the globus project
www.globus.org

# MDS-2.1 and Futures

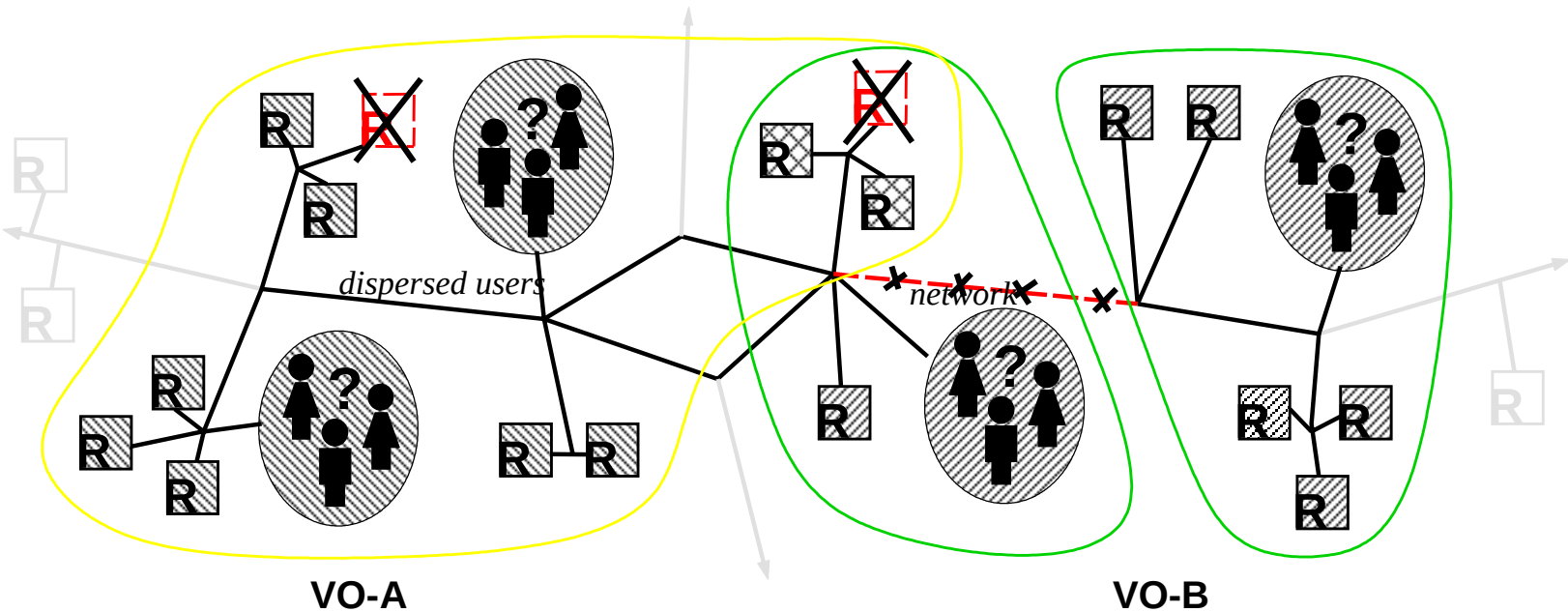
Karl Czajkowski

Information Sciences Institute
University of Southern California

# Talk Outline

- Introduction
  - Problem, history, etc.
- MDS-2 Architecture
  - Protocols
  - Features, services
- MDS-2.1 Software
  - Technology map
- Information Model
  - Additional background

# Resource Discovery/Monitoring




- Distributed users and resources
- Variable resource status
- Variable grouping and connectivity

# Basic Grid Acquisition Phases

- Resource *Discovery*
  - "What resources are relevant?"
  - Bootstraps planner state
- Resource *Status Inquiry*
  - "How do resources compare (now)?"
  - Refines planner knowledge
- Resource *Control*
  - "Did I acquire the resources?"
  - **Not part of information service**!

# MDS History

- MDS-1 (classic)
  - Centralized database
  - Globus 1.1.2 and earlier
  - Did not scale
- MDS-2
  - Distributed services
  - MDS 2.0 in Globus 1.1.3
  - New MDS 2.1 development in beta release

# Base Features

- Virtual Organizations (VOs)
  - Group together resources and users
  - Support community-specific “discovery”
  - Specialized “views”
- Scalability
  - Many resources
  - Many VOs
  - Graceful degradation of service

# Virtual Organizations

- Collaborating individuals and institutions
  - Shared goals
  - Enable sharing of resources
  - Non-locality of participants
- Dynamic in nature
  - VOs come and go
  - Resources join and leave VOs
  - Resources change status and fail
- Community-wide goals

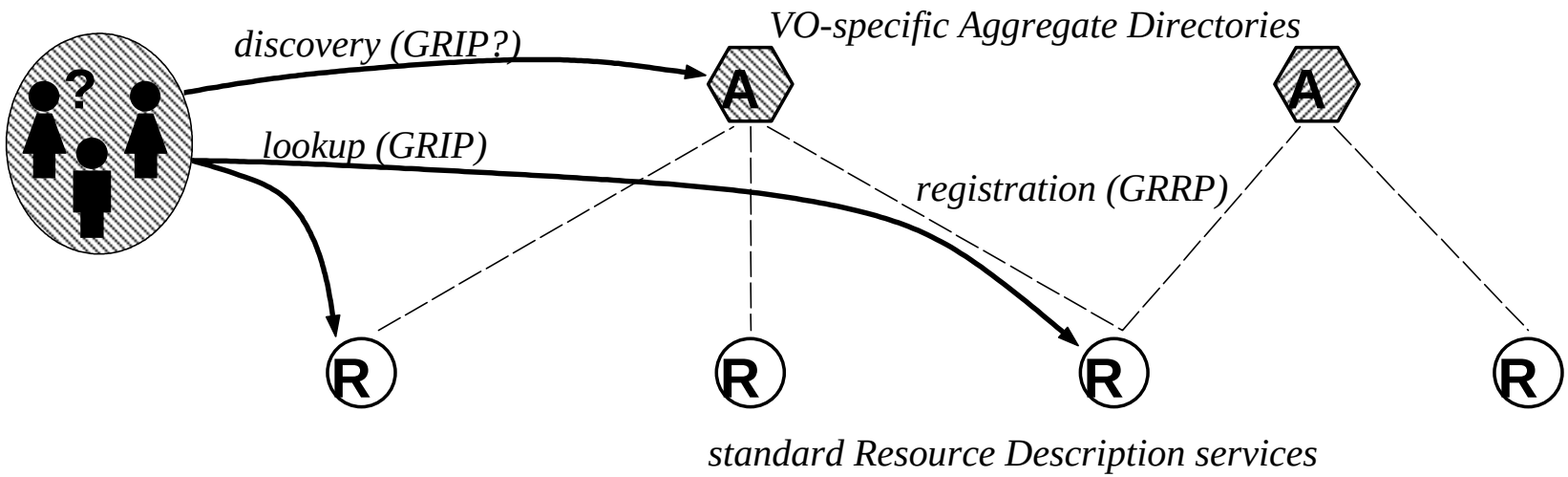
# Scalability

- Large numbers
  - Many resources
  - Many users
- Independence
  - Resources shouldn't affect one another
  - VOs shouldn't affect one another
- Graceful degradation of service
  - "As much function as possible"
  - Tolerate partitions, prune failures

# New MDS-2.1 Features

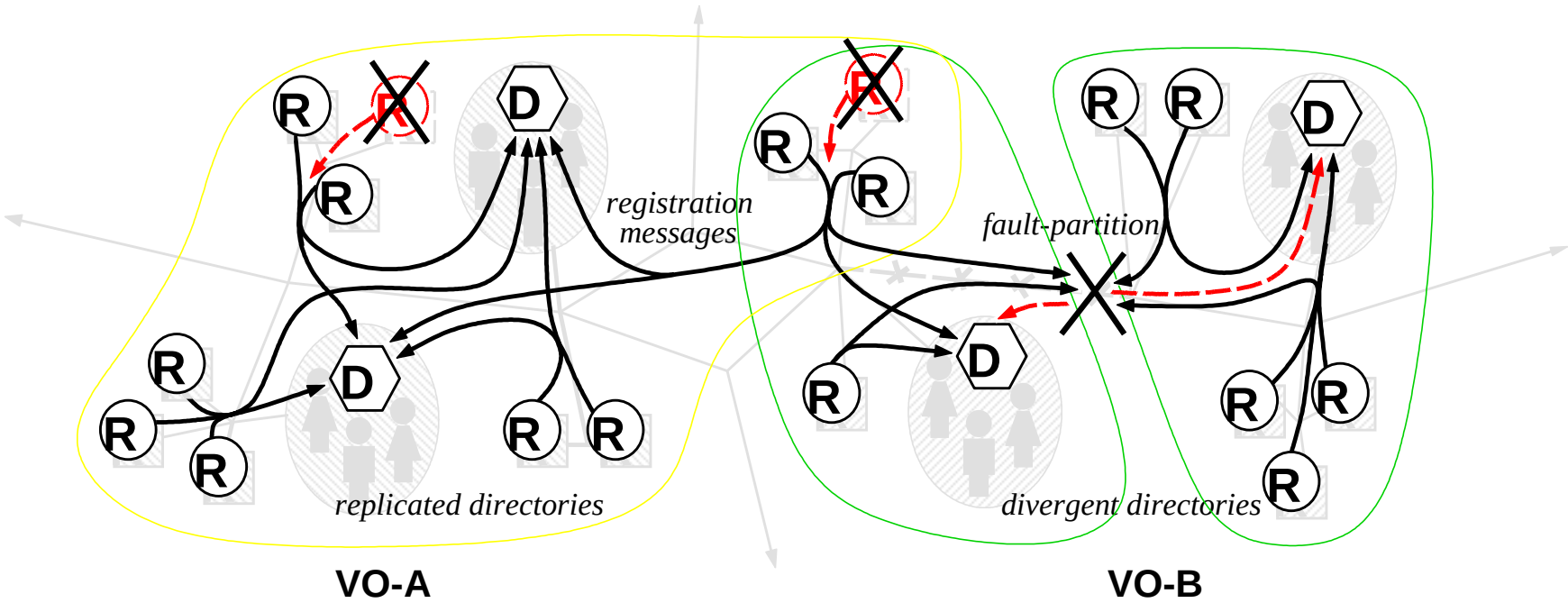
- Security Mechanisms
  - GSI mutual-authentication
  - Fine-grained access control by GSI name
- Performance Enhancements
  - Better query speeds
  - Less stale information
- New Information Model (schema)
  - Better representation of computers
  - Cleaner namespace management

# MDS-2 Service Architecture




- Dynamic Registration via Reg. Protocol (GRRP)
- Resource Inquiry via Info. Protocol (GRIP)
  - Co-located with resource on network
- Resource Discovery (via GRIP or other)
  - Using GRIP allows resource/directory hierarchy

# Distributed Services




- Service scales with Grid growth
- Loose consistency model tolerates failures
- Interoperability by GRIP/GRRP protocols

# Soft-state Registration

- Periodic notification
  - "Service/resource is available"
  - Expected-frequency metadata
- Automatic directory construction
  - Add new resources to directory
  - Invite resources to join new directory
- Self-cleaning
  - Reduce occurrence of "dead" references

# MDS-2 Implementation

- Grid Information Service (GRIS)
  - Provides resource description
  - Modular content gateway
- Grid Index Information Service (GIIS)
  - Provides aggregate directory
  - Hierarchical groups of resources
- Lightweight Dir. Access Protocol (LDAP)
  - Standard with many client implementations
  - Used for GRIP (and GRRP currently)

# MDS-2.1 Development Activities

- Incorporates external advances
  - New OpenLDAP 2.0.x code-base
  - Cyrus-SASL/GSI security integration
  - Leveraging new Globus packaging model
- Improving internal components
  - Better query servicing
  - New configuration/policy support
  - Invitation (reverse registration)
    - > Not in current beta...

# MDS-2.1 External Software Stack

- OpenLDAP 2.0.x (.14)
  - Implements LDAPv3 protocol
  - Client and server components
- Cyrus-SASL
  - Generic security
  - We provide loadable SASL/GSS plugin
- Globus GSI (repackaged)
  - Provides GSS-API interface to PKI
  - Shared library used by our SASL plugin

# MDS 2.1 Security

- PKI authentication
- Static authorization
  - Class, attribute, object name rules
- “Self” authorization
  - Semi-dynamic rule
  - Requires “owner” attribute on objects
- Dynamic authorization
  - Directory-based group lists (or future CAS)
  - Per-object access rule attributes
  - LDAP dynamic authorization (beta?)

# MDS-2.1 Internal Software

- Wrappers/tools
  - Simplify typical idioms
- Feature-specific GRIS providers
  - Resource-specific information probes
  - Modular interface to GRIS
- LDAP server “backend” modules
  - GRIS provider dispatch/caching
  - GIIS implementation
- SASL/GSS plugin

# MDS-2.1 GRIS Providers

- `grid-info-cpu` reports CPU/load info
- `grid-info-fs` reports filesystem info
- `grid-info-mem` reports RAM/VM info
- `grid-info-net` reports NIC/net info
- `grid-info-os` reports OS info
- `grid-info-platform` reports arch. info
- `grid-info-merged` merges all host info*
- Extensible for other sources, e.g. GRAM

# 2.1 GRIS Provider Times

| OS | RH 6.2 | RH7.1 | Irix | Solaris |
|---|---|---|---|---|
| Platform | 0.04 | 0.03 | | |
| Os | 0.06 | 0.04 | | |
| Cpu | 0.28 | 0.15 | | |
| Mem | 0.12 | 0.10 | | |
| Fs | 0.14 | 0.09 | | |
| Net | 0.45 | 0.17 | | |
| Merge | 0.80 | 0.37 | | |
| total | 1.89 | 0.95 | | |
| OLD | 10.01 | N/A | 8.29 | 18.27 |

# GRIS Dispatch Logic

- For each provider:
  1. Could search intersect provider? No, then skip.
  2. Is provider cache stale? Yes, then refill.
  3. Apply search filter to cache data.
- Combine all intersecting providers' results

# GRIS Response Issues

- MDS 2.0 and 2.1 are lazy
  - Probes are not issued unless queried
- Some system probes are slow
  - “Best” probe may take several seconds
- How to avoid stale data?
  - Clients set time-out per query
  - GRIS/GIIS define time-out per source
  - Fresh data found before timeout is returned
  - Cache fill continues after client time-out

# Hierarchical GIIS

- Maintain set of remote services
  - Track incoming live registrations
  - GRIS or GIIS registrants
- Cached proxy results (now), or
  - Same cache logic as GRIS
  - Refill cache with “chaining” queries
- LDAPv3 referral results (planned)
  - Do not maintain any local info cache
  - Redirect clients to active registrants

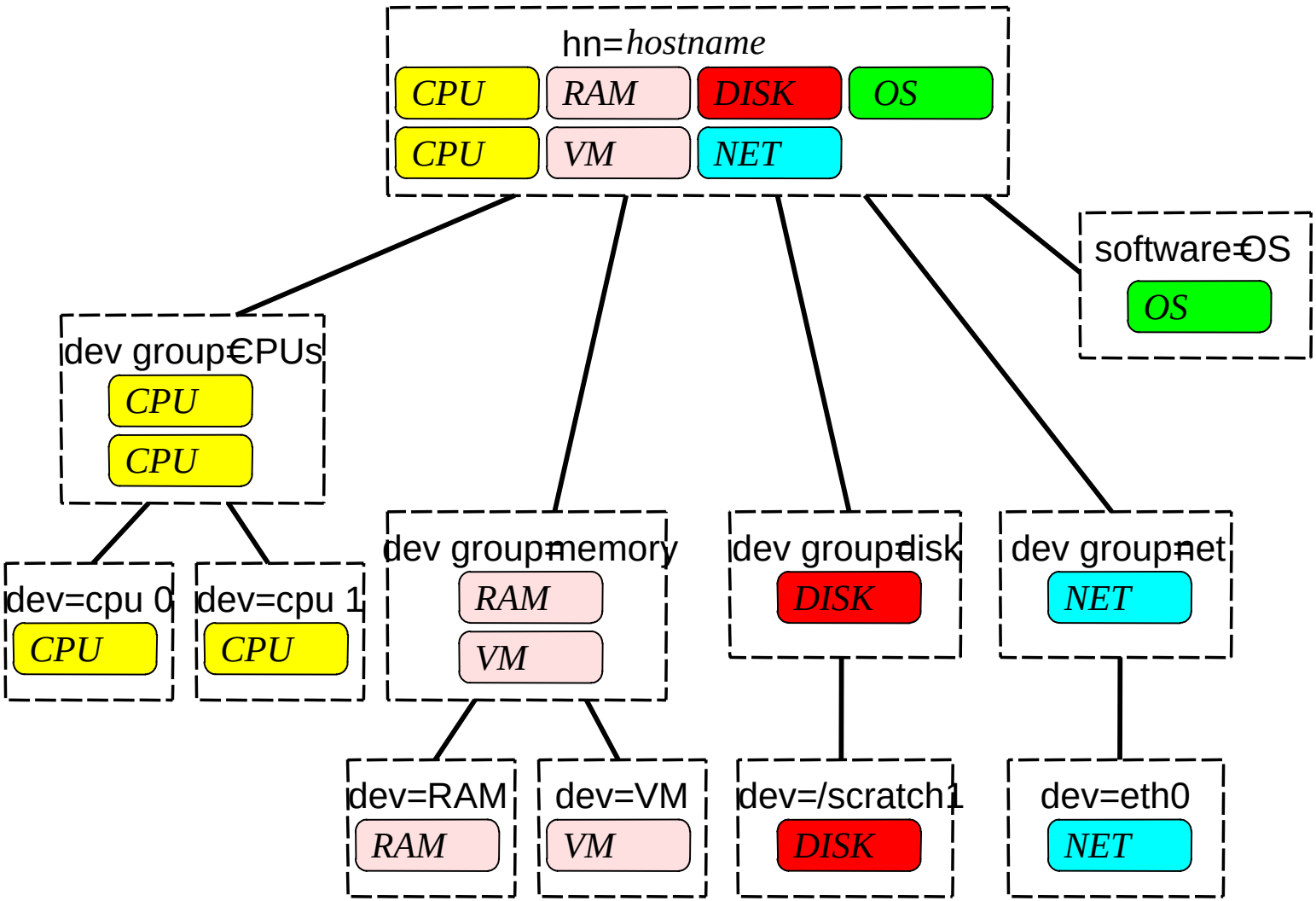
# Extensible GIIS Framework

- Modular registration actions
  - Re-use registration protocol decoding
  - Specialize directory update
  - e.g. prefetch indexable data
- Modular query actions
  - Re-use query protocol decoding
  - Specialize query handler algorithm
  - e.g. utilize precomputed indices
- Not in current beta...

# MDS 2.1 Information Model

- Structural information
  - Resource hierarchy maps to objects
  - Named positions in LDAP DIT
- Merged information
  - Some parents “join” child data
  - Simplifies common query patterns
- Auxiliary information
  - Uniform representation of leaf/parent data
  - Uses LDAP auxiliary objectclasses

# GRIS Host Objects



# GRIS Object Hierarchy

- Mds-Host-name=*hostname*
  - Mds-Software-Deployment=operating system
  - Mds-Device-Group-name=processors
    - Mds-Device-name=cpu 0
  - Mds-Device-Group-name=memory
    - Mds-Device-name=physical memory
    - Mds-Device-name=virtual memory
  - Mds-Device-Group-name=filesystems
    - Mds-Device-name=/scratch1
    - Mds-Device-name=/scratch2
  - Mds-Device-Group-name=networks
    - Mds-Device-name=eth0

# GRIS Structural Class Hierarchy

Mds

Attr: Mds-validfrom (like createtime)

Attr: Mds-validto (accuracy metadata)

Attr: Mds-keepto (discard metadata)

MdsHost

MdsDevice

MdsDeviceGroup

MdsSoftwareDeployment

- Every MDS object: name, time metadata

# GRIS Auxiliary Class Examples

## MdsCpu

Attr: Mds-Cpu-vendor

Attr: Mds-Cpu-Model

Attr: Mds-Cpu- speedMHz

- Once per CPU

## MdsCpuCache

Attr: Mds-Cpu-Cache-1kB

- Once per CPU

## MdsCpuSmp

Attr: Mds-Cpu-Smp-size

- Once per SMP

## MdsCpuTotal

Attr: Mds-Cpu-Total-count

- Once per MPP

# GRIS Auxiliary Class Examples

## MdsCpuFree

Attr: Mds-Cpu-Free-1min

Attr: Mds-Cpu-Free-5min

Attr: Mds-Cpu-Free-15min

- once per SMP

## MdsCpuTotalFree

Attr: Mds-Cpu-Total-Free-1min

Attr: Mds-Cpu -Total-Free-5min

Attr: Mds-Cpu -Total-Free-15min

- once per MPP

# More Information

- Questions?
- HPDC-10 Paper (August 2001)
  - "Grid Information Services for Distributed Resource Sharing"
- Globus Toolkit 2.0 beta
  - Early access to test code