TELECOM INFRA PROJECT

Distributed Disaggregated Backbone Router - Provider Edge (DDBR - PE)

Technical Requirements Document

- **Eva Rossi**
  - Head of Transport, Vodafone.
  - Eva.Rossi@vodafone.com
- **Jose Angel Perez**
  - Principal Engineer, Vodafone.
  - jose-angel.perez@vodafone.com
- **Ángel Cuadrado**
  - IP Transport Engineer, Vodafone.
  - angel.cuadrado@vodafone.com
- **Kenji Kumaki**
  - General Manager and Chief Architect, Technology Strategy & Planning, KDDI.
  - ke-kumaki @kddi.com
- **Ryuji Matsunaga**
  - Head of Network Strategy, KDDI.
  - ry-matsunaga@kddi.com
- **Ryouhei Fukui**
  - Network Engineer, KDDI.
  - ry-fukui@kddi.com
- **Luc-Fabrice Ndifor**
  - Specialist IP Access & Data Centre, MTN.
  - luc-fabrice.ndifor@mtn.com
- **Luis Miguel Contreras Murillo**
  - Technology Expert, Telefonica.
  - luismiguel.contrerasmurillo@telefonica.com
- **Sarah Cook**
  - Network Engineer, Orange.
  - sarah.cook@orange.com
- **Erdem Cavusoglu**
  - Network Engineer, Meta.
  - ecavusoglu@meta.com

TIP Document License

By using and/or copying this document, or the TIP document from which this statement is linked, you (the licensee) agree that you have read, understood, and will comply with the following terms and conditions:

Permission to copy, display and distribute the contents of this document, or the TIP document from which this statement is linked, in any medium for any purpose and without fee or royalty is hereby granted under the copyrights of TIP and its Contributors, provided that you include the following on ALL copies of the document, or portions thereof, that you use:

1. A link or URL to the original TIP document.

2. The pre-existing copyright notice of the original author, or if it doesn't exist, a notice (hypertext is preferred, but a textual representation is permitted) of the form: "Copyright © 2022, TIP and its Contributors. All rights Reserved"

3. When space permits, inclusion of the full text of this License should be provided. We request that authorship attribution be provided in any software, documents, or other items or products that you create pursuant to the implementation of the contents of this document, or any portion thereof.

No right to create modifications or derivatives of TIP documents is granted pursuant to this License. except as follows: To facilitate implementation of software or specifications that may be the subject of this document, anyone may prepare and distribute derivative works and portions of this document in such implementations, in supporting materials accompanying the implementations, PROVIDED that all such materials include the copyright notice above and this License. HOWEVER, the publication of derivative works of this document for any other purpose is expressly prohibited.

For the avoidance of doubt, Software and Specifications, as those terms are defined in TIP's Organizational Documents (which may be accessed at https://telecominfraproject.com/organizational-documents/ , and components thereof incorporated into the Document are licensed in accordance with the applicable Organizational Document(s).

Disclaimers

THIS DOCUMENT IS PROVIDED "AS IS," AND TIP MAKES NO REPRESENTATIONS OR WARRANTIES, EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, NON-INFRINGEMENT, OR TITLE; THAT THE CONTENTS OF THE DOCUMENT ARE SUITABLE FOR ANY PURPOSE; NOR THAT THE IMPLEMENTATION OF SUCH CONTENTS WILL NOT INFRINGE ANY THIRD PARTY PATENTS, COPYRIGHTS, TRADEMARKS OR OTHER RIGHTS.

TIP WILL NOT BE LIABLE FOR ANY DIRECT, INDIRECT, SPECIAL OR CONSEQUENTIAL DAMAGES ARISING OUT OF ANY USE OF THE DOCUMENT OR THE PERFORMANCE OR IMPLEMENTATION OF THE CONTENTS THEREOF.

The name or trademarks of TIP may NOT be used in advertising or publicity pertaining to this document or its contents without specific, written prior permission. Title to copyright in this document will at all times remain with TIP and its Contributors.

This TIP Document License is based, with permission from the W3C, on the W3C Document License which may be found at https://www.w3.org/Consortium/Legal/2015/doc-license.html

Change Tracking

Date	Revision	Author(s)	Comments
04.05.2023	v1.0	OOPT - DOR Sub-group	Initial Version

Table of Contents

Table of Figures

1

Introduction

1. Introduction

This document represents the technical requirements for a distributed and disaggregated PE router (DDBR-PE) that operators can deploy in current and future networks for the provision of backbone services. It describes the required hardware and proposes non-mutually exclusive software packages for the support of additional services or functionalities.

As part of TIP Open Optical & Packet Transport (OOPT) Project Group, the DOR (Disaggregated Open Router) subgroup formed this initiative with a mission to apply the best practices they've learnt on the DDBR (P/IGW) initiatives on the DDBR-PE routers to accelerate innovation in IP backbone networks Provider Edge routers and to help service providers provide better connectivity for their mobile & broadband customers.

The DOR subgroup members have together analysed the current challenges they face when building and scaling their IP backbone networks and have envisaged an evolution path to their backbone network which introduces innovation, efficiency and primarily openness where they can disaggregate the IP backbone devices and have the flexibility of selecting the best of breed IP products in the market.

This evolution in the IP backbone networks will allow having a full suite of open transport building blocks which can be used across the different segments of their transport networks (access, aggregation & backbone) and attain the benefits that have been achieved with the introduction of different open transport products (ex. DCSG, Cassini, …) into their transport networks.

1.1 Scope of the document

This document defines a proposal for the shift in the IP backbone architecture from monolithic chassis- based to a disaggregated Spine & Leaf architecture. The document also describes the technical aspects of a Distributed Disaggregated Backbone Router – Provider Edge (DDBR-PE) which is a versatile device that can be deployed in IP core/backbone networks as depicted in **Figure-1** below and act as an IP/MPLS provider edge (PE) router.
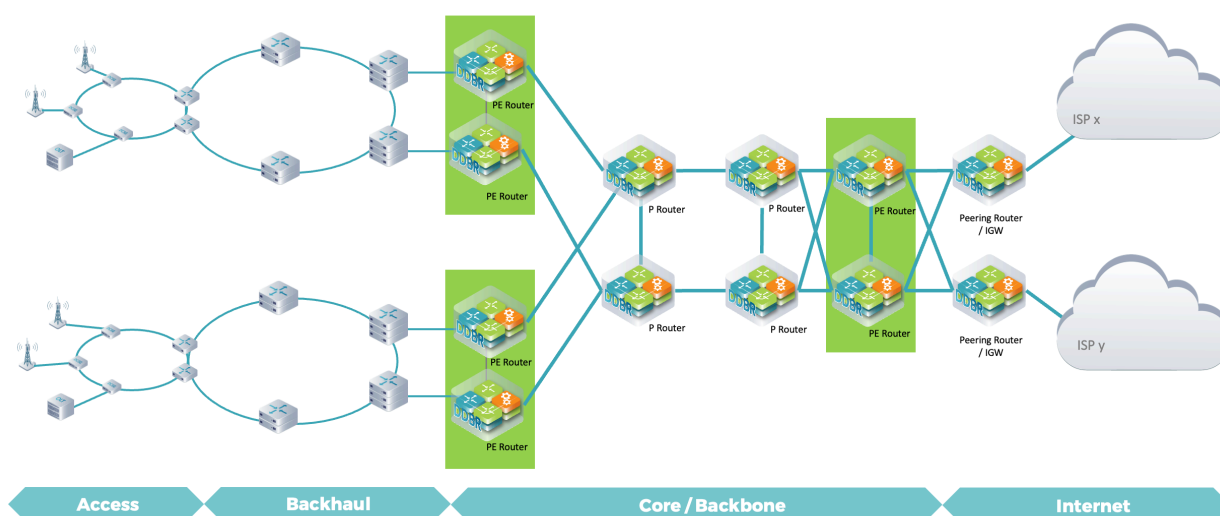
Figure 1. Distributed Disaggregated Backbone Router-PE Scope

The document describes the necessary hardware, software, management and general requirements that need to be fulfilled by a DDBR-PE compliant platform taking into consideration not only the current service providers needs when deploying core/backbone transport networks but also staying ahead of the evolving needs in terms of resiliency, capacity scaling & E2E network automation.

1.2 Document Structure

This document is structured as follows:

- Chapter 1: Introduction
- Chapter 2: DDBR-PE System Architecture & Scaling Path
- Chapter 3: Data Plane Requirements
- Chapter 4: Control Plane Requirements
- Chapter 5: Management Plane Requirements
- Chapter 6: General Requirements
- Chapter 7: Glossary

1.3 Why DDBR-PE?

The objective of this project is to continue the efforts that the DOR subgroup started with the DDBR P and IGW routers for developing a solution that overcomes some of the most relevant issues the service providers are facing when deploying or expanding their IP backbone networks.

1.3.1 DDBR Solution Description

Before diving deep into the current challenges, means to overcome the challenges and the technical requirements details, a high-level description of the key aspects to be considered in the DDBR solution is given below:

- **Disaggregation driving competition:** opening-up the market with new suppliers improving the cost savings
- **No backplane Limitations:** Transport networks need to become smarter and more flexible to meet specific customers' and service's needs and demands.
- **Pay as you grow:** reducing initial investment and optimizing the power efficiency without any growth limitation
- **Innovation:** open SW and HW to improve flexibility and innovation on SW development and reduce time to market.
- **Operational Efficiency:** Taking advantage of Software Defining Network to make the network operation simpler, give tools for automation, enhance the capabilities of our network, and introduce a set of capabilities that today are not present.
- **Reliability:** Always targeting higher availability & multi-level redundancy

1.3.2 Overview of IP backbone network challenges

The role of IP backbone networks is to route the mobile & broadband traffic between different access networks at a national or regional level within the service provider network in addition to providing connectivity with external networks ex. other service providers, public cloud providers, content data networks, Internet exchange peers and IP transit providers exchange traffic and access the internet.

New services (ex. IaaS, PaaS, SaaS, Metaverse, immersive AR/VR, cloud gaming,

eMBB, mMTC and URLLC etc.) are driving the increased bandwidth, connectivity, uptime, and latency demands and are dictating substantial transformations on the E2E network architecture and economics to cope with these new services requirements.

The IP backbone network in turn must continuously scale to support the internet traffic growth, to improve resiliency in order to meet the mission-critical type of communications and to create an evolution path for agility and automation to lower the network cost and enhance the overall customer experience.

In the following section, we will list the key challenges that currently exist in the IP backbone networking space and different proposals to address these challenges in the DOR.

1.3.2.1 Lack of Supply Chain Diversity

Like all segments of the telecom network (Radio Access, Transport & Core), the ongoing consolidations & acquisitions in the core routers supply market have led to:

- High dependency on a reduced number of suppliers.

- Less competitive market which is at greater risk from increasing costs

- Limited Innovation and time-to-market speed.

- Limited 3rd party interoperability across different hardware component

1.3.2.2 Monolithic nature

Traditionally the service providers are deploying monolithic IP Backbone routers which are based on vertically integrated proprietary components as modelled in **Figure-2** below which are a bottom-up tightly coupled.
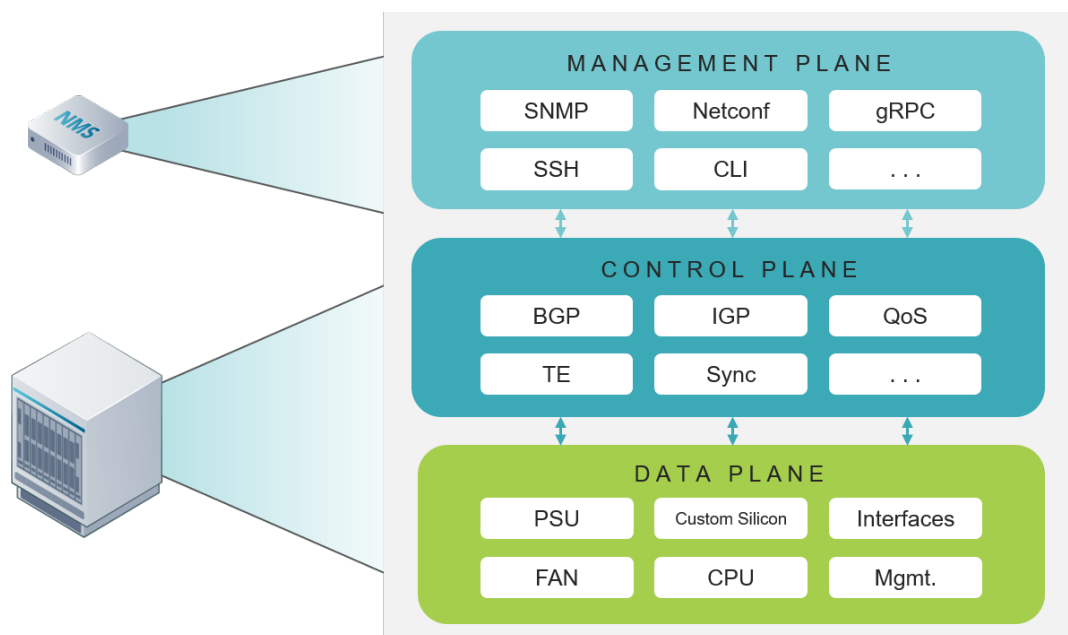
Figure 2. Monolithic IP Backbone Routers

Starting from the **data plane** hardware where custom purpose silicon chipsets are used to handle the packet processing, traffic management, forwarding and expose the fabric interfaces.

In the **control plane**, a custom networking operating system (NOS) which is purposely built to optimally run only on the custom hardware and consists proprietary code and mandatory licensing. That spans the drivers which control all the hardware components power, cooling, routing circuits, etc. The firmware is responsible for loading the networking operating system image when the router boots and the networking software stack which executes the routing protocols & different policies.

The third component is the **management plane**, which is responsible for the overall platform management for instance the interfaces configuration, services provisioning, inventory management, alarm reporting, faults handling, performance monitoring.

It worth noting that some suppliers have started to build platforms based on merchant silicon chips and port on this hardware on their proprietary NOS software, however the fact of having the hardware tied only to this proprietary NOS software still imposes on the service providers the same challenges of the monolithic nature.

These monolithic IP backbone routers have served the key performance needs (capacity, availability, etc..) and proven to be fit for purpose for a long time. However, having the platform vertically integrated impeded the service providers from unleashing the potential of open networking, limited the supply choices and generally slowed down the innovation in the IP backbone networks.

For instance, having the data plane tightly coupled with the control plane lead to a high dependency on the existing supplier's roadmap and dictated the need to completely replace the hardware plus the NOS software to benefit from a feature available in a 3rd party NOS supplier.

Additionally, for the management plane while several suppliers have offered robust Network Management Systems (NMS) and SDN (Software Defined Network) controllers to efficiently manage their monolithic products and provided support for third party products still the services providers are experiencing challenges when it comes to the complex & high integration cost to manage third party products via the existing NMS. Also, while several suppliers have implemented the promising Netconf (RFC 6241) protocol there is very limited progress in standard network data model definition and there is lot of efforts needed to create a vendor-neutral data model to describe the network or device configuration.

1.3.2.3 Chassis based

The existing IP backbone routers are predominantly built based on a chassis structure with front access where the Interfaces (NNI/UNI) cards and the control boards are plugged into a common backplane.

Given the critical role of the IP backbone routers in the network and the high volume of traffic it is carrying, this chassis had to offer:

- Extreme resiliency and robustness to maintain nonstop connectivity to the dependent mobile and broadband customers
- Powerful computing capability and routing table storage for Ipv4/Ipv6 prefixes at the size of internet
- High capacity to support the customers and services growth.

Consequently, the chassis had to be equipped with:

- Units' redundancy at all levels: control processing, switching fabric, cooling and power.
- Powerful computing capability, deeper buffers and large TCAM (Ternary Content Addressable Memory) chips
- Wide range of advanced control plane features including NSR (Non-Stop-Routing) & ISSU (In-Service Software Upgrade)
- High port density, number of slots and backplane switching capacity

That resulted in high cost, power hungry & bulky size platforms which require upfront investments, reserve larger footprint in datacenters and consume high power and does not provide a model to optimally grow based on the needs.

Moreover, in case of running out of slots a completely new chassis will be needed to add a single extra interface which leads to a complicated network topology, suboptimal traffic flows a non-linear cost per port model. Also, this upgrade model is not flexible enough to allow the service providers to promptly react to unplanned upgrade requests which leads to missing the opportunity to connect new customers due to the lengthy upgrade process.

Additionally, having all the NNI & UNI interfaces centralized into the same chassis is imposing an operational risk of losing the entire node in the case of a software glitch, undesirable state propagation due to misconfiguration, power issue, executing the wrong Method of Procedure (MoP) when rebooting the device or activating new link.

1.3.2.4 Time to upgrade the installed base

With the industry shift happening in the optical pluggable transceivers and the dawn of 400G QSFP56-DD optics, the services providers need to replace the current installed base to higher capacity, more compact dimensions, flexible thermal management ports which enable supporting higher capacity links with optimum port density per RU.

In addition to the 10G/25G/40G UNI interfaces, 100G and 400G UNI and NNI interfaces need to be supported to get the outmost of the interface capacity through the backplane bus.

Accordingly, the service providers would look for a new approach when replacing the installed base to protect their investment in IP backbone networks.

2

DDBR-PE System Architecture & Scaling Path

2. DDBR System Architecture & Scaling Path

In this section, we will describe the high-level architecture of the proposed changes to the IP backbone router and the envisioned path to efficiently scale-out the capacity and the number of interfaces based on the need for the upgrade

2.1 Scale-up in Traditional IP backbone systems

As stated earlier, currently the IP backbone routers are predominantly based on integrated modular chassis designs. Accordingly, the upgrade path for a traditional IP backbone system is achieved as shown in **Figure-3** via adding extra line cards which is a scale-up model and that go till all the available slots are consumed. However, with this scale-up model provides extra interfaces while the switching capacity and the routing engine processing & memory doesn't scale with it as its limited with the chassis fabric chips capabilities which develops to become a bottleneck for the control plane performance and the overall supported system capacity and the upgrade path for a traditional IP backbone system is to completely replace the chassis as illustrated in Figure-3 with a more powerful switching , processing and memory chips to support the anticipated growth.

Figure 3. Traditional scale-up path

2.2 Transformation to Disaggregated Spine & Leaf Architecture

One step towards the scale-out path is to move to Spine & Leaf based architecture and disaggregate the control plane from the data plane as illustrated in **Figure-4** which instantly solves the dependency on the router switching capacity and the number of interfaces.
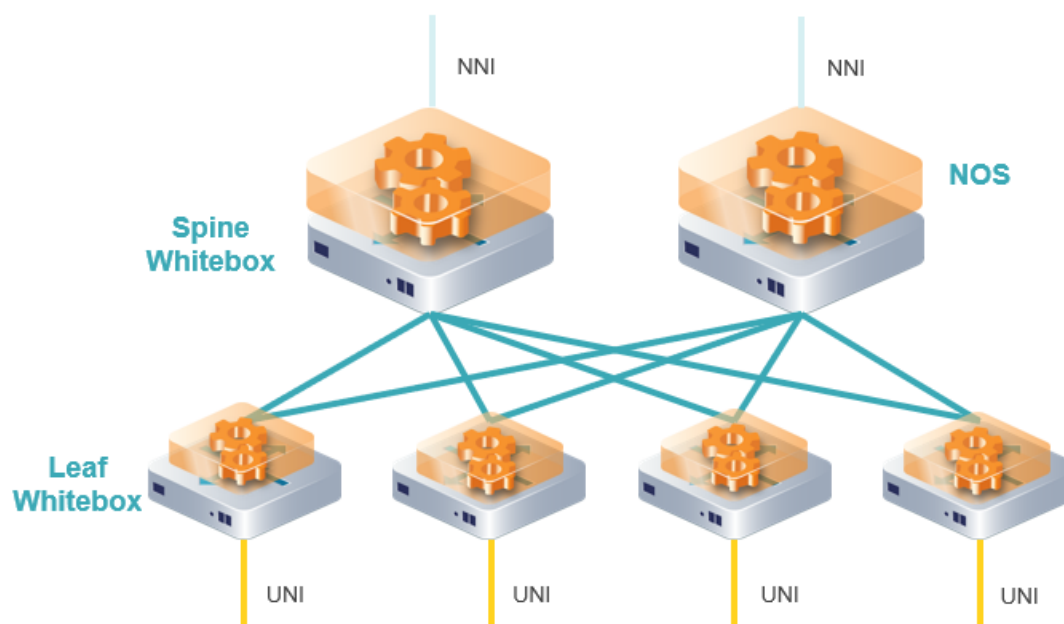
Figure 4. Disaggregated Spine & Leaf Architecture

The Spine & Leaf based architecture has been adopted in the datacenter space for a while and have helped the web-scale companies to efficiently grow their infrastructure to a massive scale that can deal with big data sets like for instance replicating copies of the internet across geographically distributed datacenters.

The combination of Spine & Leaf based architecture and disaggregation can lead to tremendous gains, including but not limited to:

- **Disjointed innovation paths**: between the data plane and control plane and allow replacing the NOS software while re-using same hardware and even mixing hardware among spine and leaf layers from multiple suppliers
- **Moving away from costly platforms**: benefiting from the low-cost merchant silicon
- **Scaling Efficiently** in a Pay-As-You-Grow model: start small and increment 1-

2U at a time based on the growth needs while capitalizing on the investment made on the initial whiteboxes

- **Increased agility in physical deployment**: with less risks compared to entirely migrating the service to a new chassis in the scale-up model, upgrades in Spine & Leaf-based architecture is imposing less operational risk for instance you just need to connect a new spine router to increase the overall system capacity and connect a Leaf router to increase the number of interfaces
- **Deterministic latency**: with a fixed number of hops between spine and leaf routers, the overall system latency (from ingress to egress) is deterministic and homogenous across all ports
- **High performance**: High capacity can be achieved with variable over-subscription 1:1 to N:1 where the total bandwidth connected to the leaf routers can surpass the overall capacity offered by the spine routers
- **High Availability**: Path redundancy with multipath thanks to ECMP (Equal Cost Multi Path)
  Relatively smaller failure domain in the network (e.g. In case of SW issue in CP, SW reboot of one node in a CLOS Topology instead of turning a full chassis down)
- **Foster competition:** disaggregation will lead to more HW & SW companies competing to build more innovative & agile solutions.

2.3 Achieving Distributed Disaggregated Cluster

A second step to get the outmost benefits of the scaling-out is to move onto a centralized control plane as shown in **Figure-5** running on an on-prem x86 server or as a container in a cloud-native fashion which does all the heavy lifting of the route computation algorithms while keeping a lighter software version on the whiteboxes CPUs for running the initial booting configuration, sending keep alive messages or responding to the control messages received from the x86 server or VM hosting the control plane.

Because of the control plane virtualization, below additional benefits will be achieved:

1. A high scale routing solution made of a cluster of white boxes managed by a

centralized control plane

2. Ability of leveraging the value of Private/Public cloud computing to provide better compute scalability through virtualization or cloud bursting and benefiting from open-source solutions like OpenStack.

3. The cluster is acting as a single network entity, which accordingly save the:

   a. IP addresses needs to assign IP addresses on both Spine and Leaf in CLOS based architecture

   b. Cost as no need to use a special optics between Spine and Leaf because of a cell-based packet

4. Advanced QoS handling, deep buffers, large TCAM
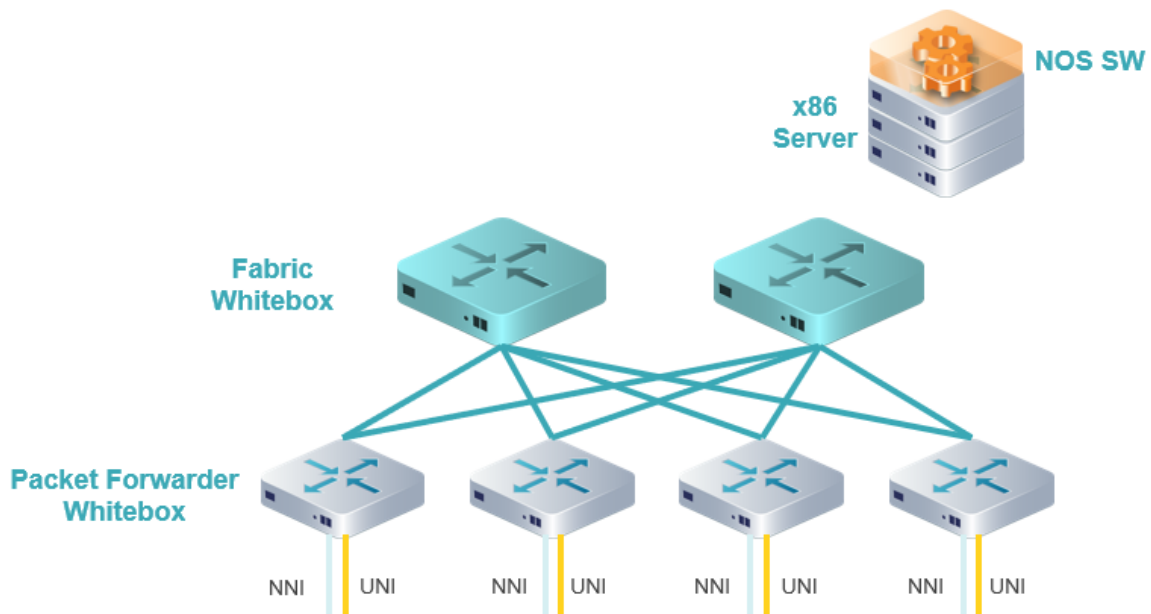
Figure 5. Distributed Disaggregated Cluster

Also, this virtualization step removes the bottleneck in the chassis-based model which was the limitation of the chassis built in CPU. Now processing can scale with servers of VM addition in a fully modular approach. It will also open the door for service chaining by adding multiple functions on the same hardware accordingly the router is not anymore, a specialized appliance but it can run multiple VNF

3

Data Plane Requirements

3. Data Plane Requirements

As shown in section-2 the platform architecture is modular and is basically a cluster of routers which consists of the following two building blocks:

- **Fabric Whitebox:** which represents the spine of the cluster and acts as the backplane
- **Packet Forwarder Whitebox:** which represents the leaf of the cluster and acts as the line cards

The whiteboxes shall be based on a 64-bit x86 architecture. At the time of this writing, this is the CPU architecture that has the most robust developer ecosystem and the most appealing roadmap to ensure platform longevity.

The forwarding capacity of the ASIC used in the whiteboxes shall support line-rate forwarding across all ports without any limitation and provide at least 4Tb/s full duplex capacity which will form the basic building block or smallest cluster size to start with,

The overall cluster capacity is the result of staging Fabric & Packet Forwarder white boxes in clos-based interconnection, The reference examples of different cluster sizes which are initially thought by the operators to be fulfilled by a DDBR-PE platform which should fit for their current needs are shown in Figure 6 in section 4.2.10, however a DDBR-PE platform shall offer seamless intra & inter cluster scaling and shall not prevent achieving higher cluster sizes with further staged whiteboxes.

There are common aspects between both the Fabric & Packer Forwarder whiteboxes, for instance the architecture, form factor, power supply, mechanical and environmental requirements while the key differences will be the merchant silicon chip which defines the type and number of interfaces that the platform will support in addition to the CPU, deep buffer, TCAM, buffer size.

3.1 Whitebox requirements

The Fabric & Packet Forwarder whiteboxes need to support Ethernet switching, extensible deep buffering with carrier-grade traffic management and subscriber-level virtual output queueing should be provided and support below different port types:

- 400GbE

- 100GbE
- 40GbE
- 10GbE / 25 GbE via native or breakout cabling solutions
- Multi-rate 1/10 GbE

DAC and AOC cable types should be supported.

Concerning the Transceiver Optics types, below types shall be considered:

- **For 400G interfaces:**
    - Support all relevant IEEE standards (IEEE 802.3bs)
    - Support all requirements concurrently while forwarding at line rate, in all conditions (independently from number/types of services/flows) on multiple port line cards each port is expected to run at line rate concurrently.
    - Support all physical interface connectors must be IETF compliant, the vendor specific interface connectors are acceptable if the majority of the 400G interfaces are also IETF compliant.
    - Support for 400GBASE-FR8
    - Support for 400GBASE-FR4
    - Support for 400GBASE-LR8
    - Support for 400GBASE-LR4
    - Support for 400GBASE-ZR
    - Support for 400GBASE-ZR+
    - Support for 400GBASE-FR8 and FR4 with QSFP56-DD
    - Support for 400GBASE-LR8 and LR4 with QSFP56-DD
    - Support for 400GBASE-ZR with QSFP56-DD
    - Support for 400GBASE-ZR+ with QSFP56-DD
    - Support for 400GBASE-FR8 and FR-4 with 3rd party QSFP56-DD
    - Support for 400GBASE-LR8 and LR-4 with 3rd party QSFP56-DD
    - Synchronous Ethernet according to G.8261 G.8262, G8264 for all types of interfaces
    - Multi rate slot, please specify details (QSFP DD/56/28)

- **For 100G interfaces:**
    - Support all relevant IEEE standards (IEEE 802.3ba)
    - Support all requirements concurrently while forwarding at line rate.
    - On multiple port line cards each port is expected to run at line rate concurrently

- o Support all physical interface connectors must be IETF compliant
- o Support for 100GBASE-SR10
- o Support for 100GBASE-LR4
- o Support for 100GBASE-ER4
- o Support for 100GBASE-SR10 with QSFP28.
- o Support for 100GBASE-LR4 with QSFP28.
- o Support for 100GBASE-ER4 with QSFP28.
- o Support for 100GBASE-SR10 with 3rd party QSFP28
- o Support for 100GBASE-LR4 with 3rd party QSFP28
- o Support for 100GBASE-ER4 with 3rd party QSFP28
- o Support for 100G-Base-CWDM4
- o Synchronous Ethernet according to G.8261 G.8262, G8264 for all type of interfaces

- **For 40G interfaces:**

  - o Support all relevant IEEE standards (IEEE 802.3ba)
  - o Support all requirements concurrently while forwarding at line rate.
  - o On multiple port line cards each port is expected to run at line rate concurrently
  - o Support all physical interface connectors must be IETF compliant
  - o Support for 40GBASE-SR4
  - o Support for 40GBASE-LR4
  - o Support for 40GBASE-ER4
  - o Support for 40GBASE-SR4 with QSFP+.
  - o Support for 40GBASE-LR4 with QSFP+.
  - o Support for 40GBASE-ER4 with QSFP+.
  - o Support for 40GBASE-SR4 with 3rd party QSFP+
  - o Support for 40GBASE-LR4 with 3rd party QSFP+
  - o Support for 40GBASE-ER4 with 3rd party QSFP+
  - o Support for 40G-Base-CWDM4
  - o Synchronous Ethernet according to G.8261 G.8262, G8264 for all type of interfaces

- **For 25G interfaces:**

  - o Support all relevant IEEE standards (IEEE 802.3by)
  - o Support all requirements concurrently while forwarding at line rate.
  - o On multiple port line cards each port is expected to run at line rate

concurrently
- o Support all physical interface connectors must be IETF compliant
- o Support for 25GBASE-SR
- o Support for 25GBASE-LR
- o Support for 25GBASE-ER
- o Support for 25GBASE-SR with SFP28.
- o Support for 25GBASE-LR with SFP28.
- o Support for 25GBASE-ER with SFP28.
- o Support for 25GBASE-SR with 3rd party SFP28
- o Support for 25GBASE-LR with 3rd party SFP28
- o Support for 25GBASE-ER with 3rd party SFP28
- o Support for 25G-Base-CWDM SFP28
- o Synchronous Ethernet according to G.8261 G.8262, G8264 for all type of interfaces

- **For 10G/1G interfaces:**

  - o Support all relevant IEEE standards (IEEE 802.3ae)
  - o Support all requirements concurrently while forwarding at line rate.
  - o On multiple port line cards each port is expected to run at line rate concurrently
  - o Support all physical interface connectors must be IETF compliant
  - o Support for 10GBASE-SR
  - o Support for 10GBASE-LR
  - o Support for 10GBASE-ER
  - o Support for 10GBASE-SR with SFP+.
  - o Support for 10GBASE-LR with SFP+.
  - o Support for 10GBASE-ER with SFP+.
  - o Support for 10GBASE-SR with 3rd party SFP+
  - o Support for 10GBASE-LR with 3rd party SFP+
  - o Support for 10GBASE-ER with 3rd party SFP+
  - o Support for 10G-Base-ZR SFP+
  - o Support for 10G-Base-CWDM SFP+
  - o Support for 10G-Base-T SFP+
  - o Support for 1/10GBASE-LX
  - o Support for 1/10GBASE-EX

- o Support for 1/10GBASE-ZX
- o Support for 1/10GBASE-LX with 3rd party SFP
- o Support for 1/10GBASE-EX with 3rd party SFP
- o Support for 1/10GBASE-ZX with 3rd party SFP
- o Synchronous Ethernet according to G.8261 G.8262, G8264 for all type of interfaces

The platform shall be fully interoperable with any 3rd-party pluggable optics, with no impact on customer/content traffic and equipment capabilities (payload, header, QoS, synchronization, etc.), able to report SFP data via Digital Diagnostic Monitoring, and support SFP dynamic data logging.

Additionally, the merchant silicon chip shall support the typical Telco protocols including but not limited to:

- IEEE 1588v2 & SyncE for frequency and phase
- L2
- OSPF & ISIS
- MPLS
- SR
- VPLS
- L2VPNs
- L3VPNs
- QoS & HQOS
- OAM
- EVPN

3.2 ASIC firmware requirements

When possible, network operating systems for this platform should be provided in the form of binary installers compatible with the Open Network Install Environment (ONIE) specification, as defined by the Open Compute Project (OCP).

4

Control Plane Requirements

4. Control Plane Requirements

4.1 NOS SW architecture

The DOR is being thought as a modular box, that can run any SW on top of the selected HW versions. In order to ensure the maximum flexibility in terms of the SW that can be loaded in the DOR, it will be equipped with ONIE. ONIE will enable any operating system to run on top of the DOR.

ONIE defines an open source "install environment" that runs on routers and switches subsystem. This environment allows end SW suppliers to install the target NOS as part of the initial system setup.

In order to enable different software packages, the DDBR solution shall rely on "trusted based" systems so there is no need to have a licensing server or internet connectivity. In most of the cases it's expected that the SDN controller or the management systems will be able to activate the different software packages based on the operator request.

The control plane flavors for the DOR shall support the PE-router functionalities. In below section we will be listing the minimum set of software features that need to be supported for both targeted functionalities. There should not be any limitation to support a customized NOS which introduces further functionalities ex. Load balancer, Firewall.

4.2 DDBR SW features

4.2.1 Interface support

- HW based BFD
- Flapping protection (carrier delay/interface dampening)
- LAG & LACP
- VLAN

4.2.2 Routing support

- IPv4/IPv6 static route
- OSPFv2/v3, including extensions for Traffic Engineering, BFD, Graceful Restart
- IS-IS, including extensions for Traffic Engineering, BFD, Graceful Restart
- BGP-4, including multiprotocol extensions (RFC4760), capabilities advertisement (RFC5492), communities (RFC1997), BGP-LU (RFC3107), deterministic-med (RFC 4721), graceful restart / non-stop Forwarding (RFC 4724), extensions for 4-byte AS number (RFC4893), confederations (RFC3065), route reflection (RFC4456), error-handling (RFC 7606), peer tracking (RFC 7854) and prefix-Independent Convergence (PIC).
- LDP including T-LDP, LDP over RSVP and LDP IGP Synchronization
- RSVP-TE, with the capability of path computation based on CSPF and supporting FRR mechanisms.
- ECMP
- Entropy-label-based load balancing
- VRRP
- IP LFA FRR mechanism
- TI-LFA
- SR MPLS
- SRv6
- PCEP
- BGP-LS
- Multicast capabilities, (PIM, IGMP) both for IPv4 and IPv6, for L3VPNs outlined in RFC6514 as the preferred option
- Policy Based Routing (this will be described further below)
- PWE3

4.2.3 Quality of Service (QoS) support

The DDBR platform shall generally handle all QoS functions like traffic classification, marking, metering, shaping & scheduling and particularly be able to:

- Classify incoming customer packets or frames into QoS classes based on 802.1p, DSCP or by static value that will follow the packet/frame through the platform
- Police customer packets/frames within classes using a dual rate policer to drop packets or to set a QoS class profile that will follow the packet/frame

through the platform

- Police all incoming customer traffic within individual services or the total traffic in a group of services, which can be configured on a per-service basis
- Where a service uses only a single QoS class, police both the service and service group simultaneously. The service policer shall also be able to set the QoS class profile.
- Set egress MPLS EXP markings based on a configurable mapping from customer QoS class & QoS class profile.
- Schedule MPLS packets using at least 6 queues with at least 2 being capable of expedited behavior and all being capable of having an assured bandwidth or ratio.
- Operate at least two WRED profiles with MPLS queues based on a packet/frame's QoS class profile.
- Classify incoming MPLS packet's class & class profile either from the MPLS EXP markings or, where the packet is destined for a connected customer access circuit, by the IP packet's DSCP markings. Both operations shall be possible simultaneously on the platform with the behavior specified on a per-service or service-group basis.
- Deploy up to 8 egress queues for QoS classes on each service with up to 2 being expedited queues and all queues being capable of assured bandwidth.
- Deploy up to 8 egress queues for QoS classes, each one of them should be independently configured as priority queue, weighted fair queueing or best effort.
- Police traffic on egress class queues and have at least 2 WRED levels (preferred 3) within service egress assured classes
- Shape service egress traffic on a per-service or total of a service group, where there is no policing within the classes, the platform should be able to egress shape on both service & service-group basis simultaneously and the egress shape service egress traffic on a per-service and per-service group basis simultaneously.
- Shape all traffic leaving a customer service ports.
- Set 802.1p COS based on QoS class
- Where the platform has multiple control plane elements, the platform shall ensure traffic forwarding continues uninterrupted throughout any failover between them. Where the platform's makes requirements of other platforms to achieve this, they should be clearly stated.
- Where the platform has multiple control plane elements, the platform should reestablish control plane signaling with other devices in a manner that results

in no impact to traffic forwarding.

- The Platform shall support Avoidance of Head of Line Blocking (HOLB), as follows:
  - o Once off the ingress line and into the buffer, no HOLB shall occur. The switch fabric must give the same consideration to QoS as the line cards.
  - o Packets of a higher priority must be given access to the fabric without being blocked by larger packets of a lower priority.
- For the QoS marking function:
  - o Marking of Platform self-generated control traffic: The Platform shall allow the IP Precedence of the following IP packets to be configured independently of each other by the operator: LDP, RSVP-TE, BGP, OSPF (all versions), ISIS, VRRP, PCEP, NETCONF, TLS & TELEMETRY
  - o QoS marking function - self-generated OAM traffic
  - o The Platform shall allow the IP Precedence of the following IP packets to be independently configured by the operator: SSH including SCP, SNMP read (all SNMP versions), SNMP traps (all SNMP versions), SYSLOG, NTP, Flow-record transport (e.g., IPFIX WG), FTP, TFTP, PCEP, NETCONF, TLS & TELEMETRY
- MPLS-Diffserv tunneling modes:
  - o Long Pipe Model of MPLS-DiffServ tunneling (in accordance with RFC3270)
  - o Uniform Model of MPLS-DiffServ tunneling (in accordance with RFC3270)
  - o Short Pipe Model of MPLS-DiffServ tunneling (in accordance with RFC3270)
  - o The Platform shall be configurable to set each of the three modes on a per egress sub-interface basis. i.e., it may be required to set Uniform Model on one sub-interface and Short Pipe Model on another sub-interface of the same common physical interface.
  - o Pipe Model, Uniform Model and Short Pipe Model shall be supported on all interface types.

4.2.4 Hierarchical Quality of Service (H-QoS) support

The DDBR platform shall support the following, in addition to the essential QOS functions mentioned in previous section:

- The platform shall support multiple hierarchical levels and the preferably 5 queues within each of the levels

- Each hierarchical level shall support as a minimum, a Low Latency Queue (LLQ)
- Each hierarchical level shall support 4 weighted round robin queues as a minimum
- Each hierarchical level shall support policing & shaping
- Simultaneous traffic shaping at VLAN, VLAN group and Port Level (access ports)
- Traffic shaping at Port Level (NNI)
- Traffic shaping at VLAN Level (NNI)
- Traffic shaping at VLAN Group Level (NNI). Please provide number of VLANs per group are supported.
- It must be possible to apply H-QoS in a pseudowire termination architecture (where a pseudowire is terminated into a VRF)
- It must be possible to apply a single H-QoS policy to both L2 and L3 services simultaneously.

4.2.5 Security support

In general, its essential that the NOS SW performs all remote operation and maintenance tasks via encrypted protocols (e.g.SSH, SSL, TLS/DTLS).

Additionally, below features need to be supported:

- ISIS MD5, BGP MD5
- BGP FlowSpec
- Compliance with IETF RFC5575 & RFC7674
- IEEE802.1x/EAP-TLS
- MACSEC

4.2.6 Services Support

- L2VPN
  - E-LAN (MEF – EVC Ethernet Services Definitions)
  - E-LINE (MEF – EVC Ethernet Services Definitions)
  - E-TREE (MEF – EVC Ethernet Services Definitions)
  - VPLS (RFC 4761, RFC 4762)
  - H-VPLS (RFC 4762, RFC 7080)
  - VPWS
- E-VPN
- L3VPN (RFC 4364, RFC 4659, RFC 6368)
  - IPv4 L3VPN
  - IPv6 L3VPN
  - CE-PE IPv4 L3VPN eBGP

- o CE-PE IPv6 L3VPN eBGP
- o CE-PE IPv6 L3VPN iBGP
- 6PE (RFC4798)

4.2.7 Performance & Management Support

- Netconf/Telnet/SSH
- Ping/Traceroute
- IP-SLA
- SNMP/Telemetry
- TWAMP
- TACACS+
- RADIUS
- DHCPv4 / DHCPv6
- NetFlow
- EEM

4.2.8 SDN

- PCEP
- NETCONF
- BGP LS
- Telemetry / gNMI, gRPC

4.2.9 Scalability Figures

As a reference, in the upcoming sections we will share a reference for the scalability figures which shall be supported by the DDBR platform

4.2.10 Hardware Scalability:

Given that the hardware parameters and scalability figures will depend mainly on the number of whiteboxes used and consequently the cluster size. We are presenting the scalability requirements for the commonly used cluster sizes (four scaling sizes and the standalone forwarding box) for the PE-router in Figure-6

Item	Standalone	Small Cluster	Medium Cluster	Large Cluster	Extra Large Cluster
Size (RU)	2 RU	16 RU	26 RU	66 RU	126 RU
Min. Bidirectional tput (Tbps), value for FULL DUPLEX *	4Tb	16Tb	32Tb	96Tb	192Tb
Port Density (Per Packet Forwarder Whitebox) * (Native or via Breakout cables for 25G and 10G interfaces) (optional 40G support)	10x400G	40x400G	80x400G	240x400G	480x400G
	40x100G	160x100G	320x100G	960x100G	1920x100G
	40x40G	160x40G	320x40G	960x40G	1920x40G
	160x25G	640x25G	1280x25G	3840x25G	7680x25G
	160x10G	640x10G	1280x10G	3840x10G	7680x10G
Packet Forwarder Whitebox**	1	4	8	24	48
Fabric Whitebox**	-	2	3	7	13
Min number of IPv4 FIB entries	2M	2M	2M	2M	2M
Min number of IPv6 FIB entries	1M	1M	1M	1M	1M

* The minimum HW configuration to get the required value.

**These values are only a reference and not a mandatory figure. Number of ports and throughput capacity requirements are discussed below.

Figure 6. Hardware scalability figures for PE-router

The requirements below shall be met when building the different cluster sizes except for the standalone configuration:

- **Scalability**: the system/configuration may be strictly adjusted and optimized to required switching and min number of interfaces per Cluster Size, and not smaller or much larger than it. E.g., Bidirectional throughput for Medium Cluster should be met and equivalent to 32TB, and not smaller or much larger than it.
- **Non-blocking**: the system/configuration shall provide a non-blocking solution to process full duplex (Tx+Rx) traffic line rate on each port including down/uplinks
- **No Single Point of Failure**: the system/configuration shall have no single point of failure (e.g., more than one spine and leaf white box per configuration)
- **Oversubscription**: an oversubscription ratio (FD Fabric / Interfaces Capacity)

higher than 100% to be non-blocking and process full duplex (Tx+Rx) traffic line rate on each port shall be provided.

4.2.11 PE-router SW Scalability:

Item	Small Cluster	Medium Cluster	Large Cluster	Extra Large Cluster
Min. number of OSPF adjacencies	500	500	500	500
Min. number of OSPF LSAs (Intra/Inter/Ext)	6000/1500/1500	6000/1500/1500	6000/1500/1500	6000/1500/1500
Min. number of prefixes per Global Routing Table	2M	2M	2M	2M
Min. number of BGP peers	210	210	210	210
Min. number of BGP prefixes	2M	2M	2M	2M
Min. number of PIM neighbors	500	500	500	500
Min. number of entries for PIM(*,G)/(S,G)	500/100	500/100	500/100	500/100
Min. number of LSP (LER/LSR) per system	1000/6000	1000/6000	1000/6000	1000/6000
Min. number LDP FECs	500	500	500	500
Min. number of T-LDP Sessions	150	150	150	150
Min. number of T-LDP Routes	9000	9000	9000	9000
Min. SR IPv4 label stack depth without recirculation	8	8	8	8
Min. SR IPv4 label stack depth	10	10	10	10
Min. number of OSPFv3 adjacencies	500	500	500	500
Min. number of OSPFv3 LSAs (Intra/Ext)	500/10	500/10	500/10	500/10
Min number of VRF instances per system	3000	3000	3000	3000
Min. no. of IPv6 prefixes per Global Routing Table	1M	1M	1M	1M
Min. number of BGP peers for IPv6	260	260	260	260
Min. number of BGP IPv6 prefixes	1M	1M	1M	1M
Min number of IS-IS adjacencies	1000	1000	1000	1000
Min number of IS-IS instances	5	5	5	5
Min number devices per IS-IS area	5000	5000	5000	5000
Min number of ACLs	16000	16000	16000	16000
Min size of RIB for IPv4	2M	2M	2M	2M
Min size of FIB for IPv4	2M	2M	2M	2M
Min size of RIB for IPv6	1M	1M	1M	1M
Min size of FIB for IPV6	1M	1M	1M	1M
Min number of EVPN instances per system	10000	10000	10000	10000
Min number of VPLS instances	10000	10000	10000	10000
Min number of MAC addresses per EVPN instance	0.5M	0.5M	0.5M	0.5M
Min number of VRF for IPv4 per system	10000	10000	10000	10000
Min number of VRF for IPv6 per system	5000	5000	5000	5000

Figure 7. PE-router Software scalability figures

4.2.12 Frequency & Time Synchronization Support*

The DDBR-PE shall be able to provide frequency and time synchronization to the mobile and fixed access networks. Additionally, the DDBR-PE shall be able receive the time sync signal (e.g. from another router in the network) and propagate it to other network elements.

The DDBR-PE platform shall support the following time sync requirements:

- Network quality model (microsecond precision) according to ITU-T G.8271.1 Network conditions and reference model where the Boundary Clock is requested to work.
- Node performance (noise generation, tolerance, transfer, and holdover) according to ITU- T G.8273.2 Section 7.1/7.2/7.3/7.4.
- Node performance (upon wander, failure, holdover) according to ITU-T G.8273.2 (7.2/7.3/annex) Boundary clock quality objectives in holdover mode
- Interoperability based on IEEE1588 profile defined in ITU-T G.8275.1 with Boundary Clock and SyncE support for holdover purposes and Grandmaster redundant sources support. This requirement corresponds to the support of IEEE1588v2 profile for telecoms – Precision Time Protocol and includes SyncE support in Ethernet interfaces as mandatory as per ITU-T G.8262 and G.8264.

  It will be also interesting to receive a detailed information on the electronics used to build up the phase sync regeneration capability (type of oscillator, type of Best Master Clock Algorithm, whether its implemented via SW or HW, etc.).

  The DDBR-PE shall also support a SFP input for GPS. The platform shall be interoperable with most of the SFP providers' solutions in the market in order to avoid interoperability issues. This GPS input shall be used in scenarios where the time sync signal cannot be received from another network element through a standard Ethernet traffic port in case time sync distribution mechanisms are used.

  The DDBR-PE shall provide a 1PPS in/out external sync interface.

  *Frequency and Time synchronization support is an optional requirement.

5

Management Plane Requirements

5. Management Plane Requirements

5.1 Telemetry & SDN readiness

As Network operators are moving away from the CLI and towards network programmability, DDBR device being a key part of the Transport network shall conform with the Open Transport Architecture represented in **Figure-8**. That network programmability can be achieved by employing a hybrid SDN hierarchical architecture, in which the management and control functionalities are split between the devices and the controller.

The main goals of such SDN solutions are:

- Agile Network Programmability, enabling full network automation and reduced time-to-market service creation.
- Network Abstraction, simplifying Operation Support Systems (OSS) and orchestrators, and their interactions, by performing the adequate level of abstraction at each layer.
- Network Intelligence, enabling Traffic Engineering (TE) and automated service provisioning mechanisms between different layers and different vendor technologies.
- Compliancy with the data models and protocols defined under the MUST specifications documents.
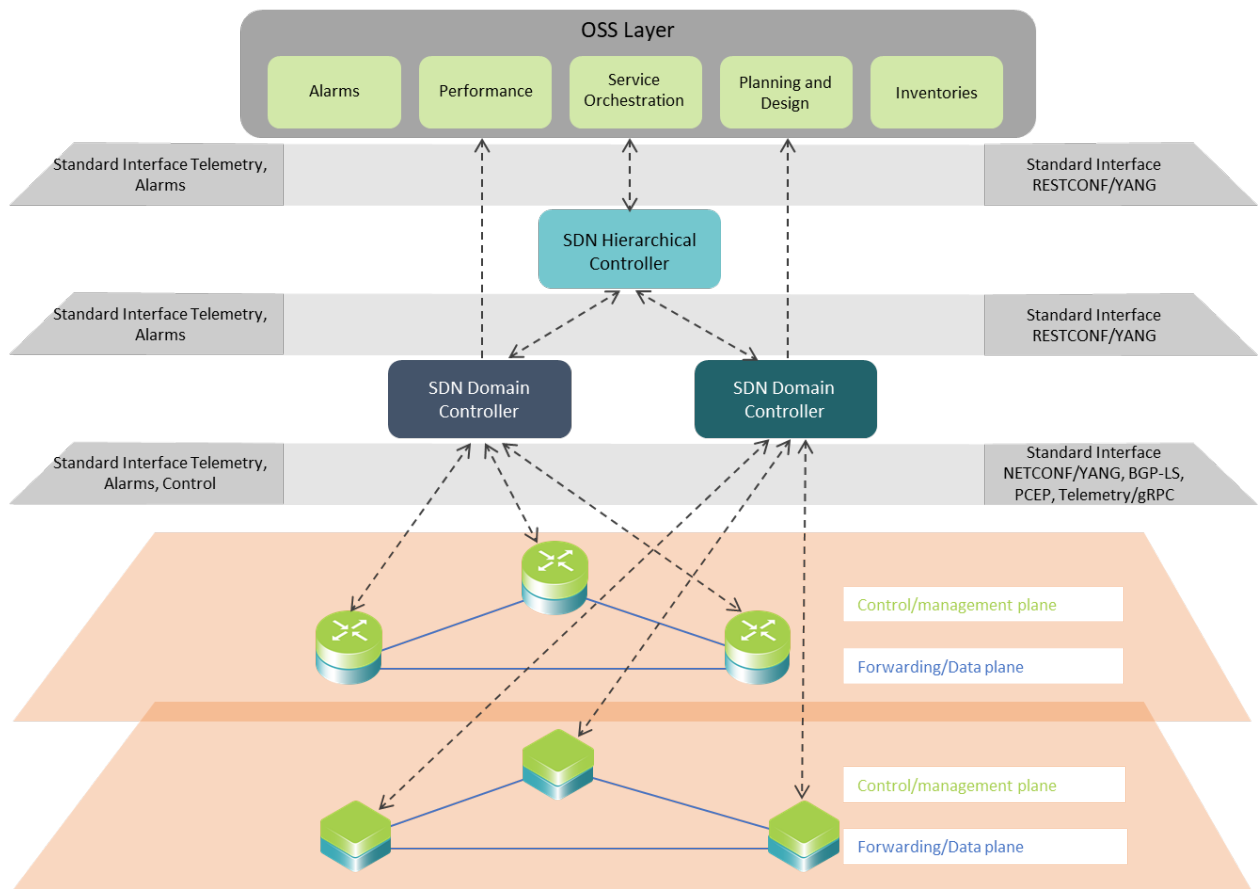
Figure 8. Open Transport SDN Architecture Vision

5.2 Standards support

The SDN & Network programmability concepts have been promising objectives for many operators and crucial for achieving agility in network provisioning and operation. In the past few years, standardization bodies and industry have spent lot of effort to standardize the SDN implementation and they have developed OpenConfig & Netconf RFC 6241

However, the network operators found a challenge to achieve network programmability and automation when it comes to a multi-vendor network. They soon realized that the programmatic interfaces available from networking vendors

6

General Requirements

6. General Requirements

6.1 Regulatory requirements

The solution shall be compliant with EU GDPR regulation: REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons regarding the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)

6.2 Access security and anti-theft requirements

In general, the solution must support the necessary security mechanisms to authenticate and encrypt communications between the network element and its management system or controller.

The network element should offer the possibility of only enabling local traffic after the device has been authenticated by the management platform/controller.

The system should also offer the possibility to enable anti-theft mechanisms that prevent the use of the equipment in any other environment than the one it was conceived in.

7

Glossary

8. Glossary

AAA   Authentication, Authorization and Accounting

AC   Alternating Current

ADMF   Administration Function

AGF   Access Gateway Function

API   Application Programming Interface

AR   Augmented Reality

AS   Autonomous System

ASIC   Application-Specific Integrated Circuit

ATM   Asynchronous Transfer Mode

BBF   BroadBand Forum

BFD   Bidirectional Forwarding Detection

BGP   Border Gateway Protocol

BGP-LS   BGP Link State

BGP-LU   BGP Labelled Unicast

BMCA   Best Master Clock Algorithm

BNG   Broadband Network Gateway

BRAS   Broadband Remote Access Server

CAC   Call Admission Control

CBS   Committed Burst Size

CE   Customer Edge

CG-NAT   Carrier Grade Network Address Translation

CHAP   Challenge Handshake Authentication Protocol

CIR   Committed Information Rate

CLI   Command Line Interface

CoA   Change of Authorization

CoS   Class of Service

CPU   Central Processing Unit

DC   Direct Current

DCSG   Disaggregated Cell Site Gateway

DHCP   Dynamic Host Configuration Protocol

DOR   Disaggregated Open Router

DSCP   DiffServ Code Point

DSL   Digital Subscriber Line

DSLAM   Digital Subscriber Line Access Multiplexer

E-VPN   Ethernet VPN

EBS   Excess Burst Size

ECMP   Equal Cost Multi Path

EIR   Excess Information Rate