WRENCH: A Simulation Workbench for Scientific Workflow Users, Developers, and Researchers

Henri Casanova[1], Rafael Ferreira da Silva[2], Frédéric Suter[3], Suraj Pandey[1], James Oeth[2], Ryan Tanaka[1]

[1]University of Hawai`i at Mānoa – Computer Science Department
[2]University of Southern California – Information Sciences Institute
[3]Centre National de la Recherche Scientifique – Institut National de Physique Nucléaire et de Physique des Particules


Henri Casanova
henric@hawaii.edu


Rafael Ferreira da Silva
rafsilva@isi.edu

WRENCH
Workflow Management System Simulation Workbench

MOTIVATION

- Scientific Workflows are key to advances in science and engineering
- Their executions are complex:
  - Workflow structures are large and can be configured in various ways
  - Workflow Management Systems (WMS) are large multi-component software systems that employ ranges of decision making algorithms
  - Workflow execution platforms are heterogeneous and diverse
- We need a strong **"experimental science"** approach to study these complex systems in a view to optimizing workflow executions

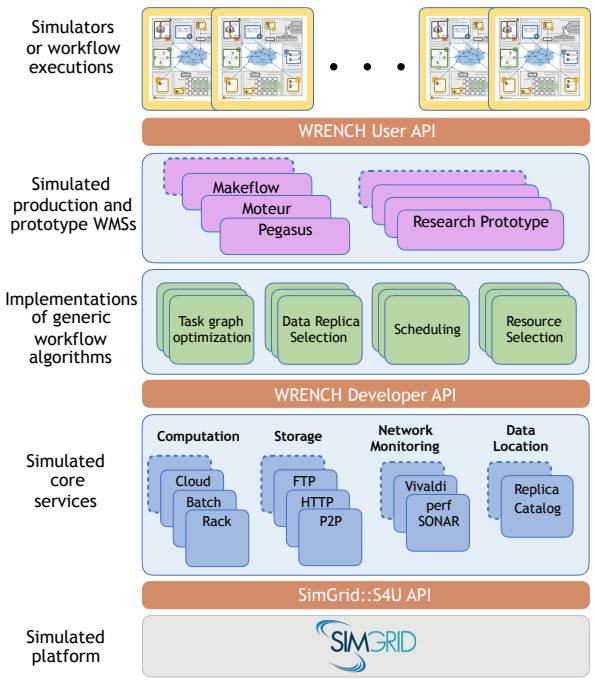
OBJECTIVE

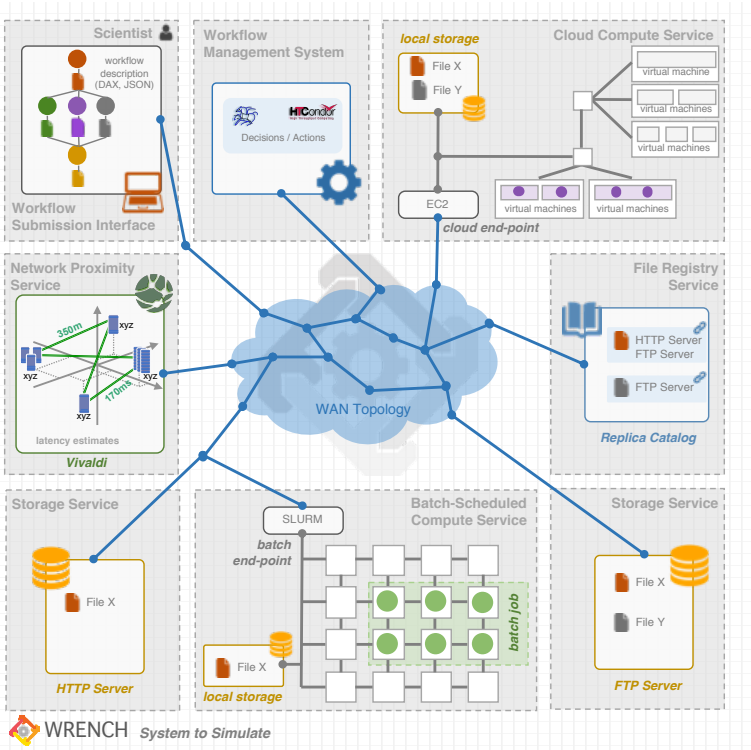
- Realize a **workflow execution simulation methodology** that has high simulation accuracy, low execution time, and low memory footprint
- This framework is to be used:
  - By workflow users to study workflow executions
  - By WMS developers to inform system and algorithm design decisions
  - By educators to teach distributed computing in the context of workflows

APPROACH

- Develop the WRENCH **"simulation workbench"**
- Develop WRENCH-based simulations of current production WMSs
- Develop WRENCH-based simulations of WMS research prototypes
- Develop stand-alone WRENCH-based pedagogic modules

WRENCH SOFTWARE STACK
WRENCH APIs: Developer and User





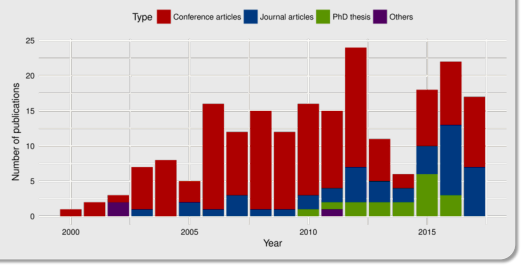
SIMGRID
Versatile Simulation of Distributed Systems

- SimGrid is a **research project**
  - Development of simulation models of hardware/software stacks
  - Models are accurate (validated/invalidated) and scalable (low computational complexity, low memory footprint)
- SimGrid is **open source usable software**
  - Provides different APIs for a range of simulation needs, e.g.:
    - S4U: General simulation of Concurrent Sequential Processes
    - SMPI: Fine-grained simulation of MPI applications
- SimGrid is a **versatile scientific Instrument**
  - Used for (combinations of) Grid, HPC, Peer-to-Peer, Cloud simulation projects
- First developed in 2000, latest release: v3.19 (March 2018)

More information: *htpp://simgrid.org*

SimGrid is well-funded, active as a research project, and widely-used as a simulation tool for research, development, and education:

- 44 journal articles
- 146 conference articles
- 17 PhD theses



ONGOING/UPCOMING WRENCH-enabled Projects
Research, development, education

Simulating Pegasus WMS

- Pegasus is a widely adopted production WMS
- Research and development in Pegasus is hindered by the difficulties and limitations of real-world experiments
- WRENCH is used to implement a simulator of Pegasus and of the complex software infrastructures it uses
- This makes it possible to make informed algorithmic, software infrastructure, and distributed system design decisions for future versions of Pegasus

Simulating VIP (Virtual Imaging Platform)

- VIP targets the execution of medical imaging workflow applications on the BioMed Virtual Organization resources provided on the EGI (European Grid Initiative) platforms
- The objective is to optimize workflow executions via better decision making strategies
- WRENCH is used to simulate novel…
  - Data-replication strategies
  - Pilot job submission strategies for batch scheduled clusters
  - Cluster selection strategies

Efficient workflows executions on batch-scheduled clusters

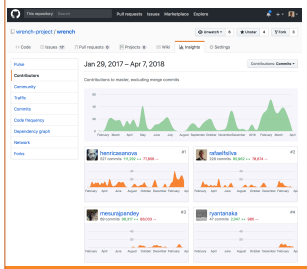
- Batch-scheduled clusters are not ideally suited to workflow applications, and yet they represent the majority of HPC execution platforms
- A key question is: how should workflow tasks be aggregated into batch jobs?
- Various static options have been proposed in the context of cloud platforms
- For batch-scheduled clusters, one approach is to design task aggregation strategies that try to account for the dynamics of the batch queues
- WRENCH simulations are used to drive the design of such strategies

Stand-alone pedagogic modules

- It is crucial to teach undergraduate students parallel and distributed computing
- But it is not easy (giving students access to sufficiently diverse and realistic software/hardware platforms, dealing with platform down-times and instabilities, dealing with time-consuming and possibly costly executions)
- Simulation resolves these difficulties and WRENCH provides the foundation for pedagogic modules on parallel and distributed computing that use workflows as a motivating context

SOFTWARE AVAILABILITY
Code Repository, Releases, Software Engineering Process

Open-source repository
https://github.com/wrench-project/wrench



Releases
1.0 (May 15, 2018 - *estimated*)
1.0-beta (April 15, 2018)
1.0-alpha (December 1, 2017)

Continuous Integration
https://travis-ci.org/wrench-project/wrench
Travis CI build passing

Tests Coverage
https://coveralls.io/github/wrench-project/wrench
COVERALLS coverage 84%

Code Review
https://app.codacy.com/app/WRENCH/wrench/dashboard
code quality A

https://sonarcloud.io/dashboard?id=wrench
sonarcloud lines of code 13435

LEARN MORE

Get in Touch
http://wrench-project.org – support@wrench-project.org

WRENCH is funded by the National Science Foundation (NSF) under grants #1642369 and #1642335, and the National Center for Scientific Research (CNRS) under grant #PICS07239.