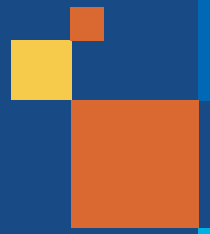
STAC Conference 2020

Move Data Faster

Intel Ethernet 800 Series Flexibility & Programmability



intel®

Notice and Disclaimers

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more complete information visit www.intel.com/benchmarks.

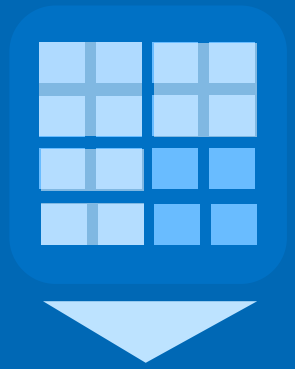
Performance results are based on testing as of dates shown in configurations and may not reflect all publicly available security updates. See backup for configuration details. No product or component can be absolutely secure.

Your costs and results may vary.

Intel technologies may require enabled hardware, software or service activation.

Changing Network Landscape

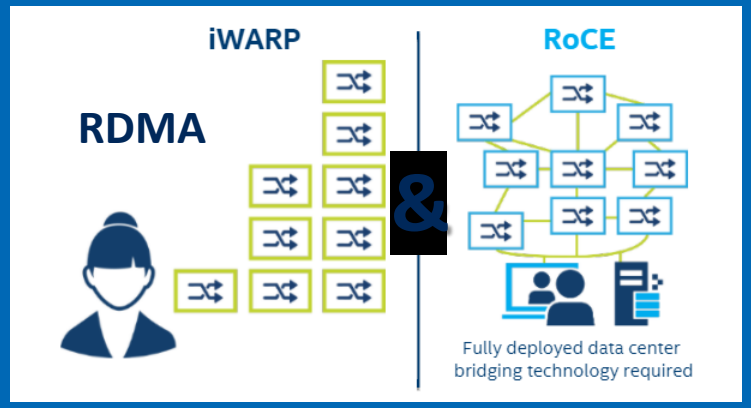
New Requirements for High Performance Networking



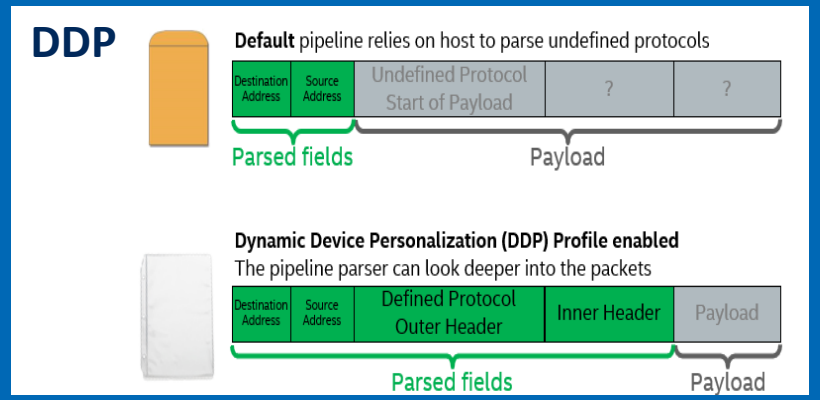
NV03 VXLAN NVGRE GENEVE
NSH C-VLAN S-VLAN
Q-in-Q MPLS GTP
IPoE L2TP PPPoE

Multiple Apps Contending For Network Bandwidth

Maintaining High-Performance Storage Access



Evolving Network Tunneling Protocol Support



Higher network bandwidth makes these requirements more challenging

Application Device Queues (ADQ) with Intel® Ethernet 800 Series

Application Device Queues (ADQ)

Meet service level agreements

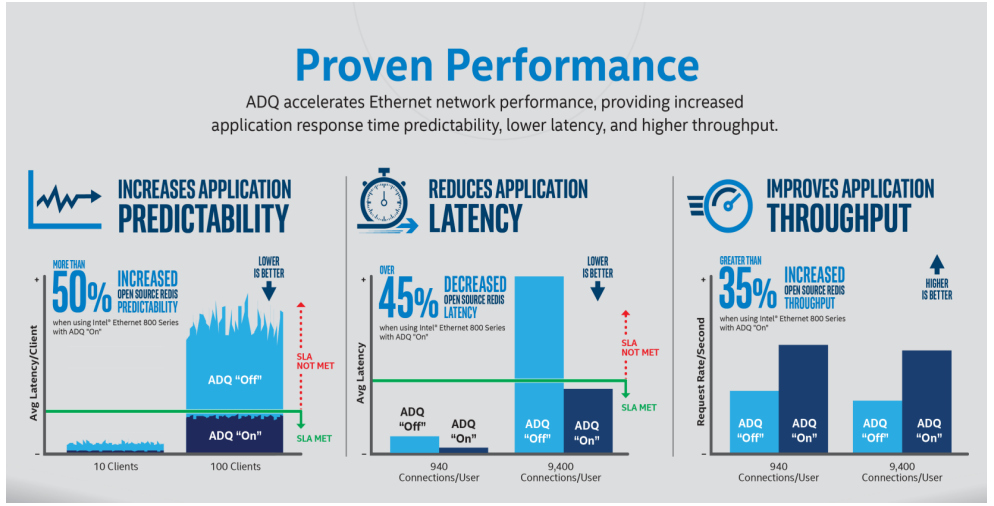
- Dedicates queues to high-priority applications to improve application response-time predictability, reduce latency, and improve throughput. Meet service level agreements better and scale service delivery to reach more end-users easily with ADQ.
- ADQ works by:
  - Filtering application traffic to a dedicated set of queues
  - Application threads of execution are connected to specific queues within the ADQ queue set
  - Bandwidth control of application egress (Tx) network traffic

Increases Application **Predictability**

Reduces Application **Latency**

Improves Application **Throughput**

Dedicates queues to high-priority applications



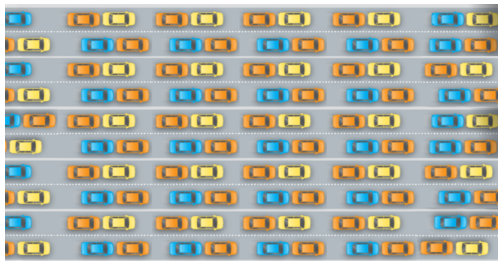
Improve application response time

https://www.intel.com/content/www/us/en/architecture-and-technology/ethernet/application-device-queues-technology-brief.html
https://www.intel.com/content/www/us/en/architecture-and-technology/application-device-queues-consistently-meet-service-levels.html

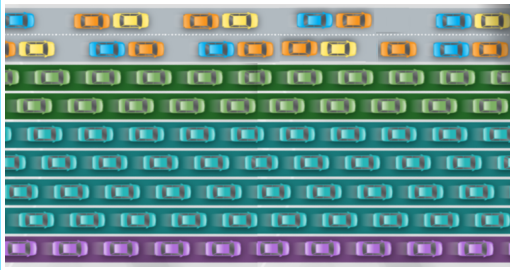
Intel® Ethernet 800 Series

Application Device Queues (ADQ) Performance Improvements

Without ADQ
Application traffic intermixed with other traffic types



With ADQ
Application traffic to a dedicated set of queues



Caching & Database[1]

MEMCACHED

Data / Key Caching

Up to 60% Predictability Increase

Up to 60% Latency Reduction

redis

>50% Predictability Increase

>45% Latency Reduction

>30% Throughput Increase

AEROSPIKE

Database Backend

>45% Predictability Increase

>15% Latency Reduction

>75% Throughput Increase

Storage[1]

Network Storage

>45% Latency Reduction

Significantly improves predictability, latency and throughput

1. Performance results are based on testing as of Feb 2020 (memcached), Feb 2019 (open source Redis), Sept 2019 (Aerospike) and Sept 2019 (NVMe/TCP) and may not reflect all publicly available security updates. See configuration disclosure for details. No product can be absolutely secure. For more complete information about performance and benchmark results, visit www.intel.com/benchmarks

https://www.intel.com/content/www/us/en/architecture-and-technology/ethernet/performance-testing-application-device-queues-with-memcached.html
https://www.intel.com/content/www/us/en/architecture-and-technology/ethernet/performance-testing-application-device-queues-with-aerospike.html
https://www.snia.org/sites/default/files/SDC/2019/presentations/NVMe-oF/Minturn_David_Vasudevan_Anil_Selecting_an_NVMe_over_Fabrics_Ethernet_Transport_RDMA_or_TCP.pdf

Application Device Queues (ADQ) Resource Center

For more information go to: http://www.intel.com/adq

Aerospike

- Aerospike Solution Brief
- Aerospike Blog
- Aerospike White Paper
- Intel Blog About Aerospike
- Aerospike Press Release
- Aerospike and Intel Joint Webinar
- Networking Tech Field Day

Memcached

- Memcached Solution Brief
- Steve OCP Summit Blog: Steve Schultz, VP CG
- OCP Summit ADQ Presentation Video

Open Source Redis

- Open Source Redis Solution Brief
- Networking Tech Field Day

NVMe/TCP with ADQ Acceleration

- SDC 2019 Technical Presentation Video
- SDC 2019 Technical Presentation
- Blog: Patricia Kummrow, VP DPG, Intel

Training

- Networking Tech Field Day
- Networking Tech Field Day with Aerospike

Additional Resources

- Intel Ethernet 800 Series Controller
- Intel Ethernet 800 Series Network Adapters
- Intel Ethernet Technologies

Intel® Ethernet 800 Series RDMA

RDMA Improvements

- iWARP and RoCE v2 RC & UD Transports
- 1 RDMA Enabled PF per Port
- **32 RDMA Enabled VFs per Device**

More RDMA resources

	Intel® Ethernet Connection X722	Intel® Ethernet 800 Series
RDMA	iWARP	**iWARP and RoCE v2**
Number of RDMA Reads	2, 8, 32, or 64	2, 8, 32, 64, **128, or 256**
Work Queue Elements (WQEs)	Fragments: 1 to 7 Sizes: 32B, 48B, 64B, 80B, 96B, 112B, 128B	Fragments: **1 to 14** Sizes: 32B, 64B, 96B, 128B, **160B, 192B, 224B, 256B**
Inline/Push Data Max Size	112B	**224B**
Host Memory Page Sizes	4KB, 2MB	4KB, 2MB, **1GB enhanced**
Outbound RDMA Read Queue Depth (ORD)	0 to 127	**0 to 255**
Protection Domains	Up to 32K	**Up to 256K**
Maximum Virtually Mapped Memory	2MB pages = 32TB 4KB pages = 1TB	**1GB pages = 256PB** **2MB pages = 512TB** 4KB pages = 1TB

OS Support

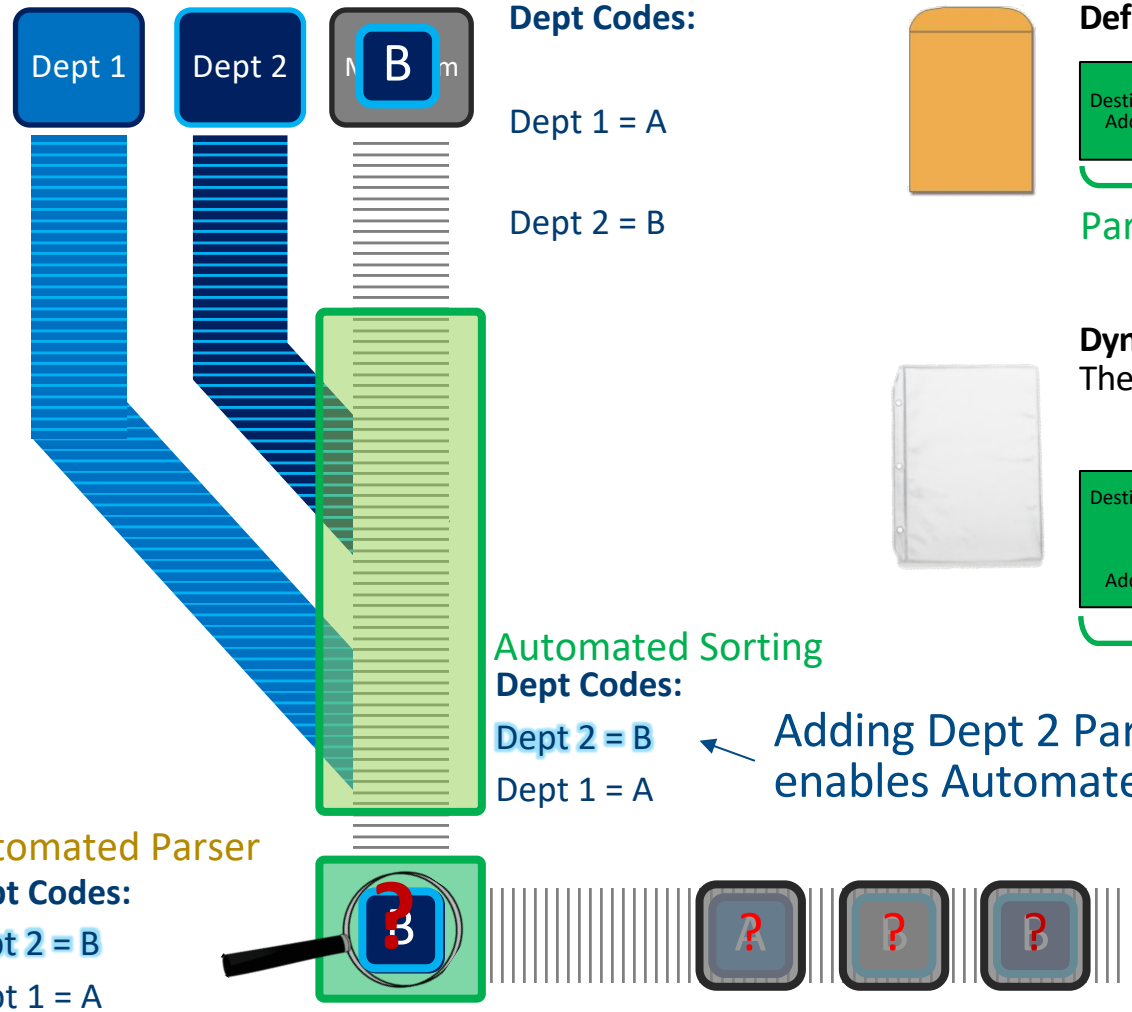
- Windows* Server 2016 (ND and NDK)
- Windows Server 2019 (ND and NDK)
- Microsoft* Azure Stack (NDK)
- Linux* OFED verbs RHEL* distribution
- Linux OFED verbs SUSE* distribution
- KVM Guests (Windows or Linux) with SR-IOV
- FreeBSD* with OFED verbs (Post PRQ)

Congestion Control Support

- TCP (iWARP)
  - Standard TCP (slow start, congestion avoidance, fast re-transmit, fast-recovery) RFC 5681
  - DCTCP (ECN-based, source estimates fraction of marked packets) http://research.microsoft.com/en-us/um/people/padhye/publications/dctcp-sigcomm2010.pdf
  - TCP-Bolt (TCP over lossless network, drops slow start) http://people.inf.ethz.ch/asingla/papers/tcp-bolt.pdf
  - TIMELY (estimates congestion based on RTT changes) http://conferences.sigcomm.org/sigcomm/2015/pdf/papers/p537.pdf
- RoCEv2
  - DCQCN (uses IBTA CNP packets, builds on ECN, QCN and DCTCP) http://conferences.sigcomm.org/sigcomm/2015/pdf/papers/p523.pdf
  - TIMELY (see above)

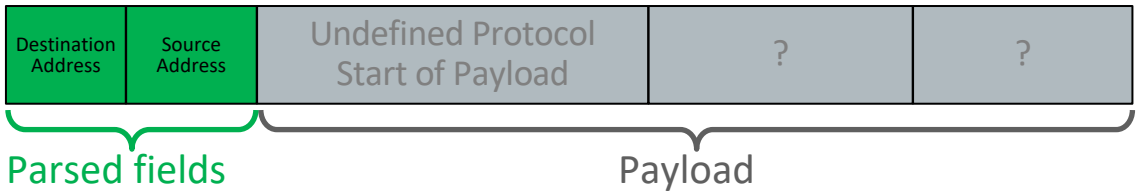
Why a Programmable Pipeline Matters

Analogy: Conveyer Belt Package Deliver

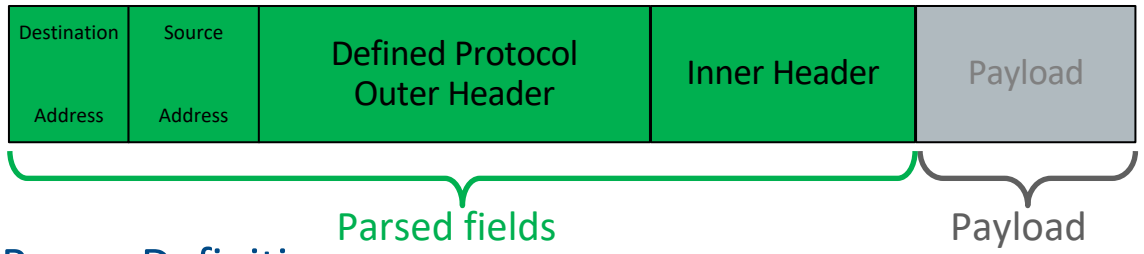


Adding Dept 2 Parser Definition enables Automated Sorting

Default pipeline relies on host to parse undefined protocols



Dynamic Device Personalization (DDP) Profile enabled
The pipeline parser can look deeper in to the packets



Summary of Dynamic Device Personalization (DDP)

Enables classification of new protocols using existing hardware

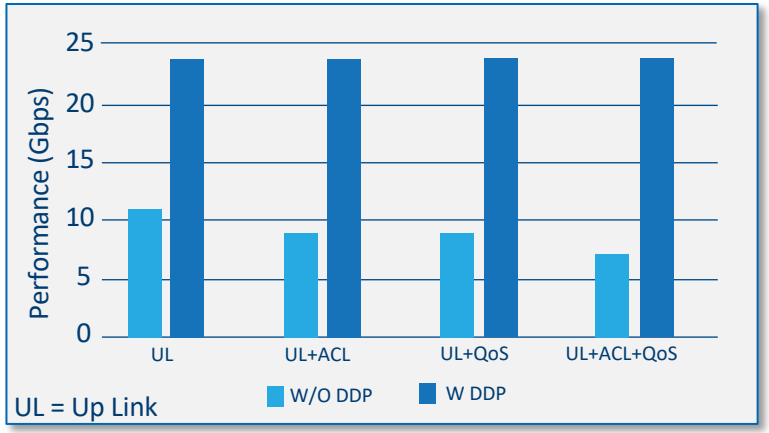
- Industry standard protocols for the Intel® Ethernet 700 and 800 Series
- Intel Ethernet 800 Series adds Enhanced DDP Packages

Available on Intel Ethernet 700 and 800 Series

- 10GbE, 25GbE, 40GbE, 50GbE, 100GbE
- PCI Express or OCP form factors
- Single, dual or quad ports

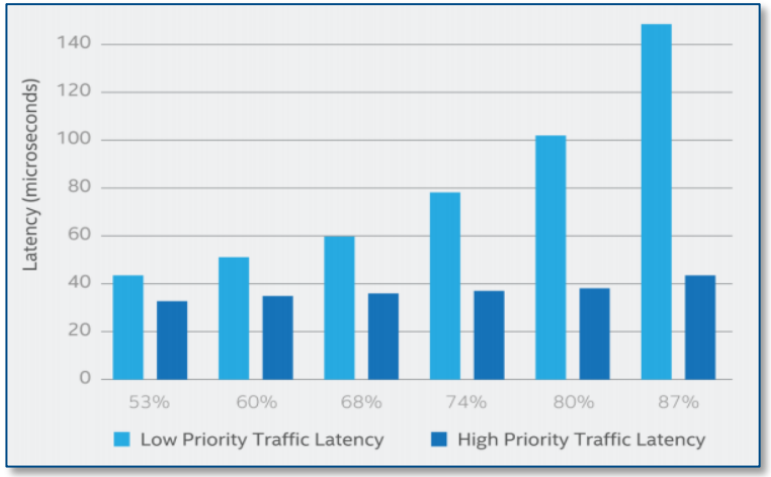
Improves network efficiency while reducing CPU utilization

- Improves packet per second processing rates
- Reduces processing latency and latency variation
- Reduces CPU utilization



Higher is better

2-3x throughput improvement in vBNG test case[1]



Lower is better

Latency vs CPU load for test case with 10% HP traffic[2]

Disclaimer: Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more complete information visit http://www.intel.com/performance.

[1]Data from Intel-NetElastic white paper [2]Data from Intel-SKT white paper

5 Timing Enhancements in Columbiaville

Constant PTP Clocking	• Link speed doesn't affect PTP precision
Timestamping in PHY	• Closer to wire time, avoids FEC timing artifacts
Native 25GbE Support	• No gearbox/PHY needed to support 25Gbps operation
Better Boundary Clock Function	• Common PHC time for upstream subordinate and downstream master clock
Timestamps for all RX Packets	• Meta-data includes PHC timestamp for all packets [at PF level], not just 1588 packets

PHC Timer Ticks

- In Fortville NIC silicon, 1588 timer tick rate depends on link rate:

Fortville Link Rate	40 Gbps	25 Gbps	10 Gbps	1 Gbps
Tick Frequency	625 MHz	625 MHz	312.5 MHz	31.25 MHz
Tick Period	~1.6 ns	~1.6 ns	~3.2 ns	~32 ns

- In Columbiaville NIC silicon, 1588 timer tick rate is constant:

Columbiaville Link Rate	100 Gbps	25 Gbps	10 Gbps	1 Gbps
Tick Frequency	812.5 MHz	812.5 MHz	812.5 MHz	812.5 MHz
Tick Period	~1.23 ns	~1.23 ns	~1.23 ns	~1.23 ns


intel

Intel® Ethernet 800 Series – For Cloud, Comms and Enterprise

Next-gen Intel® Ethernet Series Product Family

Samples: NOW

Production: July-august'2020

MAIN IMPROVEMENTS OVER THE INTEL® ETHERNET 700 SERIES

Higher Bandwidth

Intel's first NIC with PCIe 4.0 and 50Gb PAM4 SerDes

Improved Application Efficiency

Application Device Queues (ADQ), Dynamic Device Personalization (DDP), and support for both RDMA iWARP and RoCEv2

Versatility

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more complete information visit http://www.intel.com/performance.

FOR CLOUD

IMPROVED TRAFFIC STEERING TECHNOLOGY (ADQ) TO IMPROVE APP THROUGHPUT BY > 30%

- Support for common VF driver (Intel® Ethernet Adaptive Virtual Function) to reduce OPEX
- RDMA support for both iWARP and RoCEv2 providing a choice in hyper-converged networks
- 2X more virtualization resources vs 700-series for VM or container-centric environments

FOR COMMS

ENHANCED DATA PLANE DEVELOPMENT KIT (DPDK) SUPPORT

- Up to 100GbE to support high-performance workloads
- Programmable pipeline for enhanced Dynamic Device Personalization (DDP) features which can improve packet processing efficiency and reduce CPU overhead
- IEEE 1588v2 Precision Time Protocol for precise clock synchronization

FOR ENTERPRISE

READY TO USE MICROSOFT SOLUTIONS: STORAGE SPACES DIRECT, AZURE STACK

- Broad and flexible physical interfaces support ease of deployment
- Thorough test and validation with broad ecosystem of devices for interoperability
- Support for iWARP enabling ease of use in storage applications

Source: Intel internal testing as of February 2019; Redis Open Source on Cascade Lake with E810 100GbE on Linux 4.19.18 kernel