The Open Cloud Testbed (OCT): A Platform for Research into new Cloud Technologies

Michael Zink, David Irwin, Emmanuel Cecchet, Hakan Saplakoglu
Dept. of Electrical and Computer Engineering
University of Massachusetts Amherst
Amherst, MA, USA
{mzink|deirwin|cecchet|hsaplakoglu}@umass.edu

Orran Krieger, Martin Herbordt, Michael Daitzman
Dept. of Electrical and Computer Engineering
Boston University
Boston, MA, USA
{okrieg|herbordt|msd}@bu.edu

Peter Desnoyers[*], Miriam Leeser[+], Suranga Handagala[+]
[*]*College of Computer and Information Science,* [+]*Dept. of Electrical and Computer Engineering*
Northeastern University
Boston, MA, USA
{p.desnoyers|m.leeser|s.handagala}@northeastern.edu

2021 IEEE 10th International Conference on Cloud Networking (CloudNet) | 978-1-6654-3538-3/21/$31.00 ©2021 IEEE | DOI: 10.1109/CloudNet53349.2021.9657109

***Abstract*—The NSF-funded Open Cloud Testbed (OCT) project is building and supporting a testbed for research and experimentation into new cloud platforms – the underlying software which provides cloud services to applications. Testbeds such as OCT are critical for enabling research into new cloud technologies – research that requires experiments which potentially change the operation of the cloud itself.**

This paper gives an overview of the Open Cloud Testbed, including an overview on the existing components OCT is based on and the description of new infrastructure and software extension. In addition, we present several use cases of OCT, including a description of FPGA-based research enabled by newly-deployed resources.

***Index Terms*—cloud computing, testbeds, bare metal, FPGA**

I. Introduction

Cloud Computing is having as large an impact on our society as has the Internet. It has changed the way we develop, release and consume software. It has made possible new models of sharing information and enabled discovery for many disciplines at an unprecedented rate. New hardware is being developed and deployed at an accelerated pace by cloud providers drawing on resources from a massive world wide customer base.

Unfortunately, many areas of innovation are limited to researchers and developers working for today's public clouds, where they can perform experiments at massive scale, with real users, taking advantage of detailed telemetry from cloud infrastructure, and where they are ultimately able to transition successful innovation into products made available to their customers.

Realizing the critical importance of enabling the systems research community to participate in this fundamental transformation in IT, national research agencies have made significant investments to build testbeds to enable such research (e.g., the National Science Foundation in 2014 and 2017 invested ∼$40M).

These testbeds have been enormously effective. For example, CloudLab—which is supported by the National Science Foundation (NSF) CISE Research Infrastructure: Mid-Scale Infrastructure - NSFCloud program—supported 4,000 users who have run over 79,000 experiments on 2,250 servers in the 2015 - 2019 time frame [6]. The use of CloudLab led to numerous publications in many top-level research venues, including SIGCOMM, OSDI, SOSP, NSDI, and FAST.

Yet if it is our goal to enable the CISE community to participate in inventing the next generation of cloud technology, these testbeds are not sufficient, for several reasons. While they have been successful in enabling research at a scale previously unavailable, they are 100% utilized before major conference deadlines, limiting the ability and scale of research that can be performed. It is difficult for these testbeds to keep up with constant updates in cloud technology—e.g. FPGA accelerators [17], which are now deployed with all Microsoft Azure servers, for high-performance networking and other purposes. Finally, researchers using these testbeds do not have direct access to a real cloud, with real users running real applications, or detailed operational data from such a cloud.

Identify applicable funding agency here. If none, delete this.

Authorized licensed use limited to: University of Massachusetts Amherst. Downloaded on December 06,2022 at 18:43:01 UTC from IEEE Xplore. Restrictions apply.

In this paper, we present the Open Cloud Testbed (OCT), a system which combines a research testbed with cloud and HPC services. At its onset OCT was created as a small cluster of servers which has been significantly expanded by donations from industry. Currently OCT consists of a total of 5,172 cores and 63 TB of RAM distributed over 237 servers, eight of which are equipped with high-end FPGA accelerators. OCT is located at the Massachusetts Green High Performance Compute Center (MGHPCC), a 15 megawatt, 90,000 square foot academic computing facility that houses a variety of compute clusters, large-scale storage solutions, and offers high-bandwidth connectivity into research and commercial networks.

By federating with CloudLab [6], OCT resources are available to that part of the CISE community which is already enthusiastically using the existing NSFCloud testbeds. In addition, we have created tutorials and other outreach activities to allow new users to obtain the required skills to run experiments on OCT, including a workflow for the usage the FPGAs offered in the testbed.

One of the major features of the OCT is the flexible and secure sharing of hardware between different uses, via the *Elastic Secure Infrastructure* (ESI) that we have been developing [4], [8], [16]. We are currently in the process of extending ESI and bringing it to production quality such that it can be used for OCT and production cloud and HPC clusters in MGHPCC. This will allow OCT to offer elastic access to large amounts of hardware—far beyond the resources dedicated for testbed use—for short periods of time. In future work, we plan to introduce the capability to suspend and resume experiments, allowing a more fine-grained sharing of testbed hardware. The combination of these capabilities will allow a modest amount of resources to support bursts of much higher resource usage.

The close association of OCT with other computing services located at MGHPCC is key to much of its value. Researchers can run pilot services on production cloud services stood up in the MGHPCC (NERC and MOC, described below) and offer them to real users. In addition, researchers will be able to access telemetry data (logs, monitoring, etc.) of not only OCT but other production clouds running in MGHPCC. Over the past year, a group of researchers and operators has been formed that is implementing a data use agreement, allowing user-related data (i.e. traces and other telemetry) to be shared with researchers. In addition, this group is in the process of developing solutions to capture and store telemetry information in efficient and easy-to-access ways.

The OCT is designed to enable the use of institutional resources for sustainability and growth. For example, when fully deployed, OCT will allow existing institutional resources to be shifted between HPC clusters, a production cloud (e.g., NERC), and the testbed depending on current researcher needs. Essentially, a testbed experiment becomes “just another application”, consuming entire machines much like a large HPC job.

The methods and tools we use to share hardware between testbeds, HPC, and cloud use can be replicated to other institutional data centers, creating a series of federated Open Cloud Testbeds, each sharing hardware between systems research and other uses. This is in fact the ultimate goal of this project —the ability for systems researchers to be “normal” users of computing facilities at their institutions, using resources funded via various mechanisms in use today, rather than being restricted to using a single resource dedicated to this community. This is in contrast to the current situation, where systems researchers are often entitled to significant allotments on shared compute facilities at their respective institutions, but are unable to use them for most of their research and must turn to testbeds like CloudLab and Chameleon [11].

The remainder of this paper is outlined as follows. In Section II, we give an overview on the existing work and resources the Open Cloud Testbed builds on. Section III presents a detailed descripton of the current status of the testbed, and Sect. IV introduces a set of use cases. Section V concludes the paper.

II. Background

In this section, we give an overview of existing infrastructure and technologies that were either used in the design and implementation of OCT or are part of the larger ecosystem in which it is situated.

A. MGHPCC

The Massachusetts Green High Performance Computing Center is a $95 million, 90,000 square foot, 15 megawatt compute center in Holyoke MA, with the space, power, and cooling capacity for approximately 750 racks of computing equipment on a single shared floor. The facility is owned, operated, and used for research computing by a 5-university consortium (University of Massachusetts, Harvard University, Boston University, Northeastern University, and the Massachusetts Institute of Technology), and currently houses roughly 300,000 cores of compute and 45 PB of storage from member universities. The MGHPCC consortium has continued to invest in infrastructure and completed a multi-100Gbit/s fiber ring between the MGHPCC and the national research networks. The NSF DIBBs-funded Northeast Storage Exchange (NESE) [27] provides key hardware for regional and national data science, providing many petabytes of high-speed Ceph-based storage to create a “data lake” at MGHPCC. In addition, the MGHPCC hosts a GENI rack, which provides access to Internet2’s Advanced Layer 2 Service (AL2S).

B. Mass Open Cloud (MOC) and Open Research Cloud Initiative (ORCI)

The Mass Open Cloud (MOC) is a best-effort cloud developed by a partnership of academia (Boston University, Harvard University, Northeastern University, Massachusetts Institute of Technology, and the University of Massachusetts), government (Mass Tech Collaborative, USAF), and industry (Red Hat, Intel, Two Sigma, NetApp, Cisco). The existing MOC physical infrastructure includes around 2200 cores of commodity Intel compute, 160 Power9 cores, 40 GPUs, and 1.2PB of storage.

Authorized licensed use limited to: University of Massachusetts Amherst. Downloaded on December 06,2022 at 18:43:01 UTC from IEEE Xplore. Restrictions apply.

Services offered by the MOC (in collaboration with industry partners) includes an OpenStack IaaS service, Ceph Volume and Object storage, an OpenShift/Kubernetes PaaS service, an experimental AI platform (Open Data Hub), and a Dataverse service for hosting datasets. The MOC has been used by thousands of students and researchers over the last four years, supporting courses and numerous research projects that are end users of the various MOC services. It currently has around 400 active users and over 10,000 users of services deployed by its active users.

The MOC was initiated as a soft money project by a small operations team on a heterogeneous infrastructure obtained from a variety of industry donations. It provides a best effort experimental service and requires a fair amount of technical knowledge from its users with very little ability to support those users directly. As a result of its increasing usage, Boston University and Harvard are created a production cloud service, the New England Research Cloud (NERC) [26], which is supported by professional research IT staff from those universities on homogeneous hardware purchased for this purpose.

The MOC, OCT, and NERC are all part of the recently created Open Research Cloud Initiative (ORCI) [12]. ORCI has the mission to create an open production cloud that provides domain researchers with predictable low cost resources and facilitator support while enabling academic researchers and the open source community to participate in the kind of close interactions between research, development, and production operations that has resulted in so much innovation in today's public clouds.

The production services of ORCI will enable important services to be supported. For example, Harvard Dataverse [5], a large data set repository for the research community, is planning to move from AWS to NERC. With this move, users will have in-situ access to a large repository of datasets (and Harvard Dataverse will be able to expand to supporting large research data sets for which AWS is cost prohibitive). We will be able to use Elastic Secure Infrastructure service (ESI, see Sect. II-D) to shift resources between OCT and other ORCI services and expose to OCT researchers telemetry from the production services. At the same time, production services will be informed about new cloud technologies developed within OCT (or through the usage of OCT) and decide when they might be ready for adaption into the NERC production environment.

C. CloudLab

The original CloudLab physical facilities include three clusters (Utah, Wisconsin, and Clemson), which, in combination, offer 15,000 cores. The software used to operate and manage the CloudLab testbed has been in use for almost 20 years [9], derives from software originally developed for Emulab [19], [25], and was further extended for ProtoGENI and InstaGENI [14]. This software is designed specifically for reproducible research [18], with network isolation provided as part of a full-featured mechanism for describing and deploying experiments in the RSpec language [20]. Due precisely to these features, the CloudLab framework cannot easily support most production compute services, which typically use their own tooling to deploy software using low-level mechanisms such as PXE boot.

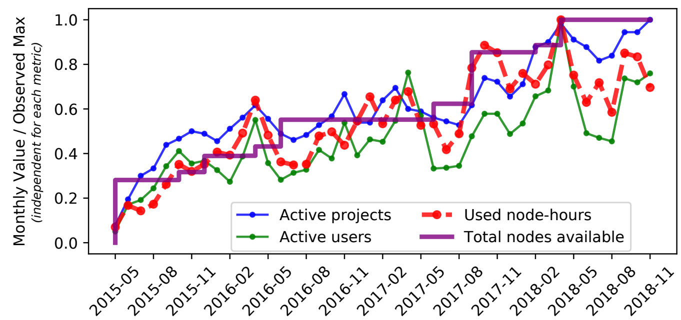

Fig. 1. CloudLab capacity and utilization over its three year life. To produce a consistent scale, all metrics are normalized to their maximum value. Utilization (orange) peaks at conference deadlines in January.

Figure 1, which shows testbed capacity and utilization since CloudLab became available to researchers, shows the challenge of an isolated systems testbed. Resource usage is strongly correlated to submission deadlines for conferences; the testbed's utilization is 100% at the end of January which coincides with the SIGCOMM submission deadline. It is likely that users in these peak periods are experiencing long wait times to run experiments. In addition, we see extended periods of time when the average utilization is relatively modest (e.g. $\sim$0.4) showing that valuable resources are going underutilized during large parts of the year. Clearly it would be enormously valuable if the testbed could itself be elastic in order to better meet peak demand without wasting resources at other times.

We refer the interested reader to Duplyakin et al.'s paper [6] for a detailed description of the design and operation of CloudLab.

D. Elastic Secure Infrastructure (ESI)

The Elastic Secure Infrastructure (ESI) service [16] is a set of services for securely managing and provisioning physical servers designed for production, rather than experimentation, developed at the MOC with funding from the NSF (MACS: A Modular Approach to Cloud Security) and the US Air Force. All higher layer services in the MOC are deployed on top of ESI, allowing computers to be moved between different uses (e.g. Open Stack, Open Shift, direct hardware experimentation) based on demand. ESI (Figure 2) is composed of a set of micro-services that include: 1) an isolation layer [8], allowing users to allocate machines, perform simple out-of-band management (OBM) operations on them, and attach those machines to different networks, 2) a stateless provisioning service [15], [24] that exploits the isolation layer to provision machines rapidly using network mounted storage, 3) an (experimental) integration with an attestation service [21], developed in collaboration with Two Sigma, MIT Lincoln Labs, and Red Hat, that allows a customer to attest that the firmware and software provisioned on a server are not compromised.

Authorized licensed use limited to: University of Massachusetts Amherst. Downloaded on December 06,2022 at 18:43:01 UTC from IEEE Xplore. Restrictions apply.

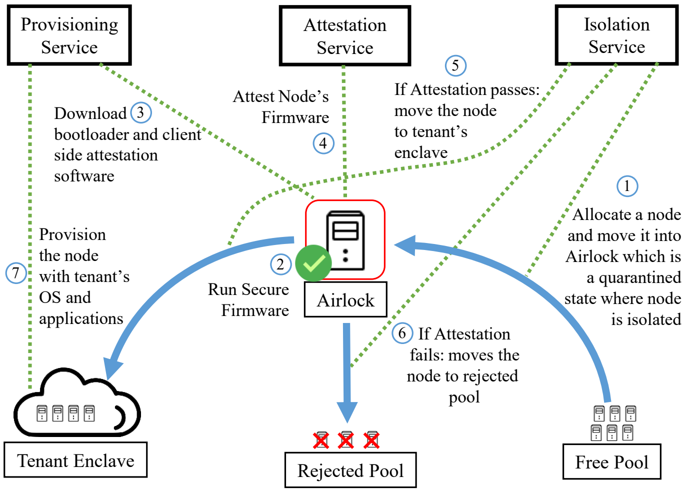

Fig. 2. Elastic Secure Infrastructure: Blue arrows show state changes and green dotted lines show actions taken by each service.

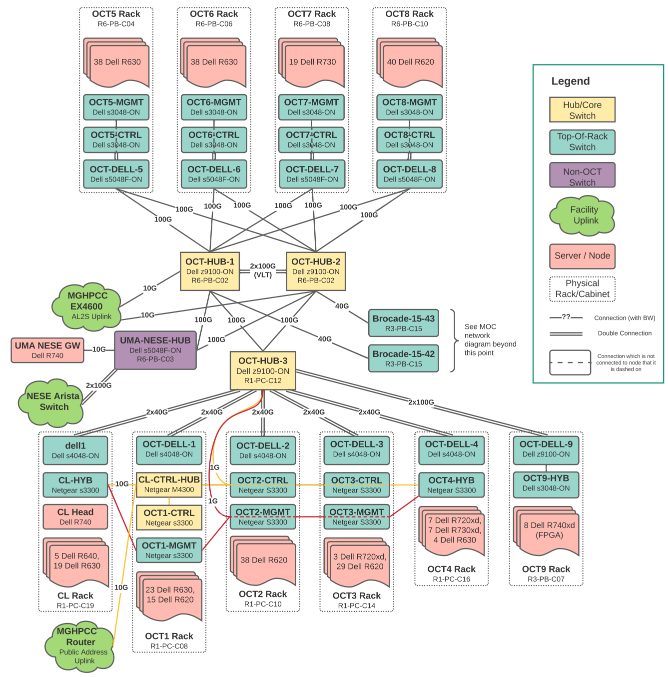

Fig. 3. Overall OCT topology, currently composed of ten racks of compute equipment. Servers are connected via 2x40G or 100G Ethernet connection; each FPGA is connected via 2x100G Ethernet.

While lacking the experiment support of CloudLab, ESI provides key capabilities needed for production support. First, because the isolation layer and provisioning system are decoupled, a tenant can use their own provisioning system, allowing services with their own provisioning system to share hardware with other services. (E.g., for the MOC Open Stack cloud, this allows the use of the Red Hat-supported Foreman deployment manager.) Second, ESI supports efficient stateless booting over the network [24], allowing for rapid movement (e.g., under five minutes) of servers between services, greatly reducing the cost of migrating resources from one use to another. (This is also of great value to experimenters, allowing virtual machine-like control of experiment state.) Finally, ESI includes an attestation service [16], which allows ESI users to protect against attacks from prior users of the hardware—a risk remote enough to be ignored in today's cloud testbeds but critical to production teams considering pooling hardware with the MOC or OCT.

The existing ESI implementation is robust enough to be used by the current MOC best-effort *production* service for rapid multiplexing of resources between different services of the MOC and a few trusted researchers. In Section III-C, we describe current development activities we are engaged with our industry partners to make ESI robust enough to support production clouds and HPC clusters which are used by thousands of researchers on a daily basis.

III. Open Cloud Testbed

In this section, we describe the current state of the testbed and its features. We will describe the current hardware available for researchers, highlight the newly-deployed FPGA accelerators and their usage, and provide an overview on adjacent resources from which researchers can benefit while conducting experiments. Figure 3 give an overview of the OCT topology and the network infrastructure.

A. Hardware

Since the inception of OCT we were able to constantly grow its footprint by adding two very significant increments of hardware. Originally, the testbed started as a small cluster of only five nodes. Based on two rounds of generous donation of equipment from industry partner Two Sigma, donated to MOC/ORCI and made available to OCT, we were able to increase the size of the testbed significantly. As of September 2021, the testbed consists of a total of 5,172 cores and 63 TB of RAM distributed over 237 servers. Since this hardware has been donated by industry we have the opportunity to make it available to not only the CISE research community, but to select open source communities as well. Through ESI (see section II-D, the hardware can be made available to different control frameworks. Section IV presents several use cases that demonstrate how the OCT hardware ca be used under different control frameworks.

The operation of a testbed that supports systems research and offers bare metal servers requires three separated networks. First, a 1Gbps management network, providing a private network that connects the management instance with the IPMI interfaces of the servers. Second, a public 1Gbps control network, which allows experimenters to connect to the individual servers via the public internet. Third, a data plane network that establishes high-bandwidth connections between the individual nodes. As can be seen in Fig. 3, half of the racks are connected via 2x40Gbps, while the other half is connected via 2x100Gbps connection. (This difference in bandwidth is based on the growth of the testbed infrastructure

Authorized licensed use limited to: University of Massachusetts Amherst. Downloaded on December 06,2022 at 18:43:01 UTC from IEEE Xplore. Restrictions apply.

in two increments.) Finally, the use of VLANs enables the creation of isolated networks, where a separate VLAN is assigned to each experiment.

In the following section we present further details on the FPGAs that are installed in several OCT nodes.

B. FPGAs

At the onset of the OCT project, we surveyed a broad group of researchers from the FPGA community to identify their needs for a testbed that offers such devices. Based on their feedback and our own experience in FPGA-related systems research we integrated FPGAs in OCT and created a set of tools readily usable by experimenters.

We have successfully integrated eight Xilinx Alveo U280s into OCT. Alveo U280s are top of the line FPGA accelerator cards for the cloud, with High Bandwidth Memory to support data intensive applications. The FPGAs use PCIe connection to the host processor and are also directly connected to the network via two independent 100Gbps connections. This allows direct FPGA-to-FPGA communication via TCP or UDP. Figure 4 depicts the current FPGA setup in OCT.

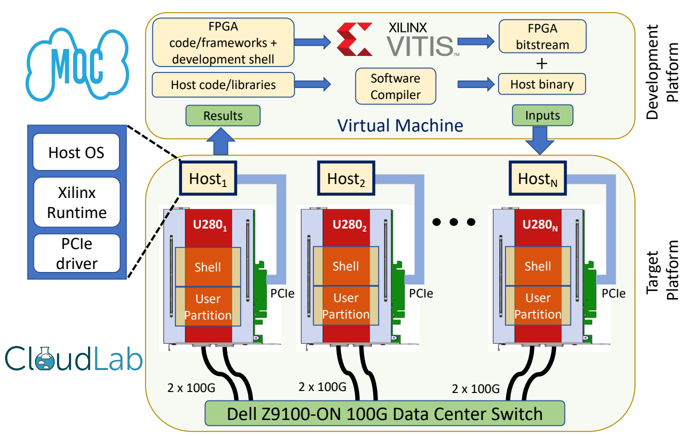

Fig. 4. Overview of the implementation of the development and target platform for the FPGAs offered in OCT.

Currently, the nodes that host the FPGAs can be allocated via CloudLab. In addition, we have create a virtual machine image that includes the Xilinx Vitis tools and can be instantiated in MOC (see Fig. 5). To better support the research community, we created tutorials for all steps required – from the creation of a bitstream to the execution of an experiment in the testbed [7]. In addition, Xilinx's FINN is installed—an experimental framework from Xilinx Research Labs to explore deep neural network inference on FPGAs [3]. Tutorials for running FINN examples have been adapted to the OCT setup and are also available. This toolchain provides researchers with the components necessary to implement an application that can run on FPGAs to evaluating such applications in an actual testbed.

Another eight Alveo U280s have been recently donated by Xilinx and will be integrated into OCT in the next month.

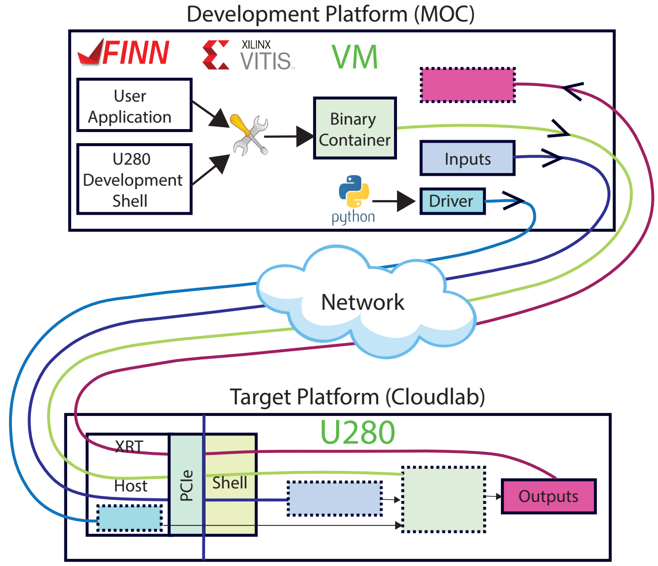

Fig. 5. OCT FPGA tool workflow.

One of the goals of the next phase is to deploy Intel FPGAs with a basic tool chain, analogous to the one currently supported for the Xilinx FPGAs, and also running COPA (COnfigurable network Protocol Accelerator). COPA provides a customizable framework that integrates communication and computation on an FPGA and incorporates SmartNIC capabilities [13]. The hardware environment provides networking and accelerator infrastructure while the software abstracts the FPGA from the underlying application or middleware.

C. ESI 2.0

The Elastic Secure Infrastructure (ESI) service is mature enough to be used to move resources between bare-metal experimentation and the MOC Open Stack (IaaS) and Kubernetes (PaaS) clusters. However these MOC offerings are provide on a best-effort service. Such a best-effort service is sufficient in case ESI is used in a single cloud-like cluster, but further development and hardening is required to support production use and the exchange of servers between clusters and the OCT testbed.

For example, a set of servers that were temporary put into the CloudLab cluster of OCT by removing them from a lightly loaded HPC cluster, would need to be seamlessly returned to the latter once the lease time has expired. Failures in this process would manifest as hardware failures to the HPC cluster scheduler, requiring manual intervention—precisely what this system is intended to avoid.

Security is also a concern in this scenario. Securing systems on the Internet is difficult, and researchers might accidentally allow a bare metal instance to be compromised by attackers. Such attacks typically involve installation of rootkits for persistent access; in the worst case firmware-level rootkits would enabling attacks against any future tenants of the system. While this risk may be tolerated for cloud testbeds, it is no

Authorized licensed use limited to: University of Massachusetts Amherst. Downloaded on December 06,2022 at 18:43:01 UTC from IEEE Xplore. Restrictions apply.

tolerable on most large HPC clusters. The ESI attestation system is designed to verify low-level firmware, with potentially-compromised systems isolated for examination and repair. (by e.g. re-flashing firmware)

To accelerate the hardening of ESI, we are working with Red Hat and the broader IRONIC team (the OpenStack bare metal provisioning service) to integrate ESI into OpenStack. This integration is a two-way street: ESI will gain the support of a broad open source community for testing, future maintenance[1], and development of new features, while IRONIC is gaining ESI's suspend/resume, multi-provider, and attestation features.

While this work will take several years to complete, work on the CloudLab - ESI integration is proceeding with the goal to implement a single-provider deployment. In parallel we are working with both production ORCI services and HPC cluster operators at the MGHPCC to integrate ESI into different, existing, institutional clusters. We will also take advantage of the diurnal pattern of production ORCI cloud service usage to allocate a pool of hardware for large-scale experiments during off hours.

D. Ecosystem

In this section, we give a brief overview on additional projects and resources located at MGHPCC that can be used by researchers to perform large-scale experiments including storage and next generation networks.

1) FABRIC: FABRIC [1], [2] is an everywhere-programmable nationwide instrument comprised of novel extensible network elements equipped with large amounts of compute and storage, interconnected by high speed, dedicated optical links. It will connect a number of specialized testbeds (5G/IoT PAWR, NSF Clouds) and high-performance computing facilities to create a rich fabric for a wide variety of experimental activities. MGHPCC is in the process of becoming a FABRIC facility with a node becoming operational as of the writing of this paper. As soon as this FABRIC node is operational, OCT will be connected to it. Initially, this will offer high-speed connectivity (100Gbps) between OCT and compute resources at TACC, LBLL, SDSC, NCSA, and PSC. In the longer term, it will allow researchers to use FABRIC's compute and storage capabilities for the execution of large-scale, distributed compute tasks. Being able to tie OCT resources into FABRIC will allow researchers to set up large-scale topologies not only in the sense of compute nodes but also in terms of wide-area distribution. For example, FABRIC nodes also host FPGAs. Thus, experiments that use direct FPGA-to-FPGA communication over wide area networks with end systems in data centers will be possible.

2) Northeast Storage Exchange (NESE): The Northeast Storage Exchange is self-sustaining storage facility serving both regional researchers and national and international scale science and engineering projects [28]. It is also physically located in MGHPCC and (as indicated in Fig. 3) directly connected to OCT via a 10G Ethernet connection. Currently, it servers as a shared facility within the MGHPCC consortium members and CephFS based storage and a tape library for long-term archival. While use is currently limited to researchers that are part of the MGHPCC consortium NESE offers storage solutions in close proximity to OCT. This offers researchers to transfer data sets to and from NESE in a very efficient way, enabling the usage of large-scale data sets in OCT-based experiments.

[1]E.g., supporting new switches and out-of-band management interfaces as they are introduced.

IV. Use Cases

A. CloudLab

Currently, a fixed allocation of OCT servers is made available to the CISE research community via the CloudLab framework. This approach has the advantage that this portion of OCT can be easily used by researchers that have experience in performing experiments in CloudLab. In addition, users need not to create any new accounts or projects but can use the existing Cloudlab authentication system. Access to CloudLab for new users is simple and easy, requiring only the request and approval of a new project. Furthermore, users can access existing profiles as a basis to create cloud environments for their research or modify them if needed.

Figure 6 illustrates the weekly node usage (in percent) of the CloudLab resources offered by OCT. The figure clearly shows that usage over time fluctuates significantly, reaching 100% utilization at times but also showing periods of very little usage.

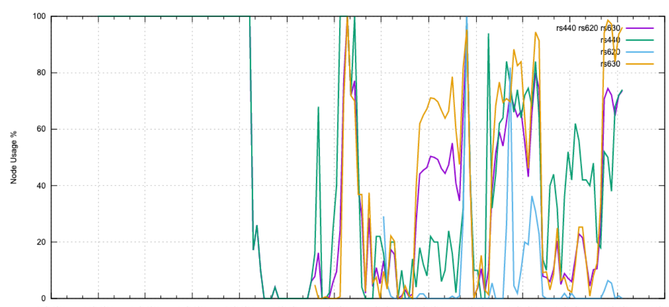

Fig. 6. This figure illustrates the node usage for the different server types that are offered as part of the CloudLab portion of OCT.

These data clearly prove our conjecture that resource allocation for research testbeds should not be static but more dynamic based on demand. For example, Fig. 1 shows a period of up to 100% usage in May 2020, which coincides with the OSDI'20 deadline (originally May 5th, then extended to May 20th). OSDI is a major venue for the CISE systems research community, and CloudLab has been often used by papers presented at the conference. Right after the OSDI deadline usage drops significantly.

In the current state, these resources stay unused. If the ability to re-deploy resources had been available at that time, idle CloudLab resources could have been added to a

Authorized licensed use limited to: University of Massachusetts Amherst. Downloaded on December 06,2022 at 18:43:01 UTC from IEEE Xplore. Restrictions apply.

production cloud (MOC or NERC), an HPC cluster (e.g., the ATLAS Northeast Tier 2 (NET2) cluster), or a cluster for open source development (e.g., OpenInfraLabs) during those periods. Conversely, resources from these other clusters could be made available to the CloudLab cluster during periods of high demand, allowing more researchers to use the testbed for their experiments.

ESI (see Sects. II-D & III-C), allows for an easy and secure transition of compute resources between different clusters. In addition, ESI manages the configuration of network switches such that computes nodes will be connected to the appropriate VLANs.

B. FPGAs

Based on the resources and toolchain described in Sect. III-B users can perform experiments in OCT using FPGAs. Compared to other solutions that provide FPGAs to researchers, the solution offered by OCT offers several advantages. First, users have bare metal access to the host systems of the FPGAs. This allows them to choose from a variety of operating systems and does not lock them into the one provisioned by the operator of the testbed.

The bare metal approach offered through OCT enables research with the goal of providing better system and operating system support for FPGAs. In addition, this approach enables research in the area of disaggregated computing, the latter being not only be supported by the bare metal provisioning of the host servers but also by the direct connection of FPGAs to the network (in contrast to the case where the FPGA is only connected via the PCIe bus). For example, OCT can support distributed applications that run completely on a distributed set of FPGAs, using the host systems only to load bitstreams on them. The second advantages also stems from the direct connection of the FPGAs to the network. This enables research in the area of advanced networking in addition to system research. FPGAs can be used as very powerful network adaptors that enable operations like encryption and decryption, compression and decompression, in-network telemetry, and programmable forwarding planes.

We are currently researching a number of applications that take advantage of the unique features available through OCT. Popular uses of FPGAs for acceleration include implementation of inference for deep neural networks (DNNs) and secure function evaluation. DNNs can be huge with millions of activations and weights, and FPGAs are a popular platform for DNN inference. The tool flow FINN [3], helps users customize large DNNs on FPGAs by implementing optimizations including finite precision implementations and pruning. However, even with these optimizations, the full DNN model may not fit on a single FPGA. We are investigating running FINN across two or more FPGAs that are directly connected over the network to be able to accelerate large machine learning problems.

In another project we are investigating secure function evaluation in the form of Garbled Circuits (GC) across multiple FPGAs. In GC an evaluator processes data in encrypted form; the only party to the computation that has access to the unencrypted data is the data owner. While the approach is promising for data privacy, it comes at the cost of high computational overhead and latency. We have used FPGAs in the cloud to accelerate GC [10]. While these results are promising, due to growth in memory usage and computational overheads, we would like to apply GC to problems split across multiple FPGAs. OCT is the only platform with FPGAs in the cloud that allow us to conduct these research projects. GC was initially done using cloud FPGAs in AWS; however the AWS environment is more restrictive than in OCT and the FPGAs are not directly connected to the network. OCT allows us to conduct experiments with FPGAs in direct communication, thus solving larger compute problems with lower latency than was previously possible.

Molecular Dynamics simulations (MD) is another application that appears particularly amenable to acceleration by the cluster of network-facing FPGAs in the OCT and their low-latency interconnects. MD is fundamental to biochemical modeling with special significance in drug discovery. It is also well-known to pose difficulties in strong scaling, which has led to, e.g., heroic solutions involving ASIC clusters [22]. Previous work on a Catapult testbed has shown that tightly coupled cloud FPGAs have the potential to provide a publicly available alternative [23]. Again, OCT allows us to continue these experiments.

V. Conclusion

In this paper, we presented the Open Cloud Testbed which has the goal rto support research in novel cloud technologies. To achieve this goal, we built a testbed that can grow and shrink based on user demand by using resources from other clusters for certain periods. A portion of OCT is made available via the CloudLab testbed framework to allow researchers to use OCT in a familiar way. In addition, OCT makes FPGAs available in bare metal servers, allowing researchers to perform experiments currently not possible on other testbeds. Based on several use cases we illustrate the features OCT offers to the systems research community.

Acknowledgment

This work was funded by National Science Foundation (NSF) grants CNS-1925464, CNS-1925504, CNS-1925658. All opinions and statements in the above publication are of the authors and do not represent NSF positions. We also thank the industry partners of the MOC and ORCI, especially Red Hat, Intel and Two Sigma for their financial, engineering, and in-kind support that have played a key role in the OCT, and Xilinx, Inc. for their generous donations to the OCT.

References

[1] I. Baldin. FABRIC. https://fabric-testbed.net/. Accessed: 2021-02-16.
[2] I. Baldin, A. Nikolich, J. Griffioen, I. I. S. Monga, K. Wang, T. Lehman, and P. Ruth. Fabric: A national-scale programmable experimental network infrastructure. *IEEE Internet Computing*, 23(6):38–47, 2019.
[3] M. Blott, T. B. Preußer, N. J. Fraser, G. Gambardella, K. O'brien, Y. Umuroglu, M. Leeser, and K. Vissers. Finn-r: An end-to-end deep-learning framework for fast exploration of quantized neural networks. *ACM Trans. Reconfigurable Technol. Syst.*, 11(3), Dec. 2018.

Authorized licensed use limited to: University of Massachusetts Amherst. Downloaded on December 06,2022 at 18:43:01 UTC from IEEE Xplore. Restrictions apply.

[4] T.-M. Chen. ESI Github Repository. https://github.com/CCI-MOC/esi. Accessed: 2021-10-16.

[5] Dataverse. Dataverse. https://dataverse.org/.

[6] D. Duplyakin, R. Ricci, A. Maricq, G. Wong, J. Duerig, E. Eide, L. Stoller, M. Hibler, D. Johnson, K. Webb, A. Akella, K. Wang, G. Ricart, L. Landweber, C. Elliott, M. Zink, E. Cecchet, S. Kar, and P. Mishra. The design and operation of cloudlab. In *2019 USENIX Annual Technical Conference (USENIX ATC 19)*, pages 1–14, Renton, WA, July 2019. USENIX Association.

[7] S. Handagala. OCT FPGA Tutorials. https://github.com/OCT-FPGA.

[8] J. Hennessey, S. Tikale, A. Turk, E. U. Kaynar, C. Hill, P. Desnoyers, and O. Krieger. HIL: Designing an exokernel for the data center. In *Proceedings of the 7th ACM Symposium on Cloud Computing (SoCC'16)*, Santa Clara, CA, Oct. 2016.

[9] F. Hermenier and R. Ricci. How to build a better testbed: Lessons from a decade of network experiments on emulab. In *Proceedings of the 8th International ICST Conference on Testbeds and Research Infrastructures for the Development of Networks and Communities (Tridentcom)*, June 2012. **Awarded best paper**.

[10] K. Huang, M. Gungor, X. Fang, S. Ioannidis, and M. Leeser. Garbled circuits in the cloud using fpga enabled nodes. In *2019 IEEE High Performance Extreme Computing Conference (HPEC)*, pages 1–6. IEEE, 2019.

[11] K. Keahey, J. Anderson, Z. Zhen, P. Riteau, P. Ruth, D. Stanzione, M. Cevik, J. Colleran, H. S. Gunawi, C. Hammock, J. Mambretti, A. Barnes, F. Halbach, A. Rocha, and J. Stubbs. Lessons learned from the chameleon testbed. In *Proceedings of the 2020 USENIX Annual Technical Conference (USENIX ATC '20)*. USENIX Association, July 2020.

[12] O. Krieger. Open Research Cloud Initiative. https://www.bu.edu/hic/centers-initiatives-labs/orci/. Accessed: 2021-10-16.

[13] V. Krishnan, O. Serres, and M. Blocksome. COnfigurable Network Protocol Accelerator (COPA). *IEEE Micro*, 41(1):8–14, Jan. 2021.

[14] R. McGeer and R. Ricci. The InstaGENI project. In R. McGeer, M. Berman, C. Elliott, and R. Ricci, editors, *GENI: Prototype of the Next Internet*, chapter 14. Springer-Verlag, New York, 2016.

[15] A. Mohan, A. Turk, R. S. Gudimetla, S. Tikale, J. Hennesey, U. Kaynar, G. Cooperman, P. Desnoyers, and O. Krieger. M2: Malleable metal as a service. In *Cloud Engineering (IC2E), 2018 IEEE International Conference on*, pages 61–71. IEEE, 2018.

[16] A. Mosayyebzadeh, G. Ravago, A. Mohan, A. Raza, S. Tikale, N. Schear, T. Hudson, J. Hennessey, N. Ansari, K. Hogan, et al. A secure cloud with minimal provider trust. In *10th {USENIX} Workshop on Hot Topics in Cloud Computing (HotCloud 18)*, 2018.

[17] S. Moss. Report: Microsoft to use Xilinx FPGAs in more than half of its servers, in blow to Intel's Altera, Nov. 2018.

[18] L. Nussbaum. Testbeds support for reproducible research. In *Proceedings of the Reproducibility Workshop (Reproducibility '17)*. ACM Press, 2017.

[19] R. Ricci. Emulab. In R. McGeer, M. Berman, C. Elliott, and R. Ricci, editors, *GENI: Prototype of the Next Internet*, chapter 2. Springer-Verlag, New York, 2016.

[20] R. Ricci and T. Faber. Resource description in GENI: Rspec model, Mar. 2008. Presentation at the Second GENI Engineering Conference.

[21] N. Schear, P. T. Cable, II, T. M. Moyer, B. Richard, and R. Rudd. Bootstrapping and maintaining trust in the cloud. In *Proceedings of the 32Nd Annual Conference on Computer Security Applications*, ACSAC '16, pages 65–77, New York, NY, USA, 2016. ACM.

[22] D. E. Shaw, J. Grossman, J. A. Bank, B. Batson, J. A. Butts, J. C. Chao, M. M. Deneroff, R. O. Dror, A. Even, C. H. Fenton, A. Forte, J. Gagliardo, G. Gill, B. Greskamp, C. R. Ho, D. J. Ierardi, L. Iserovich, J. S. Kuskin, R. H. Larson, T. Layman, L.-S. Lee, A. K. Lerer, C. Li, D. Killebrew, K. M. Mackenzie, S. Y.-H. Mok, M. A. Moraes, R. Mueller, L. J. Nociolo, J. L. Peticolas, T. Quan, D. Ramot, J. K. Salmon, D. P. Scarpazza, U. B. Schafer, N. Siddique, C. W. Snyder, J. Spengler, P. T. P. Tang, M. Theobald, H. Toma, B. Towles, B. Vitale, S. C. Wang, and C. Young. Anton 2: Raising the bar for performance and programmability in a special-purpose molecular dynamics supercomputer. In *SC '14: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, pages 41–53, 2014.

[23] J. Sheng, C. Yang, A. Caulfield, M. Papamichael, and M. Herbordt. HPC on FPGA Clouds: 3D FFTs and Implications for Molecular Dynamics. In *27th International Conference on Field Programmable Logic and Applications*, 2017.

[24] A. Turk, R. S. Gudimetla, E. U. Kaynar, J. Hennessey, S. Tikale, P. Desnoyers, and O. Krieger. An experiment on bare-metal bigdata provisioning. In *8th USENIX Workshop on Hot Topics in Cloud Computing (Hot-Cloud 16), Denver, CO*, 2016.

[25] B. White, J. Lepreau, L. Stoller, R. Ricci, S. Guruprasad, M. Newbold, M. Hibler, C. Barb, and A. Joglekar. An integrated experimental environment for distributed systems and networks. In *Proceedings of the USENIX Symposium on Operating System Design and Implementation (OSDI)*. USENIX, Dec. 2002.

[26] S. Yockel. New England Research Cloud. https://nerc.mghpcc.org/.

[27] S. Youssef. Northeast Storage Exchange. http://nese.mghpcc.org/.

[28] S. Youssef, S. Yockel, C. Hill, G. John, D. Tiwari, and M. Zink. NESE: The Northeast Storage Exchange. http://egg.bu.edu/ne/NESE_White_Paper.pdf.

Authorized licensed use limited to: University of Massachusetts Amherst. Downloaded on December 06,2022 at 18:43:01 UTC from IEEE Xplore. Restrictions apply.