



the globus alliance

[www.globus.org](http://www.globus.org)

# Science Clouds: Early Experiences in Cloud Computing for Scientific Applications

Chicago, October 2008

Kate Keahey, Renato Figueiredo, Jose  
Fortes, Tim Freeman, Mauricio Tsugawa

University of Chicago

University of Florida



# Science Clouds

- Make it easy for scientific projects to experiment with cloud computing
  - ◆ Can cloud computing be used for science?
- Evolve software in response to the needs of scientific projects
  - ◆ Start with EC2-like functionality and evolve to serve scientific projects: virtual clusters, diverse resource leases
  - ◆ Federating clouds: moving between cloud resources in academic and commercial space



- University of Chicago (Nimbus):
  - ◆ first cloud, online since March 4th 2008
  - ◆ 16 nodes of UC TeraPort cluster, public IPs
- University of Florida
  - ◆ Online since 05/08
  - ◆ 16-32 nodes, access via VPN
- Other Science Clouds
  - ◆ Masaryk University, Brno, Czech Republic (08/08), Purdue (09/08)
  - ◆ Configurations in progress: Vrije University (Amsterdam), Clemson University, FZK, ORNL
- Using EC2 for overflow
- <http://workspace.globus.org/clouds>

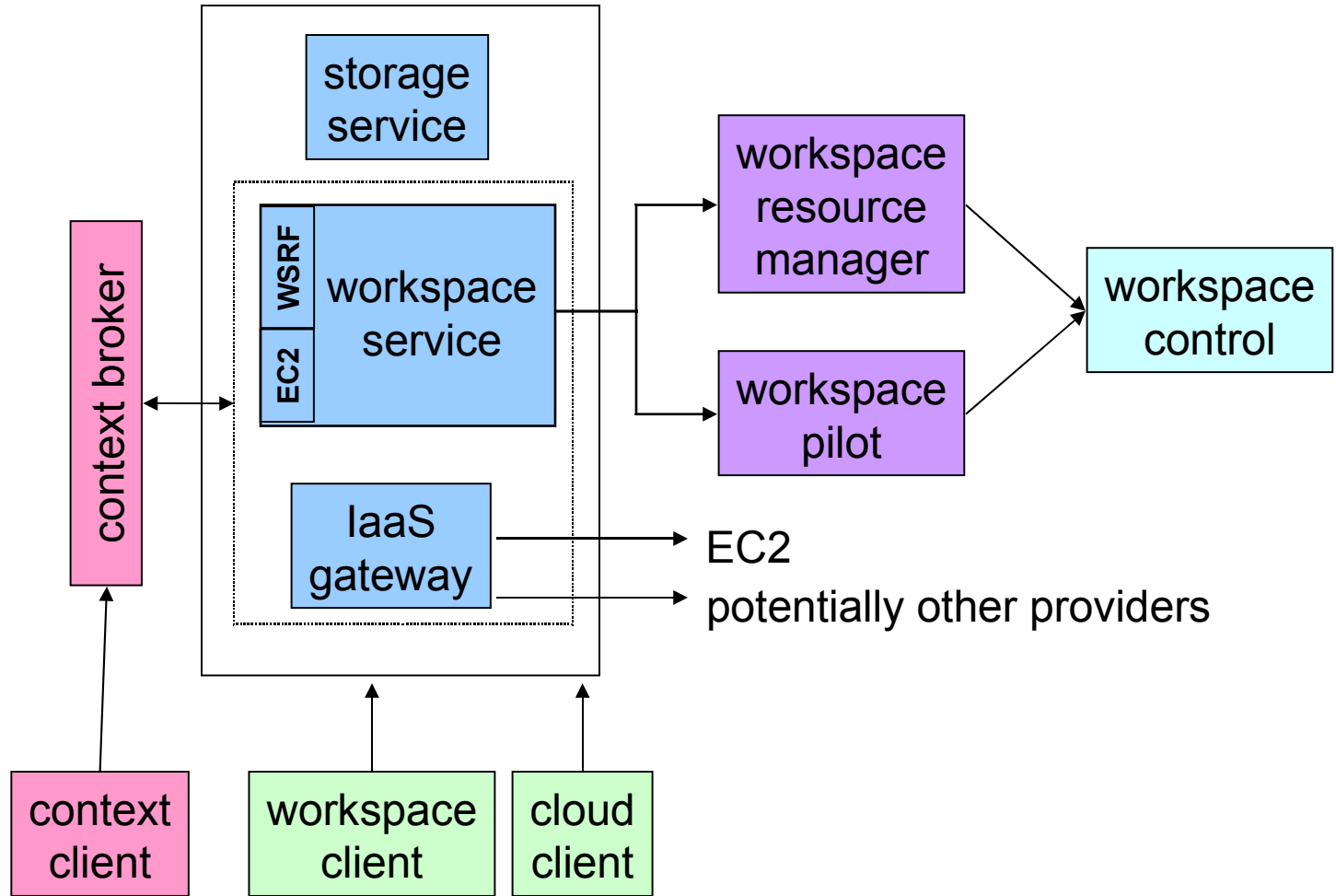


# The Nimbus Toolkit

- Provides implementation of Infrastructure-as-a-Service (IaaS)
  - ◆ WSRF interfaces and EC2 interfaces
  - ◆ Originally based on Xen, now also KVM
  - ◆ Formerly called “workspace service” (one of the Nimbus components)
- Provide infrastructure serving the needs of the scientific community
- Started mid-2003, first release in 09/2005
- Open source, extensible architecture, allows us to experiment with different capabilities and SLAs
- Current release is 2.1 (October '08)
- Available from: <http://workspace.globus.org>



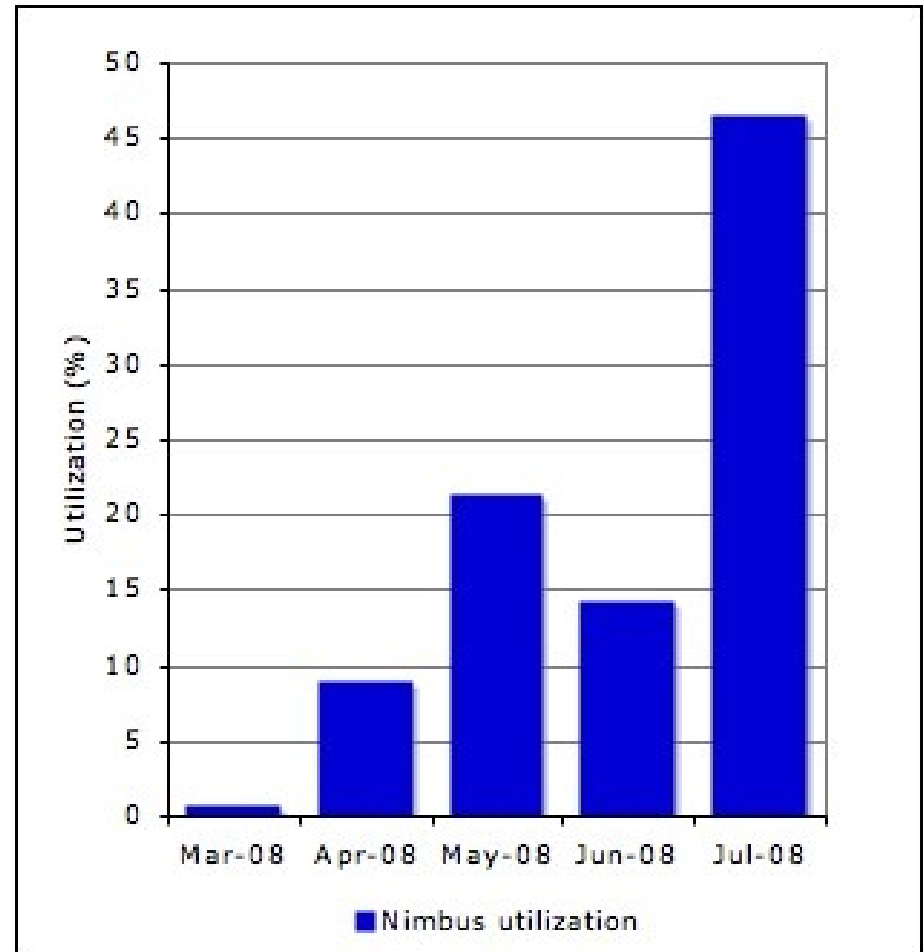
# 5-click Guide to Nimbus





# University of Chicago Cloud: Some Numbers

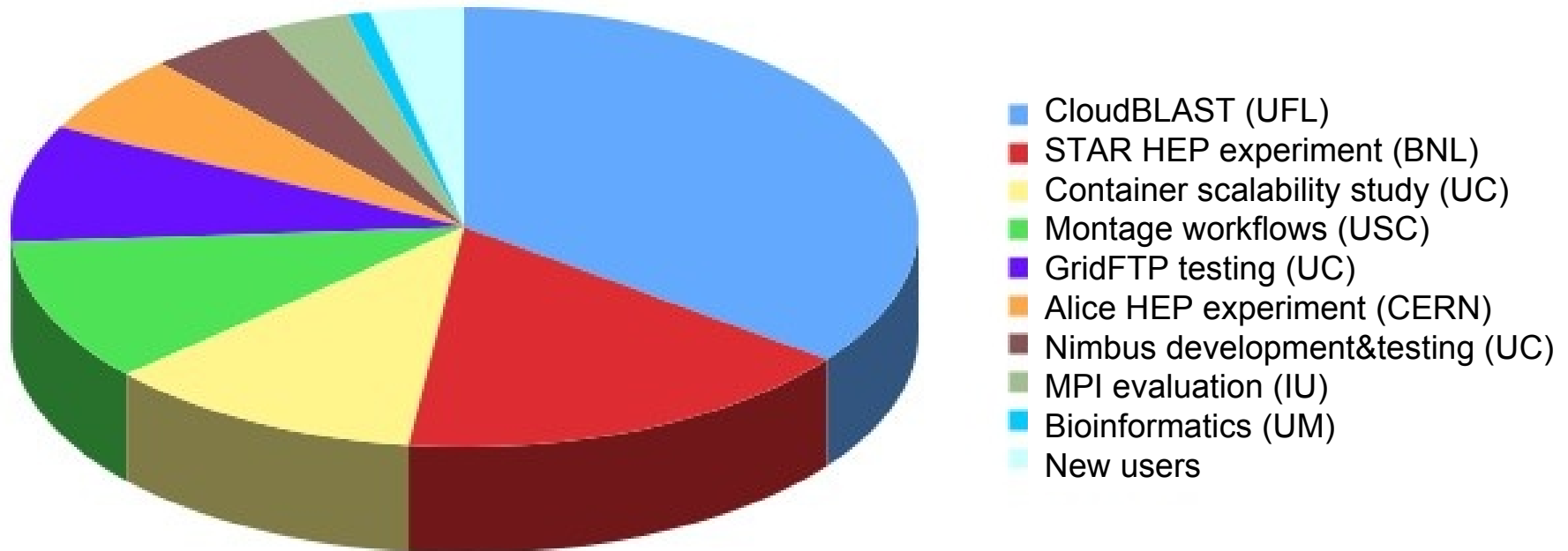
- ~100 DNs
- Utilization:
  - ◆ Overall: 20%
  - ◆ Peak pw: 86% (week of 7/14)
- Requests rejected:
  - ◆ None till 7/14
  - ◆ 65 after 7/14



*Data from March 4th to August 4th scaled to the number of days per given month*



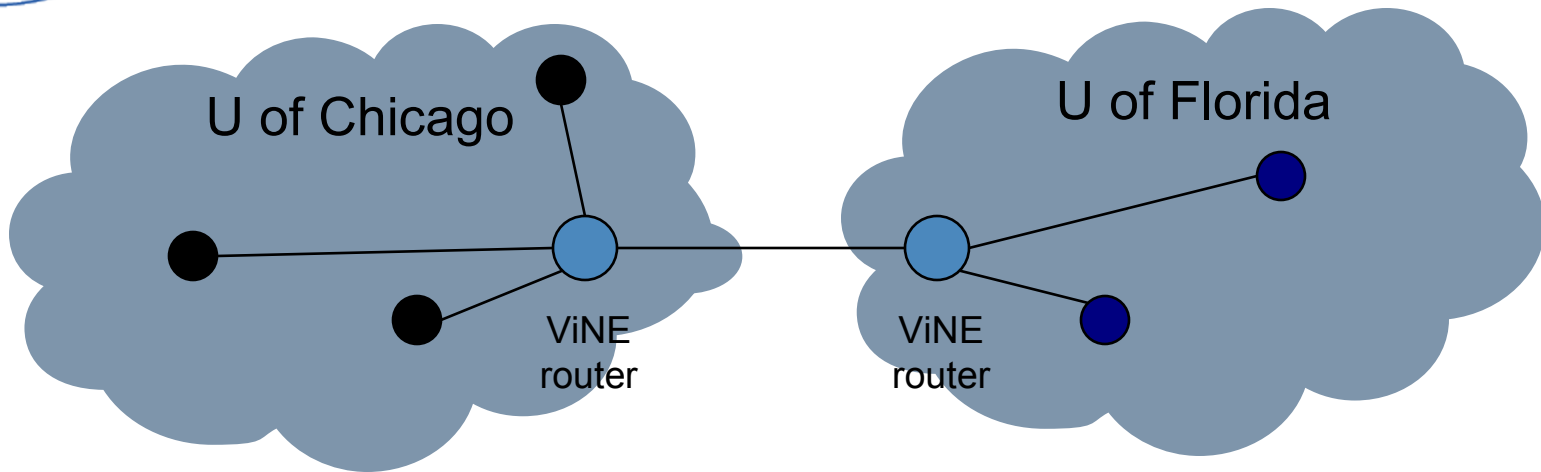
# Who Runs on Nimbus?



*Project diversity: Science, CS, education, build&test...*



# Hadoop Study

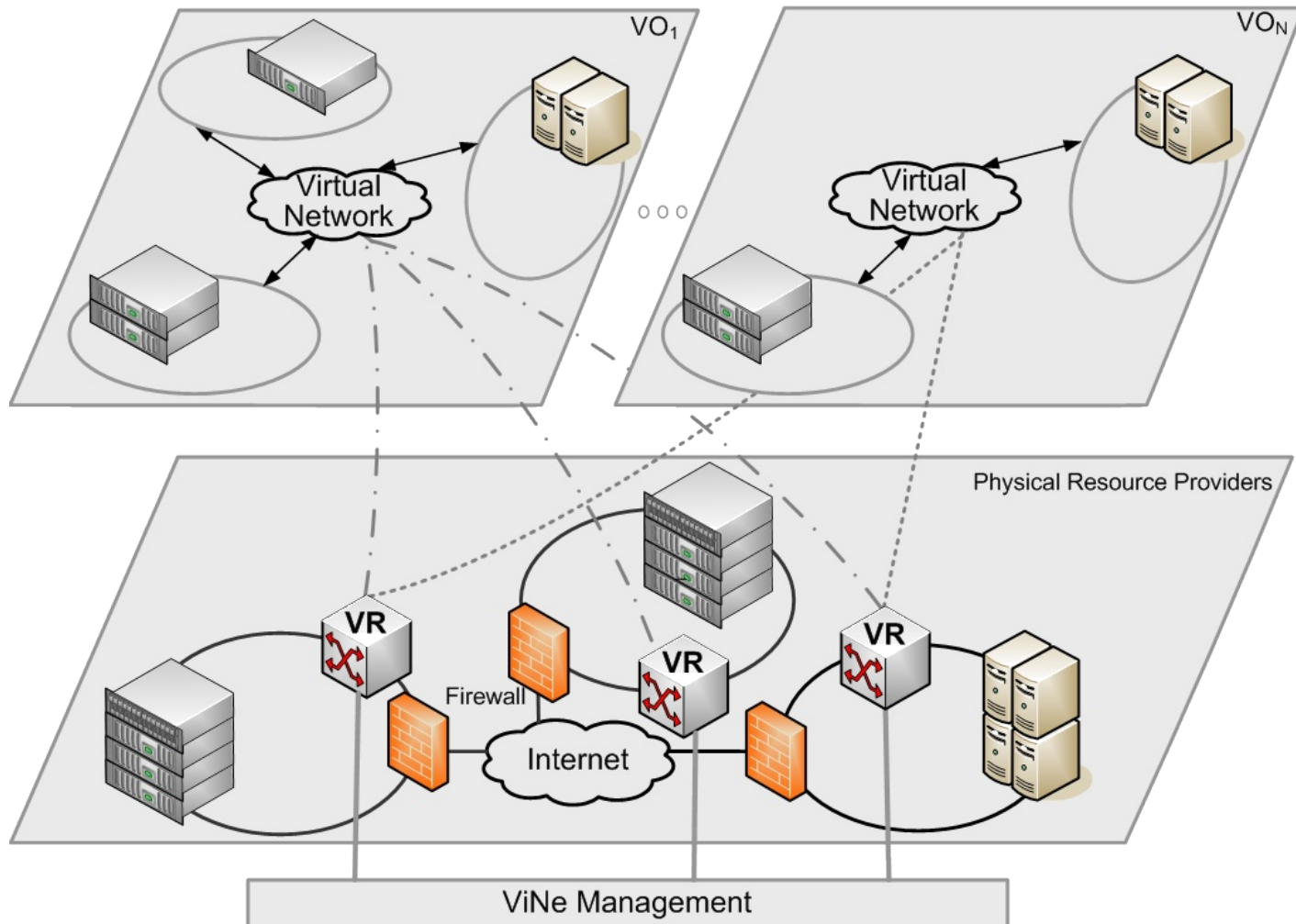


- CS research: investigate latency-sensitive apps, e.g. hadoop
- Need access to distributed resources, and high level of privilege to run a ViNE router
- Virtual workspace: ViNE router + application VMs
- *Paper: "CloudBLAST: Combining MapReduce and Virtualization on Distributed Resources for Bioinformatics Applications" by Andréa Matsunaga, Maurício Tsugawa and José FortesFirst, accepted to eScience 2008.*



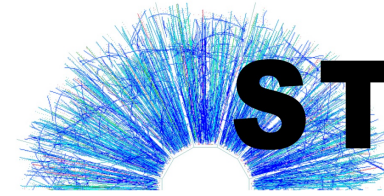


# Hadoop Study: Creating a Seamless Networking Domain with ViNE





- STAR: a high-energy physics experiment
- Need resources **with the right configuration**
  - ◆ Complex environments: correct versions of operating systems, libraries, tools, etc all have to be installed.
  - ◆ Consistent environments: require validation
- A virtual OSG STAR cluster
  - ◆ OSG cluster
    - OSG CE (headnode), gridmapfiles, host certificates, NSF, PBS
  - ◆ STAR worker nodes: SL4 + STAR conf
- Requirements
  - ◆ One-click virtual cluster deployment
  - ◆ Migration: Science Clouds -> EC2



- From proof-of-concept to production runs
  - ◆ ~2 years ago: proof-of-concept
  - ◆ Last September: EC2 runs of up to 100 nodes (production scale, non-critical codes)
  - ◆ Testing for critical production deployment
- Performance
  - ◆ Within 10% of expected performance for applications
- Work by Jerome Lauret, Doug Olson, Leve Hajdu, Lidia Didenko



# Scalability Testing

- Project: testing scalability of various Globus components on different platforms
- Need short-term but flexible access to diverse platforms
- Workspaces: Globus 101 + others
- Work by various members of the Globus Toolkit (Tom Howe and John Bresnahan)
- Resulted in provisioning a private cloud for Globus



# Montage Workflows

- Evaluating a cloud from user's perspective
  - ◆ *Paper: "Exploration of the Applicability of Cloud Computing to Large-Scale Scientific Workflows", C. Hoffa, T. Freeman, G. Mehta, E. Deelman, K. Keahey, SWBES08: Challenging Issues in Workflow Applications*



# Alice HEP Experiment at CERN

ALICE Repository

- ALICE Repository
- Google Map
- Shifter's dashboard
- Running trend
- Production info
- Job Information
- SE Information
- Services
- Network Traffic
- FTD Transfers
- CAF Monitoring
- SHUTTLE
- LCG exp. monitoring
- Build system
- Dynamic charts

close all

This page: bookmark, URL

Running jobs trend

Running jobs trend

7761

Jobs

Map Satellite Hybrid

2000 mi  
2000 km

Imagery ©2008 TerraMetrics, NASA - Terms of Use

● Running Jobs ● ML Service Down ● No Active Jobs ● ML Service Down & no running jobs

Map options

Show xrootd transfers  Show site relations

Jump to: Europe North America South America Asia World Save position and options

- *Preparing a CHEP paper*



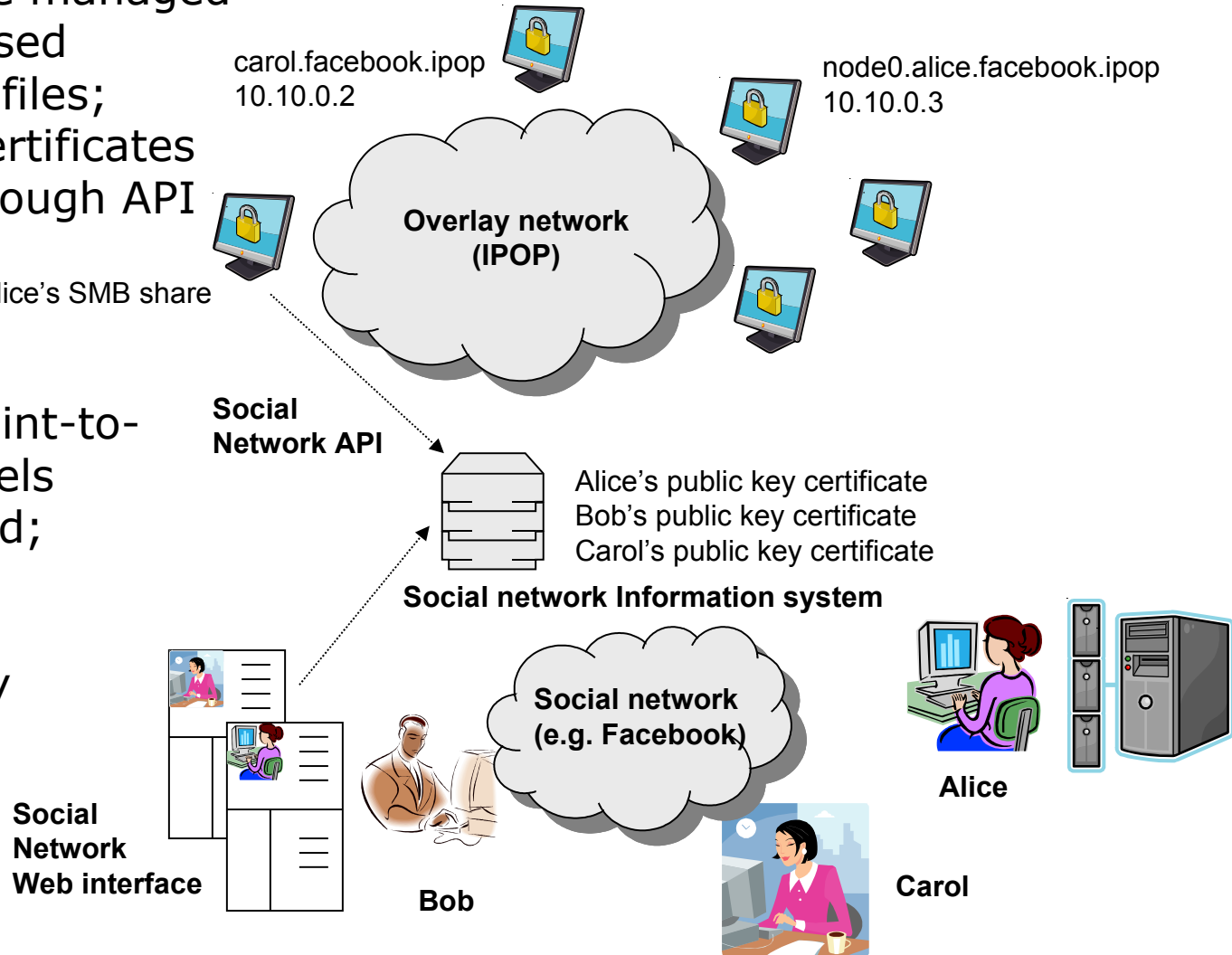
# How to Make It Easy: Social VPNs

Identities are managed with web-based interface profiles;  
Public key certificates retrieved through API

Bob: browses Alice's SMB share

Symmetric keys exchanged and point-to-point private tunnels created on demand;

Multicast-based resource discovery





# Thoughts

- **Impact:**
  - ◆ Utilization, time used per project, etc.
  - ◆ Scientific results, papers written and in preparation, ongoing discussions
- **Another kind of impact:**
  - ◆ We are doing things we could not do before:
    - Deploying network routers on remote platforms
    - Easily finding the right environment in distributed environment
    - Provisioning resource when we need them
- **We're learning what's possible**





- Kate Keahey: [keahey@mcs.anl.gov](mailto:keahey@mcs.anl.gov)
- Renato Figueiredo: [renato@acis.ufl.edu](mailto:renato@acis.ufl.edu)
  - ◆ Social networks
- Jose Fortes: [fortes@ufl.edu](mailto:fortes@ufl.edu)
- Tim Freeman: [tfreeman@mcs.anl.gov](mailto:tfreeman@mcs.anl.gov)
- Mauricio Tsugawa: [tsugawa@ufl.edu](mailto:tsugawa@ufl.edu)
  - ◆ ViNE