STAC Summit Update

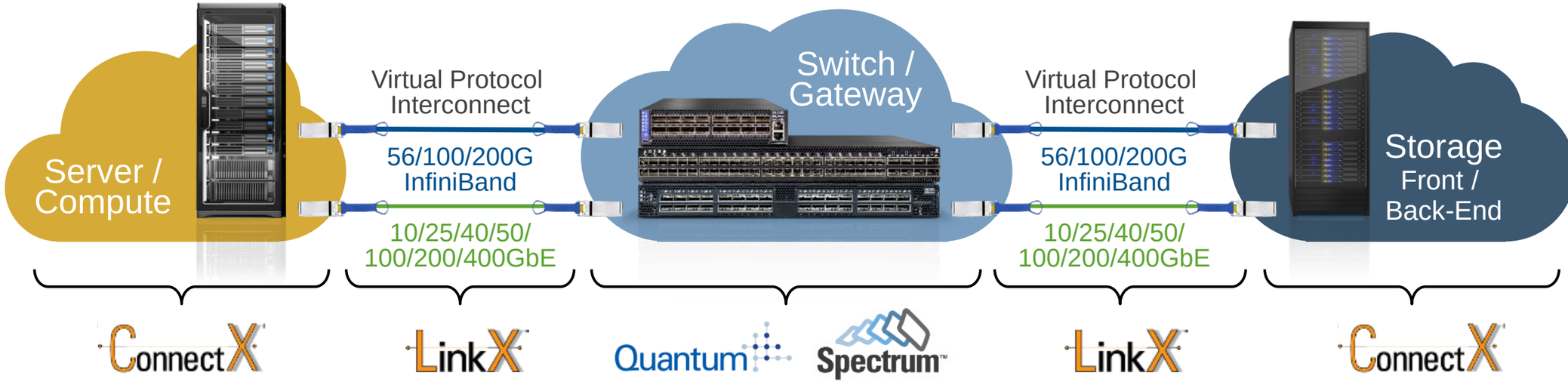
Low Latency Networking Update

Asaf Wachtel, November 2017

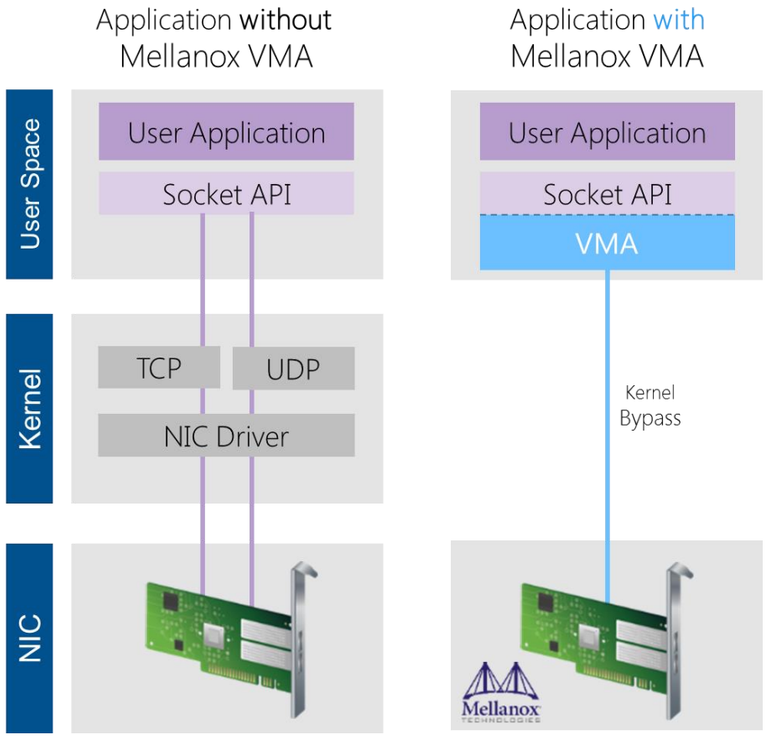
Scalable End-to-End Low Latency Interconnect Solutions

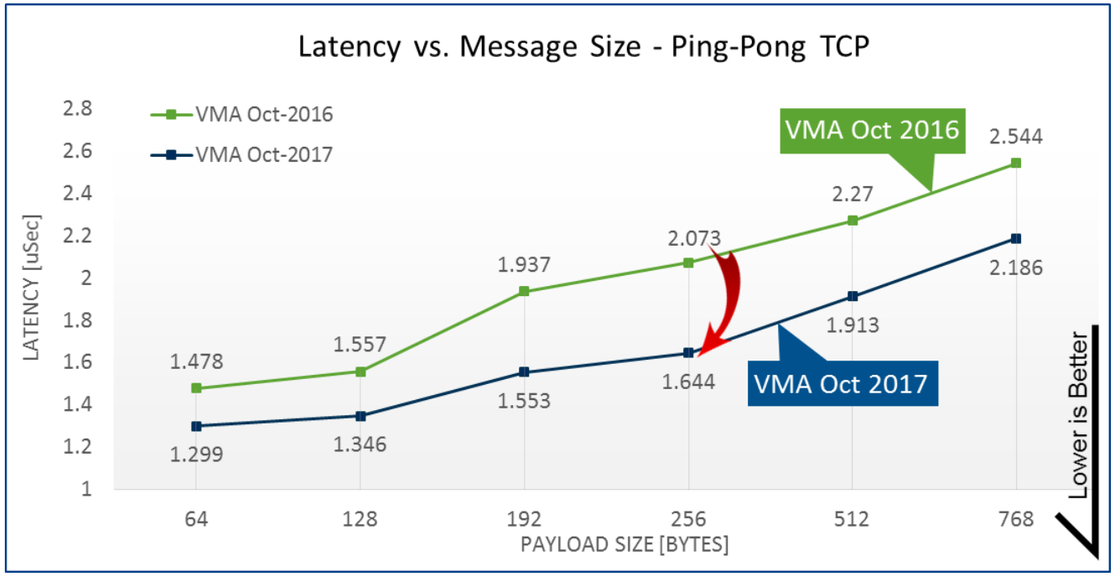
Applications	**Trading / Market Data**	**PTP / Time Sync.**	**Capture / logging**
Software APIs	**VMA**	**RDMA / RoCE**	**DPDK**
Hardware APIs	**Raw Ethernet / Verbs**		

ConnectX

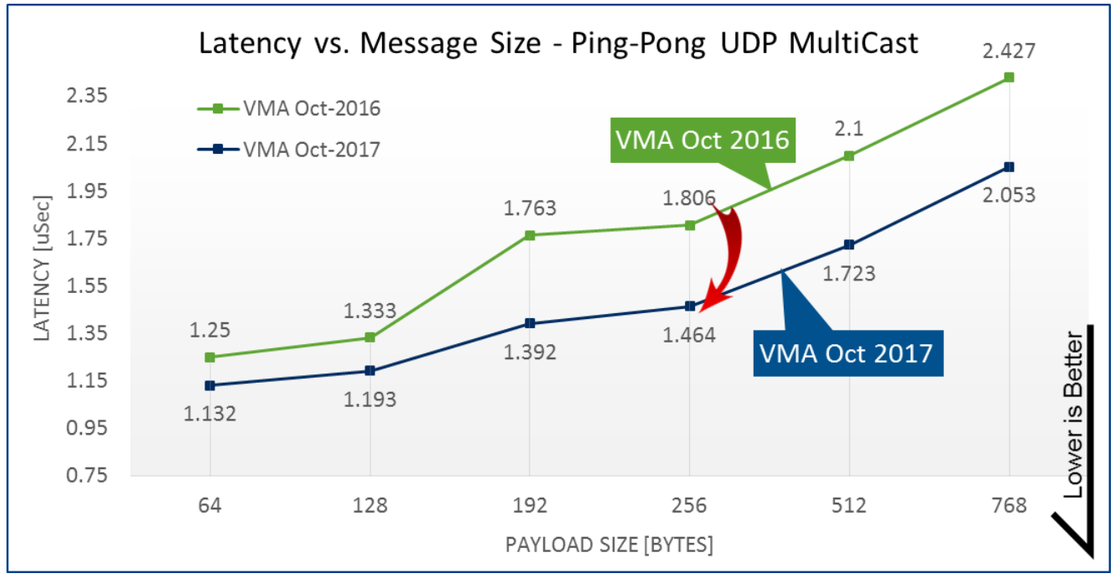


Additional 20% latency reduction in 2017 from Software Only





Not STAC Benchmarks



No Application Changes – Standard Socket API

Single Sided – Plug & play, fully interoperable

Fully open source – no licensing cost

Latest Public STAC Validation October 2017

STAC Report™

STAC-N1™ Benchmarks β1 October 2017
Mellanox ConnectX-4 Lx running Mellanox VMA 8.4

Compared to all other public STAC-N1 reports of Ethernet-based SUTs, this SUT demonstrated:

- **Lowest mean latency at both the base rate and the highest rate tested.**
- **Lowest max latency at the base rate.**
- **Lowest max latency at or above 1 million messages per second.**
- **Highest max rate reported: 1.6 million messages per second.**
- **99.9999th percentile latency (six nines) of just 5 µsec at the base rate and 6 µsec at 1.6 million messages per second.**

Unmatched STAC-N1 Results

Deterministic Latency @ Scale

@ 100K Msg/sec

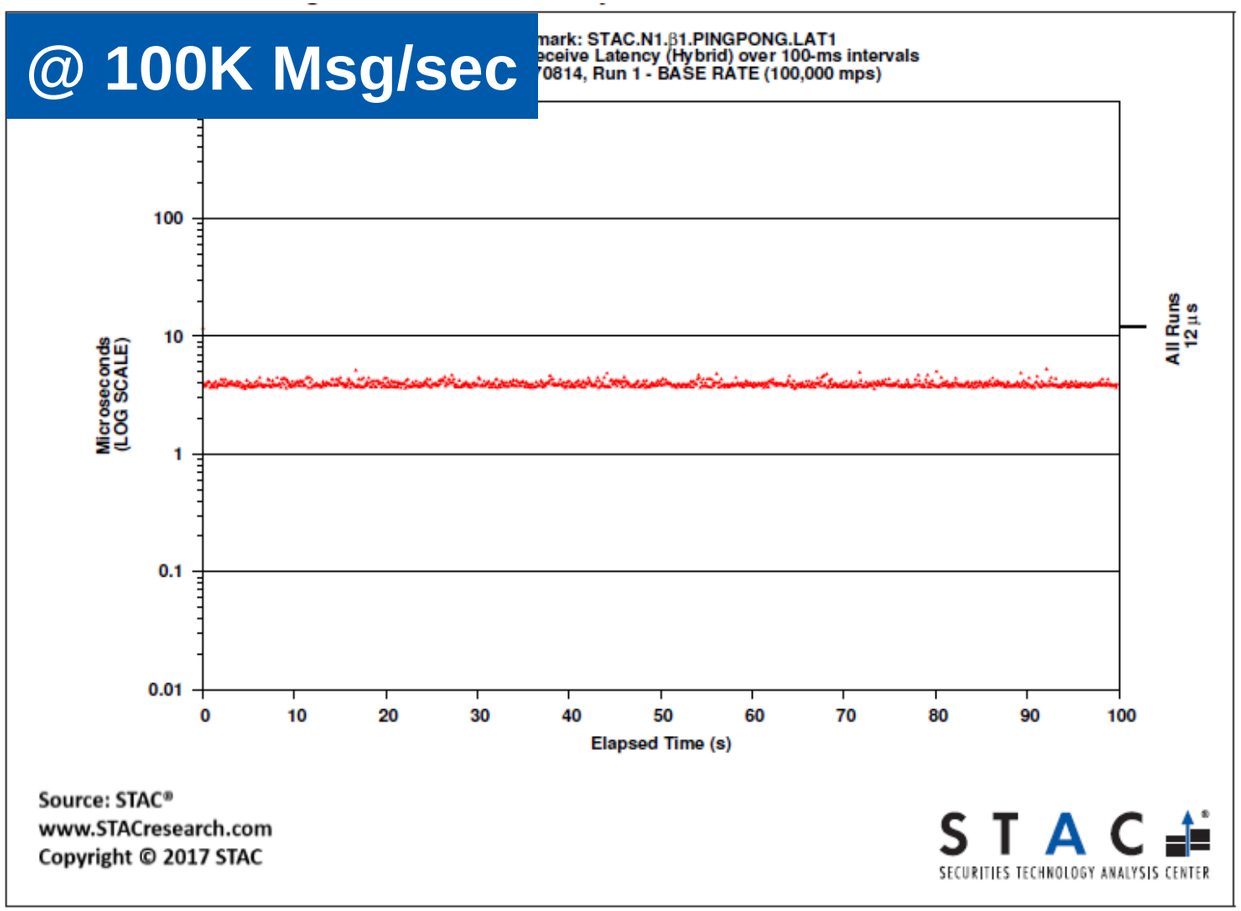


Figure 8 - Max Latency over time at BASE RATE

@ 1.6M Msg/sec

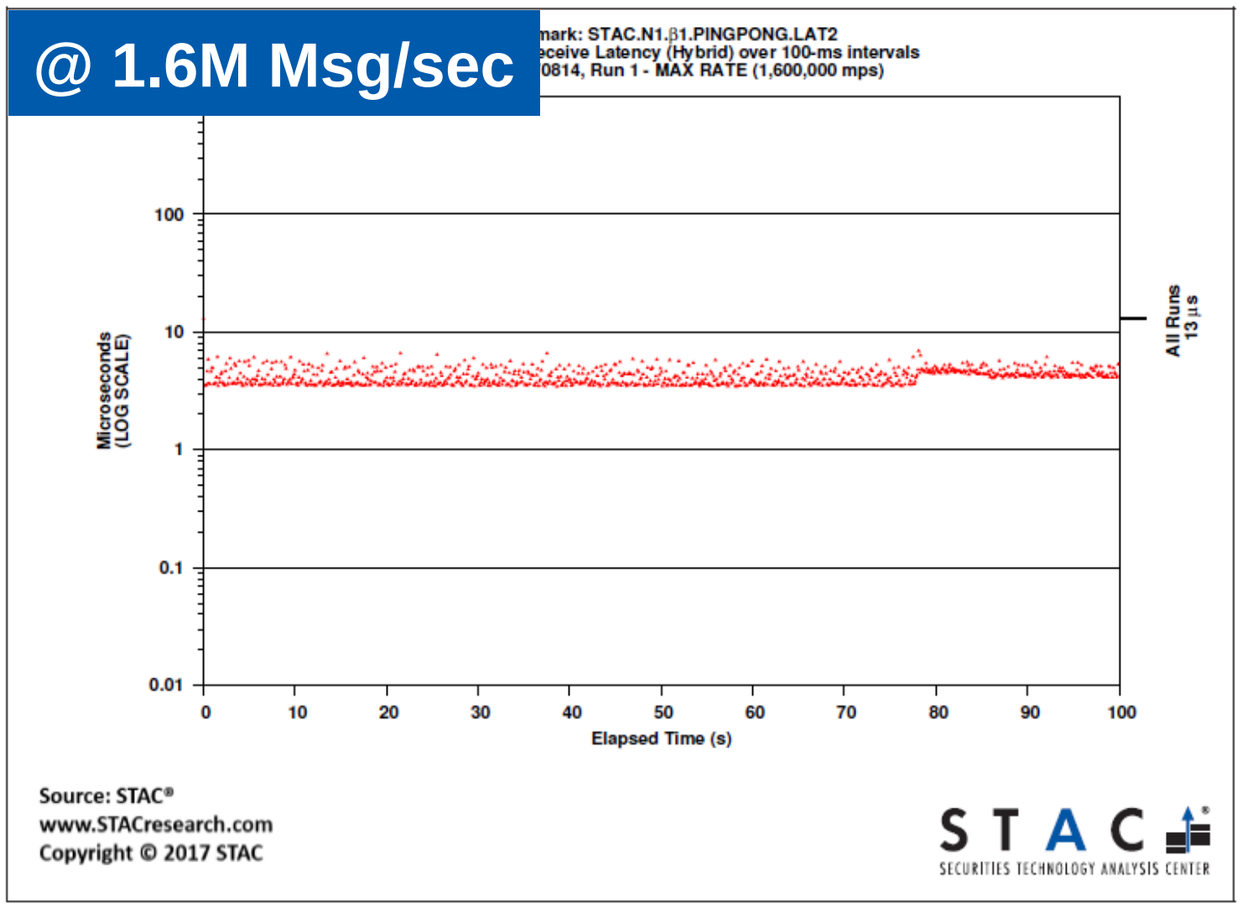
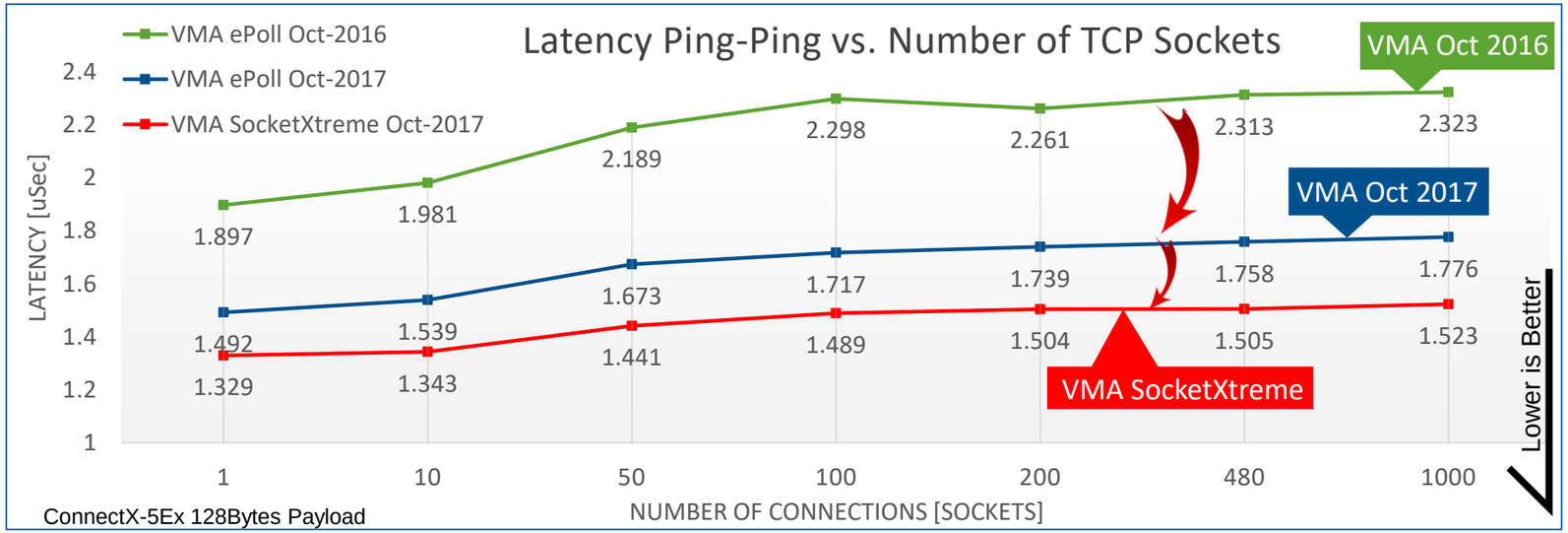


Figure 13 - Max Latency over time at HIGHEST SUCCESSFUL RATE

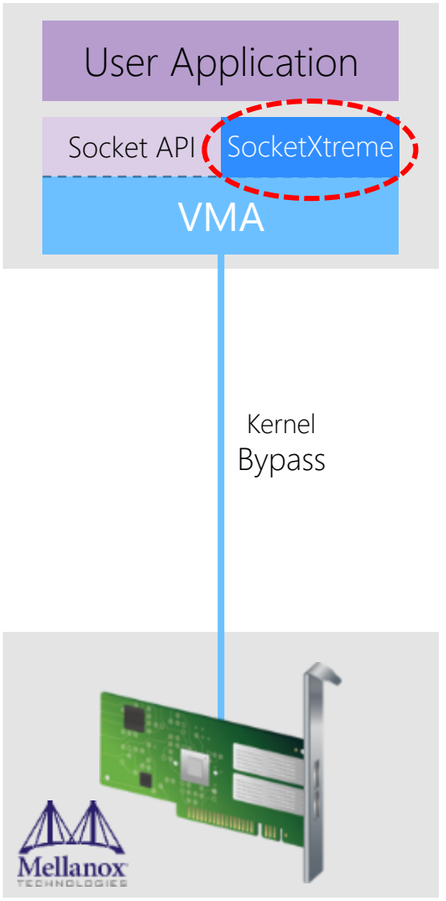
Sustaining Tightest Latency Distribution @ Highest Message Rates

VMA SocketXtreme

- Extending Socket API for achieving lower latency
- Supporting TCP, UDP and MC
- Free to use, no extra licensing
- Already deployed in production



Application with
Mellanox VMA SocketXtreme



Not STAC Benchmarks

VMA SocketXtreme

- Reduced overhead with direct access to hardware ring objects
- Socket API control interface kept (connect(), bind(), accept(), ..)
- Zero Copy - Packets & Events copied directly into the user buffer
- Accelerating ‘epoll’ with bypass
- Supporting thousands of sockets
- ConnectX-4 / Lx and ConnectX-5 / Ex

TCP Rx Multi-connections

Common Code

```
fd = socket(PF_INET, SOCK_STREAM, IPPROTO_IP);
rc = setsockopt(fd, SOL_SOCKET, SO_REUSEADDR, (char *) &flag, sizeof(int));
rc = bind(fd, (struct sockaddr *) addr, sizeof(*addr));
listen(fd, SOMAXCONN);
```

SocketXtreme

```
Wait for event
    while (0 == n)
    {
        n = _vma_api->vma_socketxtreme(_vma_ring_fd, vma_comps, max_events, 0);
    }
Receive Data
    memcpy(conns.msg, vma_comps[j].packet.buff_lst, vma_comps[j].packet.total_len);
    ret = vma_comps[j].packet.total_len;
   _vma_api->free_vma_socketxtreme(&vma_comps[j].packet, 1);
```

Standard BSD

```
Wait for event
    while (0 == n)
    {
        n = epoll_wait(efd, events, max_events, 0);
    }
Receive Data
    ret = recv(fd, conns.msg, sizeof(conns.msg), 0);
```

Scalable End-to-End Low Latency Interconnect Solutions

STAC Report™

STAC-N1™ Benchmarks β1 October 2017
Mellanox ConnectX-4 Lx running Mellanox VMA 8.4

Compared to all other public STAC-N1 reports of Ethernet-based SUTs, this SUT demonstrated:

- **Lowest mean latency at both the base rate and the highest rate tested.**
- **Lowest max latency at the base rate.**
- **Lowest max latency at or above 1 million messages per second.**
- **Highest max rate reported: 1.6 million messages per second.**
- **99.9999th percentile latency (six nines) of just 5 µsec at the base rate and 6 µsec at 1.6 million messages per second.**