WHITE PAPER

Communications Service Providers
Network Functions Virtualization

Software-Defined Infrastructure on Bare Metal Boosts Performance

A10 Networks shows how ADC software on Intel-based bare metal server can deliver up to double the number of L4 connections per second than KVM*-virtualized server.[1]

A10's ACOS software on bare metal environment achieves 40 Gbps throughput[1]

Table of Contents

Executive Overview

An unprecedented level of network traffic is driving enterprises and communications service providers (CommSPs) to look for equipment that delivers greater operational agility and economics. The industry is responding with equipment that implements network functions virtualization (NFV) on high-volume, industry-standard servers. A key enabler of this approach is the Data Plane Development Kit (DPDK), supported and originally developed by Intel, which greatly improves packet processing throughput on these servers.

Satisfying the need for higher performance, A10 Networks has raised the bar on throughput and connections per second with bare metal application networking solutions (e.g., ADC - application delivery controller) that greatly reduce the latency associated with typically virtualized environments. Bare metal is a new delivery method for A10 Networks' family of application networking and security products.

This white paper discusses A10 Networks' ADC software on bare metal environment and its performance compared to a KVM SR-IOV (Kernel-based Virtual Machine, Single Root I/O Virtualization) implementation.

Software-Defined Infrastructure

Enterprises and CommSPs, seeing the compelling agility and operational expenditure savings that virtualization brings to data centers, would like to replicate these benefits in their networks. This transition requires moving away from purpose-built network appliances to software-based network functions running on high-volume servers, referred to as Software-Defined Infrastructure (SDI). SDI can be easily added to existing server racks or even deployed on existing servers, thus eliminating the need to figure out how to mount, connect, power, and cool a dedicated appliance.

With SDI, new network functions and services can be deployed in minutes via a software download, compared to possibly weeks needed to order, ship, and install a dedicated appliance supporting new features. Based on concepts from NFV and software defined networking (SDN), SDI gives enterprises and CommSPs greater flexibility, agility, and scalability than traditional architectures through a nimble, software-based service delivery model.

Trends Affecting Application Networking

The modern data center is undergoing a transformation, primarily driven by the rapid adoption of cloud and mobile computing technologies. Enterprises and CommSPs are experiencing swift growth of data center traffic, which demands high performance, scalable, and software-based data center infrastructure. To keep up with the data traffic growth and emerging cloud architectures, application networking technologies have to significantly evolve, and deliver new capabilities and software flexibility to improve the performance, security, and operational agility of modern cloud data center architectures.

Today's Barriers

A massive gap exists today between non-automated traditional appliances and the highly programmable cloud experience that developers have come to expect from public clouds. Many organizations today look at automation efforts and focus on the virtualized data center.

It is no longer sufficient to overprovision the underlying infrastructure and leave resources idle. Layering a hypervisor on top of performance workloads increases agility, but at the cost of a slight performance hit. Organizations looking to drive efficiency are forced to look into bare metal server orchestration.

In addition, the traditional use case applications tend to remain static in capacity and configuration. There are instances where deploying a hypervisor adds limited value, as in analytics applications and high performance computing (HPC) that require bare metal servers (without hypervisors) in a data center for performance and efficiency reasons.

Mature software-defined infrastructures should consider bare metal provisioning as an option, in addition to virtualization. In today's cloud-first environment, bare metal workloads play a significant role in the data center. While legacy hardware is driving the requirements to keep bare metal workloads running in the data center, the likes of vertical-specific software means that bare metal workloads will exist in the data center for the forseeable future.

Bare Metal Benefits and Adoption Challenges

Bringing bare metal servers to the cloud and powering highly available performance workloads through a unique, never congested network significantly reduces costs. The flexibility of bare metal allows vendors to deliver software solutions with metered, pay-by-the-hour cloud consumption model for customers. Users can also benefit from an enhanced security profile. Bare metal serves as a great option for existing enterprise application deployments in the cloud. The main barrier to bare metal adoption is around orchestration capabilities; however, many cloud orchestrators already support bare metal.

Virtualized and Bare Metal Environments – Choice Matters

Although virtualization technology provides many operational benefits, they come at a cost – one of which is increased latency that ultimately lowers performance. This is due to the time needed to make context switches, copy data, handle privileged instructions, emulate guest operating system functions, and maintain multiple page tables, among other operations. Software and hardware vendors continually work to minimize virtualization overhead, but for networking applications with many connections and sessions, the impact remains considerable.

A10 Networks created application networking software for bare metal environment running on Intel® processor-based servers as an alternative to virtualized envirnoments that supports up to double the number of L4 connections per second than a KVM-virtualized server.[1] In a bare metal environment, no virtual machine monitor (VMM) or hypervisor, like KVM or Xen,* is used. The solution takes advantage of the A10 Advanced Core Operating System (ACOS) to run multiple software-based network functions, which efficiently send and receive network traffic to/from Intel® Ethernet Controllers.

A10's application delivery controller (ADC) for bare metal delivers a full set of application and security services, giving CommSPs their choice of hardware without having to sacrifice the benefits of A10's ACOS Harmony platform for open programmability, policy enforcement, and telemetry. Customers get lower total cost of ownership (TCO) by decoupling the software from the hardware, which improves portability and longevity.

The packet flow is described in more detail in a following section.

Tradeoffs: Virtualized and Bare Metal

The term bare metal is a metaphor for running software applications directly on a CPU, implying applications will run unencumbered from other system software or components that could slow it down. Hence, a bare metal solution is expected to be faster than alternatives, which indeed is the case when comparing A10 Networks' bare metal environment to a KVM SR-IOV implementation, as indicated in Table 1.

CRITERIA	A10 NETWORKS' BARE METAL	VIRTUALIZED
PERFORMANCE	✓ Supports higher connections per second and throughput[1]	✓ Has constrained performance due to virtualization overhead
DEVICE MANAGEMENT & WORKLOAD CONSOLIDATION	✓ Increases software management complexity	✓ Simplifies software updates and loading of new functions and diverse applications
PLATFORM REQUIREMENTS	✓ Requires specific Ethernet Controllers and CPUs	✓ Runs on nearly any x86 platform
SYSTEM COST	✓ Comparable	
PHYSICAL FOOTPRINT		
POWER USAGE		

Table 1. Tradeoffs between virtualized and A10 Networks' bare metal environment

Since bare metal is a minimalist approach, it is not surprising that a virtualization environment provides more flexibility with respect to software. Consequently, a virtualized environment may be better equipped than bare metal to handle software updates and the loading of new functions and diverse applications.

Most virtualized solutions are also designed to run on a wide range of Intel® architecture-based platforms with standard drivers. In contrast, a bare metal environment may be limited to a small number of platform configurations that require particular CPUs, network interface cards (NICs), memory configurations, etc. Therefore, virtualized environments typically have less stringent platform requirements than bare metal environments.

Other system aspects, such as hardware and software cost, physical footprint, and power usages tend to be comparable for the two environments.

Architectural Comparison: Virtualized and Bare Metal

The following is a brief and simplified overview of KVM SR-IOV and A10 Networks' bare metal environments. The purpose is to provide an explanation of why there is an intrinsic performance difference between two environments. The two metrics used to measure performance are throughput and connections per seconds (CPS):

- **Throughput** is a measure of the bitrate that passes through the network device. Devices tend to have higher throughput when packet sizes are large. This is because there are fewer connections and headers to process, compared to traffic comprising smaller packets.

  ADC throughput performance testing typically employs stateful (client-server) traffic such as HTTP requests from clients and HTTP responses from server. In order to provide a realistic measure of a device's throughput, testing can be performed using Internet Mix (IMIX) traffic, which is stateless traffic and contains a real-world blend of packet sizes.

- **Connections per second** (CPS) is a measure of how many stateful connections a network device can support per second. This includes a TCP connection setup (3-way handshake), one HTTP GET request and response within a TCP connection, and TCP connection teardown.

This metric is important because it is a measure of the active user traffic a network device can support over a given timeframe.

KVM SR-IOV

Early attempts to virtualize network devices resulted in an unacceptable amount of latency due to packets having to pass through a hypervisor or VMM as they travelled between NICs and virtual machines (VMs) containing networking functions.

This issue was remedied when Intel Ethernet Controllers began to support the PCI-SIG Single Root I/O Virtualization (SR-IOV) standard, which enabled different guest operating systems to directly access the Ethernet controller ports without hypervisor/VMM intervention. SR-IOV provides a way to bypass the hypervisor's involvement in data movement by providing an independent memory space, interrupts, and DMA streams for each virtual machine (VM).[2]

This enhancement was further aided by the Data Plane Development Kit (DPDK), which provides a set of software libraries that make it easier and faster to process packets. Figure 1 shows a possible example configuration where three VMs directly access dedicated resources within the Ethernet controller via virtual functions.

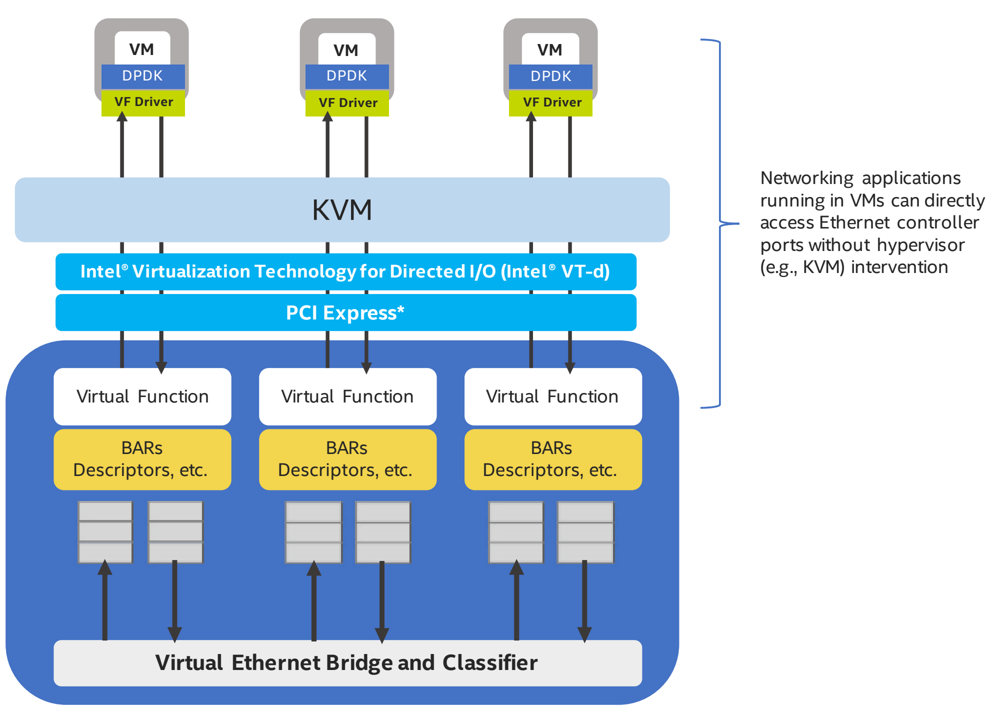

Figure 1. Example of KVM SR-IOV environment[3]

Virtualization Technology

Intel® Virtualization Technology for Directed I/O (Intel® VT-d) provides a hardware-based mechanism to translate addresses for DMA transactions issued by I/O devices. This translation mechanism is important in a virtualized environment where the address space seen by the guest OS is not the same as the underlying physical address of the host machine.

Poll Mode Driver

The Poll Mode Driver (PMD) in the DPDK greatly speeds up the packet pipeline by receiving and transmitting packets without the use of asynchronous, interrupt-based signaling mechanisms, thereby eliminating much overhead. The poll mode driver directly accesses packets from the queues in the Intel Ethernet Controller, so there is no need for the controller to generate time-consuming interrupts. For more information about the DPDK, visit DPDK.org.

Performance Limiters

- Hypervisor architecture
  Although the DPDK and SR-IOV eliminate a significant portion of the delay associated with the hypervisor/VMM, some latency remains for packet processing workloads.
- Ineffective resource allocation
  The Hypervisor/VMM may poorly allocate resources, such as assigning virtual CPU cores (i.e., Intel® Hyper-Threading Technology) instead of physical cores to time-critical workloads. Another example is using shared memory, which could become overbooked, instead of dedicated memory for critical functions.
- VM adjacency
  The Hypervisor/VMM could create a VM adjacent to itself, potentially reducing the VM's performance if the Hypervisor/VMM dominates their shared resources, like cache.
- Memory sharing
  The Hypervisor/VMM may impose memory sharing between VMs that create additional latency from inter-VM communication, memory coherency mechanisms, or extra memory copies.

A10 Bare Metal Software

From a performance perspective, A10 Networks' bare metal software environment overcomes several performance limiters found with typical KVM SR-IOV solutions.

- Hypervisor architecture
  The bare metal environment does not have a hypervisor or VMM, and software natively runs on the server.
- Ineffective resource allocation
  The bare metal environment proactively assigns resources to network functions to optimize their performance. Resource assignments are not arbitrary.
- VM adjacency
  This is not an issue because there is no hypervisor or VMM.
- Memory sharing
  As part of the A10 Advanced Core Operating System (ACOS) software, A10 Networks implements a distributed architecture with high-speed shared memory that minimizes interdependencies between multiple processes cores. Shown in Figure 2, the architecture ensures shared state is available to all cores, eliminates traditional design bottlenecks, and linearly scales performance with increasing numbers of processor cores. It provides a consistent view of memory, allowing immediate enforcement of any policy or configuration changes without the need for copy overhead across CPU cores, saving a significant amount of time and thus maximizing the processing of network traffic.

ACOS also includes the ACOS Flexible Traffic Accelerator (FTA), a high-performance network I/O technology that can distribute application flows intelligently across cores on deterministic paths.

Additionally, the ACOS platform architecture separates forwarding and management planes to maximize concurrency. For traffic forwarding functions, multiple dedicated CPUs are allocated for highly-efficient parallel

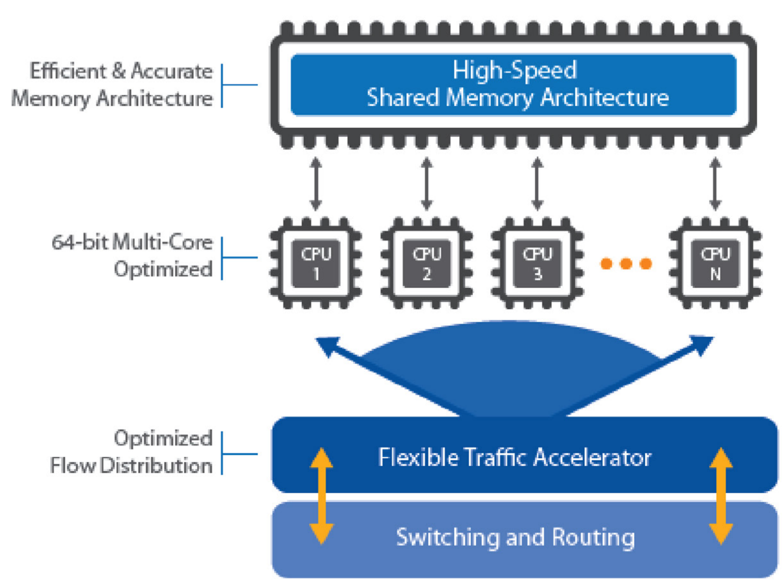

Figure 2. A10 Network's ACOS architecture

processing of traffic. For management functions, ACOS allocates a dedicated CPU for management purposes. As a result, administrators can reach devices and perform supervisory and health checks even under peak traffic conditions. The separation of management functions onto a separate core enables management flexibility and improves overall system robustness.

Performance Comparison

Figure 3 compares the Layer 4 and Layer 7 connections per second delivered by bare metal and KVM SR-IOV platform[1] per testing by A10 Networks using stateful connection traffic.[4] The bare metal platform had 93 percent higher Layer 4 connections with 4 cores (10 Gbps throughput license) and double the connections with eight cores (20 Gbps throughput license). The bare metal platform also supported more Layer 7 connections, higher by 53 percent and 25 percent for four cores and eight cores respectively.

Currently, A10 offers 10 Gbps, 20 Gbps and 40 Gbps throughput licenses for bare metal, which are associated with CPU cores: four cores, eight cores, and 14 cores respectively. For vThunder* KVM SR-IOV, the available license was with 20 Gbps throughput for KVM SR-IOV.

Bare Metal Use Cases

A10 bare metal solutions benefit a wide range of organizations, as described in the following use case examples:

Use Case 1: Service Provider Data Center

Bare metal ADC allows service providers to implement advanced application delivery functionality on Intel® architecture. This software-based ADC solution delivers the performance required for heavy workloads and is fully-programmable using A10's open API for automated application delivery service deployments in SDN and software-defined data center (SDCC) environments. A10's Bare Metal ADC can also enable application delivery partition support to isolate traffic and administration for multi-tenancy use. Furthermore, A10's aFleX policy engine performs granular traffic inspection and control, enabling administrators to construct advanced correlation policies that control access based on patterns in web server responses, source IP addresses, user agents, and more.

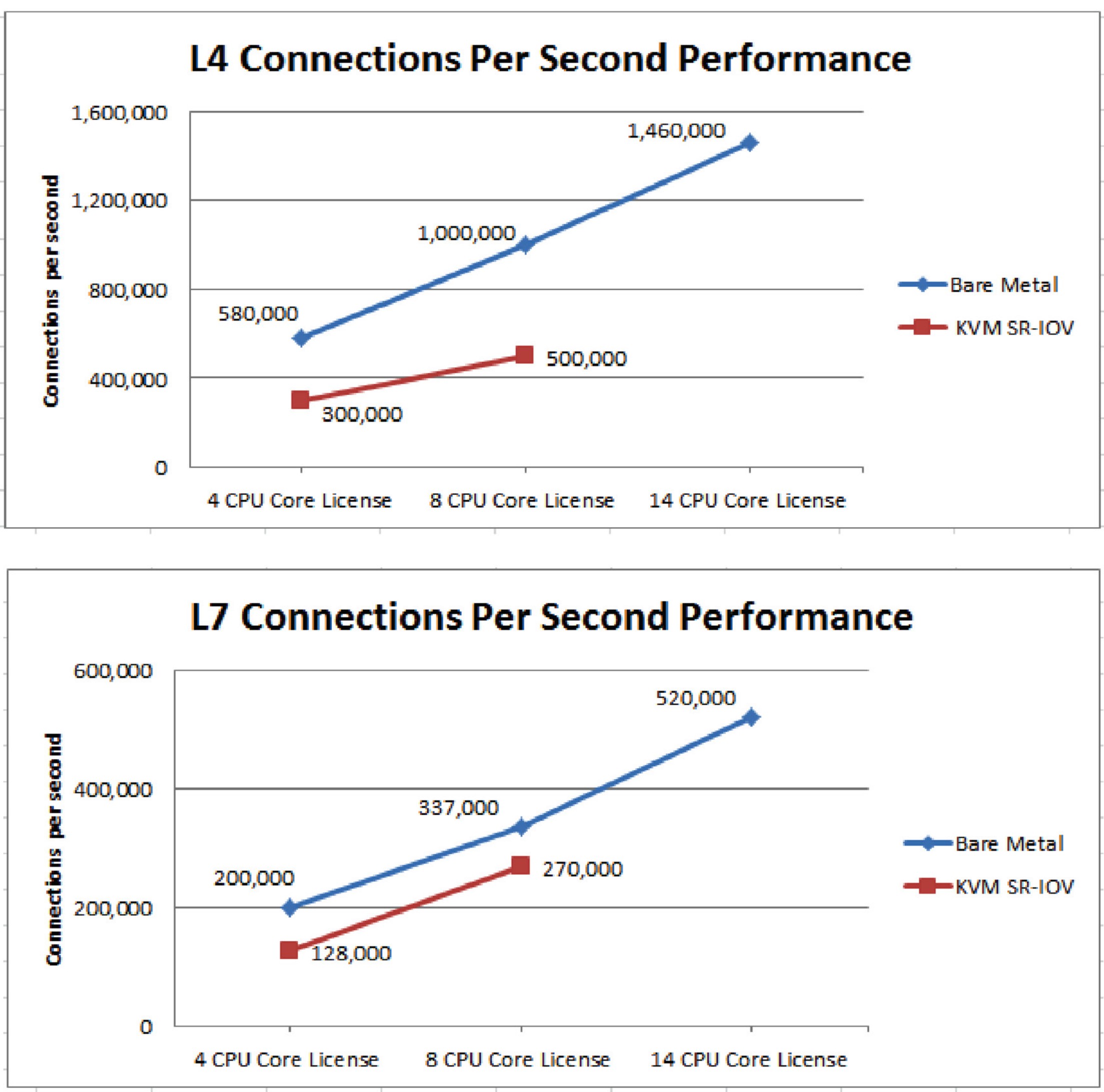

Figure 3. Performance data: bare metal vs. KVM SR-IOV[1,4]

Use Case 2: Hosting Provider

In a managed hosting provider scenario, A10's bare metal solution enables CommSPs to expand their service portfolio by supporting tiered pricing models, such as:

- Standard service: Limited SLA/shared equipment service for small- and medium-sized customer environments
- Premium service: Increased SLA/dedicated equipment service for large customer environments

These models take advantage of public cloud hosting, which can help CommSPs stay competitive and reduce their costs. Moreover, CommSPs can offer dedicated services using bare metal ADCs with off-the-shelf hardware and licenses that further lower costs compared to purchasing purpose-built ADC appliances. This approach allows CommSPs to lower service contract costs and satisfy the performance levels needed to address opportunities across shared servers (hypervisor/VMs), dedicated servers (bare metal), and a dedicated appliance.

With an infrastructure as a service (IaaS) public cloud, CommSPs can offer a range of performance levels:

- General performance: A cost-effective approach using a virtualized ADC running in a VM on a shared server
- High performance: A premium and highly secure service approach using a bare metal ADC on a dedicated server

Conclusion

The transition to software-defined infrastructure (SDI) based on NFV and SDN principles is expected to deliver more agility and flexibility to enterprises and CommSPs, but this cannot be at the expense of performance. A10 Networks' bare metal application networking solutions overcome some of the performance issues associated with virtualized solutions.

In addition, the solution, compared to traditional appliances, lowers hardware and spares costs through the use of high-volume, Intel processor-based servers. The solution also provides flexibility and agility with the ability to install software on-demand versus the added logistics associated with hardware appliances. With its bare metal environment, A10 Networks is delivering high performance software for application networking and security solutions.

For more information about network and communicaiton infrastucture solutions, visit https://networkbuilders.intel.com.

To learn more about solutions from A10 Networks, visit www.a10networks.com.

[1] Test performed by A10 Networks. Platform Configuration:
Hardware (Bare metal and KVM SR-IOV testing): Custom server, CPU: Dual Intel® Xeon® processor E5-2650 v4 @ 2.20GHz (12 cores, 30M cache), System Memory: DDR4 DIMM ECC Memory (256 GB), Intel® SSD DC S3700 Series, Intel® Ethernet Converged Network Adapters XL710 (10 GbE ports)
Software (KVM SR-IOV testing): Platform: x86_64, Kernel info: Linux kickseed 3.5.0-39-generic, KVM version: QEMU emulator version 1.2.50 (qemu-kvm-devel), OS: Ubuntu 12.04.2 LTS, virt-manager version: 0.9.1.

[2] Intel White Paper, "PCI-SIG SR-IOV Primer: An Introduction to SR-IOV Technology," pg 13, http://www.intel.com/content/www/us/en/pci-express/pci-sig-sr-iov-primer-sr-iov-technology-paper.html.

[3] Ditto #2, pg 12.

[4] Test performed by A10 Networks. Simulated Network Traffic:
All these three test cases were general ADC test cases using stateful connection traffic. To be exact, HTTP request packets from client(s) and HTTP response packets from the server.
- Layer 4 Throughput: measures the maximum traffic (in Gbps) the A10 ADC can sustain when the Layer 4 VIP (TCP) is configured. The client sends multiple HTTP GET requests within 1 TCP connection. and the target HTTP file size is 128KB.
- Layer 4 CPS: measures how fast the A10 ADC with an L4 VIP can handle new TCP connections (including 1 HTTP request) within 1 second. Target file size is small (i.e., 2B).
- Layer 7 CPS: measures how fast the A10 ADC with an L7 VIP (w/ content switching) can handle new TCP connections (including 1 HTTP request) within 1 second. Target file size is small (i.e., 2B).

For more details, please refer to https://www.a10networks.com/blog/adc-performance-metrics.