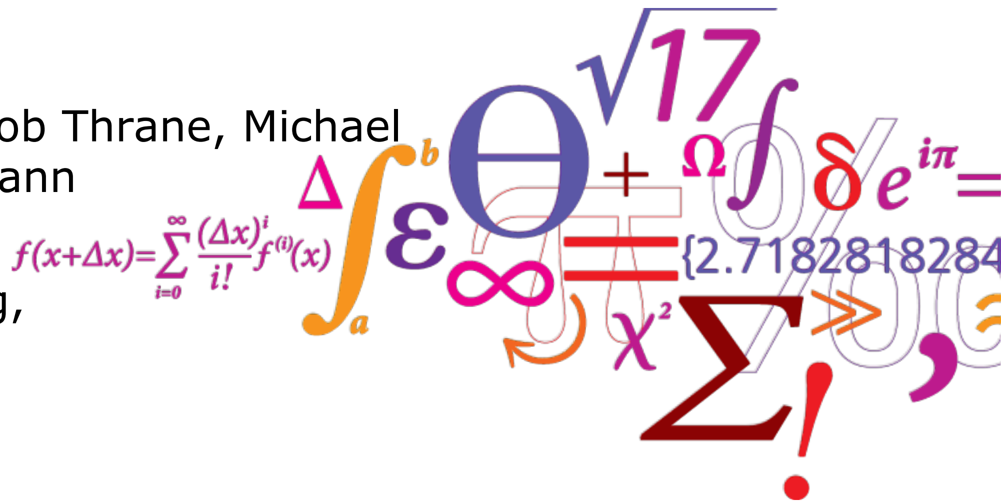
Combining Hardware and Simulation for Datacenter Scaling Studies

ONDM 2017, 21st International Conference on Optical Network Design and Modeling

Sarah Ruepp*, **Artur Pilimon***, Jakob Thrane, Michael Galili, Michael Berger and Lars Dittmann

Department of Photonics Engineering,
Technical University of Denmark,
Kgs. Lyngby, Denmark
E-mail:
artpil@fotonik.dtu.dk



DTU Fotonik
Department of Photonics Engineering

Agenda

Introduction

Goals

Work methodology:

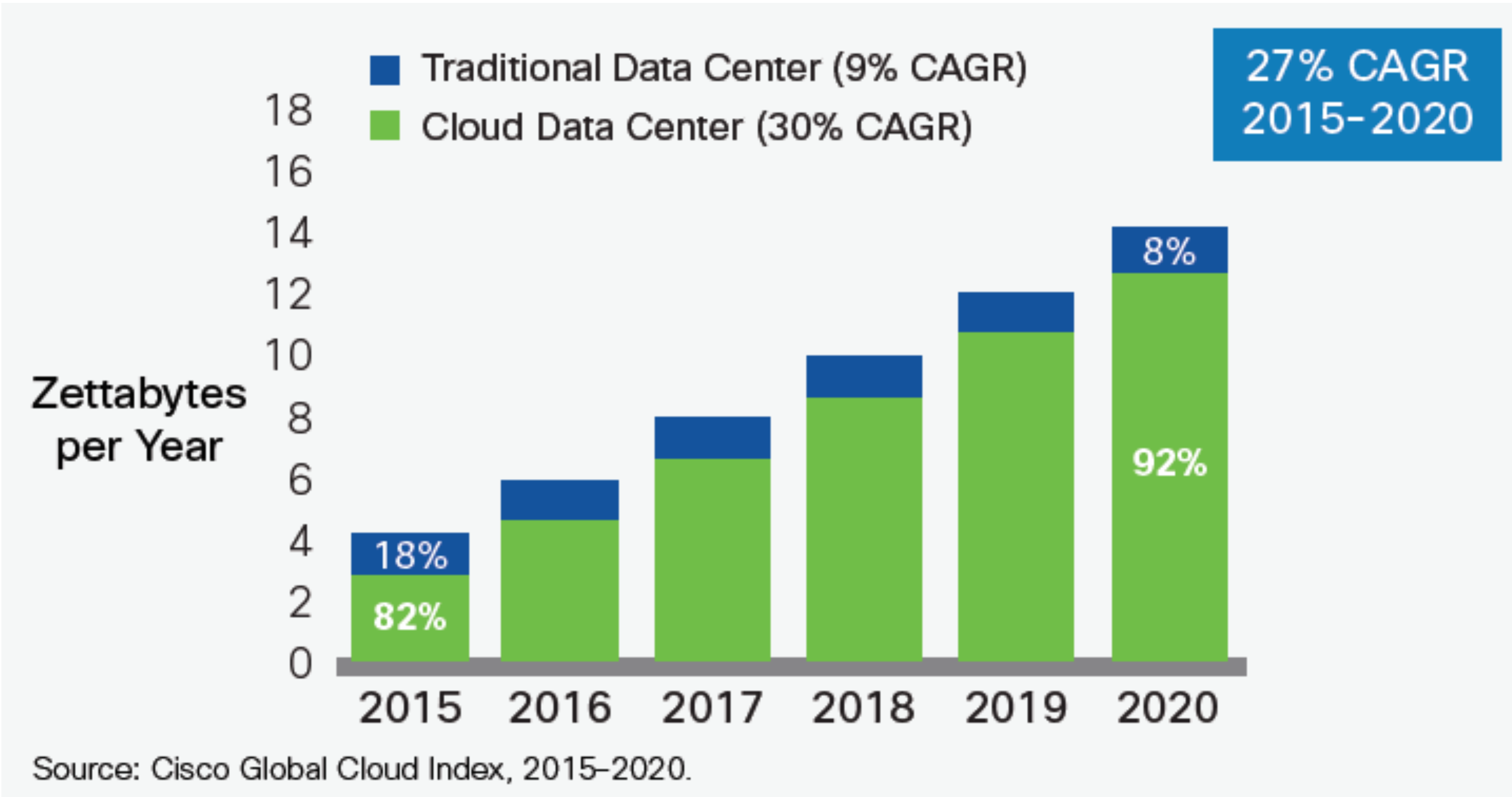
- A Simulation and Hardware setup
- Packet conversion latency measurements

Results

Takeaway points and future work

Introduction. DC traffic growth (1)

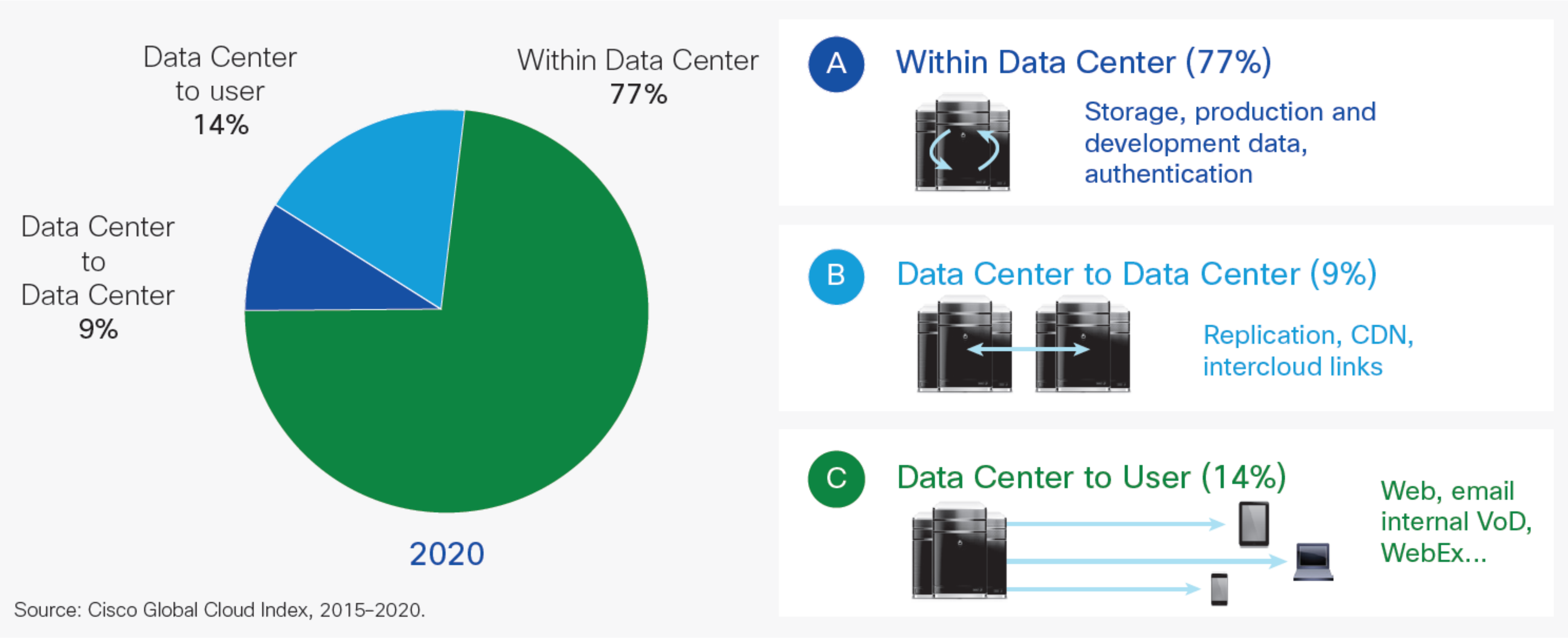
- Tremendous increase of intra-DCN traffic globally
- Fast proliferation of Cloud and network virtualization technologies



Source: Cisco Global Cloud Index

Introduction. DC traffic growth (2)

- Most (~77 %) of the generated traffic remains within a DC !!!



Source: Cisco Global Cloud Index

Introduction. Global trends

- ✓ **Increase in DCN scale and power** → Cloud DCs, MDCs
- ✓ **Diverse traffic patterns** → Mice/Elephant flows → Multi-tier applications
- ✓ **New DCN architectures** → Jupiter, RoR, Helios, c-Through, OPMDC, COSIGN, etc.
- ✓ **Optical switching** → 2-D/3-D MEMS, optical cross-bar, etc.

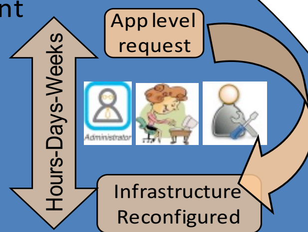
Introduction. Research: COSIGN approach

COSIGN

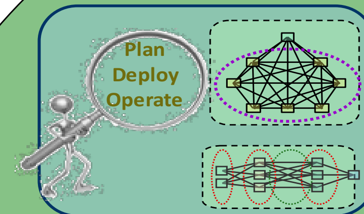
Orchestrated Management

Manual, Painful, Error-prone

- Workload deployment requires time and intervention of different admin roles
- Connectivity of workload components is static
- Performance is achieved with over provisioning, questionable scaling characteristics



Automated, Streamlined, Optimized



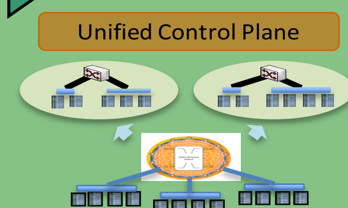
- Rapid and flexible workload management
- On demand resource allocation and release
- Scalable and efficient virtualized workload components connectivity

Control Plane

Discrete, Heterogeneous, Complex

- Multiple management roles and domains
- Unable to make global decisions efficiently and dynamically
- Requires expertise in different vendor technologies

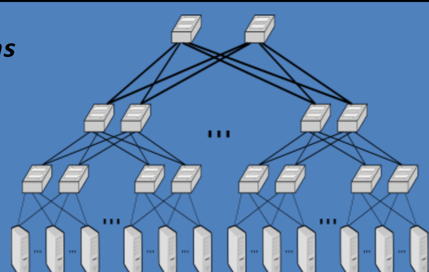
Consolidated, Converged, Programmable



- Unified IT and SDN orchestrator
- Converged IT + Network virtualization for data centers
- Efficient and optimized virtual resource utilization and allocation
- Enhanced abstraction mechanisms for emerging optical technologies

Traditional fat-tree designs

- Over provisioned
- Inefficient for east-west communications
- Restricts (virtual) server placement decisions

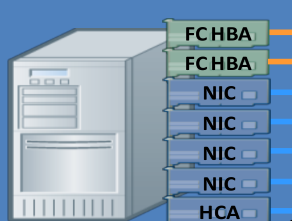


High-performance optical solutions

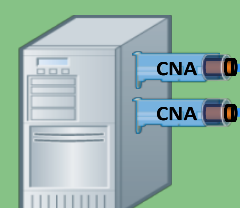
- Flattened DC network architecture
- Fast TOR switch
- 3D stacked transceiver
- InP fast switch
- Low loss beam-streering switch
- Hollow-core and Multi-core fibers

Cabling mess

- Many different incompatible interconnects



Data Plane



Converged DCN interconnect

Challenges and Solutions

- Datacenters for ”trying out things” are not really available for researchers
- New research approaches and technologies may not be implemented in commercial products
- Mass-scaling of datacenters: high CAPEX

SOLUTION: Use simulation, and combine it with the hardware that is already available in the datacenter

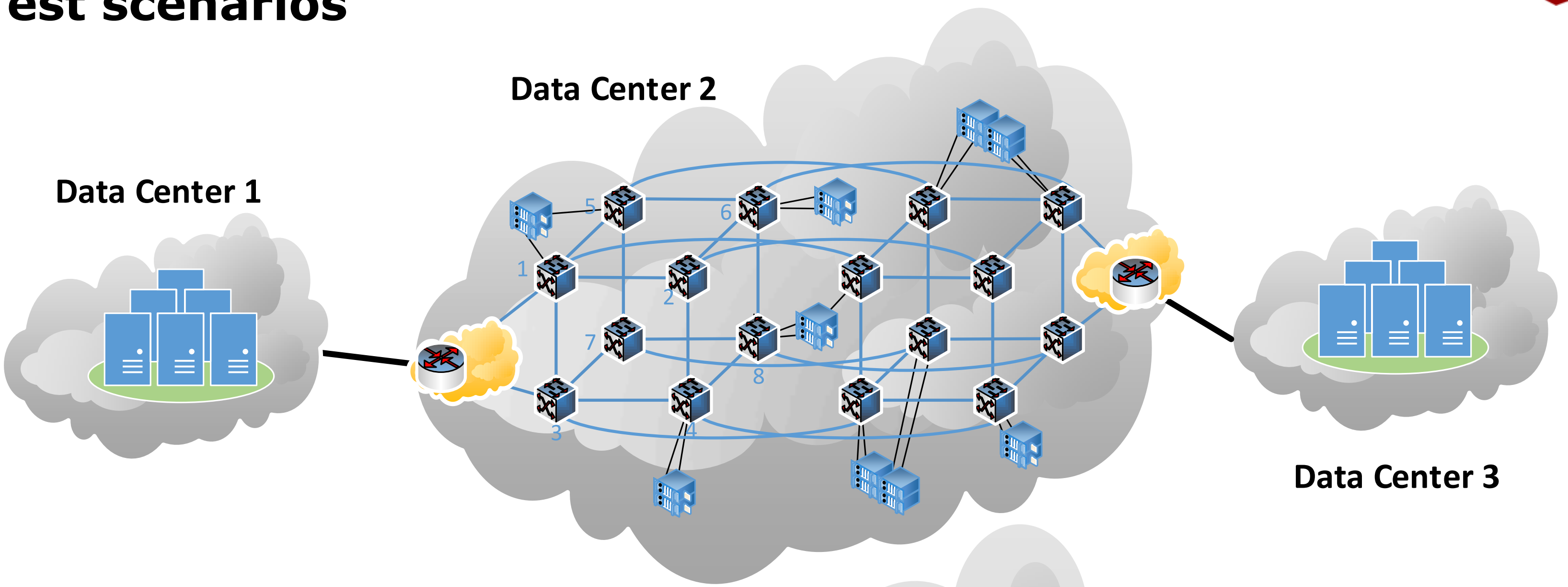
Goals

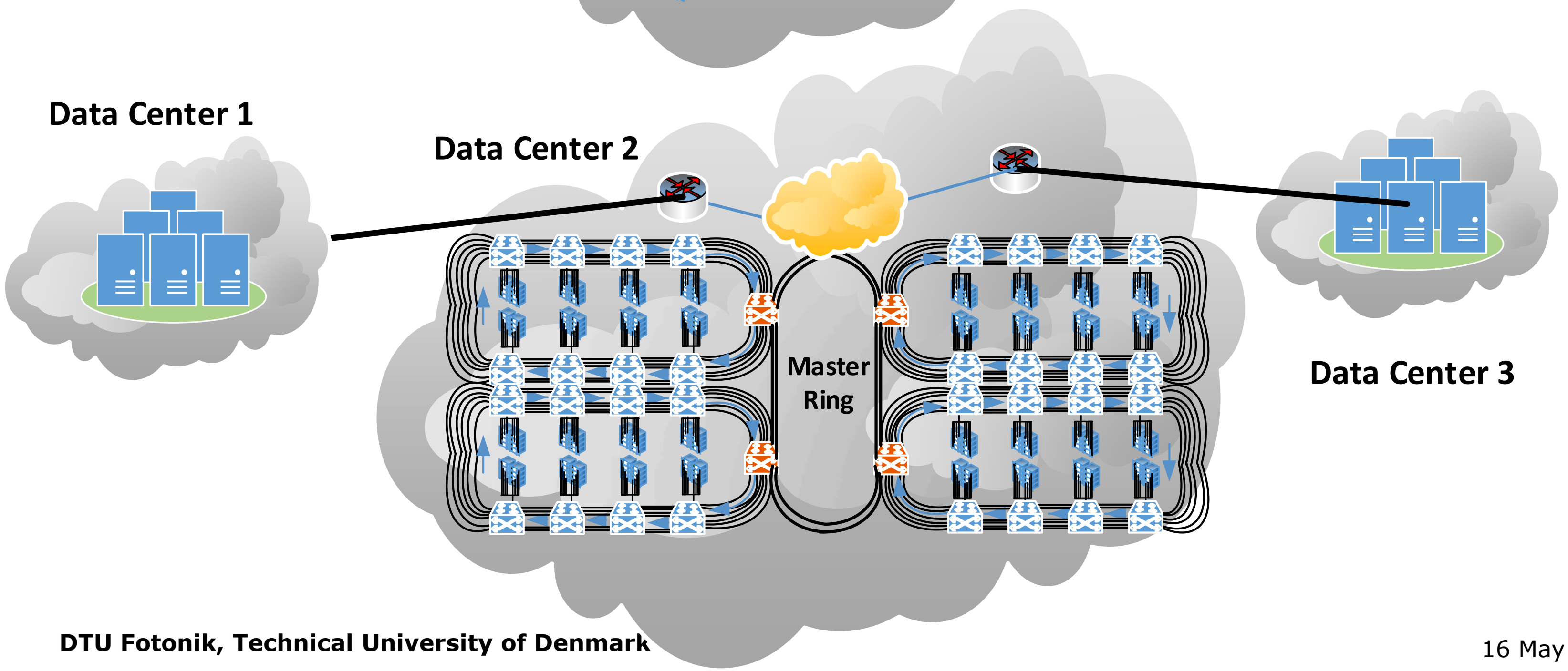
Analysis of the functional capabilities of a simulation tool

Feasibility of building a hybrid (simulation - real hardware) research testbed

Preliminary performance evaluation (Packet conversion latency measurements)

Work methodology
Test scenarios



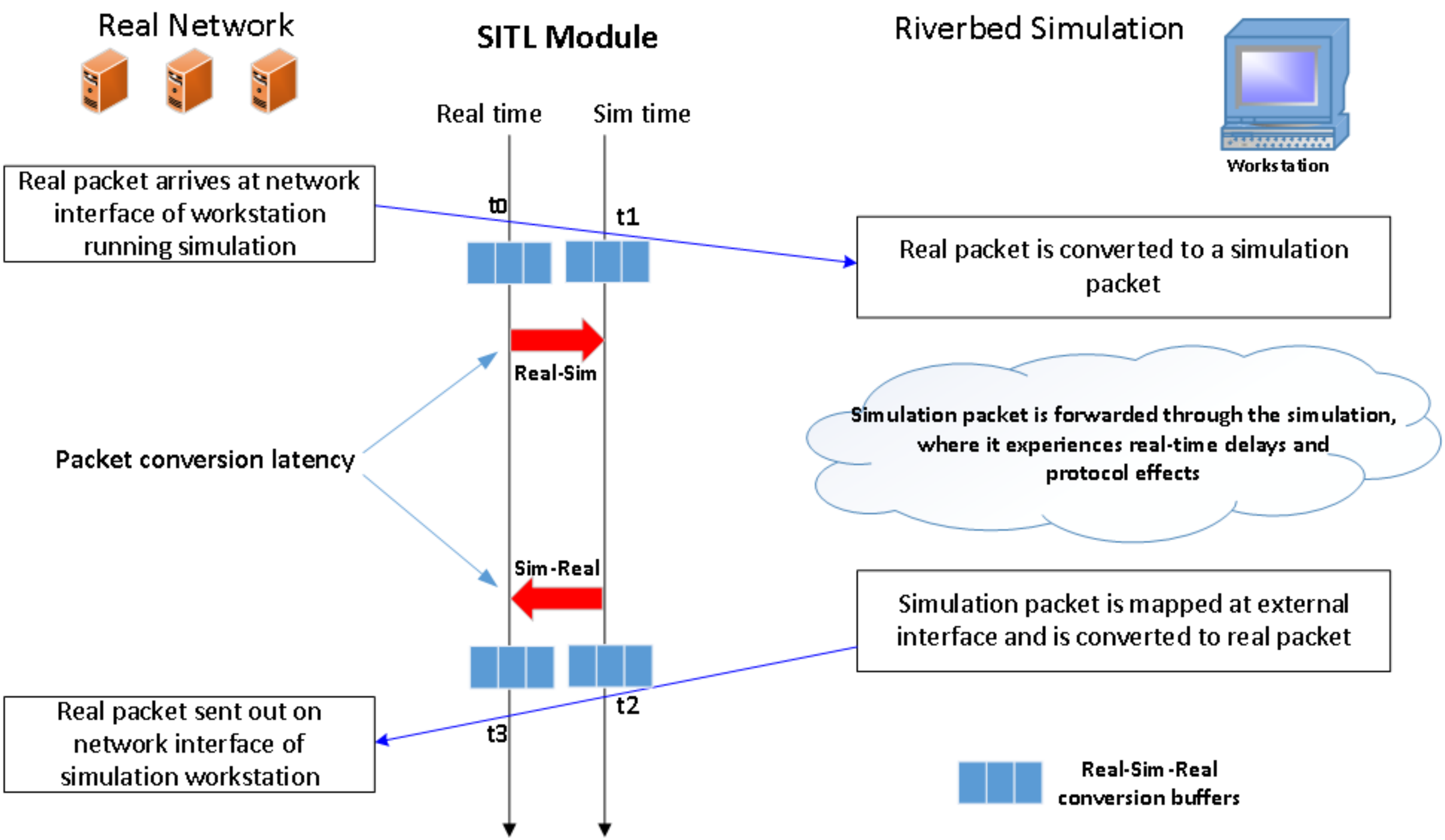


Work methodology

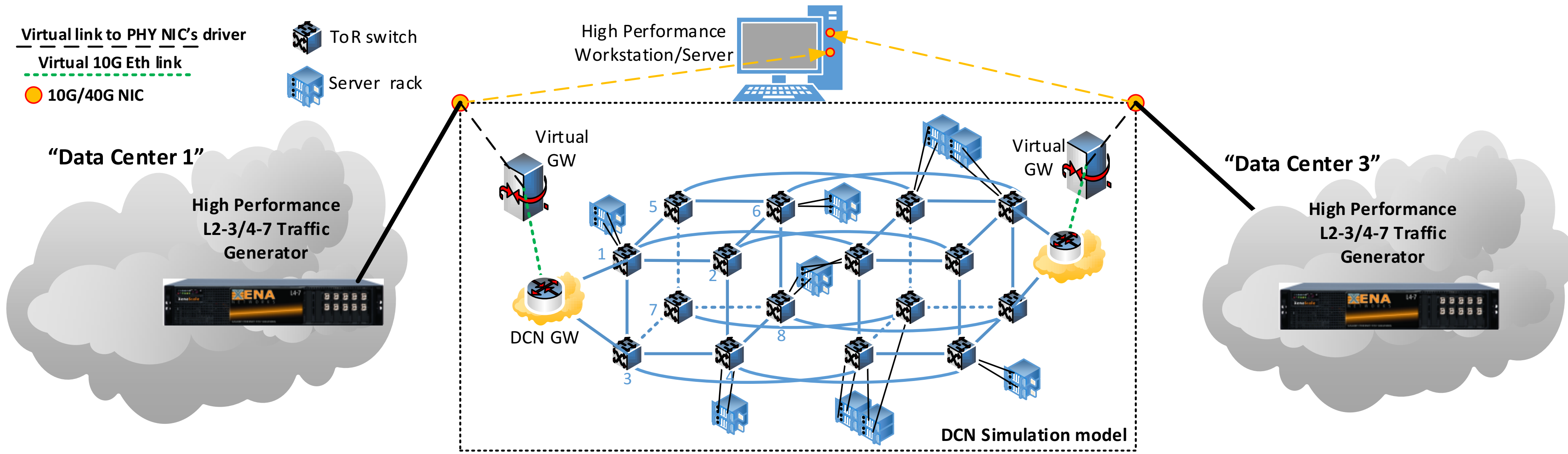
Choosing a M&S environment: Hardware-in-the-Loop

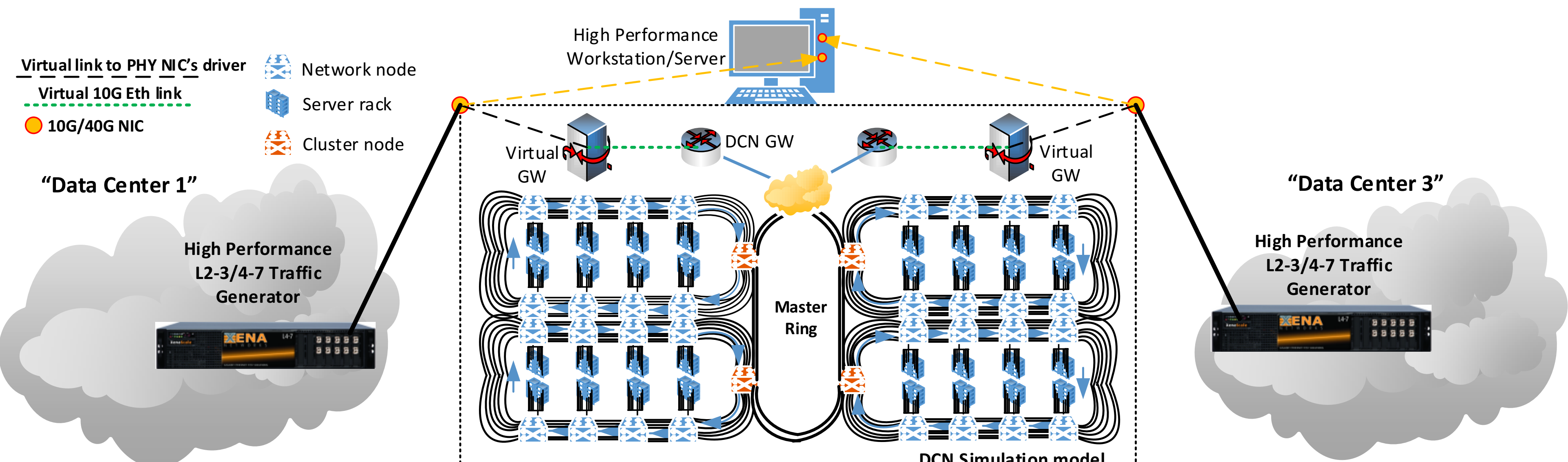
Simulation tool	Considered features				
	Real-time	HIL	SDN	Complexity	License
OMNET++	+/-	+/-	-	+/-	APL/GNU GPL*
Qualnet	+/-	-	-	+/-	Commercial
CORE	+/-	+/-	-	+/-	GNU GPL
Riverbed/OPNET	+	+	+	+/-*	Commercial
NS2	-	-	-	+/-	GNU GPL
NS3	+/-	+/-	-	+/-	GNU GPL
DCSim	-	-	-	+/-	GNU GPL

Work methodology
Riverbed modeler: System-in-the-Loop principle

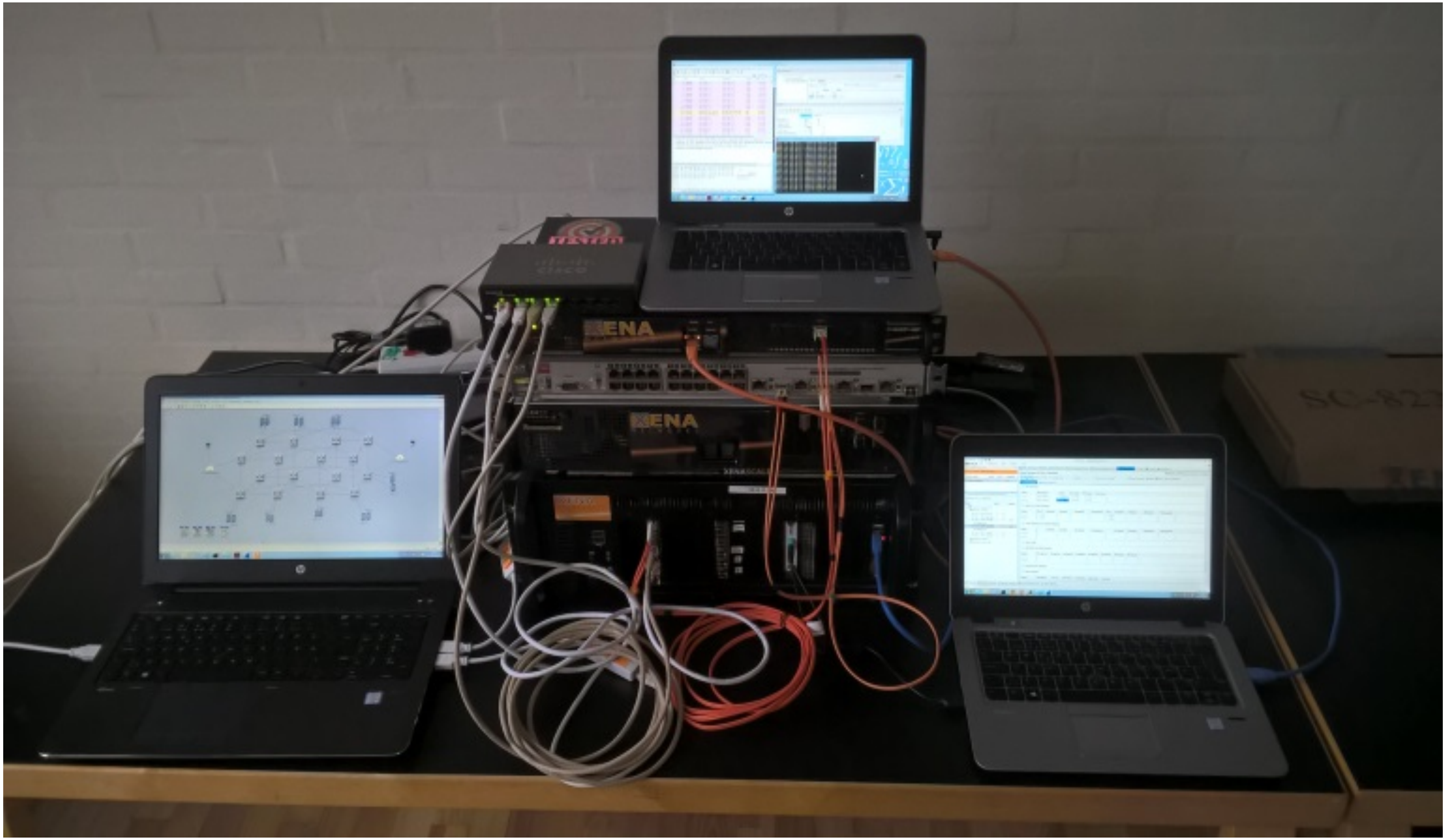


Work methodology
A Hybrid simulation-real-hardware testbed (1)

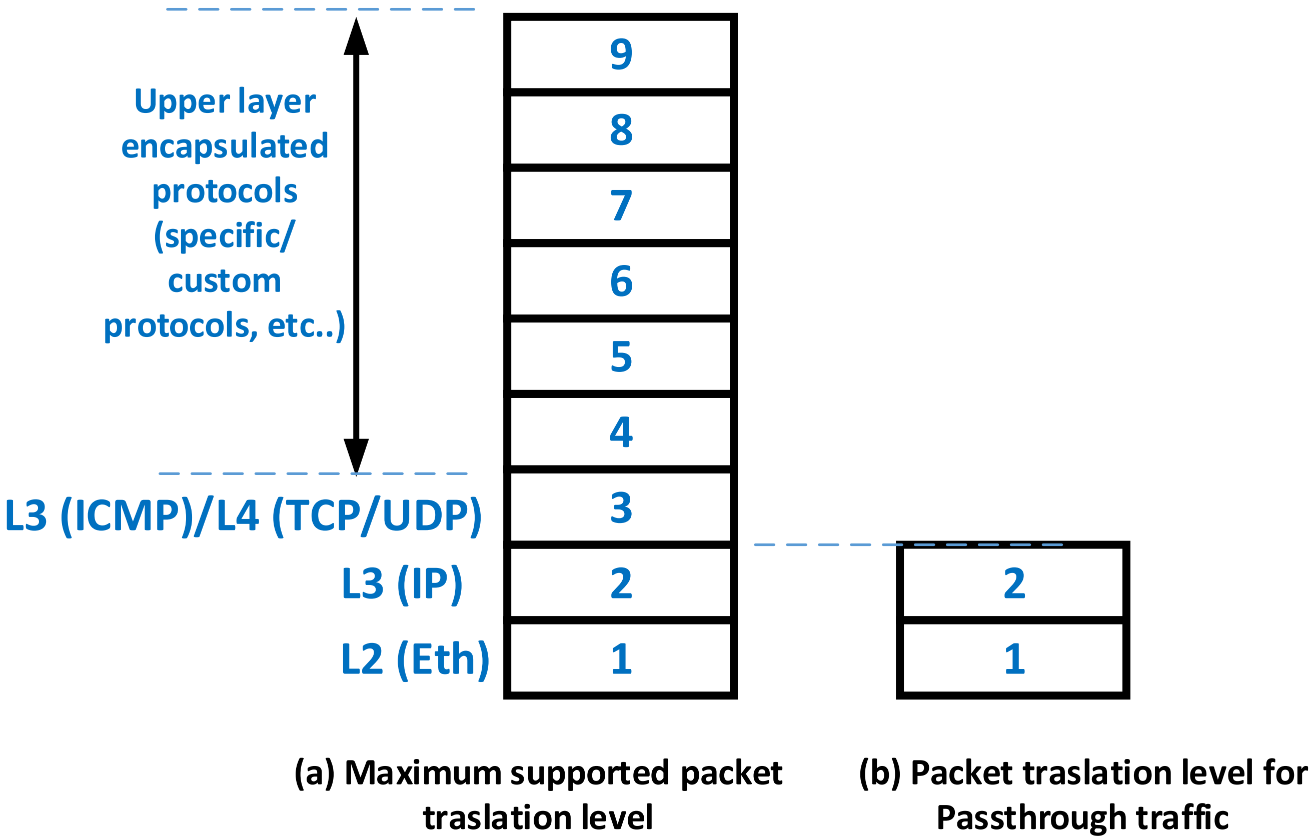




Work methodology
Experimental setup (2)



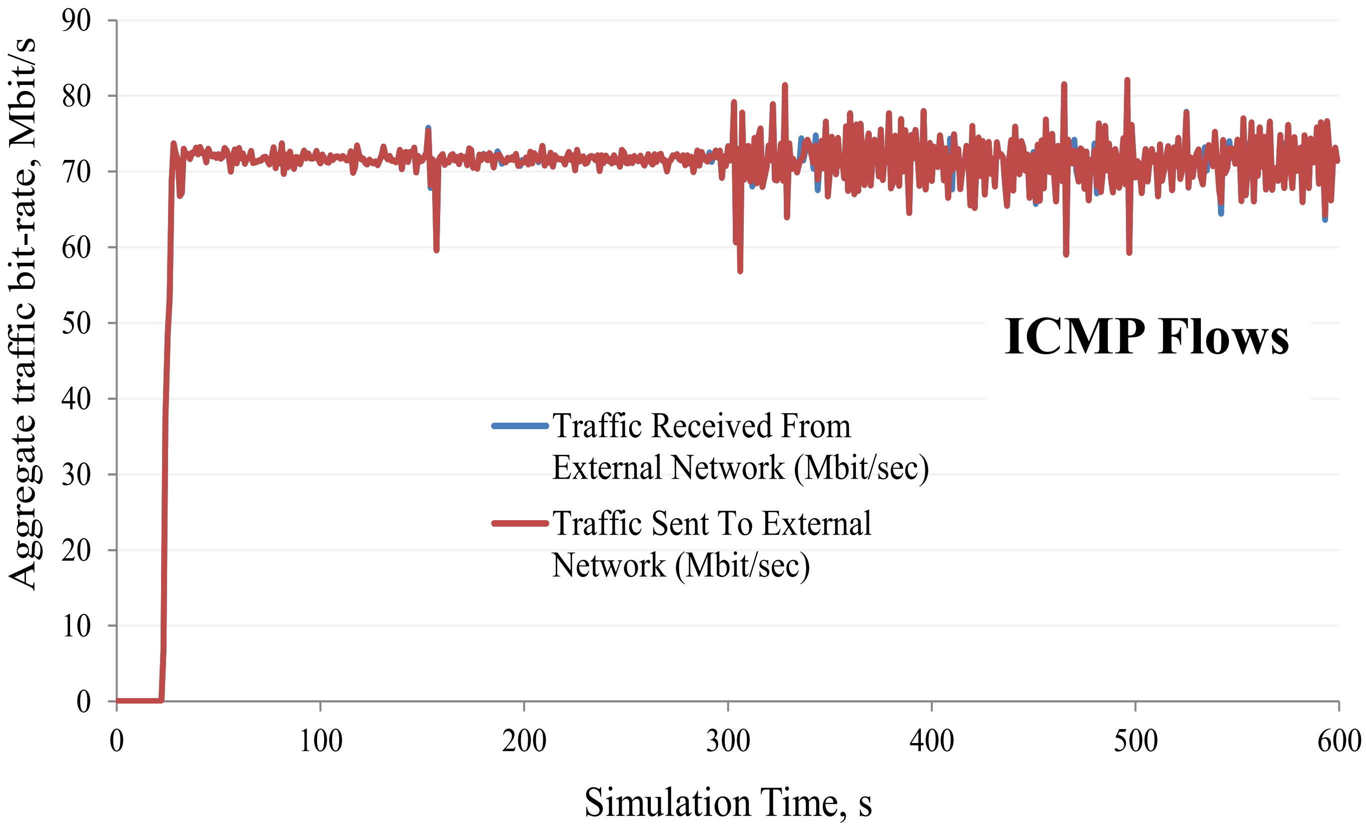
Work methodology
Packet conversion efficiency



Packet translation depth (level) at the real/simulated interface

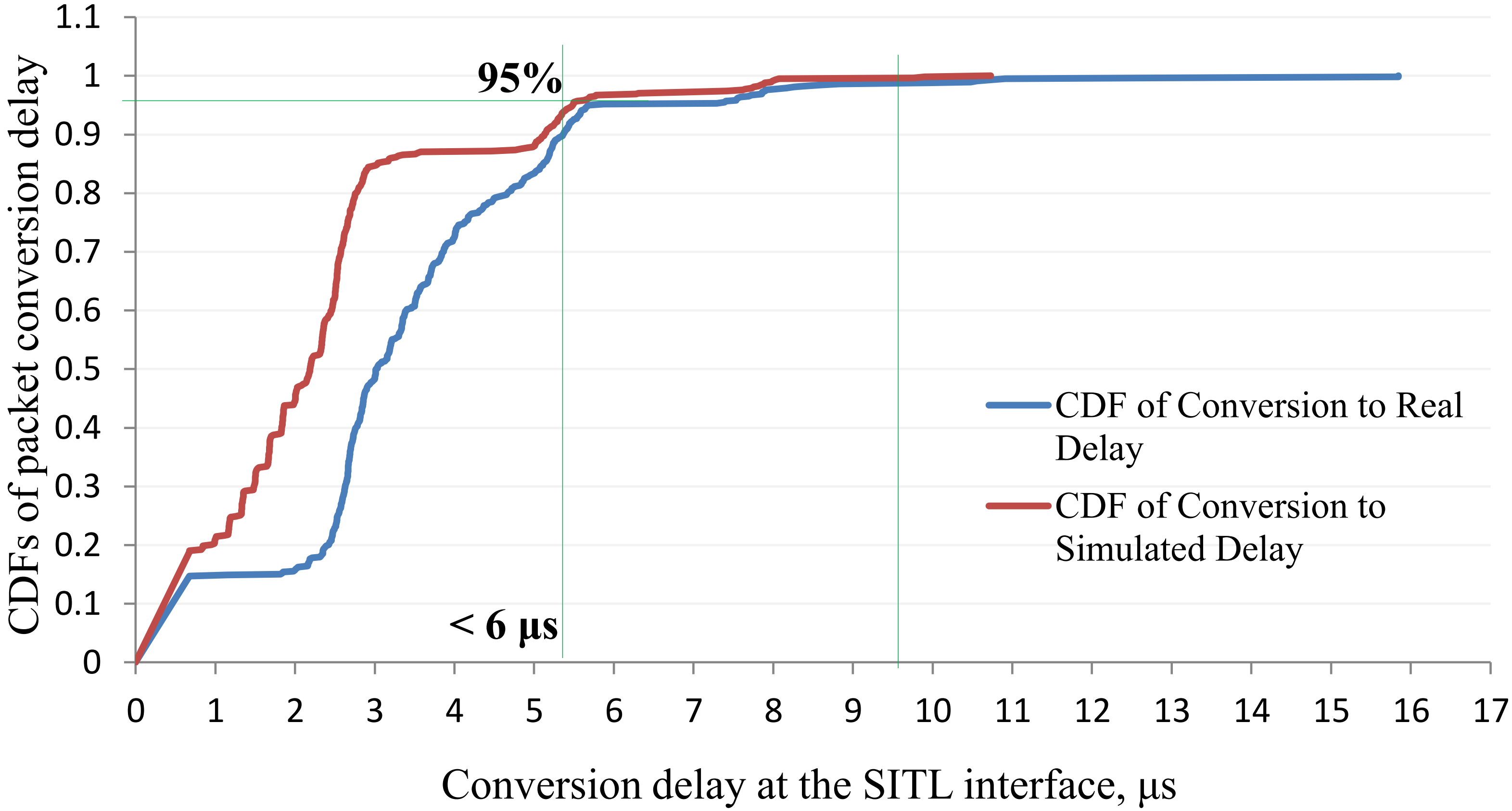
Results

Real traffic generation experiments (1)



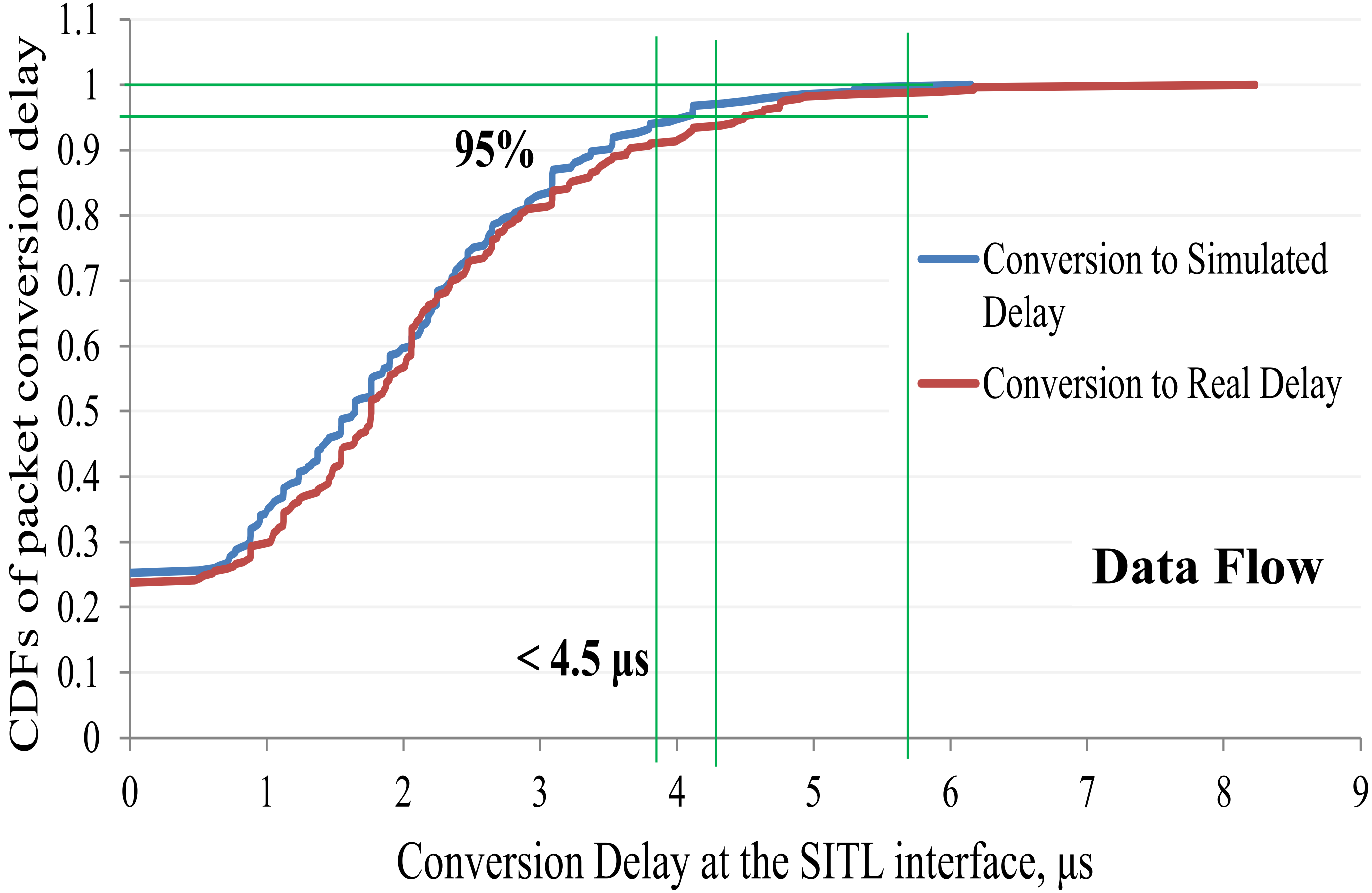
Results

Packet conversion latency measurements



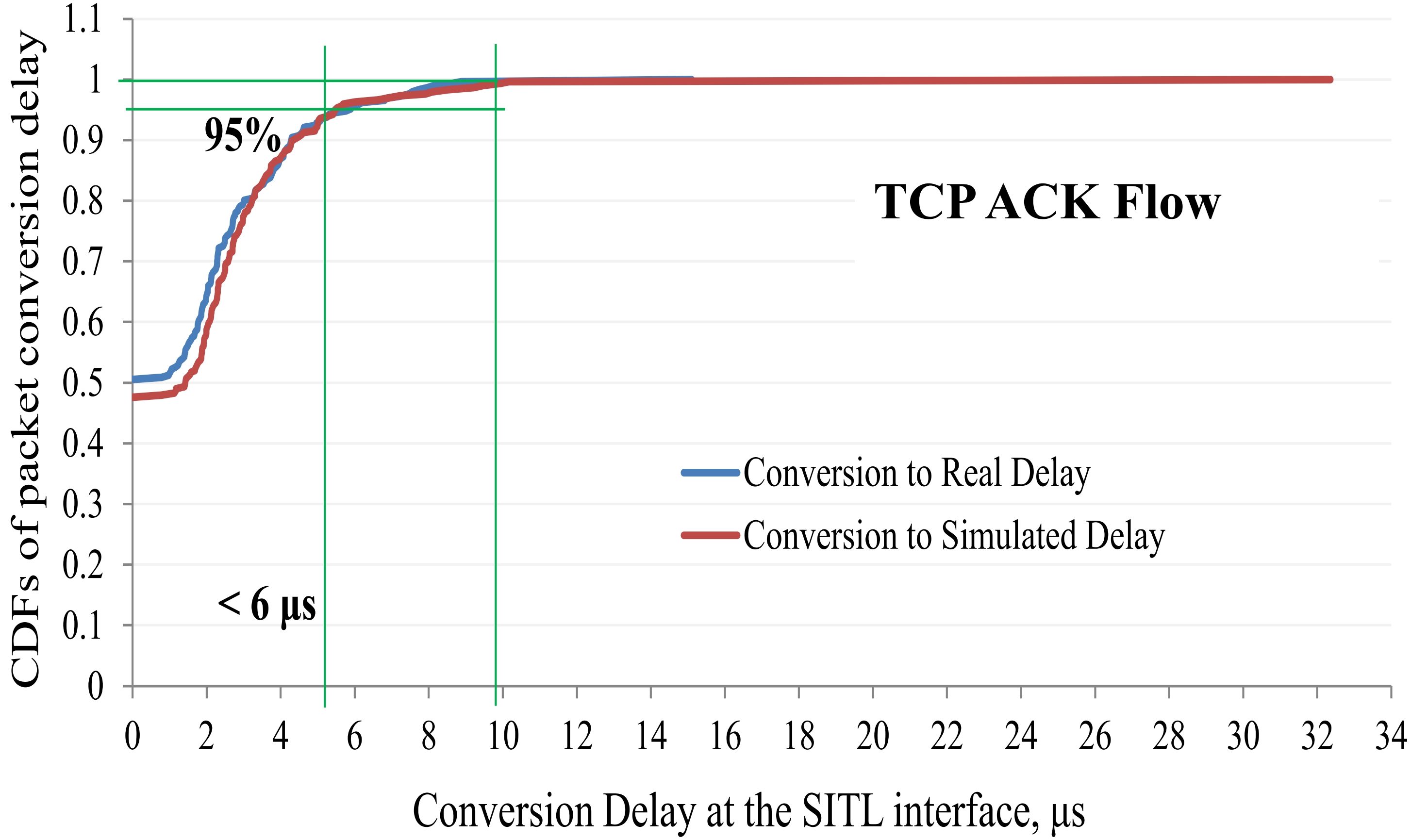
Results

Packet conversion latency measurements

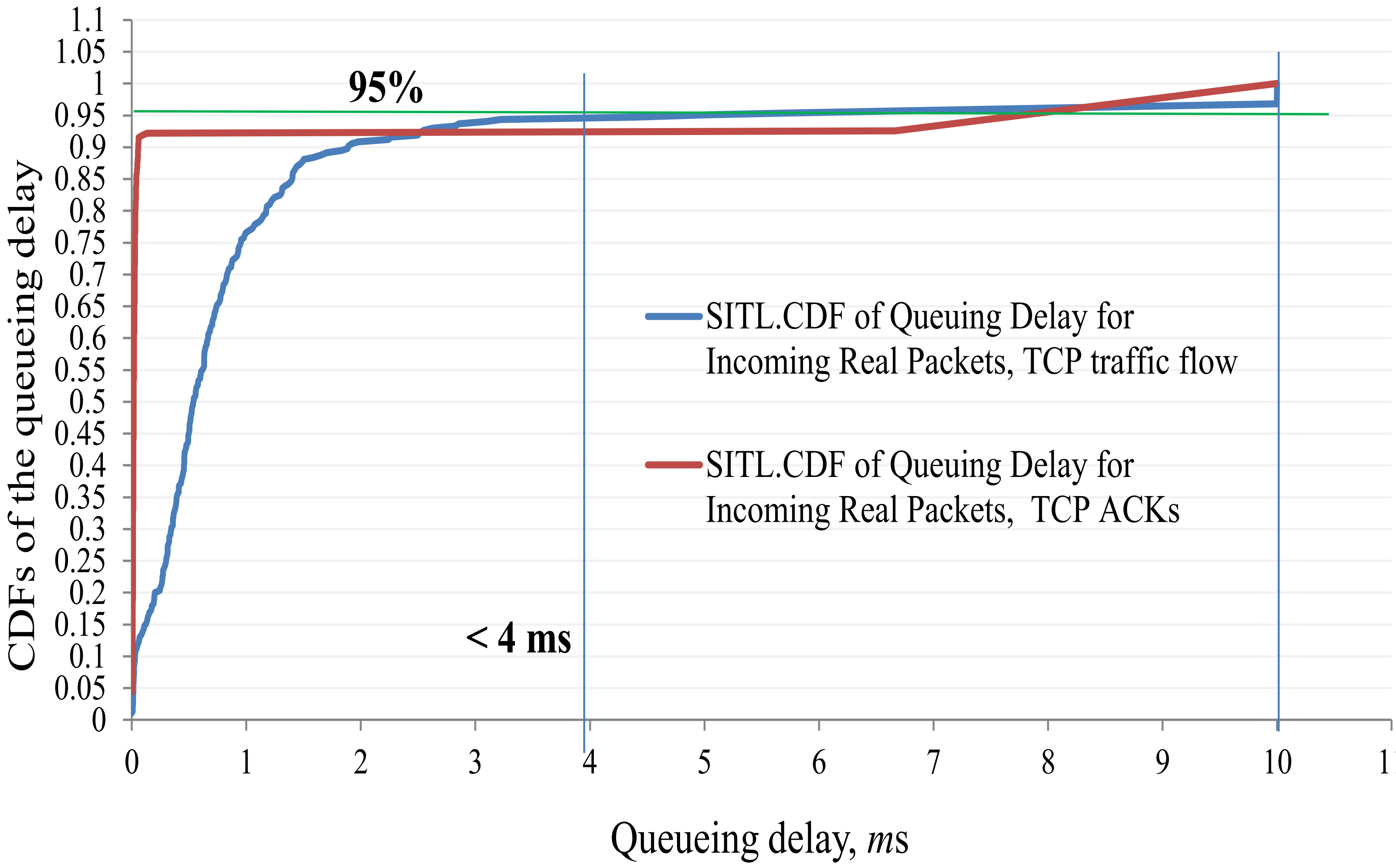


Results

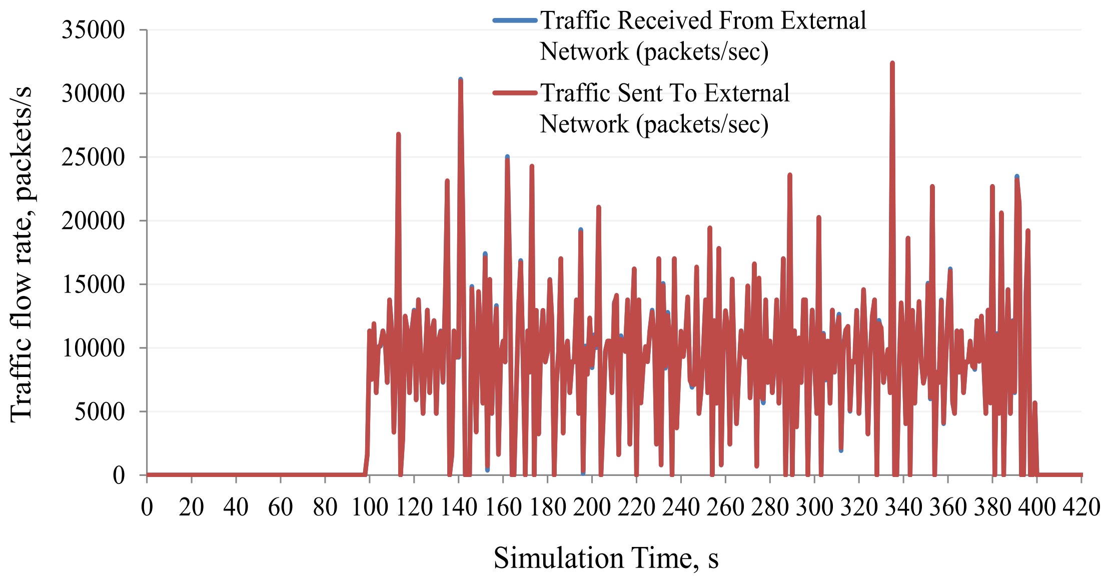
Packet conversion latency measurements



Results
Packet queueing latency measurements

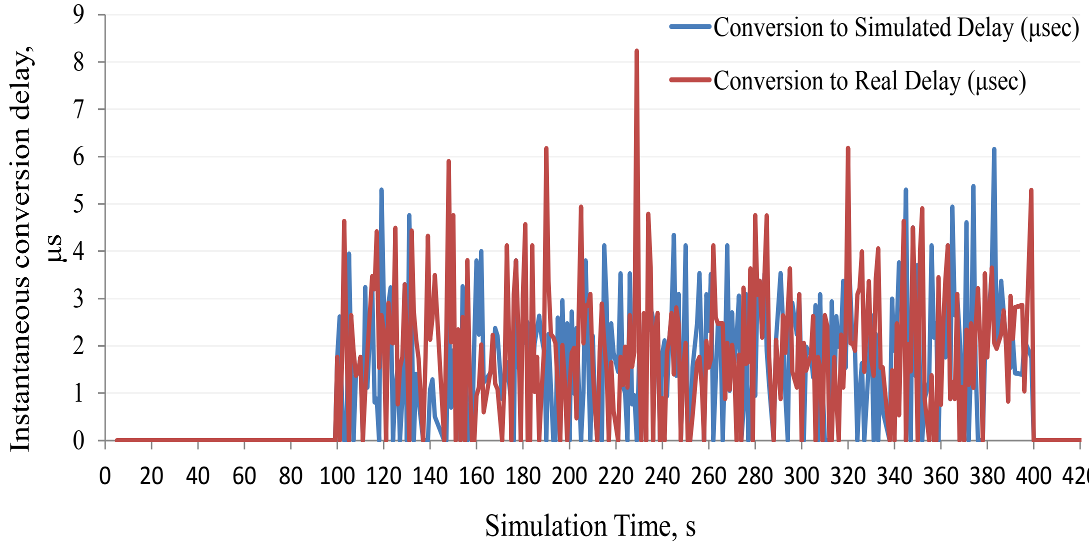


Results
Real traffic generation experiments (2)

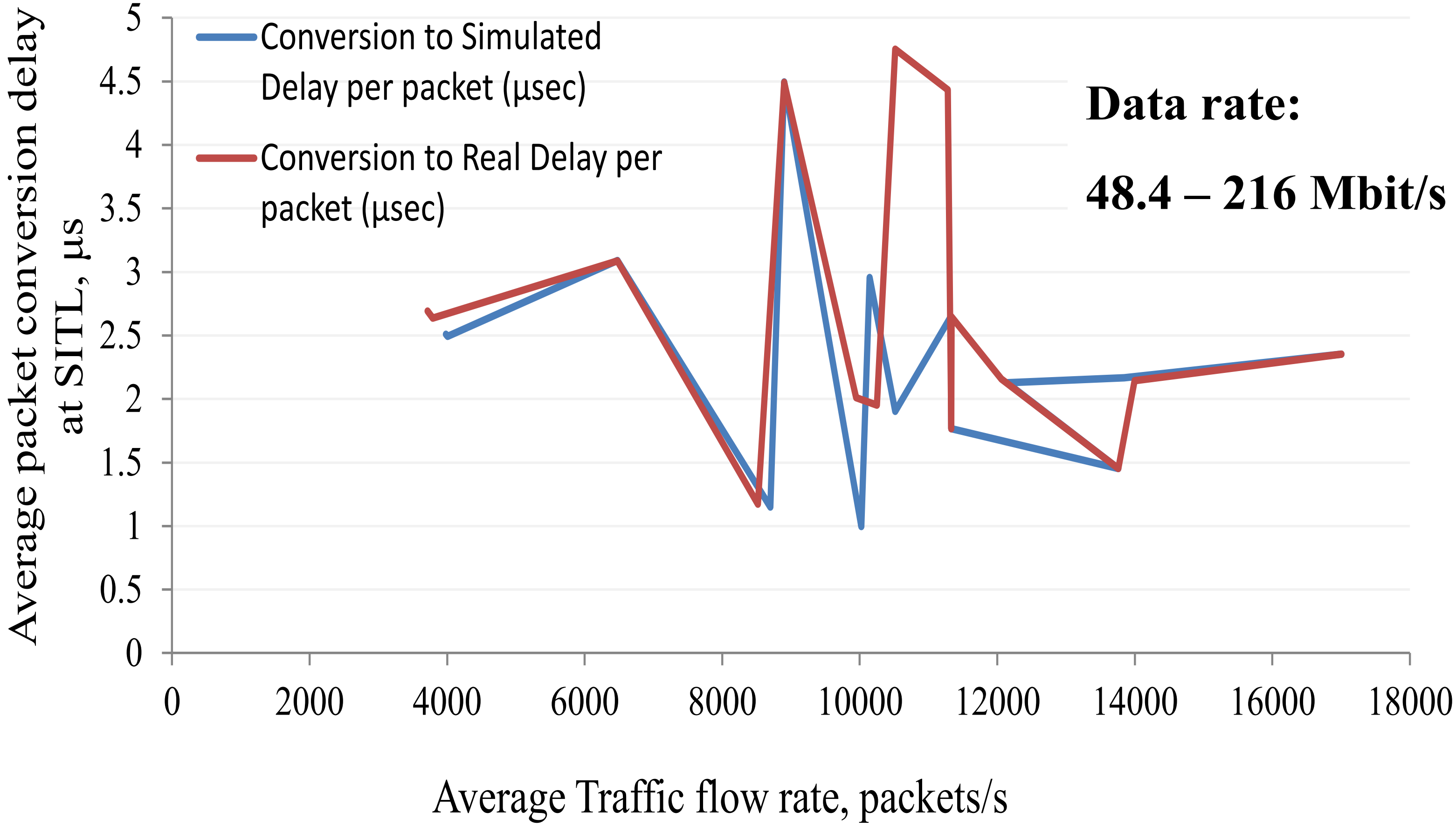


Data Flow:

3.41 GB File transfer



Results
Packet conversion latency measurements



Data rate:

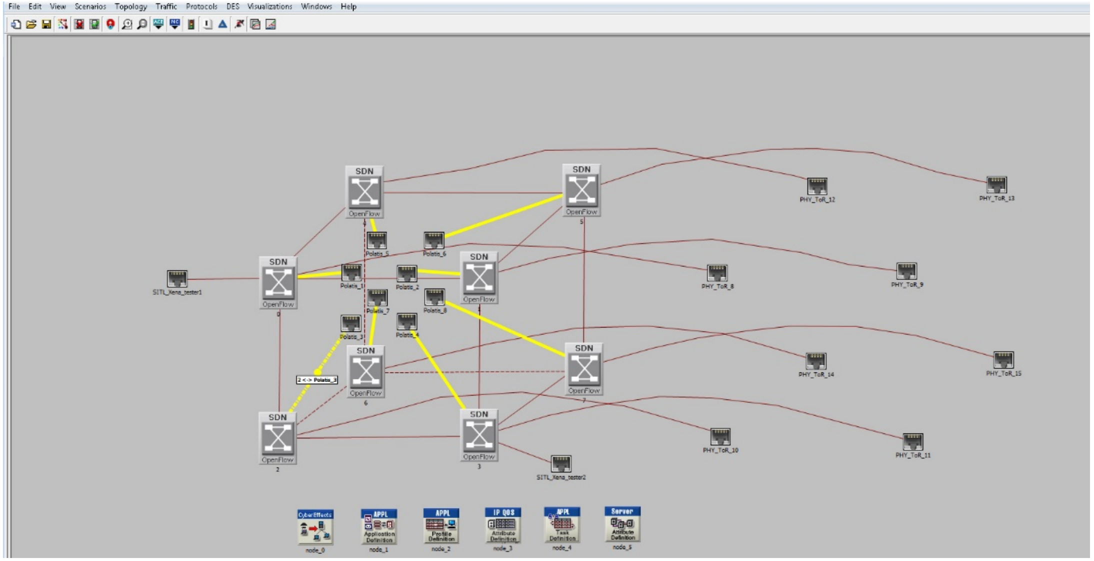
48.4 – 216 Mbit/s

Takeaway points

- Building a hybrid simulation-real-hardware experimental setup for DCN performance studies – not a trivial task!
- Two critical latency components must be taken into consideration: packet conversion delay and packet queueing delay (SITL gateway).
- Packet conversion delay:
  - Packet translation depth (traffic type)
  - Specifics of packet capture by the WinPcap [9] (libPcap for Linux) module
  - Conversion functions (code efficiency)
  - NIC characteristics/functionality
- SITL gateway adds a conversion delay in the order of microseconds (***μs***) as well as load-dependent buffering delays (***ms***)

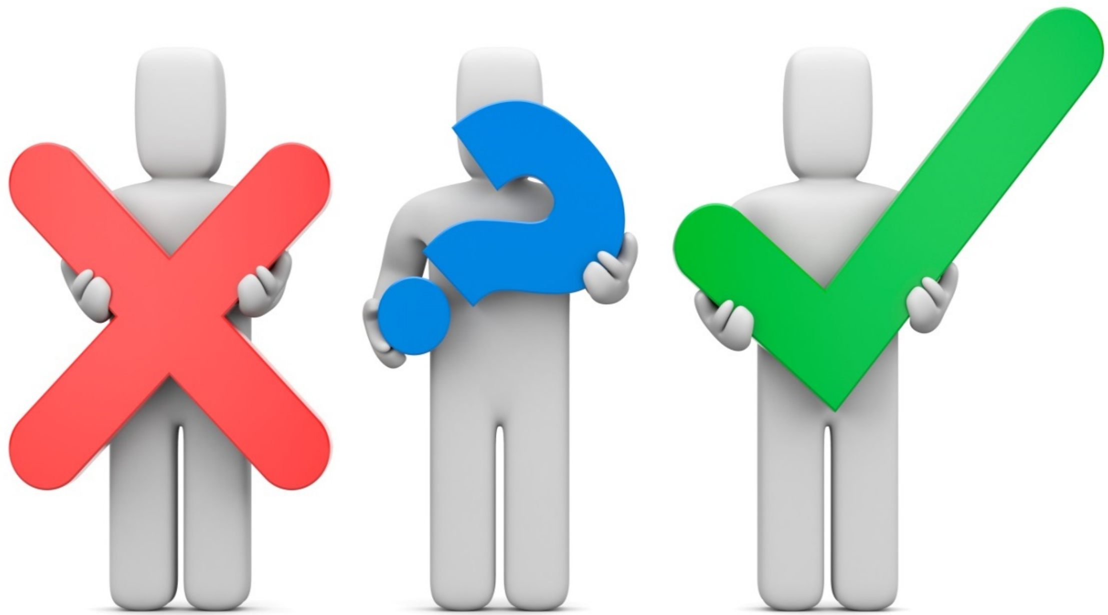
Future steps

- Consider the impact of more realistic high bit-rate mixed traffic patterns, bursty workloads (DC workload traces or a real Map-Reduce cloud application)
- Simulation model on a powerful workstation with multiple electro-optical interfaces (10G/40G)
- Wider range of performance evaluation scenarios and metrics
- Intergation with an SDN framework (hybrid SDN-controlled setup) and OCS physical nodes



Thank you for your attention!

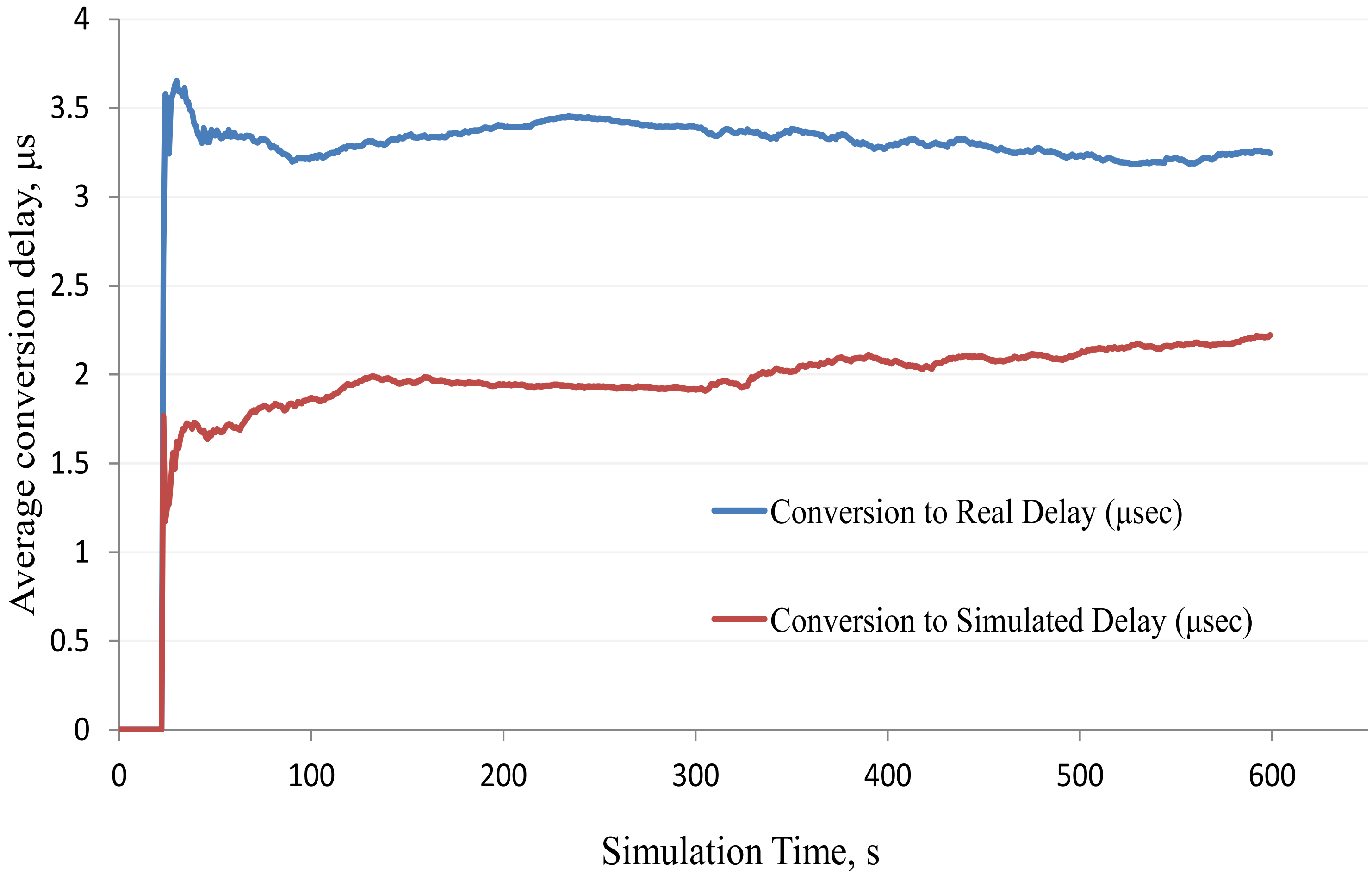
Questions?
Suggestions?



Contact: Artur Pilimon, artpil@fotonik.dtu.dk

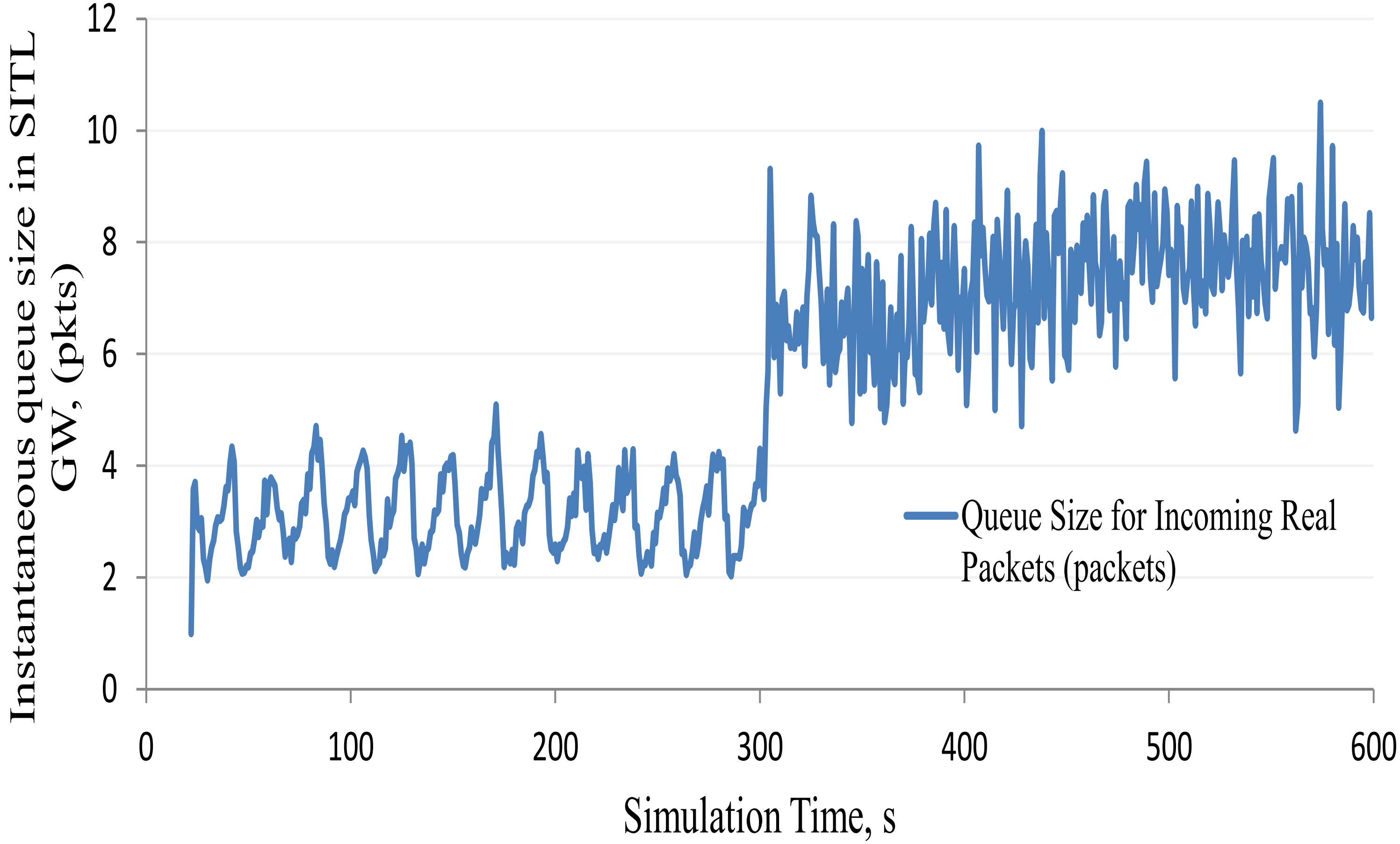
Thank You for Your attention

Results
Packet conversion latency measurements



Results

Packet conversion latency measurements



Software models and Tester interface