First Demonstration of Ultra-low Latency Intra/Inter Data-Centre Heterogeneous Optical Sub-lambda Network using extended GMPLS-PCE Control Plane

Bijan R.Rofoee [(1)], George Zervas [(1)], Yan Yan [(1)], Dimitra Simeonidou[(1)], Giacomo Bernini [(2)],Gino Carrozzo [(2)], Nicola Ciulli [(2)], John Levins[(3)] , Mark Basham[(3)], John Dunne[(3)], Michael Georgiades[(4)], Alexander Belovidov[(4)], Lenos Andreou[(4)], David Sanchez [(5)] , Javier Aracil [(5)], Victor Lopez[(6)], Juan. P. Fernández-Palacios [(6)]

[(1)] High-performance Networks Group, University of Bristol,UK, (Bijan.R.Rofoee@gmail.com)
[(2)] Nextworks, via Livornese 1027, 56122 San Piero a Grado, Pisa, Italy
[(3)] Intune Networks Limited, Blocks 9B-9C Beckett Way, Park West Business Park, Dublin 12, Ireland
[(4)] Primetel, The Maritime Center, 141 Omonia Avenue, 3045 Limassol, Cyprus
[(5)] Universidad Autónoma de Madrid, Campus Cantoblanco, Madrid, Spain
[(6)] Telefónica I+D, c/ Don Ramón de la Cruz 82-84, Madrid, Spain

Abstract *This paper reports on the first user/application-driven multi-technology optical sub-lambda intra/inter Data-Centre (DC) network demonstration. Extended GMPLS-PCE controls two heterogeneous intra-DC optical sub-lambda networks to deliver dynamic and guaranteed data transfer of ultra-low latency (<270µs) and jitter (<10µs) for end-to-end services.*

Introduction

Cloud based applications and Network Centric services and consumers such as PC virtualization, Video/Game on Demand (VoD, GoD), storage area network (SAN), Data replication, and etc, have transformed traditional data-centres to massive scale computing infrastructures[1,2] with highly complex interconnectivity requirements. The current hierarchical electrical L2/L3 Data-Centre (DC) networks can highly suffer from scalability, resource inefficiency, and high latency and non resiliency in delivering application services[3]. As such they would considerably benefit from flexible ultra-low latency finely granular optical network technologies which integrate seamless provisioning of combined intra/inter DC cloud-based computing and network services while facilitating resource usage efficiency and network scalability.

In this paper we present, to our best knowledge, for the first time, a full implementation of multi-technology sub-lambda ultra-low latency/jitter intra/inter DC optical network controlled by a technology-agnostic unified GMPLS-PCE. It consists of two different optical sub-lambda switched intra-DC research prototype testbeds: a) a synchronous multi-wavelength and topology-flexible Time-Shared Optical Network (TSON), and b) an asynchronous tunable Optical Packet Switch Transport (OPST) ring. The extended Generalised Multi-Protocol Label Switching (GMPLS), Path Computation Engine (PCE) and Sub-lambda Assignment Element (SLAE) provide user/application -driven dynamic end-to-end sub-lambda network services addressing intra-DC dynamic networking environments. The inter-DC connectivity is through a pre-established WSON network.

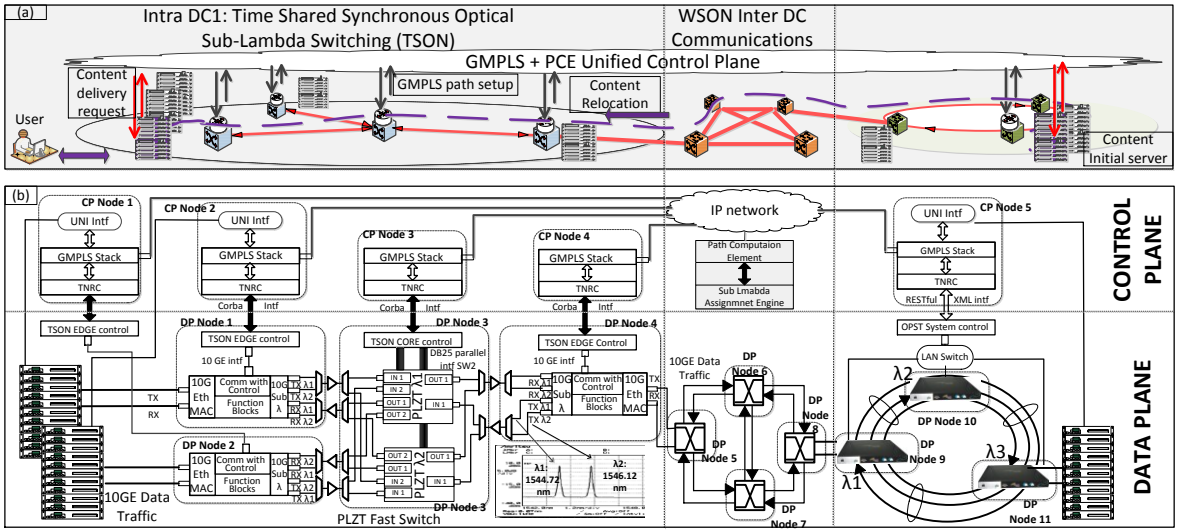

Fig 1: (a) Experimental intra/inter DC network topology with content migration scenario (b) Testbed

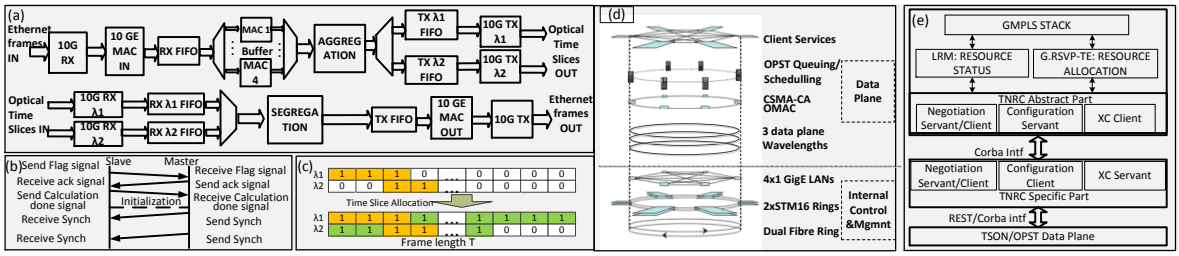

Fig 2: (a) TSON FPGA function blocks; (b) TSON synch protocol; (c) TSON time slice allocation; (d) OPST prototype system; (e) Control plane vertical structure

The multi-layered, multi-technology intra/inter DC networking solution along with its enhanced unified control-plane has been evaluated for all operational layers and processes individually and combined. It demonstrates dynamic path setup over ultra low latency (<270 µs)/jitter (<10 µs) and fine-granular (100Mbps up to 5.7Gbps with 100 Mbps step-size) optical sub-wavelength switching testbeds.

Intra/Inter DC Test-bed and scenario

Fig 1(a) displays the implemented IT+network testbed containing in total 11 optical nodes and several servers for intra/inter DC networks. The two intra-DC sub-lambda technologies are the TSON[4], implemented in a 4-node star topology, and the 3-node ring of tunable OPST system[5]. The two intra-DC networks are interconnected via a 4-node partial mesh WSON network. Cloud-based (Virtual PC migration) user/application request, initiates the GMPLS-PCE-SLAE unified control-plane, to set up and tear down lightpaths at the sub-lambda granularity within and between DC networks, for Virtual PC and VoD content migration/relocation over the cloud.

Fig 1(b) provides more detailed view of the implemented data-plane (DP) and control-plane (CP) nodes of TSON, OPST, and WSON networks. TSON is a fully bi-directional synchronous and frame based yet flexible system, with 1ms frame and 31 time-slices. TSON is implemented using high performance Nx10Gbps (control and transport) Virtex-6 FPGAs boards as well as active and passive optical components - four 2x2 10-ns PLZT switches[6], six DWDM 10G SFP+, EDFAs, MUX/DEMUXes, etc - with edge and core node functionalities. Each TSON edge node (Node 1,2,4) uses four SFP+ transceivers, two 1310nm 10km reach for end-point server traffic and control, and two DWDM 80Km reach transceivers at 1544.72nm and 1546.12nm. Ethernet-TSON and TSON-Ethernet FPGA functions are displayed in Fig 2(a): ingress 10GE traffic into TSON domain is buffered based on the traffic Dest MAC address (up to 4 MACs), aggregated in TX FIFOs to form optical signal burst (up to 8 1500 byte Ethernet frames in a time-slice), and are released on the allocated time slices and wavelength(s), providing flexible bitrate support from 100 Mbps up to 5.7 Gbps with 100Mbps rate granularity. On receiving optical signals in TSON Edge nodes Ethernet packets are extracted and sent out by segregating the optical signal. TSON core node (Node 3) on the other hand, controls four 2x2 PLZT switches, directing the incoming optical time-sliced signals on the appropriate output port, using parallel DB25 interfaces. TSON requires frame/time-slice synchronisation among TSON nodes. We have implemented a 3-way frame synchronisation protocol shown Fig 2(b) to tune and maintain a global frame synch at 1 FPGA clock-cycle (6.4 ns) accuracy between TSON FPGA nodes. The master clock sends synchronisation frames to the slave clocks regularly. So the slaves use the time stamp and the delay between the nodes to compensate for clock variations and drifts. Time-slice synchronization is performed by having fibre link lengths multiple of time-slice duration.

The second sub-lambda system, the OPST* (Node 9-11) collapses layers 0 to 2 under the same internal ring network control-plane, transforming the entire ring into a distributed switch that operates as a single new network element Fig 2(d)). The collapsing of layers 0 to 2 is achieved by using ultra-fast nsec tunable laser transmitters on the line side of each externally facing client port. The ring uses a wavelength per destination routing scheme to address packet flows. This is implemented using a wavelength selective switch and ns speed tunable transmitter. When the transmitters are used on optical burst mode, virtual wavelength paths can be set up and pulled down in response to incoming packet flow requirements. The result is an ability to merge packet flows from different sources optically, so that they arrive multiplexed in time at the destination. The system uses a new Optical Media Access Control system (OMAC), which employs a Carrier Sense Media Access with Collision Avoidance (CSMA-CA) to avoid ring-wide synchronisation. Finally, the inter DC 4-node (Node 5-8) bidirectional WSON partial mesh

* This OPST system is a research prototype testbed

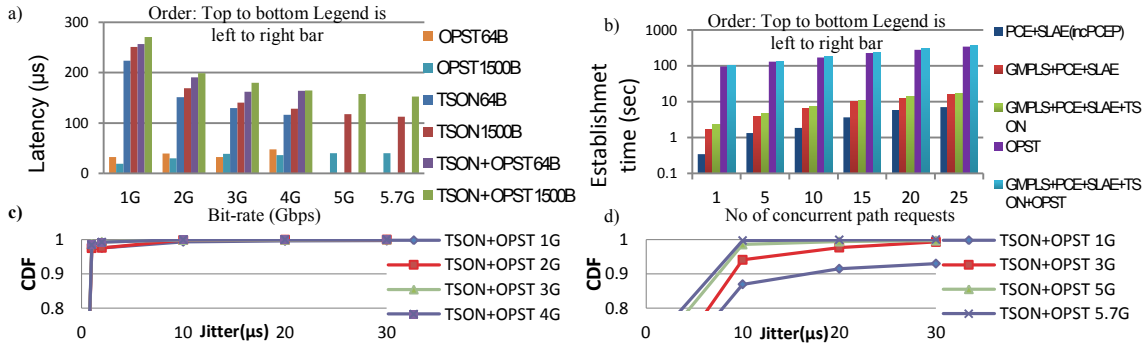

Fig 3: (a) Data-plane results for different bitrates; (b) control-plane results for different parallel path requests;(c) data-plane jitter for 64 Byte packets;(d) Data-plane jitter for 1500 Byte packets;

network is built using 3D MEMS Switched Network.

Extended GMPLS-PCE-SLAE Control Plane

The implemented multi-technology GMPLS stack (Fig 2 (e)) for the first time delivers specific extensions and procedures to support the sub-lambda switching granularity: sub-lambda network resource modelling, Sub-Lambda Assignment Engine (SLAE) for TSON, enhanced GMPLS+PCE routing algorithms and RSVP-TE protocol extensions for sub-lambda resource reservation. In action, the GMPLS edge controller is triggered from the UNI interface for setting up an end-to-end sub-lambda lightpath. It invokes the PCE for a TSON+OPST multi-layer route. PCE then calls SLAE for time slice allocation over TSON region, and SLAE allocates free time slices using its data-base (Fig 2 (c)). After path and time-slice computation, the GMPLS edge controller starts RSVP-TE signalling for setting up the multi-layer path over the TSON and OPST domains. GMPLS stack at each hop (whole OPST ring constitutes a single hop, while each TSON node is controlled as an independent entity) communicates with the DP node for resource reservation, using developed Transport Network Resource Controller (TNRC) module (as CP to DP translator making GMPLS DP technology agnostic) with Corba (for TSON), and XML RESTful (for OPST) interfaces.

Evaluation and results

Individual and integrated comprehensive end-to-end L2 results for various bitrates up to 5.7 Gbps are presented in Fig 3(a). Latency and jitter of OPST system delivers ultra low latency (<40 µs) and low jitter (<10 µs) independent of the traffic load. TSON system delivers increased by yet very low latency (<260µs) and ultra-low jitter (<5 µs) due to time-sliced aggregation. It should be noted, the higher the bitrate, the faster the aggregation and buffering, therefore the end-to-end intra-inter DC TSON- WSON-OPST latency drops from 270µs (at 1 Gbps) to ~150 µs (at 5.7 Gbps) for 1500B Ethernet frames having the jitter being < 10 µs. Also, the jitter delay is dependent on packet size, where the 1500B Ethernet frames will get most varying delays queuing in buffers (Fig 3(c) and (d)).

Complete end-to-end setup time for parallel and concurrent lightpath requests from GMPLS invocation (at the UNI-gateway) until transmission of data has been measured for different phases and technologies of operation in Fig 3(b). It can be seen the path computation and control-plane operations in the busiest scenario of 25 concurrent lightpath requests take up to 10 seconds for all the operations. Adding TSON DP with SLAE to the measurements, the latencies rise to 12 seconds for 25 parallel requests scenario. Added OPST system, this value increases to around 400 seconds due to internal OPST operations.

Conclusion:

We have demonstrated for the first time a fully implemented multi-technology, multi-layer, intra/inter DC heterogeneous sub-lambda network containing advanced optical control and DP solution. The demonstration is an 11-node network testbed containing two sub-lambda packet and time-shared systems as two intra DC technologies, interconnected through a WSON network. They are controlled by IT/Resource aware, technology agnostic extended and unified GMPLS-PCE-SLAE CP enabling end-to-end data delivery over the ultra-low latency (<270µs) and jitter (< 10 µs) sub-lambda multi-wavelength (100Mbps-5.7 Gbps) data planes.

Acknowledgment: This work is supported by the EC through IST STREP project MAINS (INFSO-ICT-247706), as well as EPSRC grant EP/I01196X: Transforming the Internet Infrastructure: The Photonics Hyperhighway.

References:

[1] C. F. Lam et al., Coms. Mag., July (2010).
[2] A. Vahdat et al., OFC'11, OTuH2 (2011).
[3] Data Center Networking Enterasys (2011)
[4] G. S. Zervas et al, J OPEX, (19) 26, (2011)
[5] G. Zervas et al, FUNMS,July (2012)
[6] K. Nashimoto, OFC'11,OThD3, (2011)