Independent market research and competitive analysis of next-generation business and technology solutions for service providers and vendors

HEAVY READING
WHITE PAPER

Platform for Change:
Building Scalable, Flexible, Secure 5G

A Heavy Reading white paper produced for VMware

vmware®

AUTHOR: JENNIFER CLARK, PRINCIPAL ANALYST, HEAVY READING

INTRODUCTION: THE CARRIER PATH TO CLOUD NATIVE 5G

Cloud native is a tough term to pin down. Heavy Reading sees the goal of cloud native as enabling enterprises to build and run highly scalable and flexible applications for deployment in a public, private, or hybrid cloud. This is accomplished through the use of containers, service meshes, microservices, DevOps and continuous integration/continuous deployment (CI/CD) development practices, and declarative APIs. The key benefits of cloud native include the following:

- Faster time-to-market (TTM) for new services and applications
- Ability to decouple the application from the infrastructure, simplify application development, and enable applications to run in a highly distributed fashion
- Improved, automated, and comprehensive lifecycle management
- Increased cadence of small and regular updates to applications enabled by the microservices architecture and the use of CI/CD
- Lower total cost of ownership (TCO) through the use of containers and microservices, enabling users to deploy only what is needed, rather than entire monolithic network functions

As compelling as these benefits may seem, cloud native represents a fundamental change in the way communications service providers (CSPs) design, deploy, and manage applications and services. It is far more challenging for CSPs to transition from virtual network functions (VNFs) to containerized network functions (CNFs) than it was for them to transition from appliance-based solutions or physical network functions (PNFs) to VNFs.

In this report, Heavy Reading examines key concerns facing the CSPs as they deploy container-based networking—especially in the challenging, highly distributed 5G RAN. We examine the relative advantages of implementing containerized solutions in the 5G RAN today over a virtualized infrastructure versus a bare-metal solution.

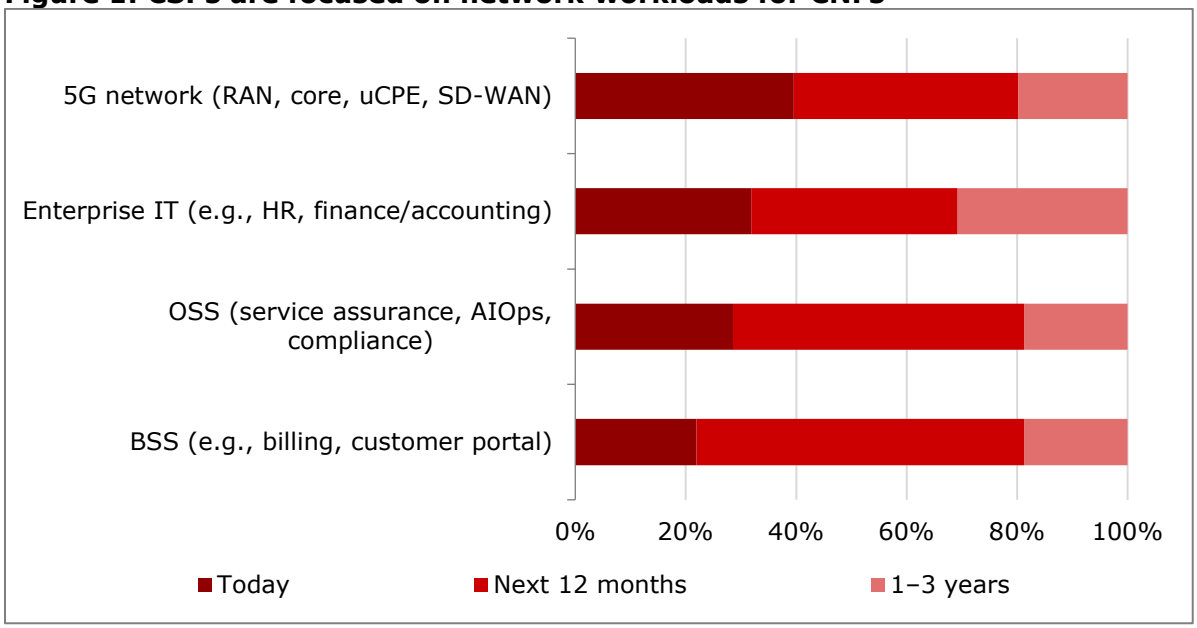
THE MOVE TO CONTAINERS IS UNDERWAY

A cloud native network function (CNF) is a software implementation of a network function that has been divided into microservices, each running inside a Linux container (typically Kubernetes). Containerized microservices communicate with each other via standardized RESTful APIs.

CSPs are committed to a migration to containers as part of their overall cloud native transition strategy. However, carriers are experiencing some of the same challenges they faced in the early days of network functions virtualization (NFV) when vendors dropped minimally altered application software into off-the-shelf hardware and called it a VNF. CSPs are complaining that some vendors are taking the shortcut of dropping an entire VNF into a single container and calling it a CNF—failing to realize any of the advantages of microservices, service meshes, etc.

However, it is still early days, even though Heavy Reading's survey of 92 service providers in 4Q21 shows that containers and microservices are currently being implemented throughout the CSP organization (see **Figure 1**). Many of these implementations are trials, are confined to a few nodes, and/or are for centralized mobile core functions, with the 4G Evolved Packet Core (EPC) and now the 5G standalone (SA) core consistently earmarked in Heavy Reading's carrier surveys as the leading use case in the near term. Of equal focus, but on a two- to five-year implementation schedule, is the move to implement the RAN in CNFs.

Figure 1: CSPs are focused on network workloads for CNFs



Q: When do you plan to deploy container-based workloads in production for the following areas of your business? n=92
Source: Heavy Reading

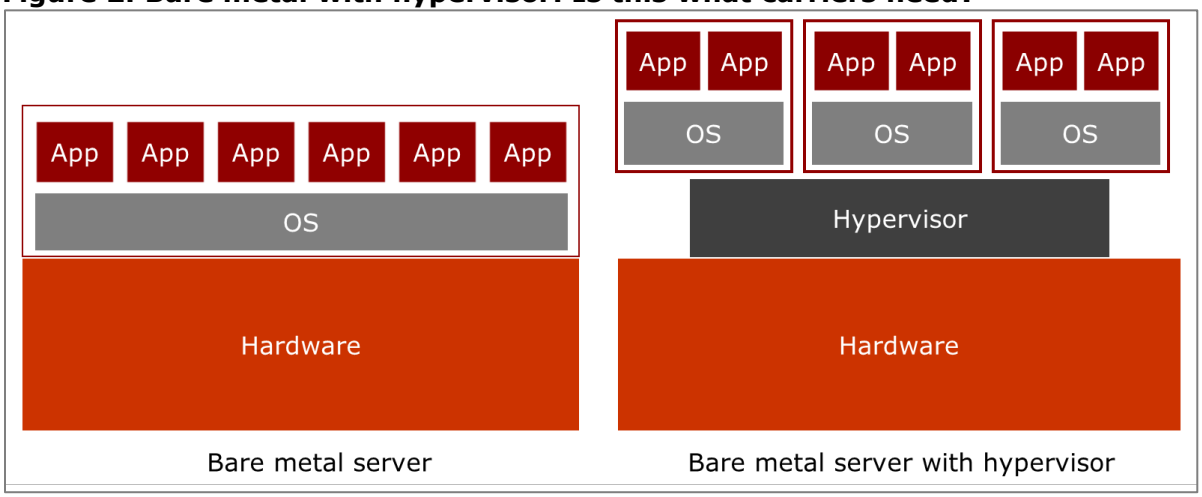
BARE-METAL INFRASTRUCTURE VS. VIRTUALIZATION: EITHER/OR?

Traditional bare-metal servers are dedicated servers in which the OS, whether Windows or a Linux distro, is specific to the platform or server. All user applications run directly on that OS. Bare-metal servers are dedicated to a single tenant and can reside in a hosted data center or an internal enterprise (e.g., CSP) data center. Bare-metal infrastructure uses configuration software, such as Kubernetes, to manage network functions, applications architected as microservices, and functions instantiated as containers. It does not rely on virtual machines (VMs) to host applications. It is important to note that Kubernetes is not virtualization. Kubernetes orchestrates applications at the software layer while virtualization focuses on the infrastructure.

Bare-metal servers with embedded hypervisors

Bare-metal servers assume many of the advantages of a virtualized infrastructure by employing a Type 1 hypervisor (see **Figure 2**). The key difference between Type 1 and Type 2 hypervisors is that Type 1, which this report examines, runs on bare metal and Type 2 runs on top of an OS. The hypervisor creates and runs VMs and provides the administrative interface for managing the VMs. A hypervisor enables one computer (the host) to support multiple VMs by virtually sharing its resources, such as memory and processing. Common hypervisors are KVM, the native virtualization feature of the Linux kernel, Microsoft Hyper-V, and the VMware ESXi hypervisor, which is included in the vSphere platform. Bare-metal infrastructures with embedded hypervisors run applications in VMs with virtualized guest OSs.

Figure 2: Bare metal with hypervisor: Is this what carriers need?



Source: Heavy Reading

The benefits that network operators cite for why they wish to deploy bare metal are similar to those for containerization and cloud native networking in general. Carriers expect the bare-metal infrastructure to be compact, draw less power, be highly automated, and be both easier and faster to deploy. The assumption is also that bare-metal infrastructure will be easier to manage, particularly for highly distributed applications such as edge computing and the RAN.

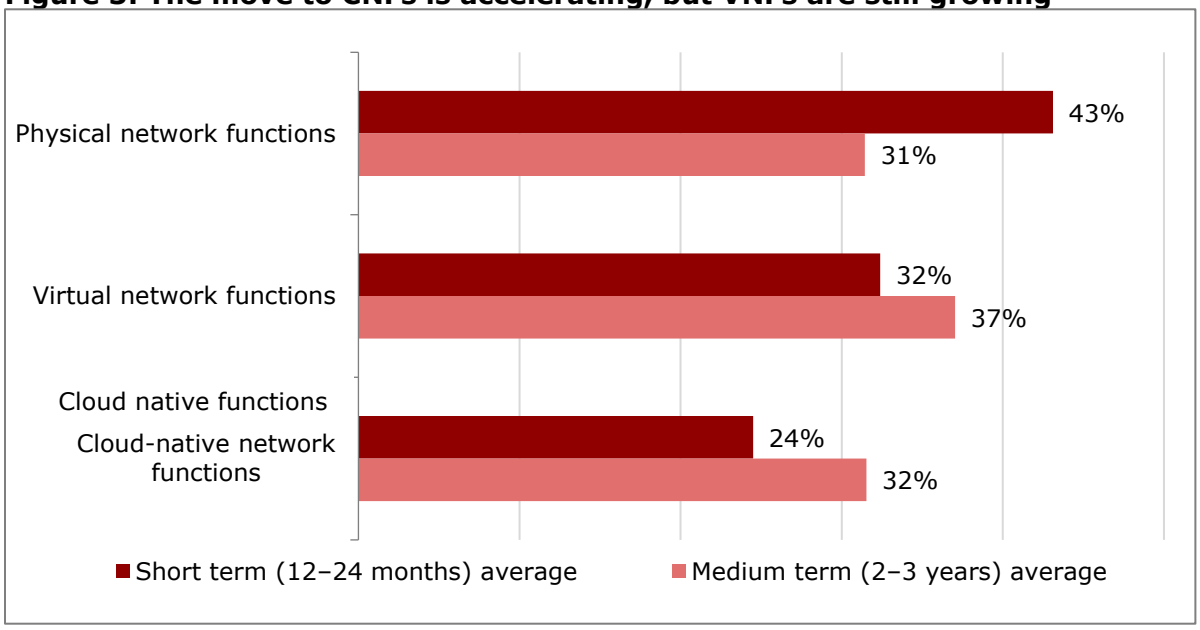
Large, Tier 1 service providers initiated the demand for bare-metal infrastructure (as they did for NFV). The Cloud Native Computing Foundation (CNCF) and Anuket (formed in early 2021 by combining the Common NFVi Telco Taskforce [CNTT] and the Open Platform for NFV [OPNFV]), both parts of the Linux Foundation, are key organizations from a standards and open source perspective.

As the CSPs start to deal with the practicalities of bare-metal infrastructure implementation, they face a variety of hurdles:

- **Leveraging existing investment in virtualization infrastructure:** Carriers have already invested tens of billions of dollars in NFV, and expenditures are still growing. In its *NFV Tracker – 2H21 Analysis Update*, Omdia projects the market for global NFV software, hardware, and services will grow from $26.7bn in 2020 to $45.5bn in 2025. VNFs continue to make up the majority of operator spend in NFV. However, the demand for cloud native solutions and CNFs is driving global NFV infrastructure (NFVI) investment from $2.5bn in 2020 to $3.9bn in 2025. Key drivers for the overall growth in the NFV market are multi-access edge computing (MEC), network slicing monetization, 5G core migration, and 5G RAN. It is critical to carriers with existing NFV infrastructure to extend the useful life of these implementations and use them as a springboard to CNFs and cloud native implementations.
- **Cost of scaling:** Current bare-metal implementations are single tenant and single OS. The vertical vision of bare metal is convenient for vendors, as the stack, like a PNF appliance, is optimized for their own applications. The carrier reality, however, is that they must be able to support diverse types of network (or/and IT) applications requiring different OS configurations and colocated in the same data center. The need to support multiple tenants, multiple vendors, and multiple OSs demands a true, virtualized cloud approach.
- **Managing hardware lifecycles:** Bare-metal infrastructure must be managed in terms of firmware, drivers, and software. While some platform vendors have worked to ease this burden with separate tooling (HPE OneView, Dell OpenManage, etc.), hardware lifecycles become additional overhead that must be managed.
- **CSPs' internal skill sets:** Carriers lack personnel with experience in CI/CD and a DevOps style of development. These software engineers are difficult to find, tough to attract, and expensive to hire. This is a key concern for the carriers.
- **Concerns about the security and isolation of applications on bare metal:** Containers, particularly for highly distributed applications such as the RAN, do not provide a secure boundary and must be protected with a layered security solution.
- **Kubernetes implementation:** Kubernetes is increasingly being used in both VM and bare-metal infrastructure use cases. However, while broadly adopted in enterprise data centers, both large and small, it is not widely deployed in the highly geographically distributed environments of the CSPs. From a carrier perspective, Kubernetes has yet to mature to the point of handling mission-critical workloads in a very large, highly scalable environment.
- **Support for end-to-end network slicing:** This type of slicing demands automation and very flexible infrastructure. Although network slicing is not yet widely deployed, carriers consider it to be essential for 5G overall and for private 5G specifically.

These hurdles are delaying bare-metal infrastructure from becoming mainstream in a carrier environment. However, the issue is not really bare metal versus virtualization. Hybrid infrastructures supporting both VNFs and CNFs will be the norm for the next two to five years (see **Figure 3**). This timeframe covers the most accelerated period of 5G deployment. If, or when, bare-metal infrastructure without an embedded hypervisor gains momentum, the industry will have already moved on to 6G trials and early implementations. Today's rising investments in 5G infrastructure, together with the functional requirements for CNFs needed by the CSPs today, make bare-metal infrastructure without a virtualization layer and hypervisor a deployment for the future or for greenfield network operators.

Figure 3: The move to CNFs is accelerating, but VNFs are still growing



Q: What is the ratio between network function types in your organization's network in the short and medium term? n=65
Source: Heavy Reading

Most NFV implementations, whether the network functions are realized in VNFs or CNFs, are in brownfield environments where operators must virtualize infrastructure, network operations, services, and orchestration across all network layers on top of existing physical infrastructure. This is required from both a functional and a financial perspective—the physical, virtual, and containerized world must be connected and managed as one network, and the investment already made in NFV cannot be cannibalized in the transition to cloud native and CNFs.

MEETING CARRIERS' OBJECTIVES FOR CLOUD NATIVE?

Heavy Reading's carrier surveys over the past five years have provided us with insights into the CSPs' objectives with cloud native. Are these objectives better realized with a bare-metal infrastructure or with a virtualized bare-metal /hypervisor implementation? With this perspective in mind, Heavy Reading discusses the CSPs' top eight objectives below.

Simplify

Carriers cannot absorb the continued and rapid growth in network traffic in RAN and edge locations and in connected devices without simplifying network operations. "Simplify" has become a top objective cited for implementing any emerging technology, often edging out perennial favorites, "performance" and "security." Simplicity is an advantage frequently claimed with bare-metal infrastructure. No hypervisor or VMs, one OS—that is, indeed, a pared down architecture. For applications that are always on with a consistent and high demand for resources, a single-tenant bare-metal infrastructure solution can be the answer. The same is true for applications with regulatory constraints that may demand dedicated hardware resources, such as some defense, healthcare, and utility applications (most of which are unlikely to employ public or hybrid cloud in the first place). These implementations are comparatively static and likely to lean toward a bare-metal infrastructure implementation when transitioning to CNFs.

However, a simplified architecture does not mean simplified operations or better security. In order to support an environment that adapts easily to constant changes in applications/ workloads, the CSPs need to abstract applications and OSs from the underlying hardware. The use of virtualization and a hypervisor allows the server to support multiple tenants and multiple OSs. Frequent software upgrades and rollbacks that are part of a CI/CD style of development can be supported, and customers can run multiple versions of the OS or of Kubernetes simultaneously. In addition, the hypervisor provides another layer of security compared to a bare-metal implementation. With bare-metal infrastructure, once the hardware layer is compromised, there is no additional holistic layer of security and thus there is access to everything residing on the server.

Scale

Overall network traffic continues to double every three years. This staggering growth applies to fixed, fixed wireless, and mobile networks. It will also drill down into the RAN environment: there are an estimated 7 million physical cell sites worldwide, a total of 10 million logical sites, and about 1.5 million outdoor/macro three-sector cell site deployments each year (including refreshes). In addition, with 5G, mobile operators are separating RAN functions into two physical entities: the centralized unit (CU) and the distributed unit (DU). The functional split between the CU and the DU is not fixed and can vary according to any number of factors. In this highly distributed environment—not homogenous in either type of location or workload—carriers need to be able to start small and scale only as needed. Operators cannot afford to leave resources underutilized or deploy a separate bare-metal server for each new site. The use of a highly virtualized environment with VMs and a hypervisor provides a more agile environment where capacity can be shared across VMs (controlled by how the policies governing the VMs are set up). In addition, new sites, in the form of VMs, can be instantiated in minutes when and where needed.

Performance/latency

This is another area where bare-metal infrastructure and dedicated hardware are often assumed to have an advantage by virtue of a lower software load (no application-linked OS, no hypervisor) and a dedicated hardware platform. On virtualized, shared-tenant platforms, there is the threat of neighboring applications on the same server causing disruptions in performance and stability, a problem known as the noisy neighbor effect.

However, performance evaluations of RAN workloads on existing bare-metal (Type 1) hypervisor platforms versus bare-metal infrastructure have shown no performance hit and no increased latency. The fact that these tests were conducted with RAN workloads is noteworthy, as the performance of the DU has been shown to be particularly vulnerable if denied needed compute capacity due to a noisy neighbor problem.

Security

The move to containers and microservices expands the application's attack surface. VM-based infrastructure has strong embedded security as part of the separation of applications into independent VMs. VMs can guarantee application isolation along with high levels of service. In addition, the hypervisor provides another layer of security compared to a bare-metal implementation where, once the hardware layer is compromised, there is no additional holistic layer of security, and there is access to everything residing on the server. Operators planning to deploy bare metal must overcome additional hurdles that affect security, including Kubernetes' lack of infrastructure management maturity, service providers' internal skill sets, and the security and isolation of applications on bare metal.

Manage/automate

There is an expectation that bare-metal nodes will be simple to manage, particularly for edge deployments. This is largely tied to the adoption of Kubernetes with its more declarative nature. Although Kubernetes has matured to become the de facto orchestrator of container workloads, it still has gaps to fill—particularly around physical resource management—before operators can adopt at scale. In addition, Kubernetes is also increasingly used in a VM architecture, making it more of a potential bridge between container and VM solutions, rather than a differentiator.

The scarcity of automation and tooling available for bare-metal infrastructure implementations is a significant challenge for operators, particularly when contrasted with rich management, visualization, analytics, and capacity planning/control available with virtualized solutions. These mature management capabilities also enable better resiliency and faster recovery than Kubernetes on bare metal—allowing the host to be restarted before Kubernetes is aware that the pod has gone down.

Proponents of bare-metal infrastructure recognize that there is a need to focus on automating bare-metal deployments to ensure the industry can efficiently manage the complexities of an open and disaggregated architecture, particularly in use cases like the vRAN or open RAN. These challenges have been addressed with virtualized implementations. The abstraction of application and OS from the hardware simplifies updates and enables the support of multiple versions of the OS and/or Kubernetes. The automation available with a virtualized platform also accelerates the onboarding and instantiation of network functions as well as the deployment of the underlying infrastructure, improving TTM for new services.

Support for multicloud

The limitations and lack of maturity of bare-metal infrastructure from a management perspective are magnified when extending the virtual environment to a hybrid cloud. A multicloud environment requires management, tools, and orchestration that can be used across multiple clouds and managed from a single location. These solutions exist for virtualized environments. Bare-metal infrastructure implementations that support Kubernetes deployments, however, are tied to the host, the host OS, and the cloud host. The tools that simplify the deployment and management of a bare-metal solution across multiple clouds, be they public, private, telco, or edge clouds, are not yet generally available.

Avoiding vendor lock-in

Bare-metal infrastructure is a little odd because it is easily perceived as a step backward in the direction of appliance-based solutions. The greenfield operators that have deployed a bare metal solution have optimized their COTS platforms for their implementations. It is unclear how much latitude they have to change COTS suppliers and how disruptive such a move would be.

The software/application side is even more complex. Carriers have worked through the challenges of implementing VNFs and VNF service chains. They will face the same challenges with CNFs, but they will be compounded by the splitting up of applications into multiple containers and microservices, the accelerated cadence of software updates, and the open APIs that link microservices. Despite these challenges, by abstracting an application and its associated OS from the hardware, CSPs are able to build an ecosystem of CNFs and avoid being locked into a limited set of one or two vendors.

Lower TCO

Heavy Reading has observed that multi-tenant, multi-OS virtualization solutions optimize server investment. While support for multi-tenancy on bare-metal infrastructure is being explored in standards workgroups, deployments are likely to remain immature for several years. Greater container pod density on virtualized platforms means more efficient use of hardware resources, which is non-trivial both from a capex perspective and in dealing with the current global chip shortage (and resulting supply chain delays).

Fewer servers on a virtualized infrastructure mean savings in opex as well as capex due to the smaller footprint and reduced power draw. There is more at stake than lowering cost. All major carriers today have sustainability goals aimed at reducing their carbon footprint. Telefónica, Vodafone, BT, and Verizon, to name a few, have all pledged to be carbon neutral—some as early as 2030 and all by 2045.

CONCLUSIONS

Virtualization allows carriers to leverage cloud economics by sharing a pool of resources. Bare-metal infrastructure, on the other hand, specifies that resources be dedicated to specific applications/ functions. The addition of virtualization via a hypervisor to bare metal allows operators to scale up and down, distribute workloads, and plan for peak traffic without nailing up resources. By continuing to use VMs, operators can leverage day two automation, management, optimization, and interoperability with an expanding ecosystem of network functions available for download. A move to bare-metal deployments tacks on an integration tax to compensate for the scarcity of management, multicloud, support, and automation tools and capabilities. Operators will adopt bare-metal infrastructure as the technology, security, and internal resources allow or as required for specific, constant demand, relatively static use cases. The investments that carriers have already made in NFV, along with the management and tooling expertise that they have acquired, will encourage them to deploy containers in a virtualized, VM-enabled platform.