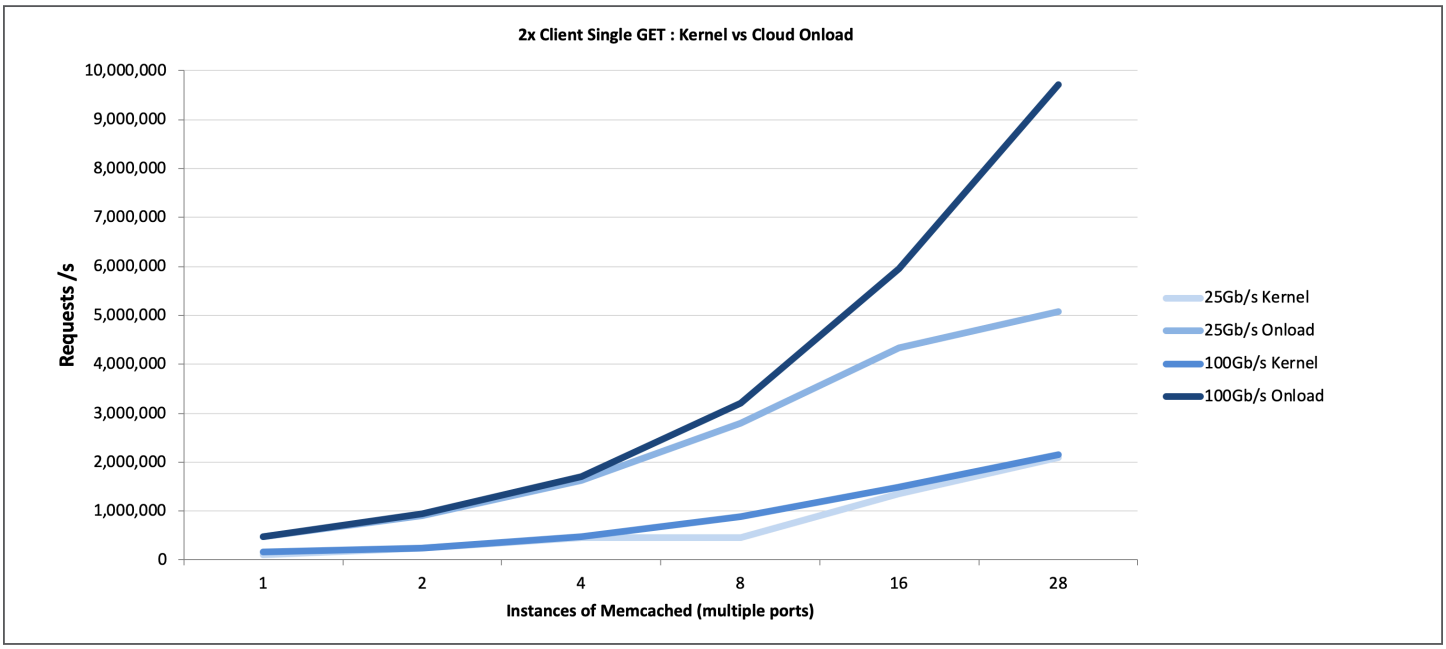
Memcached with Cloud Onload™ Sees a 300% Average Performance Gain

What is Memcached?

Memcached is a high-performance, distributed memory object caching system, generic in nature, but originally intended for use in speeding up dynamic web applications by alleviating traditional database latency. You can think of it as a short-term memory for your web applications. Structurally it is an in-memory key-value store for small chunks of arbitrary data (ex. strings, images).

Memcached is heavily network dependent by design, so its performance can be significantly improved through enhancements to the underlying networking layer.



Key Observations from Performance Testing

- Solarflare's Cloud Onload™ delivers an average performance gain of 300% for Memcached when measuring 128byte GET requests per second (RPS) across the entire range of one to 28 instances using multiple ports over 25GbE with two physical clients each spawning up to five million parallel connections.
- Using 100GbE and all the same settings as above the average benefit was an impressive 280%.
- When we moved to a single physical client machine the average results were a consistent 290% gain for both 25GbE and 100GbE.
- Looking at Memcached's SET function with 128byte messages the average performance gains ranged from 230% to 270% which across both 25GbE and 100GbE and when using a single versus two physical clients.

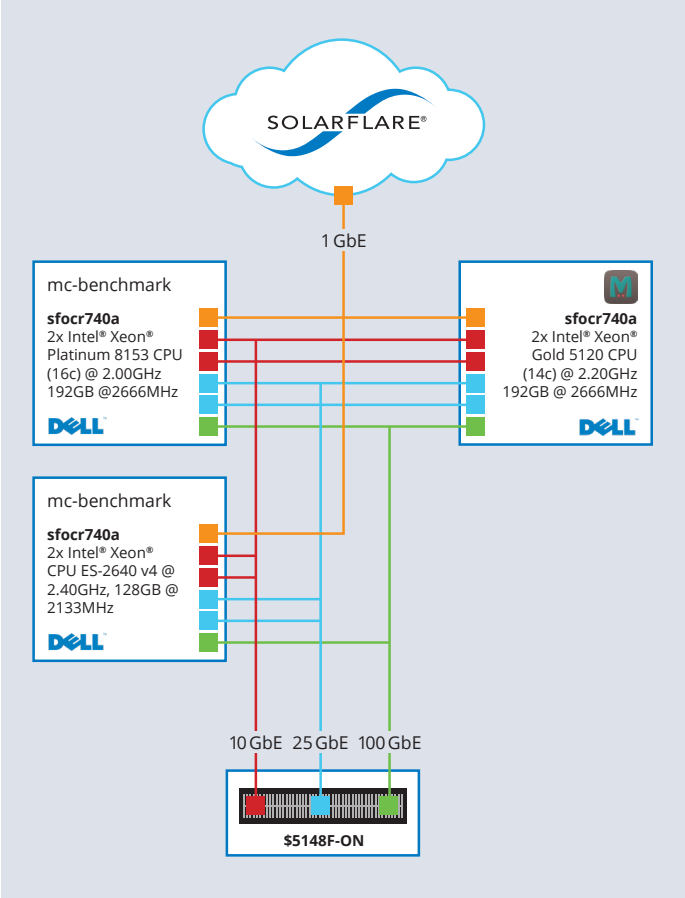
Why Memcached Benefits from Kernel Bypass

Since Memcached is network intensive, every request includes network processing overhead. Whenever an application like Memcached touches hardware, other than the CPU or memory, and in this case the network, it must make at least one, and sometimes several calls to the operating system kernel. Each request is additional overhead that requires both CPU cycles and processing time. Solarflare's Cloud Onload moves the network processing required by Memcached from the kernel into Memcached's own application space in memory. This single modification improves Memcached performance by 300% on average as can be seen in the above graph.

Description of Test Platforms

For this testing, we used two Dell EMC PowerEdge R740 dual-socket Intel Xeon systems and a single R630 system, each with a 25GbE X2522 and a 100GbE X2541 card. These cards were then connected back-to-back for this testing.

The R740 systems had two Intel Gold 5120 processors, clocked at 2.20GHz with 14 cores per processor, and two Platinum 8153 CPUs clocked 2.0GHz, with 16 cores, both with 192GB of memory. The R630 system included two E5-2460 v4 CPUs clocked at 2.40GHz, with 10 cores and 128GB of memory as shown in the block diagram.



Tuning Configuration

We tested both a "Balanced" and a "Performance" profile with Cloud Onload and found that the Performance profile yielded the best results. The Cloud Onload Performance profile is designed to achieve maximum performance, giving the best throughput and transaction rate, as well as best average and 99th percentile transaction response times. However, this profile assumes that the application threads have exclusive use of physical CPU cores. To get best performance, the user may need to explicitly pin applications threads to physical cores (i.e. avoid threads sharing hyperthreaded CPU cores) and ensure there are enough unused CPU cores for scheduling other applications. Because this profile may use busy polling (also known as spinning), CPU utilization metrics do not provide a usable indication of system load.

This profile had some of the following characteristics:

- Enable event polling without interrupts within epoll_wait calls to reduce latency and avoid context switches (EF_POLL_USEC and multiple EF_*_SPIN options)
- Use Epoll mode 3 (EF_UL_EPOLL=3) which gives best epoll performance when a large number of sockets are in the epoll set.

Observations

Memcached relies on the operating system's communications stack to process network I/O requests, but in high core count environments, this stack has become the new bottleneck. Here are some additional points to consider:

- When running Memcached with Cloud Onload over 25GbE, there is a significant and relatively consist average gain of 280% when using the Performance profile. With the Balanced profile we see a 180% average improvement. Again, tests run over 25GbE and 100GbE are comparable, so network speed is not a limiting factor up to 28 instances.

- Running mc-benchmark from one or two client servers produces kernel results over both 25GbE and 100GbE which are comparable, showing no advantage of 25GbE over 100GbE when using GET requests scale which from 78K to 1,455K. This also shows that a single client running mc-benchmarks tests pointing to a single Memcached server can saturate the Memcached processes at 25GbE up to 28 instances.
- Cloud Onload using the Performance profile running over 25GbE or 100GbE, and using the GET or SET function with anywhere from 1 to 28 instances, and multiple ports, consistently delivers performance in the range of 275%

For More Testing Details

Check out Solarflare's **Cloud Onload Memcached Cookbook** for the exact installation and testing process along with the specific tuning and tweaking commands executed above.

The above benefit statements are the result of benchmarking designed to focus on the value of optimizing networking through Cloud Onload kernel bypass. Real world use cases are not the same as benchmarks and as such the role that networking plays may vary, so your overall measurable benefits may be different.

For more information please visit:

solarflare.com

Contact Us:

US +1 949 581 6830
UK +44 (0) 1223 477171
HK +852 2624 8868
Email: sales@solarflare.com

SF-122430-CD Issue 1
Memcached Benchmark Results 071619