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

HEAVY READING

WHITE PAPER

Configurable Service Capacity:
The Future of VNF Management

A Heavy Reading white paper sponsored by F5 Networks

AUTHOR: JIM HODGES, PRINCIPAL ANALYST, HEAVY READING

INTRODUCTION

A fundamental premise of network functions virtualization (NFV)-based networks is that software functions must be executable anywhere in the cloud. To support this vision, NFV standards define virtualized network functions (VNFs), which enable any network function to execute in a distributed software environment as a virtual machine instance.

However, while VNFs are ubiquitous in that they will need to support a mix of edge and cloud instantiations depending on service requirements, they must be implemented with a flexible but rigorous suite of management techniques to optimize hardware resources and ultimately customer experience outcomes.

Therefore, VNF management, which is accomplished through the use of a VNF manager (VNFM), represents a core element of any successful network implementation. While the functions of the VNFM should in theory be relatively static, since reference architectures are now well-defined, in reality the dynamic and multi-faceted appearance of the cloud and the shift to "service-aware" networks mean that VNFMs will continue to evolve – hence the need for flexibility and programmability.

Accordingly, this white paper documents both the current and future role of the VNFM, including how it is evolving from simply managing VNFs to focusing on the optimal approach to scale individual service instances.

CLOUD VNF MANAGEMENT & ORCHESTRATION

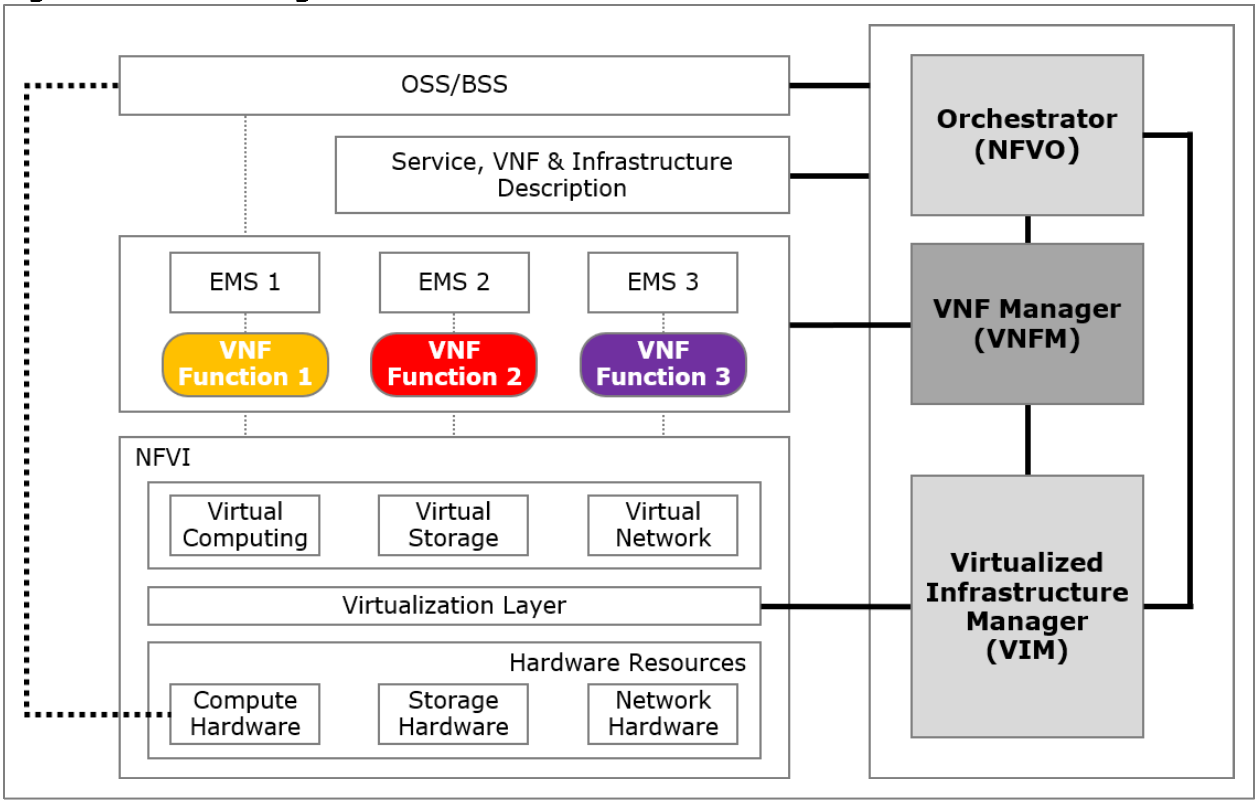
The concept of a VNFM for a telco cloud first emerged with the release of the seminal NFV introductory white paper in October 2012. This white paper – authored by leading service providers such as AT&T, CenturyLink, China Mobile, Colt, Deutsche Telekom, KDDI, Orange, Telefónica, Telstra and Verizon – even back then had the foresight to realize that breaking traditional telco services into software instances would require careful orchestration and lifecycle management.

Hence, when the first release of European Telecommunications Standards Institute (ETSI)-based NFV specifications were released in 2013, they formally defined the VNFM and other related functions, such as the Management and Orchestration (MANO) function and Virtualized Infrastructure Manager (VIM), to enable vendor-agnostic interworking. This architecture is depicted in **Figure 1** (next page).

In the ensuing years since the completion of the formal standards definition process, the VNFM has continued to be a key component of any NFV deployment, since it controls instantiation, onboarding and management of VNF lifecycles.

While some may consider these functions to be simply performing basic and static VNF "plumbing" functions, we adopt a different stance since, like all things cloud, the technology curve is so dynamic that even functions that were defined only a few years ago are being reshaped to support the implementation of more sophisticated management functions.

Figure 1: VNF Manager Reference Architecture



Source: Heavy Reading

Two factors are driving this charge: cloud-native adoption and 5G definition. In response to these factors, there are two distinct corresponding phases of NFV deployment (although not formally defined). In the first phase, the focus was on porting existing feature software from underlying hardware platforms and optimizing the code to a more layered approach, creating distinct VNF software partitions. While this is still valuable, in that it leverages the cloud for scale and empowers rapid service onboarding, it also inherits and pulls forward some of the limitations associated with managing legacy non-virtualized physical network functions (PNFs).

Accordingly, the focus on NFV deployments has now shifted to the creation of new software applications that decompose existing telco services into distinct service strata that can be assembled and reused as service "building blocks." This approach results in the creation of microservices, which without question represents the future of service delivery. Microservices not only possess superior scale and reusability attributes, they are optimally suited for supporting edge deployments due to their small single-instance software footprint.

Another related and significant consideration in this push to adopt microservices is that they support a more service-aware architecture, which forms the core of 5G networks. For instance, 5G standards specifications define and assume that 5G services will be more distributed and focused on API exposure, so that software applications can be created by a much broader software ecosystem and not constrained by single-vendor software delivery cycles.

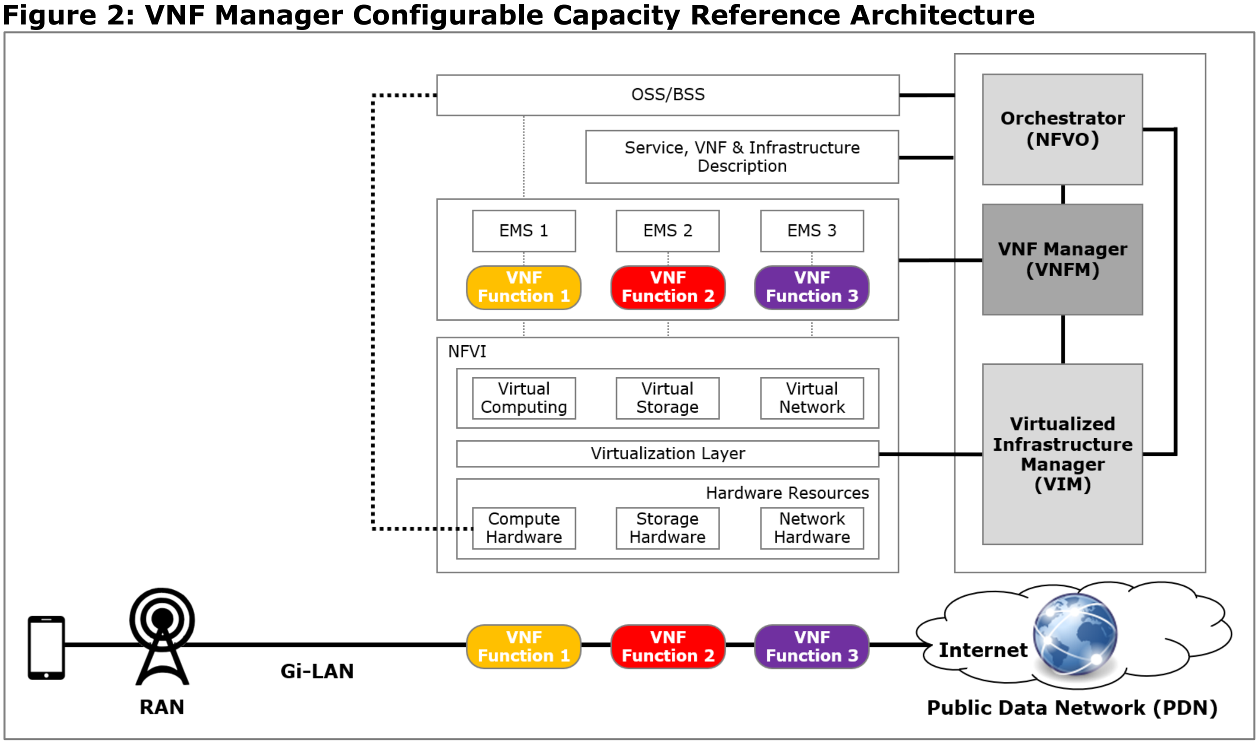
This model, which was not defined when NFV first emerged, can be considered a leading example of how the industry in a very short period of time has taken steps to fully optimize the cloud's potential to deliver massive scale, with low-latency services not possible in the traditional IP-based telco domain. However, it is important to note that 5G and the Service-Based Architecture (SBA) model it defines is access-agnostic and will, therefore, drive this API exposure model into fixed and cable networks – essentially any network that is virtualized.

Given this profound services transformation, there is now a much greater focus on the role of VNFMs and optimally managing more sophisticated service delivery demands. For service providers, this translates into the need to adopt a more flexible software-based VNFM that can integrate policy and analytics to optimally apply policy control as well as use network capacity more effectively. This entails not simply implementing a pooled resource approach, but also implementing a configurable model that can be used for *any service VNFs* – based on network requirements. We refer to this approach as the configurable capacity service model.

This approach is now more important than ever, because networks are now running a much broader range of services. Moreover, these services by default can only be classified as high-bandwidth consumption services, which makes the ability to support configurable capacity for services running on the Gi-LAN even more critical. Since the Gi-LAN is the primary access point to the Internet, it also represents the interface where policy management enforcement and traffic steering optimization must take place. However, it's important to note that the support of configurable capacity is not simply applicable to Gi-LAN services; it is now foundational to any advanced cloud-based services, including 5G network slices.

At the heart of this new management philosophy is the ability to apply more advanced modeling techniques not only to enable the scaling of VNFs anywhere in the network, but also to support the modeling and impact definition of VNF service instances on the computing resources wherever they are executed. In turn, given that these services are VNF-based, new, more advanced and programmable VNFM capabilities are rapidly becoming mandatory to support advanced orchestration and VNF lifecycle management, as shown in **Figure 2**.

Figure 2: VNF Manager Configurable Capacity Reference Architecture



Source: Heavy Reading

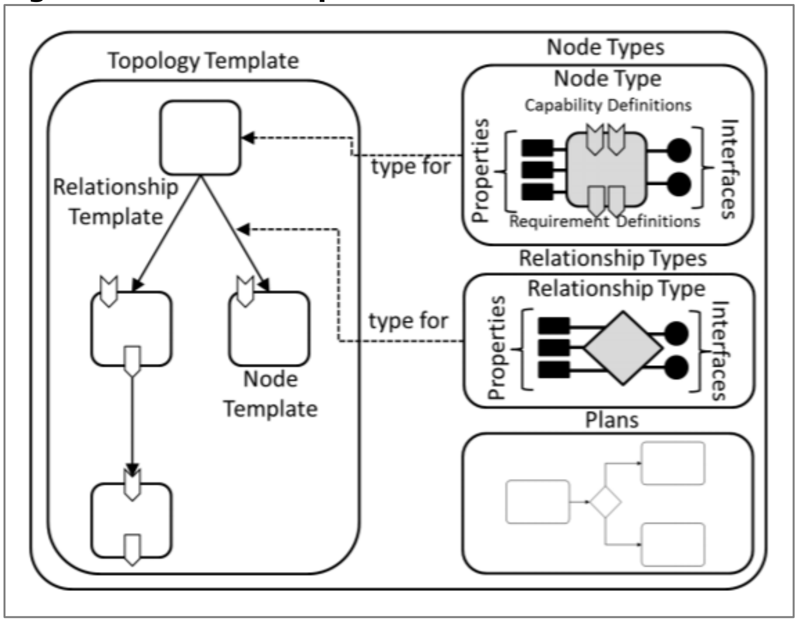
IMPLEMENTING CONFIGURABLE CAPACITY MODELING

In order to adopt a configurable capacity model, the implementation of advanced templating capabilities is essential. This section of the paper considers not only the role of templates and their relationships to VNFMs, but also the detailed relationships to service execution and composition. Service composition is now becoming increasingly important for several reasons. On a basic level, templates are now considered an integral component of service composition, since they provide a blueprint for how each service is constructed. For example, in a cloud-native microservices context, templates will document all the sub-components that make up the service execution model. These sub-components include not only the microservices, but also the types of virtual machines used, whether software containers are used (and what type of containers are specified), which orchestrator(s) are supported, and the interworking capabilities of third-party and potentially open source microservices.

Templates are also used to document the lifecycle flow of any cloud-native cycle. In essence, templates can be considered as providing a total end-to-end visual topology representation of any cloud-native service. This includes defining relationships with other services, the compute requirements necessary to support and scale, interworking potential, specific orchestration requirements, software extensibility tools supported and lifecycle management processes.

To support such a diverse set of vital requirements, templates must be standards-based to foster interworking and vendor-agnostic service onboarding through the adoption of a common descriptive language. Several standards have defined such templates, including TOSCA (Topology and Orchestration Specification for Cloud Applications), an open source standard developed by Organization for the Advancement of Structured Information Standards (OASIS); and OpenStack HEAT templates, which support a translator to integrate TOSCA templates as well. Templates such as those based on TOSCA, as shown in **Figure 3**, fulfill these requirements by creating a flexible structure for documenting the key elements of any service: node type, relationship type, properties and interfaces.

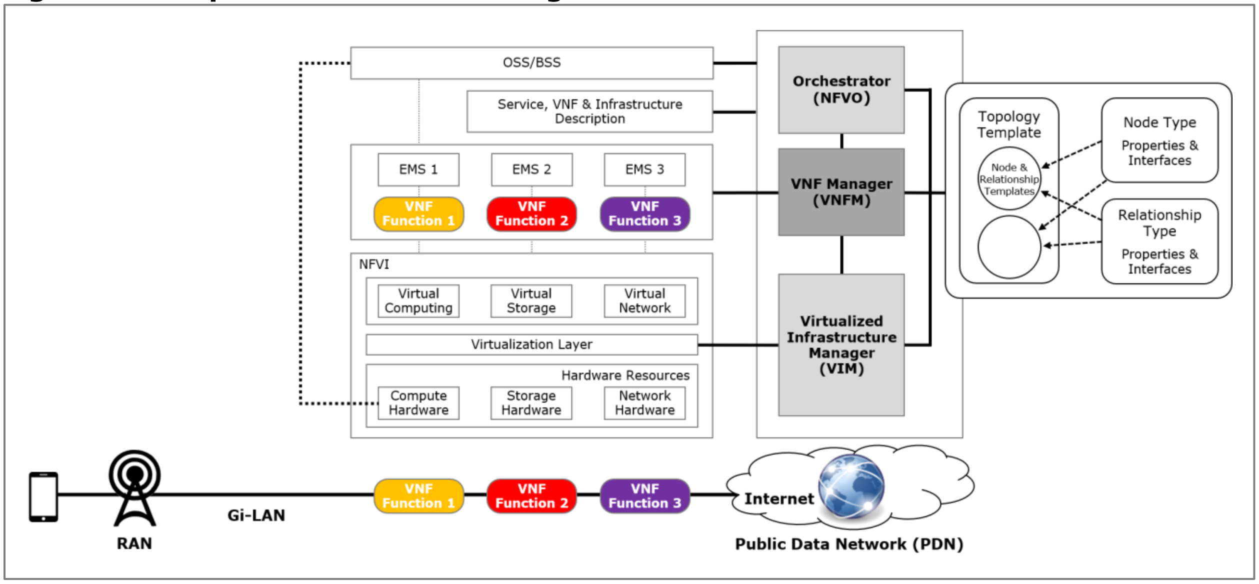
Figure 3: TOSCA Template-Based Structural Elements



Source: OASIS – TOSCA v1.0

In turn, to fully operationalize the benefits of template-based modeling, VNFMs must also be designed to support the seamless managing and modeling of template-based services and capacity configurations, as shown in **Figure 4**.

Figure 4: Template-Based VNF Manager



Source: Heavy Reading

The impact of infusing template-based intelligence into the VNFM is profound. In addition to maintaining responsibility for current functions such as managing VNF instantiation, scale and lifecycle, with templating the VNFM is positioned to support these functions in a specific service and relationship context.

This also empowers the VNFM to assist the network in executing performance capability decisions aligned with an intent-based automated and configurable capacity network performance model that is not currently supported but will be increasingly necessary as communications service providers (CSPs) take steps to commercialize 5G services.

CONFIGURABLE CAPACITY MODELING BEST PRACTICES: TEMPLATES FOR SUCCESS

The advent of template-based configurable capacity models represents a significant step in the maturation of the cloud. In response, the implementation cycle of this new, advanced type of VNFM requires new approaches and best practices. In a best-practice context, we view the associated decision points as falling into five specific categories.

Best Practice 1 – Focus on Services Capacity Monetization

As we have noted, cloud adoption and commercialization are ushering in a new era of service innovation. And in many ways, we have yet to reach even the base of this arc of service innovation. So, unlike legacy networks, in implementing an advanced VNFM, it is crucial to assume that fluid customer service and associated capacity requirements are the new norm.

Of course, the challenge here is that CSPs must not only consider this volatile service model, but also be able to most effectively monetize these services via more flexible billing models. For instance, some customers are now focusing on purchasing realized capacity instead of statically predicted capacity, so they can use this capacity for new services introduction utilizing an on-demand model.

This approach will only gain wider acceptance, since it is well-suited to cloud's distributed scale model and is aligned with 5G's services-based and programmable capacity-specific, slice-based architecture design. The end result is that the VNFM must also become capacity- and service-aware.

Additionally, it must be noted that the service landscape is now much broader as well. Because of techniques such as 5G slicing, CSPs' focus on new revenue streams is moving well beyond the telco services domain, to *any* domain to address adjacent and untapped service markets such as IoT and connected car.

In these segments, since CSPs will be by default supporting service activation, security enforcement and capacity provisioning, there is much less of a risk of them falling into the "dumb pipe" provider model that characterized the 4G/over-the-top (OTT) period.

Finally, CSP capacity monetization must also address opex and capex reduction. In other words, the entrenched principle of matching force to load still represents a fundamental tenet for cloud operators, except in this case they must adopt new approaches to optimally match this cloud-based load to virtualized resources to optimize capex and opex levels.

Best Practice 2 – Consider the Impact of Hybrid Networks

Even though the cloud undoubtedly represents the growth and service innovation engine, it will take time for CSPs to methodically migrate all services and data traffic off existing networks. Thus, not only will VNFs be distributed, there will also be a mix of VNFs and PNFs running in a hybrid environment.

While PNFs don't have the same scale characteristics as VNFs, since they will play a role in white-box deployments, they must also be seamlessly integrated into the corresponding cloud environment. Even though they don't have the exact same orchestration and management requirements, they must still support an end-to-end service delivery model.

While this hybrid environment must be supported by the VNFM, and it must also be intelligent enough to apply a configurable capacity model to hybrid networks. This is important to enable onboarding and management of new VNFs to provide configurable capacity when PNFs cannot scale due to cost or time constraints. Stated another way, the realization of effective and configurable VNF capacity management constitutes a significant step in workload harmonization.

Best Practice 3 – Prepare for a Services Anywhere Model

Services are now becoming distributed and supported at the edge more than ever before, which is driving new requirements for VNFMs. As a result, the value associated with modeling templates will also become increasingly important, since the boundary between Node

Type, Relationship Type and Topology Template depicted in **Figure 3** is now effectively unrestrained.

This means that modeling must also be flexible and extensible, and be paired with a VNFM that can manage a fully distributed environment.

Best Practice 4 – Factor in Continuous Performance Improvement

One of the fundamental attributes of a cloud network is that in response to service requirements, performance requirements will be subject to continuous improvement pressure.

One way to accomplish this in a VNFM design context is to adopt a design model that is capacity-based to share resources, rather than relying on dedicated resources for specific VNFs. This design attribute – which is discussed in greater detail in the next section of the paper – is crucial, since this continual performance improvement thrust is driven by the relentless push to reduce network speeds and latency thresholds.

For the VNFM, this translates into a requirement not only to manage relationships in a boundaryless landscape, but also to perform VNF management at a very high packet rate with extremely low latency, which consumes a lot of capacity. In this scenario, the VNFM can also play a role in directing traffic to the right servers for stateful services to optimize server performance.

Best Practice 5 – Develop a Cloud-Agnostic Strategy

A final measure of flexibility that must be factored into the creation of best practices is the need to develop a cloud-agnostic strategy. Specifically, we are referring to the decision to rigidly commit and implement only one type of standards-based, VNFM-compliant manager.

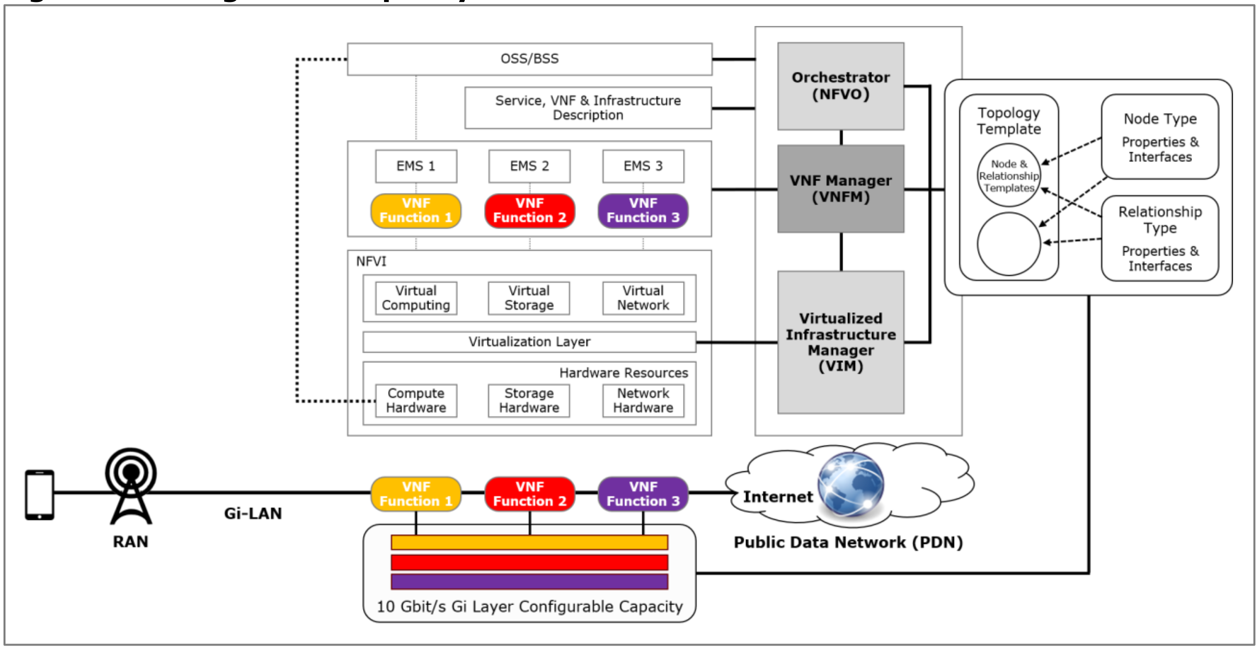
Although we acknowledge there is an initial decision point to implement either an ETSI- or ONAP-based VNFM to minimize complexity, the VNFM design methodology must not be hard-coded to support only a single approach. This is particularly important since as we have seen, the VNFM will play a fundamental role in management of 5G slices, which will likely evolve beyond the CSP's private clouds to the need to manage VNFs in public clouds as well, which could rely on either standards-based approach.

Consequently, the bottom line is that regardless of which standard a VNF implementation is based upon, the VNFM must be flexible enough to support either alternative agnostically, by possessing a rich portfolio of extensible software-based tools that enable it to adapt based on market adoption forces.

CONFIGURABLE SERVICE CAPACITY USE CASE

The ability to configure service capacity represents a game-changer for how the cloud is optimized and ultimately monetized. However, as illustrated in this case, the monetization process is not complex. As shown in **Figure 5** (next page), it is optimally suited for meeting customer billing requirements. For example, as captured in the use case, this model enabled by an advanced VNFM supports the ability to optimize not only cloud capability, but also Gi-LAN capacity.

Figure 5: Configurable Capacity Use Case Reference Architecture



Source: Heavy Reading

Specifically, what is depicted in this use case is the ability to dynamically allocate 10 Gbit/s of Gi-LAN bandwidth to specific VNFs by utilizing the VNFM to manage this bandwidth allocation, based on specific node and template requirements modeled for each specific service VNF. With this straightforward but powerful capability, it is possible to avoid the current dilemma in which the 10 Gbit/s of capacity is statically applied to specific VNFs, which results in monetization challenges associated with overprovisioning bandwidth.

Moreover, enabling configurable capacity via the VNFM is also optimally aligned with customer billing expectations. For example, as we have touched upon, CSP customers are now starting to demand the implementation of more flexible capacity-based billing models. In this use case, since each VNF has been configured and allocated bandwidth based on template-modeled requirements, it is now possible to bill based on actual bandwidth consumed for specific services.

In other words, the adoption of configurable capacity-enabled VNFMs will make it possible for customers to purchase capacity based on service mix and real-time requirements, thereby enabling CSPs to further distance themselves from the 4G "dumb pipe" impasse.

This capability will also enable CSPs to adapt and implement billing patterns for individual high-value, low-latency services that are now emerging. An associated but also important consideration is that CSPs can also streamline the process of selling or even upselling bandwidth to their customers.

In this scenario, due to the configurable capacity model, CSPs can easily sell more capacity to customers based on their service requirements. In this case, the value proposition is actually two-sided: The CSP doesn't have to wait for a formal service request from the customer, while the customer is billed for exactly the capacity they consume, rather than some form of non-configurable, threshold-based model.

CONCLUSION

Regardless of which manifestation a fully commercialized cloud assumes, the requirement to configure network capacity to service requirements must be considered a core trait for measuring implementation efficacy. Therefore, the management of VNF-based services must evolve to meet the demands of this new services archetype. In response, as documented in this white paper, key management elements such as the VNFM are continuing to advance through the adoption of new capabilities that support the introduction of a flexible and functional template-based, configurable capacity model.

ABOUT F5 NETWORKS

This section offers an overview of F5's product portfolio. The information in this section was provided by F5.

Simplicity, Velocity, Scale: F5's Ready-to-Install NFV Packaged Solutions

Deployment and management of individual VNFs has proven to be a significant challenge for service providers. To help solve this problem, F5 has developed a set of unique pre-packaged, preconfigured NFV solutions that simplify the virtualized network.

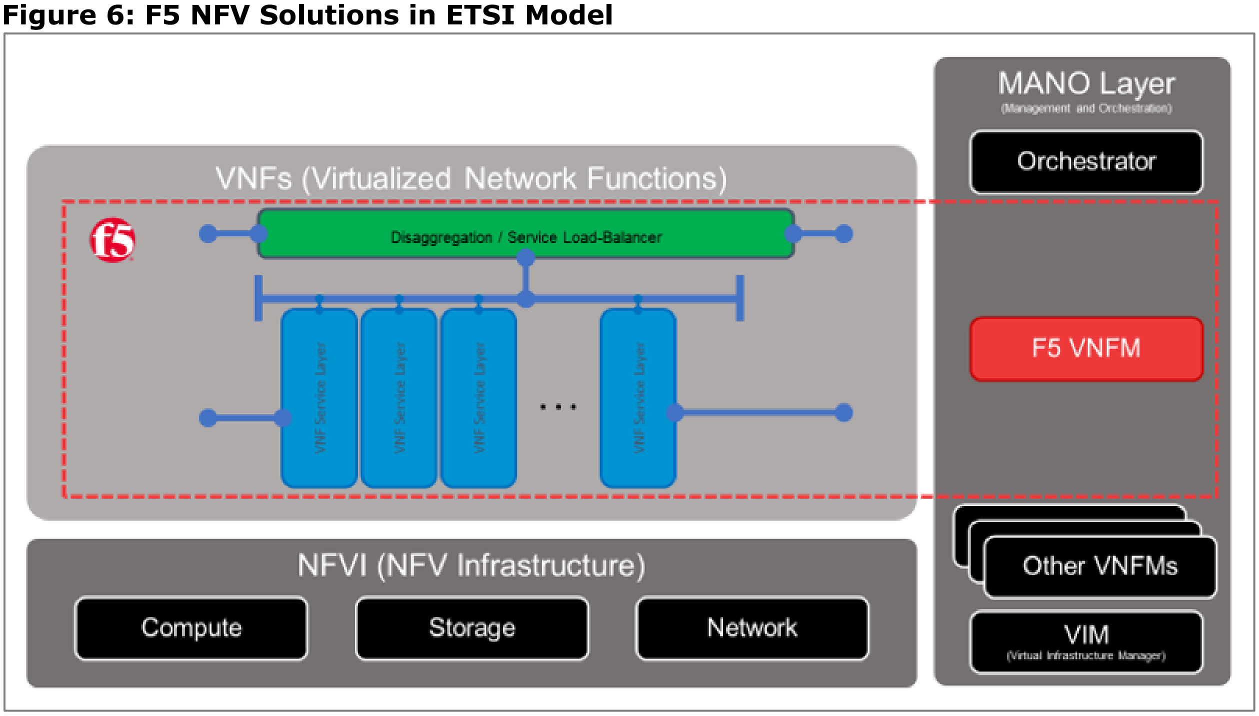
These ready-to-install solutions automatically unpack and self-configure, saving time and ensuring application performance out of the box. F5's NFV solutions include the new F5 VNFM, which delivers fully automated, end-to-end lifecycle management from service instantiation to auto-scaling and decommissioning.

F5's NFV packaged solutions deliver an automated NFV solution, providing greater business agility and accelerated new service deployment. These packages help simplify network planning, sizing and deployments with a unique capacity-based consumption model, and our use-before-buy model simplifies purchasing and reduces costs.

F5 offers two NFV solutions – the F5 Gi Firewall VNF Service and F5 Gi-LAN VNF Service. These are delivered in 5, 10 and 50 Gbit/s increments. This consumption-based model allows you to provision based on outcomes rather than instance licenses.

Figure 6 illustrates where F5 NFV solutions fit in the ETSI NFV architecture.

Figure 6: F5 NFV Solutions in ETSI Model



Source: F5

F5's VNFM controls the lifecycle of the VNFs. It automates deployment, drives configuration and manages autoscaling, updates and upgrades. It helps reduce overprovisioning, efficiently move capacity from one data center to another, and quickly spin up, spin down or add new resources. The F5 VNFM is TOSCA-based, ONAP-aligned and interfaces with northbound orchestration.

F5's Gi Firewall VNF Service provides flexible and powerful protection from attack for the network and subscribers. The Gi-LAN VNF Service allows integration of key S/Gi network functions into one platform, helping to reduce TCO by deploying far fewer devices and streamlining the network. Multi-layered security can handle the most demanding network attacks, and the policy enforcement capabilities help generate new revenue streams.

For more information on F5's NFV product portfolio: [f5.com/nfv](f5.com/nfv)