cac bulletin

The Newsletter of the Center for Autonomic Computing

Vol. 1, Summer 2010

UF UNIVERSITY of FLORIDA

THE UNIVERSITY OF ARIZONA.

RUTGERS

NATIONAL SCIENCE FOUNDATION

The Center for Autonomic Computing is sponsored by the National Science Foundation Office of Industrial Innovation and Partnerships' Industry/University Cooperative Research Center (I/UCRC) Program under award 0758596, and by our industry partners.

intel CITRIX ISCA TECHNOLOGIES xerox Raytheon Microsoft imaginestics AVIRTEC AUTONOMIC MANAGMENT SOLUTIONS Ball Aerospace & Technologies Corp.

In this issue:

Save the date! The next CAC review meeting will be held October 5 & 6 in Gainesville, FL. Read more about the upcoming meeting and about the Center at http://www.nsfcac.org. Interested in becoming an industrial partner? See page 11 of this newsletter for details!

Dear Friends,

I hope you enjoy reading this CAC bulletin in which you will learn of ongoing research and other activities related to autonomic computing at the three CAC sites—the University of Florida, the University of Arizona and Rutgers University. In this bulletin, CAC researchers share their perspectives on timely topics and provide updates on projects, people and events.

Already in its third year of existence, the CAC has made remarkable progress as a three-university center, in spite of difficult economic times which negatively affected our industrial partners and limited their ability to support the center for the last two years. Currently the CAC has fifteen industry and government memberships, has funded partnerships with two other NSF centers and has secured several additional research and instrumentation awards. All together, these activities ensured a strong economical basis on which a large number of students and faculty have been able to build industry-relevant projects whose results have led to a large number of publications and software artifacts. In addition, new laboratory and office spaces have been deployed to host the center equipment, activities and researchers, including visitors from industry.

A recurring question from those who first hear about CAC is “What is autonomic computing?” In this issue you will find technical answers to this question in specific contexts, but I often explain that a system is autonomic if it operates properly without anyone having to manage it or even know that it exists—it just works! While many such systems exist in living organisms, few man-made artifacts can claim to be fully autonomic. Airplanes are a good example of IT-rich systems where autonomic systems have enabled highly autonomous behaviors that make it technically possible for an airplane to safely take off, fly and land without human assistance. In practice, however, human intervention is preserved to allow for unforeseen scenarios and keep a human-centric view of autonomic computing.

A timely example where autonomic computing plays a key role is that of computer clouds. The figure at right shows my view of cloud engineering, cloud architecture and cloud computing. Clouds are engineered by designing and deploying middleware that exports to cloud consumers a set of abstractions of datacenter capabilities. The abstractions are provided as services, e.g., in clouds that provide Infrastructure as a Service (IaaS) the abstractions might consist of virtual machines, storage and network connectivity. When clouds provide Platforms as a Service (PaaS) the abstractions are library APIs and programming environments, whereas Software-as-a-Service (SaaS) clouds deliver end applications and/or high-level domain-specific application development platforms. Just as a microprocessor architecture is captured by its instruction set, a cloud (architecture) is characterized by the set of abstractions it provides as a service. The abbreviation XaaS is used to denote all these types of clouds, where X stands for I, P or S. Cloud computing refers to the creation of cloud applications, newly developed and/or composed of existing applications, built using cloud abstractions as their basic components. These abstractions are expected to “just work”, i.e., cloud middleware must ensure that the functionality expressed by the abstraction is always provided—correctly, dependably, securely and at cost. Autonomic computing is therefore a must, if cloud abstractions are to be used as the basis for other IT systems. This is an area of CAC research—we invite you to join us in these efforts.

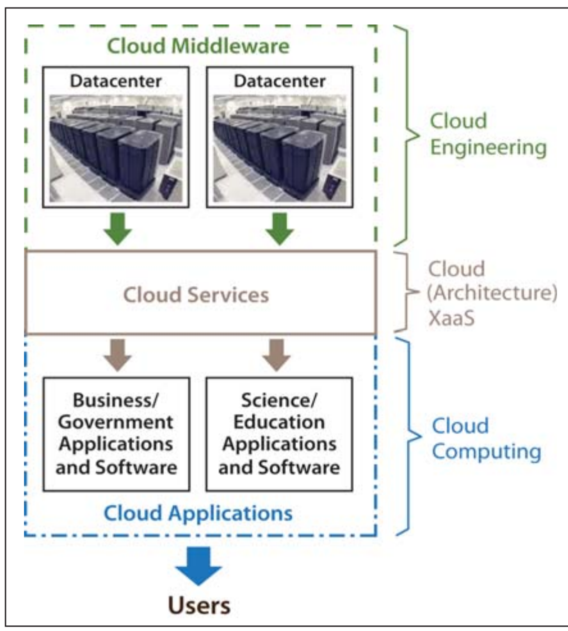


Figure: An overview of autonomic processes in cloud computing. Cloud engineering involves seamlessly linking remote resources via cloud middleware. The cloud architecture provides cloud services based on the needs of the end user. Cloud computing deals with running applications on the aggregate computing resources and returning results to the end user.

Please do not hesitate to provide comments or ask questions by sending me a message at fortes@ufl.edu. Happy reading!

Sincerely,
José A. B. Fortes, CAC Director

Challenges and Opportunities in Cloud Computing

by Renato Figueiredo, UF Site Director

The demand for cloud services is on the rise, and so is the supply of computational resources deployed in data centers provisioning clouds. A key appeal of cloud services is enabling enterprises to reduce costs associated with the ownership and management of computing resources and software. In particular, infrastructure-as-a-service (IaaS) allows enterprises to outsource the provisioning and management of computers, with the flexibility of dynamic, elastic allocation. Virtualization is a key enabling technology that has made this model a reality and led to infrastructures including Amazon's EC2: virtual computers that can be configured, instantiated, resized, and terminated based on demand, on the order of minutes.

How widely this model of provisioning will be adopted is still an open question—even as today's technology enables effective instantiation of virtual resources, there are challenges in security, privacy, interoperability, software licensing, scalable resource provisioning, interconnectivity, and storage management that need to be addressed. Autonomic computing techniques are poised to play a role in addressing several of these issues, from the standpoint of both resource providers and resource consumers.

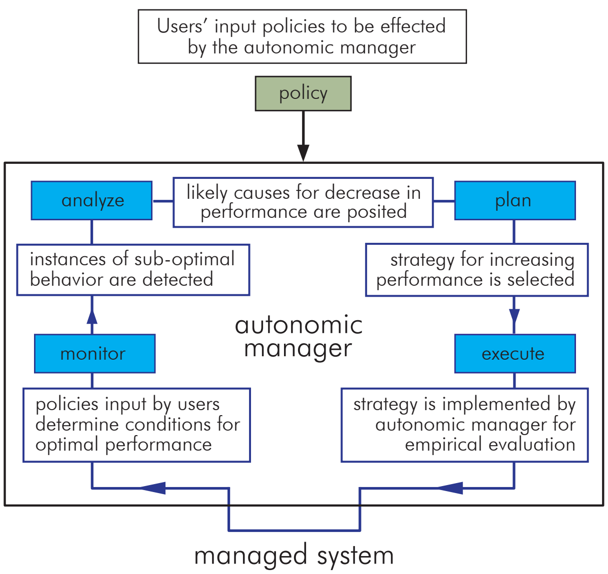


Figure: Schematic of a generalized autonomic manager. The MAPE loop is the iterative process by which user-defined policies are implemented to deal with constantly changing computing resources.

From the provider's standpoint, maximizing profits requires minimizing operational costs. In the large-scale infrastructures of IaaS providers, the costs of power, cooling, and management of thousands of computers distributed across multiple data centers have become larger than the cost of acquiring the actual resources. Furthermore, the very nature of this provisioning model implies that demand is dynamic and elastic. The autonomic computing framework provides a basis for addressing problems such as reducing power and cooling costs through self-optimization under dynamic workloads and improving availability through self-healing and self-protection.

From the user's standpoint, the ability to dynamically configure and scale their infrastructure presents the opportunity to control the supply of resources based on observed or predicted demand, but managing these resources without the aid of information technology can be complex and can create a barrier of entry for end-users. Autonomic computing provides a basis for approaches whereby users are able to define high-level policies regarding the allocation of IaaS resources which are tailored to characteristics of their own workloads and usage patterns, while the mechanisms responsible for enforcing the desired policies are autonomous (see figure at left for a schematic of a generalized autonomic manager). Current IaaS providers such as Amazon EC2 already enable a basis for monitoring and actuation through programmatic APIs (CloudWatch and EC2 system console APIs), while the planning and execution subsystems of an autonomic monitor-plan-analyze-execute (MAPE) loop can incorporate data summarization, prediction and control actions that are tailored to user-defined policies.

Moving forward, it will be crucial that the interfaces supplied by IaaS providers that allow the configuration, creation, migration, and interconnection of computers and storage resources are standardized such that users are not locked in to a single provider, opening a competitive market among multiple providers. The portfolio of ongoing CAC projects addresses challenges in future IaaS infrastructures, such as cross-layer self-optimization, power management, autonomic network security, data streaming, and virtual networking. These are exciting times for research on autonomic systems, and the synergy of academic and industrial perspectives places CAC in a unique position to address these challenges and produce innovative approaches with broad applicability and impact.

Sensor-driven Autonomic Management of Instrumented Data Centers

by Dario Pompili, RU Site Director

Due to the increasing demands for computing and storage, energy consumption, heat generation and cooling requirements have become critical concerns in data centers, in terms of both their growing operating costs (power and cooling) as well as their environmental and societal impacts. Many current data centers are not following a sustainable model in terms of energy consumption growth as the rate at which computing resources are added exceeds the available and planned power capacities. For these reasons, there is a need for realizing environmentally friendly computing ecosystems that maximize energy and cooling efficiency. Technical advances are leading to a pervasive computational ecosystem that integrates computing infrastructures with embedded sensors and actuators, thus giving rise to a new information/sensor-driven autonomic paradigm for monitoring, understanding, and managing natural and engineered systems.

One of the fundamental problems that CAC is studying is the uneven heat generation in data centers, which may be due to several causes, such as uneven workload distribution, poorly designed airflow circulation, failures of AC systems and heterogeneity of physical resources used. Because CPU efficiency is very

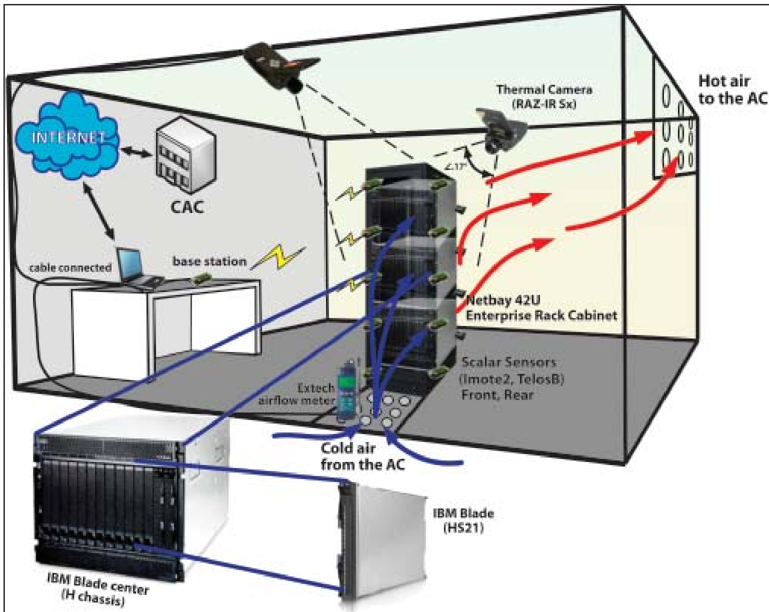


Figure: A schematic of the infrastructure for studying sensor-driven thermal management in data centers planned at Rutgers University. Autonomic redistribution of workload based on temperature profiles could decrease costs and improve performance in data centers used in a wide range of industries, from banking to utility services.

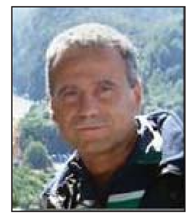
sensitive to hardware temperature increase, it is critical to optimize the effectiveness of cooling systems by monitoring airflow and temperature and humidity of hardware in order to (1) maximize computation performance to meet increasing demands for computing and storage, (2) minimize energy consumption by scheduling jobs in real time, and (3) maximize cooling efficiency to bound growing operating costs.

Within CAC, Rutgers University is developing efficient computing and communication solutions that will enable distributed networks of low-power sensors (monitoring, for example, temperature, humidity, and airflow) to function not only as passive measurement devices but as intelligent data-processing instruments capable of data quality assurance, statistical synthesis, and hypothesis testing as they stream data from the physical environment to the computational world. Specifically, data fusion and uncertainty estimation problems need to be addressed to efficiently deal with the huge amount of sensed data measured at each CPU in a data center. To fully understand the scale of the problem, note that an air-cooled data center may comprise several hundreds or thousands of water-cooled containers, each with tens or hundreds of racks, which are composed of hundreds of blades, each potentially mounting tens of CPUs. Several sensors are finally wired on each CPU and collect a broad range of physical (e.g., temperature, humidity) and logistical (e.g., CPU utilization) measurements. In addition to the sensed data collected at each CPU, sensing devices monitoring the three-dimensional airflow in several strategic sites of the data center may be used to measure the efficiency of the hybrid water- and air-based cooling system of the data center. In order to collect this type of data efficiently, we utilize wireless sensor networks, which are composed of autonomous networked sensing devices (such as airflow meters, which are used for measuring the flow rate or quantity of a moving fluid or gas, e.g., the air) that cooperatively monitor physical or environmental conditions of data centers at different points in space and time. In addition, due to the lack of a priori statistical information of the multiple manifestations of the phenomenon being monitored, pre-deployment optimization of the number and locations of sensor nodes is often not possible (e.g., manifestations of a thermal fugue in a data center caused by overheating would be characterized by an often unpredictable quick change in time and space of temperature and humidity).

In this effort, the main challenge will be to bridge the gap between the computing and communication research areas in such a way as to make progress in solving the multi-disciplinary problem of sensor-driven autonomic management of data centers. Given the complexity and the intrinsic multi-disciplinary nature of the problem at hand, it is necessary to establish synergies between CAC faculty and industrial members and tackle the problem from different angles. Only in this way the pieces of the puzzle will begin to fall into place and the pervasive computational ecosystem paradigm will be enabled.

Survivable Computing System Design Methodology

by Salim Hariri, University of Arizona



Autonomic computing is a promising paradigm for the design of highly survivable computing systems and applications. Autonomic computing is inspired by the human autonomic nervous system that is, to the best of our knowledge, the most sophisticated example of survivable system existing in nature today. It is the body's master controller that monitors changes inside and outside the body, integrates sensory input, and effects appropriate responses that allow the system to survive any drastic changes while performing its essential functions. The nervous system is able to constantly regulate and maintain homeostasis, one of the most remarkable properties of highly complex systems. A homeostatic system (a large organization, an industrial firm, a cell) is an open system that maintains its structure and functions by means of a multiplicity of dynamic equilibriums that are rigorously controlled by interdependent regulation mechanisms. Such a system reacts to every change in the environment, or to every random disturbance, through a series of modifications that are equal in size and opposite in direction to those that created the disturbance. The goal of these modifications is to maintain internal balances.

For a biological system to survive, it must adapt to maintain its essential variables within physiological limits. In a similar argument, we can make the case that for any computing system to be survivable, it must maintain its essential services within acceptable operational regions regardless of anomalous events or disturbances that could have been triggered by cyberattacks, faults, or disaster.

At the University of Arizona site, we are developing innovative autonomic management technologies important to build survivable computing environments, including autonomic security management (self-protection), and autonomic fault management (self-healing). The primary functions of our autonomic survivable computing system architecture include online monitoring, feature selection and correlation, multi-level behavior analysis, decision fusion, automated/semi-automated actions, visualization, and adaptive learning. Using these features, modeled on the human nervous system, we hope to move towards computing systems that are as resilient and robust as the human brain.

Youssif AlNashif
University of Arizona

Youssif AlNashif earned his B.S. in Computer Engineering and his M.S. in Communications and Electronics from the Jordan University of Science and Technology. He earned his Ph.D. degree in Computer Engineering from the University of Arizona under the advising of Arizona Site Director Salim Hariri. His dissertation entitled "Multi-level Anomaly Based Autonomic Intrusion Detection System" was one of the leading projects supported by the NSF CAC. This project provides a fresh solution in detecting and protecting against network attacks in an autonomic and proactive manner. Dr. AlNashif is currently an assistant research professor in the department of Electrical and Computer Engineering at the University of Arizona.

Andréa Matsunaga
University of Florida

Andréa Matsunaga received B.S. and M.S. degrees in Electrical Engineering from the Polytechnic School of the University of São Paulo (Brazil) and a Ph.D. in Electrical and computer Engineering from the Department of Electrical and Computer Engineering at the University of Florida, under the guidance of CAC Director José Fortes.

Her dissertation entitled "Automatic Enablement, Coordination and Resource Usage Prediction of Unmodified Applications on Clouds" provides methods for automating the processes of making existing unmodified applications available as services, coordinating the parallel execution of large problems in a scalable manner, and predicting application resource requirements with a newly developed machine learning algorithm.

Dr. Matsunaga has accepted a postdoctoral position at the University of Florida.

Andres Quiroz
Rutgers University

Andres Quiroz obtained his Ph.D. and M.S. degrees in Computer Engineering at Rutgers University under the direction of Center Co-director Manish Parashar, during which time he contributed to the planning and early development of CAC. His research in the field of autonomic computing, specifically in decentralized online clustering for the autonomic management of distributed systems, has been part of a continuing project at the Center since its launch in 2008. His interests are in topics of data analysis and learning, efficient resource management and provisioning, and middleware for large-scale distributed systems.

Andres will now join Xerox, where he will likely continue to have involvement with CAC through its industry membership.

Arijit Ganguly
University of Florida

Arijit Ganguly received his Ph.D. in Computer and Information Science and Engineering from the University of Florida in 2008, under the supervision of UF Site Director Renato Figueiredo. His dissertation focused on creation of self-managing IP-over-P2P virtual networks to provide to applications an environment of all-to-all connectivity. His doctoral research is continued in the IPOP project, which has been a CAC project since the center's inception in 2008.

Arijit currently works as a Software Development Engineer for the Elastic Compute Cloud (EC2) at Amazon.com in Seattle, where he has been involved in research and development of the Virtual Private Cloud (VPC). His topics of interest include cloud computing, P2P systems, autonomic computing, virtual machines and networks.

With the rapid growth of servers and applications spurred by the Internet, the power consumption of servers has become critically important and must therefore be effectively managed. Power consumption is an increasingly significant percentage of the cost of operating large data centers, which are essential to the operation of banks, investment firms, IT service providers and other large enterprises. The EPA estimated in 2007 that data centers in the U.S. consumed about 61 billion kilowatt-hours (kWh) in 2006, an energy expenditure roughly equivalent to the annual consumption of 5.8 million American households, or five percent of U. S. housing stock. CAC researchers are investigating several approaches to decreasing power consumption in large-scale data centers, making resource usage more efficient and adapting to environmental changes in real time.

Accelerating Hadoop/MapReduce for Heterogeneous Moderate-Sized Datasets Using CometCloud

PI: Manish Parashar [parashar@cac.rutgers.edu]
Lead student: Hyunjoo Kim [hyunjoo@cac.rutgers.edu]

CometCloud is a decentralized computational infrastructure developed at Rutgers that supports distributed applications that require asynchronous coordination and communication. In this project, we use CometCloud and its services to build a MapReduce infrastructure that addresses the dynamic nature of both distributed computing environments and asynchronous applications.

We have deployed real-world applications using the CometCloud-based MapReduce/Hadoop framework in a cloud computing environment that integrates both TeraGrid and Amazon EC2 resources. Overall, the CometCloud-based MapReduce solution can accelerate computations on heterogeneous, medium-sized datasets by delaying or avoiding the use of file reads and writes.

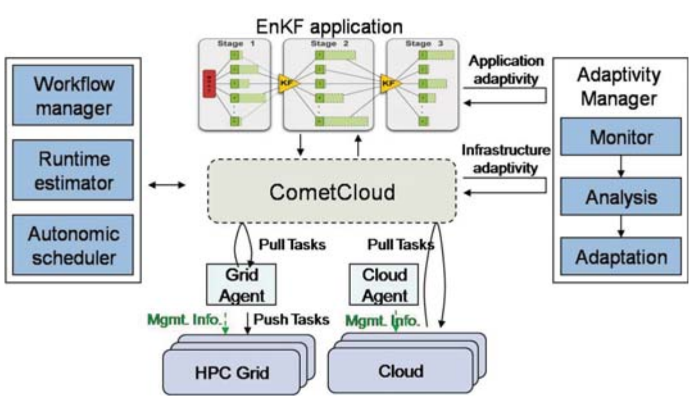


Figure: The MapReduce/CometCloud infrastructure: autonomic managers can adapt to changing resources and changing demands of applications. The Ensemble Kalman Filter (EnKF) application manages the multi-threaded workers that take advantage of multi-processor nodes.

Because many nodes in typical cloud computing systems have multiple processors, we have also developed an interface of multi-threaded workers to accelerate computation further by maximizing the utilization of multi-processor nodes. A representative worker takes the responsibility to communicate with Comet space to designate tasks and communicate with the master to send results, while other threaded workers concentrate on computation. The applications tested so far have demonstrated the performance improvement associated with multi-threaded workers. In this research, we have shown preliminary results, and ongoing efforts are focused on extended evaluation of various aspects of application performance and of MapReduce/Hadoop-Comet-Cloud overhead. We are also working on an event-notification-based consumption model and load balancing on Hadoop-CometCloud.

Demand-driven Service and Power Management in Data Centers

PI: Jose Fortes [fortes@ufl.edu]
Lead student: Jing Xu [jxu@acis.ufl.edu]

One possible approach to reducing power consumption in data centers is to keep machines in low-power states (e.g., standby or off) except when the data center workload requires them to be fully on. This approach depends on a system controller being able to monitor performance, workload and resource demands and to anticipate the need for resources in order to meet users' service-level agreements.

This project aims to devise mechanisms to: monitor and predict workload associated with individual services; profile and model a data center's power consumption and thermal distribution; model and predict global resource demand; dynamically allocate virtual machines to physical machines; devise methods based on control theory and/or market-based approaches to use the above-described mechanisms to minimize the cost of providing individual services while globally minimizing power consumption and delivering contracted service levels; and develop and evaluate software that implements these methods.

We have designed a two-level resource management system in a virtualized data center to dynamically adjust resources to individual virtual machines with local controllers at the virtual-machine level and a global controller at the resource-pool level. The biggest advantage of this two-level control is that it allows independent designs and optimizations for different levels, servers and applications. We have demonstrated that the use of fuzzy logic-based modeling and prediction in local controllers can efficiently and robustly deal with the complexity of uncertainties related to dynamically changing workloads and resource usage. The global controller determines the virtual machine placement and resource allocation using the information from profiling and modeling of the data center's system activities, power consumption and thermal distribution. Experimental results obtained through a prototype implementation demonstrate that, for the scenarios under consideration, the proposed resource management system can significantly reduce resource consumption while still meeting application performance targets.

More Datacenters and Cloud Computing

Go to nsfcac.org to learn about other CAC projects related to data centers and cloud computing, including:

- Scale-right Provisioning Architecture for Next Generation Data Centers
- Sensor-based Autonomic Management of Data Centers
- Autonomic Scaling of Virtual Machines in Next Generation Data Centers

Computing resources and data stores are linked together in increasingly widespread and complex networks. Trusted collaborative systems—which allow networked users to share hardware resources, software under development and private communication—require the use of authenticated end-to-end channels. Approaches to providing secure channels, such as virtual private networks (VPNs), exist; however, these require considerable setup and management (essential tasks include establishing trust and creating and maintaining private network tunnels among end users). Management costs can hinder the establishment of ad-hoc collaborative environments and present a barrier to entry for small research groups with limited resources. Researchers at CAC are investigating new techniques for autonomically managing networked resources that will facilitate collaboration and allow distributed systems to function more reliably and more efficiently.

Self-organizing IP-over-P2P Overlays for Virtual Networking

PI: Renato Figueiredo [renato@acis.ufl.edu]
Lead student: Pierre St. Juste [ptony82@acis.ufl.edu]

This project builds upon and extends the self-configuring IP-over-P2P (IPOP) overlay developed at the University of Florida, which enables scalable, robust, self-configuring virtual network overlays interconnecting physical or virtual resources within a LAN or across a WAN, and supports virtual private networking. In the first year of this project, a novel IP overlay architecture, SocialVPN, which enables self-organizing virtual private network (VPN) links based on relationships established via Web-based social networking infrastructures was conceived. We also conceived a novel system for integrating virtual appliances with SocialVPN-based overlays to enable simple bootstrapping of distributed virtual clusters where core management functions of a virtual organization's certificate authority (CA) can be performed in a user-friendly manner.

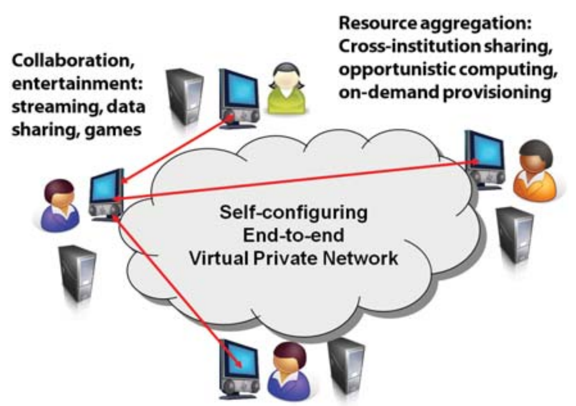


Figure: SocialVPN provides self-configuring virtual networks for collaboration.

Recently, we have shown that SocialVPNs can be bootstrapped without using a centralized online social network backend, by using the P2P backend and distributed hash table of the SocialVPN for storage and lookup of certificate fingerprints and supporting the XMPP protocol used in various online chat networks. Results from the evaluation of a 650-node overlay distributed across PlanetLab, Amazon EC2 and the University of Florida show that 90 percent of node pairs are linked in one second or less, with an average TCP end-to-end throughput of 30Mbits per second.

The software developed for this project is released under the GPL open-source license, and a bootstrap overlay for SocialVPN runs on the PlanetLab infrastructure continuously as a public service. The latest release of SocialVPN (version 0.3) has attracted external developers and has been downloaded over 400 times since September 28, 2009. This software allows users to activate IP-over-P2P virtual networks with enhanced functionality providing for self-configuring virtual private tunneling, application-transparent NAT traversal and social network bindings. Virtual machine appliances integrated with the group-oriented SocialVPN and documentation (including video tutorials) are also available.

Remote and Asynchronous Data Transfers

PI: Manish Parashar [parashar@cac.rutgers.edu]
Lead student: Ciprian Docan [docan@cac.rutgers.edu]

This project addresses the problem of autonomic data streaming at three levels: (1) a data extraction level tries to minimize the impact of I/O operations on application execution while extracting data from running applications, (2) a data sharing/redistribution level provides a virtual, distributed memory shared space that supports online data indexing, flexible data querying and data processing, and (3) a data transport/streaming level tries to optimize the data transport over wide-area networks with in-transit data processing, to satisfy strict data coupling constraints. Our three-level approach is outlined as follows:

Data extraction is implemented using DART. DART is a communication infrastructure built on advanced network interconnects (e.g., RDMA), and uses an asynchronous data transport paradigm provided by the Portals Library. DART provides flexible, asynchronous APIs that allow an application to overlap computation with data transfer and thus reduces the CPU overhead spent in I/O operations and dedicates more time to application computations.

Data sharing is executed using DataSpaces, a dynamic and asynchronous interaction framework. It provides the abstraction of a distributed memory system through a semantically specialized shared data space framework. The framework has decoupled and asynchronous data sharing semantics, enabling cooperative interactions between distinct and heterogeneous parallel applications that run on different resources and can progress at different rates.

Finally, the ADAPT architecture provides services for high-throughput data streaming services and in-transit data manipulation, and provides management strategies for large-scale data-intensive scientific and engineering workflows. This step was driven by demands of data-intensive workflows associated with fusion simulations and develops a metric that estimates the time between when the data was produced and when it is required at the sink and determines to what extent data can be processed in-transit. ADAPT focuses on scheduling the in-transit processing of data using available resources while ensuring that the end-to-end QoS constraints are satisfied and the data arrives at the sink "in-time".

More Networking

Go to nsfcac.org to learn about other CAC projects related to networking, including:

- Virtual Networking for Cross-cloud Computing
- End-to-end Services for P2P Virtual Networks in Collaborative and Cloud Computing Environments

a-CMF: Autonomic Cyberspace Management Framework

PI: Youssif AlNashif [alnashif@ece.arizona.edu]

Advances in computing, networking, software and services will lead to the development of cyberspace services that are pervasive and ubiquitous and touch all aspects of our life. These pervasive services will revolutionize the way we do business, maintain our health, conduct education, and how secure, protect and entertain ourselves. However, along with these advances, we are experiencing grand challenges to ensure that our cyberspace resources and services are highly trustworthy; that means it can effectively tolerate epidemic-style cyber attacks such as viruses and worms, spams, and denial-of-service attacks; deliver applications that can survive hardware/software failures and attacks; and manage autonomically its cyberspace resources and applications by being self-aware, self-adaptive, self-healing, self-*. (i.e., autonomic control and management).

The current management techniques to secure and protect cyberspace resources and services are mainly labor intensive (e.g., patch update), signature based, and not flexible enough to handle the complexity, dynamism and epidemic-style propagation of cyberattacks. Furthermore, the organization boundaries are gradually disappearing so that the idea of creating a defendable perimeter becomes useless. On top of that, the attackers that we need to protect against can be insiders who are trusted and have full access to system resources and services.

In this project, we are developing an autonomic cyberspace management framework that provides tools and services 1) to detect any type of cyberattack (known or unknown) using online monitoring and multi-level behavior analysis techniques; and 2) to minimize their impacts and prevent their epidemic propagations using our risk impact analysis capabilities and automated actions. The a-CMF will provide management capabilities to build survivable critical applications (e.g., healthcare system, smart power grid) that can tolerate failures and cyberattacks. Our approach to implement the a-CMF architecture is based on autonomic computing. The autonomic paradigm is inspired by the human autonomic nervous system that handles complexity and uncertainties, and aims at realizing computing systems and applications capable of managing themselves with minimum human intervention.

Decentralized Online Clustering for Autonomic System Management

PI: Manish Parashar [parashar@cac.rutgers.edu]
Lead student: Andres Quiroz [aquirozh@cac.rutgers.edu]

Autonomic techniques based on dynamic policy application provide a promising approach for the effective management of distributed computational infrastructures, by reducing management complexity and allowing human administrators to focus primarily on the definition of these policies at a high level. However, these polices (called meta-policies) are typically defined with static constraint thresholds and are associated with specific system goals or with known states of the managed entities obtained through feedback. This limits their applicability where the appropriate management actions depend on dynamic system properties, which require adapting application thresholds and parameters without modifying absolute policy definition constraints. We consider policies to be management actions to be applied to events composed of multiple attributes, according to the attribute values. A policy is then defined by a region in a multidimensional attribute space, an entity or entities to which it applies (i.e., the source of the event), and a management action. We consider two distinct policy spaces: one composed of attributes that are known indicators of system goals and where policy boundaries can be clearly delineated by an administrator (policy definition space), and another composed of managed attributes that are subject to management actions (policy application space). These two spaces are often not the same and are related in a dynamic way, dependent on the design of the system, workload and client interactions, available resources, etc.

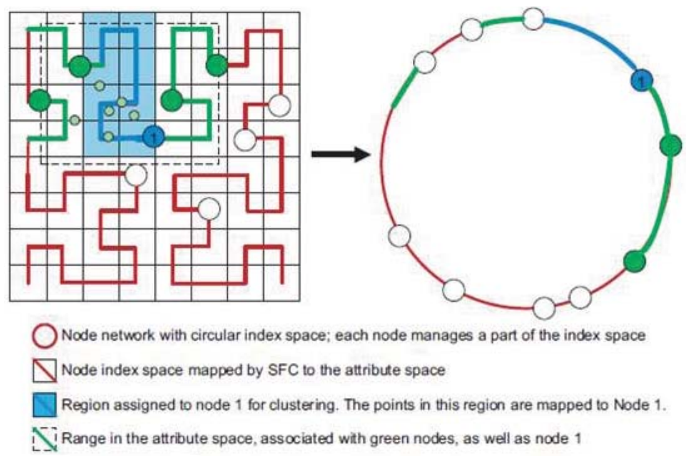


Figure: Overlay-based implementation of decentralized clustering technique

Our approach is to correlate both spaces by observing the feedback attributes for a sequence of monitored events and finding temporary mappings by finding clusters of events in these sequences that occur in both spaces. These clusters determine the dynamic boundaries of newly generated policies in the application space. In order to find clusters that occur in both spaces, we perform clustering in a combined space, and then project the results into each of the subspaces. The meta-policy whose region of application contains the cluster in the definition space can be used to generate the policy in the region of the application space that corresponds to the same cluster. Policies can also be associated with particular managed entities, as opposed to individual clusters of events. This is done by maintaining a profile for each managed entity in the system, composed of the clusters of events (profile elements) that correspond to that entity. With the possibility of performing clustering online in a decentralized manner, this approach can be applied to autonomic management scenarios such as distributed resource management and usage control.

More Applications and Services

Go to nsfcac.org to learn about other CAC projects related to applications and services, including:

- Finding Objects in Images Using 3-D Models
- Real-time Scheduling of Mixture of Expert Systems
- Improving Timer Accuracy in Virtualized Systems for Real-time Computing

The rapid growth and deployment of network technologies and Internet services have made reliability, security and management of networks a challenging research problem. Attacks on networked systems have become more complex and organized, more dynamic, and more disruptive than ever. Existing techniques that respond to such attacks, such as intrusion detection systems and firewall hardware/software systems, are manual-intensive; this makes them too slow to respond efficiently to complex and interacting organized network attacks. Furthermore, countermeasures such as signatures for intrusion detection cannot detect new kinds of attacks. The successful operation of cyberinfrastructure requires the ability to detect previously unseen attacks in real time, to avoid system failures, to provide accurate risk and impact analyses, and to take appropriate action.

Anomaly-based HTTP Attack Detection System

PI: Salim Hariri, University of Arizona
Lead student: Ishtiaq Hossein [ihossein@email.arizona.edu]

HTTP has become a universal transport protocol being used for file sharing, Web services, media streaming, payment processing, and even for protocols such as SSH. With the advent of Web Services and Cloud Computing technologies, more and more businesses are being hosted in the Internet, meaning that we can expect increased used of HTTP in future. There have been many application-level attacks using HTTP in the past, and new attacks are emerging continually. This work aims at developing a robust Anomaly Based HTTP Attack Detection System. Our current prototype collects data, trains the system based on the data collected and then observes the network behavior for any deviation from normal traffic behavior. Currently we consider only HTTP headers and use multiple features to capture the behavior. The RIPPER Association Rule Generation technique is employed to build normal and abnormal profiles. We have performed a small scale testing with an attack library consisting of 15 attacks. The initial results are very encouraging with over 90% detection rate and very few false negatives. We are currently refining our detection approach to improve the performance of our system further by including Temporal Behavior Analysis and using a richer attack library.

Management of the Health of IT Systems

PI: Jose Fortes [fortes@ufl.edu]
Lead student: Selvi Kadirvel [selvik@acis.ufl.edu]

Self-caring IT systems are systems capable of monitoring and managing their own health and, indirectly, their useful lifetime. Unlike self-healing systems, which are reactive to faults and failures, self-caring systems are aware of their health and hence can potentially circumvent and adapt to impending faults or recover from them. This early detection and handling of health deteriorations results in significant benefits in terms of both cost savings and down time reduction. The use of a systematic step-by-step methodology for system modeling using Petri nets followed by the incorporation of remaining useful life (RUL) extension techniques is proposed as the core of our health management solution.

IT systems are discrete event systems in that state changes are initiated by events rather than by time. For such a system, Petri nets serve as a convenient tool to capture system properties of interest and facilitate analysis, simulation and system control. Also as a graphical model, Petri nets allow for easy visualization of the current and future states of the system. The basic system model is incrementally augmented to incorporate health management. The final model obtained after augmentation serves as a global manager to control and manage system operation. In this project, the focus is on the management of health degradations caused by potential resource exhaustion faults. A resource in this context is defined as any software, hardware or operational entity that is available in limited supply and is essential for system functioning.

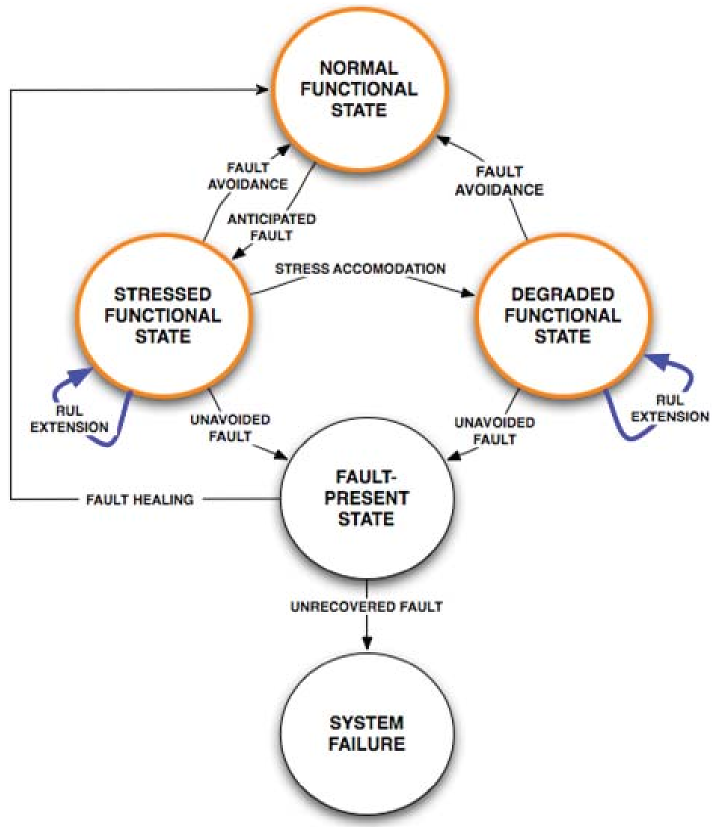


Figure: Health state diagram for a potentially complex IT system. If the system is self-caring, it can extend its remaining useful life (RUL) once the system reaches a stressed or degraded functional state. The RUL extension can make the difference between, for instance, quitting a process abruptly and shutting down the process safely, thereby avoiding a fault and possibly a system failure.

We have applied our proposed methodology to a batch-based virtualized high-performance computing environment to show its quantitative and qualitative benefits. Classical control theory as applied to computing systems is utilized as one of the techniques to perform remaining useful life extension at the subsystem level. As a simple illustration of this, an RUL controller was built for the application execution stage of the batch system that is made to contain an injected memory exhaustion fault.

More Security and Reliability

Go to nsfcac.org to learn about other CAC projects related to security and reliability, including:

- Autonomic Protection System for DNS Protocol
- Highly Secure Scada System

Events

Indraneel Kulkarni, of Rutgers University, Arjun Hary, and Ram Prasad Viswanathan, all of the University of Arizona, graduated with Master's degrees in the past year.

Dr. Maurício Tsugawa, of the University of Florida, graduated with a Ph.D. Dr. Tsugawa defended his thesis, entitled "On the Design, Performance, and Management of Virtual Networks for Grid Computing," on July 10, 2009. Since graduating, Dr. Tsugawa has rejoined the Center as a research scientist at the University of Florida site.

Dr. Andréa Matsunaga completed her Ph.D. studies at the University of Florida under the guidance of Dr. José Fortes and defended her thesis on March 26, 2010. She is profiled in the CAC People section on page 5.

The CAC Spring 2010 semiannual meeting was held in Bellevue, Washington on April 28 & 29. The meeting was held on the Microsoft Advanta Campus, and was the first semiannual meeting hosted jointly by a university site and an industry member company. The agenda for the meeting and links to conference materials can be found at nsfcac.org.

Facilities

The University of Florida site is creating a new suite of offices for CAC personnel and visitors. Photos of the renovation in progress are below; stay tuned for a feature on the new space in the next issue of the CAC Bulletin.

The computing machine room at Rutgers University has been entirely renovated and a new cooling unit has been installed. This room will host a new state-of-the-art cluster consisting of two Dell M1000E Modular Blade Enclosures, necessary interconnect/management infrastructure, and a supervisory node. Each enclosure is maximally configured with sixteen blades, each blade having two Intel Xeon E5504 Nehalem family quad-core processors at 2.0 GHz, forming an eight-core node. Each node has 6 GB RAM and 73 GB of local disk storage (10,000 RPM). The network infrastructure is comprised of an integrated 16-port Mellanox InfiniBand switch within each blade chassis, each switch having eight uplink ports and linked via eight InfiniBand lanes to the uplink ports on the switch in the other chassis. All blades will have Mellanox Quad-Data-Rate (QDR) InfiniBand interface cards. There will also be an integrated (redundant) 1-Gigabit Ethernet within each chassis, with two pairs of 10-Gigabit uplink capability in each chassis. In the aggregate, the new system will consist of 32 nodes, 256 cores, 192 GB memory and ~2.5 TB disk capacity, with a 200-Gigabit InfiniBand network and two 1-Gigabit Ethernet networks. More on the Rutgers machine room renovation will be included in the next issue.

The University of Florida site has considerably extended its computational resources, within the center and across multiple institutions. UF is in the process of expanding a shared-memory cluster funded by the ongoing NSF Major Research Infrastructure (MRI) project "Acquisition of Instrumentation for Coupled Experimental-Computational Neuroscience and Biology Research" to approximately double its processing and memory capacity. The cluster currently has a 32-socket (128-core) IBM eServer 1350 organized as two 64-core, 512GB NUMA systems. This system is fully virtualized and runs the XenServer software from founding CAC member Citrix.

An ongoing project funded by the NSF Computing Research

Figure: Photos of the new suite of offices under construction at the University of Florida site. The photos at left show early stages of the renovation. The photo above was taken on June 15, 2010, with the remodeling near completion. Glass partitions make the new suite a highly configurable space, allowing it to serve as an office for visitors from industry and other universities, a conference room, graduate student offices, and a meeting space for CAC's semiannual workshops.

Infrastructure (CRI) program, "An Instrumented Datacenter Infrastructure for Research on Cross-Layer Autonomics" is also contributing to the UF site's growing computing capability. The cross-layer project is a collaboration among CAC's three university sites, each of which is deploying a distributed set of clusters that will feature temperature and power sensors internal and external to the clusters to enable research on autonomic computing in virtualized, distributed data centers. The data centers established by this project are representative of cloud infrastructures, a major topic of interest for CAC researchers and member companies.

Another CRI-funded project, Archer is a multi-year, multi-university effort to provide shared resources to benefit researchers and educators in the field of computer architecture. The Archer community gives access to not only computing resources, but also a repository of user-created tools and algorithms to foster collaboration and accelerate scientific discovery. The Archer distributed system is up and running with clusters at University of Florida, Florida State University, Northeastern University, the University of Texas at Austin, and University of Minnesota.

Figure: Some of the additions to the computing facilities at the University of Florida site: a) UF's contribution to the Distributed Research Testbed (DiRT), funded by the NSF's Computing Research Infrastructure (CRI) initiative; b) the expanding resources dedicated to the Coupled Experimental-Computational Neuroscience and Biology project; c) all-new instrumentation for the Cross-layer Autonomics project—temperature sensing and monitoring equipment will be added in the project's second year; and d) the testbed created for the FutureGrid project, a multi-university effort facilitating cross-cloud computing.

A third project also funded by the CRI program, the "Distributed Research Testbed (DiRT)" creates a testbed across five U.S. universities (the University of Mississippi, Notre Dame, University of Hawaii, University of Chicago, and UF). The testbed consists of cluster and storage nodes that will provide researchers with access to a heterogeneous, geographically distributed set of resources with flexibility in the configuration of virtualized and data-intensive wide-area experiments. In addition to physical local resources, UF has obtained access to the Elastic Cloud (EC2) infrastructure through an Amazon Web Service grant. The resource credits will be used for research of systems "at scale," giving us the flexibility of dynamically bringing up and down pools of hundreds of virtual machines on the distributed data centers hosted by Amazon.

Awards

CAC alumnus Viraj Bhat received the prestigious 2009 Grid Award at Yahoo!. The Grid Award is an yearly award that is presented within Yahoo! for evangelizing grid (Hadoop, Pig, Zookeeper) technologies, profiling and optimizing MapReduce and Pig applications within Yahoo!, and migrating various mission critical applications which have very strict SLA requirements to the Grid and development of various grid tools to help users migrate to the Grid. Viraj completed his Ph.D. under the guidance of Center Co-director Manish Parashar in Spring 2008 and joined Yahoo! in May 2008.

In Fall 2009, the National Science Foundation awarded the $10.1M FutureGrid project. FutureGrid is led by Indiana University, with Center Director Fortes and UF Site Director Figueiredo serving as co-investigators. The FutureGrid project involves the creation of a large-scale testbed that is comprised of a geographically distributed set of heterogeneous computing systems, and that will support virtual machine-based environments as well as native operating systems for experiments on cloud- and grid-computing middleware. The University of Florida team contributes to the project with expertise both in open-source virtual networking software tools (such as Virtual Network (ViNe) and IP-over-P2P (IPOP), developed by University of Florida researchers) that enable resource connectivity across cloud providers, and in virtual machine grid appliances that provide pre-packaged educational modules on parallel/distributed computing that are simple to replicate, deploy and use. Partners in the FutureGrid project include the University of Texas at Austin/Texas Advanced Computing Center, the University of Chicago/Argonne National Laboratory, the San Diego Supercomputing Center at the University of California at San Diego, the University of Southern California Information Sciences Institute, the University of Tennessee at Knoxville, and Purdue University.

The University of Arizona site, in collaoration with CAC member company AVIRTEK, Inc., received a phase II award of $750,000 from the Air Force to develop an autonomic Cyber Battle Management System (CBMS) to autonomically manage the configration of cyberspace resources and services and protect them against any type of cyber attacks, known or unknown. The project will commence in September 2010 and continue through September 2010 under the direction of PIs Salim Hariri and Youssif AlNashif.

Become a CAC industrial member

CAC members collaborate with and advise researchers to create a diverse, industry-relevant research program. Members are afforded access to leading-edge developments in autonomic computing and to knowledge accumulated by academic researchers and other industry partners. The annual membership fee of $35K allows industry partners to reap the full benefits of CAC's full research program, with an operating budget of over $1 million annually. To inquire about membership in CAC, please contact Center Director Jose Fortes at fortes@ufl.edu.

Face to Face

CAC researchers are committed to promoting the Center, addressing industry-relevant research problems, and furthering collaboration. We achieve this in part by attending conferences to increase the visibility of CAC projects and attract new members, and by visiting CAC member companies to boost industry participation. This section of the Bulletin highlights the meetings between CAC personnel and the world at large.

September 2009

On Sept. 29, UF Site Director Figueiredo presented the seminar "Self-configuring P2P Virtual Networks and Applications in Cloud Computing and Social VPNs" as part of the Distinguished Lecture Series at the Florida International University School of Computing and Information Sciences.

December 2009

On December 18, 2009, Center Director José Fortes visited Northrop Grumman Aerospace Systems in Bethpage, New York, where he gave a presentation on CAC projects and discussed topics related to the management of health of IT systems.

January 2010

On January 13, Center Co-director Manish Parashar presented the eSI Visitor Seminar entitled "Addressing Complexity in Emerging Cyber-Ecosystems—Experiments with Autonomic Computational Science" at the e-Science Institute in Edinburgh, UK.

March 2010

On March 9, 2010, Dr. Fortes, Dr. Figueiredo and Arizona Site Director Salim Hariri visited Microsoft in Seattle, WA, where they met with several technical groups and presented ongoing CAC work.

On March 10, 2010, Dr. Fortes visited Citrix in Seattle, WA, where he reported on new technical projects being planned by CAC researchers. He visited Citrix again on April 30.

April 2010

Dr. Parashar presented a technical plenary talk, "Addressing Complexity in Emerging Cyber-Ecosystems—Exploring the Role of Autonomic in E-Science," on April 13 at the 5th Enabling Grids for E-sciencE (EGEE) User Forum, held in Uppsala, Sweden.

Dr. Parashar gave the keynote address, entitled "Computational Science and Engineering & Clouds," at the 7th High Performance Grid Computing Workshop on April 19 in Atlanta, GA. The workshop was held In conjunction with the 24th International Parallel and Distributed Computing Symposium (IPDPS 2010).

Dr. Figueiredo visited Intel in Hillsboro, OR on April 30 to present "Autonomic Virtual Networks and Applications in Cloud and Collaborative Computing Environments" and meet with Intel researchers to discuss topics of common interest. Intel has been a member of the University of Florida site since the Center's inception.

CAC Personnel at the University of Arizona organized the SCADA Cyber Security Workshop, held at Biosphere 2 in Tucson, AZ on April 20-21. At the workshop, Dr. Nabil Adam, from the Department of Homeland Security, gave a keynote speech about the importance of cyber security on industrial control systems controlled and managed by SCADA systems.

May 2010

Dr. Fortes traveled to Melbourne, Australia on May 13 to deliver a keynote address at the 10th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid 2010). Dr. Fortes' address, entitled "Sky Computing: When Multiple Clouds become One" focused on virtual networking technology that enables users to access resources across many clouds for single simulations or applications. Dr. Parashar also attended CCGrid 2010 to serve on the Panel on Cloud Deployment Trajectories for National Goals.

June 2010

On June 7, Dr. Fortes delivered the first of two keynote addresses (photo at right) at the 2nd Grid Meets Autonomic Computing (GMAC) Workshop, held in conjunction with the 7th International Conference on Autonomic Computing and Communications (ICAC 2010) in Washington, DC. Dr. Parashar served as general chair for ICAC 2010, and served on the panel "Towards grid/clouds benchmarks for Autonomic Computing". Dr. Figueiredo served as program co-chair with Emre Kiciman of Microsoft Research.

Dr. Figueiredo was a panelist in the "Cloud versus Cloud: how will cloud computing shape our world?" panel at the 2010 ACM International Symposium on High Performance Distributed Computing (HPDC 2010), held in Chicago, IL on June 20-25. The Center co-sponsored the dinner cruise at HPDC on Wedensday, June 23, along with Northwestern University. Dr. Hariri introduced the Center to the more than 130 attendees of the dinner fucntion. Dr. Hariri also participated in the "Cloud vs. Cloud" panel, highlighting the importance of security solutions for cloud computing to facilitate the rapid deployment and acceptance of cloud services.

Rutgers Site Director Dario Pompili served on the Panel on Experimental Research: Fashionable or Fundamental at the 7th IEEE Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON), held on June 21-25 in Boston, MA.

On June 22, Dr. Figueiredo served as general chair for the 4th Workshop on Virtualization Technologies in Distributed Computing (VTDC 2010) in Chicago, IL.

Dr. Fortes attended the International Advanced Research Workshop on High-Performance Computing, Grids and Clouds, held June 21-25 in Cetraro, Italy. He gave an invited talk, entitled "Cross-cloud Computing," on June 23 in a session on grid and cloud technology and systems.

On June 24, Dr. Fortes delivered the second keynote address, entitled "Virtually Networking the Clouds," at the 9th IFIP Annual Mediterranean Ad Hoc Networking Workshop in Juan-les-Pins, France.

1. Lee, E. K., I. S. Kulkarni, D. Pompili, and M. Parashar, "Proactive Thermal Management in Green Datacenters," to appear in the Journal of Supercomputing (Springer), Special Issue on Energy-efficient Computing, 2010.

2. Liu, H., E. K. Lee, D. Pompili, and X. Kong, "Application of Infrared Camera Networks for Datacenter Thermal Management," submitted for publication.

3. St. Juste, Pierre, David Wolinsky, Kyungyong Lee, P. O. Boykin, Renato Figueiredo, "SocialDNS: Trusted P2P Name Resolution through Social Networking", submitted.

4. St. Juste, Pierre, David Wolinsky, P. Oscar Boykin, Michael J. Covington, Renato Figueiredo, "SocialVPN: Enabling Wide-Area Collaboration with Integrated Social and Overlay Networks", to appear in the Elsevier Computer Networks Journal, 2010

5. Guim, Francesc, Ivan Rodero, Julita Corbalan, Manish Parashar, "Enabling GPU and Many-Core Systems in Heterogeneous HPC Environments using Memory Considerations," submitted to the IEEE 12th International Conference on High Performance Computing and Communications, September 2010.

6. Rodero, Ivan, Juan Jaramillo, Andres Quiroz, Manish Parashar, Francesc Guim, S. Poole, "Energy-Efficient Application-Aware Online Provisioning for Virtualized Clouds and Data Centers ," submitted to the First Annual IEEE International Green Computing Conference, August 2010.

7. Rodero, Ivan, Eun Kyung Lee, Dario Pompili, Manish Parashar, Marc Gamell, Renato J. Figueiredo, "Towards Energy-Efficient Reactive Thermal Management in Instrumented Datacenters," submitted to the 5th Workshop on Virtualization in High-Performance Cloud Computing, August 2010.

8. Docan, C., M. Parashar and S. Klasky, "DataSpaces: An Interaction and Coordination Framework for Coupled Simulation Workflows," presented at the High Performance and Distributed Computing (HPDC'10), Chicago, Illinois, June 2010.

9. Hary, Arjun, Ali Akoglu, Youssif AlNashif, Salim Hariri, and Darrel Jenerette, "Design and Evaluation of a Self-healing Kepler for Scientific Workflows," presented at the ACM International Symposium on High Performance Distributed Computing, Chicago, IL, June 2010.

10. Quiroz, A., Parashar, M., Gnanasambandam, N., and Sharma, N, "Autonomic Policy Adaptation using Decentralized Online Clustering," to appear in Proceedings of International Conference on Autonomic Computing (ICAC), June 2010.

11. Rodero, Ivan, Juan Jaramillo, Andres Quiroz, Manish Parashar, Francesc Guim, "Towards Energy-Aware Autonomic Provisioning for Virtualized Environments," presented at The ACM International Symposium on High Performance Distributed Computing (HPDC) 2010, Chicago, IL, June 2010 (poster).

12. Docan, C., F. Zhang, M. Parashar, J. Cummings, N. Podhorszki and S. Klasky. "Experiments with Memory-to-Memory Coupling for End-to-End Fusion Simulation Workflows", presented at the 10th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid'10), Melbourne, Australia, May 2010.

13. Matsunaga, Andréa and José Fortes, "On the Use of Machine Learning to Predict the Time and Resources Consumed by Applications," presented at the 10th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, May 2010.

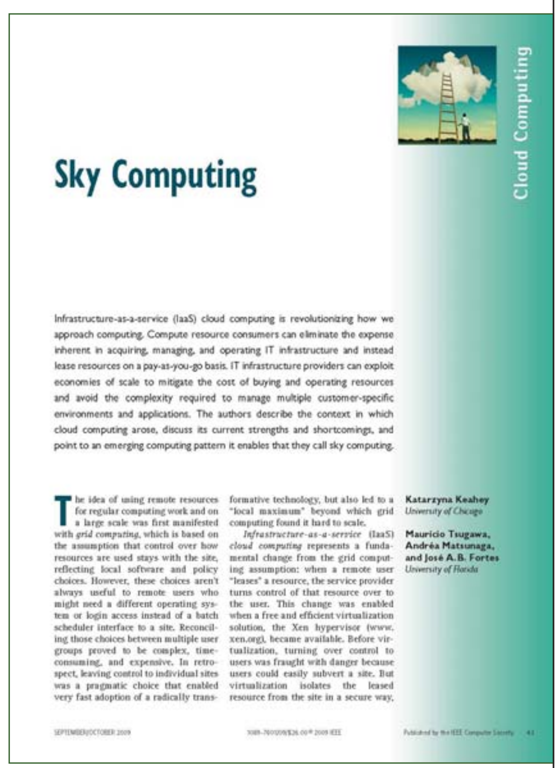
14. Docan, C., M. Parashar and S. Klasky, "Enabling High Speed Asynchronous Data Extraction and Transfer Using DART," Concurrency and Computation: Practice and Experience, DOI:10.1002/cpe.1567, March 2010.

15. Kadirvel, Selvi and José Fortes, "Towards IT Systems Capable of Managing Their Health," preliminary extended abstract, presented at the 16th Monterey Workshop, Redmond, WA, March 2010.

16. Chaudhari, Shivangi, "Accelerating Hadoop Map-Reduce for Small/Intermediate Data Sizes using the Comet Coordination Framework," MS Thesis, Department of Electrical & Computer Engineering, Rutgers, The State University of New Jersey, 2009. Advisor: Manish Parashar.

17. Kim, Hyunjoo, Yaakoub el-Khamra, Shantenu Jha, and Manish Parashar, "An Autonomic Approach to Integrated HPC Grid and Cloud Usage," presented at the 5th IEEE International Conference on e-Science, Oxford, UK, December 2009.

18. Kumar, Mahendra, Richard E. Newman, José Fortes, Dennis R. Durbin, Flaura K. Winston, "An IT appliance for remote collaborative review of mechanisms of injury to children in motor vehicle crashes," in Proceedings of the 5th International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom 2009), pp. 1-9, November 2009.



Cloud Computing

Sky Computing

Infrastructure-as-a-service (IaaS) cloud computing is revolutionizing how we approach computing. Compute resource consumers can eliminate the expense inherent in acquiring, managing, and operating IT infrastructure and instead lease resources on a pay-as-you-go basis. IT infrastructure providers can exploit economies of scale to mitigate the cost of buying and operating resources and avoid the complexity required to manage multiple customer-specific environments and applications. The authors describe the context in which cloud computing arose, discuss its current strengths and shortcomings, and point to an emerging computing pattern it enables that they call sky computing.

Katarzyna Keahey
University of Chicago

Mauricio Tsugawa, Andréa Matsunaga, and José A.B. Fortes
University of Florida

The idea of using remote resources for regular computing work and on a large scale was first manifested with *grid computing*, which is based on the assumption that control over how resources are used stays with the site, reflecting local software and policy choices. However, these choices aren't always useful to remote users who might need a different operating system or login access instead of a batch scheduler interface to a site. Reconciling those choices between multiple user groups proved to be complex, time-consuming, and expensive. In retrospect, leaving control to individual sites was a pragmatic choice that enabled very fast adoption of a radically transformative technology, but also led to a "local maximum" beyond which grid computing found it hard to scale.

Infrastructure-as-a-service (IaaS) *cloud computing* represents a fundamental change from the grid computing assumption: when a remote user "leases" a resource, the service provider turns control of that resource over to the user. This change was enabled when a free and efficient virtualization solution, the Xen hypervisor (www.xen.org), became available. Before virtualization, turning over control to users was fraught with danger because users could easily subvert a site. But virtualization isolates the leased resource from the site in a secure way,

SEPTEMBER/OCTOBER 2009 1089-7801/09/$26.00 © 2009 IEEE Published by the IEEE Computer Society

19. Xu, Jing, Ming Zhao and José Fortes, "Cooperative Autonomic Management in Dynamic Distributed Systems," in Proceedings of the 11th International Symposium on Stabilization, Safety, and Security of Distributed Systems, pp. 756-770, November 2009.

20. Keahey, Katarzyna, Maurício Tsugawa, Andréa Matsunaga and José A. B. Fortes, "Sky Computing," IEEE Internet Computing, September/October 2009.

21. Kim, Hyunjoo, Manish Parashar, David J. Foran and Lin Yang, "Investigating the Use of Cloudbursts for High-throughput Medical Image Registration," presented at the 10th IEEE/ACM International Conference on Grid Computing (GRID2009), Banff, Canada, October 2009.

22. Quiroz, A., N. Gnanasambandam, M. Parashar, H. Kim, and N. Sharma, "Towards Autonomic Workload Provisioning for Enterprise Grids and Clouds," presented at the 10th IEEE/ACM International Conference on Grid Computing (GRID2009), Banff, Canada, October 2009.

23. Xu, Jing, José Fortes , "Multiobjective Virtual Machine Placement and Migration", ACIS Technical report TR-09-001, September 2009.

24. Xu, Jing, José Fortes, "Optimization in Autonomic Data Center Resource Management", ACIS Technical report TR-09-002, September 2009.

25. Khargharia, Bithika, Salim Hariri and Mazin S. Yousif, "An Adaptive Interleaving Technique for Memory Performance-per-Watt Management," IEEE Transactions on Parallel and Distributed Systems, Vol. 20, no. 7, July 2009.

26. Tsugawa, Maurício, "On the Design, Performance, and Management of Virtual Networks for Grid Computing," Ph.D. dissertation, [Gainesville, Fla.]: University of Florida, July 2009.

27. Gnanasambandam, N., M. Parashar, A. Quiroz, H. Kim, and N. Sharma, "Towards Autonomic Workload Provisioning for Enterprise Data Centers and Clouds," presented at Hot Topics in Autonomic Computing Workshop at ICAC 09, Barcelona, June 2009.

28. Jararweh, Yaser, Arjun Hary, Youssif Al-Nashif, Salim Hariri, Ali Akoglu, Darrel Jenerette, "Accelerated Discovery through Integration of Kepler with Data Turbine for Ecosystem Research," presented at the 2009 IEEE/ACS International Conference on Computer Systems and Applications, May 2009.

29. Quiroz, A., N. Gnanasambandam, M. Parashar, and N. Sharma, "Robust Clustering for the Management of Self-monitoring Distributed Systems," Springer Journal on Cluster Computing, vol. 12, vol. 1, p. 73, February 2009.

30. St. Juste, Pierre , David Wolinsky, Jiangyan Xu, Michael Covington, Renato Figueiredo, "On the Use of Social Networking Groups for Automatic Configuration of Virtual Grid Environments," presented at Grid Computing Environments (with SC-2008), November 2008.

31. Xu, Jing, Ming Zhao, José Fortes, Robert Carpenter, Mazin Yousif, "Autonomic Resource Management in Virtualized Data Centers Using Fuzzy Logic-based Approaches", Cluster Computing, vol. 11, no. 3, September 2008.

32. Figueiredo, Renato , P. Oscar Boykin, Pierre St. Juste, David Wolinsky, "Social VPNs: Integrating Overlay and Social Networks for Seamless P2P Networking," presented at IEEE WETICE/COPS 2008. Best paper award, June 2008.

33. Hariri, Salim, Yaser Jararweh, Yeliang Zhang, Talal Moukabary, "Physics Aware Programming Paradigm: Approach and Evaluation," in Proceedings of the 6th International Workshop on Challenges of Large Applications in Distributed Environments, June 2008.

34. Khargharia, Bithika, Salim Hariri, Wael Kdouh, Manal Houri, Hesham El-Rewini, Mazin Yousif, "Autonomic Power and Performance Management of High-performance Servers," presented at IEEE International Symposium on Parallel and Distributed Processing, Miami FL, April 2008.

35. Xu, Jing, Ming Zhao, José Fortes, Robert Carpenter, Mazin Yousif, "On the Use of Fuzzy Modeling in Virtualized Data Center Management", in Proceedings of the Fourth international Conference on Autonomic Computing, June 2007.

To request additional copies of this newsletter, please send an email to Julie at jooly@acis.ufl.edu