SOLUTION BRIEF

Telecom Workload Placement and Affinity Model

An analysis to guide decision making for workload placement

Authors

Ahmed Ibrahim

Eslam Kandiel

Petar Torre

Introduction

As network transformation progresses, it has become the focus of Communications Service Providers (CommSP) in the Telecom vertical as well as Cloud Service Providers (CSP). CommSP are targeting or evaluating the implementation of parts of Network Functions Virtualization (NFV) plans in the cloud, while adopting cloud principles. At the same time, the CSPs are interested in using Telecom's physical locations to host services and getting closer to their end customers.

Understanding the characteristics of a network workload, including how it fits into the overall network architecture, and the factors affecting how the workload operates (connectivity, performance, security and integration) is fundamental when making a CSP services hosting decision.

We wrote this document to provide telecom planners and architects with an analysis that guides the decision making for such workload placement.

This document is part of the Network Transformation Experience Kit, which is available at https://networkbuilders.intel.com/network-technologies/network-transformation-exp-kits.

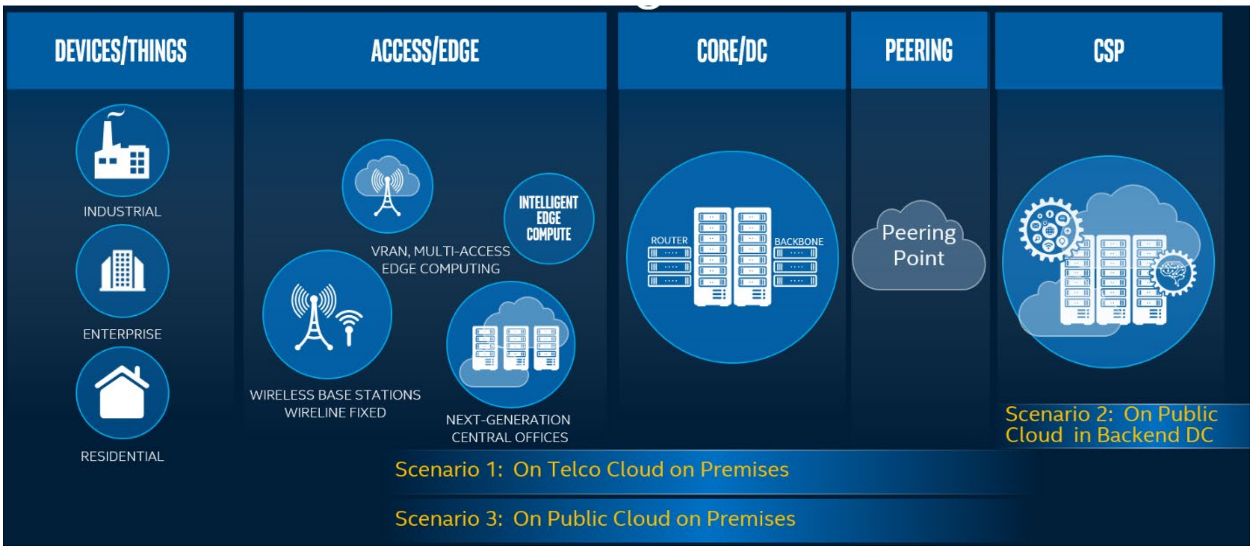
Telecom Workloads Implementation Models

This section describes several key implementation models for placing workloads in a telecom environment:

- **Telecom Cloud on premises** is the telecom-owned private cloud network. It supports the delivery of telecom services at scale, by leveraging the cloud. Telecom cloud is based on hardware and software infrastructure designed to enable flexible and efficient deployment of the applications used by the telecom to manage and deliver services. A telecom cloud is a highly distributed infrastructure capitalizing on the emerging edge computing model. It deploys virtualized and programmable platforms; leverages NFV, Software-Defined Networking (SDN), Artificial Intelligence (AI), and automation technologies; and adopts cloud business practices that change the way they innovate. The telecom cloud enables operators to become more open, dynamic, agile, and efficient.
- **Public Cloud in backend data center (off premises)** is based on virtualized compute, network, and storage resources that are offered and managed by a third party outside of the telecom's private network. Resources are typically hosted in a multi-tenant configuration in external geographically distributed data centers.

- **Public Cloud on premises** provides an approach that combines workload components from both the telecom cloud and public cloud solutions. This is a true hybrid cloud computing solution, which is an extension of the public cloud, allowing telecoms to provide public cloud services using their own on-premises data centers.

The figure below shows a high-level view of several workload hosting scenarios.



Telecom Workloads Placement Criteria

The study in this paper defines a methodology and criteria for mapping telecom workloads into the most appropriate cloud implementation model. Four technical aspects have been defined to help assess a workload suitability to a selected deployment-model. A score, based on the operators' strategy, can be assigned to each of the four aspects. The sum of all four scores can guide the operator's decision about the match of workload to implementation model. The following sections describe the four technical aspects that should be considered: physical connectivity, integration, security, and performance.

Note: Non-technical aspects learned from other verticals are described in Learnings from Other Verticals or Telecom Infrastructure Domains, later in this brief.

Physical Connectivity

Physical location and physical connectivity play a major role when placing the telecom workload, in particular data plane functions, which are performance-sensitive. The infrastructure that connects the user and the digital services must ensure efficient digital services reachability and meet the performance criteria, including throughput and latency Key Performance Indicators (KPIs).

Throughput of the connection is one of the main requirements to be achieved by the service provider to meet modern network serviceability targets, which includes 5G in the mobile domain and Gigabit-capable Passive Optical Network (GPON) in the fixed connectivity domain, for example. The service provider data plane core network functions aggregate traffic from thousands and millions of users [examples include the User Plane Function (UPF) in the mobile domain and the Border Network Gateway (BNG) in the fixed connectivity domain]. The connectivity throughput of the data plane core function become a major challenge with these increased throughput requirements and is one of the biggest factors for bottlenecks.

The connectivity of the core network function is achieved through the IP technology, which is running on top of the Wavelength-Division Multiplexing (WDM)/fiber technology. In this situation, the physical location plays a major role in defining how the connection is built. From the WDM/fiber perspective, moving the traffic away from the core location can be very challenging for the core network/data plane function. In addition, it impacts the costing model associated with the data connectivity offered to the users. The **Public cloud on Premises** model is one of the options that overcomes the challenges associated with the connectivity for the data plane core functions, if it is compatible with other requirements.

Latency is another aspect to be considered for placing telecom functions. Many of the emerging digital services are very sensitive to the latency between the user and the service. Also, some network functions are very sensitive to latency issues for correct operation. One particular example of a latency-sensitive network function is the Radio Access Network (RAN). RAN can be interrupted and lose its capability to connect users to the network if the latency is increased to a certain limit defined by the targeted technology (for example, 4G and 5G). With the emerging Virtualized RAN (vRAN) and O-RAN Alliance initiatives, the telecom cloud becomes the cloud of choice to host different components of the vRAN.

Integration

A service provider network consists of many functions that are tightly integrated with each other to perform the needed connectivity functions. These functions have a variety of characteristics and requirements, which are related to the service type and the purpose that is fulfilled by this function. Fixed services related functions [such as BNG, Authentication, Authorization and Accounting (AAA), and Optical Line Termination (OLT)] have relatively fewer integration requirements, compared to 3rd Generation Partnership Project (3GPP) mobile services related functions [such as Policy and Charging Rules Function (PCRF), Mobility Management Entity (MME), Serving Gateway (SGW), and Packet Data Network Gateway (PGW)]. It is worthwhile to highlight that the future of the 3GPP releases will cover fixed mobile convergence, which will provide a tight integration between both fixed and mobile related network functions.

The integration requirements are not only tied to the telecom functions required to deliver the services, but are also extended to cover the integration with Operation Support Systems (OSS) and Business Support Systems (BSS). OSS is the computerized system that helps operators manage the network and perform different operational functions, including fault detection, fault correlation, and network service healing, for example. All network functions must be integrated with the corresponding OSS system. BSS is the computerized system that ensures the mapping between the network service and the user, doing tasks such as service fulfillment and billing required activities. Only the nodes that have a role in service activation must be integrated with BSS related functions/APIs, which includes Home Location Register (HLR)/Home Subscriber Server (HSS), PCRF, and AAA, for example.

The 3rd Generation Partner Project (3GPP*) defines the details of the EPC architecture, functional elements, and interface requirements. 3GPP TS 23.002 provides an overview of the architecture of the 3GPP system. In particular, it describes all the network elements used in the EPC and also in legacy core networks. This specification is available at: http://www.3gpp.org/DynaReport/23002.htm

The following diagram shows the basic 3GPP configuration and interfaces to support CS and PS Services.

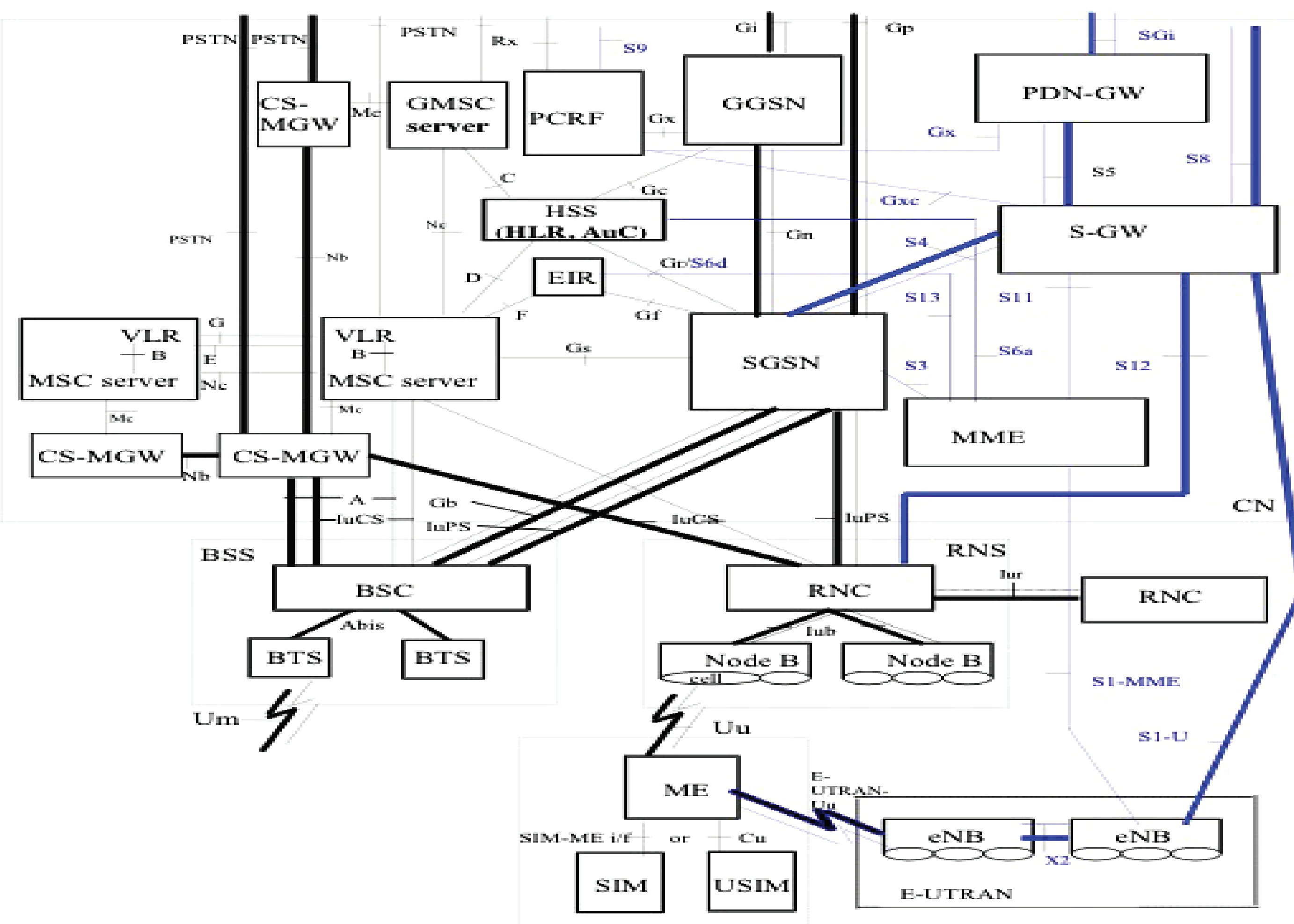


Image source: http://www.3gpp.org/DynaReport/23002.htm

Security

Security is a significant concern for the service provider because the telecom network resides between the user and their targeted applications. Additionally, the service provider has a significant amount of customer's private information, such as their location and billing information. Service providers use different levels of security hardening to ensure their network and their customer information are protected. One of the most common security practices followed by service providers is to isolate critical telecom functions in the network connectivity domain, and strictly limit the connections to their domain. Network database functions like Home Subscriber Server (HSS) and Home Location Register (HLR) are examples of critical functions that not only have user data which can be compromised, but if the ability to reach it is lost or intercepted, then the telecom service itself can malfunction.

Voice-related functions, such as IP Multimedia Subsystem (IMS), can be critical as well. While intercepting voice calls can be critical, the amount of signaling information available in voice-related functions can be misused in a very dangerous way. Although all the telecom functions need to be treated in a secure manner, not all of them have the same level of criticality. On the other hand, pure data plane functions that are controlled by other signaling functions, like UPF, tend to have less security restrictions compared to signaling and database functions.

Regulations are a very important aspect that must be considered in this analysis. Laws and regulations defined by communication authorities in each country play a significant role in defining security measures. In many cases, the physical security of "where the location is applied" is mandated by the regulation authorities. All these security related elements are considered during the scoring process used in this paper.

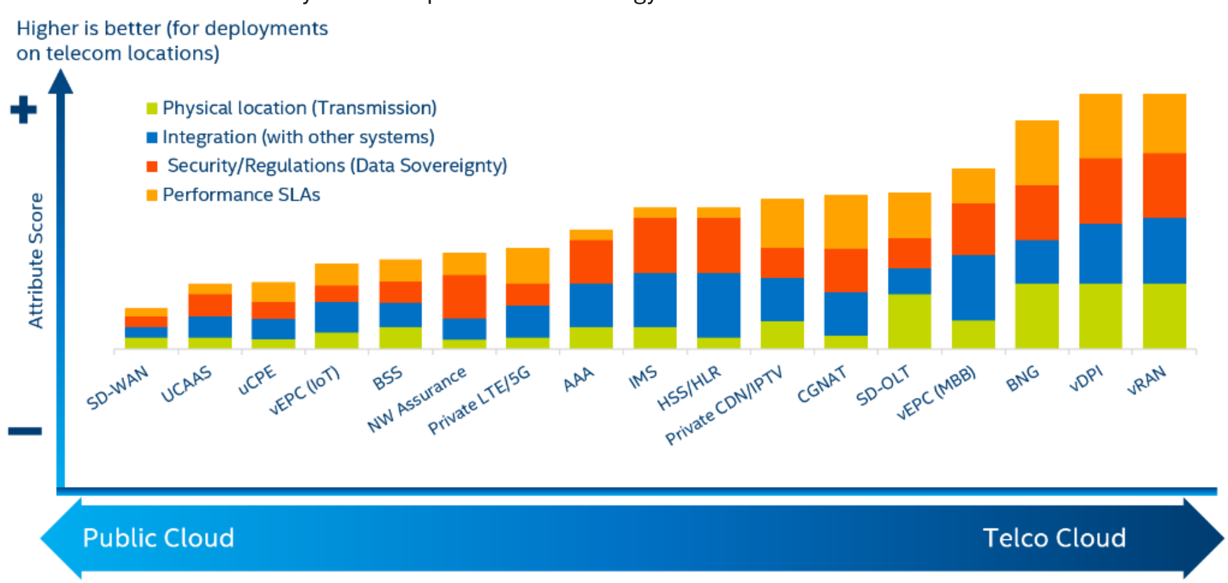
Performance

Performance is another aspect we used in our scoring method. Since the telecom function requires high performance hardware, performance is scored higher when compared to the other aspects. Not all telecom functions have the same level of performance requirements. Data plane functions like BNG and UPF are very demanding from a networking resources perspective. On the other side, the signaling plane functions such as MME and Signaling Management Function (SMF) push the processing requirements to the limit, leading to a noticeable compute aspect that is directly proportional to the number of concurrent connections and/or number of users. Some workloads are demanding for both the networking and compute power. RAN virtualization is a clear example of this type of workload, alongside the Deep Packet Inspection (DPI). It is noteworthy that not all the telecom functions have heavy demands on the hardware; some functions are low-demand and can be easily hosted in wide range of cloud implementation models.

It is important to highlight that some telecom functions and workloads can benefit from special hardware accelerators to more easily achieve requirements like latency or get to better cost by increasing throughput per node. One clear example of this is using Intel® QuickAssist Technology (Intel® QAT) to enhance the IPSec Gateway functionality for the S1 interface, using the IPSec performance improvements in the 3rd Generation Intel® Xeon® Scalable Processor, and using Intel® FPGAs (Field Programmable Gate Arrays) and Programmable Devices for RAN virtualization. In our methodology to evaluate the performance impact on the placement strategy, using hardware accelerators to optimize the telecom functions enhanced the performance score.

Benefits of Solution and Use Case Examples

In creating the **Telecom Workloads Placement and Affinity Model** we evaluated the four aspects described above and scored the different telecom workloads against them. The diagram below shows an example of equally weighted attribute scores for each of the workloads. Exact weighting of particular score category can vary to more accurately guide telecoms to make workload placement decisions in the best way for their operation and strategy.



Learnings from Other Verticals or Telecom Infrastructure Domains

This section summarizes certain learnings that are relevant for telecoms, including design principles, applied methodologies, success criteria, and calculating Total Cost of Ownership (TCO) and Return on Investment (ROI).

For some time, companies in different vertical market segments than telecoms have had the choice of running workloads on public cloud backend and edge platforms (like CloudSP offerings in major regional DCs or Edge DCs) or building their own centralized or distributed platforms. This is also true for telecoms in other infrastructure domains, such as IT. Over time, many vertical companies in both of these scenarios have moved their services between different cloud infrastructure offerings.

Lessons Learned When Moving from On-Premises to Public Cloud and Vice-Versa

Strategy and Planning:

- The link between core business strategy and cloud strategy is critical. It can allow companies to outsource capabilities that are seen as secondary to their core business; however, when this link is misaligned, even a technically successful move to the cloud can fail to deliver the expected business results.
- Cloud migration must be combined with a comprehensive organizational approach to allow for a much more effective agile transformation.
- Once a decision is made, nothing is more impactful than time and planning. Success is a direct function of both of these characteristics.

Financial Consideration:

- Financial consideration is not usually the lead driver for moving services, it's often more of an outcome when done correctly. The economics are evolutionary, not straightforward, and often are not "one to one", with longer time frames and variable cost models creating challenges. The economics of migration include the obvious cost of infrastructure and service, however there are significant soft costs (for example, perceived flexibility) which are less tangible.
- Cost management is complex, but critical for success. Companies are leveraging Proof of Concepts (PoCs) to estimate and manage costs before operationalizing the move. Investments in cloud management tools are recommended to manage cloud spend, scale up cost management staff, and negotiate enterprise discount programs.
- Comparing operating costs can be challenging because on-premise organizations lack visibility into the application-level cost. In the public cloud, many costs may not be obvious from the cloud provider fee schedule. Also, very often cloud implementations cost more than necessary, due to poor implementation and adoption.

Talent and Adoption:

- Focus in-house talent: Organizations benefit from supporting business priorities such as digital transformation, time to market and revenue growth, while minimizing low-value and labor-intensive activities.
- For better adoption, implement Change Management, extensively train both commercial and technical teams on cloud-based offerings, and have leaders set tone and approach for the project.

Workload and Use Cases:

- Cloud is not the right solution for every organizational need: The type of cloud service selected by an enterprise is critical. How the cloud service is managed is also critical. Thinking strategically about benefits, costs and risk is paramount and must be done upfront, before any contract is signed.
- When looking to migrate, a bottoms-up approach and workload-by-workload analysis should be completed for every application to be migrated to the cloud. The optimal environment for each application (for example, on premises, co-location, private cloud, or public cloud) and its corresponding use case should be evaluated.
- Leverage refactoring of applications to take advantage of current architectures, characterized by microservices and containerization to enable companies to balance the projected cost, focus on the pace of innovation and enhancements, and improve responsiveness of the fast-changing needs in dynamic markets.
- Continue to evaluate applications that have already been re-hosted or re-platformed for the cloud and decide whether further adaptation would be beneficial.
- For cases where on-premise workloads were moved to the public cloud and then brought back to on-premise: Factors for these decisions were workloads that cannot auto-scale, lift-and-shift without cloud optimizations (resulting in higher costs and instability), and not adequately training the support teams.

How to calculate TCO or ROI

Two common ways to do in-depth financial analysis are:

1. Total cost of ownership (TCO): Accounts only for the cost that is associated with an acquisition for its entire life span or a predetermined period.
2. Return on investment (ROI): Considers all the tangible and intangible business benefits (for example, increased employee efficiency, faster time to market, and others) when compared to the hard costs of migration, including migration, operational expenses, etc.

ROI is harder to estimate but is becoming the preferred evaluation method over TCO. Calculating ROI does not need to be complex and it can be a directional estimate because cost is not the primary driver. A simple and effective ROI calculation enables the enterprise to support an investment decision and measure whether the expected costs and benefits occur. An overly complex calculation can make it hard to understand why a decision was made or measure its effects.

Full methodology needs to include financial KPIs like on-premise, migration and cloud operations cost, revenue, profitability and cash flow, and take input for all benefits and costs categories.

Technologies Implemented

The presence of identical cloud infrastructure hardware instances across multiple locations minimizes the efforts of application migration, validation and supporting of multiple environments, and makes operations easier. Intel offers an unmatched portfolio for the unique requirements of cloud and edge implementations including 3rd Generation Intel® Xeon® Scalable Processors, Intel® Xeon® D-2100 Processor, Intel® Ethernet 800 Series, Intel® Optane™ Solid State Drive DC P4800X Series and Intel® Optane™ DC Persistent Memory 200 Series.

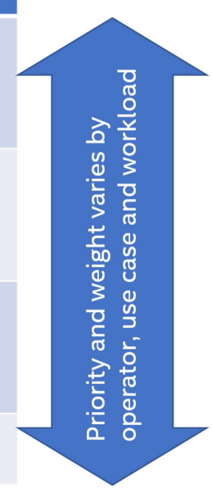
The 3rd Generation Intel® Xeon® Scalable Processor sets a new level of platform convergence and capabilities across compute, storage, memory, network, and security. It includes Intel® Advanced Vector Extensions 512 (Intel® AVX-512) that boosts performance and throughput for the most demanding computational tasks in applications, such as modeling and simulation, data analytics and machine learning, data compression, visualization, and digital content creation. The processor also supports Intel® Software Guard Extensions (Intel® SGX) as hardware-based isolation and memory encryption that provides more code protection to help you develop and deliver more secure solutions.

These hardware ingredients are the foundation for enabling Kubernetes environments as described in the Container Bare Metal Reference Architecture.

Summary

Telecom, public (on/off premise), and hybrid cloud positioning is evolving. Telecoms are moving from private to public and public to private, as well as taking hybrid positions. The motivations for transition in all of these directions are varied and nuanced. Intel recognizes that advising our customers cannot take a 'one-size-fits-all' approach and is endeavoring to understand the universe of criteria and factors that influence these cloud transitions. Not all the factors to determine the workload placement are technical - there are other areas to consider discussed in this paper and summarized in the diagram below.

Area	Capabilities
Business and Services	• Service provisioning and operations • Time to Market • Digital Services enablement • Security and compliance
Technology	• Platform openness and APIs • SLAs • Workloads throughput, latency and performance requirements
Organization and Governance	• Organizational agility • Leadership ownership • Network operations • Governance
Financials	• Total Cost of Ownership • Return on Investment



Telecoms should conduct extra analysis and studies involving their technical, commercial, and architectural teams to determine the workload placement and the impact technically and commercially on the service offering.

The authors believe that the methodology used in this study can be used to certain extent for other topics relevant to cloud unification, emerging placement for workloads, and building future clouds. For more discussion on these topics, please contact your Intel representative.

0920/DN/PTI/PDF 632093-001US