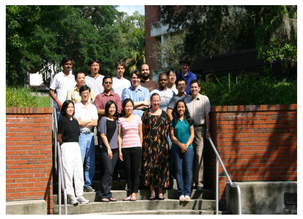


ACIS/CAC Research Activities

The mission of the Advanced Computing and Information Systems Laboratory (ACIS) is to conduct fundamental and applied research on all aspects of systems that integrate computing and information processing. Current ACIS research falls under the broad categories of Cloud/Grid-computing middleware, Cyberinfrastructure for e-science, Autonomic computing and Peer-to-peer computing

Faculty: José A.B. Fortes,Renato Figueiredo, Andy Li, Tao Li, Maurício Tsugawa, Andréa Matsunaga

Students: Prapaporn Rattanatamrong, Selvi Kadirvel, Tae Seung Kang, Meng Han, Giljae Lee, Joseph Makkar, Pierre St. Juste, Jiangyan Xu, Kyungyong Lee, Yonggang Liu, Heungsik Eom

Center for Autonomic Computing

2008-2013 (IIP-0758596)
http://nsfcac.org

Industry-academia research consortium funded by the NSF, industry member fees and university funds;

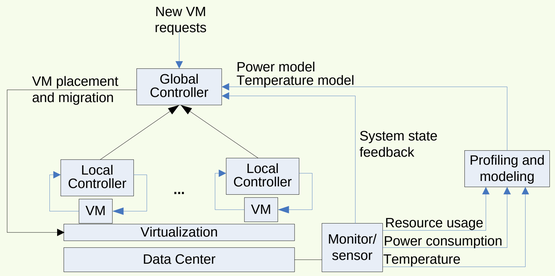
- Conducts cooperative research on autonomic IT
- U. Florida, U. Arizona, Rutgers U., Mississipi St. U.
- Industry members including Raytheon, Intel, Xerox, Citrix, Microsoft, ERDC, Avirtec, Mobilabs.

Main ongoing technical thrusts in IT systems:

- Performance, power and cooling management
- Autonomic computing in virtualized systems
- Application modeling for policy-driven management
- Self-protection and system-health management
- Virtual networking
- Self-managed peer-to-peer private networking
- Cloud and grid computing
- Cross-layer performance tuning in datacenters
- Web-based collaborative systems

- PIs: José Fortes, Renato Figueiredo, Manish Parashar, Salim Hariri, Sherif Abdelwahed and Ioana Banicescu

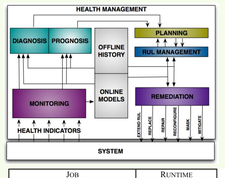
Datacenter Resource Management



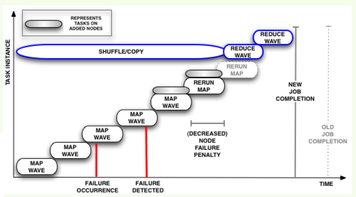
- Controllers predict + provision virtual resources for applications
- Multiobjective optimization (30% faster with 20% less power)
- Use fuzzy logic, genetic algorithms and optimization methods
- Use cross-layer information to manage virtualized resources to minimize power, avoid hot spots and improve resource utilization

- L. Wang, J. Xu, M. Zhao, Y. Tu, and J. Fortes, "Fuzzy Modeling based Resource Management for Virtualized Database Systems,", MASCOTS, 2011.
- Jing Xu and José Fortes, A multi-objective approach to virtual machine management in datacenters. 8th Int. Conf. on Autonomic Computing (ICAC '11).
- J. Xu, J. Fortes, "Multi-Objective Virtual Machine Placement in Virtualized Data Center Environments," Int. Conf. on Green Computing and Communications, 2010.

Self-Caring IT Systems



	Job	Runtime
Single Node Failure	Built-in Fault Tolerance	7.8% (Penalty)
	Scaling: 2 nodes	0.9% (Penalty)*
	Scaling: 4 nodes	6.8% (Gain)
	Scaling: 8 nodes	14.6% (Gain)
Two Node Failures	Built-in Fault Tolerance	84.0% (Penalty)
	Scaling: 2 nodes	9.1% (Penalty)
	Scaling: 4 nodes	5.2% (Gain)*
	Scaling: 8 nodes	16.4% (Gain)

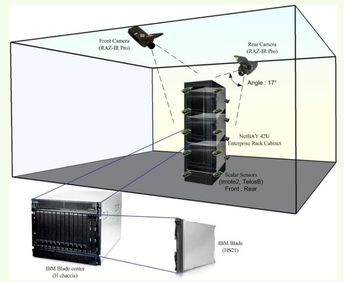


- **Need:** High likelihood of failures due to scale, heterogeneity, distribution, configuration errors, complexity, etc.
- **Objective:** Self-Caring IT systems capable of proactively managing their health in addition to reactively responding to failures.
- **Case Study:** MapReduce on a Cloud environment, mitigate performance (job runtime) penalties under node and task crashes through dynamic resource scaling, migration and rejuvenation.

- S. Kadirvel, J. Fortes, "Towards Self-Caring MapReduce: Proactive Fault Handling for Reducing Execution-Time Penalties", High Performance Computing and Simulation, 2011.
- S. Kadirvel, J. Fortes, "Self-Caring IT Systems - A Proof-of-Concept Implementation in Virtualized Environments", Int. Conf. on Cloud Computing Technology and Science, 2010.

Cross-Layer Autonomics Intercloud Testbed

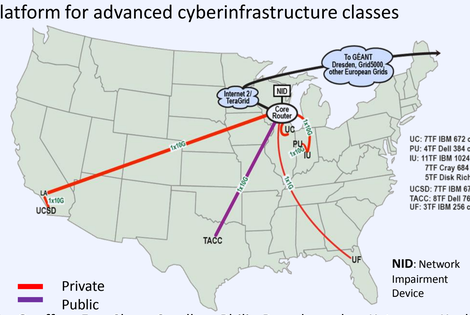
- Distributed testbed at UF, UA and RU
- Framework for cross-layer optimization studies
- Thermal modeling
  - Sensors in different locations
  - Report temperature and humidity
- Short-term/long-term predictions to improve controllers' performance
- Intercloud protocols
- Virtual networking
- Interoperability studies
- Sky computing



FutureGrid

2009-2013 (OCI-0910812)
http://futuregrid.org

- FutureGrid is an international testbed that supports computer science and computational science research in cloud, grid and parallel computing (HPC)
- Provides a flexible development and testing platform for middleware and application users looking at interoperability, functionality, performance or evaluation, and a rich education and teaching platform for advanced cyberinfrastructure classes



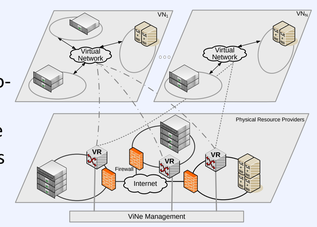
- PIs: Geoffrey Fox, Shava Smallen, Philip Papadopoulos, Katarzyna Keahey, Richard Wolski, José Fortes, Ewa Deelman, Jack Dongarra, Piotr Luszczek, Warren Smith, John Boisseau, and Andrew Grimshaw

University of Florida Roles in FutureGrid

- Key roles in middleware, teaching, education and outreach with emphasis on overlay networks and virtual appliances.
- Lead roles in operations team, and education and outreach team
- UF FutureGrid hardware: IBM iDataPlex – 256 cores, 768GB of RAM, 20TB of storage

Virtual Network (ViNe) project

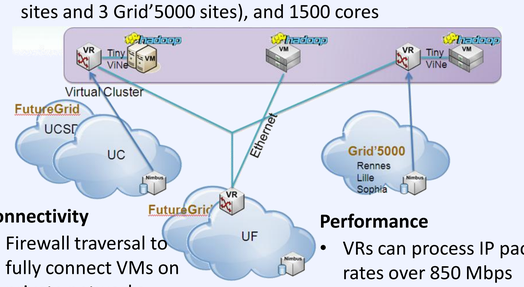
- **Need**: Enable communication among cloud resources overcoming limitations imposed by firewalls
- **Objective**: Develop an easy-to-manage intercloud communication infrastructure
- **Case Study**: Interconnect VMs deployed on different FutureGrid clouds.



- M. Tsugawa; J. Fortes. A Virtual Network (ViNe) Architecture for Grid Computing. In Proc 20th Intl Parallel and Distributed Processing Symposium (IPDPS-2006), Rhodes Island, Greece, p.10, April, 2006.
- M. Tsugawa; J. Fortes. Characterizing user-level network virtualization: performance, overheads and limits. International Journal of Network Management, 2009.

Sky Computing

- **Need**: Large scale scientific computation requires resources that a single cloud provider cannot offer
- **Objective**: Efficiently combine cloud technologies (ViNe, Nimbus, Hadoop, etc) to form an intercloud virtual cluster
- **Case Study**: Execution of CloudBLAST on 750 VMs (across 3 FG sites and 3 Grid'5000 sites), and 1500 cores



Connectivity

- Firewall traversal to fully connect VMs on private networks

Performance

- VRs can process IP packets at rates over 850 Mbps
- CloudBLAST speedup: ~870X

- K. Keahey; M. Tsugawa; A. Matsunaga; J. Fortes. Sky Computing. Internet Computing, IEEE, vol.13, no.5, p.43-51, Sept.-Oct. 2009.
- M. Tsugawa; P. Riteau; A. Matsunaga; J. Fortes. User-level Virtual Networking Mechanisms to Support Virtual Machine Migration Over Multiple Clouds. In IEEE Intl Workshop on MENS, 2010, pp. 568-572.

ACIS Facilities

- State-of-the-art computing, storage and networking facilities
- Unique environment for experimental research and design of distributed systems that use virtualization software developed by commercial and open-source projects
- ~200 servers, ~1250 cores, ~4.8TB of memory, ~260TB storage
  - FutureGrid cluster: IBM iDataPlex connected to Florida Lambda Rail.
  - NUMAcloud: allows up to 64 cores and 512MB of memory in a single image
  - Autonomic Testbed: enables research on cross-layer autonomics for datacenter management
  - VM and cloud: rich set of VMMs (VMware, Xen, KVM, QEMU), and cloud software (Nimbus, OpenStack)
  - Storage: centralized (IBM DS4800) and cloud-based (OpenStack)

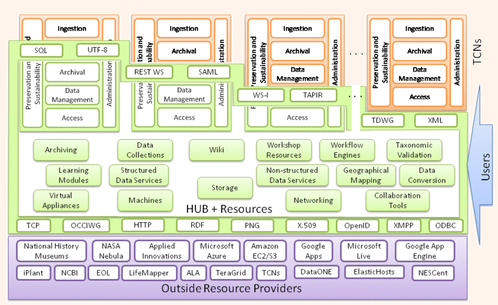
iDigBio

2011-2016 (EF-1115210)
http://idigbio.org

- The Home Uniting Biocollections (HUB) established and funded by the NSF Advancing Digitization of Biological Collections (ADBC)
- ADBC is a national effort to support digitization of all vouchered biological and paleontological collections housed in US institutions
- iDigBio coordinates this challenging undertaking by fostering partnerships, training, and innovations, and serving as a central site for integrating data and techniques, and establishing cohesion and interconnectivity among all the digitization projects, especially with Thematic Collections Networks (TCNs)
- The resource will provide access to information critical to scientific research and education, including that designed to understand biodiversity and societal consequences of climate change and other environmental issues
- iDigBio oversees data integration, support for computational needs and assessment of new technologies and programs to facilitate digitization efforts including directing the development of "appliances" and the development of cloud storage capabilities

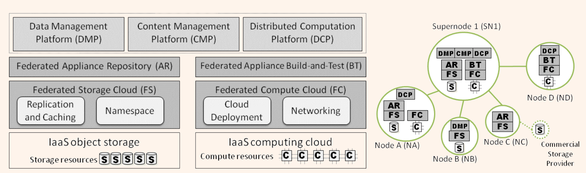
- PIs: Lawrence Page, Jose Fortes, Pamela Soltis, Bruce McFadden, and Gregory Riccardi

Appliance-based Cloud-Oriented Architecture



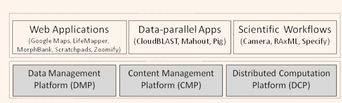
- **Need:** Software appliances and reliance on cloud computing to be adaptable and handle a large set of tools, scenarios and partners
- **Objective:** Creation, dissemination, deployment and sharing of "virtual appliances" encapsulating software environments and tools found in the iDigBio in a "sandbox" environment
- **Case study:** Platform for cyber-learning packages with self-contained executable educational modules tailored by educators and seamlessly deployed by students

Federated Storage Cloud (iDig*)



- **Need:** Distributed storage infrastructure with large capacity and common namespace that is reliable, fast, accessible and affordable
- **Objective:** Design, build and deploy a customized cloud storage environment for collections based on off-the-shelf disks and offering distributed object storage services including distributed federated namespace registry/mapper and bucket/object caching and replication manager
- **Case study:** Support efficiently different platforms with heterogeneous access patterns (e.g., data management platform, content management platform, distributed computation platform, federated appliance repository).

Federated Compute Cloud (iDig*)



- **Need:** Platform to perform operations on large digitized collections housed in US institutions

- **Objective:** Design, build and support Platform-as-a-Service integration modules for domain-specific scientific workflows and data parallel applications (e.g., Camera, CloudBLAST, Condor, Kepler, Pegasus, Pig, Specify)
- **Case study:** Emphasis on platforms that use standard vocabularies for metadata, interoperable APIs for software and data management, and perform operations on geographically distributed data