CHAMELEON: REPRODUCIBLE EXPERIMENTS IN THE EDGE TO CLOUD CONTINUUM

Kate Keahey

Mathematics and CS Division, Argonne National Laboratory
CASE, University of Chicago

keahey@anl.gov

January 7th, 2022
IEEE FNI Testbed Workshop
CHAMELEON IN A NUTSHELL

- Chameleons like to change: a testbed that adapts itself to your experimental needs
  - **Deep reconfigurability** (bare metal) and isolation + KVM cloud (different cost/isolation trade-off)
  - Capabilities: power on/off, reboot, custom kernel, serial console access, etc.
- **Balance: diversity and scale** – from large to small
  - Large to small: from 2 supercomputing sites (UC, TACC) connected with 100G network to **edge devices**
  - Diverse: FPGAs, GPUs, NVMe, NVDIMMs, Corsa switches, edge devices via CHI@Edge etc.
  - **CHI-in-a-Box** sites at Northwestern, in preparation: NCAR, IIT, and other places
- **Cloud++**: CHameleon Infrastructure (CHI) via mainstream cloud tech
  - Powered by **OpenStack** with bare metal reconfiguration (Ironic) + “special sauce” (50/50 split)
  - Blazar contribution recognized as official OpenStack component
- **Reproducibility, repeatability, and sharing**
  - **Jupyter integration** for imperative and non-transactional experiment packaging, **Chameleon daypass** for easy access, **Trovi** for sharing and finding experiments, integration with **Zenodo** for publishing
OPEN TESTBED – BY THE NUMBERS

450+ Papers published

800+ Unique projects

6,000+ Users

160+ Institutions

45 Countries

6+ Years Old

~3 more years to grow!
CHAMELEON HARDWARE

**CHI@UC**
- Skylake, CascadeLake, IceLake, and AMD nodes
- GPUs (A100, V100, RTX6000), FPGAs (Xilinx Alveo U280)
- Storage (NVMe SSD, NVDIMM)
- Network (Corsa SDN, 25G Ethernet, 200G InfiniBand)

**Storage**
- ~0.5 PB

**CHI@TACC**
- Haswell, Skylake, CascadeLake, and AMD nodes
- GPU (K80, M40, P100, MI100), FPGAs (Altera)
- Storage (NVMe, SSD, NVDIMM)
- Network (Corsa SDN, 25G Ethernet, 200G InfiniBand)

**Storage**
- ~3.5PB

**Network**
- 100Gbps public network

**Commercial Clouds via CloudBank**
- FABRIC and other partners

**Chameleon Associate Sites:** Northwestern (3x 100G NICs) and other partners

**CHI@Edge**
- Raspberry Pi, Jetson Nano, Jetson Xavier NX

**Chicago**

**Austin**

www.chameleoncloud.org
CHI EXPERIMENTAL WORKFLOW

**discover resources**
- Fine-grained
- Complete
- Up-to-date
- Versioned
- Verifiable

**allocate resources**
- Allocatable resources: nodes, VLANs, IPs
- Advance reservations and on-demand
- Expressive interface
- Isolation

**configure and interact**
- **Bare metal**, KVM, containers for edge
- Image catalog
- Snapshotting
- Orchestration (Heat)
- Jupyter integration
- Networks: stitching and BYOC

**monitor**
- Hardware metrics
- Fine-grained data
- Aggregate
- Archive

Authentication via federated identity, accessed via GUI, CLI and python/Jupyter

*Paper: “Lessons Learned from the Chameleon Testbed”, USENIX ATC 2020*
NEW IN P3: CHI@EDGE (PREVIEW)

- A lot like a cloud! All the features we know and love but for edge!
- Not at all like a cloud! Location, location, location! Not server-class! IoT: cameras, actuators, SDRs! And many other challenges!

- CHI@Edge: all the features you know and love plus
  - Reconfiguration via container deployment
  - Support for peripherals based on an extensible plug-in model
  - **Mixed ownership** model via an SDK with devices, virtual site, and restricted sharing
  - And more... Chameleon@Edge Community Workshop in September 2021 https://chameleoncloud.org/chiedge-community-workshop/

[Image of a device with the Chameleon logo]
LEAVING NO EXPERIMENT BEHIND!

Supporting research projects in architecture, operating systems design, virtualization, power management, real-time analysis, security, storage systems, databases, networking, machine learning, neural networks, data science, and many others.

Check out user experiment stories on our blog:
https://www.chameleoncloud.org/blog/category/user-experiments/
ARA: WIRELESS LIVING LAB FOR SMART & CONNECTED RURAL COMMUNITIES

- **ARA objectives**
  - Enable research to achieve a factor of 10+ reduction in broadband cost and make rural broadband as affordable as urban broadband!
  - Support broadband use cases for rural communities

- **ARA wireless living lab**
  - Deploy advanced wireless platforms in Central Iowa (>600 square miles); capture systems and application and community contexts of rural broadband
  - Mainstream open-source platforms for living lab management and experimentation: OpenStack, CHI-in-a-Box & CHI@Edge, ONF (SD-RAN, SD-CORE, ONOS), srsRAN, OpenAirInterface etc
  - CHI@Edge: collaborating on spectrum reservations for management of wireless networks and CHI@Edge in a Box

---

Hongwei Zhang, ARA PI
Iowa State University

www.chameleoncloud.org
arawireless.org
EDGE FOR MARINE BIOLOGY

- Goal: map existing fish populations and thereby understand better how pollution impacts their habitat and the general Biscayne Bay ecosystem
- Challenges: What is the best cloud/edge strategy for collecting and analyzing data from the autonomous vehicle (AV)? How does the resolution of video data and quality of network connection influence them?
- CHI@Edge: using CHI@Edge for developing edge to cloud data processing workflows via Jupyter notebooks
PRACTICAL REPRODUCIBILITY

- Can experiments be as sharable as papers are today?
- Could it be as easy to provide conditions for reviewers to repeat experiments or data analysis in a paper as it is to organize a PC meeting?
- Can I simply integrate somebody’s model into my research instead of reinventing the wheel?
- Can I have so much fun playing with somebody’s experiment that discover a new result?
- Can I develop exercises for my class based on most recent research results?

The existence of powerful open testbeds is a fundamental requirement for practical reproducibility
**TESTBED AS SHARING PLATFORM**

- **Instruments held in common** are a reproducibility imperative
  - Hardware and hardware versions: >105 versions over 5 years
  - Expressive allocation

- **Sharing via cloud pattern**
  - Disk images, orchestration templates, and other artifacts
  - Chameleon >130,000 images, >35,000 orchestration templates and counting

- Testbed as “player” for environments

*Paper: “The Silver Lining”, IEEE Internet Computing 2020*
WHAT IS MISSING?

- Packaging: complete, imperative, non-transactional, integrated (literate programming)
- Get access for reproducibility
- Discover/find experiments through various channels
- Package experiment in a way that is cost-effective but also user-friendly
- Give access for reproducibility
- Share work in progress; publish and advertise completed work
PACKAGING SHARABLE EXPERIMENTS

Literate Programming with Jupyter

Experimental storytelling: ideas/text, process/code, results

Complex Experimental containers

- Repeatability by default: Jupyter notebooks + Chameleon experimental containers
  - JupyterLab for our users: use jupyter.chameleoncloud.org with Chameleon credentials
  - Interface to the testbed in Python/bash + examples (see LCN’18: https://vimeo.com/297210055)

**Paper:** “A Case for Integrating Experimental Containers with Notebooks”, CloudCom 2019
Authors create a subproject with multiple short-term leases that are long enough to reproduce the experiment.

Readers click through data of a published experiment, request a daypass, and reproduce either the experiment or data analysis.
SHARING AND FINDING EXPERIMENTS

Digital publishing with Zenodo: make your experiments citable via Digital Object Identifiers (DOIs)

Trovi: sharing work in progress
- BINs to collect all the artifacts, fine-grained sharing, versioning
- Portal to browse, filter, and find interesting experiments
- Integrated with Jupyter/Chameleon, Swift, Zenodo, and github (in progress)
PARTING THOUGHTS

- Constantly in motion: scientific instruments are laying down the pavement as science walks on it

- **Testbed evolution**: from cloud to edge
  - Before: expensive *provider-owned* hardware as the main draw
  - Now: *user-owned* inexpensive hardware using testbed *sharing and connecting* mechanisms
  - Testbeds == effective *sharing and connecting* mechanism + residual resources

- Sharing your research digitally is more important than ever!
  - Make it easy with Chameleon: public platform, environments as images, packaging, access, and sharing mechanisms at the ready
  - Biggest benefit in emergent area == real incentives
Think Big!
(with the help of a small reptile)

www.chameleoncloud.org