README file revision:

Suggested tuning for systems with ATTO 10 Gigabit Ethernet Adapters

With ATTO Fast Frame 10GbE Network adapters, the default operating system configurations will very likely limit the total available bandwidth due to TCP auto-tuning stacks inherent to the available operating systems. To overcome this, there are many sets of possible variables that, when applied together, will increase the overall ability for the operating systems to transmit and receive data. Depending on your workflow needs, TCP tuning is subjective and not an all encompassing fix all to performance issues. Each network needs are different and each application behaves differently so it is recommended to investigate the various settings within your own network to find what works best for your needs.

The settings covered in this document are only suggestions and act as a starting point for tuning your 10GbE performance. Your “mileage may vary” is cited by many 10GbE forums and support sites and it is recommended that you “tweak” the settings to find what works best for your particular environment. Each operating system has an “auto-tuning TCP Stack”, performance tuning and file parameters may be different for each specific setup and platform. All parameters discovered are from multiple sources and vary in degree of affect on performance.

The main steps you need to follow to ensure you are taking advantage of the available 10 Gigabit Ethernet bandwidth, are listed here in order of priority and effect on performance:

1. Verify that the adapter is installed in a proper PCI Express (PCIe2.0) slot with a width greater than or equal to 8X. The minimum requirement to achieve 10-Gbps in one direction is to install the card into aPCIe2 x 8 slots. This is a common oversight during the installation and will avoid performance issues if the slot width and speed are confirmed. *Of special note is the fact that many manufacturers may label the slot as x8 but physically wire it with fewer channels. In these cases, what may be marked as x8 is only running at x4, x2 or x1 slot speeds.*

Number of Lanes	**Bandwidth Per Direction for PCIe 1.0**	**Bandwidth Per Direction for PCIe 2.0**
1	250 MB/s, 2Gbps	500 MB/s, 4Gbps
2	500 MB/s, 4 Gbps	1GB/s, 8 Gbps
4	1GB/s, 8Gbps	2GB/s, 16Gbps
8	2GB/s, 16 Gbps	4GB/s, 32 Gbps
12	3GB/s, 24 Gbps	6GB/s, 24 Gbps
16	4GB/s, 32 Gbps	8GB/s, 64 Gbps
32	8GB/s, 64 Gbps	16GB/s, 128 Gbps

Fig.1: *Common PCIe bandwidth chart.*

2. Ensure that “Jumbo” frames (i.e. MTU=9000) are enabled on ***ALL*** 10GbE devices in your network subnet. (Clients, Servers, Switches, Routers, NICs…etc) The MTU parameter for the ATTO 10GbE NIC/CNA is adjusted via the driver properties in your operating system. (i.e. Linux, uses ifconfig <adapter#> mtu=9000, in Windows: adjust MTU from the “Advanced” tab in the NIC driver properties page.)

3. RSS (receive-side scaling) configuration. Make sure that the TCP traffic generation and reception is distributed across ALL the available processor cores by selecting the appropriate RSS configuration. For example; if you have a processor with 4 CPU cores, then you should configure 4 RSS queues in the NIC driver properties. RSS facilitates distribution of traffic across the available processor cores by separating the traffic into multiple queues equal to the number of cores in the system. (In Windows, this is adjusted in the advanced tab of the driver properties. In Linux, the modprobe command[7].)

4. Make sure that large segment offload and TCP checksum offloading are enabled,(these are enabled by default in the ATTO 10GbE NIC/CNA’s but confirm them anyway) The TCP Checksum Offload option enables the network adapter to compute the TCP checksum on transmit and receive operations to save the CPU from having to process the checksum. The performance advantage will vary by packet size. Smaller packets have little to no benefit from this, while larger packets will have larger benefits to performance. The use of TCP offload in concert with RSS improves server performance significantly.

5. Operating Systems in general tend to offer less performance in receive mode than in transmit mode, so in typical performance test with two machines having identical hardware, the sender can overwhelm the receiver causing it to drop frames. To alleviate this issue, you should make sure that the TCP selective acknowledgements (TCP_SACK) are enabled on both the adapter and the Operating System.

With TCP offloading disabled, the TCP/IP stack in the Windows operating systems, automatically implements selective acknowledgments.

When enabled ***AND*** the TCP/IP stack is implemented on the network adapter card, you may need to verify SACKS is enabled at the network adapter layer.

Make sure that all options enabled on the ATTO 10GbE NIC/CNA's such as - large segment offloading, TCP offloading, RSS, and TCP scaling…etc - are enabled in the same way on all operating systems in the 10GbE network subnet.

(i.e. In Microsoft Windows Server 2008, you can check this configuration by using the command **netsh int tcp show global.** In Linux, use the modprobe command).

6. Modify the Operating System Kernel parameters to allow the operating system to handle TCP auto tuning more effectively. In Linux and OSX this is done by editing or adding the ***sysctl.conf*** file to the operating system. Not all settings will have an effect and some may not be needed at all. It depends on your use of the systems, your performance needs and the versions of the network stacks in use on your systems. Below is one recommended starting point:

   You can use the following command to make temporary modifications to the kernel in order to test each tuning parameter and see the effects; a reboot will reset your changes. Editing the sysctl.conf file with the values will make them persistent across rebooting.

   Run*: sudo sysctl -w setting.name=newvalue to set the setting until you reboot.*

   Once you know the results of your changes, you can create a file at ***/etc/sysctl.conf* –Linux** or ***/private/etc/sysctl.conf-* MacOSX**, and drop the settings in one line at a time.

ATTO recommended sysctl.conf settings for OSX:

```
kern.ipc.maxsockbuf= 4194304 – will not go higher in OSX due to system limitations.
net.inet.tcp.sendspace= 2097152
net.inet.tcp.recvspace=2097152
net.inet.tcp.maxseg_unacked=32
net.inet.tcp.delayed_ack=2
kern.maxnbuf=60000
kern.maxvnodes=280000
net.inet.tcp_sack=1
```

Other notes:

1. All File systems "auto-tuning" TCP stacks greatly affect the performance of 10GbE NICS.
2. To get full TCP performance the TCP windows need to be large enough to accommodate the Bandwidth Delay Product. The BDP tells us the optimal TCP window size needed to fully utilize the line. To keep the pipe full, you must push data onto the wire, at your given bandwidth, for as long as an entire Round Trip Time (RTT). That is, the receiver must advertise a window size big enough to allow the sender to keep sending data right up until he begins receiving acknowledgments.
3. With the 10 Gigabit network adapters, the default operating system auto-tuning TCP/IP configurations will most likely limit the total available throughput artificially. Re-tune your systems with settings that work best for your network and workflow requirements.
4. OSX has set limitations on the ability to "tune" the kernel beyond certain limits, whereas other operating systems allow more flexibility. (i.e. Linux allows the kern.ipc.maxsockbuf to be set at 16MB, OSX 10.6.x and above limits this to 4MB.)
5. With "Out of the box" operating systems, it appears the Linux NFS and Windows sharing (SMB) does the best at taking advantage of 10GbE performance with little to no modifications of the auto-tuning TCP/IP stacks.

Other Performance Tuning Suggested References:

- http://support.microsoft.com/kb/224829/en-us
- http://www.redhat.com/promo/summit/2008/downloads/pdf/Thursday/Mark_Wagner.pdf
- http://msdn.microsoft.com/en-us/windows/hardware/gg463392.aspx
- http://www.intel.com/support/network/sb/CS-025829.htm

10GbE Document Resources:

- http://www.intel.com/support/network/sb/CS-030881.htm
- http://www.cisco.com/en/US/prod/collateral/switches/ps9441/ps9670/C07-57282810_10Gb_Conn_Win_DG.pdf