# Architectures and Technologies for Data-Center Interconnection

A. Stavdas, C. Matrakidis, T. Orphanoudakis
Dpt of Informatics and Telecommunications, University of Peloponnese
Greece

Antonio Manzalini
Strategy and Innovation Future Centre
Telecom Italia,
Turin, Italy

Ricardo Martínez
Centre Tecnològic de Telecomunicacions de Catalunya,
Spain

***Abstract*— Socio-economic drivers, progress in IT technologies, reduction of hardware costs and availability of open source software are steering the evolution of current Telco infrastructures towards a highly dynamic and flexible environment of virtual resources capable to serve multiple applications. Emerging paradigms such as Software-Defined Networks (SDN) and Network Functions Virtualisation (NFV) deeply integrated with optical technologies will create the favorable conditions for a change of paradigm. We propose a novel optical cloud architecture where IT and Telecom resources are used interchangeably as common infrastructure. Key assets are the shared access passive networks for distributed multiplexing and grooming and a node architecture integrating transmission and switching**

***Keywords—architectures, networks, data-centrers***

## I. INTRODUCTION

Three main innovations are driving the evolution of the Internet today. The first is the emergence of *Big-Data* leading to resource delocalization that are subsequently distributed in the net. The second is *Virtualization*, as both machine processing commoditization and network functions abstraction are enabled by the availability of cheap and powerful computers. The third driver is *Software Defined Networking* (SDN) which is freeing the control plane of telecom equipment from proprietary software so they can be adapted to the specific needs of the user. As it is elaborated in this paper, these trends are pushing optical technology deeper into the Data-Centers (DCs) leading to the merging Telco and IT Industries which must work in harmony to deliver real-time results. So far, the two industries -that traditionally had separate supply chains, organizations and methods- will have to integrate with each other to support the future of the ICT services.

We propose an Optical Cloud architecture amalgamating IT and Telco architectures, infrastructures, systems, control mechanisms, services and applications in a flexible and manageable infrastructure where the available resources from both brands are used interchangeably, regardless of physical location. The proposed architecture allows integrating, by design, this infrastructure into an elastic optical transport network, federating islands of IT and Telco domains into one single distributed IaaS. To this end, an orchestrator entity is then able to coordinate, slice and optimize any of the resources at will.

## II. OPTICAL CLOUDS

IaaS requires an "architecture of architectures"; an Optical Cloud that may: a) host and efficiently handle a gigantic pool of digital resources regardless of their physical location; b) offer high flexibility and programmability; c) able to guarantee QoS performance and in particular, ultra-low latency; d) offer high level of security; e) rely on low CapEx/OpEx solutions.

Therefore, the proposed clean-slate architecture is aiming: a) to reduce -as much as possible- the protocol stacks and the number of interfaces. This would reduce costs and increase efficiency; b) to limit -as much as possible- routing and switching. Consequently, this would reduce costs and delays as well as make the platform more manageable; c) to delegate as many networking functions as possible to the optical layer exploring transmission and passive multiplexing. This will enhance performance and will simplify the network processing and node interconnection; d) To exploit -as much as possible- the advances in technology for “storage” and “processing” of digital data but not for data-forwarding; e) To execute, by means of virtualization, higher-functions and services (L3-L7) that are implemented on logical resources Virtual Machines (VMs) which are flexibly instantiated and/or migrated across a number of collocated or disjoint physical resources

The proclaimed integration of IT and Telco infrastructure, is realized by means of an Optical Cloud Interconnect Node (OCIN). This is an “one-purpose” infrastructure which serves as both DC and Telco node at the same time: the IT and Telco services are provided in parallel and can change dynamically by slicing portions of IT and/or Telco h/w and s/w resources within a said node as it is schematically shown in Fig. 1. As final step, the federation and integration of node into a flexible and dynamic transport network is necessary for the realization of a ubiquitous Optical Cloud. To coordinate these processes, the role of a Global Orchestrator is essential. Specifically, it relies on open interfaces and a flexible and dynamic control-plane entity, following the SDN principles, to implement the required control functions as shown in Fig. 2.

## III. ARCHITECTURAL SELECTIONS

To ensure scalability and manageability, we adopt a layered intra-node architecture consisting of three Tiers (Fig.

1). Moreover, within a single node, we proceed to the separation between "processing" and "storage" functions from the corresponding data-forwarding functions; the latter being implemented without resorting to the former.

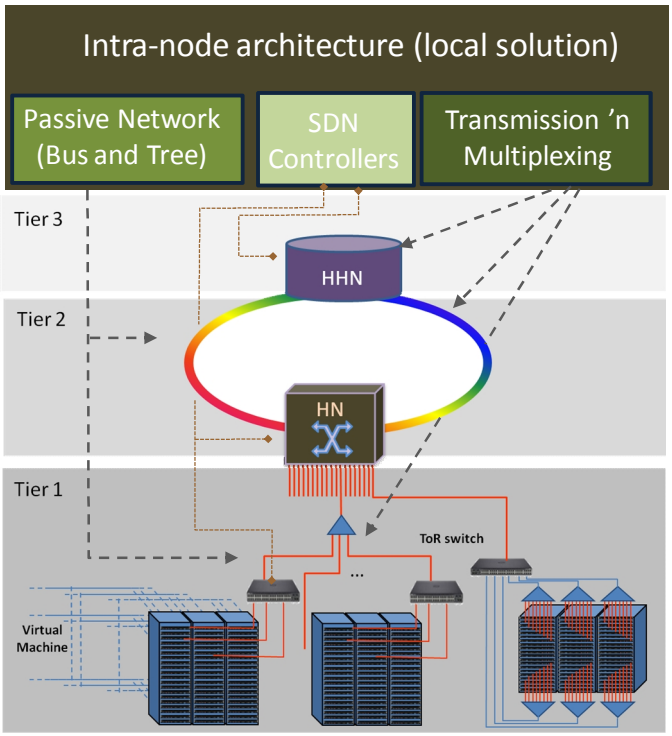


Fig. 1: Overview of the proposed layered architecture

This approach combined with the separation between data and control planes sets the necessary framework to reduce the total number of interfaces and the multi-layer protocol stacks involved in data-forwarding schemes (a process known as de-ossification).

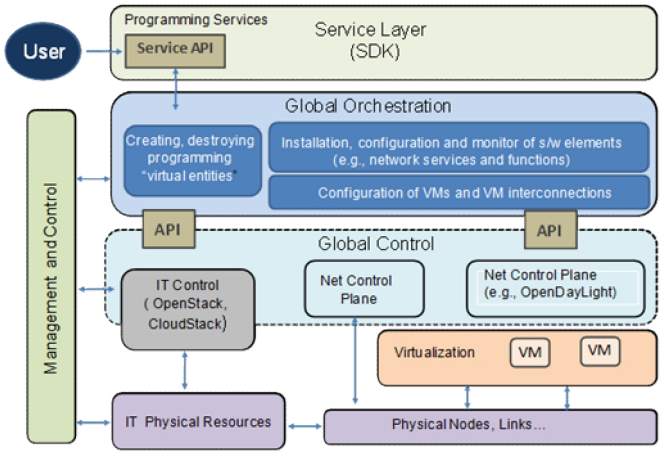


Fig.2: A Global Orchestrator is a necessity in an Optical Cloud

One may ask: what mechanisms and technology platforms we need to explore in order to replace the current paradigm that implements data-forwarding by means of excessive "processing" and "storage"? The answer is: transmission and multiplexing could be a viable alternative to the excessive multi-layer electronic routing and switching to the realization of connectivity fabrics. The conditions needed for this change of paradigm are the following:

1) *Minimize large multi-layer/protocol electronic switches.* The concept is simple: only small/medium size L2 electronic switches need to be deployed that are, then, interconnected by means of an agile optical system as follows: a) the PONs shown in Fig. 1, are operated as a distributed aggregation medium and, as such, they could flexibly interconnect the switches providing dynamic bandwidth allocation (DBA), fairness, QoS guarantees etc via a medium access control (MAC) mechanism. Specifically, a tree-topology PON is used in Tier 1 and a bus-topology in Tier 2; b) the data-forwarding mechanism in node's gateway (HHN in Tier-3) is implemented by means of a switchless architecture which is exclusively based on transceivers in an innovative way that is integrating transmission and switching.

2) *Use open and programmable controller interfaces*, for both IT and SDN controllers in all Tiers as a measure to coordinate platforms with dissimilar operating systems.

Under such framework, the three Tiers are detailed as follows:

**Tier-One:** This is the segment where the entire processing and storage functions and services of the node are implemented. A tree-topology PON is used to provide connectivity between the Racks hosting the IT/NFV/Telco processing and storage systems, to the node residing at the interface between Tier-1 and Tier 2 in Fig. 1, that is called High-Node (HN).

Following NFV principles, network functions and operations like NAT, BRAS, LB, EPC mobile core, etc. can be emulated by virtualization nodes (V-nodes) in Tier-1. The IT category also includes storage nodes (IO-nodes), small size Ethernet switches at Top-of-Racks (ToRs), intra-Rack Storage Area Network (SAN) switches, management boxes, etc. Nevertheless, not all Telco functions and operations can be NFV-based so dedicated purpose h/w is still needed like switching machinery in Routers, CPRI switches etc.

The overall traffic profile stemming from a Rack strongly depends on the ratio between the V-nodes and IO-nodes in that Rack and the traffic volume each node type is producing, so the final outcome depends on the application.

The essence of a universal node is that it may dynamically change its role, i.e. it may facilitate more as a "Data-Centre" or more as "Telco node", as a result of the high-level decisions of the Global Orchestrator. To do so, the node should be able to dynamically and interchangeably resort to either the IT/NFV or the Telco h/w placed in a node, or to add new h/w, in a straightforward way without complex multi-protocol operations and scalability problems. The unspecified mixture of V-nodes, IO-nodes and the proprietary Telco h/w is expected to create a traffic profile with high spatial and temporal asymmetry. In such framework, a PON allows using the available resources more efficiently by dynamically allocating bandwidth to "hot" Racks at the expense of the

relatively idle ones. This is done without resorting to multi-protocol packet level electronic switches and, pending on the application, the statistical multiplexing gains of a PON can be substantial. Moreover, a PON saves a large number of transceiver pairs compared to the case where point-to-point (ptp) links are employed. Last but not least, the Server-to-ToR connectivity is reaching today silicon switching limits. The use of PONs is steering away from over-loaded Leaf-and-Spine solutions to a protocol agnostic platform scalable to high throughput.

**Tier-Two**: PONs and ptp links from Tier-1 are interfaced to the lower-bound inputs of HN. The HN itself consists of a) an electronic layer with a L2 switch as the main element; b) the optical interfaces which as shown in Fig. 5 consist only of power couplers and receiver-transmitter arrays. It is important to point out that apart from the interfaces to/from Tier-1, the L2 switch is also the end point of the access networks. Therefore the L2 switch operates effectively as a backplane for Tier-1 by forwarding traffic i) to/from the access networks to the proper Tier-1 terminals; ii) from Tier-1 terminals to the rest of the Optical Cloud.

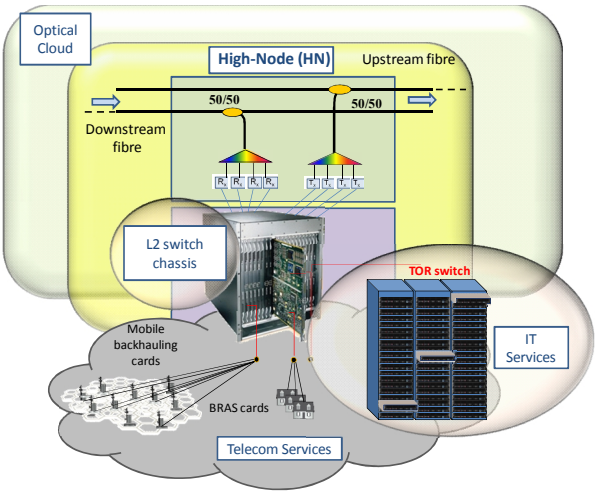


Fig.3: A schematic layout of the HN

To interconnect HNs, a bus-topology PON is deployed in Tier-2 as shown in Fig. 1. The optical bus and the HNs attached to it constitute a cluster. The packet forwarding between HNs is set-up, regulated and implemented by the Highest-Hierarchy Node (HHN) since no direct connection between HNs is possible. The source-HN identifies to the HHN the destination-HN and the HHN allocates the necessary optical resources. In Tier-2, the optical bandwidth is exploited by means of two transportation modes: a) the Time-slicing or TDMA mode as detailed in [1]; b) Spectral-slicing or SDMA mode where the HN is using sliceable bandwidth variable transceivers (S-BVT) and the optical bandwidth is explored by means of a number of optical spectral slots; the number of slots allocated to each HN changes dynamically by means of a suitable DBA mechanism implemented in the HHN to adapt to the actual traffic each HN contributes to the cluster.

**Tier-Three**: The role of this segment is central to achieve the objective for an architectural integration of a single node in the entire Optical Cloud ecosystem as shown in Fig. 4. The HHN in Tier-3 is the gateway and the physical link between Tier-2 and the flex-grid optical transport network.

The integration of Tier-2 optical bus, the HHN and the transport core is made possible by means of a) a novel HHN node architecture, and b) a novel forwarding mechanism. *HHN architecture*: The HHN is a multi-granular optical node consisting of two blocks: a wavelength-selective switch (WSS)/Bandwidth Variable (BV)-reconfigurable optical add/drop multiplexer (ROADM) block and a Switchless Elastic Rate Node (SERANO) block. The SERANO block does not only groom the traffic to the transport network but also provides connectivity between the Tier-2 clusters. Last but not least, SERANO coordinates with the DBA of Tier-2 to ensure a lossless forwarding of traffic between HNs.

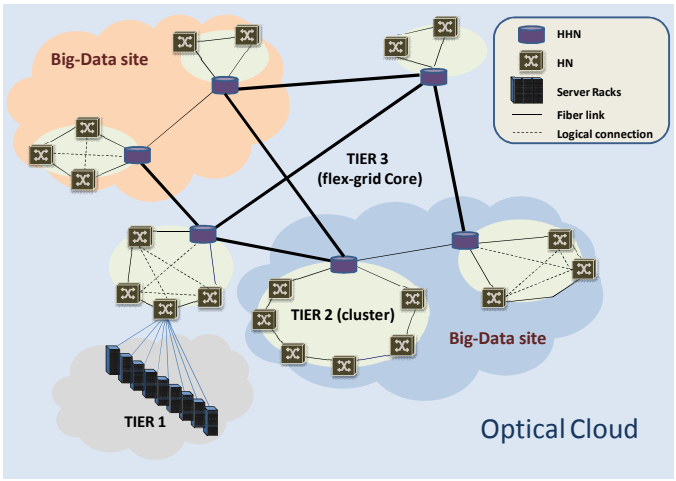


Fig.4: The multi-Tier architecture with emphasis on Cloud interconnection and the Big-Data sites

Assuming N I/O fibres in SERANO, there is an "optical blade" per fibre which consists of an 1:N optical splitter where each of the N outputs is sent to N WDM cards in parallel as shown in the inlet of Fig. 5. A WDM card consists of an optical de-multiplexer, followed by an S-BVT array where each of them is arranged in the form of fixed-receiver tunable transmitter pair connected back-to-back.

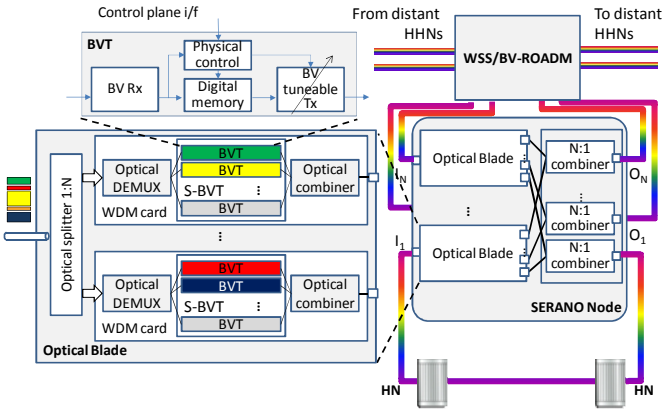


Fig.5: A schematic layout of SERANO

Therefore, the functionality of this block is to terminate the signal providing full 3R regeneration of the entire comb of flows but it re-emits only selected flows (i.e. those directed to the selected output fiber at the desired spectral slot/modulation format). The final stage of the WDM card consists of a combiner with M input ports, M being the upper number of flows that a particular input fiber may forward to a particular output fibre. In such a final stage of SERANO, the outputs of each WDM card are passively recombined by means of an additional optical combiner N:1. It is also worth noting that no switching technology is introduced since the main active building blocks are S-BVT arrays so the entire process integrates transmission and data-forwarding allowing the industry to concentrate in optimising a single element for both functions. The decoupling of the S-BV Receiver arrays from the Transmitter arrays paves the way for vendor interoperability.

*A novel forwarding mechanism:* In Tier-2 the optical spectrum is sliced either in SDMA or TDMA mode in an operation controlled from the HHN. The optical spectrum is sliced in a way that all HNs in a given cluster with traffic towards a said other cluster, share the same slice. Thus, with no further processing, a said wavelength/spectral-slot at the input of SERANO's block in the HHN has a uniquely identifiable SERANO output port/fibre, regardless if this is towards another cluster in the same node or towards a cluster in a remote node in the Cloud. This process enables both a) integrating Tier-2 and the transport network through Tier-3, and b) simplifying the overall coordination and Cloud orchestration since HHNs would only need to route large optical slices.

## IV. ADVANTAGES OF THE SOLUTION

A summary of the proposed innovations is the following: a) two different passive network topologies are introduced to implement distributed multiplexing and grooming functions so large electronic switches are obviated; b) the HHN is operates in a switchless and bufferless mode that integrates transmission and switching since the operations are used interchangeably; and c) an end-to-end arbitration and forwarding mechanism that integrates the intra-node with the inter-node routing, under guaranteed QoS performance, is attained.

Thanks to the modular architecture, the fabric scales from few hundred Gb/s to several Petabit/s throughput while the pay-as-you-grow approach minimizes the up-front capital outlay. The overall operations limits buffering to HN's only with no buffering between HNs belonging either to the same or distant clusters and, as such, any delay in slot forwarding is completely minimized. The use of a protocol-agnostic transmission as the main mean of a packet forwarding fabric also paves the way to integrate IT and Telco infrastructures.

## V. THE INTEGRATION OF THE ENTIRE OPTICAL CLOUD ECOSYSTEM:

To maximize the benefits of this architecture, a technology-independent functional model to dynamically instantiate, orchestrate and migrate multiple physical and logical IT/NFV and Telco resources like VMs, VNFs (for NFV), virtual subnets, spectral slots etc, is necessary.

The basic building blocks of this service model are the software components and the logical resources as shown in Fig. 6. The s/w components are modular s/w entities paired with their accompanying configuration parameters, capable of accomplishing specific tasks, which can be executed on top of virtualized resources. On the other hand, the logical resources are elementary logical abstractions of the underlying physical elements (e.g. VMs) with their configuration parameters.

Following the data-plane architecture of section III, an accompanying framework for the Control Plane (CP) and the Management Plane (MP) is essential for the functional integration of the entire IT/NFV and Telco ecosystem into the *Optical Cloud*. Thus, the CP/MP architecture is topologically adapted to the hierarchical layering of the data-plane. It includes a set of s/w components, *the SDN IT & Network Orchestrators (SINOs),* which are responsible for: a) the orchestration of IT/NFV and Telco resources; b) the orchestration of tasks spanning two or more Tiers or they incorporate more than one node as shown in Fig. 4 for the federated Optical Cloud; and c) ensuring technology and vendor interoperability.

The SINOs are classified into the Intra-SINO and the Global SINO. The orchestration of Tier 1-2 is exclusively conducted by the Intra-SINO orchestrator. On the other hand, the Global SINO orchestration manages Tier 3 (flexgrid optical) resources as well as the overall IT and network resource coordination of Tier 1-2 delegating in the Intra-SINO orchestrators. *The coordination between the two orchestration levels* is an essential operation and, therefore, both SINOs incorporate the orchestration of: a) the *IT/NFV resources*, which aims to decide whether and where the virtual IT resources have to be created, allocated, migrated or deleted; b) the *Telco resources* which, according to the position of the VMs, aims to suitably configure via the nodes' control plane entities (e.g., SDN controller) the targeted traffic flows, or modify the paths of existing traffic flows to maintain the service availability, etc.

***The Intra-OCIN SINO:*** It comprises the Telco-side Orchestrator realised as an Application-Based Network Operations (ABNO) module and the IT-side Orchestrator implemented by means of an open Cloud management platform (e.g., OpenStack, CloudStack, etc.).

The role of the ABNO controller in the intra-SINO orchestrator is to coordinate the provisioning of the network services in Tier-1 and 2. The ABNO controllers communicate with the underlying SDN controllers which are interfaced with the network elements by means of open APIs (e.g., RestFul, OpenFlow, etc.). In parallel, the IT orchestrator interacts with the V-nodes and IO-nodes inside the racks with OpenStack/CloudStack either via virtualization APIs provided from the Hypervisors (like for instance Xen or VMware) or via virtual network management APIs that are provided from server's OS modules to implement the communication between VMs, like OpenvSwitch. The joint IT & Telco resource slicing is carried out by the intra-SINO orchestration which enforces these decisions to the ABNO and IT orchestrators.

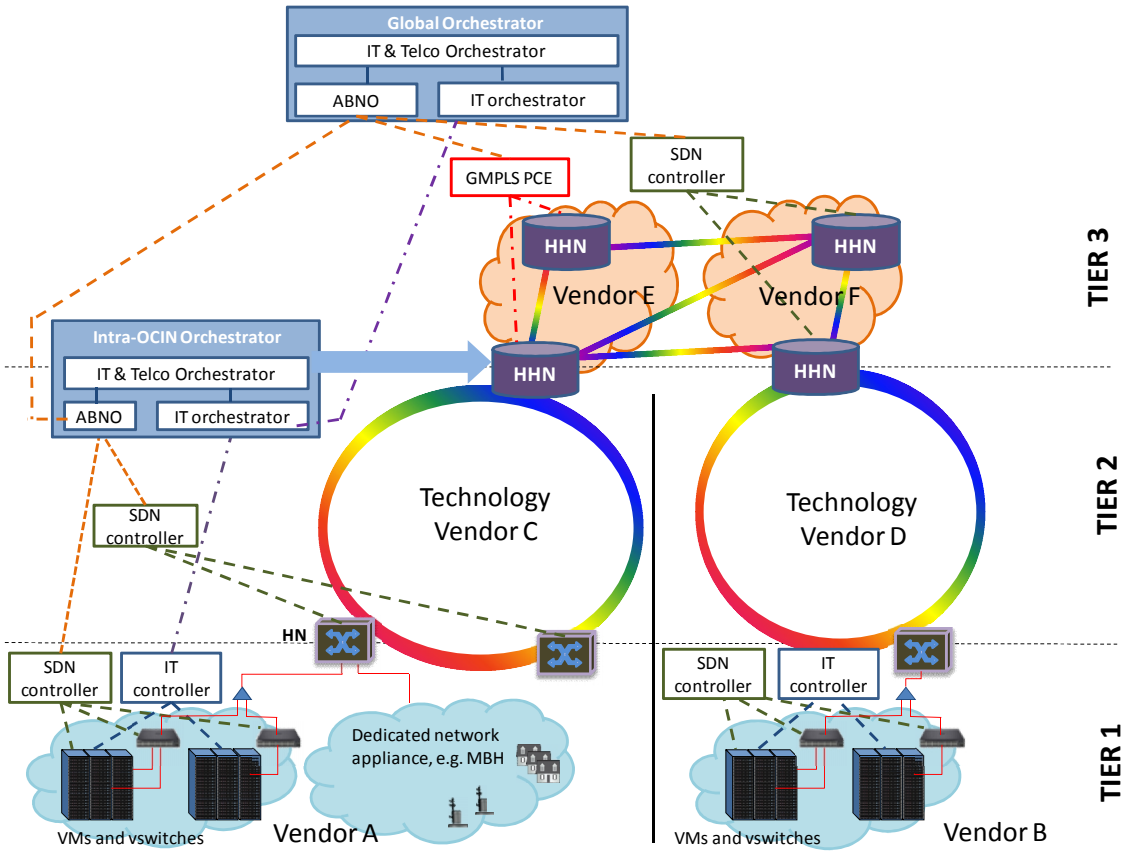


Fig.6: The Optical Cloud ecosystem integration by means of a SINO

The intra-SINO operation is completed in two stages:

In the first stage, it implements the chains/flows of control/management actions. This requires implementing tasks that possibly necessitate using both Telco and IT/NFV resources, by engaging the corresponding SDN and IT controllers. In the second stage, it implements global optimisation algorithms for the fine tuning of the Telco and IT/NFV virtual resources. This fine tuning includes specific policies that resort either to Telco or to IT/NFV in a dynamically changing relationship.

**The Optical Cloud SINO:** The Global-SINO has similarities but also differences compared to the Intra-SINO. The Global Orchestrator makes decisions that are based on the global network and resource knowledge. An important innovation is that the orchestration approach is implemented in a two-axis scheme. Across the horizontal axis, the orchestration is made possible by synchronizing and optimizing both IT/NFV and Telco (virtualized) resources, while across the vertical axis, the orchestration is implementing a hierarchical scheme through all Tiers. The Global Orchestrator incorporates the Global ABNO and the Global IT Orchestrator. In an Optical Cloud they implement the following functions:

The ABNO module coordinates the provisioning of the network services of the Global-SINO. One of the ABNO objectives is the Global network coordination and orchestration through different flexgrid optical domains that may be based on different control plane paradigms following either SDN principles or legacy GMPLS/PCE control.

The IT Orchestrator is the global controller of IT virtual resources and therefore it communicates with all intra-SINOs IT Orchestrators. Thus, the IT orchestrator of a source-cluster, in cooperation with the IT orchestrator of the destination-cluster, may dynamically migrate VMs to servers in new locations or dynamically mirror VMs.

## VI. CONCLUSIONS

We presented the main drivers that are steering the evolution of current networks towards a highly dynamic and flexible environment of virtual resources, interconnected by optical technology. To meet the requirements of emerging distributed and dynamically composed services by the underlying infrastructure we proposed a novel optical cloud architecture where IT and Telecom resources can be jointly optimized and used interchangeably from both industries as common infrastructure. The proposed architecture can achieve scalability, flexibility to adapt to new service models and high performance at reduced cost.

## ACKNOWLEDGEMENT

This work was partially supported by the FP-7 IDEALIST project under grant agreement number 317999

## REFERENCES

[1] A. Stavdas et el; ECOC 2013, Mo.3.E.4, UK