A Hybrid Testbed for Performance Evaluation of Large-Scale Datacenter Networks

Pilimon, Artur; Ruepp, Sarah Renée

Link back to DTU Orbit

A Hybrid Testbed for Performance Evaluation of Large-Scale Datacenter Networks

Artur Pilimon *and* Sarah Ruepp

Department of Photonics Engineering
Technical University of Denmark
{artpil, srru}@fotonik.dtu.dk

Abstract— **Datacenters (DC) as well as their network interconnects are growing in scale and complexity. They are constantly being challenged in terms of energy and resource utilization efficiency, scalability, availability, reliability and performance requirements. Therefore, these resource-intensive environments must be properly tested and analyzed in order to make timely upgrades and transformations. However, a limited number of academic institutions and Research and Development companies have access to production scale DC Network (DCN) testing facilities, and resource-limited studies can produce misleading or inaccurate results. To address this problem, we introduce an alternative solution, which forms a solid base for a more realistic and comprehensive performance evaluation of different aspects of DCNs. It is based on the System-in-the-loop (SITL) concept, where real commercial DCN equipment (switches) is interconnected with simulated DCN nodes via the SITL virtual gateway modules in the Riverbed Modeler, forming a unified hybrid real-simulated DCN setup. This testbed is complemented with high performance network testers, which can generate a mix of different real traffic streams at L2-3 and L4-7 at the rates of 1, 10 or up to 40 Gbps. These components allow performance benchmarking tests to be conducted at large scale.**

Keywords—Hybrid Datacenter Testbed; System-in-the-Loop; real-time simulation; large-scale Datacenter

I. INTRODUCTION

IT and Communication Technologies (ICT) have become the backbone of the global economy, and Datacenters (DC) as well as associated intra- and inter-DC networks are playing a role of mission critical infrastructures. Ubiquitous network connectivity, increasing global user base, new feature-rich interactive services and applications – these are some of the key reasons, affecting the global growth of the data traffic volumes, most of which constitute internal DC traffic (East-West traffic), originating and terminating within a DC [1]. As a result, existing as well as next-generation DC infrastructures are being seriously challenged in terms of the resource utilization and energy efficiency, scalability, availability, reliability and performance requirements (ultra-high bandwidth and ultra-low latency).

In order to address the aforementioned challenges, various optimizations and innovative solutions are being proposed by the research community. However, in many of the cases, these customized or new algorithms, protocols and network architectures are evaluated in a very small-scale testing environment with very specific use-cases or in a functionally

limited simulation or emulation tool. The outcome of that is such that the obtained results may be inaccurate or misleading, and many innovative solutions and technologies, incubated in the lab environment, may never be transferred to large-scale industrial deployments. Therefore, it is important to point out some common main reasons why some research ideas are not getting proper attention and presented results are often incomplete. First, lack of convincing measurements and analysis, which are applicable at scale, is the main barrier for the transfer of innovative research. Second, assumptions and simplifications, introduced in the design phase of an experiment or a simulation model due to lack of resources, such as uniformly distributed traffic, no realistic background traffic or network processing effects, no possibility to deploy a real DC- or cloud-native application or service to test, lead to discrepancies in the results, compromising their credibility. Third, there are often two extremes of resource-limited research, namely only experimental validation using a small-scale test setup of hardware, which can hardly be extended to emulate a large-scale context, or a simulation-based analysis, which can scale better and can be reproduced, but suffers from unavailability of DC-specific traffic sources (real servers with fully-featured applications) and realistic usage patterns.

A way to rectify such a situation would be to perform the initial evaluation and small-scale testing of the developed ideas as a Proof-of-Concept (PoC) in a local research lab, and then to access a production scale DCN testing environment in order to conduct an extensive large-scale validation. However, only a limited number of research groups have access to such high-end testing facilities, and building up anything similar in a local research lab is unrealistic both financially and timewise.

In this work we introduce an alternative solution, which could form a solid base for a more realistic and comprehensive performance evaluation of different aspects of DCNs. It is based on the System-in-the-loop (SITL), also known as Hardware-in-the-loop (HIL), concept, where real commercial DCN equipment (e.g., switches) is interconnected with simulated DCN nodes via the SITL virtual gateway modules in the Riverbed Modeler [2], forming a unified real-simulated DCN setup. This testbed is complemented with high performance network testers, generating a mix of different traffic streams at L2-3 and L4-7 at the rates of 1, 10 or up to 40 Gbps. The PoC of building and feasibility of using such a hybrid setup for DCN performance studies was demonstrated in [3], where the communication of high performance L2-7 network testers and software traffic generators through a

simulated DCN environment was assessed. The main contributions of this work are the following: 1) a hybrid (real-simulated) DCN testbed, consisting of real commercial and simulated network switches, and high performance hardware traffic generators (network testers) for more realistic stress-testing; 2) an outline of potential application scenarios of using such a testbed for large-scale experimental scalability and performance studies.

The remainder of this paper is organized as follows: section II presents the state of the art hybrid solutions for the DCN-specific large-scale studies. In section III, the hybrid DCN hardware and simulation setup is explained. Section IV outlines the use-cases considered for large-scale DCN studies using this setup and the paper is concluded in section V.

II. RELATED WORK

There are different types of network research testbeds proposed, which can be categorized into local physical [4], emulation-based [5][6] and hybrid [7][8] testbeds, as well as globally distributed purely virtualized testbeds such as PlanetLab [9] or OneLab [10]. These experimental network research environments have been built to enable more comprehensive studies of a larger spectrum of network communication problems. However, the focus of this overview is on the existing combined (using a combination of hardware and some form of emulation or simulation techniques) systems, created for large-scale experimental studies of DCN environments, where the scalability and functional capabilities of the setup are the most important features. Therefore, the aforementioned testbed setups [4-10] are out of the scope of this discussion, because they are not specifically targeting DCN-oriented large-scale environments, and a DCN-specific setup, presented in [11], will be described next.

Benet in [11] recently proposed a Cloud testbed, called OpenStackEmu, which is built as a combination of a well-known OpenStack platform, a CORE (Common Open Research Emulator) network emulator and an external SDN (Software Defined Networking) controller [11]. The real-time network emulator uses virtual TUN/TAP (network TUNnel and tap) interfaces (virtual network kernel devices) in order to connect external physical devices to it and to be able to inject external traffic into the emulated DCN. The presented test setup consists of a set of external physical devices, connected to the machine running the CORE emulator: a basic set of OpenStack infrastructure nodes, as well as an external node with an SDN controller and one more node with background traffic generation software. This hybrid system allows testing various realistic DC networking scenarios with VM deployment, live VM migration, service orchestration and it can be used to study the impact of the DCN topologies, network characteristics and other aspects. However, it is not obvious how well such an emulated setup would scale, because the network devices (Open virtual Switch, OVS) as well as hosts are emulated using the OS-level virtualization principle, where each virtual device is modelled as a light-weight Linux container (represents a VM). This approach is more resource-demanding than just a simulated counterpart, and that impacts the scalability of the setup. Our work is different in such a way that the DCN itself is composed of real physical (electronic and one optical) and simulated network switches, which are communicating via a set of electrical-optical interfaces, installed on the high performance server, hosting the simulation environment with modelled devices.

III. A HYBRID DCN TESTBED ARCHITECTURE

A high-level architecture of the hybrid real-simulated testbed setup is shown in Figure 1. As it can be seen, the setup consists of a complete (symmetric) 16-Hypercube structure, half of which is composed of 8 commercial HP Aruba ToR (Top of Rack) switches, and half is simulated in a Riverbed Modeler tool with real-time simulation kernel and using virtual SITL gateway modules for connectivity to the external physical hardware (or software processes). The important components, differentiating this hybrid setup from the others, are summarized in Table I.

TABLE I.   KEY FUNCTIONAL ELEMENTS OF A DCN TESTBED

Functional Element(s)	Purpose
Simulation platform with SITL	Virtual hardware-software linking for real-time communication with real DC equipment
High performance Xena Bay/Xena Scale network testers [12]	Accurate generation of large volumes of data with mixed traffic patterns, many simultaneous connection-oriented and connection-less flows. Creating a more realistic DCN communication context, difficult to achieve with software gen.
Integration of simulated, electrical and optical DC network equipment	Possibility to evaluate the scalability of different DCN architectures by stress-testing with real communication flows; assessment of the benefits of optical switching in hybrid DCN architectures.

A short summary of the technical characteristics of this hybrid DCN setup is provided in Table II. However, these technical specifications are not definitive and any combination of powerful enough hardware can be used. The same applies to the choice of a DCN topology: we modelled and assembled this topology, because Hypercube, HyperX, Flattened Butterfly and other direct-connection structures [13] have been widely used in HPC (High Performance Computing) environments due to their well-known features such as high scalability and low latency. Such flattened, high-radix topologies in combination with optical switching may be considered as promising candidates for the next-generation large-scale architectures.

The "physical" part of the DCN structure, namely one 8-node sub-cube (Figure 1, *left*), has two types of physical layer connections: interconnections with the physical and simulated ToR switches within the real-simulated 16-Hypercube using Multi-Mode Fiber links (MMF) and optical shortcut connections through Polatis 24x24 free-space Optical Circuit Switch (OCS) using Single-Mode Fiber (SMF) links.

The simulation environment is hosted on a high performance server (Figure 1, *right*), capable to run a large-scale simulation model with multiple virtual SITL gateway modules to connect to the external devices via available electrical-optical interfaces. In this particular test setup, we used IBM 2U X3690 X5 server with Intel Xeon X5650 2-CPU (24 logical cores), 128 GB RAM and 4 installed PCI-e NICs (Network Interface Card) with up to 4 ports on each.
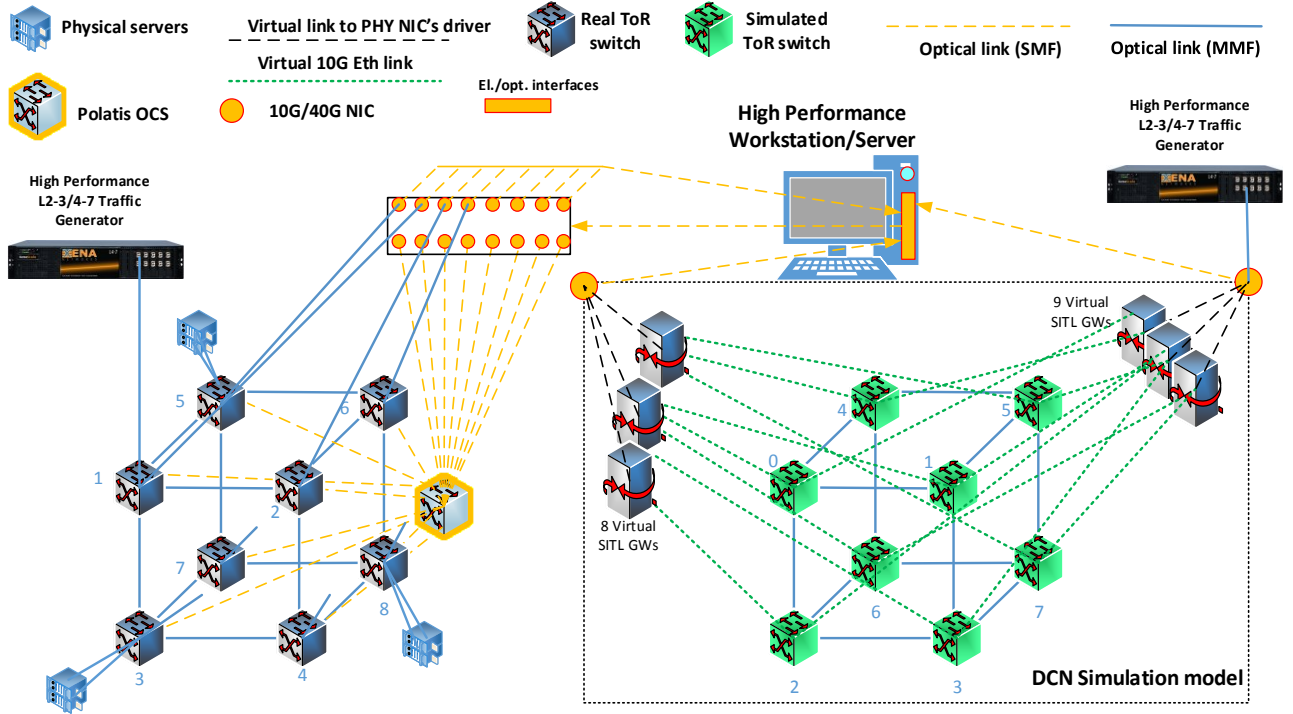
Figure 1 A high level architecture of a hybrid DCN research testbed

Please note that on Figure 1 all electrical-optical interfaces of the server machine are shown on 1 board (with 16 ports) for presentation clarity. The ToR-simulation-server connections of the bottom plane of the real 8-Hypercube (3, 4, 7 and 8 switches) are also not shown for the visual clarity of the presentation, but these connections are present.

TABLE II.   PERFORMANCE/FUNCTIONAL CHARACTERISTICS OF A DCN TESTBED

Network node type	Characteristics			
	Network interface type/rate (Gbps)	*Nr. El./opt. interfaces*	*Nr. Optical interfaces*	*Optical switching capabilities*
ToR switches	SFP/SFP+, 1/10G	16	-	-
Polatis OCS	Signal bit-rate/format independent up to 100G, 400G and beyond	-	48	OCS, Fiber Cross connect
Network testers	SFP+/1-10G, QSFP/CFP 40/100G	6-12+ (custom.)	-	-
High Perf. Server	SFP/SFP+, 1-10G	Up to 16[a]	-	-

[a.] For IBM X3690 X5 Intel Xeon X5650 24-core machine, with 4 PCI-e slots, 128 GB RAM

The network switches of the simulated 8-node sub-cube (Figure 1, *right*) maintain two types of logical connections: regular simulation logical links between simulated ToR switches and *virtual SITL 10Gbps links*. The latter are used by the simulation kernel to create a logical channel from the simulated Ethernet-enabled device (e.g., a router, a switch or a workstation) to the driver of the physical NIC's port to be able

to perform bidirectional real-time extraction and conversion of the data packets (Ethernet frames). The performance of the simulated environment can be further enhanced by splitting a large-scale simulation model into clusters and parallelizing the setup for even larger scaling. In addition, the use of Multi-Threading (MT) of the simulation model within each cluster can further increase performance and reduce such critical parameter as packet conversion and buffering latency, which was thoroughly investigated in the previous work in [3].

The simulation model of an 8-node sub-cube of the hybrid DCN is depicted in Figure 2, where we can observe multiple virtual SITL modules, associated with the corresponding simulated nodes. The scalability aspect of this hybrid setup is reflected in the capacity of the simulation environment to scale to large DCN topologies, which is only subject to the hardware limitations of the server nodes used. As it can be seen in Figure 3, we can create a sophisticated DCN architecture of a larger scale in the simulation environment and complement it with a smaller set of real DCN equipment. In this scenario we created a model of asymmetric (incomplete) Hypercube by combining a 44-node Hypercube with 4 additional physical commercial DCN switches, forming an incomplete (deemed to be more efficient from the scalability point of view) 48-node Hypercube. Moreover, it has 4 fast optical shortcut connections through Polatis OCS.

Integration of the Optical Circuit Switch (Polatis) with the hybrid testbed was performed to be able to study the potential benefits (relative gain in terms of latency reduction and increase of throughput) of introducing optical switching in DCNs that can be assessed more accurately in larger test network scenarios.
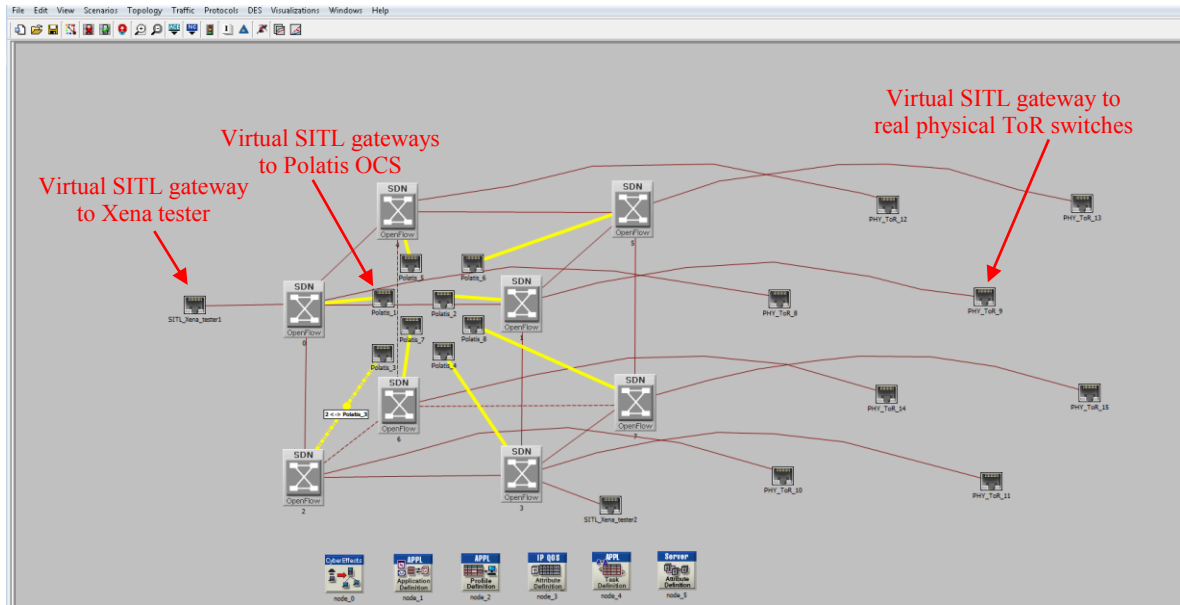
Figure 2 A SITL simulation model of 8-Hypercube structure with connections to the physical 8-Hypercube, a Polatis OCS and Xena testers
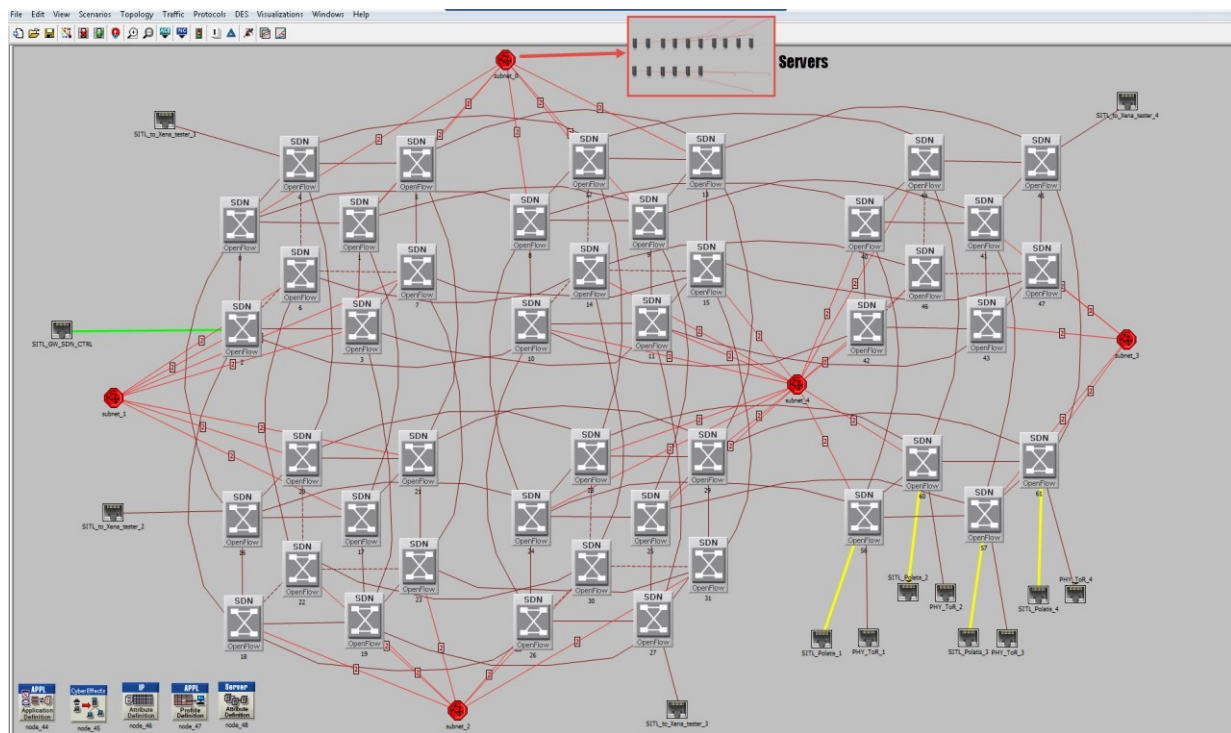

Figure 3 A SITL simulation model: 44-Hypercube, connectivity to Polatis OCS, connectivity to physical ToR switches and Xena network testers

Preliminary testing was conducted by measuring the end-to-end communication latency (using the capabilities of Xena testers) in the hybrid real-simulated Hypercube in two settings: with Polatis OCS fast optical shortcuts and without Polatis OCS. The results are shown in Figure 4, and, by using fast reconfigurable optical shortcuts, the latency, even in a 16-switch DCN setup, can be reduced by more than 50%, considering a long 6-hop path spanning the Simulated switches (0, 1, 3, 7) and Real switches (1, 2, 6) as compared to a short path switch_0 (simulated) $\leftrightarrow$ Polatis OCS $\leftrightarrow$ switch_1 (real).

Please note that this measurement result shows a normalized and 1-second averaged latency to show the relative gain of an optical bypass circuit when steering the traffic.

Additional clusters of simulated DC servers are added to the simulation model to create background traffic within the modelled part of the DCN, based on user-defined demands. Finally, the presented testbed is also SDN-ready (Software Defined Networking), since the commercial HP Aruba ToR
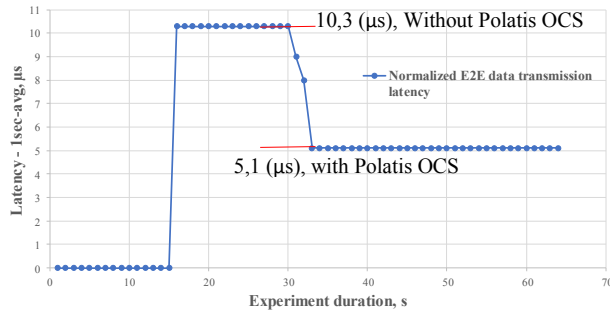
Figure 4 Normalized latency measurement in a hybrid DCN testbed

switches, simulation switches as well as Polatis OCS are all OF-enabled (Open Flow protocol).

IV. TESTING SCENARIOS AND SCALABILITY ASPECTS

Examples in Figures 2 and 3 show that we can examine a large number of communication scenarios under more realistic networking constraints. The following (but not limited to) use-cases and DCN-related aspects can be studied using this setup:

- *Real servers and simulated DCN*. A small cluster of real application servers can be deployed, running a set of test applications (e.g., multi-tier services), which can be used for more accurate performance and quality of service characterization. For instance, a video streaming server can be deployed to assess the impact of different DCN parameters on the Quality of user experience. That would be more difficult to achieve in a small hardware-only setup (no DC scale) or in a simulation-only setup, where we cannot observe the video quality.

- *Real servers, simulated DCN and Xena testers*. The same tests can be performed in a more realistic setting by specifically overloading certain parts of the DCN using high performance testers. That would be difficult to achieve using generic software traffic generators, e.g., if we would like to stress-test a DCN, which is equipped with 10/40 Gbps links. In addition, Xena testers support traffic generation using capture-replay method, which allows recording real communication sessions (e.g., skype, web browsing) and replaying them to many concurrent flows.

- *Optical Circuit Switch and simulated DCN*. There is a large potential for the emerging integration of optics in the Datacenter world, where scaling and latency effects must be studied carefully. The latency reduction gain can be assessed more accurately in the context of a large-scale DCN environment while using fast optical shortcut links to offload the bottleneck areas of the network (e.g., separating elephant from mice flows).

- *Simulated-Real-Simulated*. Inter-DCN communication aspects can be studied by combining two simulated DCNs with a real edge device (or a set of edge routers). For instance, performance aspects of geographically distributed DCNs can be analyzed using this method.

- *Integration with SDN*. This scenario offers great opportunities for comprehensive analysis of different traffic routing/forwarding/engineering mechanisms.

V. CONCLUSION

In this work, we present an approach for building a hybrid DCN testbed for large-scale performance and scalability analysis. That gives a possibility to exploit realistic communication patterns under mixed traffic flows, to stress-test particular areas of a DCN topology of interest by combining real DCN hardware with simulated equipment and high performance network testers. The setup is used to evaluate the introduction of a fast Optical Circuit Switch in terms of latency, allowing studying the benefits of optical switching in a hybrid DCN context, as well as supports SDN-based control. Further non-exhaustive use-cases are presented as well.

REFERENCES

[1] Cisco Systems, Inc., "Cisco Global Cloud Index: Forecast and Methodology, 2014–2019," Cisco, 2015.

[2] Riverbed Technology, "Riverbed Modeler," [Online]. Available: http://www.riverbed.com/products/performance-management-control/network-performance-management/network-simulation.html

[3] S. Ruepp, A. Pilimon, J. Thrane, M. Galili, M. Berger, and L. Dittmann, "Combining Hardware and Simulation for Datacenter Scaling Studies," in *Proc. 2017 IEEE 21st International Conference on Optical Network Design and Modeling, ONDM*, 2017.

[4] S. K. Dhurandher, I. Woungang, I. Uppal, H. Bhanushali, and D. Gupta, "A Dot Net Framework based Physical Testbed for Ad hoc Network Routing Protocols," *in Proc. 2011 24th Canadian Conference on Electrical and Computer Engineering (CCECE),* 2011, pp. 229–233.

[5] R. Lübke, D. Schuster, and A. Schill, "Reproducing network conditions for tests of large-scale distributed systems," *in Proc. - 13th IEEE/ACM Int. Symposium on Cluster, Cloud, and Grid Comput.*, 2013, pp. 74–77.

[6] Q. Le-Trung, "Towards an IoT network testbed emulated over OpenStack cloud infrastructure," *in Proc. - 2017 International Conference on Recent Advances in Signal Processing, Telecommunications and Computing*, 2017, pp. 246–251.

[7] A. S. Leger, J. Spruce, T. Banwell, and M. Collins, "Smart grid testbed for Wide-Area Monitoring and Control systems," *in Proc. - IEEE Power Eng. Society Transmission and Distribution Conf.*, July 2016, pp. 1–5.

[8] H. Gao, Y. Peng, K. Jia, Z. Wen, and H. Li, "Cyber-Physical Systems Testbed Based on Cloud Computing and Software Defined Network," *in Proc. - 2015 International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, IIH-MSP 2015, 2016, pp. 337–340.

[9] PlanetLab Consortium, "PLANETLAB: An open platform for developing, deploying and accessing planetary-scale services," [Online].

[10] OneLab Consortium, "OnaLab: Future Internet Testbed," [Online]. Available: https://onelab.eu/.

[11] C. H. Benet, R. Nasim, K. A. Noghani, and A. Kassler, "OpenStackEmu - A Cloud Testbed Combining Network Emulation with OpenStack and SDN," *14th IEEE Consum. Commun. Netw. Conf.*, pp. 566–568, 2017.

[12] Xena Networks, "L2-3 (Xena Bay) and L4-7 (Xena Scale) Test Platforms," [Online]. Available: http://www.xenanetworks.com.

[13] D. Abts and J. Kim, "High Performance Datacenter Networks: Architectures, Algorithms, and Opportunities," *Synth. Lect. Comput. Archit.*, vol. 6, no. 1, pp. 1–115, Mar. 2011.