(random}_c, \text{ random}_s, \ldots)]$

$\{ \text{ h(secret) } \}_{\text{secret}}$

GSI Mutual Authorization

- What is the client authorized to do on the server?
  - typically set by grid-mapfile
- Is the server trusted by the client?
  - i.e., is the server authorized by the client?
  - typically based on authenticated server identity matching the user’s request
- Client must have the ability to verify server certificates
  - must trust certificate of the CA that signed the server’s certificate
  - must have correct system clock

How to Authorize Clients?

- Access Control Lists
  - ex. Globus grid-mapfile
  - answer “Who can access this resource?”
  - need to maintain many distributed ACLs
- Capabilities
  - ex. SAML, X.509 PMI, VOMS, Akenti, CAS
  - answer “What can this person do?”
  - don’t need to distribute ACL updates
  - capability issuer maintains authorization database
- GGF OGSA Authorization WG

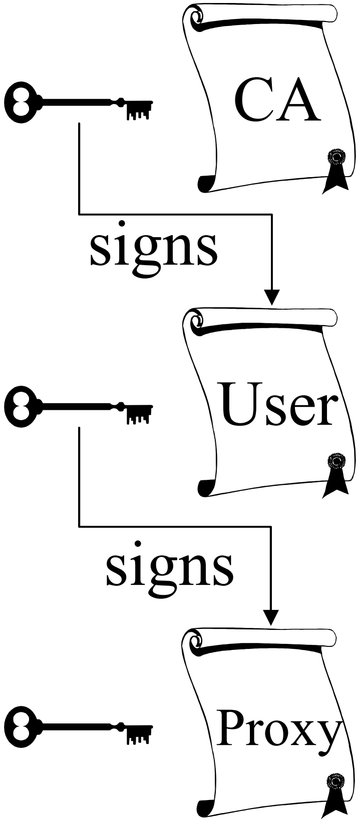
What to Authorize?

	Keys	**Names**
Examples:	SSH, PGP, SPKI	X.509 PKI, GSI
Trusted Third Party?	None	CA signs certificates
Cost of re-keying?	Update ACLs with new public key	Obtain new certificate

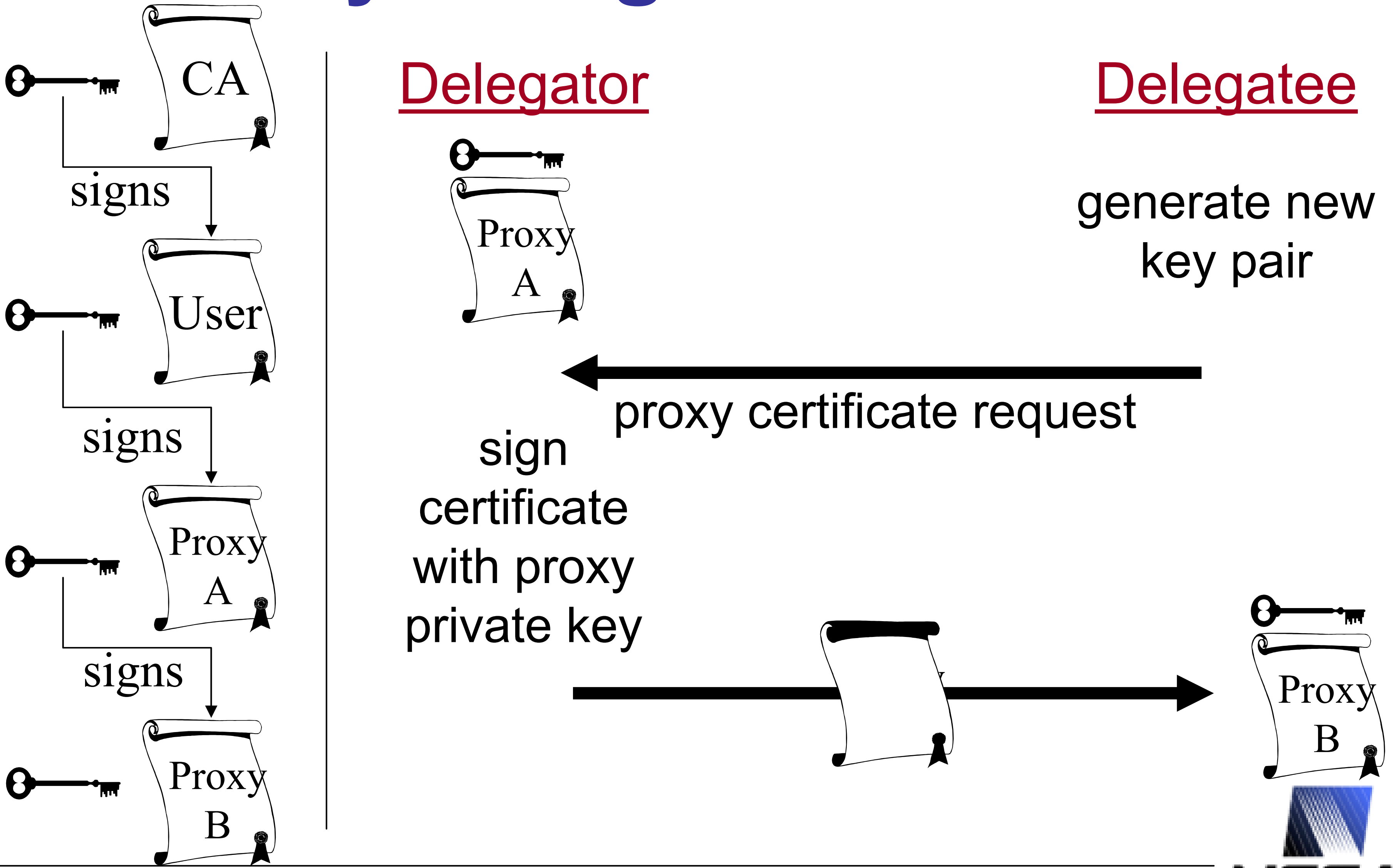
- Names can be convenient to work with but…
- Common names are not unique identifiers

Globus Proxy Credentials

- New certificate and key pair
- Proxy certificate signed by user's long-term private key
  - enter passphrase to decrypt private key
- Certificate has short lifetime
- Proxy private key remains unencrypted
- Authenticate with proxy credentials for the remainder of the session



Proxy Delegation Protocol



CA

signs

User

signs

Proxy A

signs

Proxy B

Delegator

Delegatee

Proxy A

generate new key pair

proxy certificate request

sign certificate with proxy private key

Proxy B

TeraGrid PKI

- A single TeraGrid Certificate Authority is not feasible
  - many sites already have a CA
  - distributed model is preferable for Grids
- TeraGrid PMA evaluates CA trust
  - for interoperability, all TeraGrid sites should accept TeraGrid approved CAs
  - TeraGrid PMA distributes trusted CA certificates to users and administrators

TeraGrid Online CAs

- An **Online CA** allows users to authenticate and obtain PKI credentials immediately
  - without requiring the user to visit a registration authority, fax a copy of an institutional ID, etc.
  - without requiring the CA operator to manually approve each request
  - leveraging the site’s existing relationship with its users
- Online CAs can return long-term or short-term credentials:
  - users contact the online CA infrequently to obtain / renew long-term (1+ year) certificates, or
  - users contact the online CA daily to obtain short-term (12 hour) credentials
  - TeraGrid includes examples of both types of online CAs

CACL

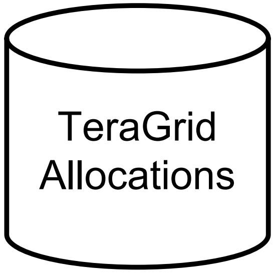
- NCSA and SDSC have online CAs that return long-term credentials
  - OpenSSL-based CACL online CA software developed at SDSC
  - at NCSA, online CA recently replaced offline CA
- Users login to NCSA or SDSC cluster and run a command to obtain 2-4 year credentials
  - credentials stored in ~/.globus as usual
  - requires users to manage their long-term key and certificate files
- For more information:
  - http://www.npaci.edu/CA/
  - http://grid.ncsa.uiuc.edu/ca/

KCA

- PSC runs a Kerberized online CA (KCA)
- Users obtain short-term (12 hour) Kerberos tickets at login
- KCA command allows users to authenticate with Kerberos ticket to obtain Globus credentials
  - KCA credentials have short lifetime equal to Kerberos ticket lifetime
  - stored unencrypted in /tmp to be used like Globus proxy credentials
- No need to issue CRLs as there are no long-term certificates to revoke
- For more information:
  - http://www.citi.umich.edu/projects/kerb_pki/
  - http://www.psc.edu/certificate-authority/

TeraGrid Account Creation

- US National Science Foundation committees evaluate research proposals and allocate TeraGrid resources to scientists
- Allocation info is entered into TeraGrid Accounting Database
- Account creation requests sent to sites
  - via TeraGrid Account Transaction System
- Scientist receives account information in the mail
  - includes username(s) and initial password(s) for the site(s)

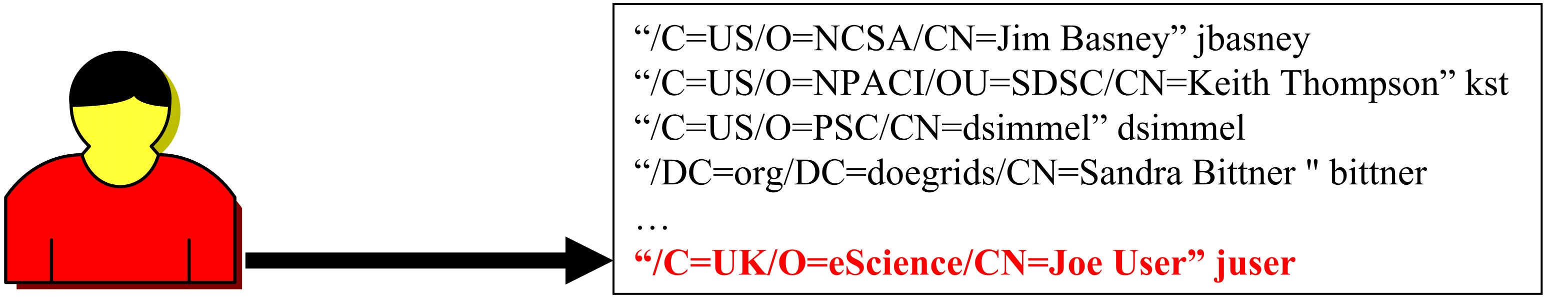


TeraGrid Grid Single Sign-on

- Users can access all TeraGrid resources using their Grid proxy credentials
  - using GSISSH, GRAM, and GridFTP
  - no need to remember different usernames and passwords
- For users with no PKI certificate
  - request a certificate from a TeraGrid CA
  - TeraGrid Account Transaction System adds user’s distinguished name to grid-mapfiles (planned)
- For users that already have a PKI certificate
  - issuing CA must be trusted by TeraGrid sites
  - gx-map command allows users to add additional distinguished names to grid-mapfiles

GX-Map

- A Globus grid-mapfile management tool
- Allows users to add distinguished names to the grid-mapfile
  - mapped only to that user’s account
- Similar to adding SSH Authorized Keys
- For more information:
  - http://www.sdsc.edu/~kst/gx-map

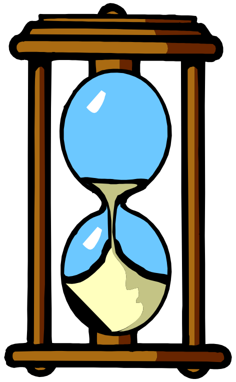


```
“/C=US/O=NCSA/CN=Jim Basney” jbasney
“/C=US/O=NPACI/OU=SDSC/CN=Keith Thompson” kst
“/C=US/O=PSC/CN=dsimmel” dsimmel
“/DC=org/DC=doegrids/CN=Sandra Bittner " bittner
…
“/C=UK/O=eScience/CN=Joe User” juser
```

Credential Management

- TeraGrid users can store their credentials in an online MyProxy repository
  - credentials encrypted with the user's passphrase
  - users can retrieve delegated proxy credentials from the online repository when/where needed
- MyProxy provides credential mobility
  - users need not manually copy certificate and key files between machines
  - long-term keys protected on the MyProxy server
- For more information:
  - http://myproxy.ncsa.uiuc.edu/

Credential Renewal



- Unsolved problem for TeraGrid
- Long-lived tasks or services need credentials
  - task lifetime is difficult to predict
- Don’t want to delegate long-lived credentials
  - fear of compromise
- Instead, renew credentials as needed during the task’s lifetime
  - renewal service provides a single point of monitoring and control
  - renewal policy can be modified at any time
  - for example, disable renewals if compromise is detected or suspected
- Possible solutions using MyProxy
  - EDG Proxy Renewal Service
  - Condor-G with GRAM proxy refresh

Managing Multiple Credentials

- Will a single identity credential per user suffice?
  - Difficult to achieve trust in a single CA across many organizations
  - Advanced services require authorization credentials
- Pieces of a solution
  - Credential negotiation protocols (WS-SecurityPolicy, …)
  - Online credential services
- Want to retain single sign-on and ease-of-use

Summary

- TeraGrid has deployed a PKI for single sign-on via the Globus Security Infrastructure
  - Online CAs (CACL, KCA)
  - user control of grid-mapfile authorization (gx-map)
  - online credential repository (MyProxy)
- Ongoing work
  - credential renewal
  - managing multiple credentials

Thank you! Any questions?

Jim Basney <jbasney@ncsa.uiuc.edu>