SwitchOn Workshop

Breakout - Clouds
Room 124

Edit this document: **http://tinyurl.com/switchoncloud**
Slides: **http://tinyurl.com/switchon-breakout**

Participants

José Fortes, UFL
Hermes Senger
Rogerio Lope
Heidi Morgan
Andy Bavier
Ram Krishnan
Sejun Song
Shaolei Ren
Malathi Veeraraghavan
Kate Keahey
Carlos Kamienski

Objectives

Challenges and Important Research Issues

- Cloud security: new model, policies, public versus private clouds, HIPA compliant -- is our data going to be private
- Computer security
    - policy what is secure in cloud?
      cloud computing changes model - devise different attacks
      exploitation of the model, policy questions
      misclassification in different models, fact public vs private cloud
- - security changes in different types of cloud
- or infrastructure vs app or services type

-
- ongoing discussion about data in cloud
- medical data in cloud dont match regulations in europe
- how to know if it is compliant? compliance updated for clouds
-
- privacy computing
- obsfuscated computation for the cloud
- online storages are a classical exampe
- multitenancy - most difficult to secure infrastructure
- abstraction for the cloud - layer to provide VM and VNets, deserve more attention
    - platform as service, seem to be a category - what secure look in that space?
    - tool biil - frameworks to build they're own infra as service
    - federating cloud, touch in the area as well
    - how to map different systems, federation is a issue
    - cloud is just means to provide resources, the model is same for sysadmin, soft developers

- build a system, with sensors, build a cyberinfrastructure that coulb ne programmed in different ways, whole combination of programmable networks, end-to-end combinations of programmable networks, storages, computing.
-
- 1 break through - data streaming detect earthquakes with smartphones, fitbit, people instrumenting cities, buy a drone for hundred bucks, possible to do an experiment, but no designed this way. imagine all data feeds available, can ask dynamic questions in that particular space, IF cannot be processed the data, answers does not count
- 2 - programability of the infrastructure to reprogram the infra to achieve SLA, feedback is important, flood is coming, and need to program.

data processing mechanisms, mechanisms to squeeze data, and heterogeneous datasets, meta-data is key. LHC has million of sensors, but in the wild, need to be calibrated to work together. No one put that in the sensors as default. unlikely to have, depends on projects, for example, biosciences the way they collect data is very diverse.

also needs to correlate data, some baseline is important for example sismic, simulate earquake data, what is changing? New astronomy project in Chile, need to every night to see what is change? need to put heterogeneous data layers together.

ocean science - real time is few days versus erathquake science in seconds. need to to end-of-end QoS in all layers, networks, storage, computing. Software Defined Networking can help to cope this problem.

climate appication that need more than thousand cores, communication is the bottleneck. particular apps that have tight requirement, MPI apps. New inovattive network

approaches for these problems. Infiniband has more 100Gbps, is not the communication, old fat stack in inifiband still there is issue, crossing layers.

How to make HPC run on clouds - light weight hypervisor, low noise, little overhead, infiniband, all sort of approaches, reliability issues in this space, can we leverage virtualization for that. In this space, people don't do virtualization in HPC in clouds.

single data centers, proprietary data centers in blue gene, interconnects cannot run VMs, even whole machine communication is a problem. part of larger problem, what is HPC cloud? Cloud means certain model, on-demand availabilty.

2 different areas of cloud - business users and scientific clouds. requirements are totally different. Could the same cloud provide services for both? NSF is working in Clouds, the way is desgined is very different, scientific clouds lots of processing, and volume different than business.

current public clouds are not engineered for HPC. very critical if the anwser is to have 2 divergent models. extremely cost to support divergent model. how to design a model that could be good for both.

how to provide services to society? Can cloud could provide services wherever you are, new model. 1000 cores, and then dont need anymore. People will need, services for society vs services for scientists? Amazon have different offering, depending on different demands.

cloud is a black hole, work on multiclouds, there are research issues on coordination, orchestrating, on different clouds could be a basis for collaboration.

management - optimization and programmability, we are still thinking in data plane, not control and management. It is issue in between clouds, old problems of grid come back. specific ideas of management, usual question, single single, load balance, data management. Muti-tenant issues become hard, software defined systems, multi-tentant and every tenant wants a different thing. Management is difficult. Trying to achieve green in network management is hard.

their intersect - aspects of management and deisgn solutions that overlap on conserve power.

security and HPC - "here come the killer microprocessors", big domain on microprocessing and HPC, then HPC ran out of business. Security - cloud is not secure, cloud has more people working in security and getting better. Eventually getting better in security than any small, medium enterprise. Same thing in terms of interconnections, software updates

come from linux. Pushing the frontiers even if now the cloud is not providing HPC. clouds are going to make better.

black swan problem - the kind of computing, small fraction thar scientist need, they are way out there. Many HPC companies died but Cray and IBM are growing. Climate application is way out there. Create computing to advance science. track 2 different issues - intra-network design for HPC, network design intra and between clouds.

- how to cope with mobile and clouds?

Existing Major Projects US and Brazil (not necessarily joint)

US: existing: nsf cloud, chameleon, cloudlab. CNS is funding cloud research, OpenScience datacloud partnership for international research and education. Project in MIT - electronic textbook, allows kids to run simulations in clouds - intuitive ideas and have “rooms” to interact.

SwitchOn, IRNC.

BR: existing: RNP cloud initiative and Cloud@USP, UNESP HPC - education cloud project - Intel is partner - provide resources for schools in state of Sao Paulo, IaaS for high schools in SP. Coordinated call between Brazil and Europe in 3rd Call in cloud theme, set up the partnerships - cloud research. 1st Call (5 projects) - secure future cloud, 2nd (4 projects) - more on scientific cloud, 3rd call now cloud with security aspects.

There is a gap between research questions and building infrastructure. However, we need infrastructure to research.

-

Existing and Future Collaborations

existing collaboration between BR and US - site shows in switchon website (**switchon.ampath.net)**, list of collaborations, subsets on cloud specific, most of them are GENI related. UFSCar and FIU - congestion control in cloud, UNICAMP and UVirginia - Hadoop schedule and SDN controllers. Jose Fortes - sharing data, eventually use cyberinfrastructure, domain of sharing data is essential and networks is already in place, but middleware may not be there, it could be opportunity. Network needs case studies, 100Gbps between SP and Miami, need more case studies once is provisioned. LHC is big drive. Other

geographic constrined projects for example Chile telescopes, decimeter band antennas for radio-telescope the SUn, etc.

future collaborations: UFABC and U Chicago - smart cities collaboration.

in terms of existing and future collaboration - planetlab has a long standing research relations, it could be more. projects that are successful and make as a foundation could be more useful to spawn collaboration.

what types of programs could be complementary? testbeds? develop education materials? research projects that could use those testbeds. Data management, system layer research in one place, and data mining research in another place, neither one of them solve the full problem. data algorithms and infrastructure.

Potentials for large-scale collaborations

- Large Scale - need to init as small scale. Seed funding (Newton) funding - scientific meetings, fund british people. This is a normal path.
    - ARPANET - 3 or 4 nodes, slowing grow, 10 years
    - L2 services - started CCNIE grants, Metcalfe - value of network in N^2. new control plane. create a seed and then large scale.
    - Need to deploy.
    - Preceding effort, sustainability - cannot be large scale in the first year, build up - CCNIE - 2 to 3 year, PIRE is 5 years. Major projects building a new telescope - 10 years.

- Barriers:
    - typically hard to get the international funding
    - in US, institution is funded in the grant,
    - everybody is obligated to work on the deliverables
    - obligation to show results, but if there is not international relation
    - difficult to find reciprocal funding together.
    - without a coordinated call this is a barrier, NSF and CNPq, FAPESP could do these type of calls.
        - Student exchange program (need funds on both sides)
        - Some use cases, of sending france students that France is paying, France needs to have abroad experience as a requirement

- We should have a program exchange graduate student (funded on both sides)
- need to provide the value that students will get latter in exchange, in US difficult - students want to go to industry, to do internships.

- faculty advisor is not in collaboration, doesn't support in what they are doing.

- Culturally, time zone is not critically different.
- Miami - cultural ties. Language.
- Global cyber-bridge to do videoconferencing - difficult Japan, US, have to 7am.

- There has to do a complementary expertise.
    - complementary expertise, problem itself is international in nature.
    - security is international topic
    - funded to do security - is going to do for a long time, US or Brazil is not secure, there is a good complementary. People works in both countries, the topic is really broad
    - unlikely in BR and US will research in different thing.
    - export control in crypt algorithms - domain that could come with new ideas
    - physical miltary - some countries share military expertise more than others.
    - why should a project should be cross-country?
        example, high energy phisics
    - there is only some many people that will analyse the data
    - large problems in physics are not solvable without so many people