NANOG Presentation

13-FEB-2023

The Operational Impacts of Supporting a Disaggregated, Distributed, Cloud-based Network Architecture

Aliraza Bhimani
Principal Network Engineer/Team Lead
Comcast Cable
Aliraza_Bhimani@comcast.com
13-FEB-2023

Introduction

Managing and maintaining highly scalable networks is complex

- Support today's IP network traffic demands driven by video, gaming, and remote working
- Future growth
  - Access technologies e.g., 5G/6G cellular traffic volumes increase port and bandwidth
  - Commercial Services port growth
  - Peak information rate growth due to DOCSIS 4.0 providing multiple symmetrical gigabit services
  - Port capacity increases are accelerating 100->400->800->1.2Tbps
  - Normal Growth rate in High-speed data services
    - Internet usage has increased by 1,355% over the last 22 years[1]!

NANOG™

Introduction

Cable operators must rethink and rearchitect their existing IP networks and operations to maximize performance and efficiency

Network operators must consider:

- Total cost of ownership of various hardware and software options
- End to End, not just that of an individual component or software option
- Time to market of new services
- Simplicity of the network operations
- Reduce the blast radius and impact of component failures
- Availability of the network

NANOG™

Introduction

In this presentation we dive into

- Disaggregated Distributed Chassis (DDC) and Disaggregated Distributed Backbone Router (DDBR)
- How solutions based on open-source specifications can solve current operational challenges when building and scaling IP backbone/aggregation networks [5, 6]
- How cable operators can:
  - Leverage open transport building blocks across various networks to implement concepts of real DDC/DDBR architecture
  - Utilize orchestration, automation, and analytics
  - Shift IP backbone architecture from single chassis to a disaggregated, distributed, scalable solution

Disaggregation as a success criterion for ISPs/CSPs

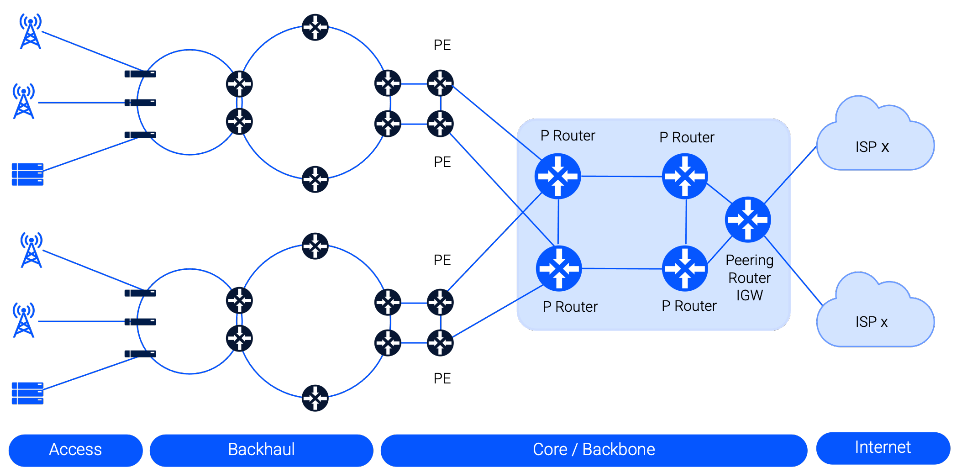
Key points of the DDC/DDBR solution [5, 6]:

- **Disaggregation is driving competition**
- **Innovation**
- **Operational Efficiency**
- **Reliability**

Figure 1 – DDC/DDBR placements in the network

DDC/DDBR

- This system can be placed:
  - In the IP/MPLS backbone
  - In the access layer
  - Internet Gateway Router (IGW)
- Network Operating System(NOS) agnostic HW



NANOG

Differences between DDC/DDBR and traditional routers

Common to both types:

- Line cards, fabric modules, Route Processor modules, PSUs, and FAN FRUs

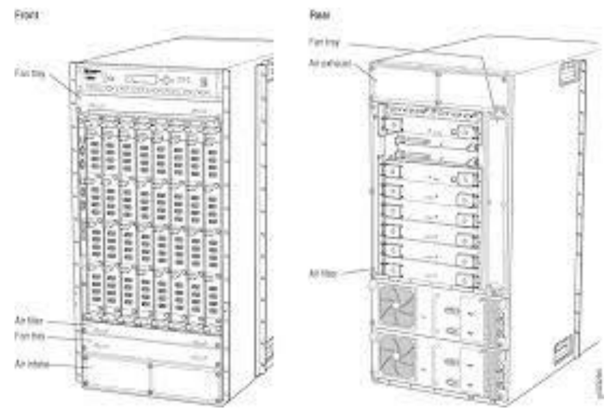
Unique:

- Modular chassis routers use pluggable line cards vs. Standalone "pizza boxes"
- Modular chassis can't span racks if Power or Cooling Per Rack is limited
- Modular chassis can only run proprietary software
    - Modular chassis vendor responsible for support of both HW and SW
    - Multiple vendors involved in DDC/DDBR
- Modular chassis Scales UP vs. DDC/DDBR scaling OUT
- DDC/DDBR has a lot of external wiring
- DDC/DDBR allows for best in breed in HW and SW

Figure 2 - Front and Back View of a Modular Chassis

Modular Chassis

Figure 2 shows a front and back view of a finished product view of a generic modular chassis routing system [5]



NANOG™

Disaggregated Routing System

- Route processors, fabric modules, and line cards are all disaggregated onto either
  - Whitebox 1RU routers
  - Commercial off-the-shelf (COTS) x86 server
- Several Fabric Forwarders (FF) which represent the spine or backplane
- Packet Forwarders (PF) which represent the leaf of the cluster or line cards
- The “brains” of the system i.e., routing stack are running on redundant servers in a cloud-native fashion
- The external ethernet switches provide OOB, distributed communications channels between all 3 described components
- Ability to scale vertically by adding more FFs
  - Increase the fabric throughput and redundancy
- Ability to scale horizontally by adding more PFs or line cards
  - Accommodate port demands and future growth

Disaggregated Routing System

The combination of the Spine & Leaf based architecture and disaggregation can lead to phenomenal advantages and innovations:

- ISPs/CSPs can easily change Routing Network Operating System vendors while using the same hardware
- ISP/CSP's do not have to be bound to a specific router vendor's hardware/software roadmap:
  - Can freely mix and match different Whitebox server hardware
  - Have custom vendor software applications/containers
- This combination also promotes more open-source concepts:
  - APIs to "hook" into this custom routing platform
  - "Virtual" chassis capability

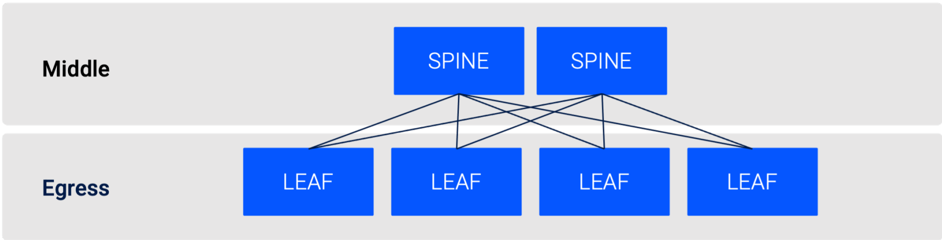
NANOG™

Figure 3 – 3-Stage Non-blocking Clos Topology

Clos Topology

3-stage Clos network is the smallest version of a Clos network:

- Relevant to modern scalable carrier and hyperscalers' networks
- Has 3 stages: ingress, middle and egress, where the ingress and egress points are folded back on top of each other [22]



Disaggregated Routing System

- Allow potential for seamless integration into the existing network while maintaining small failure domains
- N+M design allows hitless maintenances and updates

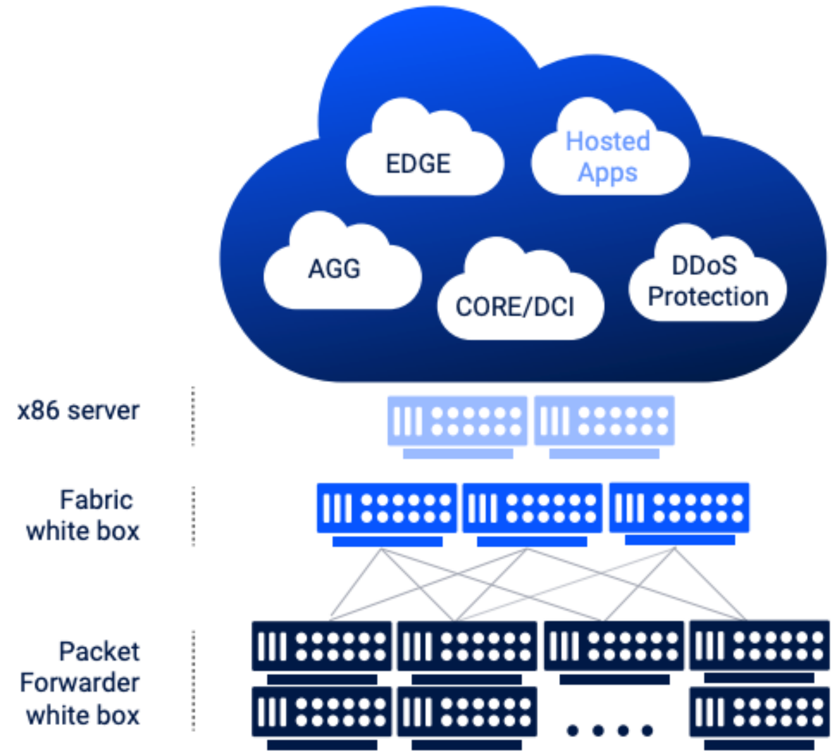
Challenge of the identification of the responsibility domain: hardware and software

- 2 vendor solution requires an understanding of which domain a problem belongs to e.g., HW or SW
- Future launches of a virtual chassis in the network will be done by the orchestration system:
  - Supports comprehensive validation methods:
    - Spaghetti-wiring fabric cabling
    - Management of the code upgrades
    - Device configurations

Figure 4 - Network Cloud

Evolution of Network Functions

- The offload of the NF to the Network Processing Unit (NPU) on a dedicated Input/Output (I/O) device.
- Similar offer is made by “smart NIC” providers who offload the NF to the Data Processing Unit (DPU) and accelerate it as a result.



NANOG™

Figure 5 - Network Cloud

DDC/DDBR Architecture

- Based on the Whitebox architecture where each component of the standalone monolithic chassis router is distributed into individual components .
- This allows the operators to scale out this re-architected chassis by simply adding white boxes that perform fabric and packet forwarding functions.

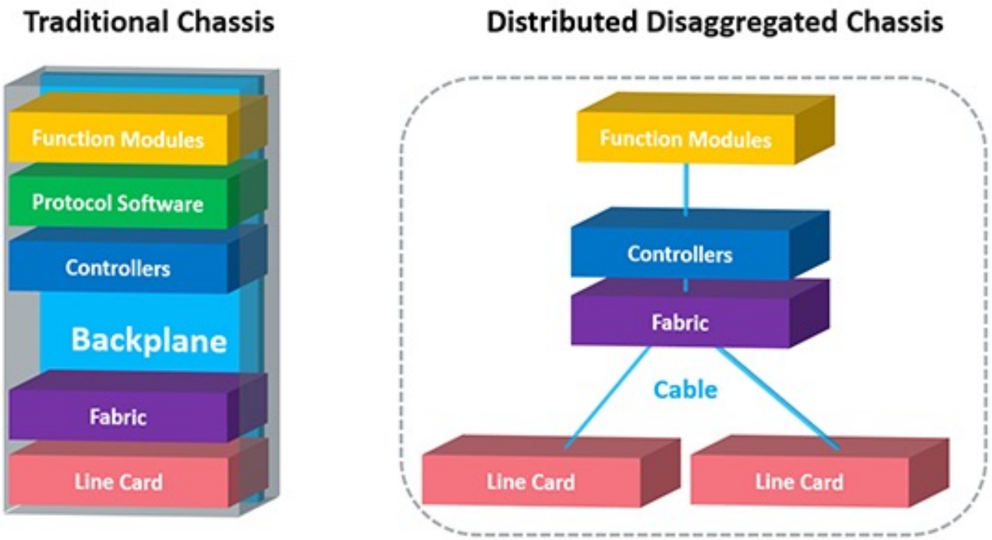
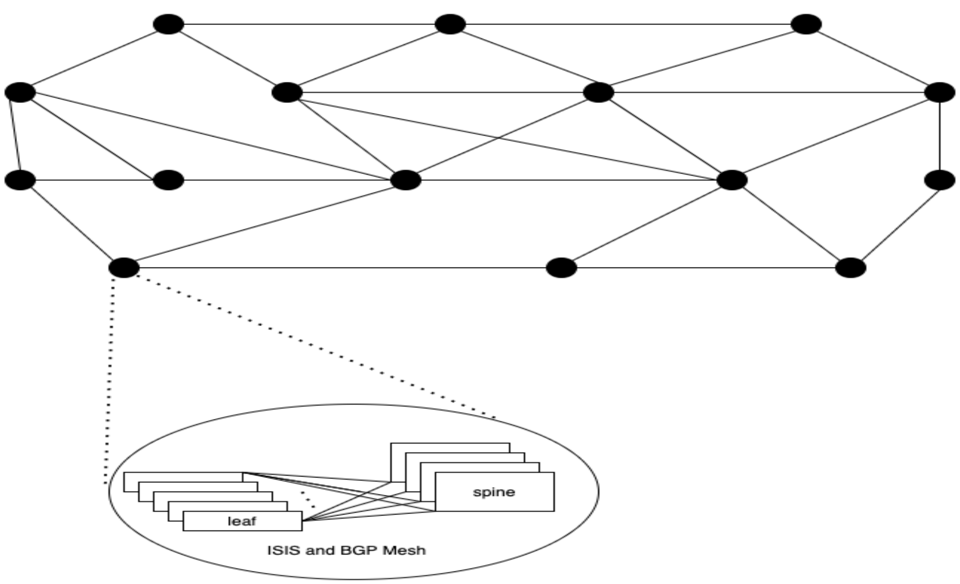
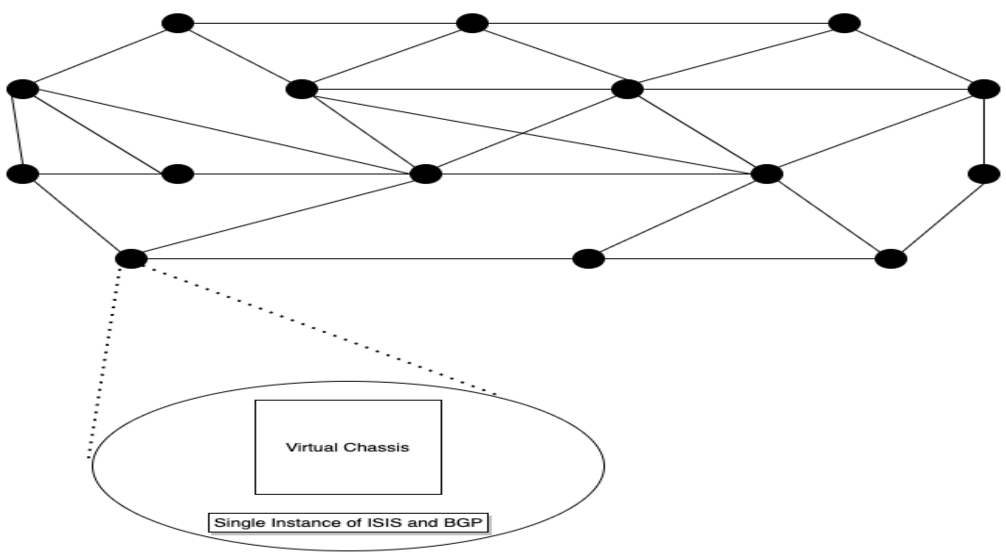


Figure 6 - S/L Cluster vs. Virtual Chassis Within a Backbone/Core Site

In Typical spine leaf an ISIS and BGP mesh must be maintained within the cluster and with external, core network.



The virtual chassis eliminates the requirement for IGP based intra-cluster control-plane.



NANOG™

Real-life deployment of the DDC/DDBR model

- DDC/DDBR architecture is a field-proven concept deployed in the core and aggregation layer of the largest telcos and cablecos in the world [26]
- There are several tools and dashboards needed to monitor production grade Clos networks

On the other hand, with a virtual cluster:

- Easier for the NOC team to manage the network and respond accordingly to the events within the system
- Reduction of the number of nodes in the IGP domain
  - Reduces complexity and load on routers
- No BGP enhancements required to improve recovery time during convergence

NANOG™

New System Admin Skills Needed For Network Engineers

- DevOps, programmability, and scripting skills to leverage automation tools and utilize server sysadmin skills
- Model driven (YANG) configuration and telemetry
- API vs. CLI driven configuration and telemetry
    - Eliminates human errors in repetitive tasks

Microservices

- Complexity of the code, scale limitations, reliability among many others
- A software code bug may affect both the SW and HW

Microservices allow a relatively large application to be divided into smaller parts, having their own autonomy

- With microservices in containers:
  - Simpler to take advantage of hardware
  - Easily orchestrate services
- Majority are based on the cloud-native, distributed architecture

NANOG™

Operational Considerations and Challenges

Challenges with traditional router architectures

- Notable reliance on a very limited number of suppliers
- Market which is extremely difficult to enter and compete
- Possible interoperability across different hardware components
- There are availability concerns when replacing or upgrading a component inside the chassis
- Data and control plane closely tied together leads to a significant dependency on the vendor’s roadmap
- Best HW vendor isn't necessarily the best SW vendor and vice versa

Operational Considerations and Challenges

Traditional router architectures chassis limitations and tradeoffs

- Resiliency to maintain uninterrupted connectivity to the mobile, broadband, and business customers
- Significant computing and storage capability to store IPv4 and IPv6 global routing tables
- Port density and capacity enough to support the growth of the customers and services

- High availability at all levels: control processing, switching fabric, cooling, and power
- Large TCAM (Ternary Content Addressable Memory), strong computing capability, deep buffers
- Variety of sophisticated control plane features; NSR (Non-Stop-Routing) & ISSU (In-Service Software Upgrade)
- Large port density, number of slots for line cards and backplane switching capacity

Operational Considerations and Challenges

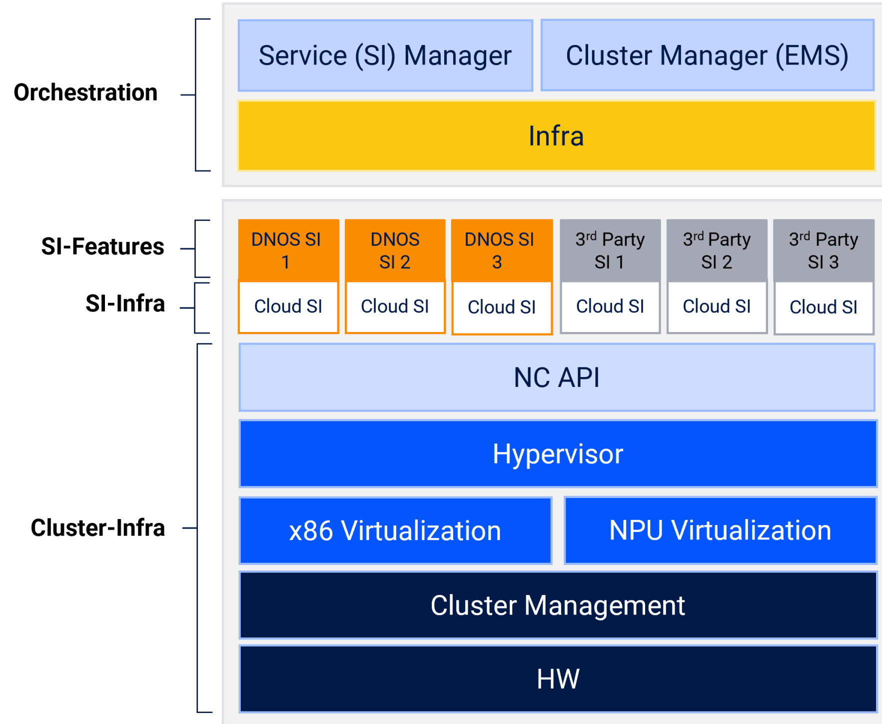
Challenges and tradeoffs with Spine/Leaf architectures

- Packet forwarders enable operators to utilize any port on the Whitebox for any service regardless of the implementation area
- Multiple NOS vendors in the ecosystem can install their software onto the open networking hardware
- Clusters are built in a way where bandwidth is equal or over-provisioned to leaf ports [23]
- Installation process is identical across all node sizes, eliminating the need to constantly train staff and avoid costly installation errors
- There is a lot of backend fabric fiber and ethernet connections (fondly referred to as "spaghetti-wiring") that must be installed to connect the nodes as a cluster
- There is a lot of up-front cost and effort in a DDC/DDBR architecture in performing the physical install work initially to obtain a longer roadmap

Figure 7 - Multi-services Network Cloud Architecture for the Edge

Software Architecture

- Combining networking and compute resources over a shared, cloud-like infrastructure.
- Allows operators to put greater functionality at the network edge, even with space and power limitations.
- Each networking function, which runs a containerized Service Instance (SI), can be allocated with its required hardware resource (Physical interfaces, NPU, CPU, TCAM, QoS etc.) out of the underlaying shared hardware infrastructure



NANOG

DDC/DDBR S/L and Virtual Chassis solutions

With the proliferation of the Internet of Things (IoT) it became apparent that it is beyond the bounds of possibility for humans to manage them [12]

MUST automate configuration management and maintenance as much as possible and streamline the deployment

- Engineers should get familiar with network automation tools and methodologies that offer an API
- Automation of configuration of each node in the S/L cluster
- Need an Orchestration, Automation and Analytics system which can now treat all the nodes like a virtual cluster
  - Best of both worlds = benefits of a traditional system with the flexibility of S/L.

NANOG™

Health Monitoring and Assurance

The orchestration tool automates event and KPI monitoring for:

  - Cluster topo/nodes states
  - Formation and connectivity across clusters

- Hardware components:
  - CPU/Memory
  - Environmental
  - Ports and interfaces
- Software components:
  - Base OS
  - Firmware, processes, containers, and microservices

NANOG™

Telemetry

Move from SNMP polling based to streaming telemetry

- Scale - Collector does not need to poll each router individually
- Define different set of counters for collection from specific routers

NANOG™

The Future of Disaggregated Solutions

"The only constant in life is change." Heraclitus

- The pace of innovation must keep up with the demand and new emerging use cases
- Telecommunications companies along with vendors will keep driving innovation in disaggregated network solutions
- Essential future developments are multiservice port functionality on top of the shared pool of resources, OpenOffload and data center sustainability

The Future of Disaggregated Solutions

With the growth of IoT and “smart” appliances, almost any type of device needs connectivity to the home network

- Examples of this are laundry machines, refrigerators, stoves, microwaves, thermostats, garage openers, and even the front door keypad of a home!
- Network operators are continuously seeking technologies to reduce carbon footprint
- Decommissioning a disaggregated cluster and re-using its hardware in a developing country where it can be practical for many years to come paves the way for more sustainability

Conclusion

Traffic growth over the last two decades requires ISP/CSP to explore alternative routing solutions

The disaggregated approach brings:

- Cost reduction
- Removal of vendor lock-in
- Service innovation

DDC/DDBR architecture brings:

- Service agility and faster innovation
- Cloud orchestration for streamlined cluster management, device configuration and critical files retrieval

Future to include OpenOffload and Sustainability goals

NANOG™

Bibliography & References

[1] https://www.broadbandsearch.net/blog/internet-statistics

[2] https://corporate.comcast.com/stories/announcing-another-10g-milestone-amidst-a-flurry-of-innovation

[3] https://www.broadcom.com/blog/800g-optical-platform-solutions

[4] https://www.youtube.com/watch?v=d90uK4WW0Po

[5] Hardware Specifications and Use Case Description for J2-DDC Routing System, Tuan Duong; Open Compute Project

[6] Distributed Disaggregated Backbone Router (DDBR) Technical Requirements Document, Eva Rossi, Jose Angel Perez, Kenji Kumaki, Ryuji Matsunaga, Yuji Sonoki, Ahmed Hatem, Diego Marí Moretón; Telecom Infra Project

[7] https://www.opencompute.org/

[8] https://telecominfraproject.com/

[9] GigaOm Radar for Network Operating Systems: Network Service Providers, https://research.gigaom.com/report/gigaom-radar-for-network-operating-systems-network-service-providers/?utm_content=208592782&utm_medium=social&utm_source=linkedin&hss_channel=lcp-11282464

[10] https://www.cncf.io/

Bibliography & References

[11] https://www.opencompute.org/wiki/Networking/ONL

[12] The Internet of Things: Catching up to an accelerating opportunity, Michael Chui, Mark Collins, Mark Patel; McKinsey & Company

[13] https://restfulapi.net/

[14] https://datatracker.ietf.org/doc/html/rfc6241

[15] https://www.w3schools.com/xml/xml_whatis.asp

[16] https://www.w3schools.com/js/js_json_intro.asp

[17] https://yaml.org/

[18] https://datatracker.ietf.org/doc/html/rfc7950

[19] https://opencomputeproject.github.io/onie/overview/index.html

[20] Optical Transceivers for Datacom & Telecom Market 2021: Market and Technology Report 2021, Dr. Martin Vallo, Pars Mukish, Dr. Eric Mounier; Yole Development

[21] A Study of Non-Blocking Switching Networks, Charles Clos; The Bell System Technical Journal, March 1953

[22] On Nonblocking Folded-Clos Networks in Computer Communication Environments, Xin Yuan, Department of Computer Science, Florida State University, Tallahassee, FL 32306, 2011

Bibliography & References

[23] https://www.ufispace.com/company/blog/what-is-a-distributed-disaggregated-chassis-ddc

[24] Engineered Elephant Flows for Boosting Application Performance in Large-Scale CLOS Networks, Broadcom Corporation, March 2014

[25] https://grpc.io/

[26] https://about.att.com/story/2020/open_disaggregated_core_router.html

[27] Multiservice: Maximizing Infrastructure Utilization; DriveNets Whitepaper

[28] https://github.com/att/sessionOffload

[29] https://datatracker.ietf.org/doc/html/rfc6071

[30] https://datatracker.ietf.org/doc/html/rfc8926

[31] Topology and component description of DriveNets Network Cloud as a sustainable solution for multipurpose networking, Run Almog; Open Compute Project

[32] Annual outage analysis 2021, Andy Lawrence; Uptime Institute https://uptimeinstitute.com/uptime_assets/25ff186d278b32c202fc782e60a0d473bd72bfbc6d4d65afedfa15dd406c7656-annual-outage-analysis-2021.pdf

[33] The Operational Impacts of Supporting a Disaggregated, Distributed, Cloud-based Network Architecture, Aliraza Bhimani, Comcast Cable; Idris Jafarov, DriveNets https://www.nctatechnicalpapers.com/Paper/2022/FTF22_OP03_Bhimani_3765

[34] The Evolution of the Edge – Why Edge Compute and Networking Should Tightly Integrate into a Cable Edge Cloud, Idris Jafarov, DriveNets https://www.nctatechnicalpapers.com/Paper/2022/FTF22_CLOUD01_Jafarov_3799

Thank you

A big "Thank you!" to all the below individuals for their contributions:

Idris Jafarov: Delivery Team Leader, DriveNets (Co-author of the white paper)
ijafarov@drivenets.com

Mannan Venkatesan: Distinguished Engineer, Comcast Cable
mannan_venkatesan@comcast.com

Tony Tauber: Distinguished Engineer, Comcast Cable
tony_tauber@comcast.com

Robert Gaydos: Sr. Fellow, Comcast Cable robert_gaydos@comcast.com

Larry Wolcott: Fellow, Comcast Cable larry_wolcott@comcast.com

13-FEB-2023

NANOG™

COMCAST