Workload-Aware NIC Profiles

Introduction
This document explains the performance profiles supported by Intel® PROSet for Windows* Device Manager. A performance profile provides a simple, one-click option for a user to configure the network adapter for a specific workload or server usage. At this time, four performance profiles are available to configure 10 Gigabit Intel® Ethernet Converged Network Adapters for optimal performance.

• Standard Server
• Web Server
• Hyper-V®
• Storage Server

It should be noted that performance tuning is a challenging task which typically includes configuration of the platform, operating system, networking stack or storage stack, and device driver. This document focuses on only the motivation behind creating these performance profiles and the performance improvements with these profiles on Microsoft Windows Server® 2012.
Standard Server Profile

The Standard Server Profile is provided to configure the adapter for optimal performance on a standard enterprise server, which can be a Microsoft Exchange® Mail Server, a File Server, etc. running on a non-virtualized server. In such workloads, the number of active TCP connections is small, which does not demand the best scalability. In general, the workload is not I/O-intensive or CPU-intensive. The performance tuning is thus focused on system efficiency. The Standard Server Profile provides good throughput with minimum CPU utilization and minimum latency. It provides a balanced performance profile for typical enterprise server usages. With modern server systems based on Intel® Xeon® processors, two CPU cores can deliver 10Gbps networking throughput for most workloads. The number of receive-side scaling (RSS) queues is thus set to 2 in this profile to reduce CPU utilization. The adaptive interrupt moderation algorithm is also fine-tuned to reduce latency for burst traffic patterns and to achieve high throughput. Below are the key settings for the Standard Server Profile:

- Receive Side Scaling Queues = 2
- Maximum number of RSS Processors = 16
- Interrupt Moderation Rate = Adaptive
- Virtual Machine Queues = Disabled
- Data Center Bridging = Disabled
- RSS Profile = Closest_Processor

When different workloads are running on the same server, quality-of-service (QoS) configurations in the networking stack should be considered for performance isolation and bandwidth reservation. We recommend enabling Data Center Bridging (DCB) when the network infrastructure is ready.

Web Server Profile

The Web Server Profile is provided to configure the adapter for optimal web server performance. For a web server, achieving maximum transactions per second with minimum CPU utilization is preferred. Transaction latency is less critical as long as the latency is not impacting user experience in terms of response time. To achieve maximum transactions per second, enable maximum RSS queues for maximum parallel processing of web requests. To reduce CPU utilization, the adaptive interrupt moderation algorithm is adjusted to medium or low interrupt rates (such as 6,000 interrupts/sec) per MSI-X vector, which reduces the total number of interrupts in the system. Below are the key settings for the Web Server Profile:

- Receive Side Scaling Queues = 16
- Maximum number of RSS Processors = 16
- Interrupt Moderation Rate = Adaptive
- Virtual Machine Queues = Disabled
- Data Center Bridging = Disabled
- RSS Profile = Closest_Processor

For web servers, QoS is not relevant, since only a single workload is running on the system. DCB is recommended to be disabled on such servers.
Hyper-V* Profile

In the Windows* Hyper-V environment, the cost of an interrupt is much more expensive due to additional context switches introduced, such as VM exit and VM resume. The CPU utilization of a Hyper-V server could be high enough that excessive interrupts in the system decrease the throughput significantly. Enabling VMQ reduces software processing overhead in the data path, but it could also increase device interrupts when each VMQ is assigned to a separate MSI-X vector. To control the total system interrupts, we recommend sharing the same MSI-X interrupt vector among multiple VMQs and limiting the number of processors for VMQ interrupts. MSI-X interrupt vector sharing for VMQs is enabled by default in the base driver. The base driver allocates only one interrupt vector on each selected CPU for VMQs. When two VMQs with the same CPU affinity are allocated in the base driver, these two VMQs are assigned to share the same interrupt vector. “Maximum number of RSS processors” is a setting used to limit the number of processors available for VMQ interrupt processing on an adapter. This networking stack uses this setting to decide the set of CPUs available for VMQ affinity. Since each CPU is allocated only one interrupt vector, the setting of “Maximum number of RSS processors” indirectly decides the number of interrupt vectors for VMQs as well. As RSS and VMQ do not coexist, the RSS Profile setting doesn’t impact VMQ. Based on the data from different benchmarks and test tools, four processors are able to deliver 10Gbps throughput for most workloads. The adaptive interrupt moderation algorithm is also adjusted to limit the maximum interrupt rate on VMQ interrupt vectors to avoid excessive interrupts when a large number of VMQs are actively running with high I/O throughput. Below are the key settings for the Hyper-V Profile:

- Receive Side Scaling Queues = 2
- Maximum number of RSS Processors = 4
- Interrupt Moderation Rate = Adaptive
- Virtual Machine Queues = Enabled
- Data Center Bridging = Disabled

Storage Server Profile

The Storage Server Profile is provided to configure the adapter for optimal iSCSI or Fibre Channel over Ethernet (FCoE) initiator performance. iSCSI runs on top of the TCP stack, so RSS settings are critical to iSCSI scalability. We recommend eight or sixteen RSS queues for iSCSI when the workload is composed mainly of small I/O requests. The Intel® Ethernet FCoE initiator software stack is designed for scalability and is optimized for multi-core systems. For maximum I/O transactions per second (IOPS), we recommend enabling eight FCoE queues. To reduce CPU utilization for large I/O transactions, the interrupt moderation rate is set to 8,000 interrupts/second. In the current Windows driver implementation, the miniport driver is managing interrupt vectors and FCoE queue affinity. At runtime, the TCP driver may do load balancing by adjusting the CPU affinity of RSS queues. When the affinity of a RSS queue is updated, the interrupt vector associated with the RSS queue is moved to a new CPU. For FCoE queues and RSS queues sharing the same interrupt vectors, the affinity of FCoE queues thus may be changed in the base driver without notification to the FCoE stack. This can cause misalignment in the FCoE stack and degrades FCoE performance. To avoid such side effects from runtime RSS queue affinity updates from the TCP stack, the current implementation allocates separate vectors for FCoE queues. QoS is critical for storage traffic. The storage stack is usually very sensitive to packet drop. With the Storage Server Profile, DCB is enabled on the adapter to provide no-drop service for storage traffic. Below are the key settings for the Storage Server Profile:

- Receive Side Scaling Queues = 8
- Maximum number of RSS Processors = 16
- Interrupt Moderation Rate = Medium (8000 interrupts/sec)
- Virtual Machine Queues = Disabled
- Data Center Bridging = Enabled
- RSS Profile = Closest Processor

In enterprise data center environments, there may be different workloads running on the same system. QoS configurations in the networking stack should be considered for performance isolation and bandwidth reservation. We recommend enabling DCB when the network infrastructure is ready.
Interrupt Moderation Settings

Interrupt Moderation is a key setting that could make a noticeable difference for most profiles. The base driver provides an adaptive interrupt moderation algorithm that considers the traffic patterns at runtime to achieve the highest throughput and lowest latency. The base driver also adjusts the adaptive interrupt moderation algorithm depending on the number of RSS queues enabled. When the number of queues is increased (such as to 16), the moderation algorithm is more aggressive to reduce total system interrupts. In a Hyper-V environment with many VMQs running, an interrupt rate limit (12,000 interrupts/second) is also applied to each VMQ interrupt vector to avoid excessive interrupts. Although the adaptive interrupt moderation algorithm is built to be very adaptive for each workload, we recommend measuring the exact performance of a specific workload. The following are the static settings of interrupt moderation available for performance tuning.

<table>
<thead>
<tr>
<th>Interrupt Moderation Setting</th>
<th>Effective Interrupt Rate (interrupts /second)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Off</td>
<td>Limited by CPU interrupt rate (usually &gt; 100,000)</td>
</tr>
<tr>
<td>Low</td>
<td>20,000</td>
</tr>
<tr>
<td>Adaptive</td>
<td>4,000 ~ 20,000 depends on runtime traffic pattern</td>
</tr>
<tr>
<td>Medium</td>
<td>8,000</td>
</tr>
<tr>
<td>High</td>
<td>4,000</td>
</tr>
<tr>
<td>Extreme</td>
<td>2,000</td>
</tr>
</tbody>
</table>
Receive Side Scaling Profile

Windows Server* 2008 R2 added an RSS Profile, which sets the policy by which the RSS cores are chosen. The profile chooses cores based on proximity to the NUMA node and whether or not the indirection table is static. The combination of these settings makes for the following profiles:

- Closest_Processor
- Closest_Processor_Static
- NUMA_Scaling
- NUMA_Scaling_Static
- Conservative_Scaling

The Closest_Processor_Static profile has the lowest memory overhead, so the driver sets this profile in Windows Server 2012. As this is already the default in Windows 2008 R2, the driver doesn’t need to set it. A detailed explanation of the RSS Profile is available at MSDN article on RSS keywords.
Performance Comparison

This section provides a high-level view of performance data from different profiles. The benchmark tool for each profile is chosen based on the targeted usages. For the Web Server Profile, the Microsoft Web Fundamental benchmark is used for testing. For the detailed system configuration and test setup, please refer to the Appendix on test configurations.

Web Server Profile:
Below is the performance comparison between the Standard Server Profile and the Web Server Profile. The Web Server Profile shows about a 65-percent increase in transaction rate compared to the Standard Server Profile.

![Web Fundamentals - Web Profile - 12 Clients/1000 Thread](chart)

Windows Server 2012/IIS 8.0 - Intel® Ethernet Converged Network Adapter x520

<table>
<thead>
<tr>
<th>Transactions/Sec</th>
<th>CPU Utilization (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,400,000</td>
<td>100</td>
</tr>
<tr>
<td>1,200,000</td>
<td>90</td>
</tr>
<tr>
<td>1,000,000</td>
<td>80</td>
</tr>
<tr>
<td>800,000</td>
<td>70</td>
</tr>
<tr>
<td>600,000</td>
<td>60</td>
</tr>
<tr>
<td>400,000</td>
<td>50</td>
</tr>
<tr>
<td>200,000</td>
<td>40</td>
</tr>
<tr>
<td>0</td>
<td>30</td>
</tr>
</tbody>
</table>

WF Stock (Random) 512B fixed 2KB B fixed

- Web Profile
- Standard Profile
- Web Profile CPU
- Standard Profile CPU
Storage Profile:
Below is the performance comparison between the two-queue Standard Server Profile with eight-queue behavior from the Storage Profile. The IOPS increased almost 80 percent at smaller buffer sizes while using eight queues.

Virtualization Profile:
Below is the performance comparison between the Standard Server Profile and the Virtualization Profile. The Virtualization Profile shows about a 60-percent increase in throughput at larger buffer sizes, which leads to the associated increase in utilization.
Summary
Performance tuning is a challenging task. Each server usage or workload can have quite different properties, which usually require different driver settings. Performance profiles provide a simple option to quickly configure a server’s Intel® Ethernet adapters or controllers for well-known workloads. For new workloads from a specific usage model, and even for workloads described above, we recommend analyzing the actual workload properties, running tests, and measuring performance to determine optimal settings. The user can then save those settings in a custom profile using Intel PROSet for Windows Device Manager. In a future software release, Intel plans to add a performance profile for FSI (Financial Service Industry) workloads.

Appendix
System configurations and test setup

- Tested the Intel® Ethernet Converged Network Adapter X520 in a Dell PowerEdge® R720 (Intel® Xeon® processor E5-2690 dual-socket) running Windows Server 2012. Data compares the indicated Profile against the Standard Profile using web traffic.
- Driver information
  - ixn63x64 = 3.2.61.0
  - Only driver changes made were to the profile setting
- System under test: Dell PowerEdge R720
  - 2x Intel Xeon processor E5-2690 (20-M cache, 2.90 GHz, 8.00 GT/s Intel® QuickPath Interconnect)
  - Intel® C602 Chipset
  - RAM: 64-GB DDR3 RAM
  - BIOS: 1.4.1
  - BIOS power management and PCI Express* ASPM Disabled

- WCAT (Web Capacity Analysis Tool) 6.4.1 with Web Fundamentals: Ran the default wf.proxy.hot.ka test (random req. size from 1-4K bytes) as well as a fixed 512 bytes and 2 kilobytes req. for transactions/sec and throughput. Testing used 12 clients running 1,000 threads each.
- Testing was against Internet Information Services 8.0 and the Web Fundamentals installed site pages

Source: Intel Networking Division SW Performance Lab