SANDVINE

# Deploying Virtualized DPI at multi-Tbps scale

## KEY HIGHLIGHTS

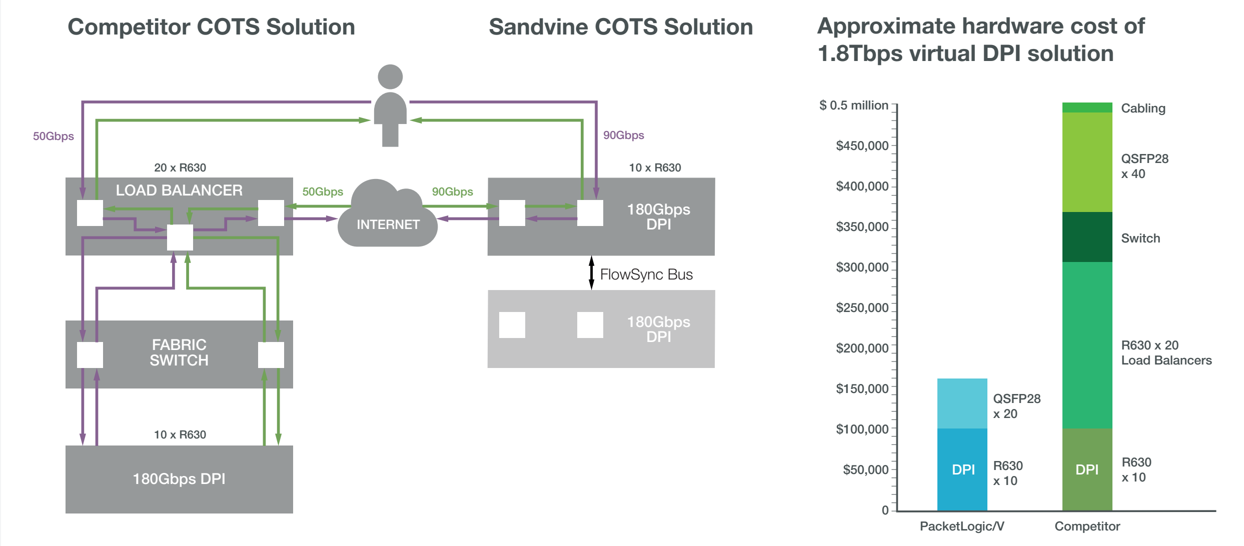
- Delivers 180Gbps in a single 1RU footprint and up to 1.8Tbps in a 10RU footprint with 57% CPU per system **using 100GE ports**
- No extra VNFs to create artificial symmetry with load balancers – cutting CPU and networking infrastructure requirements by half
- Easy path for upgrades and scale-out using live, stateful migration between systems without requiring minutes of downtime
- Utilizing "pizza box" COTS takes the hardware cost to the $50 per Gbps mark

## How do you build a network where infrastructure costs can be reduced as data traffic increases, without compromising QoE?

**A simple question, but increasingly a challenge for network operators in the era of Facebook, Netflix, and massive volumes of encrypted traffic. Network operators around the world have begun to invest in Network Functions Virtualization (NFV) and Software Defined Networking (SDN) in order to reduce cost and increase flexibility of their infrastructures. The journey from purpose-built or proprietary hardware to full Commercial Off the Shelf (COTS) hardware is still in progress.**

The nirvana of a fully orchestrated telco cloud network still faces significant scaling challenges to be able to service the multi-Tbps networks deployed today – much less the increased scalability that is expected with 5G and the rapidly escalating bandwidth used by consumer broadband connections.

Figure 1




**Why does the competitor solution need 20 boxes just to do load balancing and then just 10 boxes for DPI?** This is because each box has just 200Gbps of PCI I/O available. When load balancing the traffic, you see each packet twice - once on its way to the DPI, and again coming back.

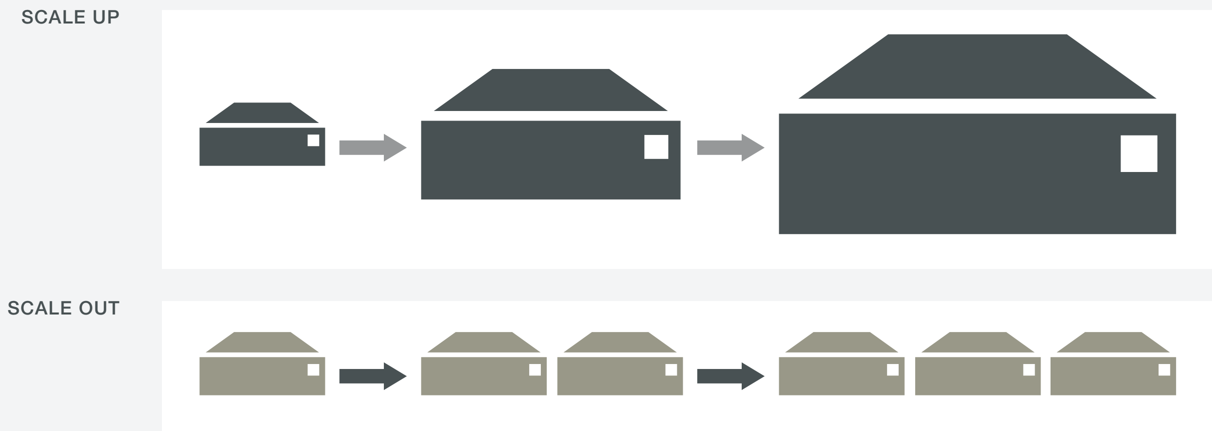
## CHALLENGE #1: CPU AND I/O EFFICIENCY

The first challenge is to deliver on the efficiency promises of NFV by minimizing the number of CPUs needed to process packets while also reducing the number of network interfaces needed for a NFV deployment. Solutions that rely on dedicated load balancing functions will be wasting valuable I/O and CPU resources that could be put to better use. This is made worse by the fact that most COTS systems are severely I/O restrained, leaving expensive CPU resources unused.

## CHALLENGE #2: SCALE-UP AND/OR SCALE-OUT WITH STATE AWARENESS

Some functions in the packet core, like traditional DPI, CGNAT or the PGW function are significantly more challenging to scale up and out. For other services, like routing, you can typically throw more ports at the problem to get more capacity, but for these services you quickly run into a situation where you're building load balancers and sending packets back and forth, consuming valuable I/O and CPU resources. And as you want to scale out, this problem becomes worse, because each new instance is interconnected with every other function, not scaling linearly. Which does not bode well for terabit scale cores.

Figure 2




## OUR EXPERIENCE

- Sandvine has been tackling some of the biggest challenges in the virtualization and NFV market since 2013, when we introduced PacketLogic/V
- Over 50 successful virtualization deployments worldwide as of December 2016.

## DEPLOYING VIRTUALIZATION AT MULTI-TBPS SCALE: A PERFORMANCE CASE STUDY WITH APPLICATION INTELLIGENCE, LEVERAGING DEEP PACKET INSPECTION TECHNOLOGY

Sandvine wanted to provide service providers with a performance test example that they could use to help guide their network planning as they seek to implement 100GE on COTS hardware, which brings COTS to the level of performance where it can compete with existing hardware solutions from vendors.

Until recently, this performance challenge was a barrier to entry for COTS hardware, as the only way to get enough I/O to scale up performance was either by using chassis based server systems, or using 40GE connectivity with 10GE breakouts to get more ports. However, many operators have already moved to 100GE at the core, and sometimes the edge of their network; 100GE support in a virtual offering substantially expands the addressable market for deployments.

100GE DPI is one of the most compute-intensive solutions on the market and one of the high priority VNFs many operators seek to virtualize in order to enhance deployment flexibility and reduce cost.

# SYSTEM CONFIGURATION

## Hardware

**COTS Platform**: 10 x Dell PowerEdge R630

## Each Platform

**Compute**: Intel Xeon E5-2697 v4 2.3GHz
**Network**: Mellanox ConnectX-4 Dual Port, QSFP28-based 100GE
**Storage**: 2x400GB SSD
**Memory**: 256GB DDR4
**Cost**: Approx. $10,000

## Software

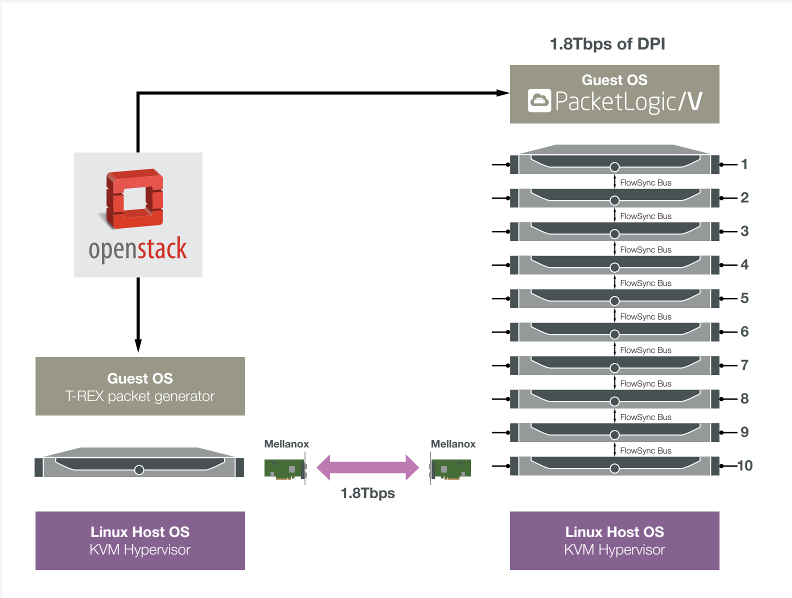
**Host Software**:
Debian 8 (Linux 4.7) KVM (2.7.0)
**PacketLogic/V Software**:
PacketLogic Version 17.1
**Core Usage**: Host OS (1 Core)
PacketLogic/V (17 Cores) w/one instance of PL/V per socket

The goal of this performance test was to demonstrate to operators, two specific virtualization dimensioning scenarios:

1) Scale Up to support line rate on 100GE links on a single system
2) Scale Out to support 1.8Tbps of traffic total across multiple systems with 100GE links

Figure 3

NETWORK DIAGRAM:




## TEST CASE 1: SCALE-UP SINGLE SYSTEM PERFORMANCE

The first test case utilizes a single system with 4x100GE ports, maximizing the throughput for a single system. The packet generator saturates the 100GE links with as much traffic as the PacketLogic/V can process without dropping packets. Each system has two instances, one on each CPU socket and each CPU is configured with one core for the hypervisor, and seventeen cores for PacketLogic/V. Each instance of PacketLogic/V is performing full application identification and is configured for traffic management and full analytics collection.

**The performance results for the single system test are**
**Raw Throughput:** 180Gbps
**PPS:** 22.2Mpps
**Per Core Throughput:** 5.29Gbps
**CPS:** 150k/s
**Concurrent Connection Count:** 6M
**CPU load:** 57%

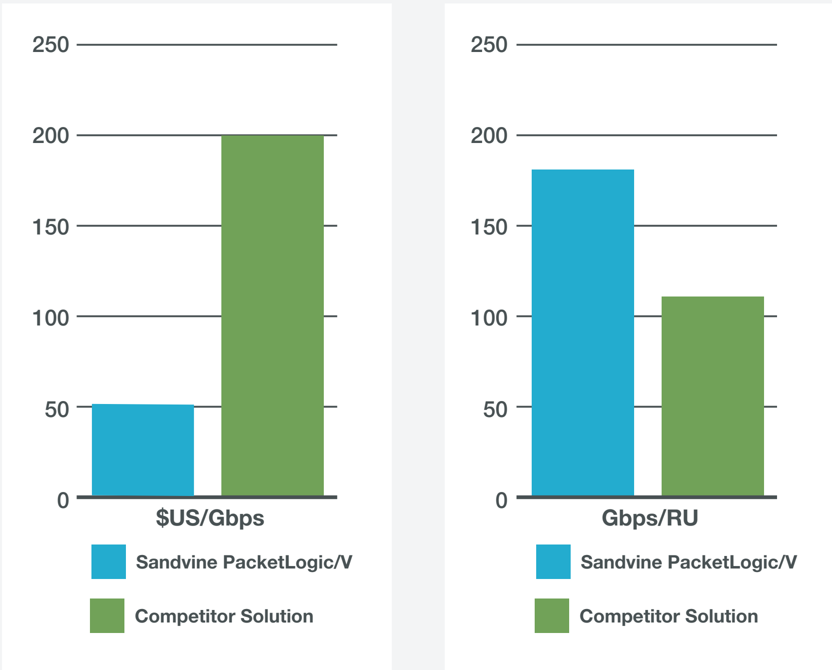
## INTERPRETING THE TEST RESULTS

Results demonstrate the scale-up performance achievable on a single, high performance, 1RU COTS server using a 100GE interface that is shipping today. These results will help give guidelines for operators that are migrating to 100GE interfaces and want to maximize throughput per rack unit. The CPU load for each instance indicates that the system is capable of significantly more throughput, or to deliver additional services in the same footprint (some use cases like Carrier Grade NAT will require more CPU cycles than a base configuration). This means that an operator can deploy on a smaller number of cores and share the COTS hardware with other VNFs.

It is worth noting that there are no additional VNFs needed in this test, no load balancers to distribute flows to individual CPUs or cores. This is a significant factor that we will discuss later in the paper. This also significantly reduces the cost per Gbps for deploying DPI-based solutions. The chart below shows comparison to another recent test result from another vendor.

Figure 3

**With PacketLogic/V's cost performance ratio of 4:1 and a Gbps/RU advantage of almost 2:1, the benefits to the operator become significant for large-scale deployments.**




## TEST CASE 2: SCALE-OUT 1.8TBPS SOLUTION

The second test case featured ten systems stacked (for a total of 10RU) in a rack to demonstrate what an operator could get if they needed to scale out to 1.8Tbps of capacity. In a scale-out deployment, the goal is to add VNFs (which in this case translates into additional servers) until the required performance can be met. The goal for operators is to minimize the footprint needed to meet their performance requirements. In this case, 1.8Tbps was chosen for convenience, however with a Sandvine solution, it is as simple as taking the single system or the single core performance and adding VNFs until you reach the desired performance.

**1.8Tbps System Performance**
**CPS:** 1.5M/s
**Concurrent Connection Count:** 60M
**PPS:** 222Mpps
**Raw Throughput:** 1.80Tbps
**Per-Instance CPU load:** 57% (indicating that more capacity could be supported with more I/O)

## INTERPRETING THE TEST RESULTS

These test results demonstrate the scale-out performance achievable on multiple, high performance 1RU COTS servers using a 100GE interface that is shipping today. They guide operators when scaling up to support multi-Tbps network deployments. As in the single system test, the CPU load for each instance indicates that the system is capable of significantly more throughput, or delivering additional services in the same footprint. Again, notice that in this test case, no additional VNFs are needed to scale the performance out for the deployment.

# Solving NFV Scaling Challenges

**We spoke of two challenges in the introduction section of the whitepaper – scale-up and scale-out and CPU and I/O efficiency. The test results above start to build a solution that can meet these challenges for the operators while the underlying PacketLogic technology solves several unstated problems that we will now explore:**

## I/O AND CPU CONSERVATION

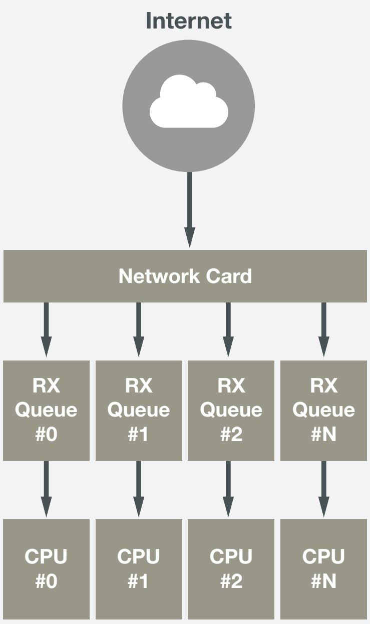
Explicitly mentioned in the challenges section is the goal of maximizing CPU and I/O efficiency. In both test cases, it was mentioned that neither configuration required additional VNFs to scale up the performance. What exactly were we driving at with this comment?

Each of the multiple scale-out systems mentioned above has 4x100GE ports, and these ports efficiently consume all the available I/O for the system and the underlying PCI bandwidth.

Many stateful processing implementations require an additional load balancer function to distribute flows across systems, CPUs and cores to ensure that all traffic from a single "subscriber" is sent to the same CPU core. In hardware-based systems, this is done inside the chassis system on the switch fabric or a dedicated CPU/FPGA/Network Processor and is totally transparent to the rest of the network. In an orchestrated NFV environment however, the goal is to create a dynamic network that can adapt to higher loads utilizing scale-up or scale-out. If you need to factor in additional I/O ports and CPU for new load balancer VNFs every time you want to add capacity, that dramatically increases the operational complexity of the deployment, not to mention "wasting" CPU cores and I/O ports. Even if you accept the added complexity, there are still some underlying issues to explore.

What if the capacity increase is driven by the need to add more bandwidth for the existing subscribers already active on the system? If the network is load balancing all traffic from existing subscribers to the existing CPUs, adding new CPUs will only help new subscribers (or possibly subscribers that have timed out of their existing CPU mappings). Unless a re-balancing algorithm is run to redistribute the load across existing CPUs each time a new VNF is added, state would be lost for existing flows.

What if a load balancer fails? Since state is maintained with load balancers, redundancy would be needed for the load balancing VNFs, expending valuable CPUs (and potentially I/O) to handle the extra processing. In the same scenario, what happens if one of the packet processing VNFs (or an entire system) fails or an upgrade is needed? The rebalancing of the load in this scenario should not take minutes, it should take seconds. Creating network and subscriber symmetry with load balancers is a very inefficient method for virtualized networks- but what is the alternative?




Unlike legacy NIC solutions with just one software RX queue per NIC port, Mellanox CX4 balances traffic across as many queues as the CPU has threads. This helps modern core-dense CPUs by eliminating the software step of arbitrage and balancing of packets across these threads.

## MAINTAINING APPLICATION STATE DURING SCALE-UP AND SCALE-OUT

In the test cases above, traffic was sent directly to CPUs/cores based on the automatic distribution of traffic from I/O modules to CPU cores without any special load balancing solution. If traffic from a specific subscriber connection crossed multiple CPUs with flow asymmetry (i.e. one direction of the flow crosses one CPU and the return flow crosses another), PacketLogic/V uses a technology called FlowSync to pass flow data between the VNFs to maintain session state across multiple systems.

This works across multiple cores on the same CPU, multiple CPUs in a single system, or multiple systems across a distributed deployment – and scales to hundreds of simultaneous VNFs in a FlowSync domain. This has been deployed on Sandvine's hardware deployments since 2008 and has been battle tested and proven even in geographically dispersed deployments, and has now made a successful transition to the virtual world. It was deployed on our very first virtual deployments in 2013, and we are continually adding new capabilities like live migration of FlowSync state between systems to enhance the overall system resiliency.

For example, let's assume you want to upgrade the software version of a VNF to a new version of software. In the new virtual world, you spin up a new VNF running the new version of software, copy the FlowSync state from the active VNF to the new VNF and then switch the traffic over to the new VNF (in a staged manner if you want to test the new software before you conduct a full-scale switch). From the subscriber perspective, there is no downtime while for the network operator there is no loss of analytics or charging data – a huge win and a big contribution to maintaining a high QoE.

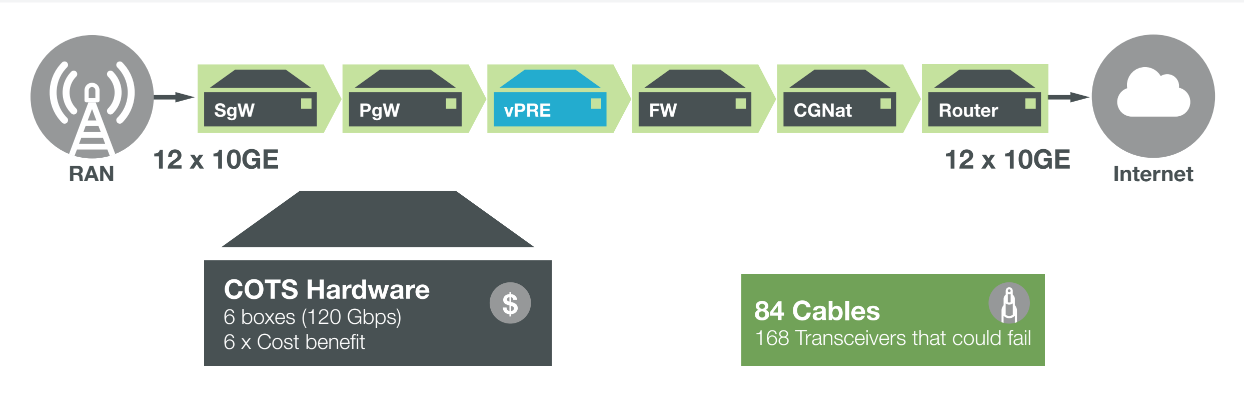
## PRACTICAL DEPLOYMENT SCALING EXAMPLE FOR A MOBILE PACKET CORE

We discussed the cost and efficiency gains from ensuring the highest possible performance per CPU in the test results above. However, that is not the only benefit that is gained with the move to a virtualized infrastructure – there is a significant physical benefit as well from moving to 100GE links combined with virtualizing solutions on COTS hardware. In a high-performance packet core today, for example, there may be 120Gbps of traffic coming into the packet core on 12 x 10GE links. There may be as many as 7 vendors in the packet core - a serving gateway, a packet gateway, a PacketLogic system, a firewall, a CGNAT function, and a border router, and then 12 links going out to the Internet.

Until recently this would be deployed with big iron, meaning vendors' own appliances, each with their own hardware and software combinations and interoperability issues. If this moves to COTS and virtualization, those systems could move to COTS hardware (like the Dell systems used in this test) with a cost of around 20% of vendor-specific hardware - a significant reduction in CAPEX.

Figure 4

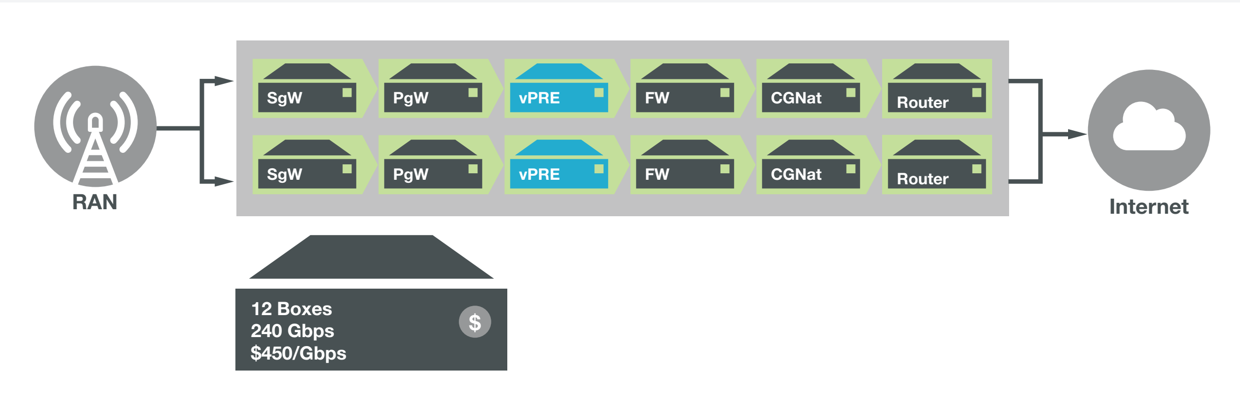
SERVICE FUNCTION CHAIN



However, there are also significant potential savings in OPEX as well. This deployment consists of 84 cables and 168 transceivers that could also fail. If you think about scaling that packet core by another 120GE up to 240GE we would have to add another one full line of 120GE. It is likely this would be done way too early; as you approach 90GE out of the original 120GE you would probably be thinking about your second line, roughly a year before you actually need it. If you want redundancy you would need a full line of redundancy, meaning your throughput costs would double. But even then, the deployment would have a line with 168 cables and 336 transceivers - twice as many things that could fail.

Figure 5

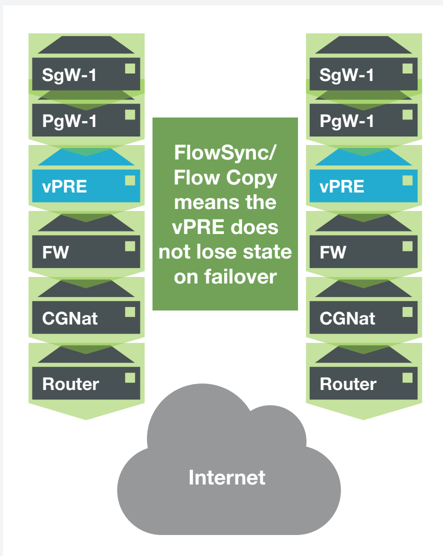
SERVICE FUNCTION CHAIN




What if, instead, all of the packet core vendors coexist with other vendors on the same hardware in a virtualized format? Now one of those COTS systems could do at least 20Gbps per VNF, with the potential for 6 different vendors utilizing several 100GE ports. Bandwidth “hardware” cost is around 20% per gigabit but the envelope is a little bit different - allocate a couple of cores and a little memory to each VNF, either equally or unequally depending on which software is the fastest, hook it up with SR-IOV and the result is a massively altered OPEX challenge. It also enables easy scale-out as you need more capacity.

Figure 6

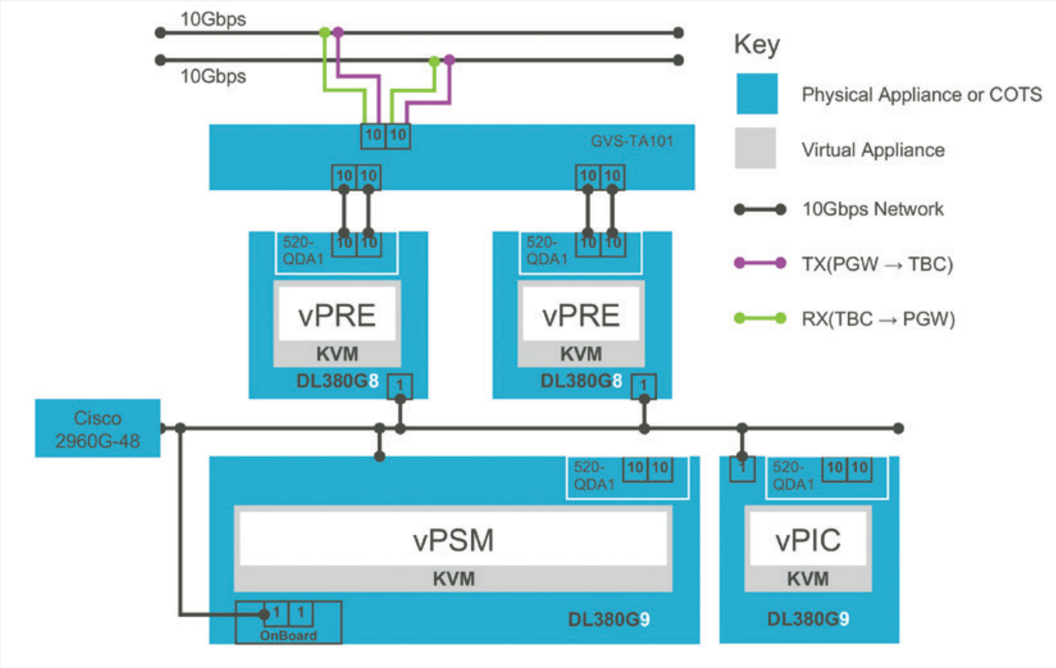
SERVICE FUNCTION CHAIN



## ANALYTICS DEPLOYMENT AND SCALE-OUT

One deployment that has been active since 2015 is an analytics-only deployment for an APAC mobile customer servicing tens of millions of mobile subscribers. The initial deployment was extremely cost effective for the operator, as they simply re-used hardware that was already deployed on their network. The network deployment (for a single site example) is shown below, with each system delivering 40Gbps of capacity utilizing the COTS hardware connected to the mobile packet core.

Figure 7

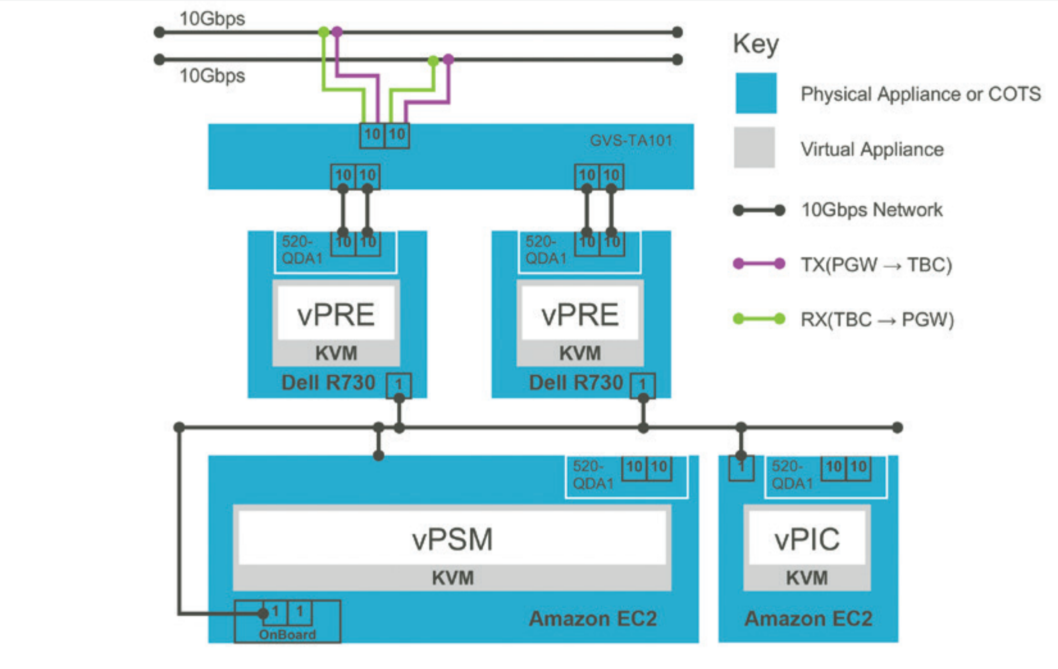



Each PacketLogic/V instance is receiving all traffic from the mobile packet core, enriching the data with policy-based subscriber information (subscriber ID, location, service plan, etc.), and sending the data via IPFix to the operator's data lake to be processed by their Big Data solutions. The Sandvine subscriber and network intelligence is being used to monitor network QoE, identify specific capacity issues that are degrading network QoE and feed the service planning teams with subscriber demographics to identify new service opportunities. When additional capacity is needed in each site, a scale-out of another system is required with no additional VNFs needed thanks to FlowSync.

## TRAFFIC MANAGEMENT AND POLICY ENFORCEMENT DEPLOYMENT SCALE-OUT

Our first virtual deployment at Boingo is a good example of a traffic management and policy enforcement scale-out across multiple geographic locations. At each site, a single COTS server is deployed with around 6 VNFs per server, including a PacketLogic/V VNF.

Figure 8




Each PacketLogic/V PRE instance is run locally on the COTS hardware and is used for subscriber management (in this case when each user authenticates, they are associated with their correct service plan – which may be a free service, or a higher priority service for mobile offload customers of Boingo or Boingo subscribers), traffic/congestion management, policy enforcement, and analytics (subscriber, network, and QoE). The PSM (subscriber integration) and PIC (analytics) VNFs are run in the Amazon EC2 cloud and can be shared between multiple deployment locations if desired. When a new site is brought online, a new instance is spun up from preconfigured templates and activated on the COTS system. This can be utilized to scale-out to thousands of sites. In a single site, if more bandwidth is needed, scale-up can simply add more cores to a single PacketLogic/V instance or another PacketLogic VNF can be added using spare CPU capacity.

## SUMMARY

There are significant challenges to overcome before operators arrive at the nirvana of a fully orchestrated telco cloud utilizing NFV. Sandvine is leading the way for operators with virtualized DPI solutions that enable significant cost savings by optimizing I/O and CPU processing power as well as flexible scale-up and scale-out options.

PacketLogic/V's use of FlowSync technology enables cross-instance flow communication that removes the need to create artificial symmetry using load balancers that consume valuable CPU and I/O resources. The FlowSync approach ensures that operators achieve maximum cost savings when deploying virtual solutions, and enables dynamic instantiation for scale-up and scale-out without complicated network routing changes. In addition to the optimization delivered by leveraging FlowSync, PacketLogic/V has been engineered to deliver the highest performance per CPU core to minimize the overall deployment cost.

If you are interested in learning more about the many use cases supported by Sandvine, please contact us at: **info@sandvine.com**

v20171221



SANDVINE

**USA**
47448 Fremont Blvd,
Fremont,
CA 94538,
USA
T. +1 510.230.2777

**EUROPE**
Birger Svenssons Väg
28D
432 40 Varberg,
Sweden
T. +46 (0)340.48 38 00

**CANADA**
408 Albert Street,
Waterloo,
Ontario N2L 3V3,
Canada
T. +1 (0)519.880.2600

**ASIA**
Ardash Palm Retreat,
Bellandur, Bangalore,
Karnataka 560103,
India
T. +91 80677.43333