NSF CISE Investments in Cloud Computing

Keith Marzullo
Director, Division of Computer and Networks Systems
Directorate of Computer and Information Science and Engineering
U.S. National Science Foundation

IFIP 10.4 Working Group Meeting
Tavira, Portugal

17-19 January 2013

Roadmap

- NSF’s mandate in cloud computing
- What NIST considers cloud computing
- Some current CISE projects
- Some new initiatives

NSF and Cloud Computing

- There is no program specifically targeted for cloud computing.

- OCI also funds projects in cloud computing.

- Projects can support foundational research, applied research, education, infrastructure development, and instrumentation and infrastructure deployment.

America Competes Reauthorization Act of 2010, Section 524

SEC. 524. CLOUD COMPUTING RESEARCH ENHANCEMENT.

(a) RESEARCH FOCUS AREA.—The [NSF] Director may support a national research agenda in key areas affected by the increased use of public and private cloud computing, including—

(1) new approaches, techniques, technologies, and tools for—

(A) optimizing the effectiveness and efficiency of cloud computing environments; and

(B) mitigating security, identity, privacy, reliability, and manageability risks in cloud-based environments, including as they differ from traditional data centers;

(2) new algorithms and technologies to define, assess, and establish large-scale, trustworthy, cloud-based infrastructures;

(3) models and advanced technologies to measure, assess, report, and understand the performance, reliability, energy consumption, and other characteristics of complex cloud environments; and

(4) advanced security technologies to protect sensitive or proprietary information in global-scale cloud environments.

America Competes Reauthorization Act of 2010, Section 524

(b) ESTABLISHMENT.—

(1) IN GENERAL.—Not later than 60 days after the date of enactment of this Act, the Director shall initiate a review and assessment of cloud computing research opportunities and challenges, including research areas listed in subsection (a), as well as related issues such as—

(A) the management and assurance of data that are the subject of Federal laws and regulations in cloud computing environments, which laws and regulations exist on the date of enactment of this Act;

(B) misappropriation of cloud services, piracy through cloud technologies, and other threats to the integrity of cloud services;

(C) areas of advanced technology needed to enable trusted communications, processing, and storage; and

(D) other areas of focus determined appropriate by the Director.

(2) UNSOLICITED PROPOSALS.—The Director may accept unsolicited proposals that review and assess the issues described in paragraph (1). The proposals may be judged according to existing criteria of the National Science Foundation.

(c) REPORT.—The Director shall provide an annual report for not less than 5 consecutive years to Congress on the outcomes of National Science Foundation investments in cloud computing research, recommendations for research focus and program improvements, or other related recommendations. The reports, including any interim findings or recommendations, shall be made publicly available on the website of the National Science Foundation.

NIST Essential Characteristics of Cloud Computing

http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf

- On-demand self-service provisioning of computing capabilities (server time, network time, etc).
- Broad network access that promotes the use by heterogeneous thin or thick client platforms.
- Resource pooling using a multi-tenant model, with different physical and virtual resources dynamically assigned and reassigned according to consumer demand. Resources include storage, processing, memory, and network bandwidth.
- Rapid elasticity of capability provisioning and release to scale with demand.
- Automatically control and optimize resource use by leveraging a metering capability at some level of abstraction appropriate to the type of service (e.g., storage, processing, bandwidth, and active user accounts).

NIST Service and Deployment Models of Cloud Computing

http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf

- Service models
  - Software as a Service
  - Platform as a Service
  - Infrastructure as a Service
- Deployment models
  - Private cloud
  - Community cloud
  - Public cloud
  - Hybrid cloud

CISE programs supporting Cloud

- CNS:
  - Computer Systems Research (CSR)
  - Networking Technology and Systems (NeTS)
  - Major Research Instrumentation (MRI)
  - Computer Research Infrastructure (CRI)
  - Cyber-Physical Systems (CPS)
  - Secure and Trustworthy Cyberspace (SaTC)
  - Future Internet Architecture (FIA)
- CCF:
  - Algorithmic Foundations (AF)
  - Communication and Information Foundations (CIF)
  - Software and Hardware Foundations (SHF)
  - Computing in the Cloud (CiC)
- IIS:
  - Information Integration and Informatics (III)
  - Robust Intelligence (RI)

Rough categorization of funded projects in cloud

- Computer Systems
- Computer Networks
- Security and Privacy
- Algorithms and data management
- Applications and software engineering
- Infrastructure
- Education

	CCF	CNS	IIS	CISE
FY 09-11	40	76	9	125
FY 12	14	63	4	81

Rough categorization of funded projects in cloud

Next 13 slides contain projects that fall into these seven categories

- Four per category
- Chosen to demonstrate breadth of topics
- Does not represent a “fair” sample

Computer Systems

Broad set of topics, including addressing the huge and growing energy cost of data centers (now above $4.5B/year and doubling every 5 years).

- 1017127 CSR: Small: A Server-Centric Approach to Data Center Networks. Songwu Lu, UCLA

  Moves function from network switches to servers, and uses low-cost COTS mini-switches for interconnection. Doing so provide the foundation for enhancing scalability, inter-server capacity, and fault tolerance.

- 1012070 CSR: Large: Storage Class Memory Architecture for Energy Efficient Data Centers. Bruce Childers, University of Pittsburgh

  Power consumption in data centers is shifting from processor to main memory. Rather than relying solely on DRAM, this project constructs a high-capacity, energy-efficient memory system for virtualized computer servers with a new Storage Class Memory Architecture that incorporates multiple memory technologies such as DRAM, Phase-change memory (PRAM) and Flash.

Computer Systems

- 1217597 CSR: Small: Bringing Predictable Low Latency and Strong Consistency to Data Center Services. Steve Gribble, University of Washington

  Exploring techniques for constructing data center services that provide low latency tails. High latency tails now result from, e.g., rare events in distributed storage systems, OS and transport protocols, and storage devices. By bringing down tail latency, both average latency will be improved and protocols that provide strong data consistency more practical.

- 1149703 CAREER: 3D Stacked Systems for Energy-Efficient Computing: Innovative Strategies in Modeling and Runtime Management. Ayse Coskun, Boston University

  Make 3D stacked systems effective agents for attaining low-power, high-throughput computing in both embedded and large-scale computing domains. Specific directions are: (1) develop a widely applicable methodology for jointly analyzing the performance, energy, and temperature characteristics of 3D multi-core systems; (2) design runtime management policies to maximize performance under thermal or power constraints; (3) optimizing liquid-cooled 3D systems to push the performance bounds while maintaining reliable and low-energy operation.

Computer Networks

Rethinking and redesigning the systems underlying the Internet and the internal networks that are parts of datacenters.

- 0904729 NeTS: Medium: A SCAFFOLD for Service Centric Networking. Mike Freedman, Princeton
  - Proposes a new network architecture, SCAFFOLD, that treats service-level objects as first-class citizens and explores a tighter coupling between object-based naming and routing. System components include programmable routers/switches, resolution services for object-based lookup and forwarding, and integrated end-hosts.
- 0917339 NeTS: Small: Topology Switching for Data Centers and the Clouds Above. Ken Yocum, UC San Diego.
  - Cloud providers must run a diverse set of client applications, each with potentially different networking demands, on shared data-center facilities. Traditionally, a datacenter network is configured to use the same routing process to choose the "best" route for each flow in a datacenter, regardless of the application. But, routing along a single tree leaves many paths unused, sacrificing potential gains in reliability, isolation and performance. Topology switching, in contrast, allows applications to create custom routing systems within a data center; they can configure multiple logical topologies that, together, are tailored to their reliability and performance requirements.

Computer Networks

- 1162088 NeTS: Medium: Collaborative Research: Optimizing Network Support for Cloud Services: From Short-Term Measurements to Long-Term Planning. Nick Feamster, UMd and Jenn Rexford, Princeton.
  - Designing, implementing, deploying, and evaluating practical techniques that allow OSPs to perform content and network routing, as well as make longer-term placement decisions, based on timely and accurate information about end-to-end performance and transit costs. Project tasks include (1) designing performance-measurement techniques and conduct measurement-driven studies of OSP traffic management; (2) designing, modeling, and prototyping protocols for joint optimization of content and network routing, and traffic management within an OSP backbone; and (3) driving long-term planning of server placement and ISP peer selection based on models of transit costs.
- 1219116 NeTS: Small: WaveCube: A Scalable, Fault-Tolerant, High-Performance Optical Data Center Architecture. Yan Chen, Northwestern.
  - Preventing localized bottlenecks in datacenter networks can be achieved by providing rapid (microseconds to millisecond) on-demand provisioning of optical links between communication hotspots in the datacenter. This project develops WaveCube, that does this. Optical wavelength selective switch technology is used to achieve cost savings while implementing multipathing and dynamic bandwidth scheduling for improved performance. The WaveCube topology is a multi-dimension cube with fiber carrying multiple wavelengths (Wavelength Division Multiplexing or WDM). WDM dramatically cuts down on the number of fibers needed in a datacenter but gives rise to a wavelength assignment problem, which they undertake as well.

Security and Privacy

Privacy, trustworthiness of cloud providers, and protecting cloud providers from threats.

- 0909980 TC: Large: Collaborative Research: Trustworthy Virtual Cloud Computing. Anna Yu, NC Agricultural & Technical State; Mike Reiter, UNC Chapel Hill; Jeff Chase, Duke; Peng Ning, NCSU
  - Envisions a new security architecture that harnesses new capabilities such as built-in out-of-band system access, processor and hardware support for trusted computing, and out-of-box examination by hypervisors. This project focuses on key research issues following this security architecture, including new security services that enhance the trustworthiness of virtual cloud computing, protection of management infrastructure against malicious workloads, and protection of hosted workloads from potentially malicious management infrastructure.
- 1017782 TC: Small: Reining in Side-Channel Information Leaks in the Software-as-a-Service Era. Xiao Feng Wang, Indiana University
  - A web application is a two-part program, with its components deployed both in the browser and in the web server. The interactions between the two reveal the program's internal states to any observer of the communication stream simply through the pattern of packet lengths and the timing of interactions, even when the stream is entirely encrypted. This research reveals that such "side-channel" information leaks are both fundamental and common: a number of popular web applications are found to disclose highly sensitive user data such as one's family income, health profile, investments and more. This research will develop an in-depth understanding of web applications' side channel vulnerabilities, particularly the design features and domain knowledge that lead to side-channel leaks. Based upon this understanding, new technologies will be developed to facilitate the detection and mitigation of such threats during the development and operation of web applications.

Security and Privacy

- 1223710 TWC: Phase: Small: Software Cruising for System Security. Dinghao Wu, Penn State
  - Develop an innovative security monitoring technology, called *Software Cruising*, to explore the use of multicore architectures for non-blocking concurrent security monitoring using lock-free data structures and algorithms. Applications include heap buffer integrity checking, kernel memory cruising, data structure and object invariant checking, rootkit detection, and information provenance and flow checking.
- 1222748 TWC: Small: Auditing PII in the Cloud with CloudFence. Angelos Keromytis, Columbia
  - Prototypes and evaluates *CloudFence*, a framework that allows users to independently audit the treatment of private data by third-party online services. Investigates novel techniques for conducting fine-grained tracking of information of interest (defined by the user of the cloud service in a flexible, context-sensitive manner) toward (a) providing increased transparency to end users of the handling of their information by the cloud, and (b) enabling the periodic or continuous auditing of such handling, either by the user or an agent acting on the user's behalf.

Algorithms and data management

- 1114809 AF: Small: Collaborative Research: Algorithms for Reallocation Problems. Michael Bender, Stony Brook; Martin Farach-Colton, Rutgers
  - (Online problems) Formalizes how computational resources can be reallocated efficiently when there is some cost for the reallocation. Investigates the algorithmic complexity of reallocation problems arising from large-scale systems design, and especially reallocation algorithms that are universal with respect to reallocation-cost functions. The research plan addresses problems in memory and storage reallocation, scaleout and sharding in storage systems, reallocation for dynamic combinatorial and geometric structures (finite-element meshes, metric spaces), and classical online allocation problems (scheduling and bin packing).
- 1217648 SHF: AF: Small: Collaborative Research: RESAR: Robust, Efficient, Scalable, Autonomous Reliable Storage for the Cloud. Ahmed Amer, Santa Clara; Darrell Long, UC Santa Cruz
  - Tackles the problem of building ever-larger data stores, and offers a novel approach to reducing the energy impact of such increases in scale while allowing easier management and adaptation of the system as it ages. RESAR offers a means to gracefully adapt a storage system to offer increased reliability or performance as demanded by the systems' age or administrator's requirements. RESAR uses novel erasure codes that allow faster layout restructuring, while offering increased scalability, and improved reliability over competing schemes. RESAR also allows for restructuring on the fly, and so has the added benefit of being complementary to data relocation tasks necessary for routine maintenance and optimization.

Algorithms and data management

- 1217890 AF: Small: Efficient Data Management Algorithms. Samir Khuller, U Md
  - Focuses on the development of algorithms that manage data storage for processing, with energy efficiency as the primary consideration. Much of the prior scheduling literature assumes a job-centric perspective -- algorithms are developed to optimize tardiness, completion time, makespan, etc. In contrast, this work is motivated by a system-centric view in which utilizing resources in an efficient way is of the utmost priority, subject to individual jobs being completed in a satisfactory manner. Such efficiencies are primarily manifested in the form of the energy cost incurred by the system. The main focus is on data of all types, ranging from multimedia data stored on a collection of disks to data collected and stored in a distributed storage system.
- 0952692 CAREER: Foundations and Extensions of Public Key Cryptography. Brent Waters, UT Austin
  - Encryption protects data stored at third party service. Unfortunately, traditional public key encryption a party encrypts data to a single known user. While this functionality is useful for applications such as encrypted email and establishing secure web sessions, it lacks the expressiveness needed for more advanced data sharing. This project lays the foundations for an new vision for encryption called Functional Encryption. Instead of encrypting to individual users, in a functional encryption system, one can embed any access predicate f() into the ciphertext itself. Functional encryption simultaneously renders completely general functionality and its data access is self-enforcing: it requires no trusted mediator. The goal is to create a system where the encrypting party can specify any access predicate over a recipient's credentials (i.e. f can be any Turing Machine). Functional encryption for any predicate opens up a world of possibilities for data sharing; one could encrypt an image such that the access function f encodes an image recognition program allowing only people in the picture to view it.

Applications and software engineering

- 1047753 RAPID: Collaborative Research: Cloud Environmental Analysis and Relief. Christopher White, VA Poly; Jeffrey Gray, U. Alabama Tuscaloosa; Doug Schmidt, Vanderbilt
  - Develops a cloud-supported mobile CPS application enabling community members to contribute via sensor deployments and direct recording of events and ecological impacts of the Gulf oil spill, such as fish and bird kills. Exploits the availability of smartphones and cloud computing infrastructures that enable collecting and aggregating data from mobile applications. The goal is to develop a scientific basis for managing the quality-of-service, user coordination, sensor data dissemination, and validation issues that arise in mobile CPS disaster monitoring applications.
- 1048125 Collaborative Research: CiC (SEA): Using the Cloud to Model and Manage Large Watershed Systems. Jonathan Goodall, University of South Carolina; Marty Humphrey, U Va
  - Leverages cloud computing for modeling and managing large watershed systems. Creates a cloud-enabled hydrologic model; creates generic cloud-based data processing workflows needed by hydrologic models and other domains; applies the hydrologic model and data processing workflows to model a large watershed system at detail and scale to address research questions related to quantifying impacts of climate change on water resources.

Applications and software engineering

- 1249722 EAGER: Scaling the Preprocessor and Making it More Intelligent in Deterministic Database Systems. Daniel Abadi, Yale
  - In general, applications that require elastic scalability are forced to abandon ACID guarantees. To overcome this problem we: (1) Implement a database system using an innovative deterministic architecture that guarantees that nondeterministic processing events will not affect database state, (2) Leverage this new architecture to avoid "commit protocols" for distributed transactions in a cluster, (3) Design a scalable preprocessor for the deterministic database that collects, analyzes, and dispatches transactions to the database cluster in order to further improve scalability, and (4) Develop a new lazy transaction evaluation approach in order to spread out load and avoid damaging effects of database load spikes.
- 1258741 RI: Small: GraphLab 2: An Abstraction and System for Large-Scale Parallel Machine Learning on Natural Graphs. Carlos Guestrin, University of Washington
  - With the growth of the Web and improvements in data collection technology in Science, datasets have been rapidly increasing in size and complexity, necessitating comparable scaling of machine learning algorithms. However, designing and implementing efficient parallel machine learning algorithms is challenging and time consuming. To address this challenge, we recently released GraphLab, a framework providing an expressive and efficient high-level abstraction satisfying the needs of a broad range of machine learning algorithms. This project develops GraphLab 2 which addresses the much more challenging online and distributed settings, tackling: 1) Cloud-based distributed machine learning. 2) Natural graphs, with very high-degree vertices that are not amenable to graph partitioning methods. 3) Online tasks, where data and queries are streaming over time. 4) Off-core computation, since huge problems may not fit into memory, even across the cloud.

Infrastructure

- 0958514 II-NEW: GreenIT: Testbeds for Real-time Data Center and Platform Energy and Thermal Management Karsten Schwan, Ga Tech
  - A measurement-based, GreenIT testbed for energy efficient IT. The testbed will operate at data center scale, coupled with ongoing efforts provide both (1) a large-scale, commodity IT infrastructure, i.e., racks of machines and (2) new facilities used for a multi-site collaboration in cloud computing. Commodity equipment and instrumentation are housed in the CEETHERM lab, in Mechanical Engineering at Georgia Tech, which offers dynamically controllable air cooling capabilities.
- 1040666 MRI: Development of a Fully Instrumented Self-Sensing and Self-Regulating Data Center. Kanad Ghose, SUNY Binghamton
  - An instrument for next generation data centers where energy consumption and management is treated as a first class resource (together with CPU, communications, etc) that needs to be scheduled and managed explicitly. Specifically involves: 1) Experimental, scaled down data center of Linux servers with dynamic cooling facilities, modified kernels. and scheduling components; 2) Large number of temperature and airflow sensors and power meters, along with software instrumentation; 3) Floor plenum based chilled air cooling system used to provide nominal cooling and remotely controlled computer room air conditioners (CRACs) that provide a quickly adjustable and directed cooling facility; 4) Software components for the facility that permit research into the development of a wide variety of techniques, both traditional and innovative, for improving the data center energy efficiency; 5)Generation and recording live data on workload levels, power consumption, temperature, and airflow distributions.

Infrastructure

- 1205699 II-NEW: Collaborative Research: Image Processing Cloud (IPC): A Domain-Specific Cloud Computing Infrastructure for Research and Education. Lei Huang, Prairie View A&M; Xiaoming Li, U Del; Yonghong Yan, U Houston
  - An integrated image processing research environment within a computing Cloud infrastructure that includes: 1) an open image processing computing Cloud to support researchers and students to conduct image processing research, share knowledge and research results, and stimulate education materials among Prairie View A&M University, University of Houston, and University of Delaware; 2) a high-level domain specific language designed to provide an abstract and productive programming model for image processing applications; 3) a general compiler optimization framework with the capability to tune into image processing applications at various levels starting from high-level representations to low level transformations in the Cloud environment.
- 1040123 MRI: Development of a Virtual Cloud Computing Infrastructure. Larry Peterson, Princeton; Michael Freedman, Vivek Pai, Jen Rexford.
  - VICCI is a programmable cloud-computing research testbed that enables a broad research agenda in the design of network systems that requires both multiple point-of-presence and significant processing/storage capabilities on the sites. VICCI has a point-of-presence at Princeton, Ga Tech, Stanford, and U Washington, along with international clusters in Europe and Japan. VICCI enables research in 1) Building block services (addressing issues of replication, consistency, fault-tolerance, scalable performance, object location, and migration) designed to be used by other cloud applications, 2) Developing new cloud programming models designed for targeted application domains, and 3) Studying cross-cutting issues at the foundation of the cloud’s design and how to build a trusted cloud platform that ensures confidentiality and integrity of computations that are outsourced to the cloud.

Education

Education is a frequent broader impact. These three projects, however, focus on undergraduate education.

- 1048711 A Curriculum Initiative on Parallel and Distributed Computing - Toward Core Topics for Undergraduates. Sushil Prasad, Ga Tech
  - Bringing all stakeholder experts working together and periodically providing guidance on restructuring standard curriculum across various courses and modules related to parallel and distributed computing.
- 1156574 REU Site: Enhancing Undergraduate Experience in Next Generation Networking Technologies. Chiu Tan, Temple University
  - Co-funded by the Department of Defense. REU students will participate in research projects that involve mobile computing, wireless communication, and cloud computing. Each summer a cohort of undergraduate students will work with Temple faculty in small groups to design, implement, and evaluate a research project. The students will have access to the WiMAX testbed that covers much of the metropolitan downtown area of Philadelphia, an internal supercomputer and a cloud cluster.
- 1005153 REU SITE: Computer Systems Research in High Performance Cloud Computing Environments. Michael Lewis, SUNY Binghamton
  - Students immerse themselves in intensive high-quality research projects directed by one or more CS Faculty Research Mentors at Binghamton, for a ten week period in the summer. As part of the REU experience, students write research-style papers and make public conference-style presentations in Binghamton's CS Department. In preparing to do so, the students make significant contributions to the department's research programs. Departmental research areas of expertise (and therefore potential areas for REU student projects) include computer architecture, energy-aware computing, graphics and image computing, virtualization, Internet search and information retrieval, security, cloud and grid computing, circuit design, algorithmic optimization, data mining, mobile and sensor networks, and more.

Dependability

For the FY 09-11 projects only:

Security Trust	Reliability Fault Tolerance Availability	Privacy	Attack Tolerance	Monitoring	Fault Models
26	22	7	3	2	2

Cloud Workshops

- Cryptography in the Cloud Workshops, August 2009 and August 2010.
- Science of Cloud Computing PI Meeting, March 2011.
- Security of Cloud Computing Services and Systems Workshop, March 2012.
- Workshop on Instrumentation Needs of Computer and Information Science and Engineering (INCISE2), July 2012.
- Workshop on Computing Clouds for Cyber-physical Systems, March 2013.

Exploiting Parallelism and Scalability

XPS: A new solicitation that addresses issues that arise from current and future architectural challenges

21st Century Promise

- ICT promises much…
  - Data-centric personalized health care
  - Computation-driven scientific discovery
  - Human network analysis
  - Much more: known & unknown

- Characterized by
  - Big Data
  - Always Online
  - Secure/Private
  - …

What will be the enablers of future cost-performance gains?

From *21st Century Computer Architecture: A Community White Paper* (Mark D. Hill, coordinator)
NSF outbrief 6/22/12

Technology's Challenges 2/2

Late 20th Century	The New Reality
Moore's Law — 2× transistors/chip	**Transistor count still 2× but …**
Dennard Scaling —~constant power/chip	**Gone.** Can't repeatedly double power/chip
Modest (hidden) transistor unreliability	**Increasing transistor unreliability** can't be hidden
Focus on computation over communication	**Communication (energy)** more expensive than computation
One-time costs amortized via mass market	**One-time cost** much worse & want **specialized** platforms

21st Century Computer Architecture

20th Century	21st Century	
Single-chip in stand-alone computer	**Architecture as Infrastructure:** Spanning sensors to clouds	**Cross-Cutting:** Break current layers with new interfaces
Performance via invisible instr.-level parallelism	**Energy First** Parallelism, specialization, cross-layer design	
Predictable technologies: CMOS, DRAM, & disks	**New technologies** (non-volatile memory, near-threshold, 3D, photonics, …) Rethink: memory & storage, reliability, communication	

From *21st Century Computer Architecture: A Community White Paper* (Mark D. Hill, coordinator) NSF outbrief 6/22/12

Spanning Sensors to Clouds

- Beyond a chip in a generic computer
- Pillar of 21st century societal infrastructure.
  - Computation in context (sensor, mobile, …, data center)
  - Systems often large & distributed
  - Communication issues can dominate computation
  - Goals beyond performance (battery life, form factor)

- Non-exhaustive list of opportunities
  - Reliable sensors harvesting (intermittent) energy
  - Smart phones to Star Trek's medical "tricorder"
  - Cloud infrastructure suitable for both "Big Data" streams & low-latency qualify-of-service with stragglers
  - Analysis & design tools that scale

From *21st Century Computer Architecture: A Community White Paper* (Mark D Hill, coordinator)
NSF outbrief 6/22/12

Exploiting Parallelism and Scalability (XPS)

NSF 13-507

Focus areas

Foundational Principles

New models guiding parallel algorithm design on diverse platforms

Optimization for resources (energy, bandwidth, memory hierarchy)

New languages designed for parallelism

Scalable Distributed Architectures

Highly scalable and parallel architectures for people and things connected everywhere

Runtime platforms and virtualization tools

Cross-layer Approaches

Re-thinking/re-designing the hardware and software stack

Coordination across all layers

New runtime systems and system architectures

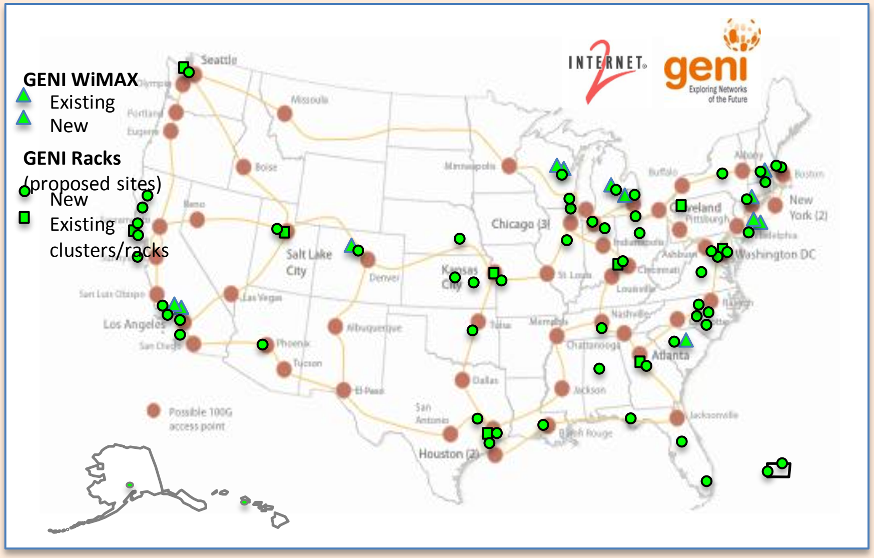
Domain-specific Design

Exploiting domain knowledge to improve programmability and performance

Program synthesis tools

Hardware-software co-design

Mid-scale Research Infrastructure



Advancing networking, distributed systems, cloud computing and cybersecurity research through experimentation at scale

Global Environment for Networking Innovations (GENI)

- A virtual laboratory for exploring future internets at-scale, now taking shape in prototype form across the U.S.
- Key GENI concepts:
  - Slices & deep programmability
  - Federation and enabling "at scale" experiments

US Ignite

- Launched June 14, 2012 at the White House
- NSF leadership
  - Leveraging GENI investments
  - Stitching together testbeds and network resources across the country
  - Jumpstarting gigabit public sector application development
- Public Private Partnership
  - Bringing industry and foundations into he mix

The future?

Discussion

- Cloud research is supported by several programs.
- Current issues include communications, power, scheduling, handling massive data, privacy, trustworthiness, programming, and specific applications.
- Growing interest in infrastructure, new architectures, and broader application.