

# A Distributed Academic Cloud and Virtual Laboratories for Information Technology Education and Research

Jorge Crichigno\*, Elie Kfoury\*, Kelly Caudle†, and Paul Crump†

\*College of Engineering and Computing, University of South Carolina, Columbia, SC, USA.

†Information Technology, Stanly Community College, Albemarle, NC, USA.

**Abstract**—This paper describes the development of an Academic Cloud and hands-on virtual laboratories that run on this cloud. The system has been deployed by the University of South Carolina, Stanly Community College, and the Network Development Group. The Academic Cloud provides remote-access capability to laboratory equipment via the Internet and seamlessly pools and shares resources (CPU, memory, storage) from four data centers. The data centers are located in South Carolina, North Carolina, Illinois, and Idaho. The virtual laboratories are disseminated via workshops that are organized by the project team. In one year, the Academic Cloud has served over 100,000 learners. Based on input from workshops’ participants who used the platform, the quality of the Academic Cloud and virtual laboratories have exceeded initial expectation.

**Keywords**—information technology; cloud system; virtual laboratories; virtual machines.

## I. INTRODUCTION

Despite the high level of unemployment observed in the U.S. after the COVID-19 pandemic erupted in March 2020 (an increase of unemployment from 4.4% in March’20 to 14.7% in April’20), the information technology (IT) sector has experienced a significant increase of job needs in areas including cybersecurity, systems administration, cloud computing, and others, as shown in Fig. 1 [1].

Gaps in the technology sector have made it difficult for businesses to fill roles. These occupations are fueled by the rise of online systems, the growth of e-commerce and online marketplaces, as well as the reliance on virtual private networks (VPNs) for safe access to resources and remote work. IT-related skills have become pivotal for the survival of businesses and institutions in general, prompting organizations to online business models that are heavily reliant on technology skills [2]. The IEEE and ACM are the main societies that guide IT education. According to them [3], the IT curriculum should emphasize “learning IT core concepts combined with authentic practice” and “use of professional tools and platforms.” COVID-19 exacerbated the need for efficient technologies for hands-on IT education.

This paper describes the implementation of virtual laboratories and a distributed academic cloud system by the University of South Carolina, Stanly Community College, and

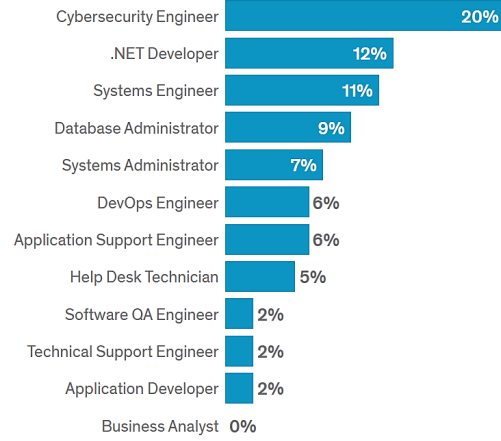


Fig. 1. Increase in demand for IT professionals after COVID-19 [1].

the Network Development Group (NDG) [4]. The latter is a leading developer of IT training solutions. The partnership created new learning material for essential IT domains that facilitates remote learning without compromising rigor nor hands-on activities. The system, referred to as the Academic Cloud, seamlessly pools and shares resources from four data centers located in South Carolina, North Carolina, Illinois, and Idaho.

## II. NEED OF AN ACADEMIC CLOUD

### A. Academic Cloud Platform

As described by the IEEE / ACM guidelines for IT programs [3], “...it is not enough to simply attend courses and read books. Hands-on learning is essential for information technology. Students that learn by doing through live lab exercises are highly prized.” In the past, physical facilities provided such environments. However, they may not be used for remote instruction, require a large capital investment and physical space, require qualified professional staff, and are not scalable.

Features of the deployed Academic Cloud are listed in Table I. It dynamically provides and releases resources required by virtual laboratories. A virtual laboratory experiment requires a pod of equipment (e.g., virtual router, virtual switch, virtual

This work was supported by the National Science Foundation under award numbers 1902397/1902401 and 1723323.

TABLE I. FEATURES OF THE ACADEMIC CLOUD.

Feature	Description
Rapid elasticity	Provide and release resources required by virtual laboratories as needed (CPU, memory, storage).
Measurements	Metering service for statistical and optimization purposes (number of users conducting experiments, hours spent per experiment).
On-demand self-service	Equipment pod to complete a virtual laboratory must be created automatically on demand, as requested by the learner. No human intervention is required for maintenance, deployment, etc.
Resource pooling	Resources are pooled from data centers in South Carolina, North Carolina, Illinois, and Idaho. The learner is oblivious of the exact location of the provided resources.

networks). Note that the laboratory environment is transparent for the learner, and each experiment uses the reserved equipment pod. The pod is created when the student wishes to conduct the experiment, on-demand and self-service.

### B. Virtual Laboratories

Virtual laboratories may have stringent requirements that may be difficult or costly to deploy on public clouds. Table II lists features of the Academic Cloud. Performance refers to the ability of emulating high-speed systems. Functional realism refers to the ability of emulating real IT hardware systems. Traffic realism refers to the capability of any device in the virtual environment to generate/receive live traffic to/from the Internet or to/from another device within the lab environment. The presentation layer refers to the ability of providing a friendly interface, even for inexperienced learners. Flexibility permits instructors to create simple and complex topologies.

## III. ACADEMIC CLOUD AND VIRTUAL LABORATORIES IMPLEMENTATION

### A. Academic Cloud Platform

Fig. 2 illustrates the Academic Cloud. The initial deployment connected three data centers located in South Carolina, North Carolina, and Illinois. The Academic Cloud was further expanded with a fourth data center in Idaho.

TABLE II. FEATURES OF VIRTUAL LABORATORIES.

Feature	Description
Performance	Virtual laboratories precisely emulate high-performance systems (e.g., high-speed networks running at 50 Gbps).
Functional realism	Virtual laboratories have the same functionality as real IT hardware in a real deployment, and execute the same code [5].
Traffic realism	Devices within the virtual environment are capable of generating and receiving real, interactive network traffic to and from the Internet, or from other devices within the virtual environment [5].
Presentation layer	Navigating through an experiment is easy for an inexperienced learner. The topology used by the experiment must be presented using a regular browser. Devices within the virtual environment must be accessible by simply pointing and clicking on them.
Topology flexibility	It must be easy to create an experiment with any topology, including inter-connecting heterogeneous virtual machines (VMs).

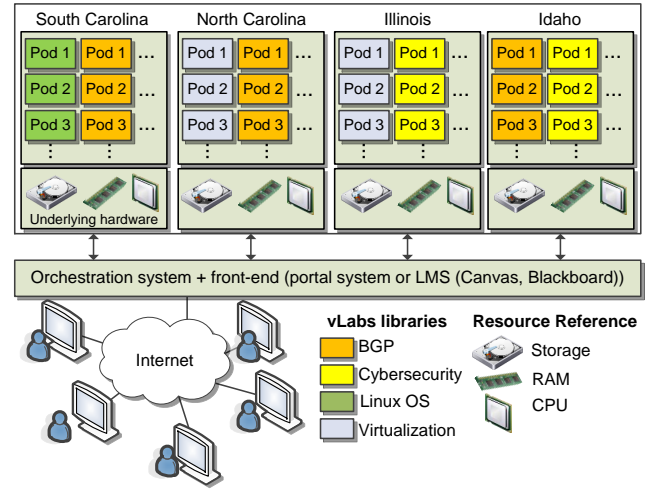


Fig. 2. Academic Cloud.

Servers are specified with large amount of resources: CPUs (32 to 40 cores per server), RAM memory ( $\sim 1$ TB), and storage ( $\sim 2.5$  TB). Each data center has  $\sim 10$  servers used to host VMs. One or more VMs are required to recreate an experiment. The system is managed and orchestrated by a customized software that communicates with VMware's vCenter [6]. vCenter enables the administrators of the system to centrally manage multiple servers and virtual machines distributed across the four data centers. Functions of the customized software include aggregating resources from different data centers, implementing a calendar interface for scheduled access to equipment pods, and routing pod requests to the nearest data center containing the requested pod type.

### B. Virtual Laboratories

Laboratory libraries cover several IT domains, including networking, cybersecurity, virtualization, Linux operating systems, and others. Few library examples are shown in Table III. Each library consists of several laboratory experiments (10-20). For each experiment, a manual is created. The manual contains the theory and step-by-step directions on how to execute the laboratory experiment.

Note that virtual laboratories are based on virtual machines exported in a standard format, namely Open Virtual Appliance (OVA) [7]. As a result, libraries created by third parties are easily imported into the Academic Cloud. The authors are currently developing additional libraries in IoT [8], blockchain [9], routing protocols [10], and other topics.

### C. Access to the Academic Cloud from a Learner's Perspective

One of the objectives of the system is simplicity to use it, from the perspective of a learner. Fig. 3 shows a screen capture of a user accessing the system, referred to as "Cyberinfrastructure Lab". In step (a), the student enters the platform using a web browser. In step (b), the student reserves lab time on a day and time he/she wishes to conduct the experiment. In step (c), the student enters the lab; by then, the scenario has been

TABLE III. EXAMPLES OF VIRTUAL LABORATORY LIBRARIES.

Library	Description
NTP	The Network Tools and Protocols (NTP) library provides hands-on experience with tools and protocols needed for emulation of wide area networks (WANs), performance measuring, congestion control protocols, enhancement of TCP throughput using pacing, and other topics related to high-speed networks [11].
SDN	The Software Defined Networking (SDN) [12] library provides a thorough understanding for administering SDN networks using ONOS controller [13], demonstrating the available applications supported by ONOS, as well as leveraging their full capabilities.
BGP	This library provides a detailed, hands-on experience to understand and configure BGP [14], adjust its attributes, and control network traffic on the Internet.
MPLS and Advanced BGP	This library provides a detailed hands-on experience to understand advanced BGP. Additionally, the library covers topics that include Multiprotocol Label Switching (MPLS) [15] and Ethernet Virtual Private Networks [16].
OVS	This library provides a detailed, hands-on experience to understand and configure Open vSwitch (OVS) [17].
perfSONAR	The perfSONAR library provides hands-on lab experience using emulated multi-domain networks. Labs include configuration details required for accurate performance measurement, monitoring and debugging using dashboard. [11].
Zeek / Bro	The Zeek / The Zeek library provides hands-on lab experience on Zeek Intrusion Detection System (IDS). Labs demonstrate Zeek capabilities such as parsing Zeek files, capturing and analyzing network attacks [18].
Linux Essential	This is an introduction to Linux as an operating system, basic open source concepts and the basics of the Linux command line.
Cloud and virtualization concepts	This library provides details of data center virtualization and types of virtualization. It describes the concepts of server, network, storage, and desktop virtualization.
CCNA Cyber Operations	This library is aligned to Cisco's certificate. It prepares for security operation centers as an analyst or incident responder.

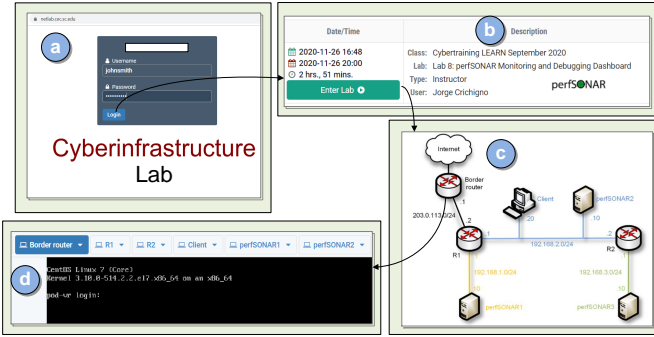


Fig. 3. Access to the platform. (a) A user enters the system. (b) The user reserves lab hours to conduct an experiment. (c) At the reservation day and time, the user enters the lab and a scenario is recreated. The scenario consists of a pod of virtual devices interconnected between each other, and to the Internet. In this example, the user clicked on the “Border router”, resulting in a window being opened. (d) The user operates the device via a command-line interface.

recreated. The scenario consists of equipment interconnected between each other, and to the Internet if needed. By pointing and clicking on a device, e.g., the border router in step (c), a console connection opens, as shown in step (d), and the student can manipulate the device and follow the lab manual directions.

#### IV. PRELIMINARY FEEDBACK AND USAGE OF ACADEMIC CLOUD

The project team continuously organizes workshops for instructors and IT professionals interested in using the Academic Cloud and adopting the material. In 2020, more than 500 instructors and IT professionals attended the workshops (there were approximately 10 workshops, and each workshop was attended by 40 or more attendees). Fig. 4 shows the attendees’ feedback from one workshop conducted between June 15-19, 2020. The workshop trained instructors on high-speed networks, Software-defined Networking, and Zeek.

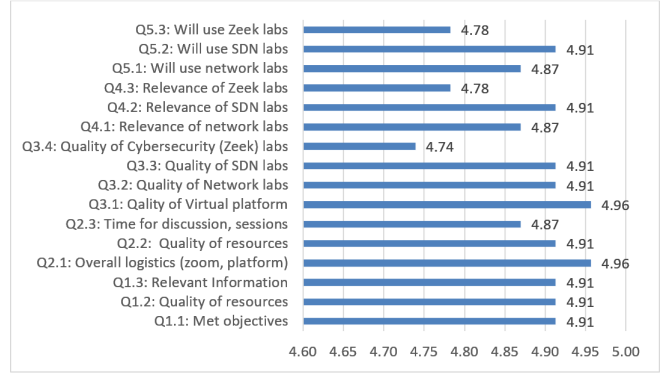


Fig. 4. Summary of the results of the online training workshop, June 15-19, 2020.

The virtual laboratories and platform were evaluated using a survey taken by attendees at the end of the workshop. Out of 45 attendees, 23 responded the survey. Participants completed the 5-item survey designed to measure the quality, relevance, and utility of the information and experience in the workshop. Each question has five levels of agreement: strongly agree (5), agree (4), disagree (3), strongly disagree (2), did not attend (1). On a scale of 1-5, the average for each question was above 4.74, which shows a strong agreement on the quality of the virtual platform, material, and future adoption.

Based on the survey, the following observations can be noted. First, the quality of the platform was noted by all attendees. The average score was 4.96, see Q3.1. Second, the quality of virtual laboratories was between 4.74 (labs on Zeek / cybersecurity, see Q3.4) and 4.91 (labs of SDN and Networks, see Q3.2 and Q3.3). Third, the adoption likelihood was very high, which was later confirmed with follow up conversations with instructors. The average score for the question on adoption was 4.87 for the Networks virtual labs (Q5.1), 4.91 for the SDN virtual labs (Q5.2) and 4.78 for the

Community Usage			
Reservations Made	Labs Attended	Hours Reserved	Hours Attended
6909	6534	42446.59	25158.03
6909	6534	42446.59	25158.03
6909	6534	42446.59	25158.03

Fig. 5. Usage (2020) of the Cloud at one institution supporting one academic program (~300 students).

### Cybersecurity / Zeek virtual labs (Q5.3).

Fig. 5 shows the usage of the platform by one institution, between January 1, 2020 – December 30, 2020. This institution supported approximately 300 students with the platform. Note the number of laboratory experiments, 6,534, and the total number of hours attended, 25,158.03 (average of 3.85 hours per lab). The support staff consisted of the principal investigator (faculty member) and one graduate student. A major advantage of the platform over physical laboratories is the scalability and availability of resources. Conducting 6,534 experiments with minimal staff personnel would not be possible using physical laboratories. Additionally, the institution would not have enough physical resources to support this number of experiments.

### Use of Academic Cloud for Research

Although the Academic Cloud has been used for teaching and training activities mostly, the platform is also used for research experiments as well.

For experiments that test the performance of a system, it is key to allocate sufficient CPUs to the corresponding equipment pod. We conducted performance measurements on topologies with over 200 nodes and 1Gbps links, using Mininet [19]. For such pod, allocating 8 Xeon 6130 cores operating at 2.1 GHz, with 8GB of RAM produces accurate results without over-utilization of resources. The platform has been used also for rapid prototyping of networks consisting of state-of-the-art P4 switches (Tofino) [20], used for congestion control [21] and voice over IP [22].

## V. CONCLUSION

This paper describes the implementation of the Academic Cloud and virtual laboratories running on it. The cloud is managed by a customized software that connects four data centers distributed across the U.S. and controls the resources in those data centers (servers, CPUs, RAM, VMs). Other functions of the customized management software include implementing a calendar interface for scheduled access to equipment pods and routing pod requests to the nearest data center containing the requested pod type.

Virtual laboratories running on the Academic Cloud can emulate complex inter-networks operating at up to 50 Gbps, multi-domain environments, and infrastructures with virtual appliances that include end devices, routers, switches, firewalls, and others. These appliances are capable of generating

and receiving real, interactive traffic to/from the Internet or to/from other devices within the virtual environment, thus exhibiting traffic and functional realisms.

The Academic Cloud is scalable. In one year, the system has served over 100,000 learners. Encouraging feedback from users highlights the quality of the platform and the virtual laboratories.

## REFERENCES

- [1] T. Perry, Tech Jobs in the Time of COVID, IEEE Spectrum, May 2020. <https://tinyurl.com/y4bpdng>.
- [2] Executive Business Education, US Job Market: Tech Skills in-demand Post COVID-19, June 2020. <https://tinyurl.com/y4kcf24n>.
- [3] ACM / IEEE IT Task Force, Information Technology Curricula Guideline 2017. <https://tinyurl.com/yxaot8w>.
- [4] Network Development Group (NDG). <https://www.netdevgroup.com/>
- [5] N. Handigol, B. Heller, V. Jeyakumar, B. Lantz, N. McKeown, Reproducible Network Experiments using Container-based Emulation, International Conference on Emerging Networking Experiments and Technologies, December 2012.
- [6] vCenter Server: Simplified and Efficient Server Management. <https://www.vmware.com/products/vcenter-server.html>.
- [7] OVF and OVA File Formats and Templates, VMware Documentation, <https://tinyurl.com/duwnab9w/>.
- [8] M. Galluscio, N. Neshenko, E. Bou-Harb, Y. Huang, N. Ghani, J. Crichigno, G. Kaddoum, A First Empirical Look on Internet-scale Exploitations of IoT Devices, IEEE Symposium on Personal, Indoor, and Mobile Radio Communications, October 2017.
- [9] G. Srivastava, S. Dhar, A. Dwivedi, J. Crichigno, Blockchain Education, IEEE Canadian Conference of Electrical and Computer Engineering, May 2019.
- [10] J. Crichigno, J. Khoury, W. Shu, M. Wu, N. Ghani, Dynamic Routing Optimization in WDM Networks, IEEE Global Communications Conference, December 2010.
- [11] J. Crichigno, E. Bou-Harb, N. Ghani, A Comprehensive Tutorial on Science DMZ, IEEE Communications Surveys and Tutorials, Vol. 21, Issue 2, 2019.
- [12] D. Kreutz, F. Ramos, P. Verissimo, C. Rothenberg, S. Azodolmolky, S. Uhlig, Software-defined Networking: a Comprehensive Survey, Proceedings of the IEEE, Vol. 103, Issue 1, January 2015.
- [13] Open Networking Foundation (2021). Open Network Operating System (ONOS). <https://opennetworking.org/onos/>.
- [14] Y. Rekhter, T. Li, S. Hares, Border gateway protocol 4 (BGP-4), Request for Comments 4271, January 2006.
- [15] L. De Ghein, MPLS Fundamentals, Cisco Press, November 2006.
- [16] A. Sajassi, R. Aggarwal, J. Uttaro, N. Bitar, W. Henderickx, A. Isaac, Requirements for Ethernet VPN (EVPN), Request for Comments 7209, May 2014.
- [17] B. Pfaff, J. Pettit, T. Koponen, E. Jackson, A. Zhou, J. Rajahalme, J. Gross, A. Wang, J. Stringer, P. Shelar, K. Amidon, M. Casado, The Design and Implementation of Open vSwitch, USENIX Symposium on Networked Systems Design and Implementation, May 2015.
- [18] Zeek, an Open-source Network Security Monitoring Tool. <https://zeek.org/>.
- [19] E. Kfoury, J. Gomez, J. Crichigno, E. Bou-Harb, An Emulation-based Evaluation of TCP BBRv2 Alpha for Wired Broadband, Computer Communications, Vol. 161, Issue 1, September 2020.
- [20] Barefoot Networks Tofino ASIC. <https://tinyurl.com/47d42ve4>.
- [21] E. Kfoury, J. Crichigno, E. Bou-Harb, D. Khoury, G. Srivastava, Enabling TCP Pacing using Programmable Data Plane Switches, IEEE International Conference on Telecommunications and Signal Processing (TSP), July 2019.
- [22] E. Kfoury, J. Crichigno, E. Bou-Harb, Offloading Media Traffic to Programmable Data Plane Switches, IEEE International Conference on Communications (ICC), June 2020.