Training and Teaching Students and IT Professionals on High-throughput Networking and Cybersecurity using a Private Cloud

Dr. Jorge Crichigno, University of South Carolina

Jorge Crichigno is an Associate Professor in the Department of Integrated Information Technology (IIT), College of Engineering and Computing (CEC), at the University of South Carolina (USC). Dr. Crichigno's research focuses on practical implementation of high-speed networks and network security, custom protocol development using P4 switches, experimental evaluation of congestion control algorithms, and scalable flow-based intrusion detection systems. He is the Principal Investigator of multiple research initiatives involving high-speed and next-generation networks. Dr. Crichigno has served as a reviewer and a TPC member of journals and conferences, such as the IEEE Transactions on Mobile Computing, IEEE Access, IEEE Globecom, and others. He has also served as a panelist for the National Science Foundation, for programs related to advanced cyberinfrastructure and undergraduate and graduate education. He is an ABET Evaluator representing the IEEE.

Prof. Elias Bou-Harb, University of Texas at San Antonio

Dr. Elias Bou-Harb is currently the Associate Director of the Cyber Center For Security and Analytics at UTSA, where he leads, co-directs and co-organizes university-wide innovative cyber security research, development and training initiatives. He is also an Associate Professor at the department of Information Systems and Cyber Security specializing in operational cyber security and data science as applicable to national security challenges. Previously, he was a senior research scientist at Carnegie Mellon University (CMU) where he contributed to federally-funded projects related to critical infrastructure security and worked closely with the Software Engineering Institute (SEI). He is also a permanent research scientist at the National Cyber Forensic and Training Alliance (NCFTA) of Canada; an international organization which focuses on the investigation of cyber-crimes impacting citizens and businesses. Dr. Bou-Harb holds a Ph.D. degree in computer science from Concordia University in Montreal, Canada, which was executed in collaboration with Public Safety Canada, Industry Canada and NCFTA Canada. His research and development activities and interests focus on operational cyber security, attacks' detection and characterization, malware investigation, cyber security for critical infrastructure and big data analytics. Dr. Bou-Harb has authored more than 80 refereed publications in leading security and data science venues, has acquired state and federal cyber security research grants valued at more than $4M, and is the recipient of 5 best research paper awards, including the prestigious ACM's best digital forensics research paper.

Mr. Elie Kfoury, University of South Carolina

Elie Kfoury is a second-year PhD student in computer science at the University of South Carolina, USA. He is a member of the CyberInfrastructure Lab (CI Lab), where he developed training materials for virtual labs on high-speed networks (TCP congestion control, WAN, performance measuring, buffer sizing), cybersecurity, and routing protocols. He previously worked as a research and teaching assistant in the computer science and ICT departments at the American University of Science and Technology in Beirut. His research focuses on telecommunications, network security, blockchain, Internet of Things (IoT), and programmable data plane switches.

Mr. Jose Gomez, University of South Carolina

Jose Gomez is a Computer Engineering PhD student at the University of South Carolina in the United States of America. For the last three years, he worked as a researcher and teaching assistant in the School of Engineering at the Catholic University in Asuncion.

Antonio Mangino, The University of Texas at San Antonio

Antonio Mangino is currently pursuing a Master's degree in Information Systems and Cyber Security at The University of Texas at San Antonio. He received his B.S. in Computer Science from Florida Atlantic University (FAU) in 2019. As a member of the Cyber Threat Intelligence Laboratory at Florida Atlantic

University, Antonio has worked on the development of various network and cyber security projects, with a focus on the IoT paradigm. His research interests include network analysis, operational cyber security and information security.

Training and Teaching Students and IT Professionals on High-throughput Networking and Cybersecurity using a Private Cloud

Abstract. This paper describes the deployment of a private cloud and the development of virtual laboratories and companion material to teach and train engineering students and Information Technology (IT) professionals in high-throughput networks and cybersecurity. The material and platform, deployed at the University of South Carolina, are also used by other institutions to support regular academic courses, self-pace training of professional IT staff, and workshops across the country. The private cloud is used to deploy scenarios consisting of high-speed networks (up to 50 Gbps), multi-domain environments emulating internetworks, and infrastructures under cyber-attacks using live traffic.

For regular academic courses, the virtual laboratories have been adopted by institutions in different states to supplement theoretical material with hands-on activities in IT, electrical engineering, and computer science programs. Topics include Local Area Networks (LANs), congestion-control algorithms, performance tools used to emulate wide area networks (WANs) and their attributes (packet loss, reordering, corruption, latency, jitter, etc.), data transfer applications for high-speed networks, queueing delay and buffer size in routers and switches, active monitoring of multi-domain systems, high-performance cybersecurity tools such as Zeek's intrusion detection systems, and others.

The training platform has been also used by IT professionals from more than 30 states, for self-pace training. The material provides training on topics beyond general-purpose networks, which are usually overlooked by practitioners and researchers. Additionally, the platform has supported workshops organized across the country. Workshops are co-organized with organizations that operate large backbone networks connecting research centers and national laboratories, and colleges and universities conducting teaching and research activities.

1. Introduction

General-purpose enterprise networks are capable of transporting basic data, e.g., emails, multimedia, and web content. However, these networks face many challenges when moving petabytes (PBs) of scientific data, e.g., genomic, climate, imaging, and high-energy physics, [1]. As a response, network architects have developed the concept of a Science Demilitarized Zone (Science DMZ or S-DMZ) [2] as parts of a vision for a "cyber-highway system without stoplights" for science data.

As the popularity of high-speed networks capable of moving data at tens or even hundreds of terabits per seconds surges, the need for trained cyberinfrastructure engineers with the requisite skills to condition these high-performance infrastructures has increased tremendously. However, today's network engineers, researchers, and practitioners are mostly trained to operate enterprise networks (referring to regular commercial networks herein). According to the 2017 NSF Campus Cyberinfrastructure (CC*) PIs meeting survey [3], many participants across the country, from large to small institutions, noted significant challenges trying to find appropriate cyberinfrastructure (CI) engineers, see Table 1.

Table 1. Concerns raised during the 2017's NSF CC* PIs meeting [3].

#	Concerns by PIs, Co-PIs, and attendees of 2017 NSF CC* meeting
1	"Very difficult to find, or nonexistent - difficult to retain (CI engineers)"
2	"Largest challenge was in the area of time to hire... ended up taking 10 months… (difficult to find CI engineers with the right skills)"
3	"Candidates should have hands-on knowledge of networking, at least bachelor degree, and certifications in networking and security"
4	"Combination of education and experience"
5	"At least one tour of duty as an intern or apprentice"
6	"System & network engineering, user support experience, good communication (written and presentation)…"
7	"Training in routing and switching (e.g., Juniper, Cisco), a minimal knowledge and/or training in security (e.g., Palo Alto or similar), cabling"
8	"Working knowledge of theory and practice underlying VLAN/LAN/WAN network operations"
9	"Working with researchers to identify areas where their research can benefit from high-end technologies such as HPC, Science DMZ, Data Transfer Node (DTN), Big Data platforms"
10	Difficult to find, preferred qualification: combination of "Bachelor degree" and "certifications in networking and security"

To address this skills gap, this paper describes a project that has deployed a private cloud and developed hands-on experiential learning, training, and research material that is scalable and cost effective. The project is led by the University of South Carolina (USC), in conjunction with the University of Texas at San Antonio (UTSA) and University of South Florida (USF). After the first year of the project, the virtual labs and companion material running on the virtual training platform has been used by hundreds of professionals and students from more than 30 states.

2. Need for Teaching / Training on High-speed / Big Data Transfer Network

2.1. Motivation and Network Elements

The main network architecture for big science data transfers is the Science DMZ [4]. Elements of the Science DMZ are also used in other network designs. The Science DMZ is intended to handle the exponentially increasing amount of data being transferred across (scientific) networks. As an example, the ESnet network [5] (U.S. national backbone connecting research centers and national laboratories) alone already transported almost 250 petabytes of science flows in 2018, see Fig. 1.

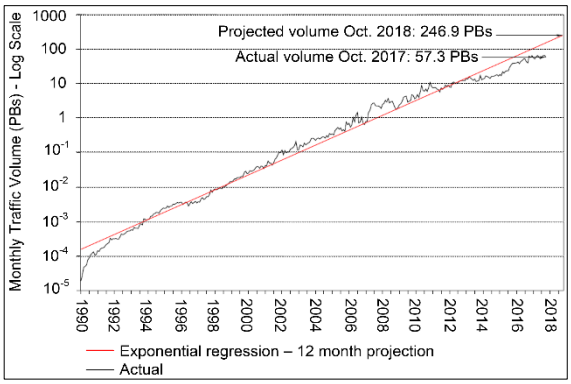

Fig. 1. Monthly average traffic volume, ESnet.

A Science DMZ is typically co-located next to a main enterprise network. However, the path from the Science DMZ to the WAN involves as few devices as possible to avoid friction between nodes exchanging data across large distances. Having a friction-free path is essential for big data transfers, otherwise, packets can be dropped or received out of order (which may cause TCP throughput reduction by half). Fig. 2 shows the consequences of friction along a wide area network (WAN) path, from [2]. Namely, this plot shows TCP throughput for a device receiving data over a 10 Gbps path (packet loss rate of 1/22,000, or 0.0046%). The purple curve shows the throughput in a loss-free environment and the green curve shows the theoretical maximum. Beyond LAN transfers, these results show very rapid throughput collapse to under 1 Gbps.

A Science DMZ example is shown in Fig. 3(a). Its essential elements are data transfer nodes (DTNs), perfSONAR nodes, offline security monitoring, and a friction-free high-speed path to the wide area network (WAN). Note the absence of CPU-intensive devices between the Science DMZ and the high-speed WAN. The rationale of this design is to avoid any device that may drop packets.

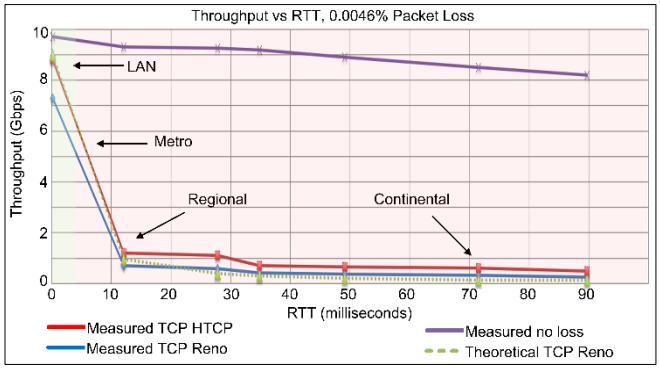

Fig. 2. Throughput vs RTT for two DTNs connected via 10 Gbps path. Performance of two TCP versions are shown: Reno (blue) and HTCP (red). The theoretical performance considering packet loss (green) and under a loss-free assumption (purple) are also shown [2]. As RTT increase, the performance decreases by more than 50% beyond the metro area.

Friction-free network WAN path: Two nodes (DTNs) are connected by a WAN composed of high-end routers and switches. This setup requires ISP engineers to understand big flows. Namely, network devices must be able to forward packets at a high-speeds (10-100 Gbps) and have large buffer sizes to absorb transitory packet bursts and prevent losses. The path should have no devices that may add excessive delays or cause packet sequencing problems; e.g., firewalls, intrusion prevention systems (IPSs). Internet2 is a typical national research and engineering network (NREN) [11]. However, most instructional materials on WAN are theoretical, i.e., having a real WAN infrastructure is very costly and impractical for teaching.

Dedicated, high-performance DTNs: These devices represent end points for data transfers [4] and are typically Linux devices built/configured for receiving WAN transfers at high speed. DTNs use

optimized TCP congestion-control methods for high-performance, large receive buffer sizes, and other extensions. However, there is no structured material on setting/managing the TCP/IP stack for high-performance.

Targeted security: Enterprise networks use inline devices such as firewalls and IPSs to collect state and inspect application-layer payloads in real-time. However, these systems are omitted in Science DMZ setups because of substantial impact on flow rates. Instead, high-speed networks can be protected by a combination of router's access-control lists (ACLs) and offline intrusion detection systems (IDSs) [4]. While there are plenty of instructional materials for enterprise firewalls and online IPSs, learning materials for friction-free security are limited or non-existent.

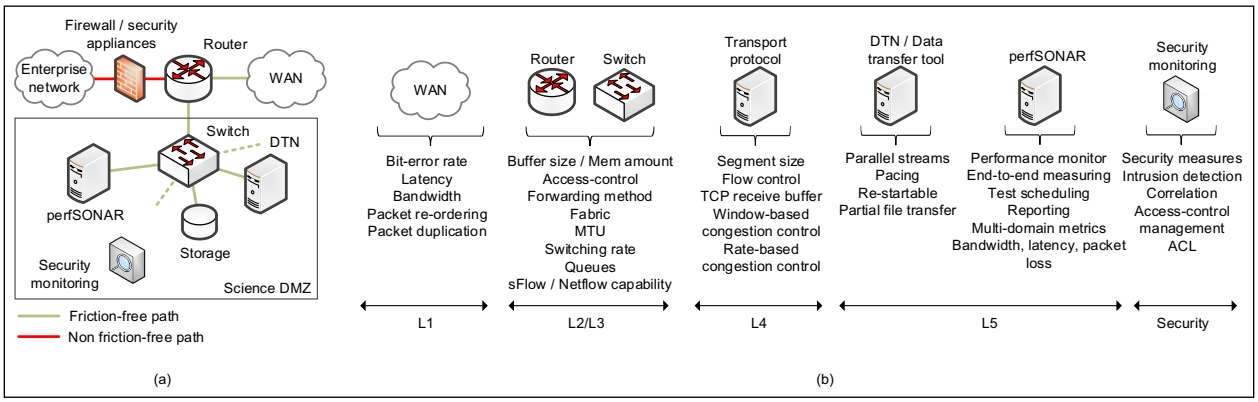

Fig. 3. (a) A Science DMZ co-located to the regular enterprise network. Notice the absence of firewall or any inline security appliance in the friction-free path. (b) Features of Science DMZ's devices that must be considered for teaching and training.

Performance measurement and monitoring point: It is essential to maintain the health of end-to-end paths during WAN transfers. This requirement is usually not implemented in enterprises as they do not cooperate with ISPs to achieve this goal (enterprise network tools are intra-domain: SNMP, Syslog, Netflow). On the other hand, perfSONAR is a multi-domain tool which allows enterprises, ISPs, and NRENs to cooperate and provide mechanisms to track WAN performance metrics such as throughput, packet loss, and latency [19]. The number of perfSONAR node deployments has also grown to over 2,000 in the last few years, see Fig. 4. However, PerfSONAR is still not covered in academic programs and trainings, e.g., key texts in the field [6], [7], [8] do not mention perfSONAR.

Fig. 4. perfSONAR nodes deployed as of June 2017.

2.2. Training and Education Workforce Challenges

Most available training materials on Science DMZs and high-throughput networks are found on the ESnet website [9] and associated workshops [10], [11]. However, the most essential topics shown in Fig. 3(a) are overlooked by cyberinfrastructure professionals, e.g., topics listed in Table 2 (Science DMZs column) are largely ignored. Hence this project deployed a private cloud and

developed hands-on vLabs along with laboratory manuals on these topics, which have been incorporated in formal and informal education environments.

Table 2. Differences between enterprise networks and Science DMZs.

Topic	Enterprise Networks	Science DMZs
L1: WANs	Limited bandwidth by commercial ISPs; routers/switches not optimized for performance; congestion; routing achieved independently by ISPs; typical frame size is 1,500 bytes	WAN is typically provided by Internet2/NRENs[1]; 10-100 Gbps paths; routers/switches optimized for large flows; predictable performance; end-to-end routing optimization; jumbo frames supported (9,000 bytes)
L2/L3: switches and routers	Rates lower than 10 Gbps; buffer memory amount equals $\frac{BDP}{\sqrt{N}}$ [12]; typical forwarding method is cut-through; many switches use shared memory buffer allocation; popular switches used shared-memory and even bus fabrics; input-only queues for buffering acceptable for some networks	Rates higher than 10 Gbps (e.g., 40 and 100 Gbps); buffer memory at least equal to BDP; forwarding method must be store-and-forward; buffer allocation should be port-based; recommended fabric is crossbar; buffering should include input and output queues
L4: TCP / transport protocol	Segment size 1,500 bytes; stop-and-wait protocol behavior is acceptable; TCP buffer size has small impact on performance; window-based congestion control widely used; no pacing; no parallel streams	Segment sizes should be as large as possible; pipelined behavior is required; TCP buffer sizes must be greater than BDP; rate-based congestion control has positive impact; pacing improves throughput; parallel streams is essential
L5: applications	Variety of applications; general-purpose data transfer tools (SCP, FTP); single-domain monitoring application (SNMP, Syslog)	Small set of applications; specialized data-transfer tools (Globus); multi-domain performance measurement and monitoring application (perfSONAR)
Security	Online devices (IPSs, firewalls) are typical; IDS and ACLs used to complement IPS and firewalls; flow-based IDS not used	Online devices cannot be used; ACL used as primary defense; flow-based IDS is attractive but not well understood by CI community; protocols used in a non-traditional way (e.g., Netflow)

[1]NREN: National research and education networks; [2]*BDP*: bandwidth-delay product. [3]*N*: number of concurrent flows thorough the router/switch, typically thousands or millions in enterprise networks.

3. Virtual Labs

3.1 Platform

The project relies on a private cloud from the Network Development Group (NDG) [13]. The platform serves as a framework where vLabs are developed. The platform supports remote-access capability (REC) to virtual lab equipment from the Internet. For example, in order to perform a comparison study of different congestion control algorithms for high-throughput high-latency networks, students are able to reserve a pod that

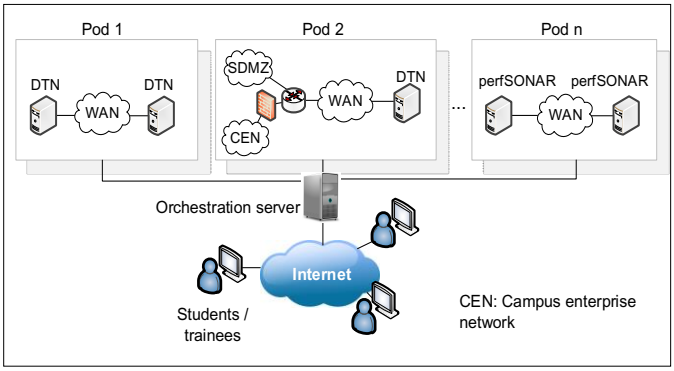

Fig. 5. Infrastructure to support pods.

automatically creates two DTNs connected by a WAN (bit-error rates, latency, jitter, and other parameters will be adjustable). Fig. 5 shows the setup. Once students have entered this environment, the virtual equipment is deployed via an orchestration server and ready to use without the need for any configuration or package installation. Detailed laboratory manuals are also available to guide students during these exercises. Note that this laboratory environment is transparent, and each experiment uses the reserved equipment pod. A pod is a set of virtual appliances used to complete a laboratory experiment. All pods are hosted at USC's datacenter and are available 24/7.

The private cloud used to support vLabs relies on physical servers. The physical resources can be classified in compute capability (CPU cores), storage (non-volatile memory), and RAM memory. Table 3 summarizes these resources.

Table 3. Servers supporting the private cloud.

Device	Cores	Storage (TBs)	RAM (GB)	Notes
Server 1 (management server)	20	4.8	128	Hosts orchestration server
Server 2 (hosting vLabs pods)	32	4.8	512	Hosts pods' VMs
Server 3 (hosting vLabs pods)	32	1.92	768	Hosts pods' VMs
Server 4 (hosting vLabs pods)	32	1.92	768	Hosts pods' VMs
Total	**116**	**8.08**	**2,176**	

3.2 Virtual Labs

The project has developed vLabs and related manuals, see Table 4. The virtual appliances (virtual routers, virtual firewalls, etc.) are exported using the standard OVA file format.

Table 4. Virtual laboratories developed during the first year of this project.

Lab	Network Tools and Protocols Lab Series
1	Introduction to Mininet
2	Introduction to Iperf3
3	Emulating WAN with NETEM I: Latency, Jitter
4	Emulating WAN with NETEM II: Packet Loss, Duplication, Reordering, and Corruption
5	Setting WAN Bandwidth with Token Bucket Filter (TBF)
6	Understanding Traditional TCP Congestion Control (HTCP, Cubic, Reno)
7	Understanding Rate-based TCP Congestion Control (BBR)
8	Bandwidth-delay Product and TCP Buffer Size
9	Enhancing TCP Throughput with Parallel Streams
10	Measuring TCP Fairness
11	Router's Buffer Size
12	TCP Rate Control with Pacing
13	Impact of MSS on Throughput
14	Router's Bufferbloat
Lab	**perfSONAR Monitoring Lab Series**
1	Configuring Administrative Information Using perfSONAR Toolkit GUI

2	PerfSONAR Metrics and Tools
3	Configuring Regular Tests Using perfSONAR GUI
4	Configuring Regular Tests Using pScheduler CLI Part I
5	Configuring Regular Tests Using pScheduler CLI Part II
6	Bandwidth-delay Product and TCP Buffer Size
7	Configuring Regular Tests Using a pSConfig Template
8	perfSONAR Monitoring and Debugging Dashboard
9	pSConfig Web Administrator
10	Configuring pScheduler Limits
Lab	**Zeek / Bro Lab Series**
1	Introduction to the Capabilities of Zeek
2	An Overview of Zeek Logs
3	Parsing, Reading and Organizing Zeek Files
4	Generating, Capturing and Analyzing Network Scanner Traffic
5	Generation, Capturing and Analyzing DoS and DDoS-centric Network Traffic
6	Introduction to Zeek Scripting
7	Advanced Zeek Scripting for Anomaly and Malicious Event Detection
8	Preprocessing of Zeek Output Logs for Machine Learning
9	Developing Machine Learning Classifiers for Anomaly Inference and Classification
10	Profiling and Performance Metrics of Zeek

The Network Tools and Protocols Lab Series. This lab series was developed using a single OVA image (one virtual machine), using Mininet [14]. Mininet is a virtual testbed enabling the development and testing of network tools and protocols, running real protocol stacks (for this lab series, a Linux lightweight Ubuntu was used). Fig. 6 shows a 10 Gbps pod used to test the queueing delay on a three-node network. The bandwidth capacity the platform is capable of emulate is ~50 Gbps. Note that the private cloud can deploy hundreds of such pods on demand, thus supporting large number of users conducting experiments at the same time.

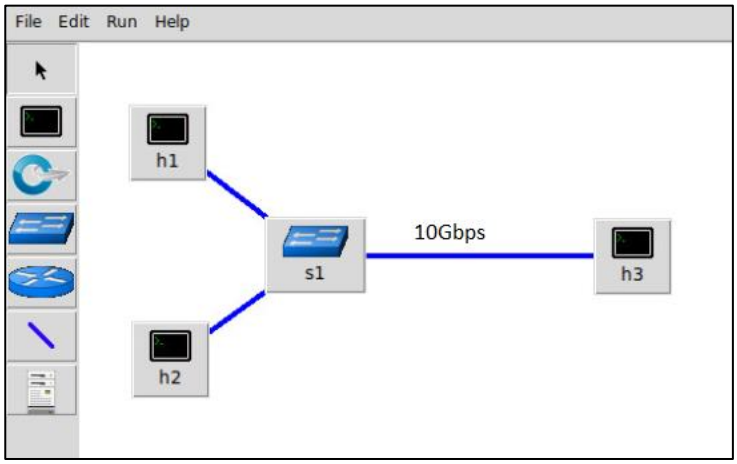

Fig. 6. A pod to emulate a 10 Gbps network. The virtual devices h1, h2, and h3 are end devices (hosts), and s1 is a virtual switch.

The main tools used for measuring performance, emulating WANs, and establishing bandwidth capacities are:

- iPerf [15]: a real-time network throughput measurement tool. It is an open source, cross-platform client-server application that can be used to measure the throughput between the two end devices. A typical iPerf output contains a timestamped report of the amount of data transferred and the throughput measured.
- Network Emulator (NETEM) [16]: a Linux network emulator for testing the performance

of real applications over a virtual network. The virtual network may reproduce long-distance WANs in the lab environment. These scenarios facilitate the test and evaluation of protocols and devices from the application layer to the data-link layer under a variety of conditions. NETEM allows the user to modify parameters such as delay, jitter, packet loss, duplication and re-ordering of packets.

- Token Bucket Filter (TBF) [17]: a Linux application implementing the token bucket algorithm. It is a queuing discipline used in conjunction with the Linux Traffic Control (tc) to shape traffic. For the network emulations, TBF is used to set the bandwidth of individual links in the network.
- Sysctl [18]: a Linux's tool for dynamically changing parameters in the operating system. It allows users to modify kernel parameters (e.g., TCP buffer size, congestion control algorithm, forwarding capability, etc.) dynamically without rebuilding the Linux kernel.

The Network Tools and Protocols Lab Series provides learners an emulated WAN infrastructure operating at high speeds, up to 50 Gbps, and devices running real protocol stacks.

perfSONAR Monitoring Lab Series. This lab series was developed using the pod shown in Fig. 7. perfSONAR [19] is a tool which offers web services-based infrastructure from collecting and diagnosing network performance. perfSONAR makes it possible to diagnose problems across multiple domains quickly and easily, providing a collection of tools for performing and sharing end-to-end network measurements. The pod of Fig. 7 emulates an inter-network with three different domains. Additionally, the pod permits learners to conduct tests against perfSONAR nodes deployed in the Internet.

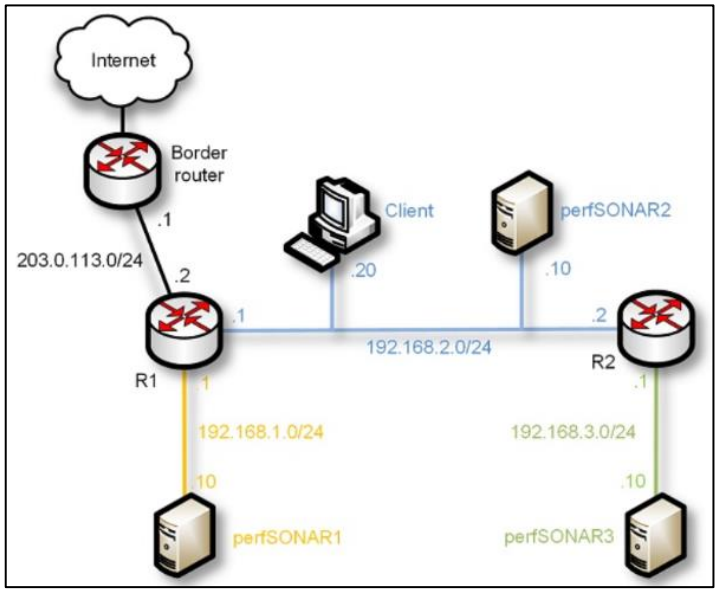

Fig. 7. Pod used to implement virtual labs on perfSONAR.

perfSONAR is a widely deployed test and measurement infrastructure that is used by science networks and facilities around the world to monitor and ensure network performance. It is a free application based on Linux's CentOS distribution.

The perfSONAR Lab Series enables users to learn perfSONAR on a multi-domain internetwork (emulating the Internet), generate live traffic, and operate multiple measurement nodes used in real deployments, such as Central Management and Toolkit.

Zeek / Bro Lab Series. This lab series was developed using the pod shown in Fig. 8. Zeek [20] is an open-source network traffic analyzer. It is primarily a security monitor that inspects all traffic on a link in depth for signs of suspicious activity. It can run on commodity hardware with standard UNIX-based systems and can be used as a passive network monitoring tool. When operating in passive mode, Zeek is appealing for high-speed networks and Science DMZs because it enables engineers to inspect packets from known protocols and to extend to other custom or new protocols, without interfering with traffic flows. The training material provides hands-on overview of Zeek logs, files, and other setups, followed by real generation and detection of attacks (scanner, Denial of Service (DoS), Distributed Denial of Service (DDoS)). It also provides activities on advanced detection techniques such as machine learning classifiers.

The Zeek Lab Series enables users to conduct cyberattacks on a controlled environment, and secure networks using offline intrusion detection suitable for high speeds.

4. Preliminary Impact

4.1 Use of Material in Academic Courses

During the first year of the project, much efforts were concentrated on developing the pods, virtual labs, and companion material. The material is currently used in programs of studies at multiple universities and colleges, including University of South Carolina, University of Texas at San Antonio, University of South Florida, Northern New Mexico College, and Fort Hayes University. There were approximately ~10 courses using the platform. We expect to expand this number in the following years of the project.

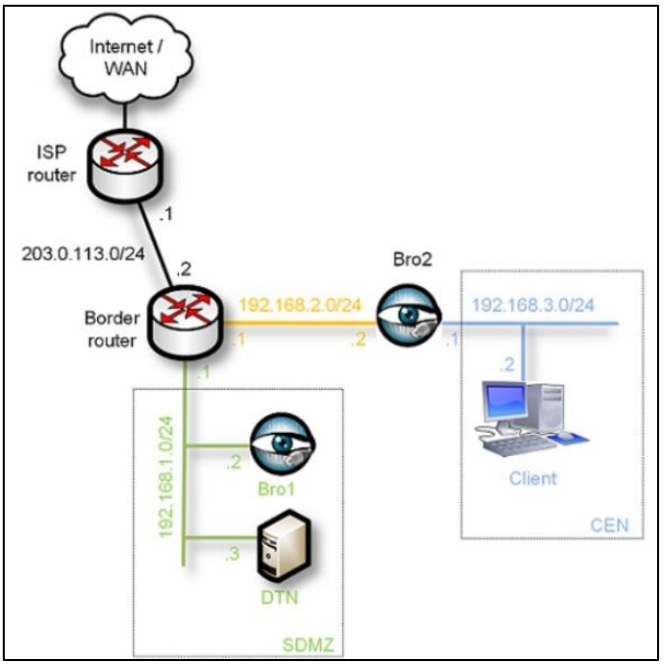

Fig. 8. Pod used to implement virtual labs on Zeek.

4.2 Use of Material for Short Courses / Training Workshops

The project also organized three two-day workshops that focused on training cyberinfrastructure professionals operating high-speed networks and Science DMZs. Two workshops were organized at the University of South Carolina and one workshop was organized in Arizona State University.

The number of attendees to the three workshops (two workshops in South Carolina and one in Arizona) was 208 instructors and professionals from 30 states and other countries (remote participants: Australia, Canada, England), see Table 5.

Table 5. Number of attendees (per state) to the workshops organized in Year 1 of the project.

State	Participants per state
AL	4
AR	1
Australia	2
AZ	33
CA	6
Canada	1
CO	7
DC	1
England	1
FL	15
GA	14
HI	1
IA	2
IL	1
IN	1
KY	1
MA	2
ME	1
MI	4
MO	2
MT	1
NC	19
NM	15
NY	1
OH	4
SC	47
TN	5
TX	3
US Navy	1
UT	1
VA	2
WA	4
WI	1
WV	1
Others	3
TOTAL	**208**

Workshops were evaluated using a survey taken by attendees at the end of the workshops. The following is the results of one of the workshops, "Training Workshop for Network Engineers and Educators on Tools and Protocols for High-Speed Networks and Cybersecurity." Out of 77 attendees, 27 responded the survey. The potential scores for each question were:

- 5: Extremely satisfied
- 4: Very satisfied
- 3: Moderately satisfied
- 2: Slightly satisfied
- 1: Poor / not at all satisfied

Questions related to logistics. Fig. 9 shows the survey results (averages) for questions related to logistics of the event. Attendees were very satisfied with the overall meeting, logistics, and quality of material. Aspects to improve next year are the time allocated for networking / discussions and food. Note, however, that even these aspects had a score above 4.

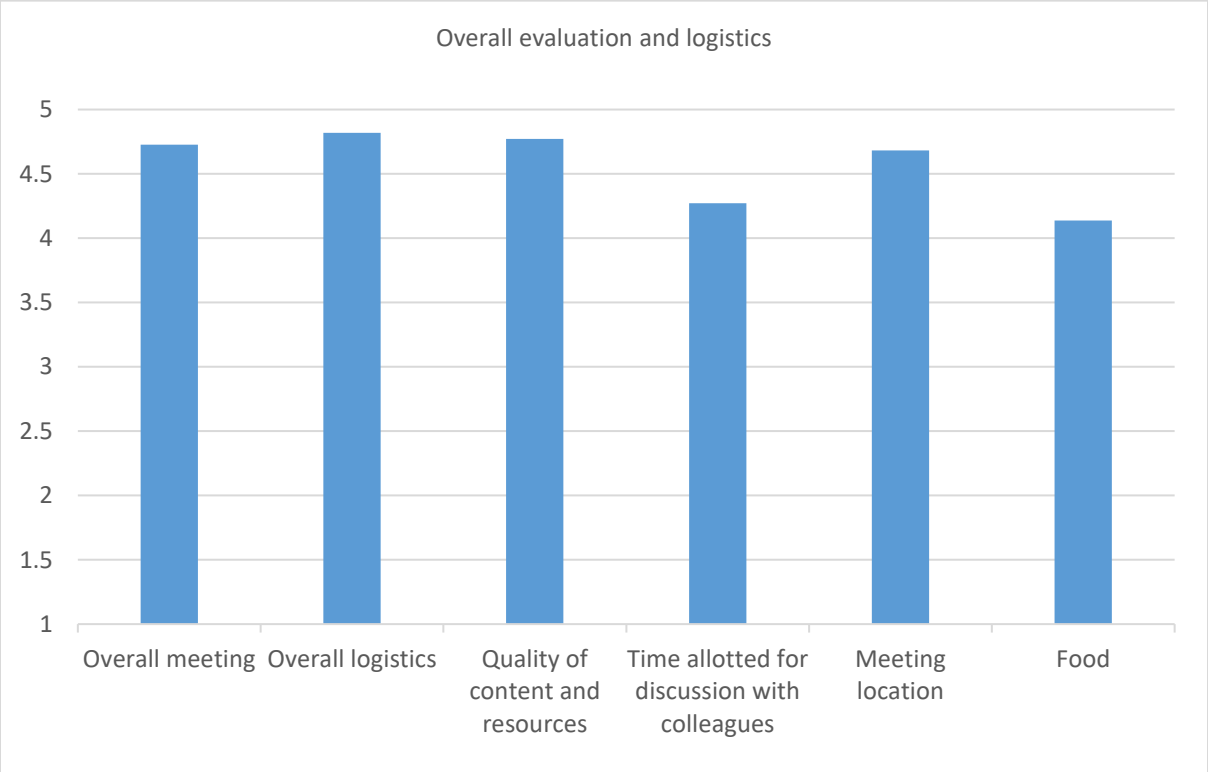

Fig. 9. Survey results for questions related to logistic and overall evaluation, for the workshop "Training Workshop for Network Engineers and Educators on Tools and Protocols for High-Speed Networks and Cybersecurity," July 22 – 23, 2019, University of South Carolina, Columbia, SC.

Questions related to virtual labs. Fig. 10 shows the average results of the survey questions related to training material and private cloud (all averages were above 4.5 / 5). Attendees were asked to rate the private cloud and the quality of the three series of laboratories: Network tools and protocols

(14 vLabs experiments), perfSONAR monitoring (10 vLabs experiments), and Zeek / Bro intrusion detection (10 vLabs experiments). Results were consisted; attendees considered that the quality of developed vLabs were about 4.5 / 5. Anecdotal evidence and comments during the workshop also confirmed the high-quality material, according to attendees.

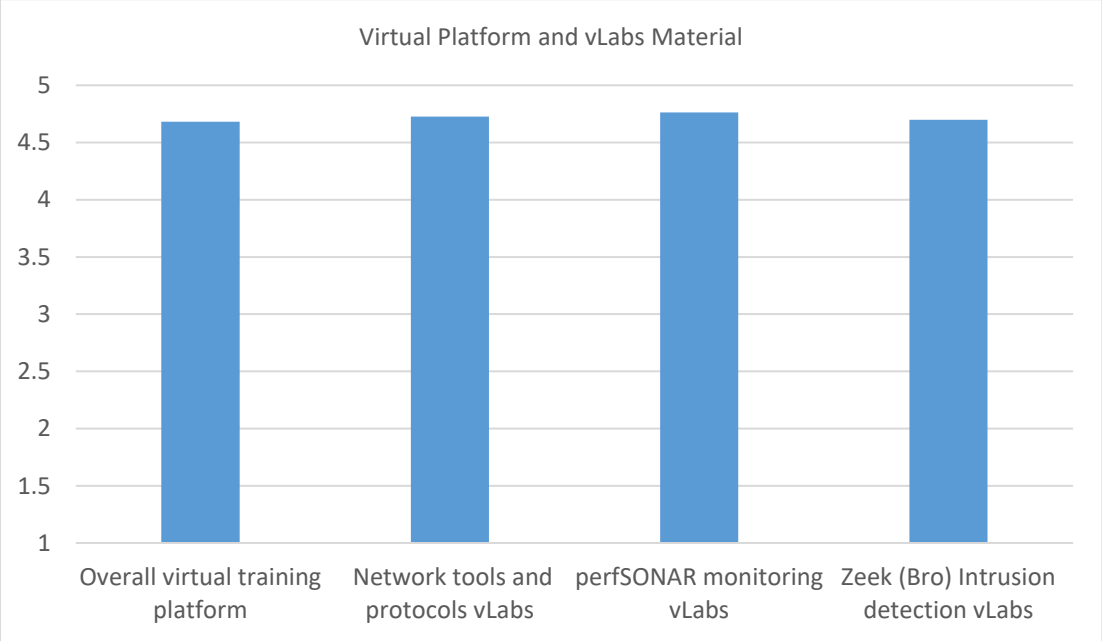

Fig. 10. Survey results for questions related to private cloud and vLabs quality, for the workshops "Training Workshop for Network Engineers and Educators on Tools and Protocols for High-Speed Networks and Cybersecurity," July 22 – 23, 2019, University of South Carolina, Columbia, SC.

4.3 Use of Private cloud for Research

The virtual platform has been also used for undergraduate and graduate research purposes. Advantages of the virtual platform on research includes:

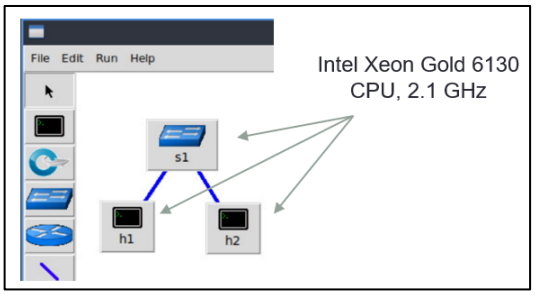

Fig. 11. Pod designed with one CPU allocated per device.

- Complex networks of hundreds of nodes can be deployed immediately
- A virtual testbed capable of recreating realistic scenarios
- Parameters such as packet loss rate, bandwidth, latency and others are easily configured
- Real protocol stacks are used; e.g., Linux, vendor specific (virtual routers, security appliances)
- Reasonable accurate at rates of tens of gigabits per second
- Resources are allocated as needed. For example, for the topology of Fig. 11, the researcher is allocating one CPU per device. The researcher may design the pod and allocates a customized amount of RAM memory, CPUs, and storage as needed.

After a year of operating the private cloud, the teams at the University of South Carolina and University of Texas at San Antonio have been able to submit and publish research work conducted in the private cloud to several well-known conferences and journals, including [21]-[38]. For research purposes, the private cloud has been used in two different capacities:

- Emulation model prior to real deployments: students use a model of a real hardware, running on the private cloud, to prototype applications prior to developing the applications into the hardware. The prototype enables them to quickly develop an application and test it into the virtual environment, reducing the testing and development cycles.

- Performance tests using virtual networks: students use the tools and emulators such as NETEM to test the performance of different algorithms, including TCP, congestion control, and others. The platform permits students to test algorithms on a variety of scenarios manipulating different parameters, such as packet loss rate, latency, buffer size, congestion control algorithms, maximum transmission units, etc. As replicating pods is trivial in the private cloud, running hundreds of tests in parallel is easy, which speeds up the evaluation process.

4.4 Use of the Private cloud and Comparison with Other Platforms

From October 1, 2018 to September 28, 2019, the use of the platform was as follows:

- Reservations made: 3,836 (a reservation indicates that a student was reserved a time block to conduct experiments)
- Reservations attended: 3,589 (actual use of the platform)
- Hours attended: 12,146.53

Fig. 12. Use of the private cloud at the University of South Carolina between October 1, 2018 to September 28, 2019. The platform was used to support academic classes, workshops, and research at the University of South Carolina.

The impact of this project is beyond cyberinfrastructure. While resources were placed for cyberinfrastructure courses, training workshops, and research, the Department of Integrated Information Technology (IIT) at the University of South Carolina is planning to extend the use of the platform to other areas, such as databases, programming, web systems, and others.

Comparison between the private cloud and public clouds. Table 6 compares features implemented in the private platform and public cloud. Deploying the private cloud required an initial investment to buy the physical servers the platform runs on (see Table 3). While the capital to deploy a virtual platform may initially seem significant when compared with the cost of renting services from a public cloud (e.g., Amazon Web Services (AWS)), the virtual platform provides much more flexibility in the design of pods, and the initial cost is amortized when the platform is used extensively. With the virtual platform, physical servers are fully dedicated and resources available as needed. Allocation of physical resources in a granular way is important when designing high-speed / high-performance pods, as they may have strict physical CPU and memory requirements. For example, emulating a data transfer across Science DMZs at 40 Gbps requires a careful allocation of physical CPUs, which may not be available in public clouds. Additionally, creating complex custom pods is easy in the private cloud, including an application layer and presentation of the scenario for pedagogy. For example, in Figs. 7 and 8, the learner has a visual presentation of the multi-domain topology and can access any device by simply clicking over the device.

Table 6. Comparison between private and public clouds.

Feature	Private Cloud	Public Cloud
Granularity to allocate physical resources	Very granular	Not granular (access to the physical resources require additional fees)
Easy to create custom pods	Easy	More difficult; hard to design complex topologies
Cost	Cost effective when used extensively	Cost effective for individual / small virtual machines; costly for large virtual machines over time
IT Staff	Higher cost	Lower cost
Application layer for pedagogy and presentation of virtual scenarios	Very flexible	Not flexible; limited to providers' interface, e.g., command-line interface
Time-sharing compute resources	The owner controls who can access resources. Easy to implement time-sharing policies	Cloud provider controls who can access resources (typically, a fee is required per user accessing resources)

A reader may rightfully note that, in contrast to a public cloud, the private cloud requires additional efforts for managing, maintaining, and updating the physical equipment (IT staff). This is one of the main advantages of the public cloud. Note that we do not claim that the private cloud model is the appropriate choice for all institutions, as IT staff / human expertise may not be available at the right cost by many institutions. However, if IT staff / human expertise is available and these resources can be aggregated over multiple institutions, then the private virtual cloud is attractive. Other features of the private cloud are listed in Table 7.

Table 7. Private cloud features.

Feature	Comment
Ready on demand	Entire scenario (WAN, DTNs, switches, routers) are instantaneously deployed on demand, as a student/trainee reserves a pod and enters the lab
Replicable	Replicable and readily available to other institutions
Calendar itf.	Trainees will be able to view the pods and timeslots to schedule lab time
Context	Visual context by deploying the complete scenario on the screen
Uniformity and readiness	At the beginning of the lab reservation a clean and tested operating system image is loaded for each device (network operating system, Linux image for DTNs, etc.)
Degrees / certs	Easily adoptable for academic degrees, workshops, and certificates
Controlled env.	The controlled platform pushes the boundaries of production and experimental high-speed networking and security
Flexibility and High Speed	Platform provides flexibility to deploy virtual networks that operates at rates of up to 45 Gbps. Virtual networks emulate complex interconnections of LANs and WANs

5. Conclusion

This paper describes a project that implemented a private cloud and developed virtual laboratories intended for teaching, training, and research on high-speed networks and cybersecurity. The platform supports customized pod and lab designs that emulate complex internetworks operating at up to 50 Gbps, multi-domain environments, and infrastructures with virtual appliances that include end devices, routers, switches, firewalls, and intrusion detection systems. The scalability of the platform permits the simultaneous on-demand deployment of hundreds of emulated WANs and LANs, thus serving hundreds of users at the same time.

The material and platform, deployed at the University of South Carolina, have been used to support regular academic courses, self-pace training of professional IT staff, and workshops across the country. The feedback from users has consistently shown that the platform and the virtualized material enhance essential hands-on IT skills on students, practitioners, and researchers. Additionally, the private cloud is a scalable and cost-effective alternative to physical laboratories and public clouds. The authors are currently expanding the topics and subjects covered by the virtual labs.

Acknowledgement and Disclaimer

Support for this project has been received from the National Science Foundation (NSF) Grants 1829698 and 1822567. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NSF.

References

[1] L. Farrell, "Science DMZ: The fast path for science data," Sci. Node, May 2016. [Online]. Available: https://sciencenode.org/feature/sciencedmz-a-data-highway-system.php

[2] E. Dart, L. Rotman, B. Tierney, M. Hester, J. Zurawski, "The science dmz: a network design pattern for data-intensive science," in Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis, Nov. 2013.

[3] "NSF 2017 PI Workshop CI Engineer Breakout Survey." [Online]. Available: http://www.thequilt.net/wp-content/uploads/NSF-2017-PI-Workshop-CI-Engineer-Survey_v4.pdf

[4] J. Crichigno, E. Bou-Harb, N. Ghani, "A comprehensive tutorial on science DMZ," IEEE Communications Surveys and Tutorials, vol. 21, issue 2, 2019.

[5] ESnet Website. Available: https://www.es.net

[6] D. Comer, "Computer networks and internets" 6th Edition, Prentice Hall, 2015.

[7] J. Kurose, K. Ross, "Computer networking: a top-down approach," 7th Edition, Pearson, 2017.

[8] D. Teare, B. Vachon, R. Graziani, "Implementing cisco IP routing," Cisco Press, Jan. 2015.

[9] ESnet Fasterdata Knowledge Base. [Online]. Available: http://fasterdata.es.net/

[10] Operating innovative networks (OIN) website. [Online]. Available: http://www.oinworkshop.com/

[11] "Campus cyberinfrastructure: operating innovative networks (CC*OIN)." [Online]. Available: https://tinyurl.com/uf3fysz

[12] G. Appenzelle, I. Keslassy, N. McKeown, "Sizing router buffers," in Proceedings of the 2004 conference on Applications, technologies, architectures, and protocols for computer communications, pp. 281-292, Oct. 2004.

[13] Network Development Group website. [Online]. Available: https://www.netdevgroup.com/

[14] Mininet website. [online]. Available: http://mininet.org/

[15] iPerf3 website. [online]. Available: http://software.es.net/iperf/

[16] S. Hemminger, "Network emulation with NETEM," in Proc. Australia's Nat. Linux Conf., Apr. 2005, pp. 18–93.

[17] Linux's Token Bucket Filter Man Page. [online]. Available: https://linux.die.net/man/8/tc-tbf

[18] Linux's Sysctl Man Page. [online]. Available: https://linux.die.net/man/8/sysctl

[19] J. Zurawski, S. Balasubramanian, A. Brown, E. Kissel, A. Lake, M. Swany, B. Tierney, M. Zekauskas, "perfsonar: on-board diagnostics for big data," in Workshop on Big Data and Science: Infrastructure and Services, Oct. 2013.

[20] Zeek Network Security Monitor website. [online]. Available: https://www.zeek.org/

[21] E. Kfoury, J. Gomez, J. Crichigno, E. Bou-Harb, "An emulation-based evaluation of TCP BBRv2 alpha for wired broadband," submitted to Computer Communications Journal, Jan. 2020.

[22] E. Kfoury, J. Crichigno, E. Bou-Harb, "Offloading media traffic to programmable data plane switches," IEEE International Conference on Communications (ICC) 2020, Dublin, Ireland, June 2020.

[23] K. Friday, E. Kfoury, E. Bou-Harb, J. Crichigno, "Towards a unified in-network DDoS detection and mitigation strategy," submitted to IEEE Conference on Network Softwarization (NetSoft), Ghent, Belgium, June 2020.

[24] M. Safei, A. Mangino, K. Friday, E. Bou-Harb, F. Iqbalc, S. Samtani, J. Crichigno, N. Ghani, On data-driven curation, learning, and analysis for inferring evolving Internet-of-Things (IoT) botnets in the wild," Computers and Security Journal, 2020.

[25] A. AlSabeh, H. Safa, E. Bou-Harb, J. Crichigno, "Exploiting ransomware paranoia for execution prevention," IEEE International Conference on Communications (ICC), Dublin, Ireland, June 2020.

[26] E. Kfoury, J. Gomez, J. Crichigno, E. Bou-Harb, D. Khoury, "Decentralized distribution of PCP mappings over blockchain for end-to-end secure direct communications", IEEE Access, vol. 7, August 2019.

[27] E. Kfoury, J. Crichigno, E. Bou-Harb, D. Khoury, G. Srivastava, "Enabling TCP pacing using programmable data plane switches", 42nd International Conference on Telecommunications and Signal Processing (TSP), July 2019.

[28] D. Oliveira, J. Crichigno, E. Bou-Harb, M. Rahouti, N. Ghani, "An efficient multi-objective survivability scheme for mapping and routing of virtual functions in failure scenarios", IEEE International Conference on Software Defined Systems, June 2019.

[29] J. Crichigno, E. Kfoury, E. Bou-Harb, N. Ghani, Y. Prieto, C. Vega, J. Pezoa, C. Huang, D. Torres, "A flow-based entropy characterization of a NATed network and its application on intrusion detection", IEEE International Conference on Communications (ICC), May 2019.

[30] Y. Prieto, C. Vega, J. Pezoa, J. Crichigno, "Shared-risk-aware design for survivable migration in SDN environments", IEEE Consumer Communications and Networking Conference (CCNC), Las Vegas, NV, January 2019.

[31] D. Oliveira, N. Ghani, M. Hayat, J. Crichigno, E. Bou-Harb, "SDN testbed for evaluation of large exo-atmospheric EMP attacks", IEEE Communications Magazine, vol. 7, issue 1, January 2019.

[32] E. Kfoury, J. Crichigno, E. Bou-Harb, V. Gurevich, "Offloading media traffic to programmable data plane switches," P4 Expert Roundtable Series abstract, April 2020.

[33] G. Srivastava, A. Fisher, R. Bryce, J. Crichigno, "Green communication protocol with geolocation," IEEE Vehicular Technology Conference (VTC2019-Spring), April 2019.

[34] D. Oliveira, J. Crichigno, N. Siasi, N. Ghani, E. Bou-Harb, “Joint Mapping and Routing of Virtual Network Functions for Improved Disaster Recovery Support,” IEEE SoutheastCon Conference, St. Petersburg, FL, USA, April 2018.

[35] G. Srivastava, J. Crichigno, S. Dhar, “A light and secure healthcare blockchain for iot medical devices,” IEEE Canadian conference of electrical and computer engineering (CCECE), Edmonton, Canada, May 2019.

[36] G. Srivastava, S. Dhar, A. Dwivedi, J. Crichigno, “Blockchain education,” IEEE Canadian conference of electrical and computer engineering (CCECE), Edmonton, Canada, May 2019.

[37] D. Oliveira, J. Crichigno, E. Bou-Harb, M. Rahouti, N. Ghani, "An Efficient Multi-Objective Survivability Scheme for Mapping and Routing of Virtual Functions in Failure Scenarios", IEEE International Conference on Software Defined Systems, Rome, Italy, June 2019.

[38] N. Siasi, M. Jasim, J. Crichigno, N Ghani, “Container-based service function chain mapping,” IEEE SoutheastCon 2019, Huntsville, AL, April 2019.