Crosscloud Computing

José Fortes

Advanced Computing and Information Systems Lab

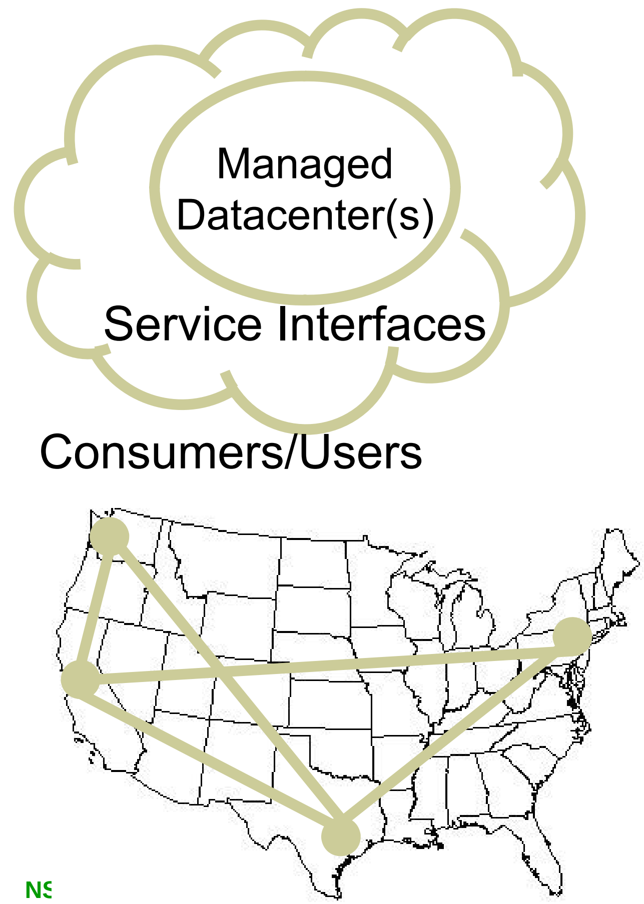
and

NSF Center for Autonomic Computing

HPC 2010 – Cetraro

Clouds

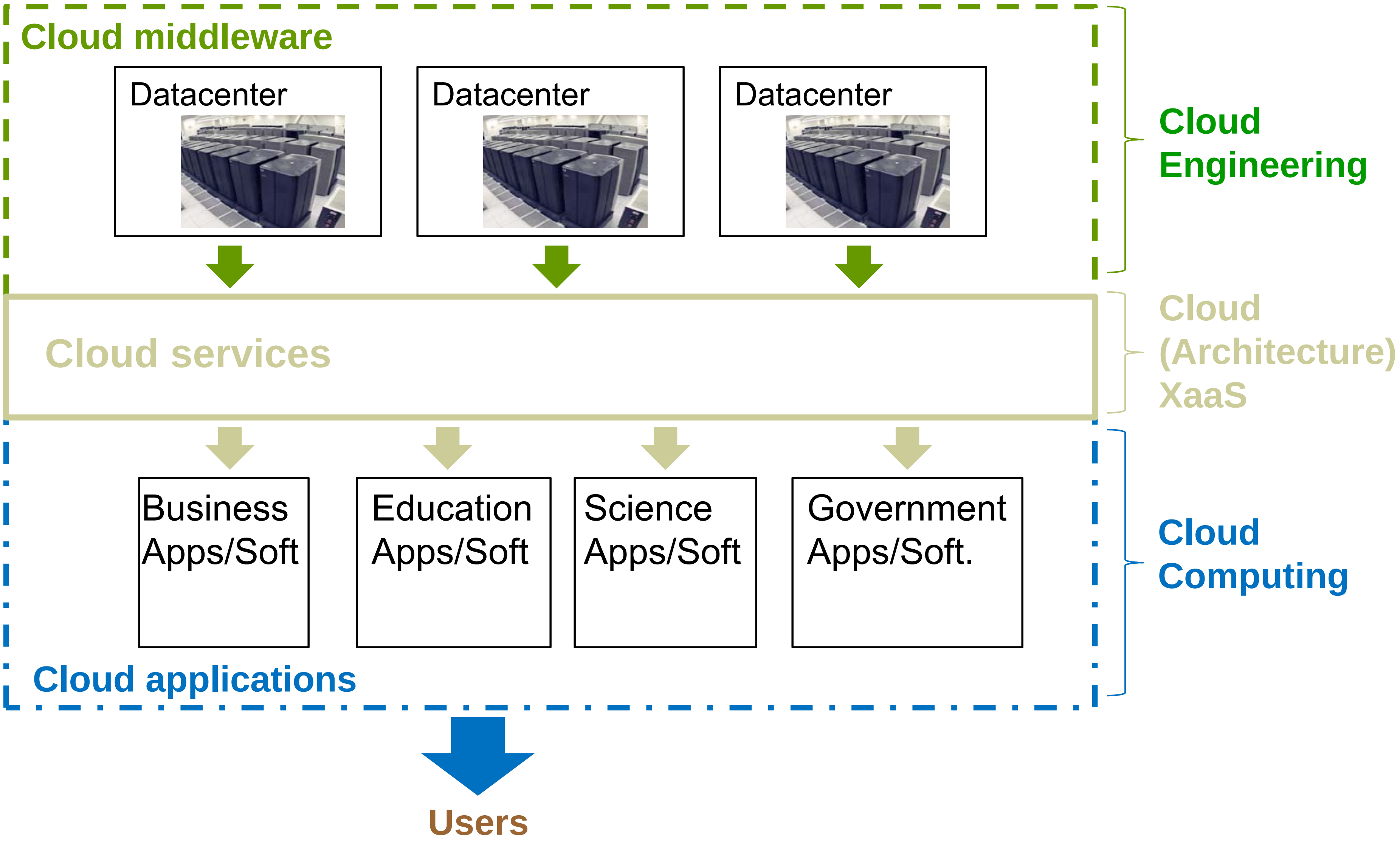
- Provider view
  - Economies of scale
  - Statistical multiplexing
  - Avoid customer-specific complexities
- Consumer view
  - No need to (over)provision
  - No operating costs
  - Pay per use
- Win-win decoupling
  - Virtualization in the large



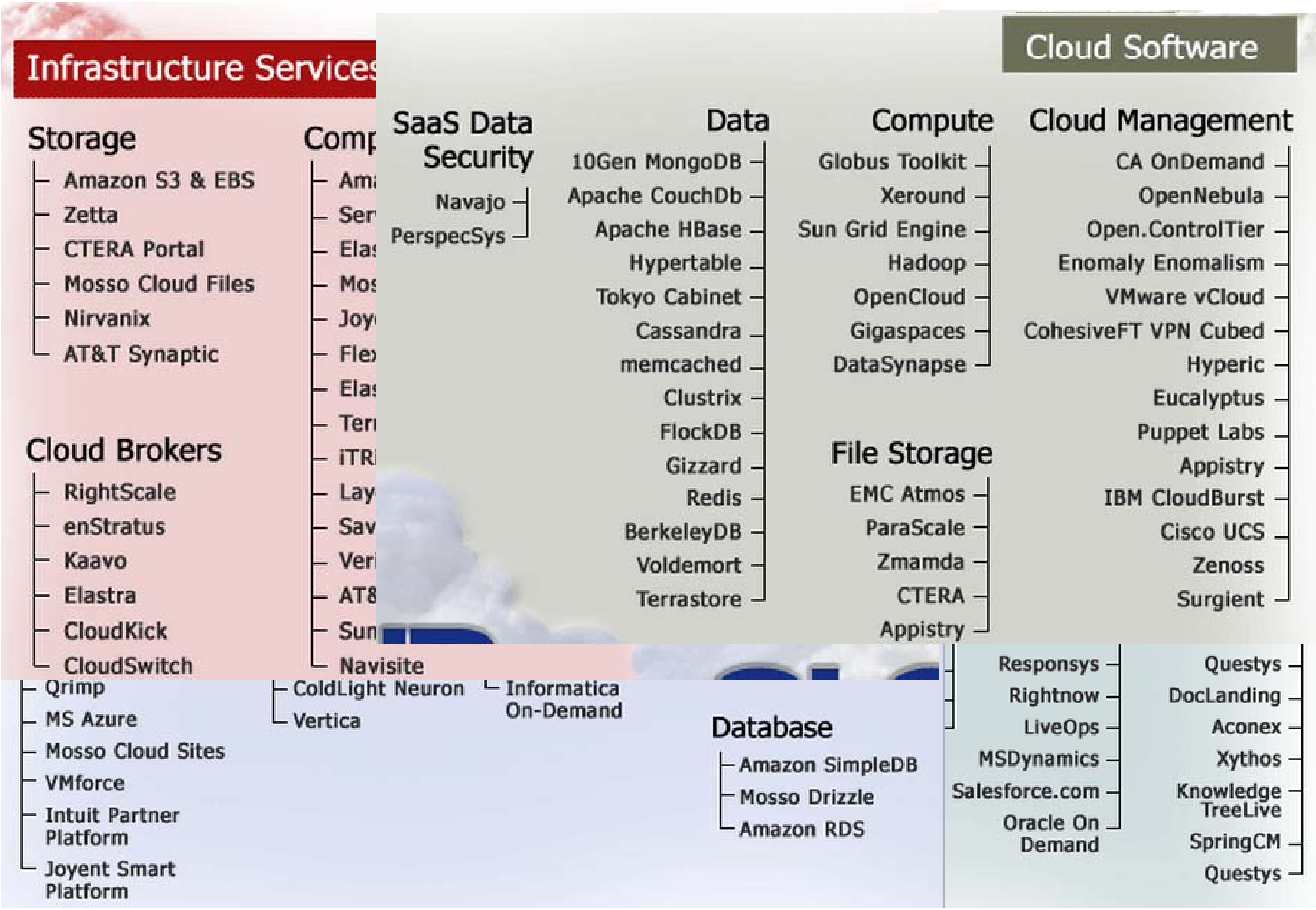
Outline

- Cloud world and ecosystem
- Sky computing
- Networking across clouds
- Other issues
  - Faults
  - Resource usage
  - Autonomics
- Conclusions

The world of a cloud

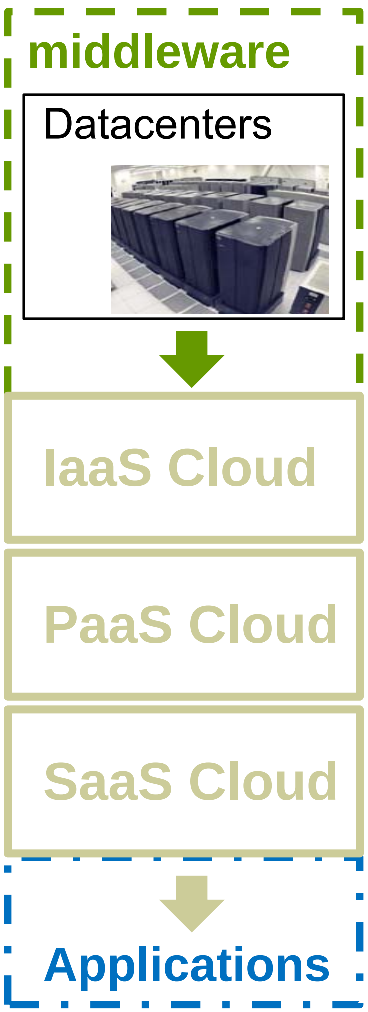


Cloud middleware and services

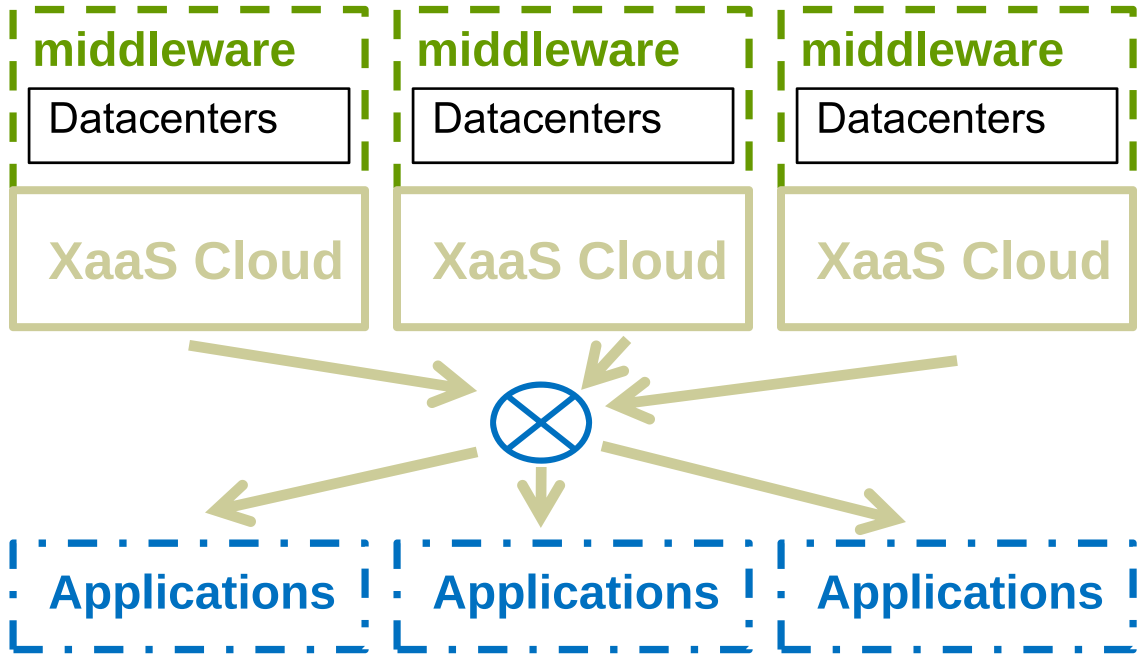


Early multi-cloud consumer systems

Nested

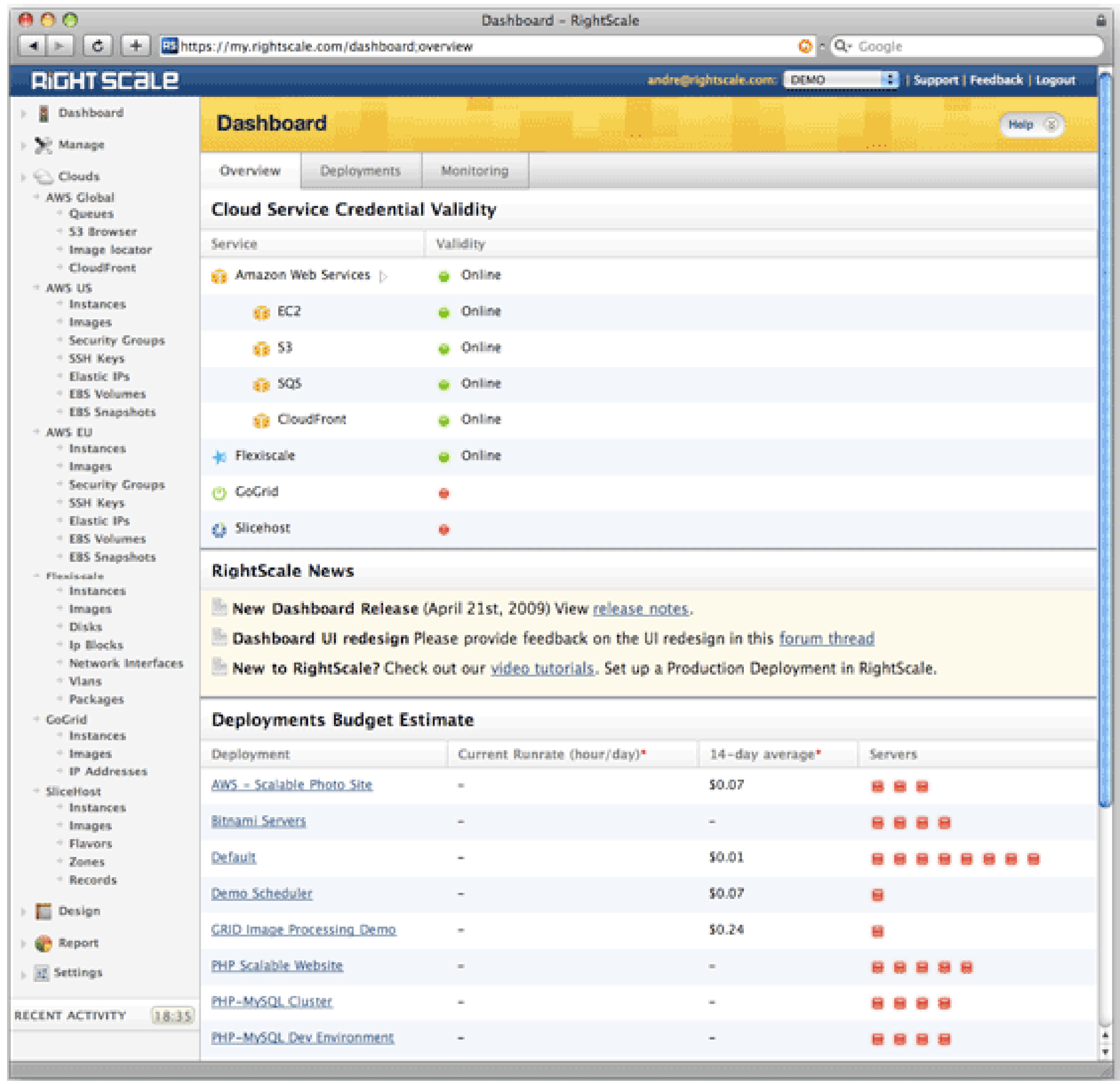


Replicated or diversified



- Geographic factors
  - Markets, location-dependent services
- Dependability/continuity
  - 24/7, disaster recovery, diversity …
- Provider independence
- On demand scale-out
- Differentiated services
- Different(iated) apps
- Hybrids

Multi-cloud management tools



Contextualization

- Nimbus

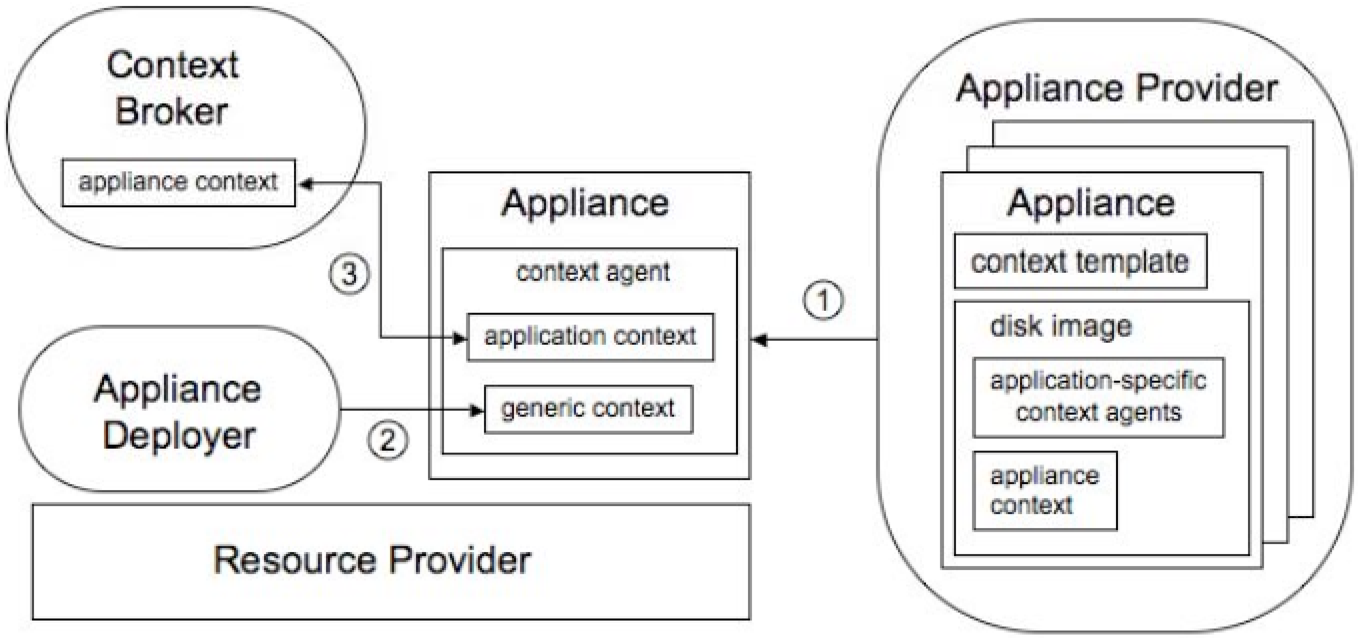
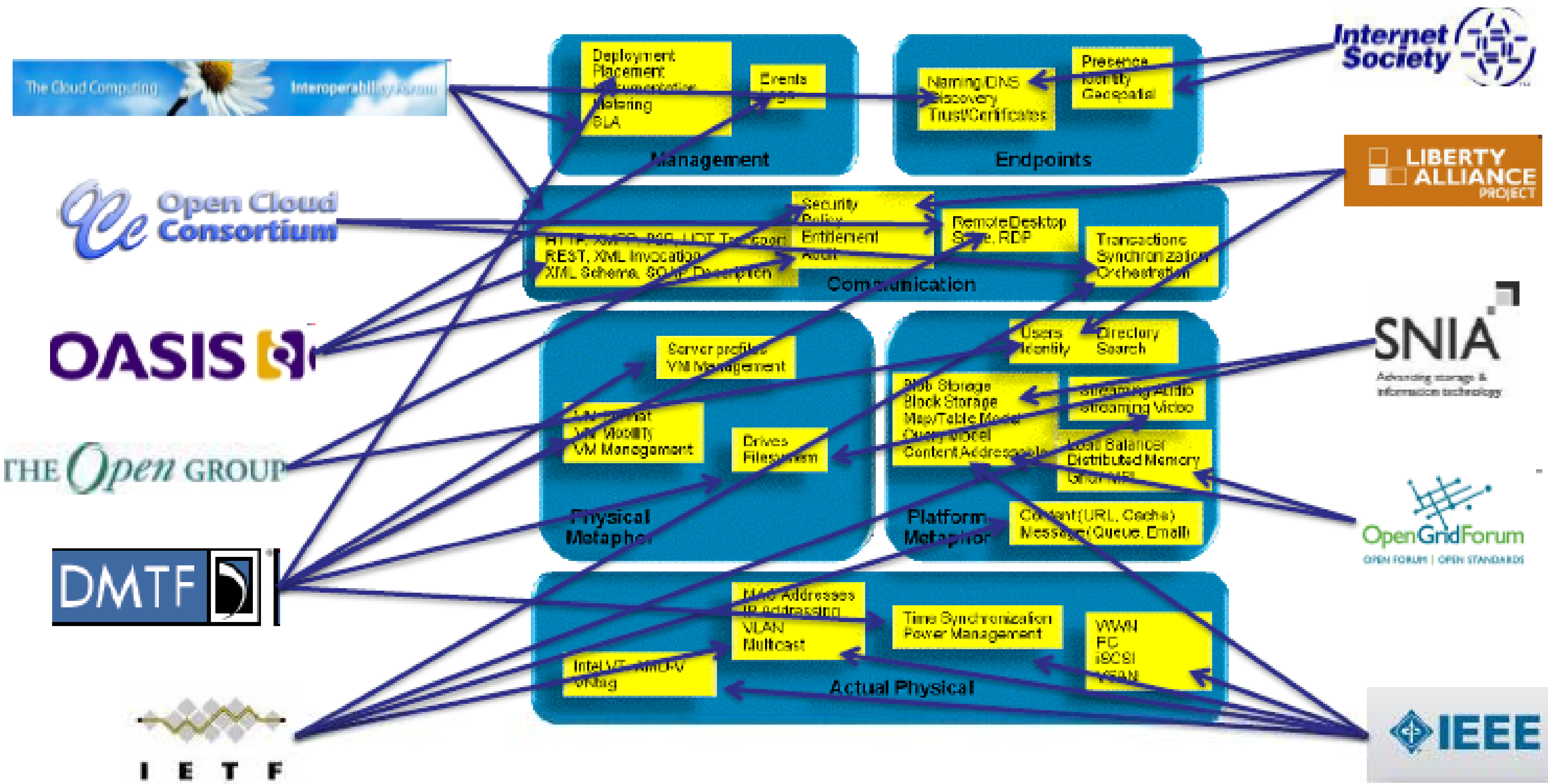


Figure 2: Relationship between appliance provider, appliance deployer, and context broker.

Contextualization: Providing One-Click Virtual Clusters, Keahey, K., T. Freeman. eScience 2008, Indianapolis, IN. December 2008.

Intercloud standards

- Protocols, formats and mechanisms for interoperability
- From David Bernstein, Hwawei Tech., www.cloudstrategypartners.com



http://ww.iaria.org/conferences2009/filesFUTURECOMPUTING09/DavidBernstein_Intro_to_Intercloud_V6.pdf

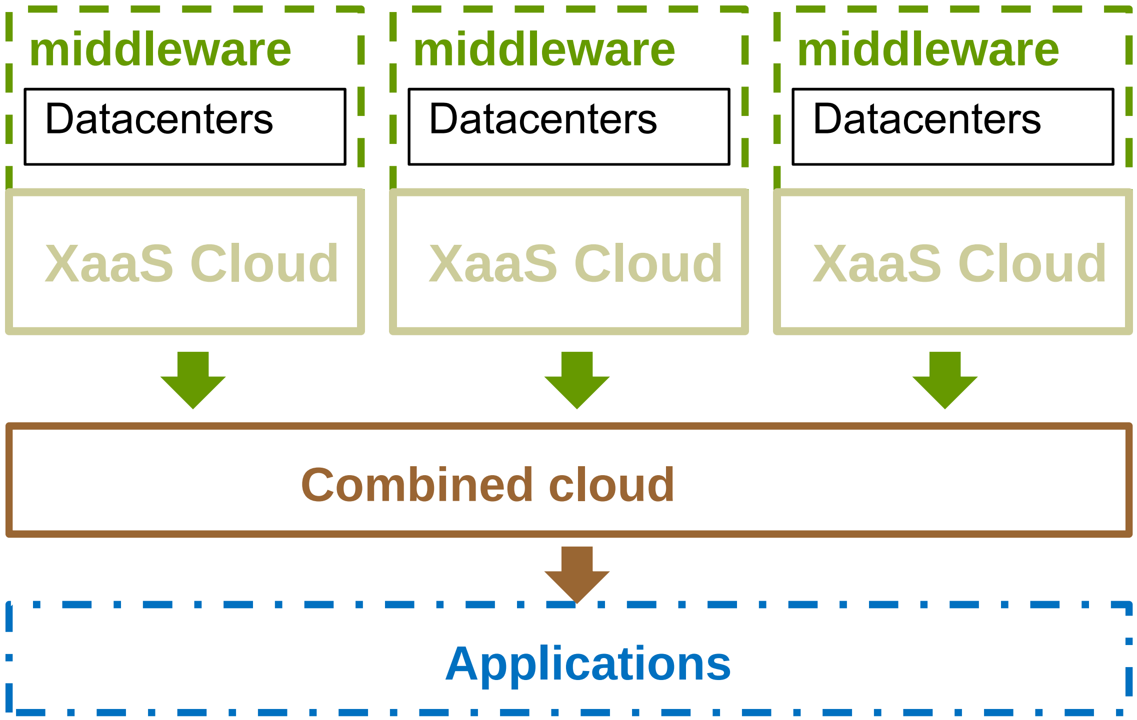
Combinatorial Innovation

"… historically, you'll find periods in history where there would be the availability of a different component parts that innovators could combine or recombine to create new inventions. In the 1800s, it was interchangeable parts. In 1920, it was electronics. In the 1970s, it was integrated circuits. Now what we see is a period where you have Internet components, where you have software, protocols, languages, and capabilities to combine these component parts in ways that create totally new innovations. The great thing about the current period is that component parts are all bits. That means you never run out of them. You can reproduce them, you can duplicate them, you can spread them around the world, and you can have thousands and tens of thousands of innovators combining or recombining the same component parts to create new innovation. So there's no shortage. There are no inventory delays. It's a situation where the components are available for everyone, and so we get this tremendous burst of innovation that we're seeing."

Hal Varian, chief Google economist and professor at UC Berkeley

Combined clouds

- Combine: to bring into such close relationship as to obscure individual characteristics



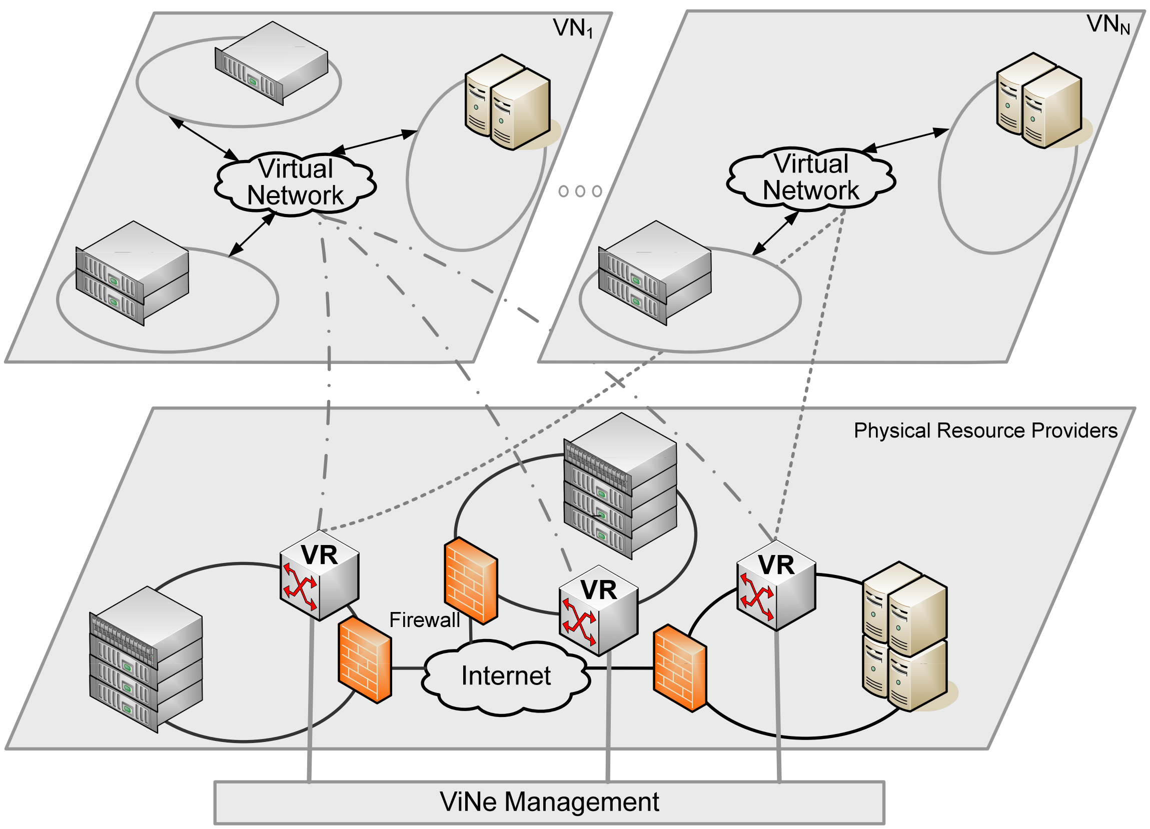
- "Heterogeneous virtual cluster on a WAN" aaS
- "(Excel-based) geospatial market analytics" aaS
- "Personalized health from multiple providers" aaS

Sky computing

- Use of combined clouds
  - Resources/apps/platforms across independent clouds are used
  - Services other than those of each individual cloud
  - Transparency of multiple clouds - single-cloud like
  - Sky providers are consumers of cloud providers
  - “Virtual” datacenter-less dynamic clouds
- Many challenges and questions
  - Communication among resources in different clouds is of key importance

Communication Problems

- Connectivity limitations due to the lack of publicly accessible addresses, firewalls, NATs …
  - solutions available for grid computing (API-based, IPOP, VNET, ViNe …), remote access (OpenVPN, CiscoVPN, …)
  - mostly based on user-level network virtualization



- E.g., ViNe general purpose overlay network solution
- Deploys user-level virtual routers used as gateways by nodes that do not run ViNe software
- Apps run unmodified
- Best performance

Cloud realities

- Dangers of VM privileged users
  - change IP and/or MAC addresses
  - configure Network Interface Card in promiscuous mode
  - use raw sockets
  - attack network (spoofing, proxy ARP, flooding, …)
- Cloud providers impose network restrictions that severely affect network virtualization

Small Instance – default*

- 1.7 GB memory
- 1 EC2 Compute Unit (1 virtual core with 1 EC2 Compute Unit)
- 160 GB instance storage (150 GB plus 10 GB root partition)
- 32-bit platform
- I/O Performance: Moderate
- API name: m1.small

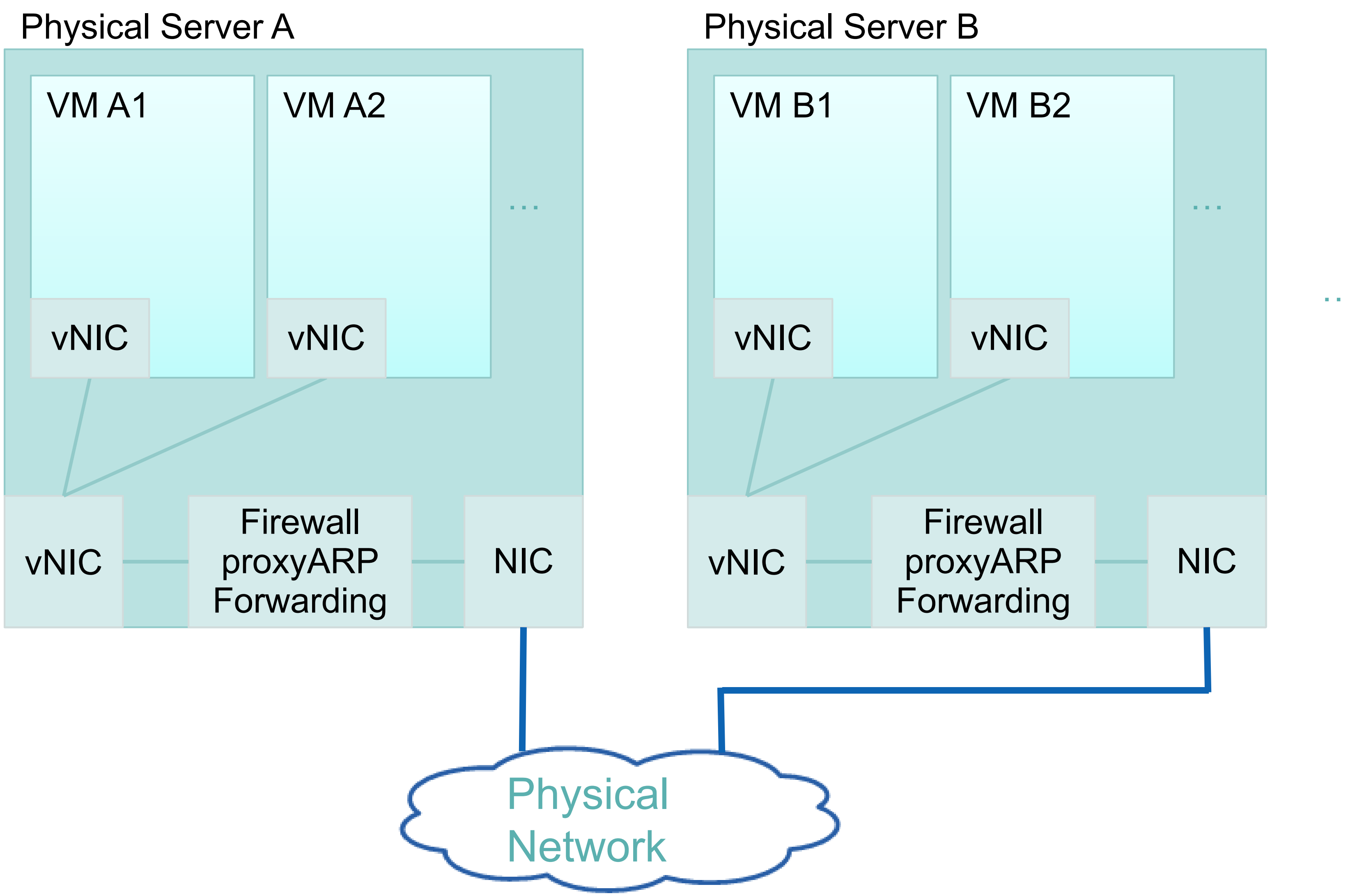
Network Restrictions in Clouds

- Internal routing and NAT
  - Assigned IP addresses (especially public) are neither visible nor modifiable from within the VMs, and NAT techniques are used
- Sandboxing
  - VMs are connected to host-only networks
  - VM-to-VM communication is enabled by a combination of NAT, routing and firewalling mechanisms
- Packet filtering (beyond usual)
  - VMs packets are inspected and only those packets containing valid addresses (IP and MAC assigned by the provider) are allowed

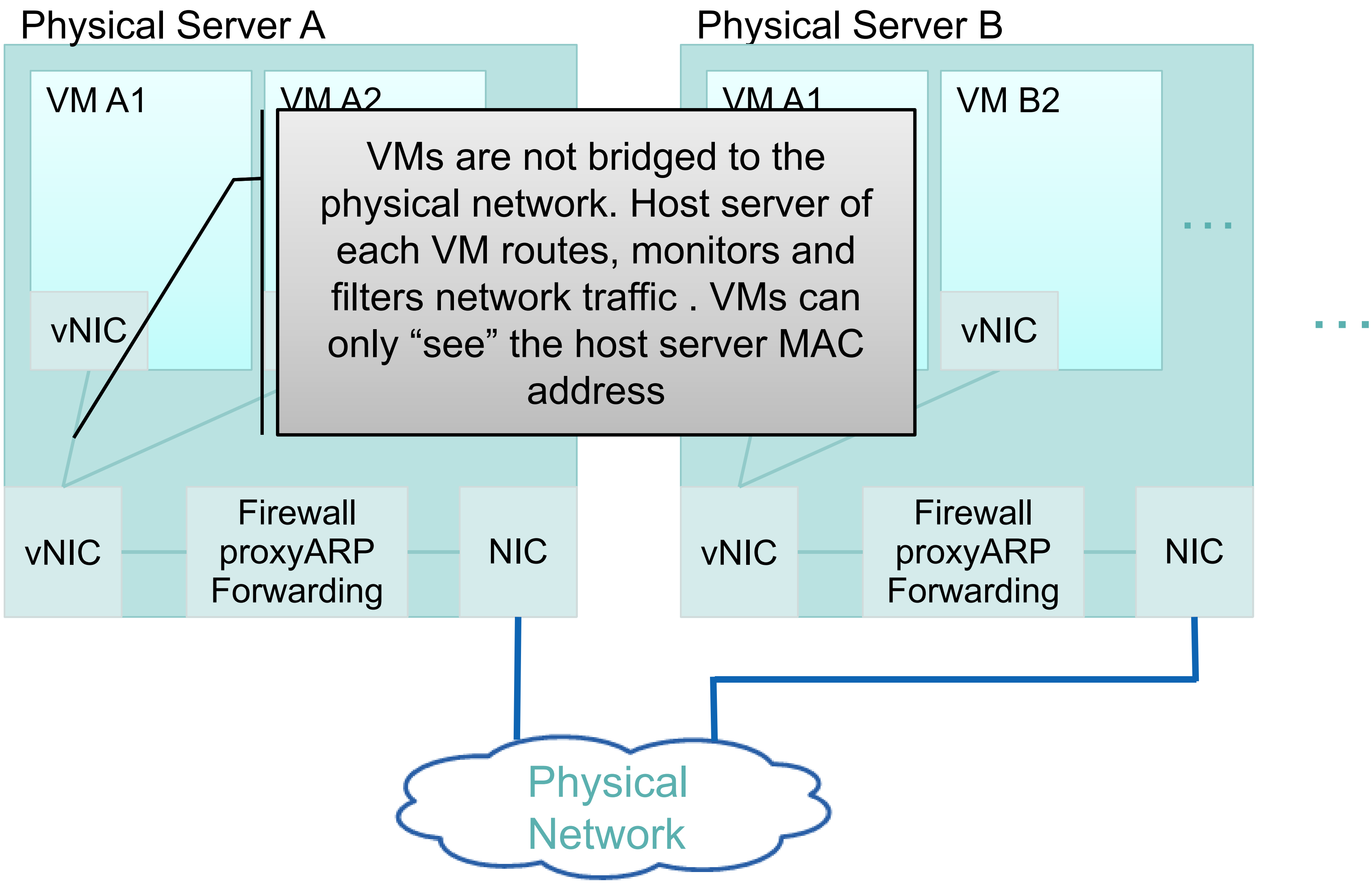
Network Challenges in Clouds

- Internal routing / NAT
  - High number of intermediate nodes (hops) in LAN communication (nodes on the same subnet, thus no hops in-between are expected)
    - EC2 public-to-public – 6+ hops
    - EC2 private-to-private – 3+ hops (better)
- Sandboxing
  - Disables direct datalink layer (L2) communication
  - Can't use VMs as routers
  - No node-to-gateway communication
- Packet filtering
  - Only allows packets w/ source IP address
  - Disables VM ability to act as a router
  - No gateway-to-node communication

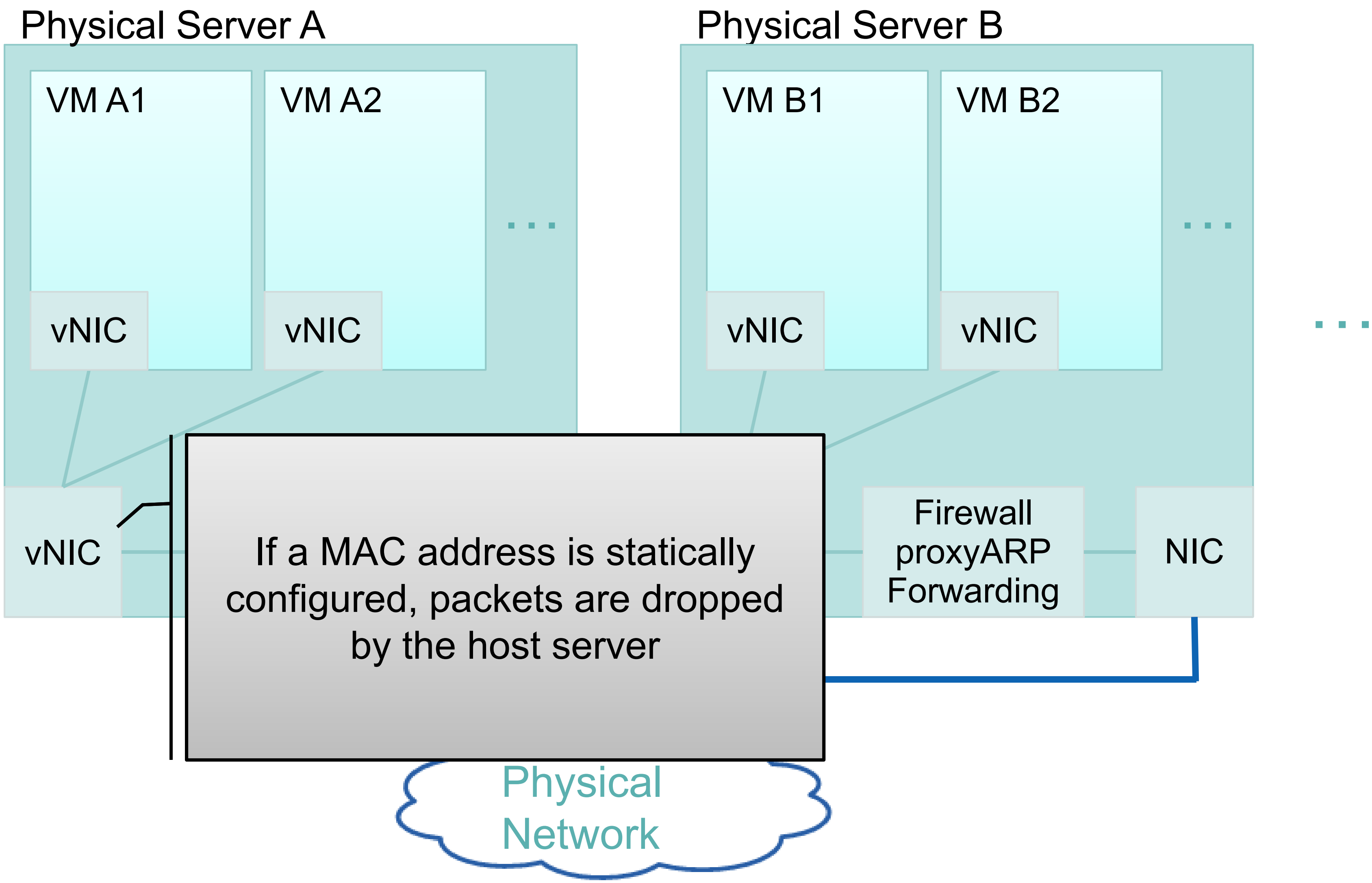
Typical IaaS Network



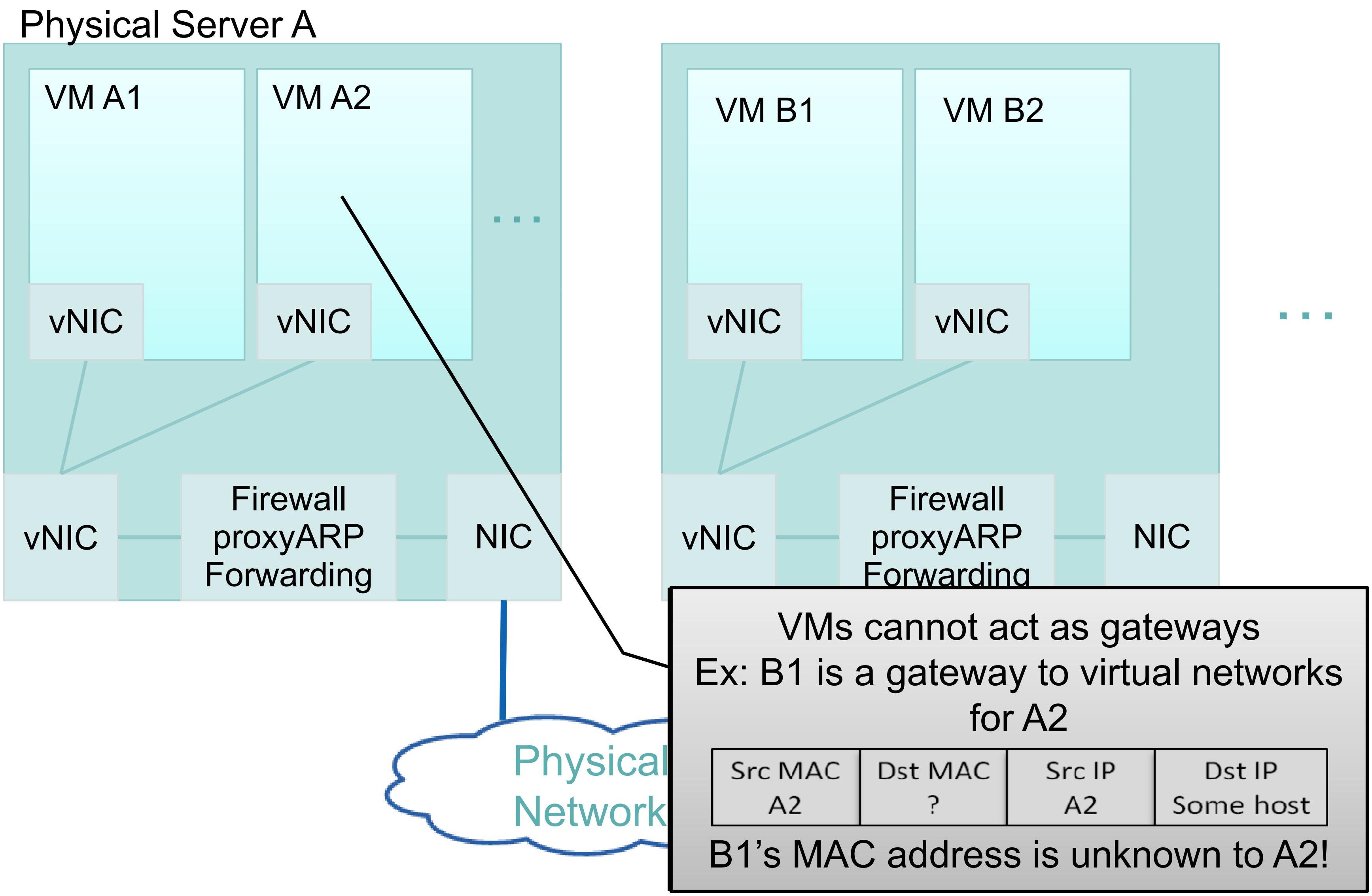
Typical IaaS Network



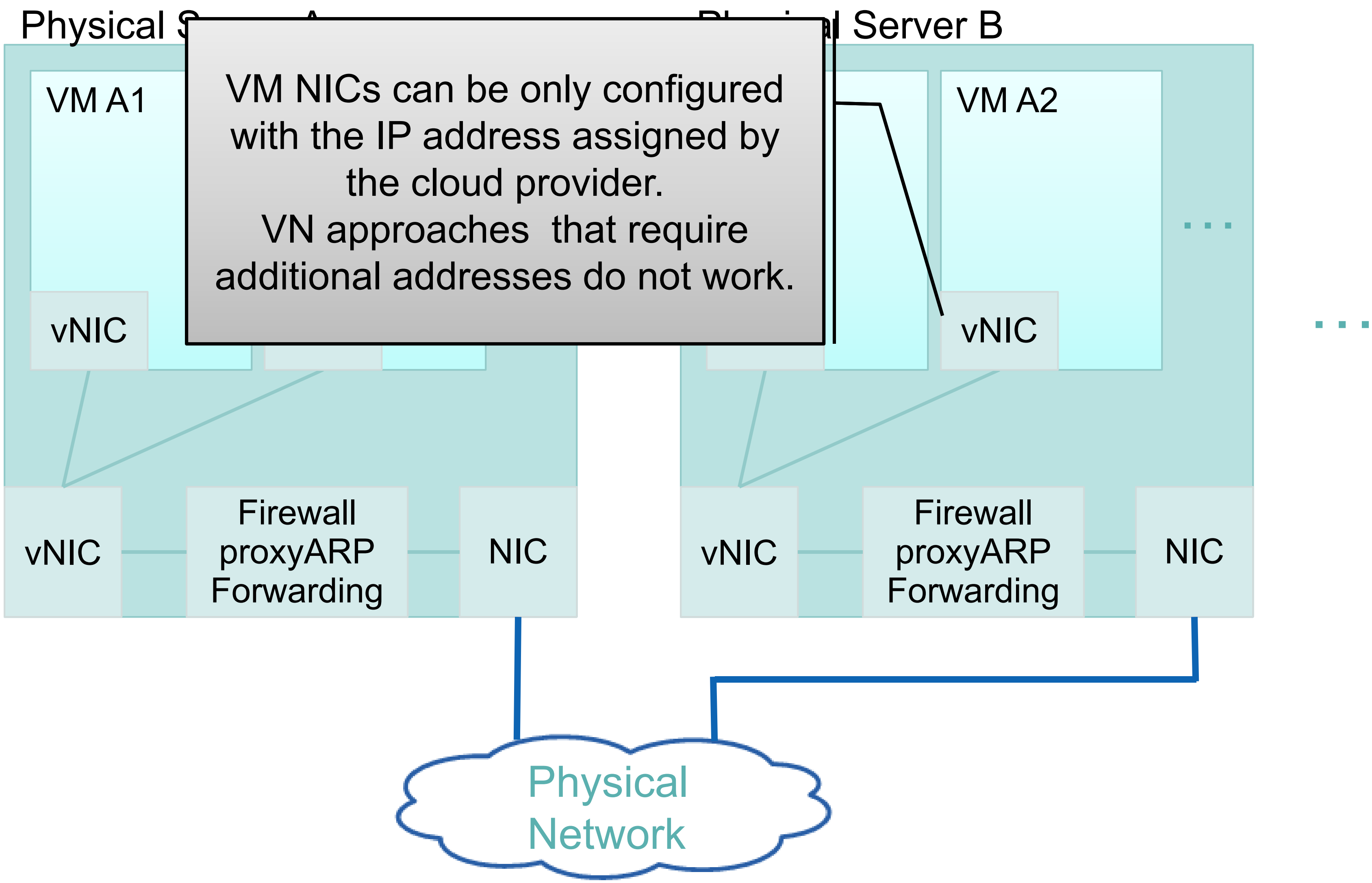
Typical IaaS Network



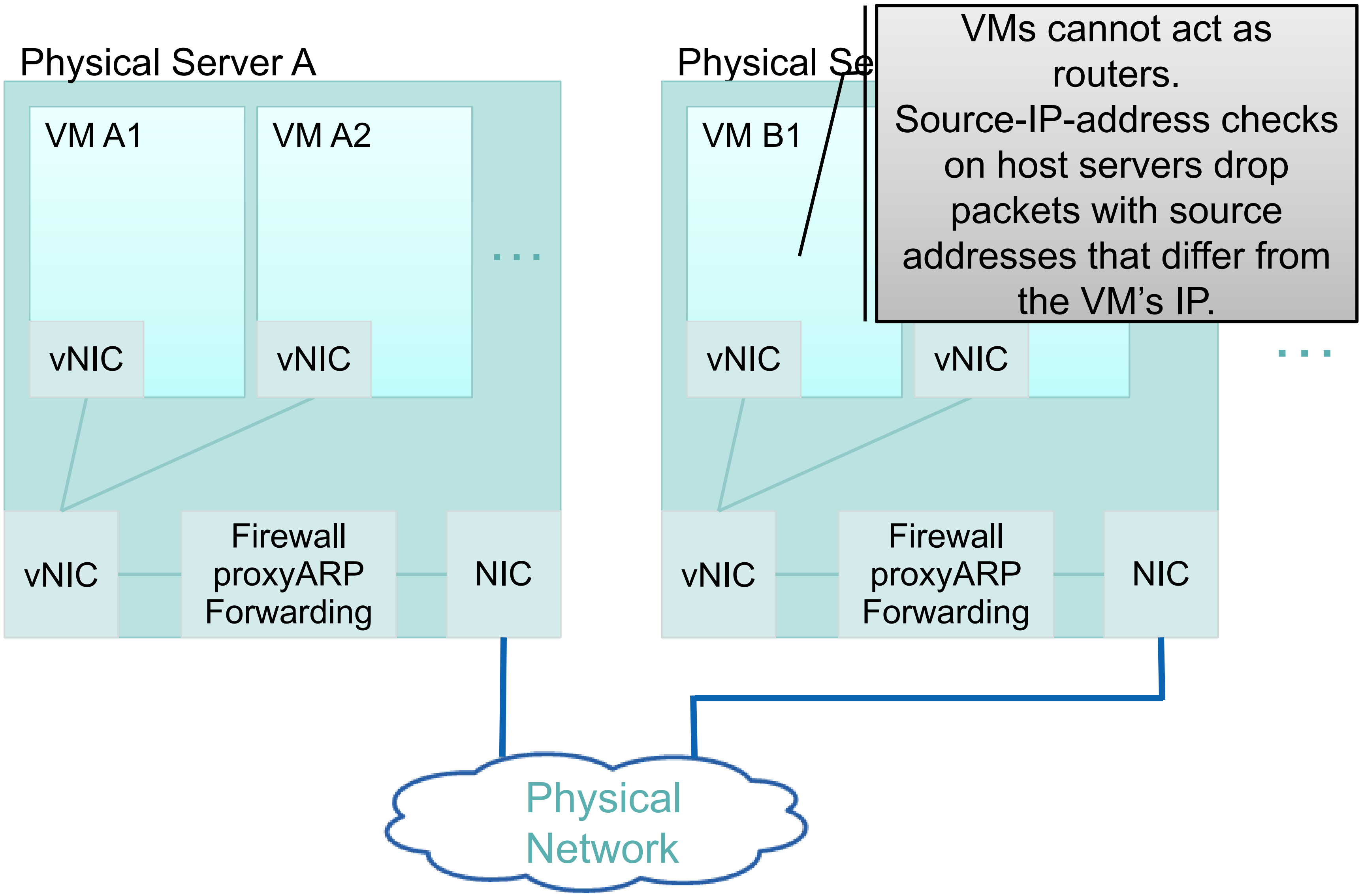
Typical IaaS Network



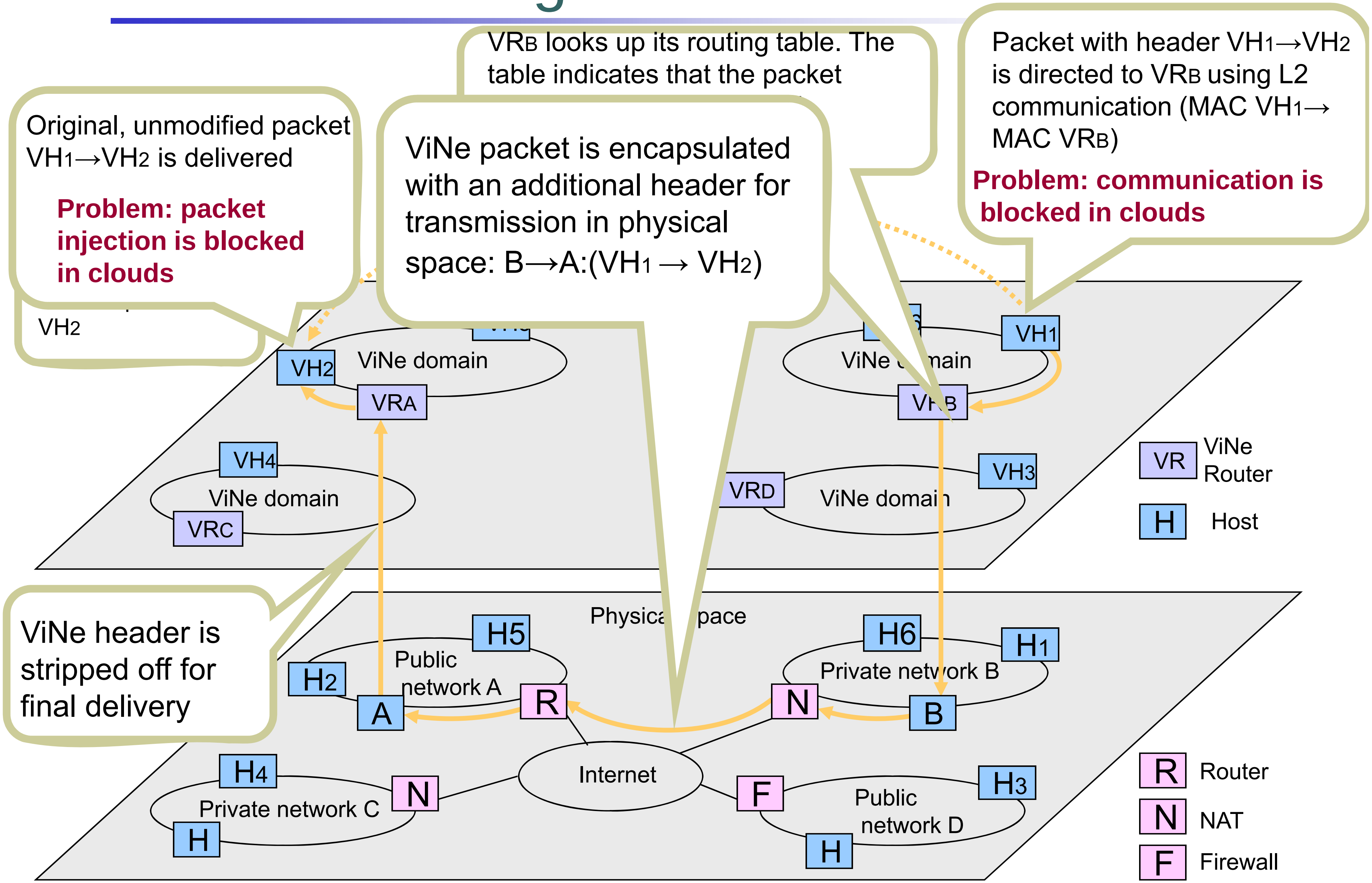
Typical IaaS Network



Typical IaaS Network



ViNe Routing

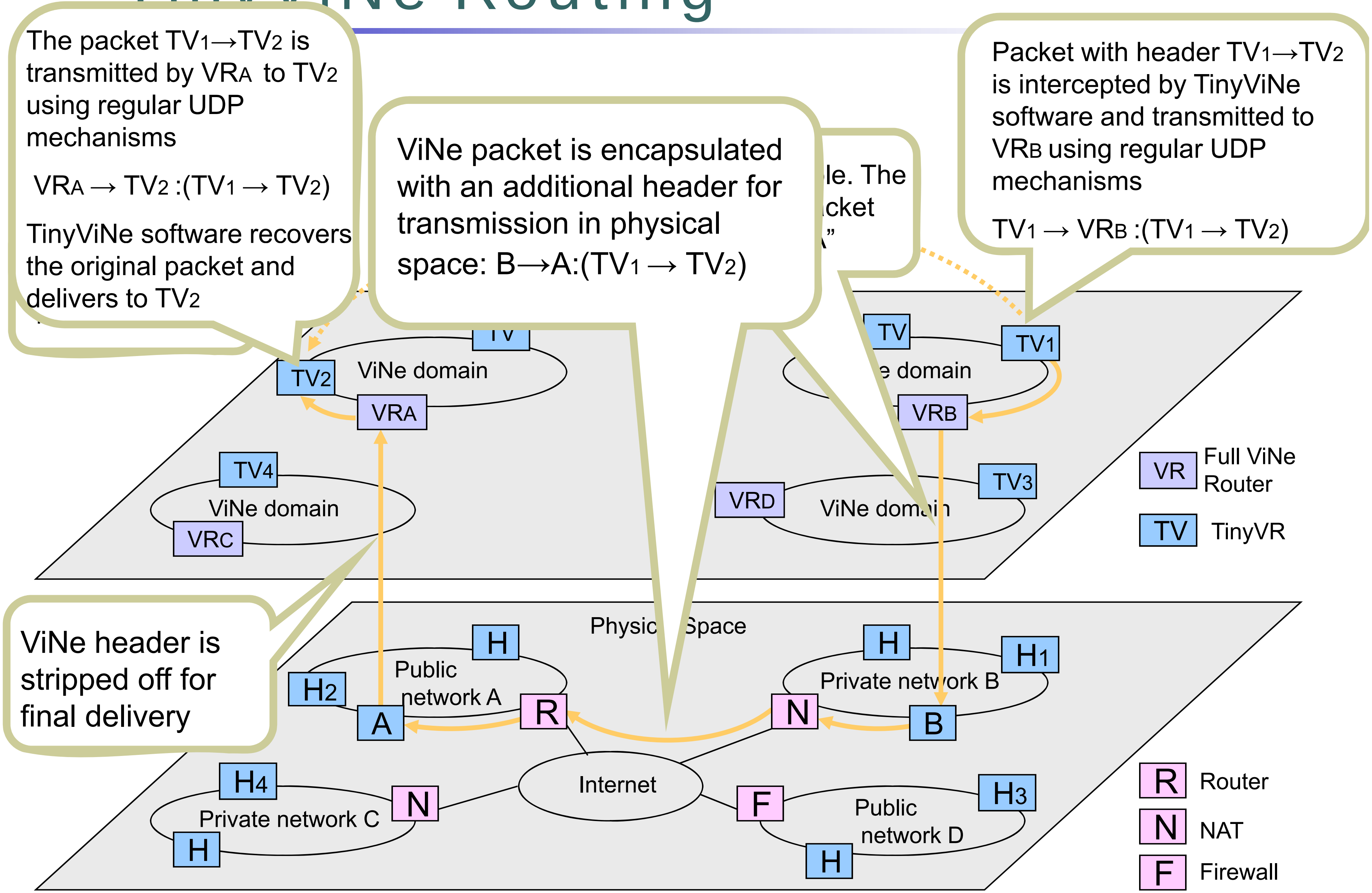


Solution

- Configure all nodes to work as VRs
  - No need for host-to-VR L2 communication
  - TCP or UDP based VR-to-VR communication circumvents the source address check restriction
- But…
  - Network virtualization software required in all nodes
  - Network virtualization overhead in inter- and intra-site communication
  - Complex configuration and operation
- TinyViNe
  - No need to implement complex network processing – leave it to specialized resources (i.e., full-VRs)
  - Keep it simple, lightweight, tiny
  - Use IP addresses as assigned by providers
  - Make it easy for end users to deploy

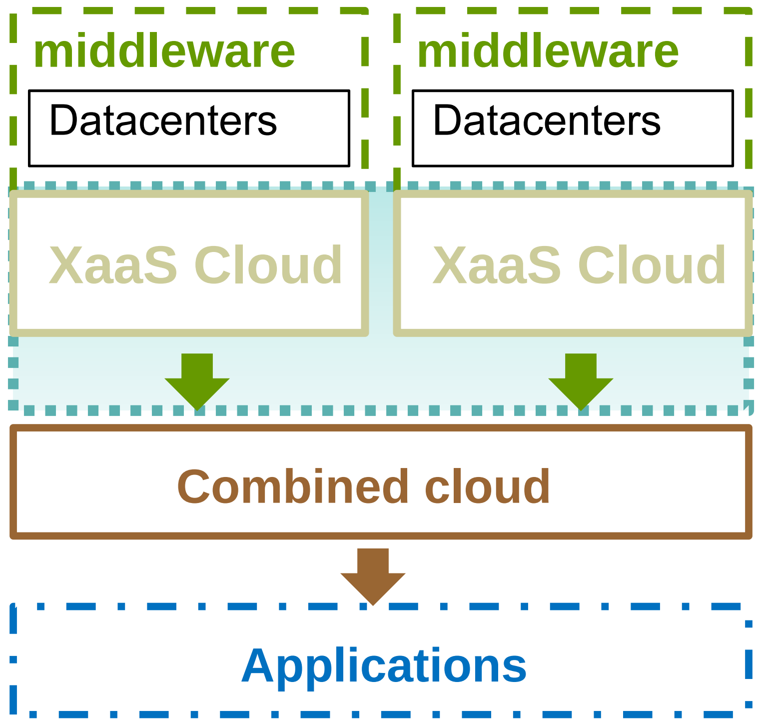
M. Tsugawa* et al.“User-level Virtual Networks Support for Sky Computing”, e-Science, 12/09.

TinyViNe Routing

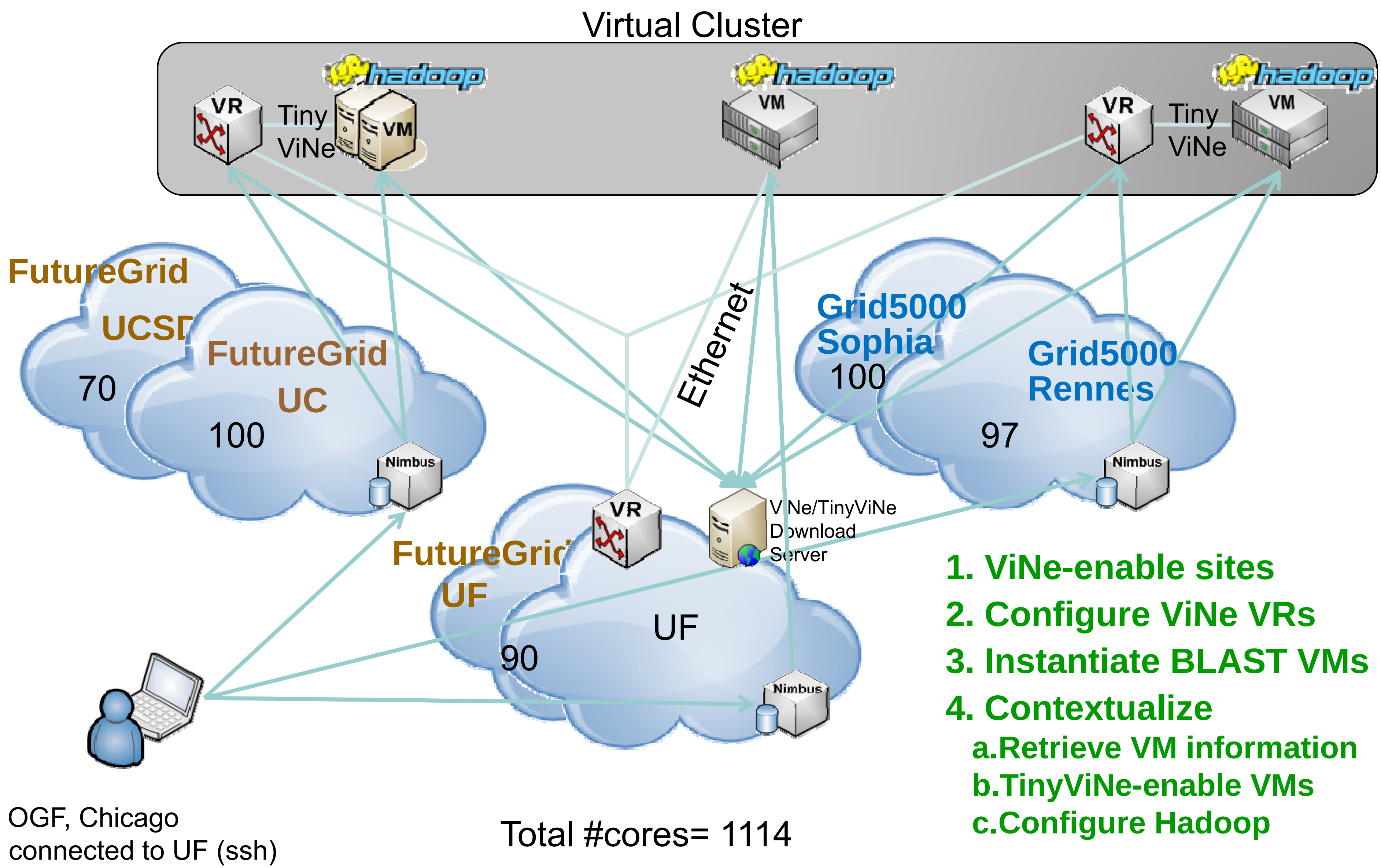


Cloud-combining software

- TinyViNe, only a possible piece of what is needed

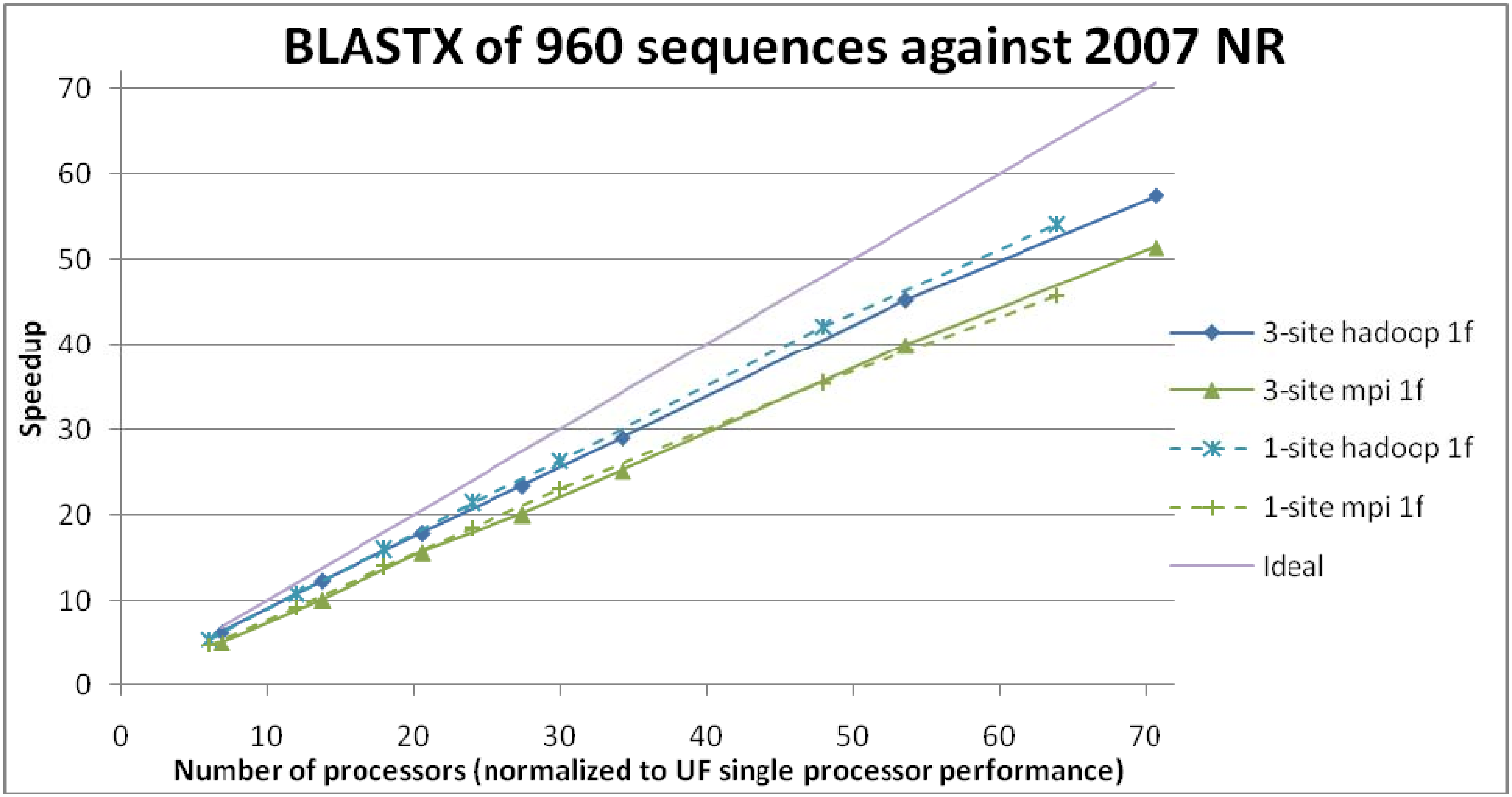


Sky Computing



1. ViNe-enable sites
2. Configure ViNe VRs
3. Instantiate BLAST VMs
4. Contextualize
   a. Retrieve VM information
   b. TinyViNe-enable VMs
   c. Configure Hadoop

3-sites experiment

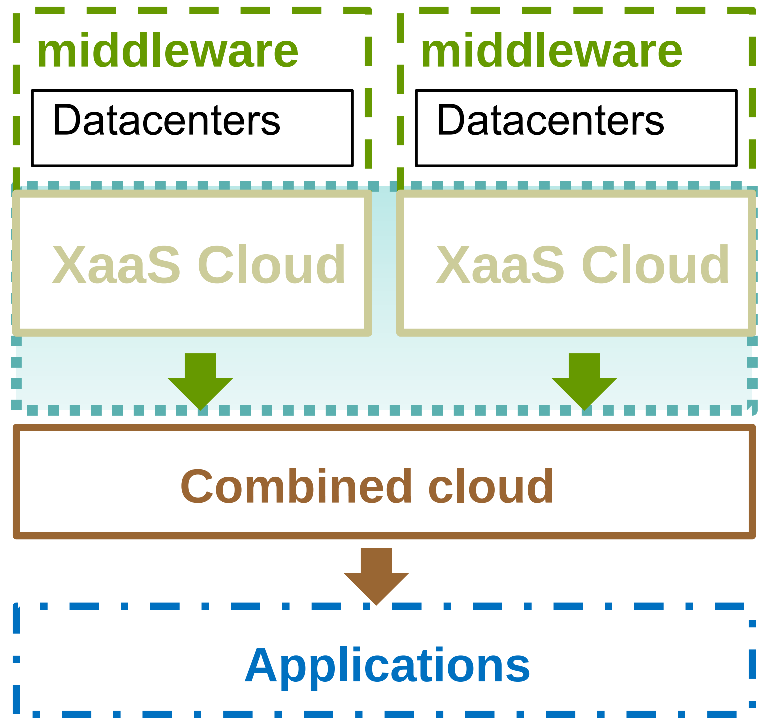


Summary

- User-level overlay networks can be used for inter-cloud communication
- Must handle cloud-specific restrictions
  - Not apparent from abstraction provided as a service
  - Overcome via network-virtualization software in VMs
  - Important to keep the software simple and light
  - Experiments with parallel bioinformatics applications show that it efficiently enables sky computing
  - Implementable as a service by a cloud provider, in the context broker of a sky provider or by the consumer
  - Many opportunities and need for autonomic networking at the provider and/or consumer levels

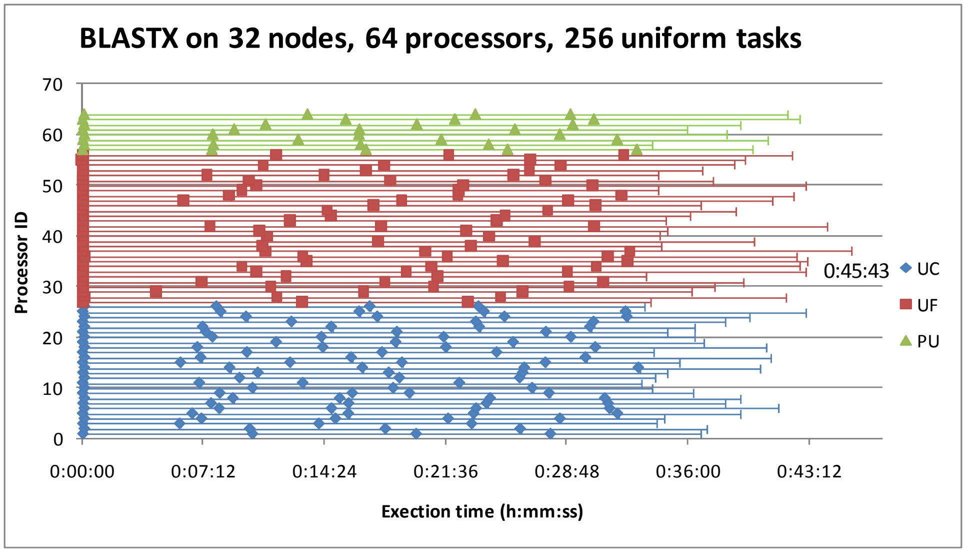
Beyond communication

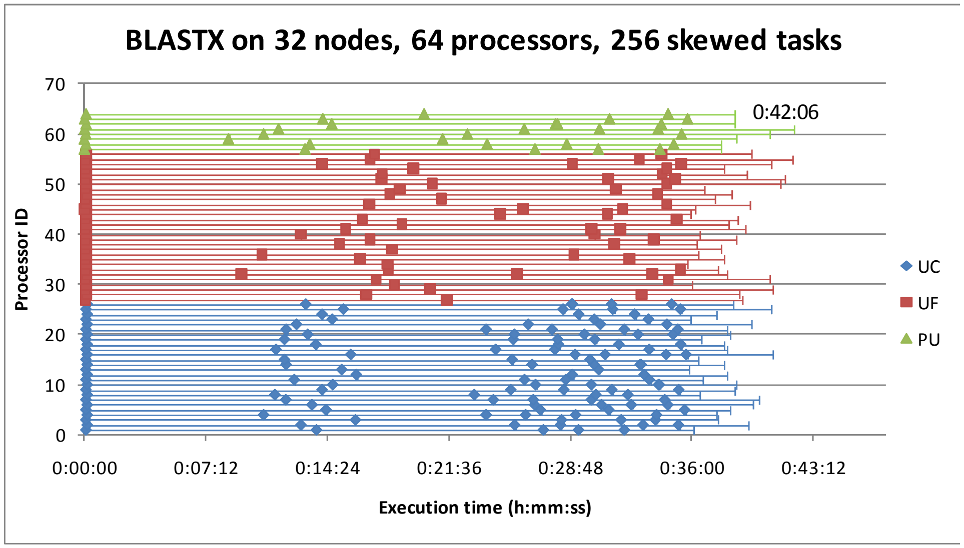
- XaaS = abstractions as a service
  - Which ones to use?
  - SLAs: what is in them and how to support?
  - Affect management of performance, complexity, dependability, …
- Contextualization, coordination and management
- Modeling is essential
- Issues: security, privacy, business models…



Resource usage estimation

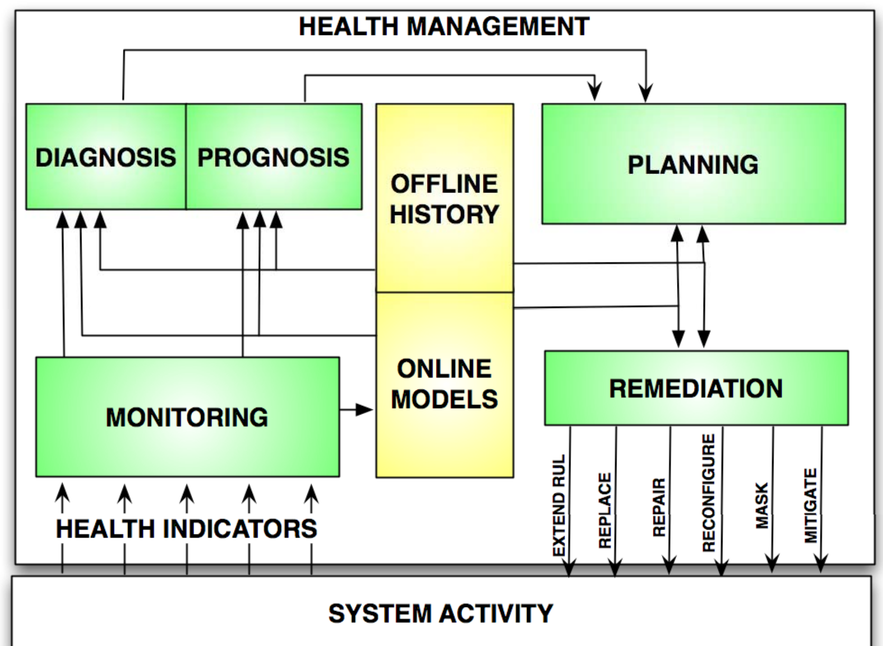
- Provider perspective
  - can improve resource utilization, as schedulers are able to fit more requests in the same resource
- Consumer perspective
  - to choose the most cost-effective cloud and resource configuration for a given problem





Fault modeling

- If components are cloud services, what is a component fault?
  - SLA violation? User-defined condition? Unusual behavior?
  - E.g. resource-exhaustion faults
- How can the health of a sky system/app be managed?
- What/how are concerns separated?
  - E.g. virtual routers
  - "A distributed system is one in which the failure of a computer you didn't even know existed can render your own computer unusable"
    Leslie Lamport
- Similar issues for security, privacy, data, performance…
- Autonomic solutions desirable



Autonomic computing in clouds?

- Decoupling implies loss of cross-abstraction information
  - Providers know little about consumer applications
  - Consumers know little about provider's infrastructure
- Scale and dynamics of clouds create management pressures - alleviated by
  - Commonalities across cloud components
  - Virtualization of application environments
- → autonomic computing needed …

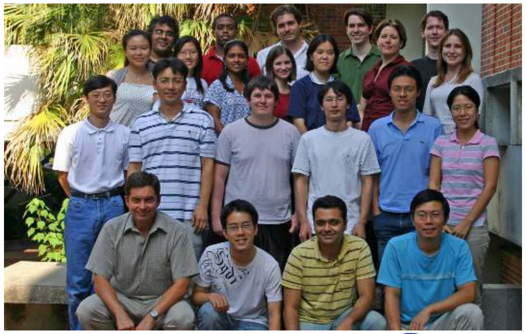
  and feasible!

Conclusions

- Clouds provide the components for novel IT systems or implementations of familiar IT systems
  - Sky-computing refers to such systems and their use
  - In particular, combined clouds capable of providing environments, workflows, enterprise IT, etc “as a service”
- Design and management of combined clouds
  - An area for fundamental and system-oriented IT research
  - Will impact standards and next-generation IT businesses
- Relevance to HPC
  - For high-throughput and pleasingly parallel apps
  - For data-centric computation
  - Models/standards may be leveraged

Acknowledgments

- Sponsors
  - National Science Foundation
  - BellSouth Foundation
  - Center for Autonomic Computing (UF site)
    - Citrix, IBM, Intel, Microsoft, Northrop-Grumman
- Collaborators
  - Andrea Matsunaga and Mauricio Tsugawa
  - Kate Keahey, Tim Freeman
  - Renato Figueiredo and others at ACIS/CAC
  - NSF FutureGrid team



Center for Autonomic Computing (nsfcac.org)

NSF
UF UNIVERSITY of FLORIDA
THE UNIVERSITY OF ARIZONA
RUTGERS THE STATE UNIVERSITY OF NEW JERSEY

Founding industry members

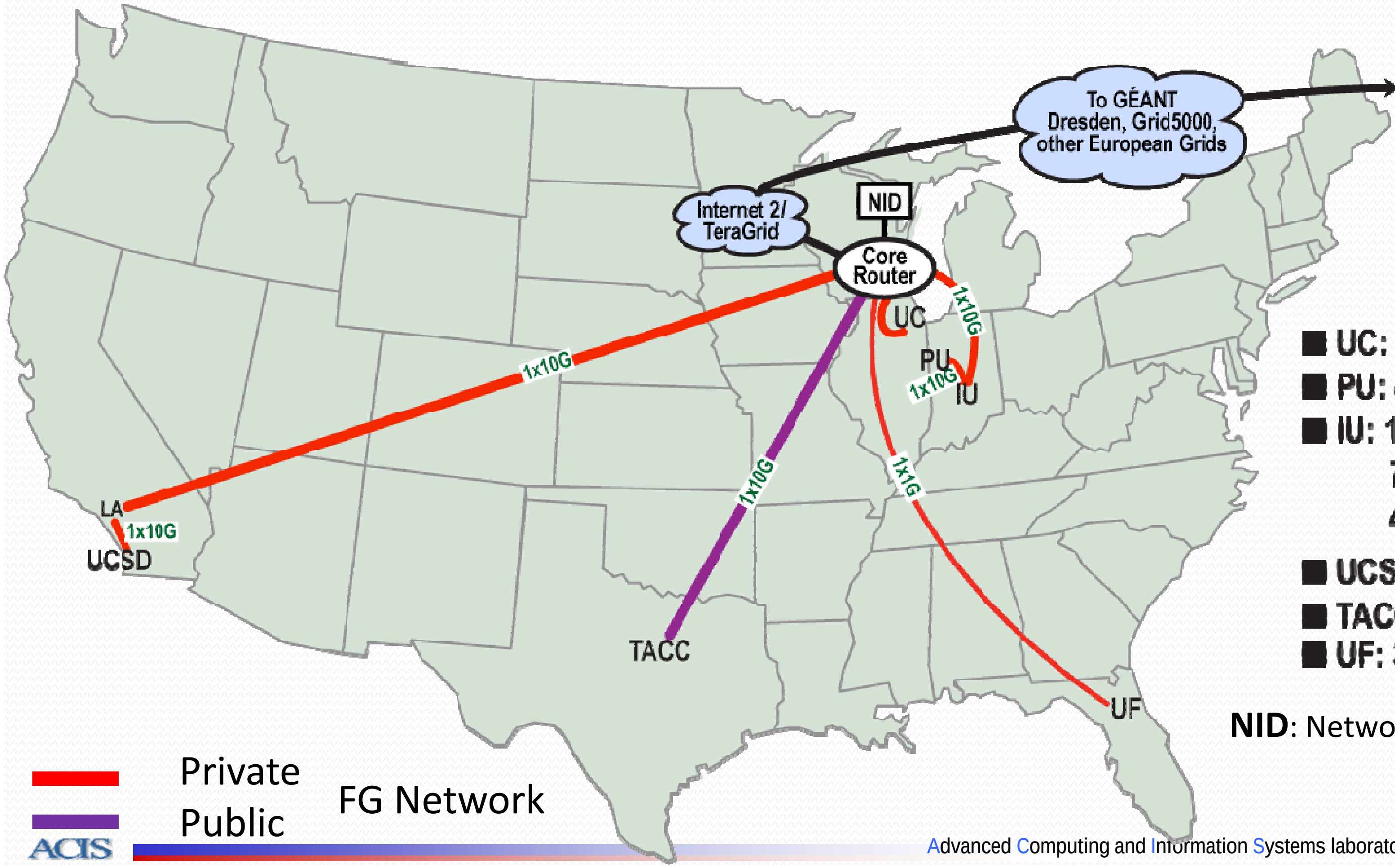
BAE SYSTEMS
CITRIX
EWA Systems
IBM
intel Leap ahead
imaginestics
ISCA
Merrill Lynch
Microsoft
NORTHROP GRUMMAN DEFINING THE FUTURE
NEC
Raytheon
xerox

- collaborative partnership amongst industry, academe, and government;
- concepts, technologies and resources for industry-relevant autonomic computing research;
- interdisciplinary education on autonomic computing;

- **Industry and government agencies invited to join as members**

FutureGrid: a Grid Testbed

- **IU** Cray operational, **IU** IBM (iDataPlex) completed stability test May 6
- **UCSD** IBM operational, **UF** IBM stability test completes ~ May 12
- **Network**, **NID** and **PU** HTC system operational
- **UC** IBM stability test completes ~ May 27; **TACC** Dell awaiting delivery of components



- UC: 7TF IBM 672 cores
- PU: 4TF Dell 384 cores
- IU: 11TF IBM 1024 cores
  7TF Cray 684 cores
  4TF Shared Memory
- UCSD: 7TF IBM 672 cores
- TACC: 12TF Dell 1152 cores
- UF: 3TF IBM 256 cores

NID: Network Impairment Device

Private
Public
FG Network

FutureGrid

- The goal of FutureGrid is to support the research on the future of distributed, grid, and cloud computing.
- FutureGrid will build a robustly managed simulation environment or testbed to support the development and early use in science of new technologies at all levels of the software stack: from networking to middleware to scientific applications.
- FutureGrid is a (small 5600 core) Science/Computer Science Cloud but it is more accurately a virtual machine based simulation environment

futuregrid.org

About | Future Grid

http://futuregrid.org/about

AfreshWiki Clock Customize Links weather TeleCenter my schedule Webster my page ACIS Google Groups » Other bookmarks

About Register Help & Support

Links

- About
  - News
  - Events
  - Sponsors
  - Contact
  - Hardware
    - Status
- Register
- Help & Support

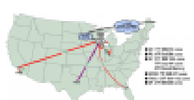
Internal Pages

Book navigation

- Education and outreach

Home

About



This project provides a capability that makes it possible for researchers to tackle complex research challenges in computer science related to the use and security of grids and clouds. These include topics ranging from authentication, authorization, scheduling, virtualization, middleware design, interface design and cybersecurity, to the optimization of grid-enabled and cloud-enabled computational schemes for researchers in astronomy, chemistry, biology, engineering, atmospheric science and epidemiology. The project team will provide a significant new experimental computing grid and cloud test-bed, named FutureGrid, to the research community, together with user support for third-party researchers conducting experiments on FutureGrid.

The test-bed will make it possible for researchers to conduct experiments by submitting an experiment plan that is then executed via a sophisticated workflow engine, preserving the provenance and state information necessary to allow reproducibility.

The test-bed includes a geographically distributed set of heterogeneous computing systems, a data management system that will hold both metadata and a growing library of software images, and a dedicated network allowing isolatable, secure experiments. The test-bed will support virtual machine-based environments, as well as native operating systems for experiments aimed at minimizing overhead and maximizing performance. The project partners will integrate existing open-source software packages to create an easy-to-use software environment that supports the instantiation, execution and recording of grid and cloud computing experiments.

One of the goals of the project is to understand the behavior and utility of cloud computing approaches. Researchers will be able

FutureGrid Partners

- Indiana University (Architecture, core software, Support)
- Purdue University (HTC Hardware)
- San Diego Supercomputer Center at University of California San Diego (INCA, Monitoring)
- University of Chicago/Argonne National Labs (Nimbus)
- University of Florida (ViNE, Education and Outreach)
- University of Southern California Information Sciences Institute (Pegasus to manage experiments)
- University of Tennessee Knoxville (Benchmarking)
- University of Texas at Austin/Texas Advanced Computing Center (Portal)
- University of Virginia (OGF, Advisory Board and allocation)
- Center for Information Services and GWT-TUD from Technische Universtität Dresden. (VAMPIR)