The City University of New York

University of Missouri

Multi-Cloud Performance and Security-driven Brokering for Bioinformatics Workflows

Saptarshi Debroy, PhD.

Assistant Professor in Computer Science, Hunter College of CUNY

Doctoral Faculty, The Graduate Center of CUNY

Project Team Introduction

NSF Program: CC* (Campus Cyberinfrastructure)

Program Area: OAC (Office of Advanced Cyberinfrastructure)

Award Number: 1827177

Project Title: End-to-End Performance and Security Driven Federated Data-intensive Workflow Management

Prasad Calyam
Associate Professor
University of Missouri

Trupti Joshi
Assistant Professor
University of Missouri

Isa Jahnke
Associate Professor
University of Missouri

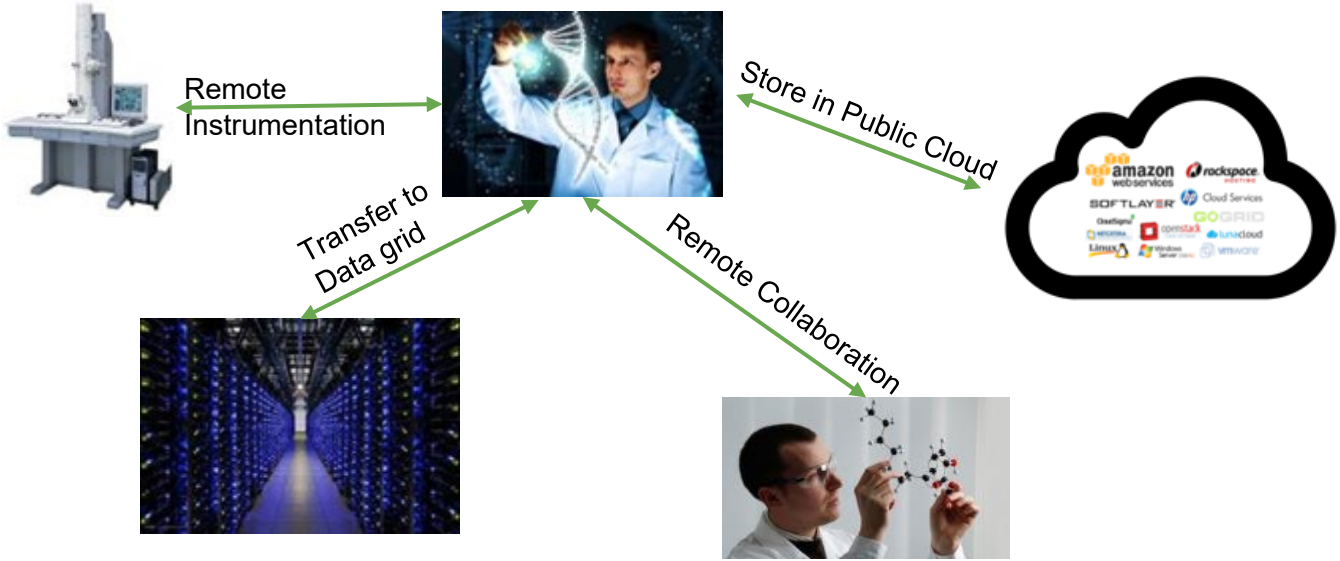
Saptarshi Debroy
Assistant Professor
CUNY – Hunter College

Timothy Middelkoop
Director, Research Computing
University of Missouri

Graduate Students:

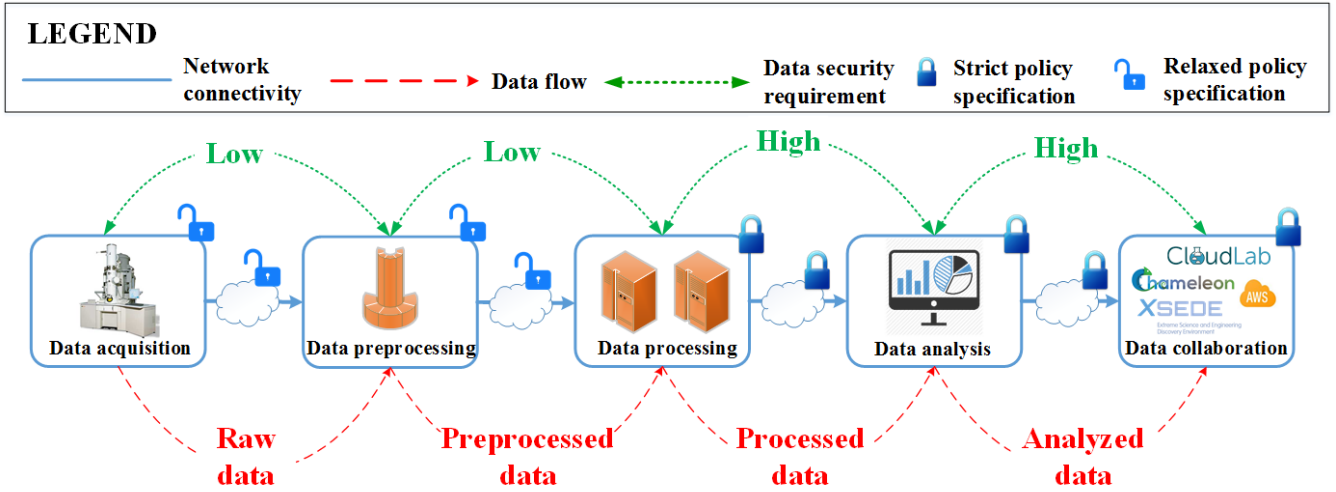
- Minh Nguyen (CUNY)
- Xiaojie Zhang (CUNY)
- Ashish Pandey (MU)
- Soumya Purohit (MU)
- Ramya Bhamidipati (MU)
- Mauro Lemus (MU)

Data Intensive Applications Today



- Science data-intensive applications require **on-demand** resources
- Motivates adoption of **hybrid cloud (private/public/community)** architectures
- Programmable technologies (e.g., SDN, OpenFlow, OpenStack) and **federated resources** (e.g., CyVerse, NSF Cloud, AWS) make such collaboration possible

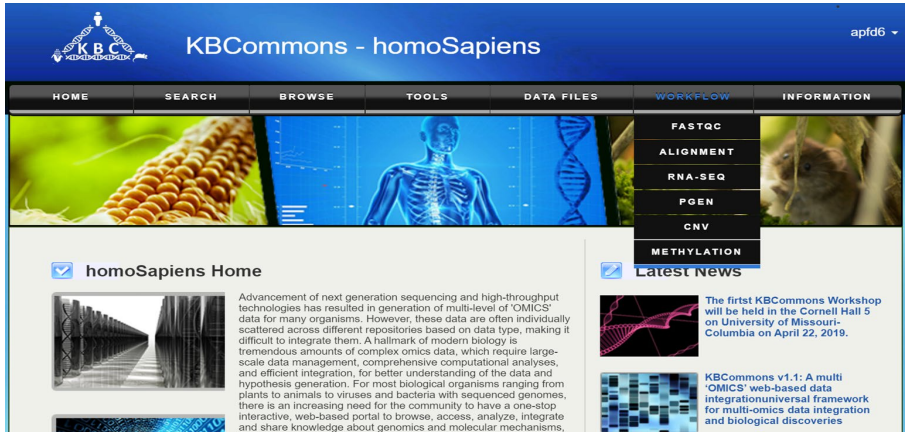
Security concerns for cross-domain resource use

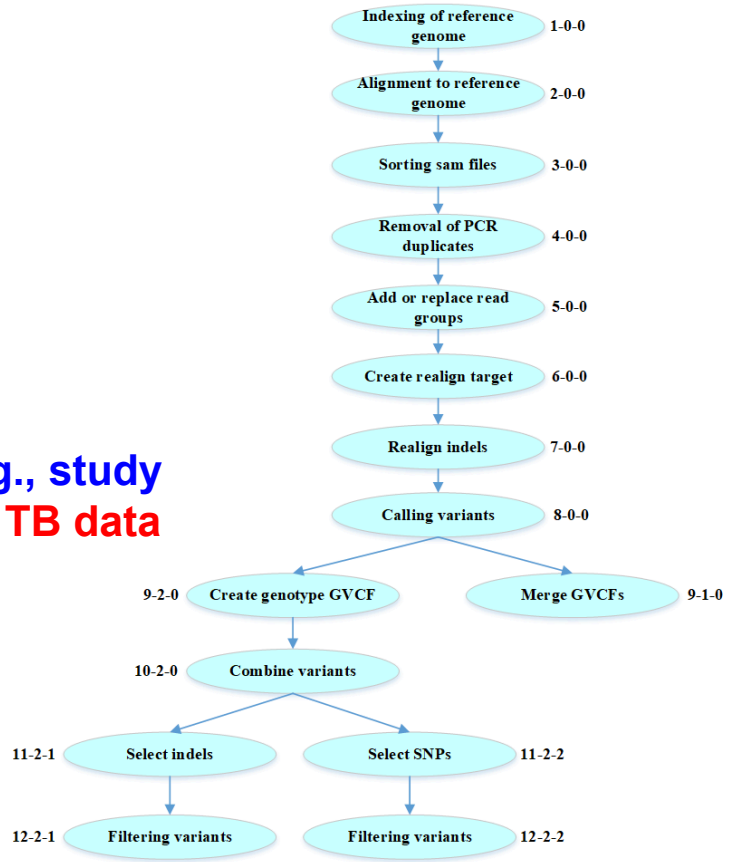


- Data has different life-cycle stages with varying security requirements
- Serious security concerns for data leaving campus or if resources are used across domains

Example of distrust at local institutions:
"*Division of IT strongly advises … to discourage if not prohibit … from using public cloud services for University related activities…*"
-- Division of IT, University of Missouri

KBCommons Science Gateway and SoyKB





Science gateway portal hosting bioinformatics workflows **e.g., study of genetic mutations in plants and organisms (involves TB data sets)**

- The multi-step workflows provide biological users with an avenue to analyze their datasets.
- Support needed to accommodate security levels to handle protected genomics data

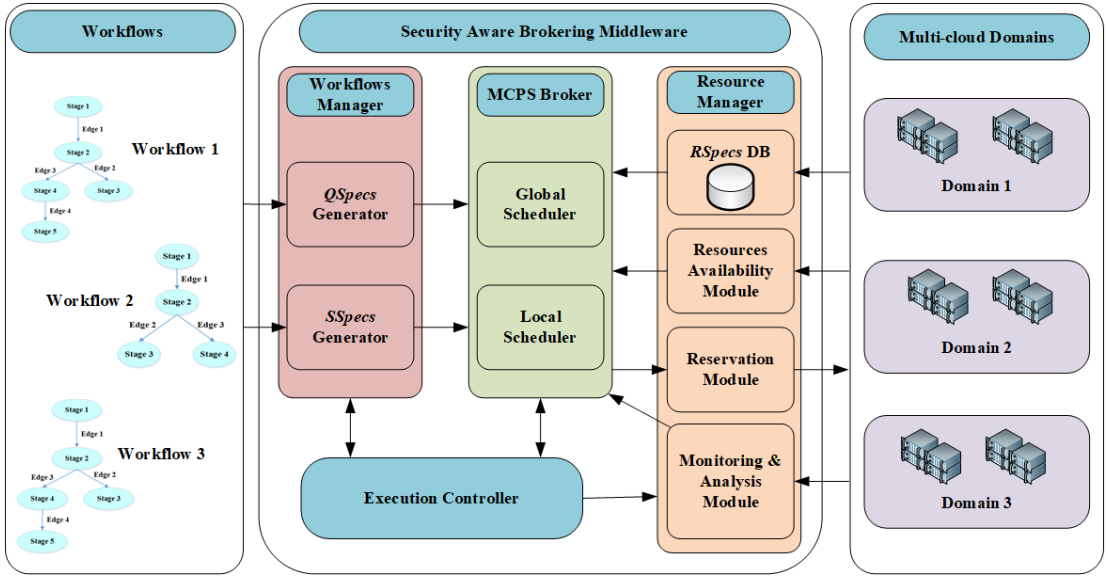
Security & Performance Inter-Conflicts Problem

We characterize the inter-conflicts problem as:

- "Friction" among user requirements and domain/ data source policies
- Strict security requirements of data custodians adherence may restrict performance
- Institutional resource policies may not satisfy security requirements

M. Nguyen, S. Debroy, P. Calyam, Z. Lyu, T. Joshi, "Security-aware Resource Brokering for Bioinformatics Workflows across Federated Multi-cloud Infrastructures", *ACM Conference on Distributed Computing and Networking (ICDCN)*, 2020.

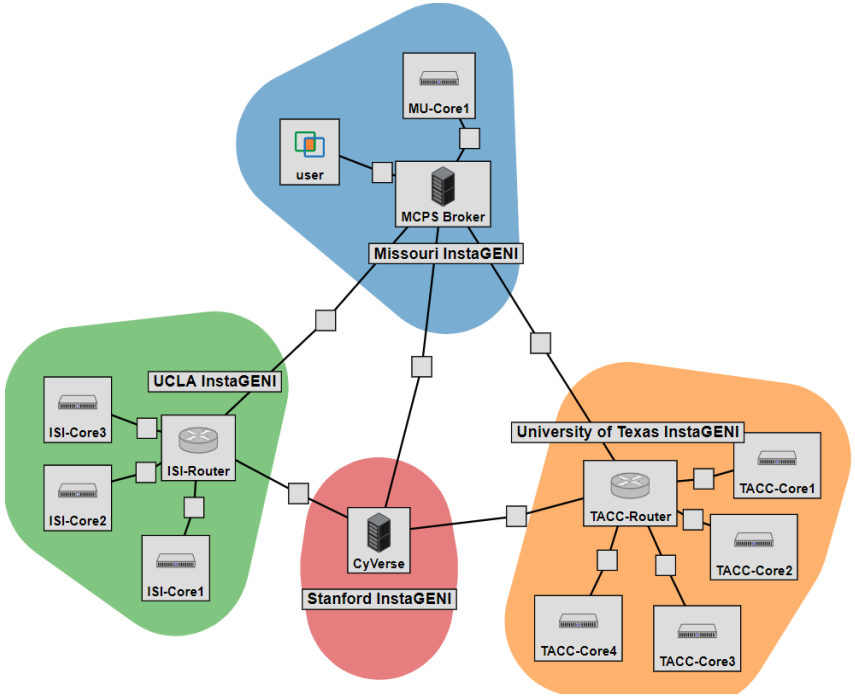
Performance and Security driven Resource Brokering



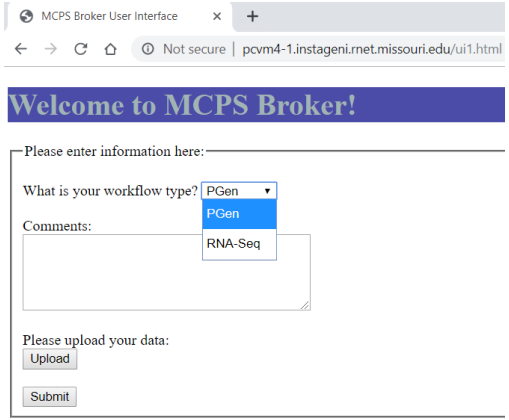
- Global scheduling algorithm allocates workflow DAG vertices to domains with *security and policy* satisfaction.
- Local scheduling algorithm chooses optimal computing core within the chosen domain for *performance* satisfaction.

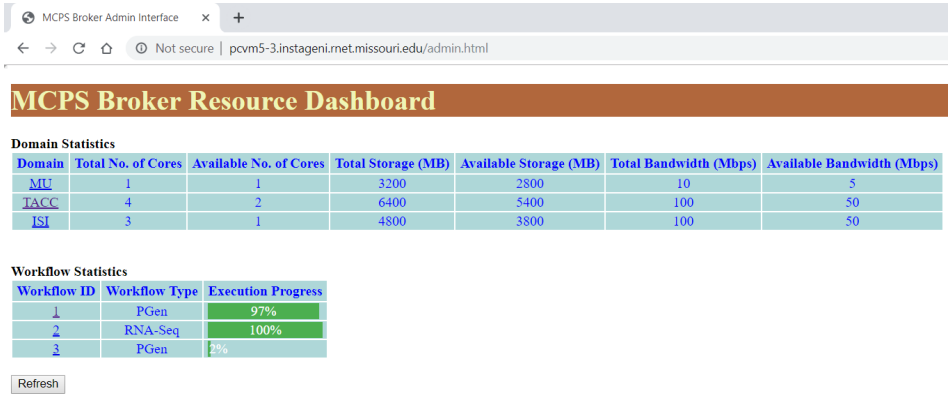
GENI Implementation

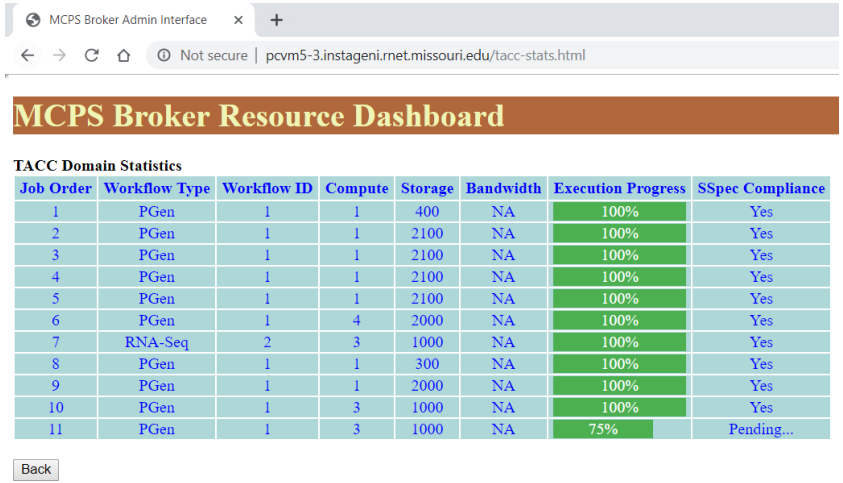
- GENI infrastructure based testbed
  - Multi-cloud resource domains approximately based on the real computing centers used for SoyKB workflows
- Compute capability and network bandwidth mismatches to mimic real-life SoyKB implementation.
- Replicates security policies of TACC, ISI, and MU domains as well as dynamic resource utilization levels.

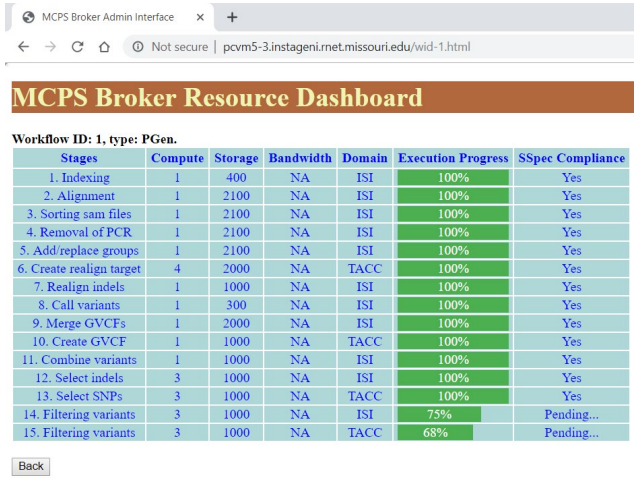


User Interfaces



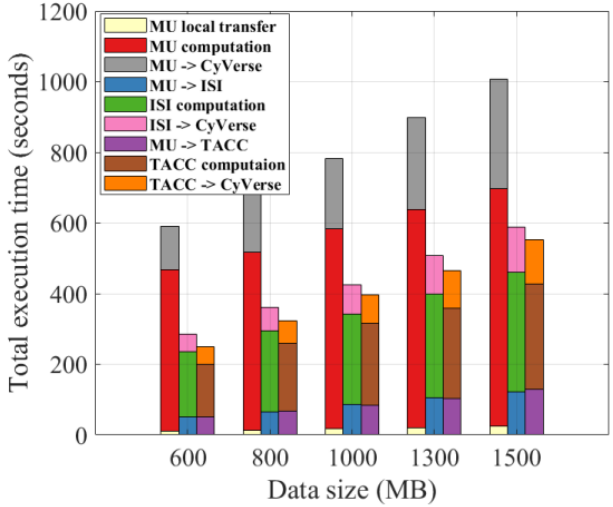


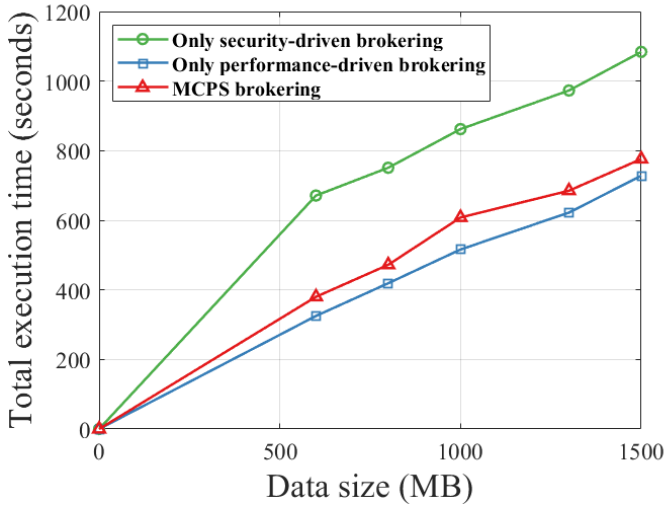




Performance Evaluation

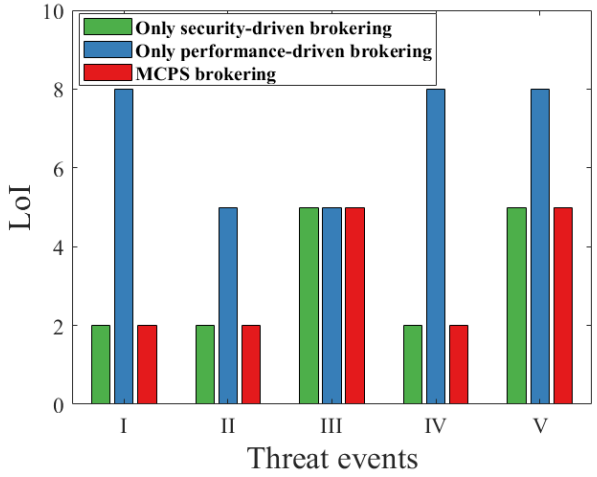
- Total workflow execution time comparison for PGen workflow
  - MCPS is almost as good as only performance-driven brokering (i.e., optimal)
- Total workflow execution time comparison for simpler RNA-Seq workflow
  - Remote computation is better

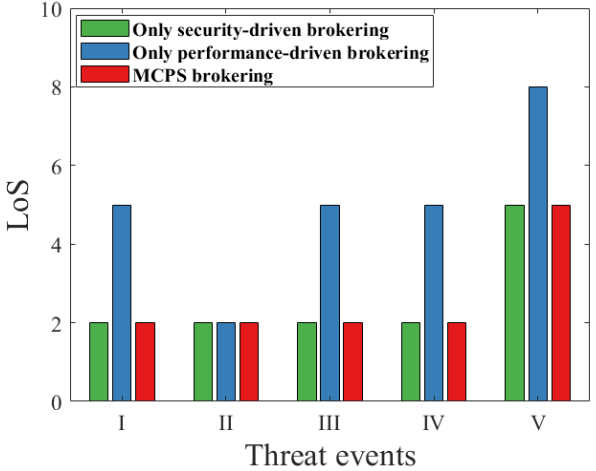


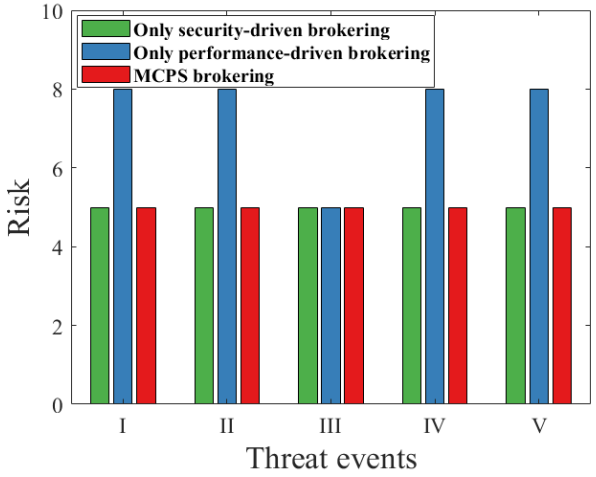


Threat Analysis and Security Assessment

- The security compliance comparison for PGen
  - Likelihood of attack success (LoS) and overall Risk are similar to security-driven brokering (i.e., optimal)







http://www.ontimeurb.net