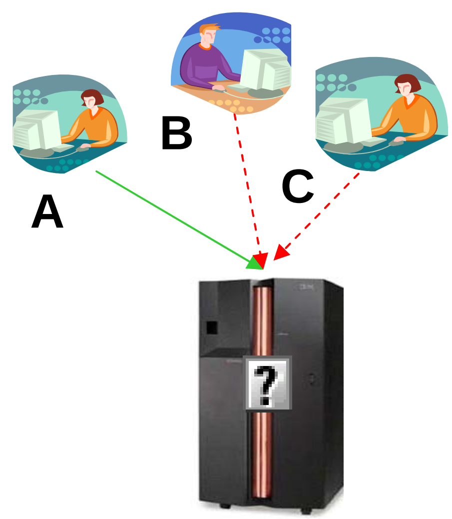
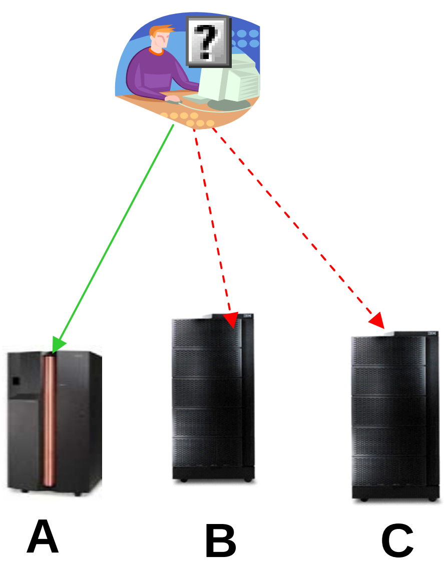
Virtualization Technology in Distributed Computing

José Fortes
Advanced Computing and Information Systems Lab
and
NSF Industry-University Center for Autonomic Computing
University of Florida

Virtualization technology

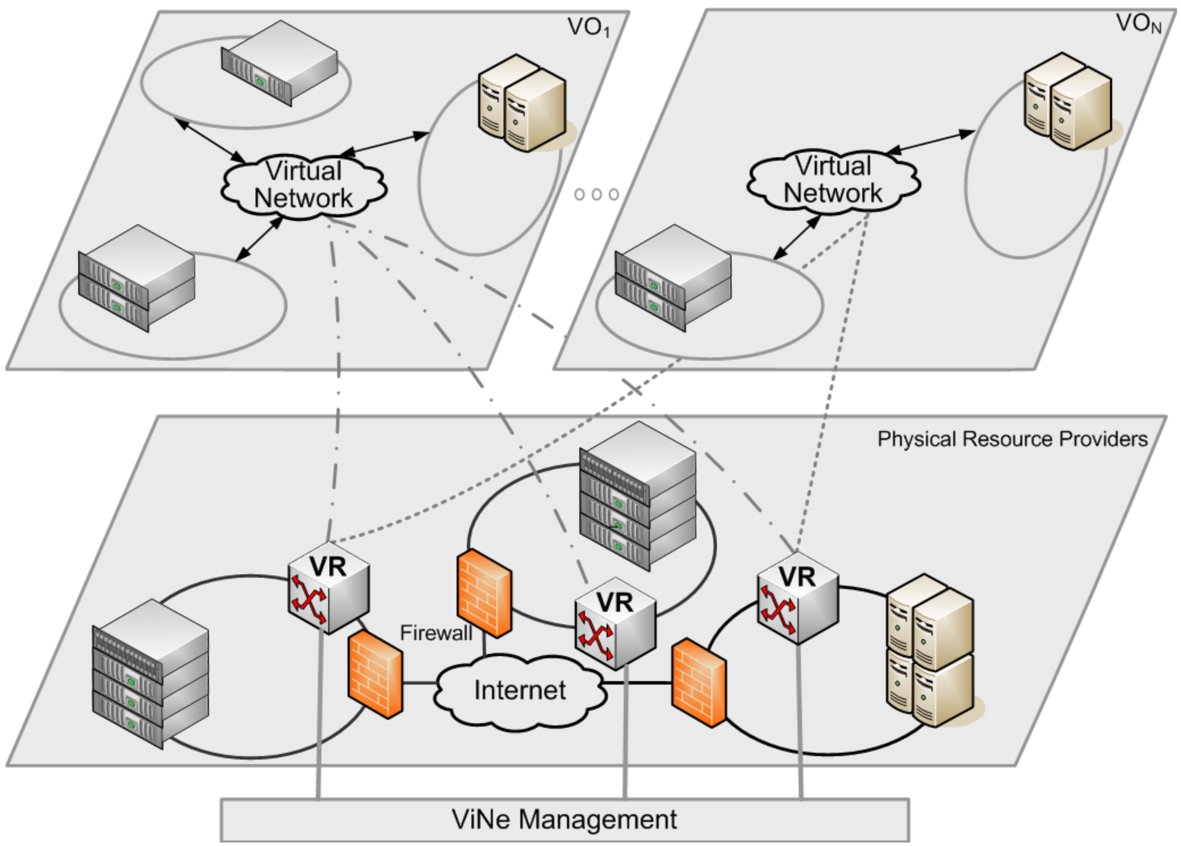
- Why is it good?
  - In one word: decoupling
  - Enables separation of concerns among IT layers
  - E.g. resource layer delivers virtual resources, application layer configures resources to deliver services, provider layer configures services to serve users …





… in distributed computing

- Why is distributed computing “bad” (i.e. hard)?
  - In one word: coupling…
  - …distinct domains is hard to do and manage unless
  - virtually connected virtual resources are used

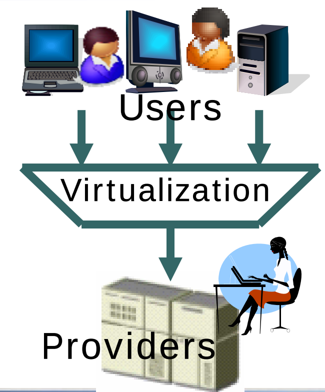


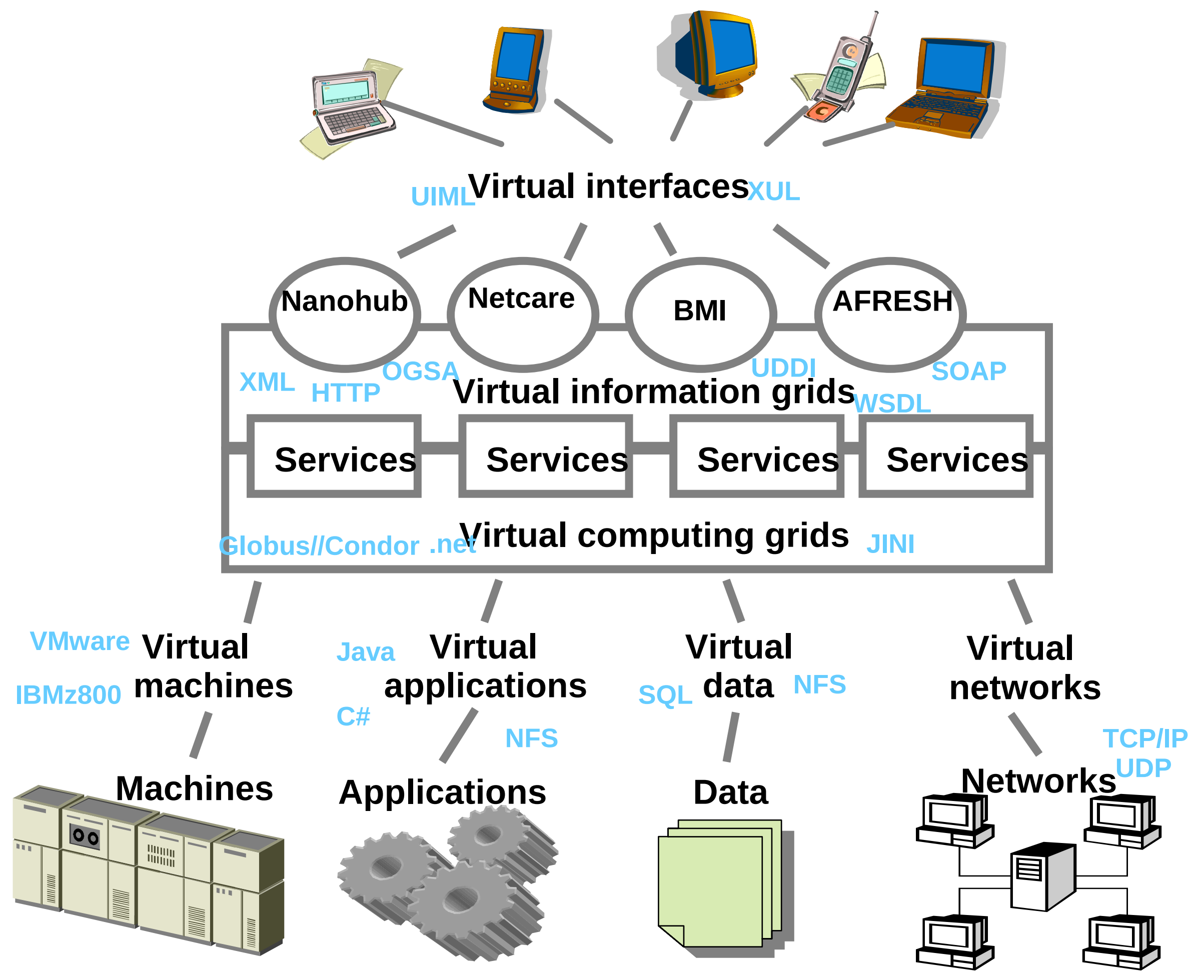
Outline

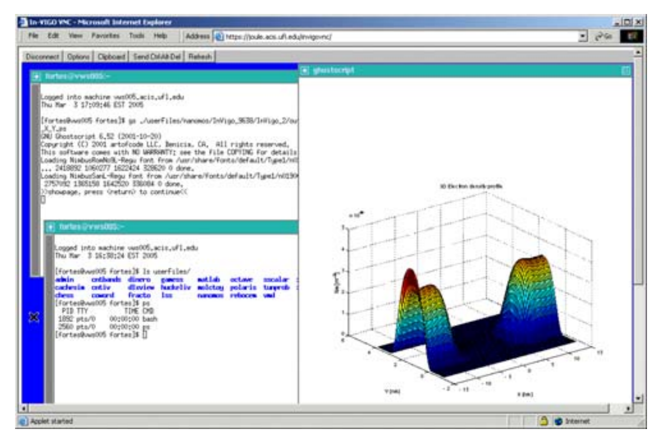
- Sample of work VT + DC
  - Virtualized grids
  - VTDC 2007
- Management: next frontier in VT + DC
  - Why
  - Where
  - How
    - Standards
    - Autonomics (self-management)
      - Examples
- Conclusions

Virtualization and grid computing

- Decoupled application environments
  - From resources and each other; security from isolation
  - Introduces new entities to manage in new ways



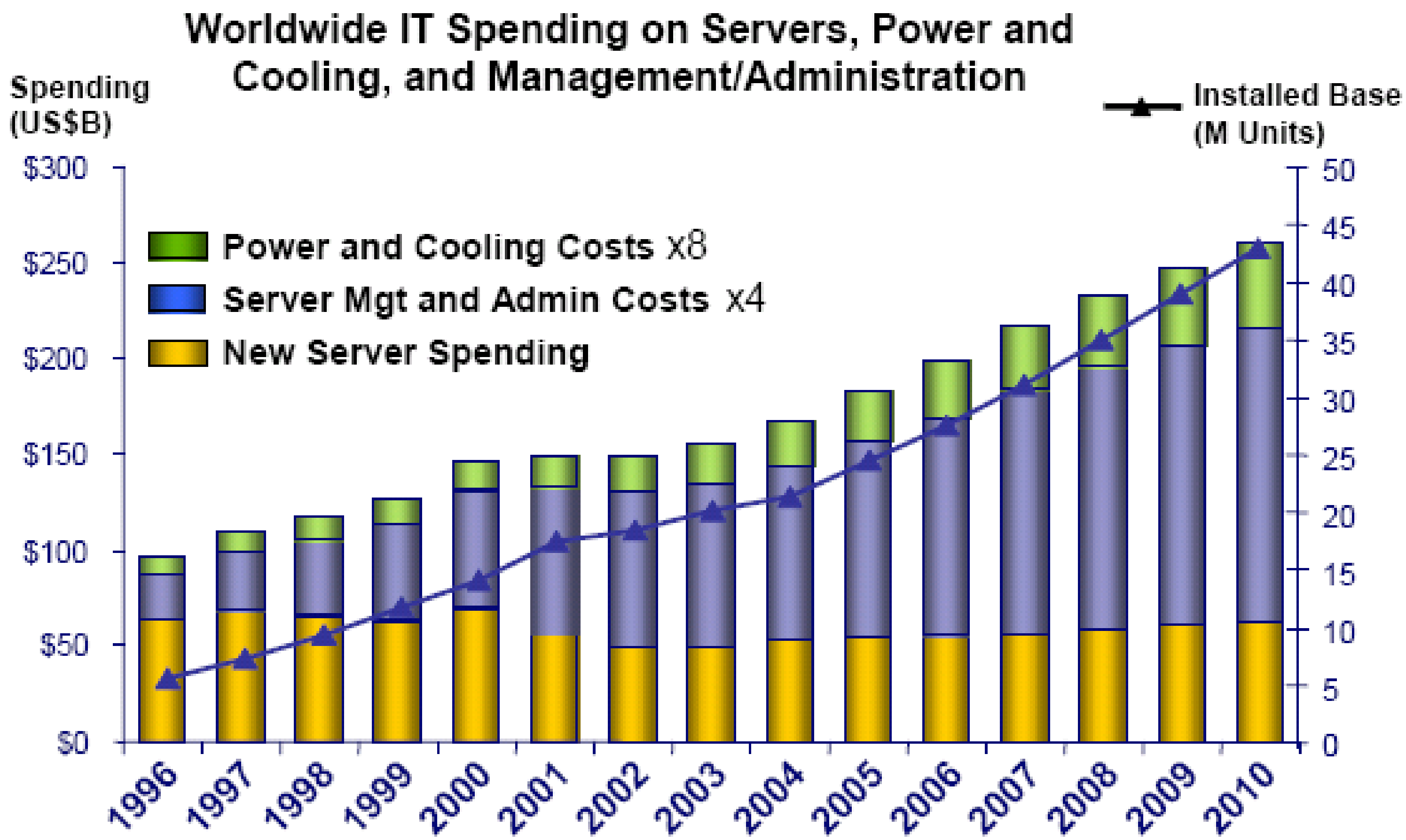




VTDC 2007 – a sample of ongoing work

- Nuts and bolts
  - Enabling Semantic Communications for Virtual Machines via iConnect
  - MemX: Supporting Large Memory Workloads in Xen Virtual Machines
  - Safe Device Driver Model Based on Kernel-mode JVM
- Tools
  - Taking Snapshots of Virtual Networked Environments
  - Experimental Study of Virtual Machine Migration in Support of Reservation of Cluster Resources
  - Model-Based Resource Selection for Efficient Virtual Cluster Deployment
  - Scheduling Virtual Grids: the Magrathea System
  - The Efficacy of Live Virtual Machine Migrations over the Internet
  - Agility in Virtualized Utility Computing

Why management matters?



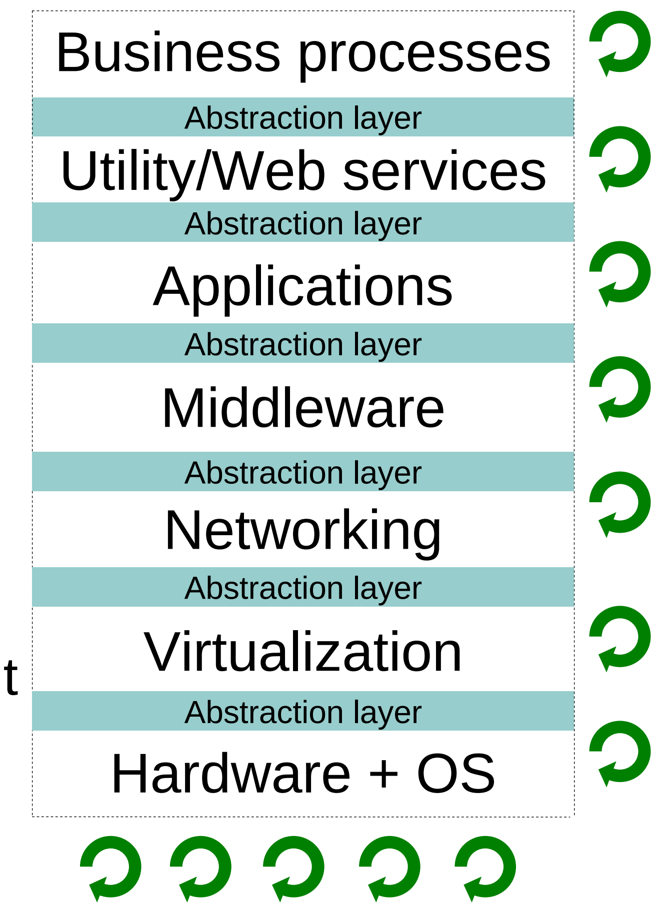
Source: IDC, Virtualization 2.0: the next step in costumer adoption, DOC#204904, Dec. 2006.

- Virtualization technologies enable new capabilities and introduce new entities
  - Both need to be managed

IT management research challenges

- Inter-layer autonomics
  - protocols and APIs, global manager design
- Intra-layer autonomics
  - AC component, local manager design, cooperation/coordination theory
- Business processes
  - process automation, policies, SLAs
- Web-services
  - IT service management, trust, composition
- Applications
  - AC design, modeling
- Middleware
  - monitoring, interoperation, synchronization
- Networking
  - (re)configuration, overlay/mobility management
- Virtualization
  - image management, modeling, configuration
- Hardware + OS
  - provisioning, power/energy minimization

Holistic approach across components and layers



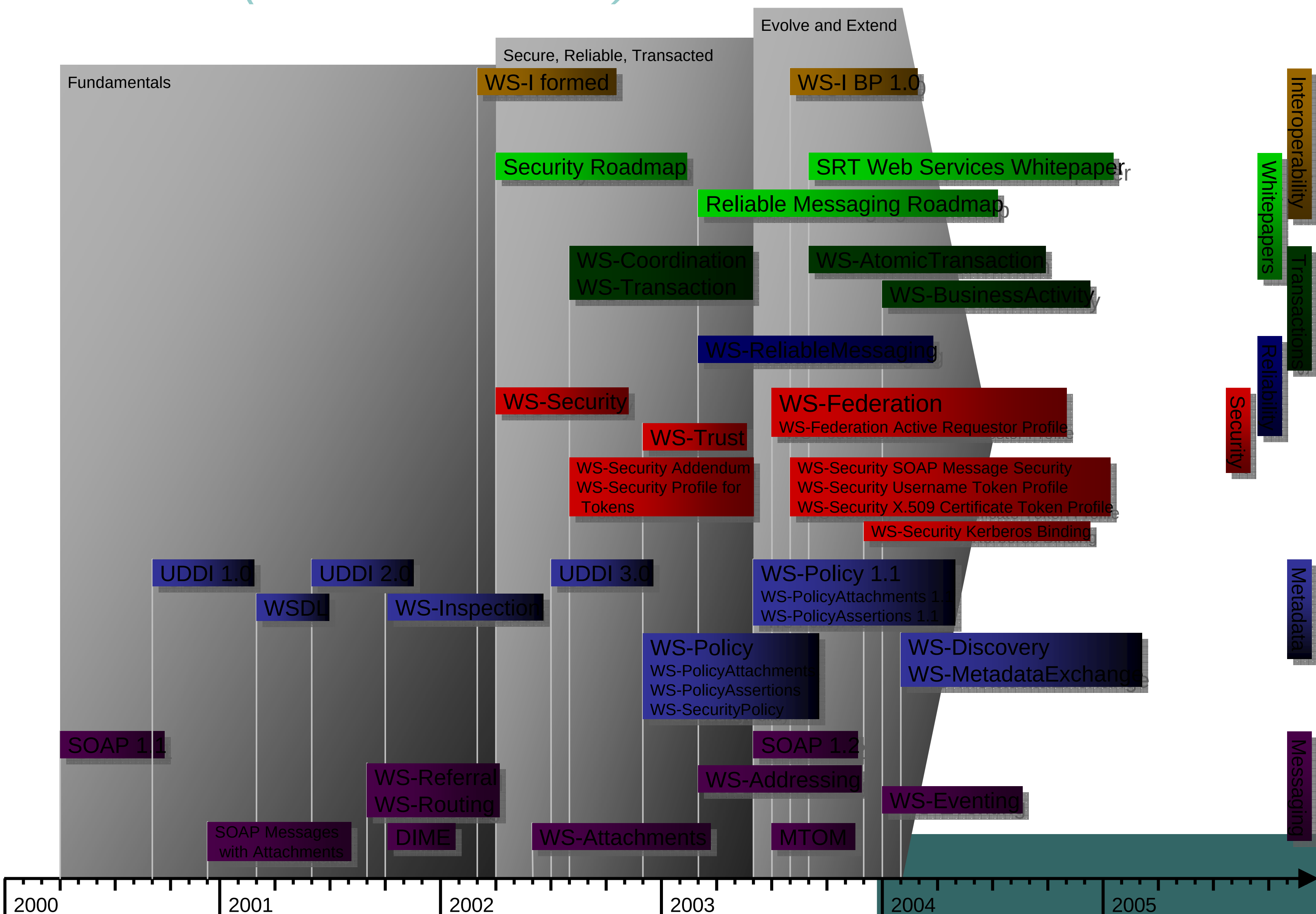
Self-tuning / healing / protecting / managing / *

From Robert Marcus presentation

Alternative	Shared Resource	Resource Users	Enabling Layer	ISSUES
Software as a Service	Application accessed as Web Services	Web clients paying per use	Multi-tenant architectures	Security across multiple uses
Utility Computing (On Demand, Cloud)	Distributed data center software and hardware	Multiple clients renting resources	Distributed resource and workload managers	Accounting, Resource management
Computational Grids	Computers across locations or organizations	Multiple groups sharing computing resources	Grid middleware	Cross-organization management
Transaction Grids (Fabrics)	Hardware/software within organization	Enterprise applications	Fabric middleware	Lack of standards
Data Grids	Data sources across locations or organizations	Multiple groups creating a shared data capability	Data source resource broker	Maintenance of metadata and data consistency lities
Storage Grids or Utilities	Storage hardware	Multiple applications and databases	Storage broker	Performance
Application Virtualization	Execution environment	Processes	Run-time support	
Virtual Server	OS and CPU	Applications running in multiple partitions	Virtual server support	
Virtual Machine Monitor	CPU	Multiple OS running on CPU	Hypervisor possibly with CPU support	
Virtual Appliance	CPU	Bundled application, OS, database	App. virtualization Packaging + run-time	

Web Services Architecture

Timeline (AS OF 2/2004)



www.nsfcac.org

CAC Center for Autonomic Computing

Home | Agenda | Information for Attendees | Presentations and Posters | Contact CAC

Home
- Goal
- Technical Scope
- Benefits of a Center Member
- Personnel
- Administrative Structure
- Facilities
- Funding

Call for Participation and Announcement of Intent to Establish the

National Science Foundation Center for Autonomic Computing

The University of Florida, The University of Arizona and Rutgers, The State University of New Jersey

Director: Dr. Jose Fortes, 352-392-9265, fortes@ufl.edu

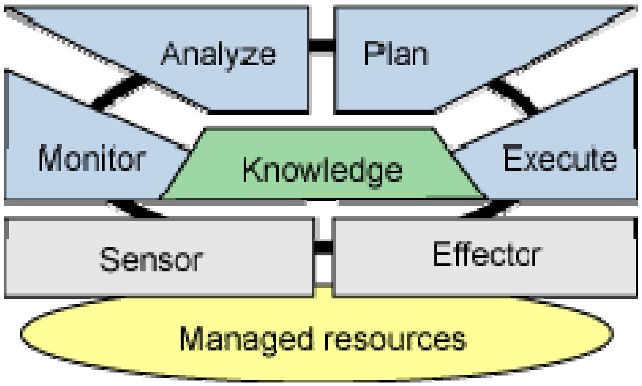
Co-director: Dr. Salim Hariri, 520-621-4378, hariri@ece.arizona.edu

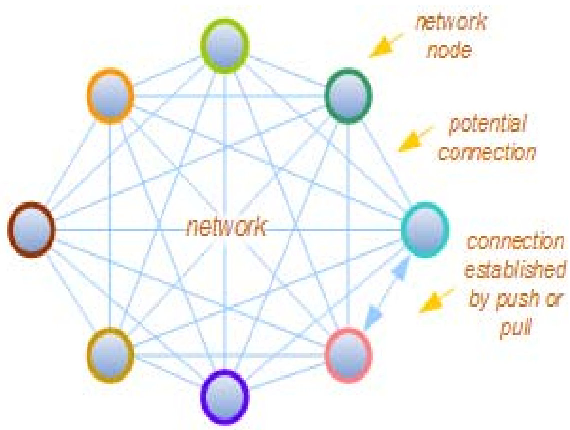
Dr. Manish Parashar, 732-445-5388, parashar@caip.rutgers.edu

The University of Florida, the University of Arizona and Rutgers (the State University of New Jersey) are in the process of establishing a national research center for autonomic computing (CAC). This center will be funded by the Industry/University Cooperative Research Centers program of the National Science Foundation, CAC members from industry and government, and university matching funds. See the LATEST VERSION OF THE AGENDA FOR THE PLANNING WORKSHOP here.

NSF IUC Center for Autonomic Computing (CAC)

- Industry/University Collaborative
  - U. Florida, U. Arizona and Rutgers U.
    - U. Florida leads center
    - Strong emphasis on virtualization
  - Funded by NSF, industry and universities
  - IBM, Northrop-Grumman, Merrill-Lynch, Intel, Microsoft, Motorola, Xerox, NEC, Avirtech, ISCA Technologies, EWA Government Systems, BAE Systems, Imaginestics, Raytheon …(Citrix, Nortel)
  - Annual reviews by NSF with industry input
  - Starting operations in January 2008
    - Kickoff meeting in February 2008
    - Additional founding memberships sought

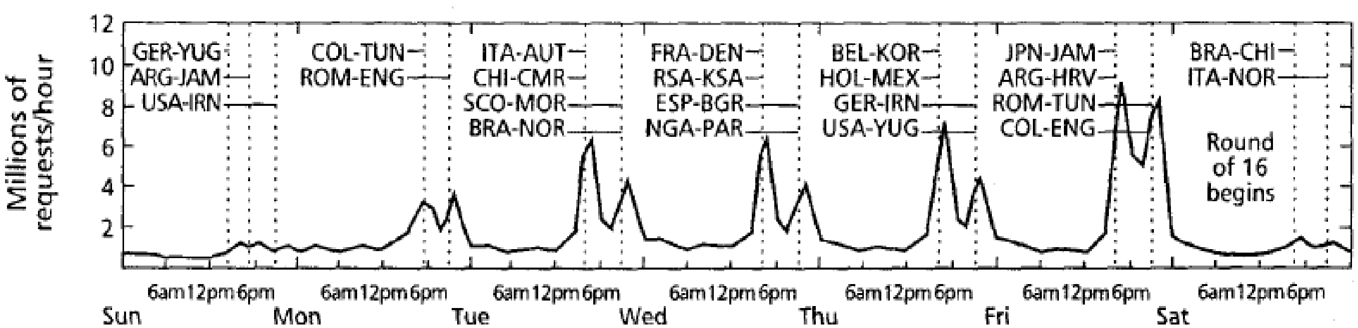




http://www.nsfcac.org

E.g. virtualized data center management

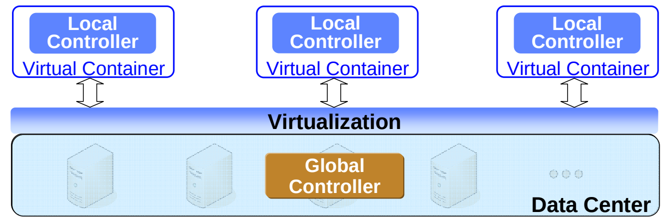
- Models, techniques and mechanisms to
  - monitor, model and predict workload associated with individual services
  - model and predict global and per-service resource demand
  - dynamically allocate/de-allocate virtual machines to physical machines
- Methods based on control theory or market-based approaches
  - to minimize the cost of providing individual services while minimizing power consumption and delivering contracted service levels
  - to develop and evaluate software that implements the methods
- Motivation
  - power savings by avoiding overprovisioning of resources
  - Quality-of-Service by avoiding underprovisioning
  - to maximize profit
- Challenge: hard-to-predict data center workloads
  - concurrent providers, multiple services, many users

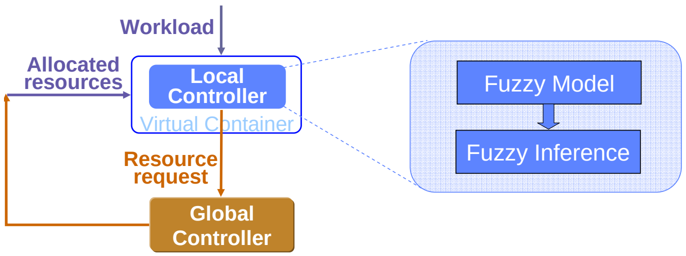


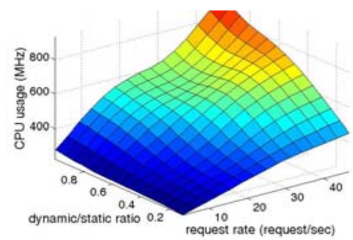
Source: A workload characterization study of the 1998 World Cup Web site, Arlitt, M. Jin, T. HP Labs.

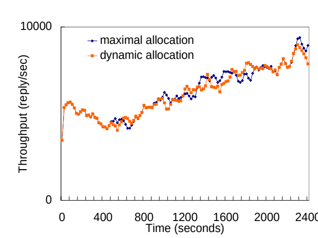
Background and Related Research

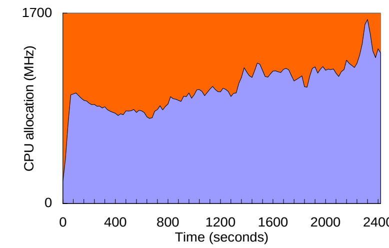
- Virtualized data centers: 2-level control (ICAC'06/07)
  - Local: VM resources for application(s) on VM
  - Global: physical resources for VMs on data center
  - Decoupled management of QoS and resources





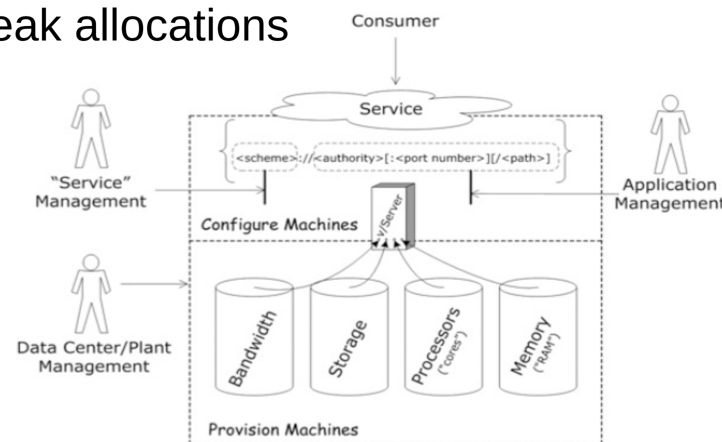






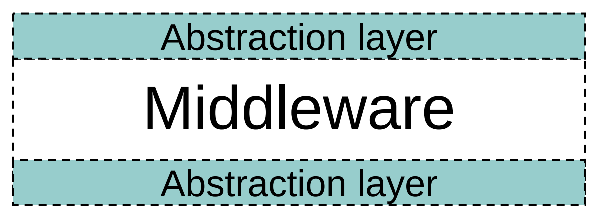
- Fuzzy modeling (ICAC'07 Xu, Zhao, Fortes, Carpenter and Yousif)
  - Resource provisioning per VM closely tracks workload demand
  - Large savings of resources (power and $) over peak allocations

- Service "Routing" with Virtual "Switch" (VTDC'06, Cohen and Schaefer)
  - Protocol that provides enough information in its <service>:// layer address to allow an application's capacity to shrink and grow based on demand
  - Software and hardware to mediate access to resources by an application
  - Market mechanism for resource allocation



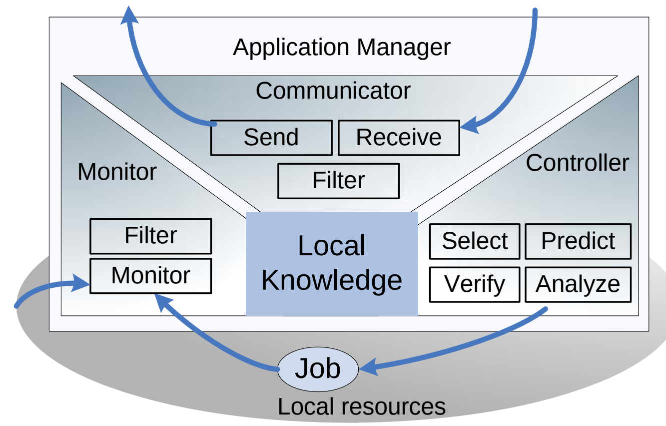
Autonomic layers as networks

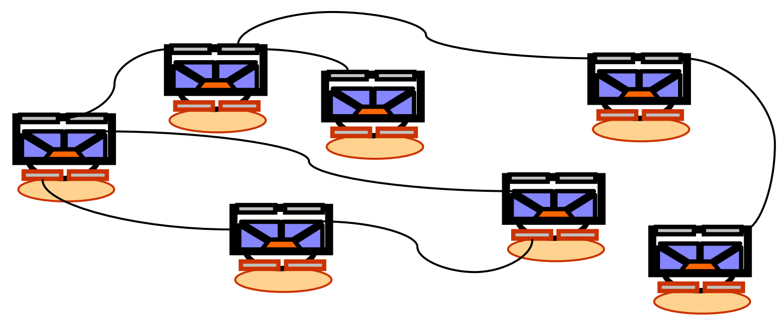
- Connecting autonomic managers
  - Multiple management nodes self-organize into a network and management “load” is distributed across the network
  - Any management node individually manages its local resources according to its local policies
  - Network enables scalable cooperation (every one provides as well as consumes information)
- Goal is to manage large-scale systems where
  - Failure happens commonly
  - System conditions change over time
  - Application QoS is hard to guarantee
  - Global information is not easy to obtain

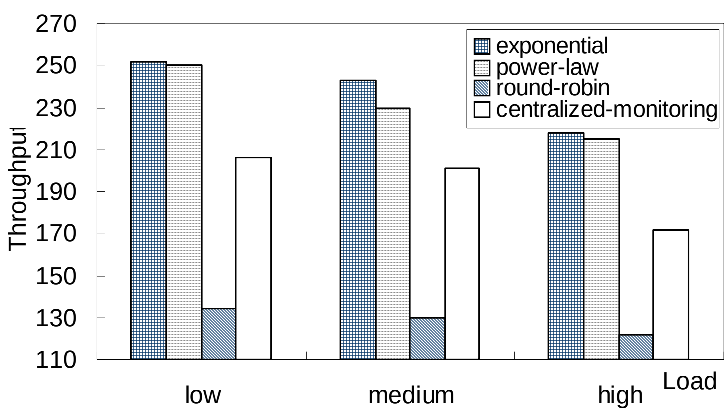


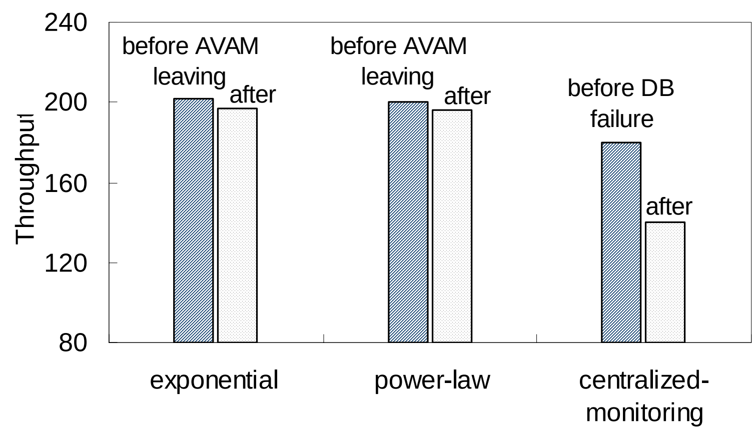
Autonomic layers as networks of autonomic managers (AMs)

- Decentralized management
  - A large number of managed elements and each AM is responsible for managing a small fraction of them
  - Each AM has a local "view" of the system and communicates with a few other AMs (neighbors)
  - AMs self-organize as a P2P management network and access to global information through the network



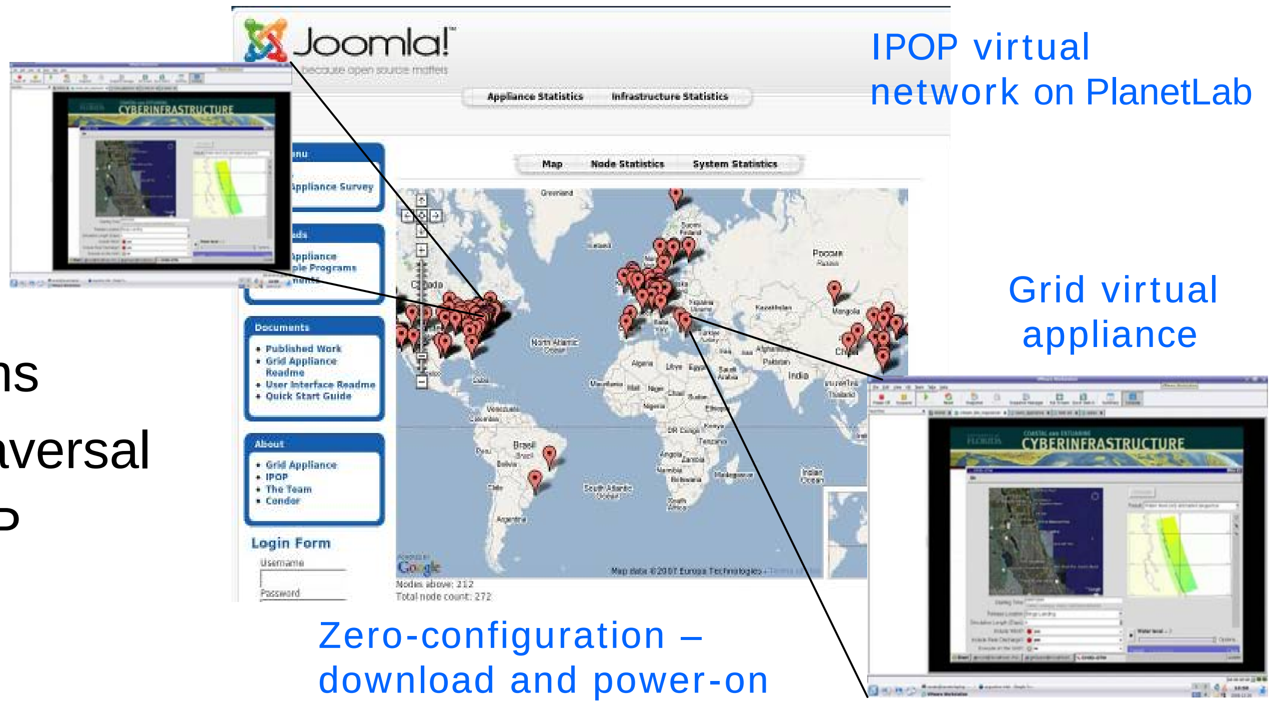






IP over P2P (IPOP) and self-organizing clusters

- Synergies of P2P and virtualization:
  - Scalability, self-management, legacy support, ease of deployment …
- P2P virtual networking – IPOP
  - Structured P2P routing, Distributed Hash Table, DHCP
  - Open-source C# implementation, deployed on 400+ PlanetLab nodes
- Key capabilities
  - Supports existing TCP/IP applications
  - Integrated NAT traversal (STUN) and DHCP



IPOP virtual network on PlanetLab

Grid virtual appliance

Zero-configuration – download and power-on

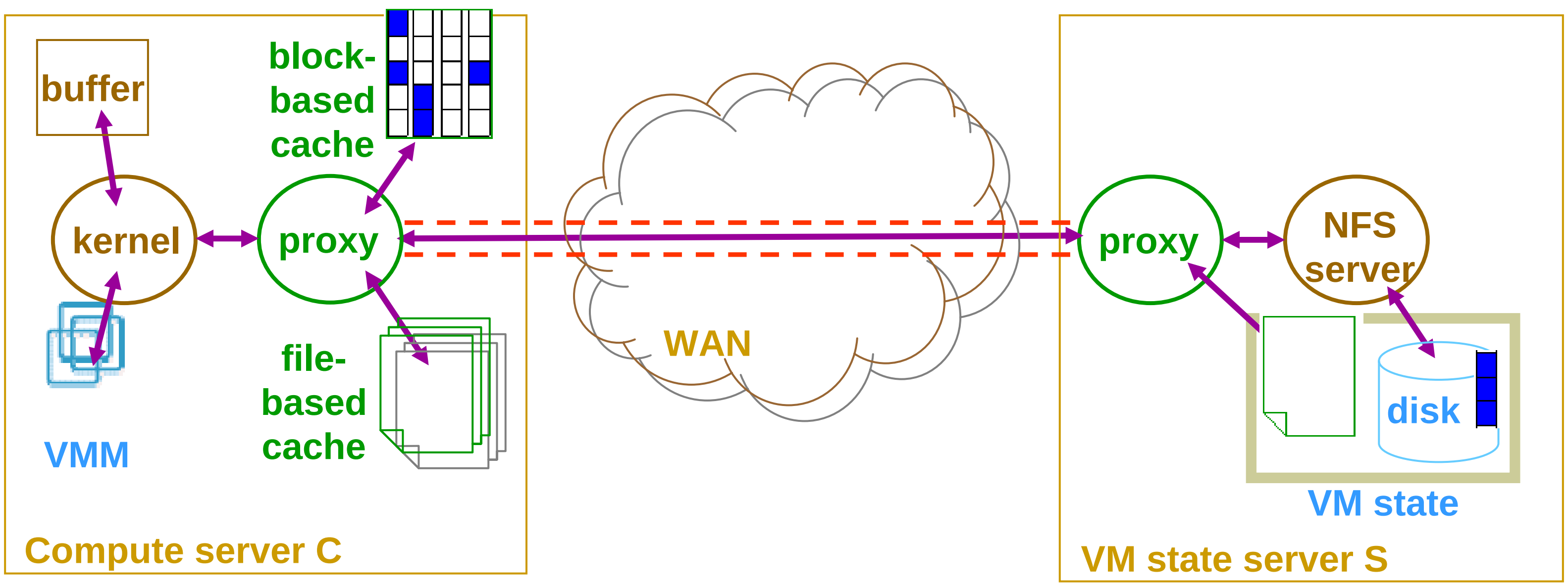
- Applications
  - VMs + Virtual networks = virtual clusters
  - Grid computing, collaborative environments

http://ipop-project.org

PIs: R. Figueiredo, P. Boykin

Grid virtual file systems

- Works with de-facto standard Network File System (NFS)
  - Virtualize NFS calls via user-level proxy
- Application-tailored enhancements at the virtualization layer
  - Aggressive caching, consistency
  - Security policies
  - Redirect-on-write
- Example application: VM image provisioning

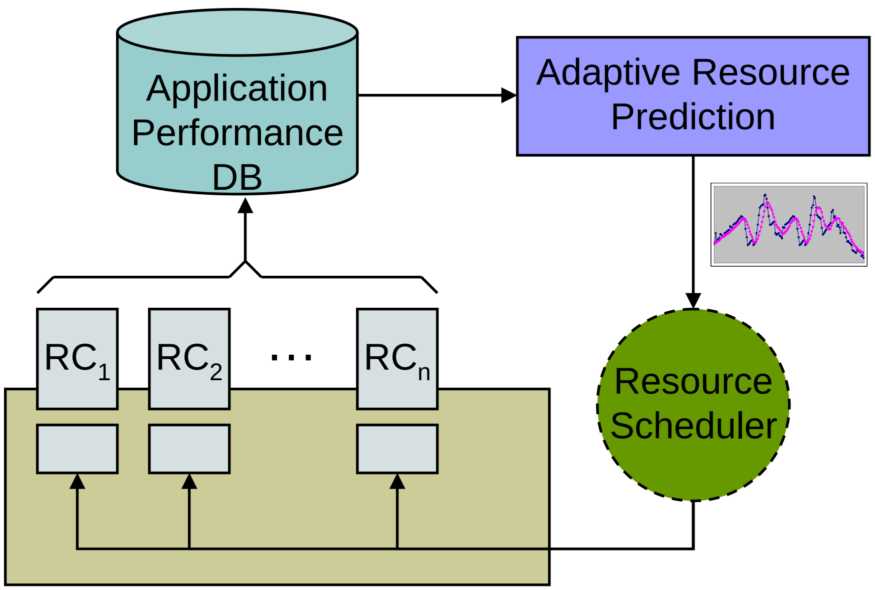


http://www.acis.ufl.edu/~ming/gvfs

PI: R. Figueiredo

Learning-aided resource management

- Dynamic resource provisioning
- Monitoring of system-level metrics
  - CPU, disk and network bandwidth
  - VM container



RC: Resource Container (VM)
DB: Database

PI: R. Figueiredo

- Dimension reduction, classification
  - PCA, k-NN clustering
- Learning-aided time-series prediction
  - Predict best model out of a pool using classification techniques
- Phase analysis and prediction

Other ongoing projects

- Virtualized data centers
  - Autonomic Resource Management in Virtualized Data Centers
  - Characterizing and improving the performance of virtualized I/O workloads in data centers
  - Learning-Aided System Performance Prediction and Workload Characterization
  - Autonomic Application Classification and Feature Selection
  - Autonomic Power Management for Grid Computing
  - Autonomic Computing Engines for High Utility Scalable Grids and Data Centers
- Middleware and cyberinfrastructure
  - BACIS: Bio-informatics at ACIS
  - Atomic-scale Friction Research and Education Synergy Hub (AFRESH)
  - Integration of Applications into SOA for Transnational Digital Government
  - Adaptive middleware management for Dynamic Data-Driven Brain-Machine Interface (DDDBMI)
  - Biosphere 2 – tackling grand challenges in the earth sciences
- Peer-to-peer computing
  - Brunet: A structured P2P system for NATed wide-area environments
  - Grid Appliances: simplifying Grid computing with self-organizing wide-area virtual clusters
  - Decentralized Scheduling in Self-organizing Networks of Virtual Machines
  - Decentralized Clustering Analysis for Online Anomaly Detection and Trust Management in Self-Monitoring P2P Systems
- Networking and file systems
  - Managed virtual networks in Grids – the ViNe approach
  - IPOP: A Self-Organizing Virtual Network for Wide-Area Environments
  - Provisioning of Virtual Environment in a Wide Area Network through Redirect-on-write Distributed File System
  - Autonomic Grid-wide Data Management with User-level Distributed File System Virtualization and Services
  - Autonomic Control and Management Environment - Autonomia
  - Autonomic Network Defense (AND) System
  - Autonomic Data Streaming and In-transit Processing
- Reliability and power
  - Hypervisor-based Byzantine Fault Tolerance in Many-core Computing Platforms
  - Mitigating microarchitecture soft error vulnerability in multithreaded execution environment
  - FPGA Based Fault Detection, Isolation and Healing for Autonomous Operation
  - Dynamic Power Cut-off Technology (DPCT) for Leakage and Switching Power Savings in Hardware

Conclusions

- Pervasiveness of virtualization technologies
  - Across IT stacks
  - Across types of resources
  - Across manufacturers
- Extensive research on
  - More and better technologies
  - Rethinking of distributed computing
  - Tools to use, deploy and manage them
- Virtualized IT infrastructure management
  - Big challenge
  - Standards play an important role (interoperability)
  - Research needed to address increase in scale, complexity, virtualization specifics
  - Self-management and integration desirable and possible
    - Focus of new NSF center for autonomic computing

Acknowledgements

- NSF
- ARO, NASA, NOAA, ONR
- Intel, IBM, Northrop-Grumman, SRC, VMware
- BellSouth Foundation