Using Intel® FM10K-Based Technology to Accelerate Cloud Network Provisioning

Open vSwitch performance surges with adapters by Silicom

PREFACE

The Open vSwitch (OVS) open source project (https://openvswitch.org) is the de-facto standard switching solution for network provisioning in virtualized environments. It is a stable, reliable market-driven solution.

Designed as a flow-oriented programmable data plane, Open vSwitch uses a multi-tiered architecture that is flexible enough to adapt to innovations in traffic engineering. Silicom has recognized that this flexibility could enable the integration of Intel® FM10K-based technology into Open vSwitch data planes, an improvement that offered significant potential for enhancing their performance.

Based on extensive experimentation, development and testing, Silicom has taken an innovative approach to the task that has proven to be uniquely successful. Rather than offloading the entire Open vSwitch data plane onto hardware, the Silicom approach is to allow each mechanism – whether software on the CPU or the Intel® FM10K switching hardware – to do what it does best, thereby optimizing the performance of the data plane as a whole. ***Silicom has implemented this approach in a wide range of Intel® FM10K- based solutions, all of which are now available.***

INTEL® FM10K EXPLAINED

In brief, the Intel® FM10K controller is a network controller (MAC) coupled with a switching fabric that sets a new standard for virtual network provisioning. An elaborate Filtering and Forwarding Unit (FFU) appears in the FM10K data path, and an extensive (32K entry) TCAM engine and tunneling engine combine to form a strong yet programmable packet processing engine. Completing the solution are mix-and-match options of 10Gbps/ 25Gbps and 40 Gbps/ 100Gbps links, with up to 24x25Gbps or 6x 100Gbps links per single adapter (in a single chipset!).

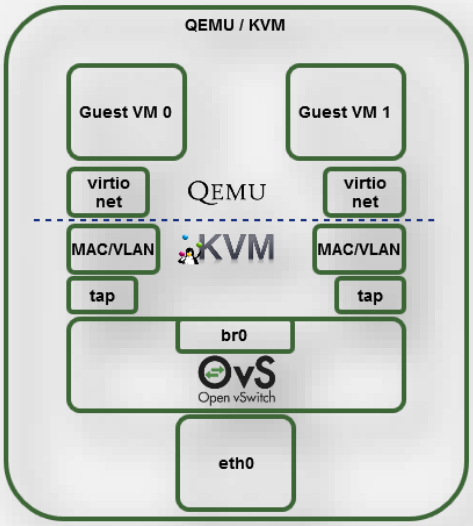


Figure 1 - Open vSwitch

OPEN VSWITCH BENEFITS

As a mature and well-maintained open source project, Open vSwitch opens the door to a long list of capabilities. Most important are the following advantages:

- **Encapsulation support** (VLAN/VXLAN/NVGRE etc.) is available, along with support for Open Flow.
- **Modular design:** the modular design of the kernel fast path and user space data plane allows for fluent progress and advancement.

OPEN VSWITCH DRAWBACKS

However, Open vSwitch also comes with some significant drawbacks.

- **Latency & bandwidth issues:** Open vSwitch's MAC/vTAP interface (the data plane interface between the kernel space and user space) adds a great deal of latency to the data path and narrows the bandwidth.
- **Encap/decap drag:** the fact that tunneling and de-tunneling (encap / decap) tasks are performed in software on the CPU takes a performance toll.
- **Drag on CPU processing cycles:** in systems that implement Open vSwitch infrastructure, the Linux OS scheduler is required to allocate a considerable number of time slots to the Open vSwitch mechanisms, resulting in **fewer CPU cycles available for processing.** The remaining processing cycles are **distributed in a non-coherent manner** across compute jobs running on a hypervisor, making it difficult to sustain SLAs in the service of compute jobs.

INTEL® FM10K BRINGS VALUE

Silicom's approach to improving the performance of Open vSwitch solutions was to transfer a portion of OVS tasks from software on the CPU to the FM10K hardware, thereby improving the performance of offloaded tasks while freeing up the CPU.

Silicom ltd., 14 Atir Yeda St., Kefar Sava, ISRAEL
 sales@silicom.co.il www.silicom-usa.com

Silicom determined that the following OVS tasks could be offloaded:

Exact Match Cache: the Exact Match Cache is a powerful flow match engine in the OVS kernel module that preserves a portion of the configured flows that must be matched within the Linux kernel, eliminating the need to forward unnecessary traffic to user space.

In the Silicom solution, the Intel® FM10K TCAM mechanism is operated to handle the exact matched cache rules far more efficiently. This single improvement improves performance significantly.

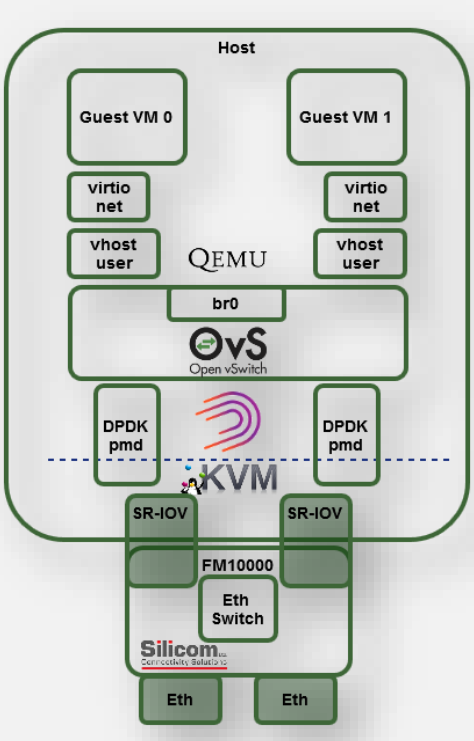


Figure 2 - Accelerated Open vSwitch with Intel® FM10K

Encapsulation and de-capsulation of 802.1Q VLAN, VXLAN, and NVGRE: Silicom solutions use a dedicated FM10K tunneling engine to handle header push and pop, a task that is beyond standard general purpose CPU's forté, thereby ***freeing up CPU cycles for VM processing***. This further enhances results:

- **Bandwidth** and throughput increase and become more coherent.
- Overall **latency** is reduced. Even more important, data flows become less jittery and more predictable.
- Inter VM compute **scheduling** becomes increasingly coherent, enabling SLA-sensitive services to run on VMs without fear of unreliable hypervisor scheduling.
- Overall **VM density** is improved.

NO LOSS OF VM CAPABILITIES

It is common in the world of offload to talk about the "cost of offload" - that is, what is lost or paid along the way when a task is offloaded from the CPU in general, to auxiliary hardware. For example, latency can increase when data is circulated back and forth to the offload engine.

In the Silicom FM10K solutions, however, there is absolutely no loss - no cost for offloading exact match and encapsulation from the OVS. In fact, the intelligent use of the low-latency SRIOV data path up to user space results in ***a tremendous reduction in overall latency.***

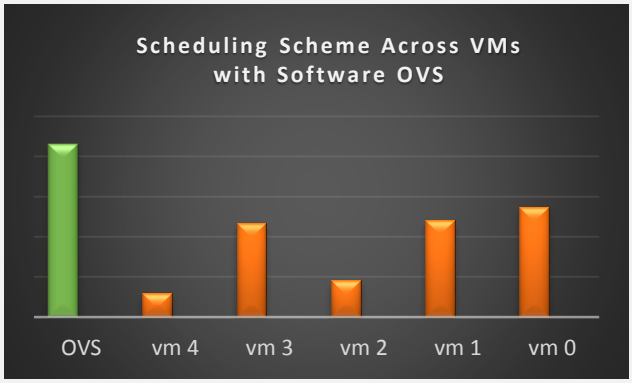
In addition to that, ***live migration, VM pause and resume***, and all other hypervisor capabilities were fully preserved with no compromise, since **`virtio`** VM front-end capability was also preserved.

For detailed tests designs and results, Silicom can be approached. However, the effect of increased coherency in bandwidth distribution across VMs, leads to a potential increase in ***VM density*** or total number of VMs that are effectively be operated on a hypervisor by 150% to 200%. Same increase could be expected in the number of VMs that effectively ***serve high bandwidth***, such as 10Gbps and up.

SUMMARY

This paper introduces Silicom's approach to accelerating OVS: the offloading of specific tasks to an FM10K-based adapter in an intelligent way that delivers a loss-less performance boost. The general intention was to relieve CPU cycles from network processing, re-allocating the cycles saved to VM compute tasks. **The underlying effect is coherency**. This leads to tangible benefits such as increased VM density, better bandwidth and optimized scheduling.

Silicom is soon to publish comprehensive test results and metrics to support these assertions.



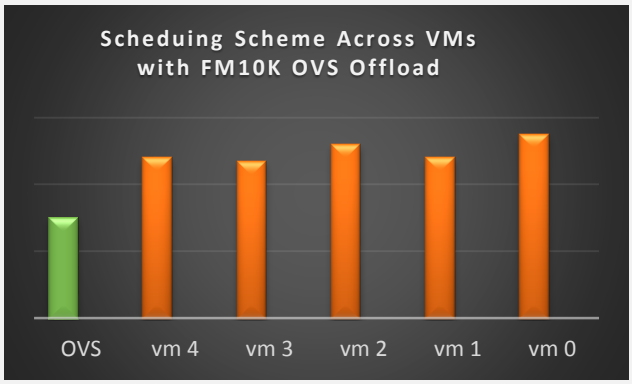


Figure3 - Comparison of VM scheduling scheme with vs. without offload. OVS takes a lot less CPU when offload to FM10K occurs, enabling better scheduling spread across VMs