Security on the U. S. Fusion Grid

Presented by Tom Fredian
on behalf of Justin Burruss

5th IAEA-TM on Control, Data Acquisition, and Remote Participation for Fusion Research

Budapest, Hungary
July 12-15, 2005

burruss@fusion.gat.com
http://web.gat.com/~burruss/

FusionGRID
www.fusiongrid.org

Acknowledgements

- **U. S. Department of Energy**
  - OFES & OASCR (SciDAC)

- **DIII-D National Fusion Facility**
  - Operated by General Atomics

- **FusionGrid collaborators**
  - MIT, PPPL, LBL, ANL, Utah CS, Princeton CS

Outline

- **Background**
  - What is FusionGrid?
  - Security goals and challenges
- **Solutions**
  - Authentication
  - Authorization
  - Secure data storage & transfer
- **Applicability to future computing infrastructures**
- **Conclusion**
  - Lessons Learned
  - Next Steps

Presentation Key Points

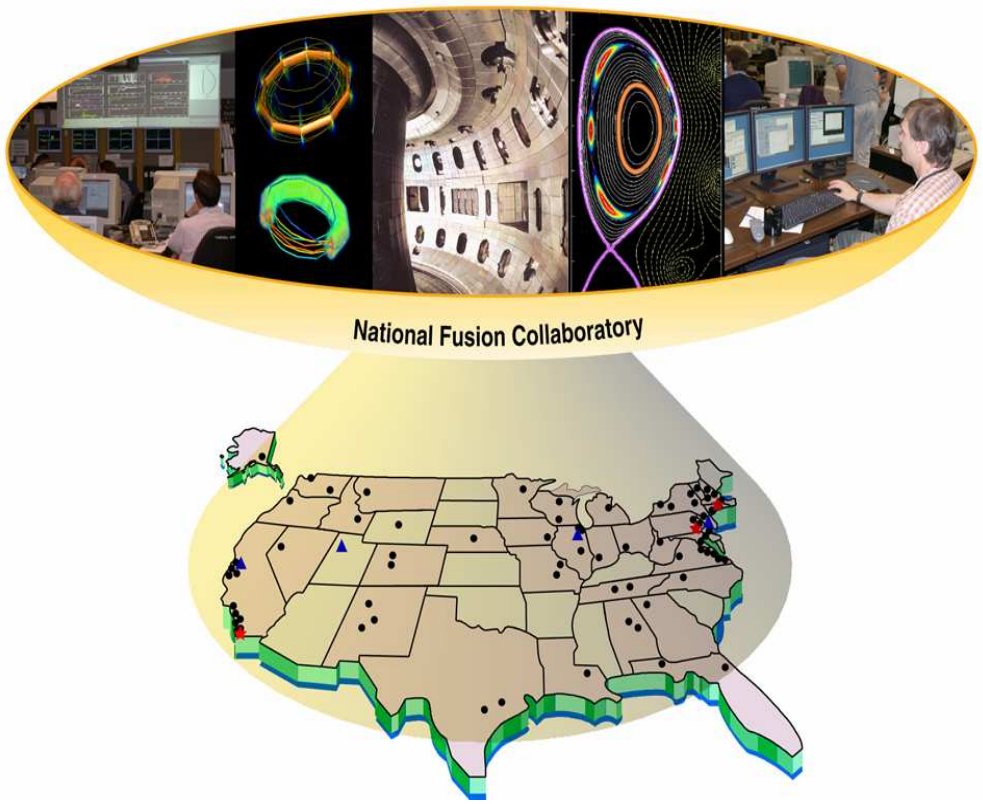
- **Security in a geographically distributed and administratively divided computing environment such as the U. S. Fusion Grid (FusionGrid) is inherently difficult**
  - Authentication across user namespaces
  - Authorization across administrative domains
  - Secure data storage & transfer

- **FusionGrid solved these problems using X.509 credentials, a credential management system, a new authorization manager, and a secure version of MDSplus**

- **These solutions may be extended to future computing infrastructures such as ITER**

FusionGrid created for better use of resources

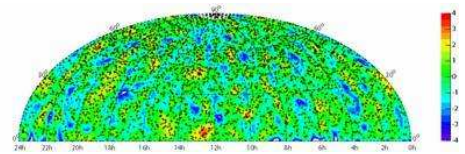
- **U. S. Fusion Grid (FusionGrid) aims to make more efficient use of computing resources**
  - Access is stressed rather than portability
  - Not CPU cycle scavenging or "distributed" supercomputing
- **Share resources between sites**
  - Reduce duplication of effort
- **Develop a common tool set**



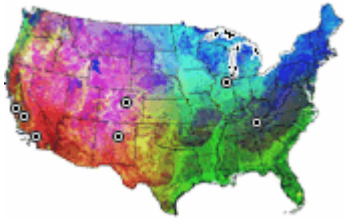
FusionGRID
www.fusiongrid.org

There are many other grids worldwide

- **Open Research**
  - TeraGrid provides teraflops of computation, petabytes of storage, and gigabits of bandwidth to scientists



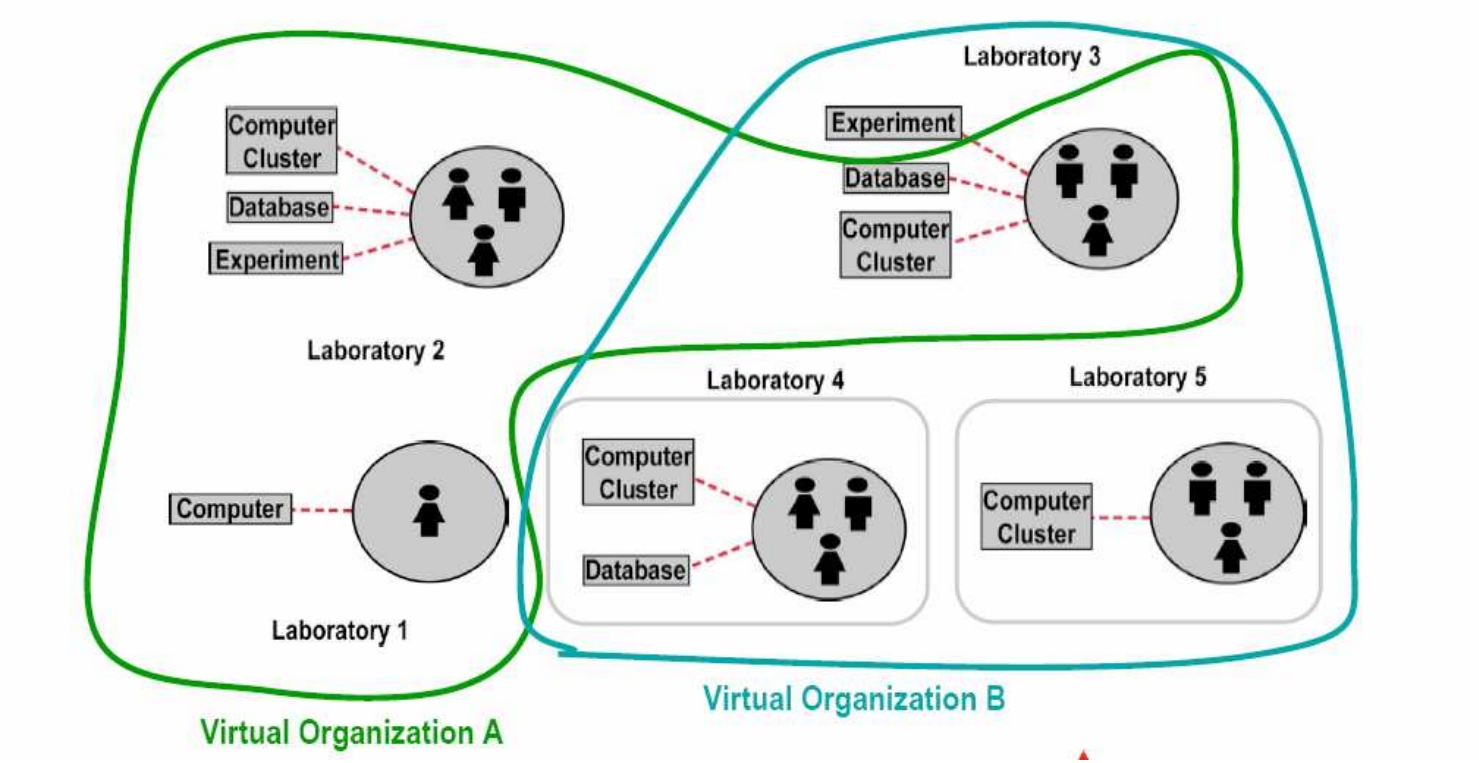
- **Climate Research**
  - Earth System Grid (ESG) used to give climate research scientists more computing power



- **High Energy Physics**
  - LHC Computing Grid (LHG) being developed to process large amounts of data for the LHC being built near CERN

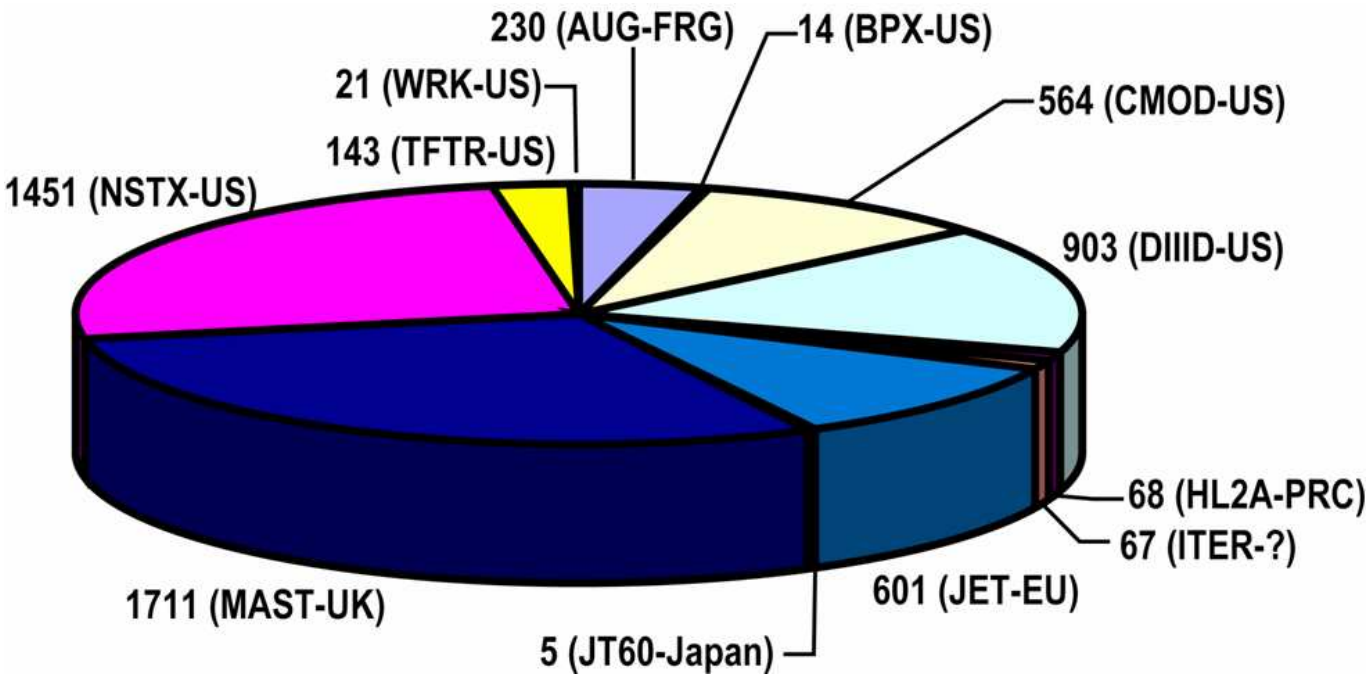
Resources pooled through grid computing

- Grid software adds a layer of abstraction
- Pool resources of multiple organizations into “virtual organizations”
- Codes and other resources made available as “services”



Example FusionGrid service: TRANSP transport code

- TRANSP is available a a FusionGrid service
- Maintained at PPPL
- Used by scientists both inside and outside the U. S.
- Many different tokamaks

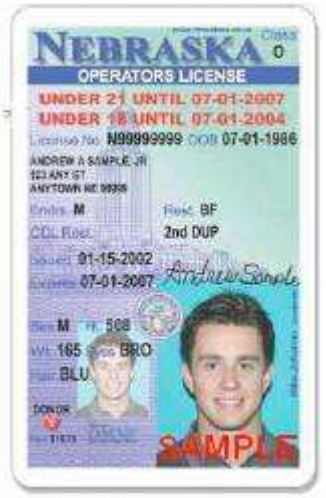


Grid computing brings new security challenges

- **In a virtual organization, how do you uniquely identify a single user across the various administrative domains?**
  - Username at site X not necessarily the same as at site Y
- **How do the various resource stakeholders control access to their resources?**
  - At the same time, how do users request access?
- **Data is no longer local-only access, but must be stored and transferred securely between sites**
- **FusionGrid developers addressed these challenges of authentication, authorization, and secure data storage & transfer**

FusionGrid authentication done with certificates

- **X.509 credentials used to uniquely identify each FusionGrid user**
  - Globally unique username assigned
- **Analogous to a drivers license**
- **Answer the question "who are you?"**
- **Implemented as text files**
- **Original FusionGrid solution required self-management of these text files**
  - Too confusing for most users



FusionGRID
www.fusiongrid.org

New FusionGrid credential management system makes authentication easier

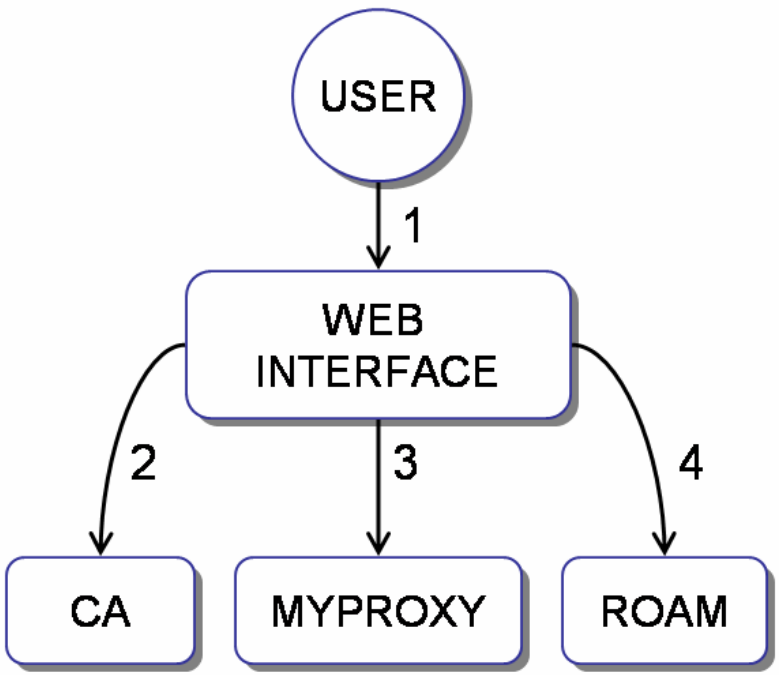
- **Credentials are now stored in a MyProxy credential server**
- **When user needs to "sign on" to FusionGrid, they retrieve a delegated copy of their credentials using their FusionGrid username and password**
- **Thus, signing on to the grid is a simple username/password procedure**
  - Well understood by all users
- **Easier for users**
- **Arguably more secure since credentials are stored on a secured server and not put in the hands of users**

User requests credentials through a web page

- FusionGrid credential manager has a web front-end

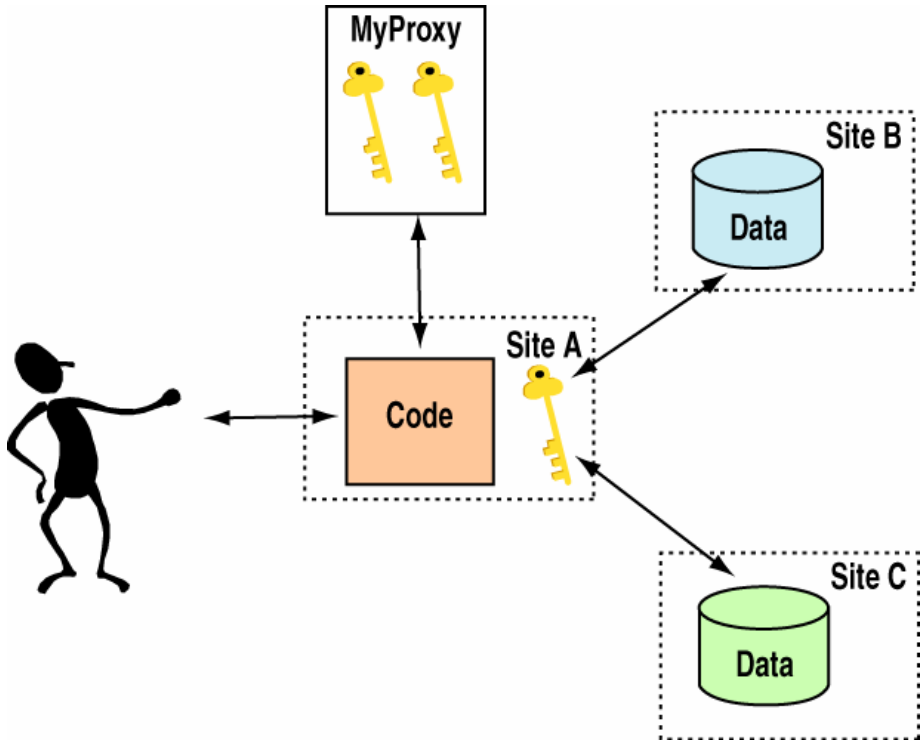
1. User requests new credential through web interface
2. New credential generated with Certificate Authority
3. Long-term delegated copy stored in MyProxy server
4. Authorization database updated

- Details are hidden from user



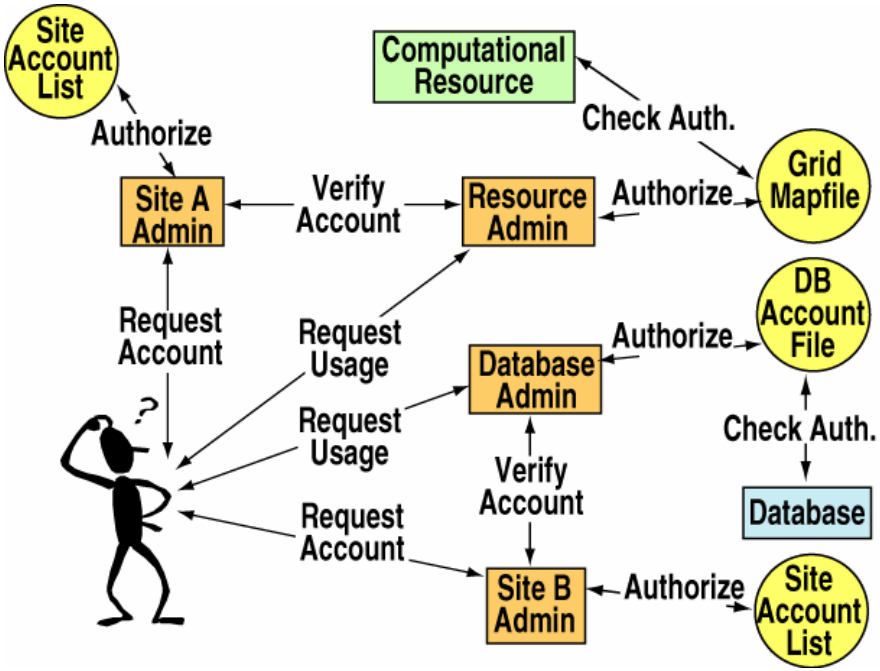
Users "sign on" to FusionGrid by retrieving delegated proxy credentials from MyProxy

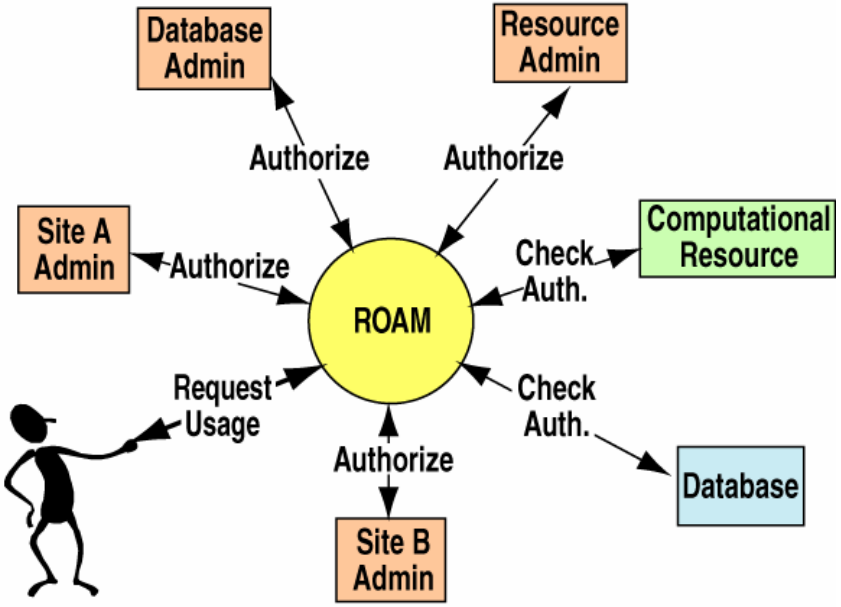
- **Before using FusionGrid resources, a scientist retrieves their credentials**
- **The fglogin script is used**
  - Prompts for password
- **Delegated credentials retrieved**
  - Valid for a limited time
- **Delegated credentials can then be used by codes to act on behalf of the user**



ROAM used to solve authorization problem

- Resource Oriented Authorization Manager (ROAM) provides a central location for authorization information
- Administrators & other stakeholders can control access
- Users can request access



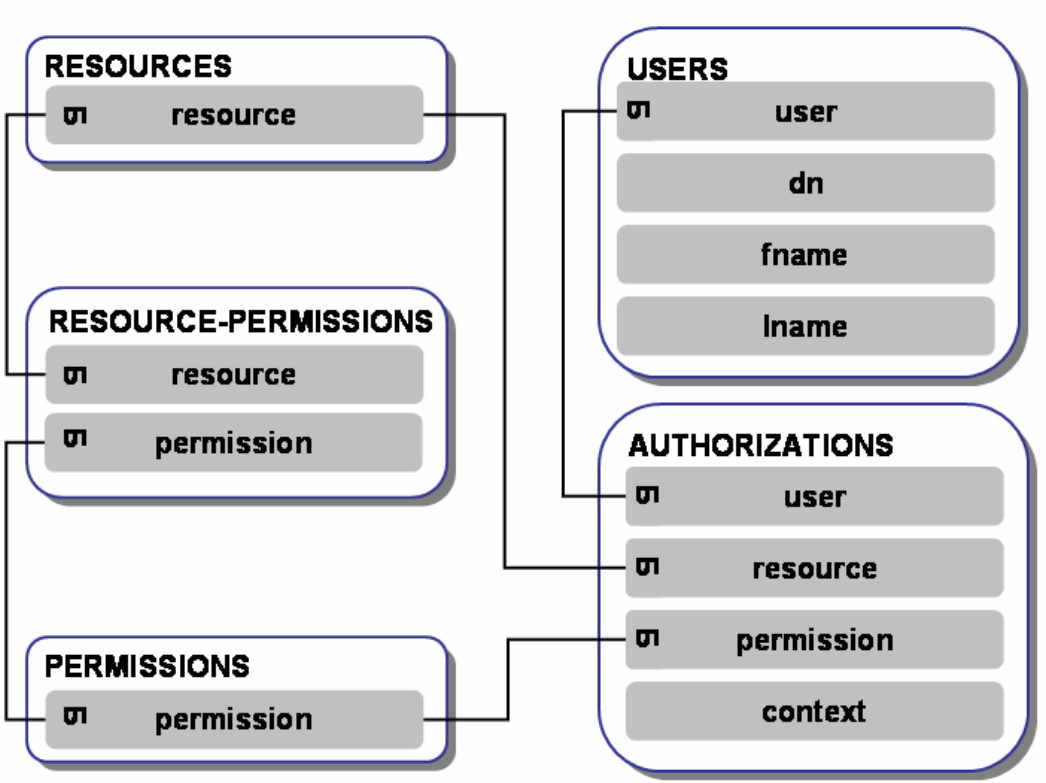


ROAM data model is fundamental to a coherent picture of authorization in the grid

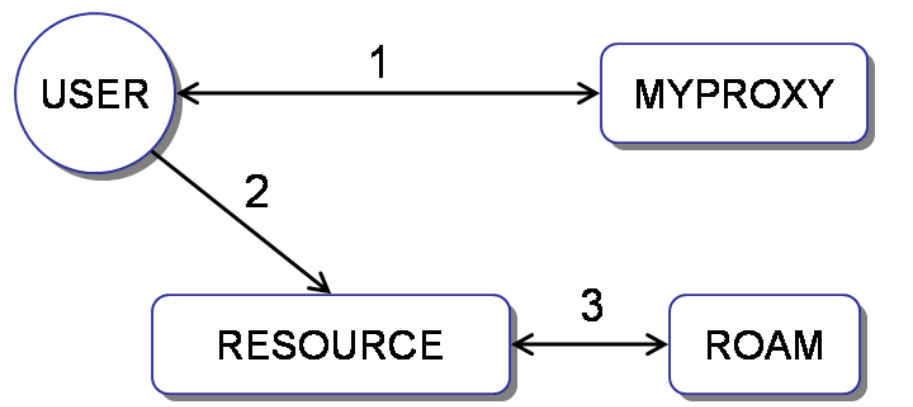
- Oriented around resources: codes, databases, entire sites

- If you have to sign a paper to get permission to use something, then it's probably a resource

- Resources have associated permissions
  - e.g. "execute" for a code, "access" for a site



ROAM avoids "push" model of authorization



- Other authorization schemes use a "push" model of authorization where users first obtain special authorization documents, then push those to the relevant party
- ROAM avoids these extra steps
- Users (1) retrieve credentials as normal and (2) use the resource as normal; the resource (3) checks authorization behind the scenes

People interact with ROAM through a web page

- An interactive web page allows users and resource owners to request and grant permissions
- Less error-prone than editing text files
- No more grid-mapfiles!
- Can take the place of mdsip.hosts for MDSplus

Welcome Justin Burruss to the FusionGrid Resource Authorization System

You are currently authorized for the following resources:

Resource	Permission	Description	Action	Logs
CMOD Data	Write	Access to CMOD experiment data	Submit Inquiry	Show log
DIIID-MDSplus	Read	MDSplus at GA	Submit Inquiry	Show log
DIIID-MDSplus	Admin	MDSplus at GA	Admin Resource	Show log
GA	Access	Site access to General Atomics	Submit Inquiry	Show log
GA	Admin	Site access to General Atomics	Admin Resource	Show log
GATO	Execute	Grid-enabled GATO	Submit Inquiry	Show log
GATO	Admin	Grid-enabled GATO	Admin Resource	Show log

Secure MDSplus used for data storage and transfer

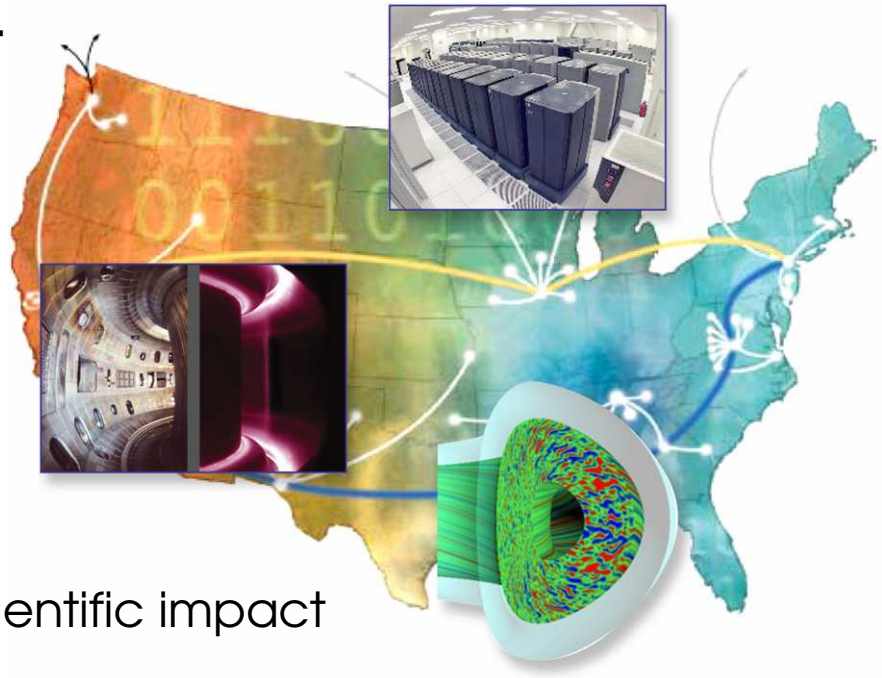
- **MDSplus was updated to work with X.509 credentials and ROAM**
- **Data made available securely**
- **Authentication is no longer host-based**
- **Added benefit: it is easier to work from home or on travel because host-based authentication runs into trouble with NAT and private IP addresses**
  - Get data from your Mac laptop on Wi-Fi at a coffee shop

Secure MDSplus available on Linux and Mac

- **A subset of the Globus Security Infrastructure (GSI) was ported to Mac OS X so that secure MDSplus would work on that platform**
  - Popular in fusion research
- **A Windows port is needed, but this will be a lot of work**
- **Was developed on Linux and is well tested on that platform**

Next steps: between-shot computing on the WAN

- **Deploying a supercomputer to support pseudo real-time analysis**
  - Network QoS
  - CPU scheduling
  - Faster CPUs and algorithms
  - Data management
  - End-to-end performance



- **Substantially enhanced data analysis**
  - Historically this had made a huge scientific impact

- **Can have a safety impact for future devices**
  - e.g. ITER: <10% of high power discharges can disrupt

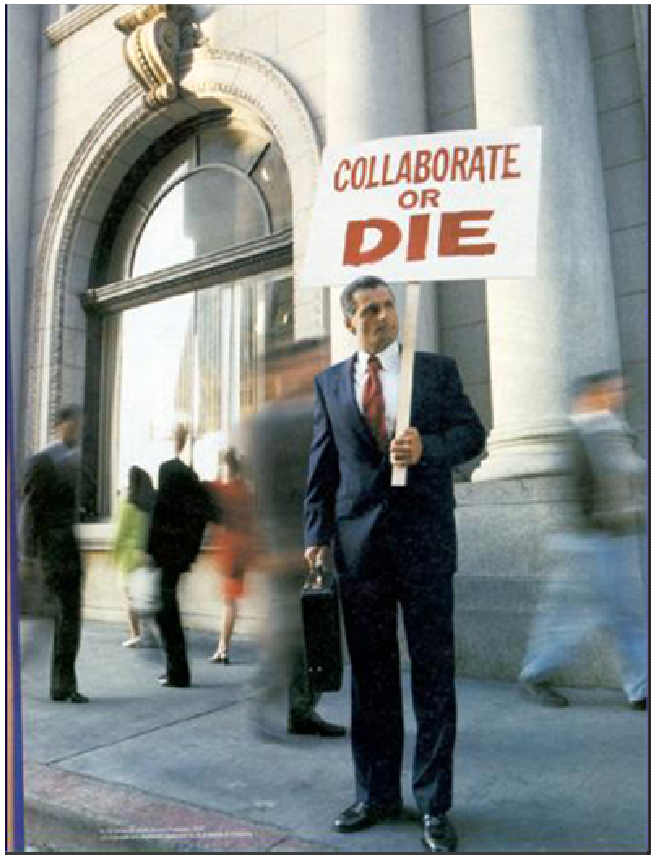
Example: between-shot TRANSP prototyped

- **A between-shot TRANSP proof-of-concept was done in March**
- **TRANSP (typically 12-36 hour code runs) adapted for shot cycle of 15 to 20 minutes**
  - Optimizations made without compromising security
- **Used canned inputs for each shot, but otherwise completely real**
  - Advance CPU reservation made
  - Launched each shot from DIII-D
  - Inputs shipped to PPPL
  - TRANSP executed at PPPL
  - Outputs written back to DIII-D

- **Six shots over two hours**

The security solutions developed for FusionGrid can be used with future computing infrastructures

- **Fusion research is and will continue to be a team sport**
- **Must collaborate between institutions**
- **Brings up exactly the security challenges addressed by FusionGrid**
  - Security in a "virtual organization"
- **Credential manager, ROAM, and secure MDSplus can be used in future computing infrastructures (ITER)**



Lessons Learned

- **Coordination is needed between site security and grid security**
  - Some site security requirements make distributed computing impossible
    - Example: strict firewall rules
  - Site security moving towards One Time Passwords (OTP)
  - Grid security moving towards Credentials

- **Tricky performance issues arise when moving from LAN to WAN**
  - Higher latency on networks
  - Some technologies for grids have poor performance characteristics (SOAP implementations)

- **Users don't get new concepts—stick to existing metaphors as much as possible**
  - Username/password, not self-managed credentials

Moving Forward

- **Web portals are a promising technology being investigated**
  - No need to distribute special client software
  - Users already understand how to use a web browser

- **Looking into technologies to ensure network Quality of Service (QoS) for remote computing over the WAN during tokamak operations**
  - Performance guarantees needed when shipping inputs/outputs between tokamak and supercomputing center
  - Investigating Multi-Layer Packet Switching (MLPS)

- **How to meet requirements of grid security and site security?**
  - OTP, Firewalls, and Credentials

- **Must not wait until *after* ITER is built to move forward**

Conclusion

- FusionGrid developers addressed the difficult security problems of authentication, authorization, and secure data storage & transfer in a grid environment
- A credential management system using X.509 credentials and MyProxy solved the authentication problem
- The ROAM authorization management system solved the authorization problem
- MDSplus was updated to work with these technologies
- These solutions may be extended to future computing infrastructures such as ITER that will have the same challenges of security in a multi-institutional environment