# Cooking the Numbers

## A Revealing Discussion of EMC's Measurements of NetApp SAN Performance

**Stephen Daniel Chris Lemmons, Network Appliance, January 25, 2007, TR 3521-6**

daniel@netapp.com

lemmons@netapp.com

## Executive Summary

On October 16, 2006 EMC published a comparison of a CX3-40 and a FAS3050C using a workload that simulated Microsoft® Exchange Server connected to a storage array via a Fibre Channel SAN. It is our belief that EMC's publication substantially misrepresents the performance of the FAS3050C under this workload.

This paper provides the results of NetApp measurements of the performance of the FAS3050C on the simulated workload described by EMC. Additionally, it discusses how the ability of the FAS3050C ability to optimize RAID writes enables it to exceed the performance demonstrated by EMC using their CX3-40.

## 1. Introduction

On October 16, 2006, EMC published a Technical Note entitled “NetApp Performance in SAN Environments”[Reference 1]. This paper represents an allegedly good-faith effort to compare a CX3-40 and a FAS3050C on a SAN workload, and it shows the FAS3050C not performing as well as the CX3-40.

EMC’s paper states that results were “configured and tuned for maximum performance following the vendor’s best practice guidelines.” (Later versions of the paper omit the words “best practice.”) No URL or other reference was provided to show what guidelines were actually used.

In this paper we show that FAS3050C performance is superior to the performance that EMC demonstrated on either the FAS3050C or their CX3-40.

### 1.1 Provenance

EMC’s paper was published without any attribution or contact information. This paper and the data behind it were written and generated by the authors, who stand behind the work. This paper has been reviewed by the performance community in Network Appliance™ Engineering and represents our consensus view of the performance of our systems. We welcome questions and comments from our readers.

## 2. Workload Description

EMC’s paper makes extensive use of a simulation of the access patterns generated by a Microsoft Exchange Server. Specifically, the workload consists of random reads and writes to a set of LUNs. All reads and writes are 4KB in size. The generated read/write ratio is 2:1.

Microsoft best-practices for Exchange require excellent storage system response time for both reads and writes. No storage system can deliver the required response time while being driven to achieve maximum possible throughput. For this reason, any Exchange simulation must be run at less than full system throughput. Rather than targeting a specific response-time threshold, EMC simply declared that a storage system’s throughput for Exchange is equal to 80% of maximum throughput. In addition, EMC simulated an Exchange I/O load of 1 I/O operation per second per user . This value is within the range of values suggested by Microsoft. As a result, in this paper EMC converts observed I/O throughput to simulated users by simply multiplying by 0.8.

Throughout this paper we use the workload described by EMC. We measure throughput in I/O operations per second and then multiply the result by 0.8 to match the scaling performed by EMC. This allows us to compare our results to EMC’s stated results. Our usage of this workload does not constitute an endorsement of measurement methods described by EMC. Specifically, our research suggests that this workload does not accurately predict storage system performance in Exchange environments and should not be used as part of a sizing exercise for an Exchange deployment.

## 3. Configuration

On the configurations described by EMC, their Exchange simulation is disk-bound. Both the CX3-40 and the FAS3050C are capable of higher performance, but only if more disks are provided. EMC describes testing the CX3-40 with 100 disks and the FAS3050C with 96 disks. Vendor best practices suggest that disks be configured in multiples of 10 on the CX3-40 and multiples of 16 on the FAS3050C. Thus 96 and 100 are as close to the same configuration as possible within these guidelines. However, on a disk-limited workload the configuration difference gives a 4%

advantage to the CX3-40. In practical deployments, NetApp customers routinely use slightly larger or smaller RAID groups to fully utilize the entire set of disks available on their system.

EMC's paper suggests that EMC tested the FAS3050C with only one active Fibre Channel loop connecting each storage controller to its disks. We configured ours according to a more standard high-performance configuration and used two active loops per controller.

EMC describes testing each storage configuration with eight Fibre Channel loops connecting the storage controllers to the load generators. This is overkill. We used four active loops.

To match EMC's configuration, all tests were performed against 12 LUNs, each 792 GB in size. EMC appears to define 12 × 792 GB as the usable capacity of the system. Tests performed at less that 100% use all 12 LUNs but restrict the test to running over a portion of each LUN.

Worth noting is the fact that the usable capacity of the two systems is approximately equivalent. The CX3-40 uses somewhat more disk space for parity drives, while the FAS3050C uses somewhat more space to manage continuous disk optimization. Both systems have approximately 80% of the raw space available to store user data.

Despite detailed documentation on some aspects of the benchmark, a close reading of the EMC paper fails to answer the following questions:

1. How many systems were used to drive load to the storage systems?
2. What tool was used to generate the load?
3. What operating system ran on the load generator hosts?
4. What switches, if any, were placed between the load generator hosts and the storage systems?
5. What multipath software was used?
6. Was any kind of ramp-up used to warm up the storage system?
7. How long was the measurement interval?

For our tests we used IOMeter (available for free at http://sourceforge.net/projects/iometer/). We drove the load from two hosts running Windows® Server 2003 Enterprise Edition. There were no switches between the load generators and the storage system. We used a 2-minute ramp-up and a 5-minute run to measure our results.

# 4. Performance at 50% Storage Utilization

## 4.1 Results

We began by attempting to use the limited disclosure information in the EMC paper to reproduce EMC's measurement of performance at 50% storage utilization. When we followed NetApp best practice guidelines [Reference 2], we found the system performance to be substantially superior to that reported by EMC. Indeed, the FAS3050C outperformed the CX3-40 by more than 14% on this test.

NetApp guidelines mandate the use of the NetApp SnapDrive® tool to provision the LUNs to a Windows system. However, by manually provisioning the LUNs and deliberately making significant mistakes in the process, we were able to generate a lower performance number, labeled in Figure 1 as "detuned." Even this detuned number exceeds the performance published by EMC. We were unable to generate a number as low as that published by EMC.

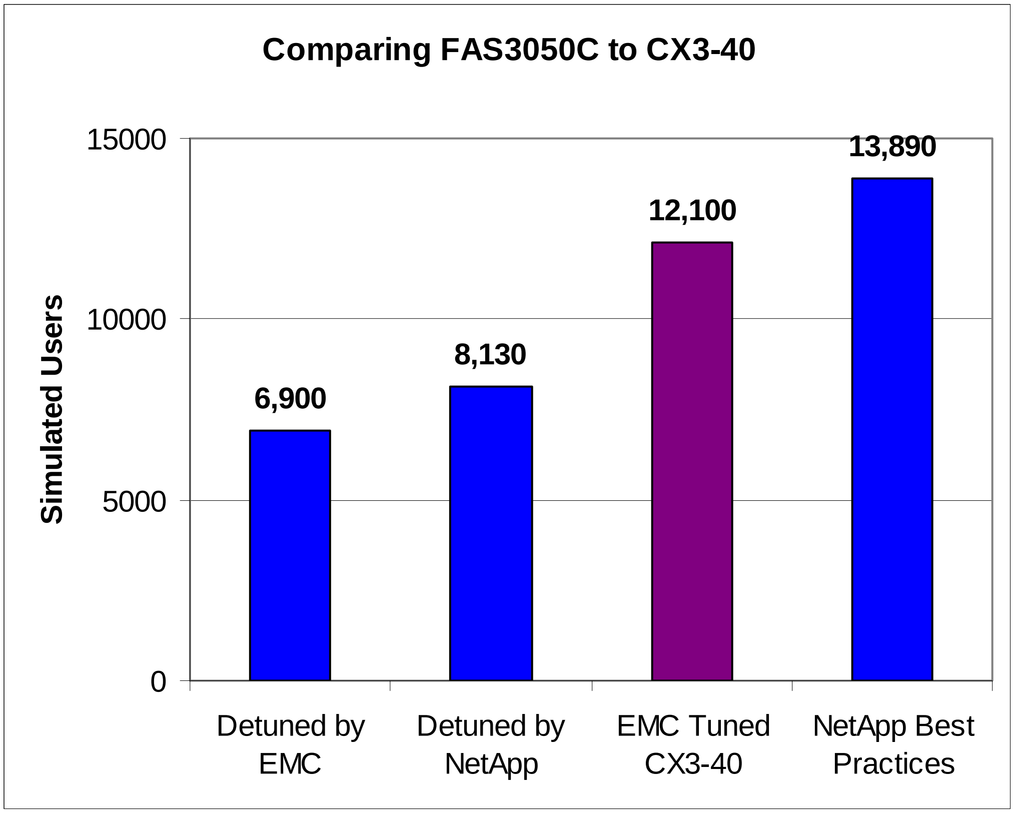

Figure 1) Detuned and tuned performance on FAS3050C and CX3-40.

### 4.2 Tuning the Memory Management System

The workload that EMC describes is quite simple, a flat random distribution of reads and writes. This workload fails to represent the vast majority of commercial applications. Such applications have significant variation in the frequency accesses across the data space.

Because the distribution of accesses described by EMC does not match real-world access patterns, the memory management system in NetApp storage systems does not optimize this access pattern. If we choose, we can inform the memory management system that the presented workload does not benefit from caching, freeing memory for other purposes. When we instruct the system to enable the “reuse” memory policy, performance improves, as shown in the following figure.

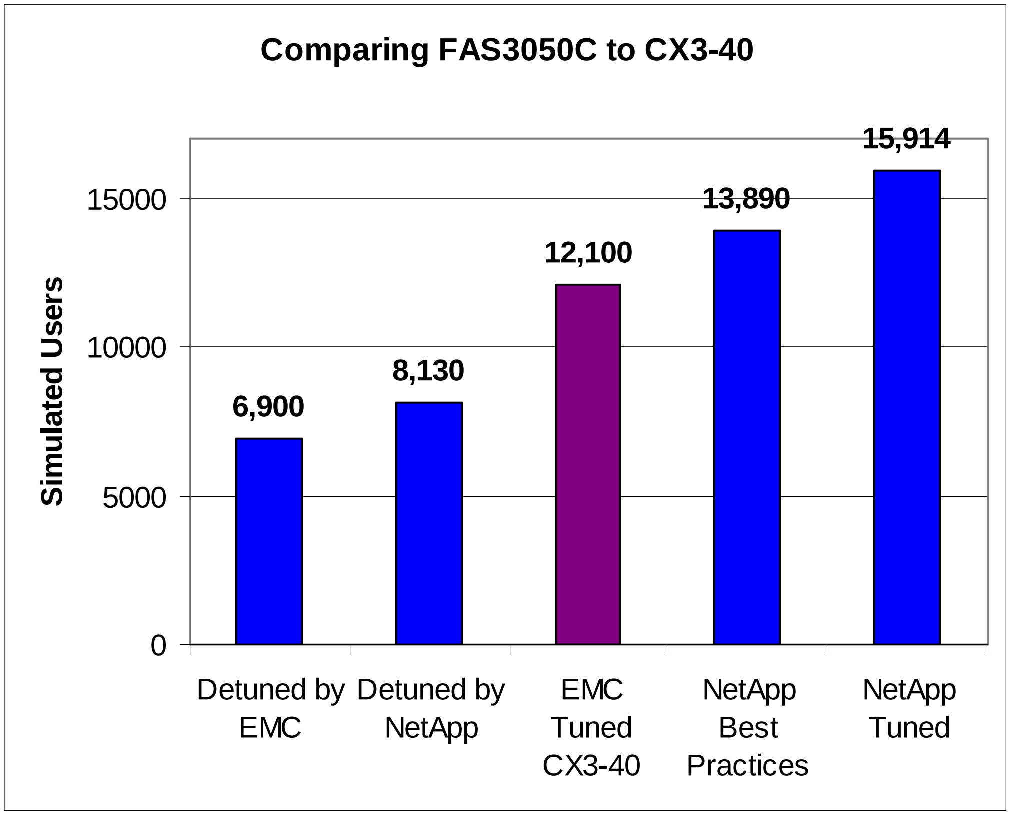

**Figure 2) Effects of tuning the NetApp system for best performance.**

The column labeled “NetApp Tuned” refers to tuning the memory management system for this unusual workload.

## 5 Performance Across Utilizations

After establishing that NetApp storage worked well at 50% utilization, we went back and measured across the entire range of utilizations referenced by EMC. For this exercise we used what we learned from the study at 50% and used the tuned memory management throughout.

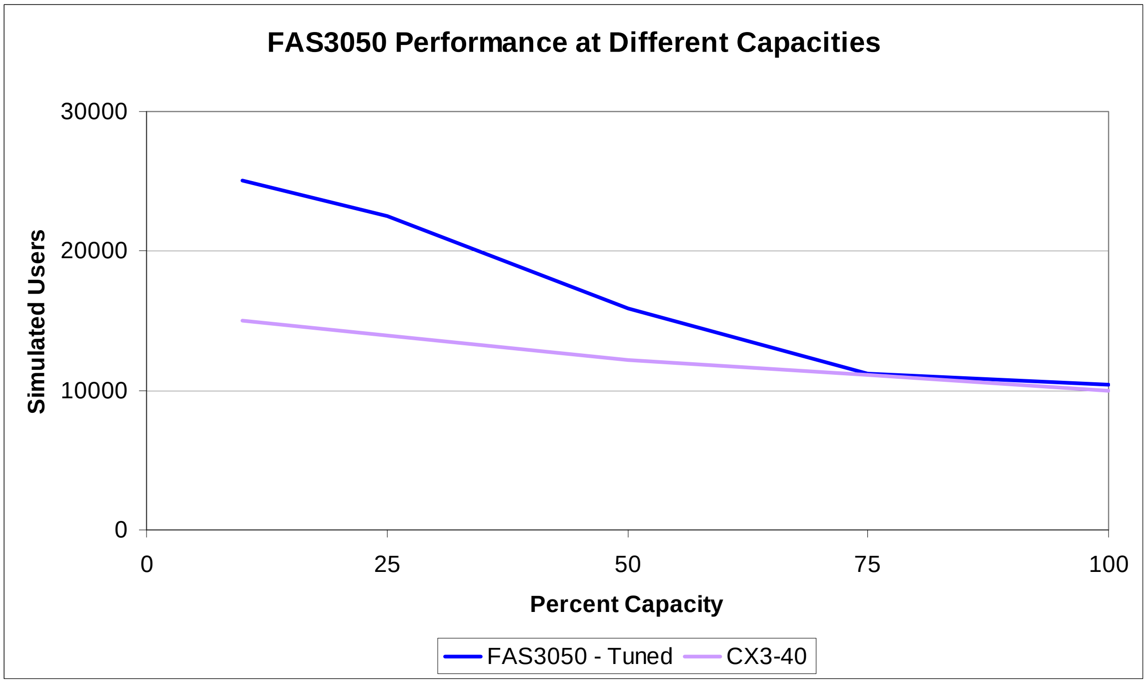

Figure 3) Relative performance over increasing capacity utilization levels **(low concurrency).**

Our initial measurements showed that the FAS3050C appeared to lose its performance advantage over the CX3-40 at high capacity utilization. We found this result surprising. There are sound theoretical reasons why NetApp technology should result in better performance on this workload independent of storage capacity.

Our analysis quickly revealed the problem. Following the description in EMC's paper, we ran this benchmark with a very modest number of outstanding I/O requests per disk drive. However, increasing the I/O concurrency provides the following results, which are more in line with our expectations.

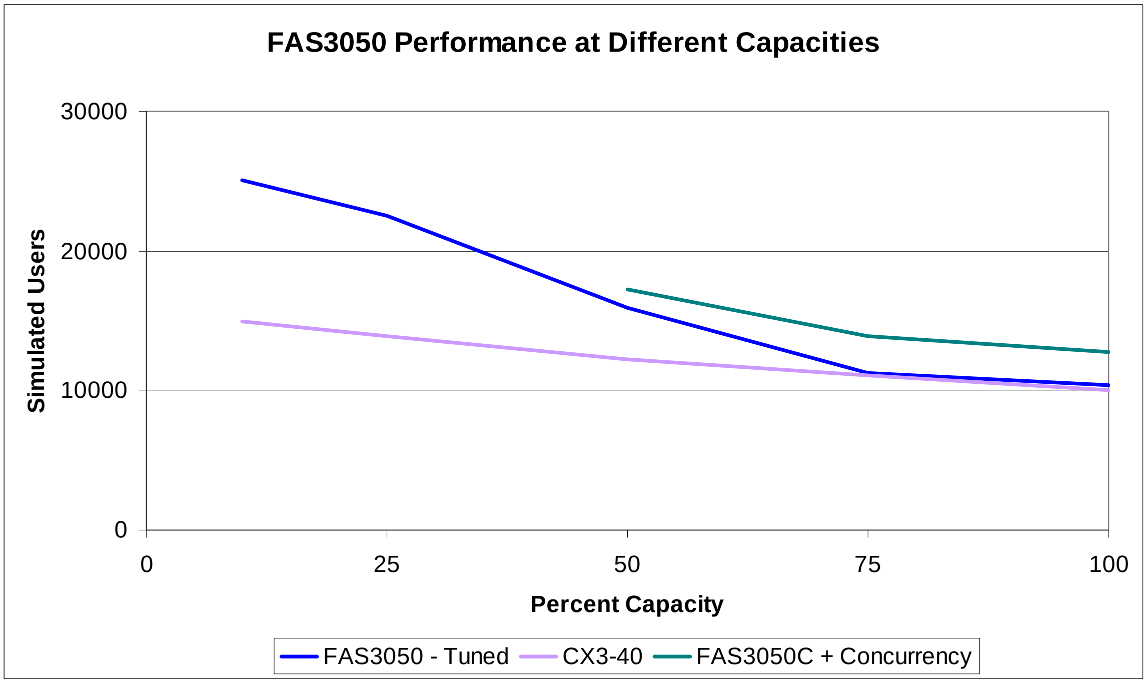

Figure 4) Relative performance over increasing capacity utilization levels (**proper concurrency**).

We were unable to measure the impact of this concurrency change on the CX3-40. Our analysis suggests that the change would have minimal impact.

Using EMC's chosen workload, NetApp has demonstrated with these results a higher level of performance compared to EMC's CLARiiON at all levels of capacity utilization. This further illustrates the automated performance-maximizing advantages of the NetApp advanced virtualization technology. These results are discussed in detail in section 7, "Analysis."

## 6 Performance Over Time

After demonstrating superior performance on a correctly configured system at various capacities, we began a measurement of the performance variation over time on a correctly configured NetApp FAS3050C.

To reproduce the measurements demonstrated in EMC's Figure 2 (not shown in this paper), we used the same storage configuration reported by EMC, 6 RAID groups of 16 disks each, two LUNs to a RAID group. We used IOMeter to generate a mix of random reads and writes, with twice as many reads as writes. All I/O operations were 4KB and scattered randomly over 25% of the available storage.

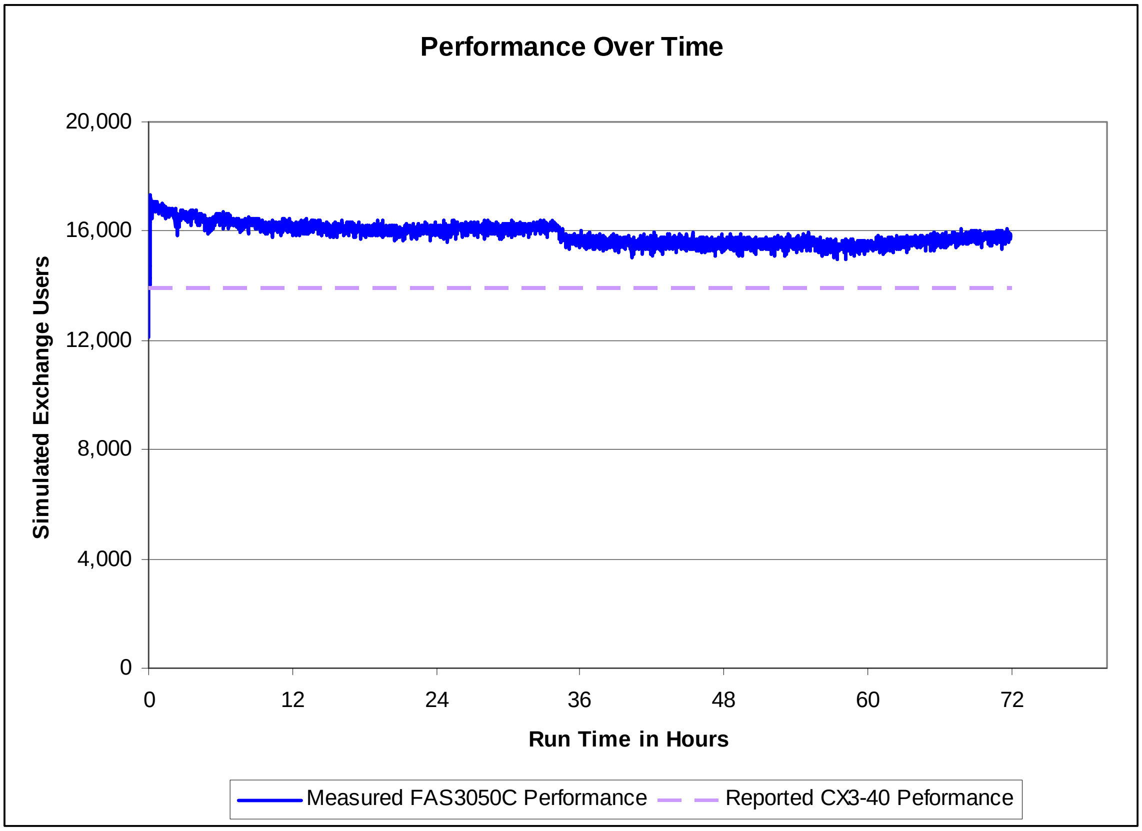

Figure 5) Simulated Exchange performance over a 3 day run

By the end of our 3-day run, performance had stabilized at approximately 15,800 users, down from an initial peak of 17,000 users measured on the newly initialized volumes, a 7% drop. EMC's Figure 2 does not show the performance of their system over time on this workload, but reports an average rate of about 13,900 simulated users.

## 7. Analysis

We believe that customers should understand why the NetApp FAS3050C and the EMC CX3-40 produce different results on this benchmark. This is a benchmark whose performance is limited by the performance of the disk drives. Since neither NetApp nor EMC makes their own disk drives, how does one vendor achieve superiority over another?

### 7.1 Hardware Configuration Issues

The NetApp and EMC systems are similar. They both use the same disk technology, and both systems used a pair of controllers (storage processors) to manage the workload. There are a few differences; however, since this workload primarily emphasizes the common disk technology, the various hardware differences are immaterial. EMC acknowledges this in their paper: "***The 4 Gb/s interconnects used by the CLARiiON CX3-40 provided no performance advantage on this test** …*"

We believe that the hardware configurations are functionally very similar. The only measurable difference should come from EMC's use of 4% more disks on the CX3-40 configuration.

### 7.2 Performance Impact of Virtualization

The most crucial difference between a NetApp SAN system and an EMC CLARiiON SAN system is the difference between the systems' dynamic vs. static virtualization technologies. NetApp FAS systems use patented dynamic virtualization technology to continuously optimize the placement of each block on disk as it is written. The NetApp dynamic virtualization technology requires more metadata to manage continuous optimization than does the EMC static (MetaLUN) virtualization approach; however, the NetApp virtualization implementation allows substantial physical I/O efficiencies in how data is managed. As a result, NetApp FAS systems implement EMC's simulated Exchange workload significantly more efficiently than do CLARiiON systems using RAID 5. This efficiency provides the improved performance shown in Figures 1 and 2. If EMC were to reproduce this test on the CX3-40 using RAID 1/0, we expect that the CLARiiON system performance would improve incrementally; however, this improvement would come at a substantial loss of usable capacity. This would be a steep price to pay for only an incremental performance improvement.

### 7.3 Performance Across Storage Utilization Levels

The "performance across increasing utilization" data in Figure 4 lines up very well with our expectations. For very small datasets, the FAS3050C significantly outperforms the CX3-40. For very large data sets, both systems are fundamentally limited by the access time of the disks, although WAFL® (Write Anywhere File Layout) virtualization continues to show an advantage.

The performance differences observed here are related to the size of the active data set, not to the amount of disk space in use. If the test had been conducted by filling the LUNs to 100% and then accessing only 10%, 25%, and so on of the LUNs, both systems would have demonstrated the same results on this test. This fact is important. Other than Microsoft Exchange Server, we know of no application that has the property that hot data is uniformly distributed across the entire data set. For typical business applications, the hot data is highly localized. This localization plays to the NetApp continuous I/O optimization strength, as shown in figures 3 and 4, where the NetApp storage *substantially* outperforming the CX3-40 for datasets with small hot regions.

### 7.4 Performance Over Time

The “performance over time” graph (Figure 5) demonstrates that the performance of the FAS3050C on this Exchange simulation varies slightly over time within a narrow band. The performance changes are a result of the same virtualization and optimization technology that enables the FAS3050C to consistently achieve 13% more performance on 4% fewer disks in this disk-limited workload.

We believe that NetApp storage systems, when properly sized and configured, deliver superior performance on Exchange workloads. EMC’s published benchmark helps to substantiate this belief.

### 7.5 Performance Impact of Snapshot™ Copies (Multiple Online Recovery Points)

NetApp technology provides far more than optimized read/write workloads; it also provides highly efficient Snapshot copies, space-efficient, low-cost clones, and a myriad of other features. The performance advantage of NetApp virtualization technology (both with and without Snapshot active) has been verified by VeriTest [Reference 3], an independent testing lab.

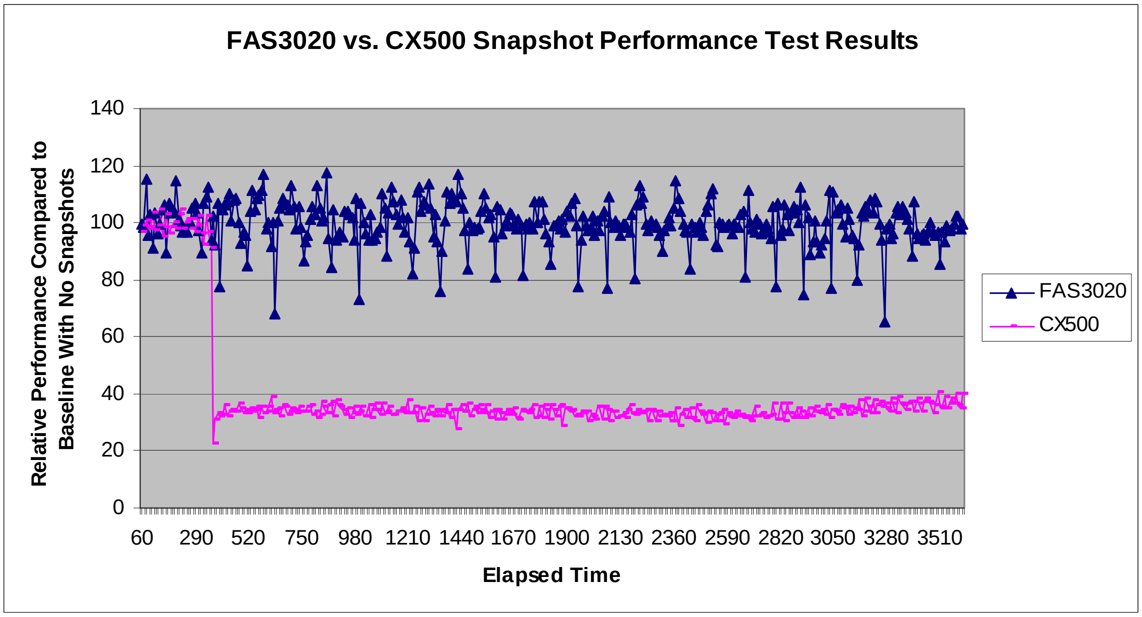

**Figure 6**) Relative performance of CLARiiON and NetApp FAS under Snapshot load.

As noted in Figure 6, from independent testing by VeriTest, NetApp performance under load remains near 100% while multiple snapshots are taken (every 120 seconds) and remain active during the entire test. On the other hand, the VeriTest report confirms that EMC CLARiiON performance drops significantly, by over a 60% margin, after the very first SnapView snapshot is activated. Only eight SnapView snapshots were possible per LUN at the time of testing by VeriTest for that report.

## 8. Performance Across Protocols

EMC's paper describes measuring the performance of the FAS3050C using various protocols and finds little difference in performance. We believe that EMC's measurement technology is fundamentally flawed by their apparent failure to follow NetApp best practice guidelines; however, their conclusion that performance of the FAS3050C is similar across protocols conforms to our general understanding of the performance of NetApp FAS systems. For random-access workloads, the performance of most systems is largely independent of interconnect protocol. As long as the majority of the work is done by the disk drives, performance is dominated by the disks' ability to process the requested workload, not by the chosen interconnect technology.

## 9. Conclusions

Working within EMC's purported framework, this paper seeks to confirm that EMC's performance measurements are incorrect. NetApp FAS products simply outperform EMC's CLARiiON systems on EMC's chosen workload.

In addition, with workload-specific tuning, NetApp FAS technology outperforms a similarly configured CLARiiON by a full 30%.

## 10. References

1. EMC Technical Note*, "NetApp Performance in SAN Environments."* (P/N 300-004-233, Rev A01) http://www.emc.com/techlib/pdf/300-004-233.pdf
2. *FCP Windows Attach Kit 3.0: Installation and Setup Guide*
3. *Network Appliance FAS3020 and EMC CX500: Comparison of Usability and Performance*. Veritest report: http://www.lionbridge.com/NR/rdonlyres/27A8C75E-1C1F-4084-A6BB-21C5CAC58A19/0/2005_NetApp_Competitive_Analysis.pdf

## 11. Revision History

| Date | Name | Description |
|---|---|---|
| 1/25/2007 | Stephen Daniel | Clarifications added |
| 11/12/2006 | Stephen Daniel | First published version |
| 10/25/2006 | Stephen Daniel | Restructured sections 5 & 7 |
| 10/24/2006 | Stephen Daniel | Sections 4.2, 5.4 added |
| 10/17/2006 | Stephen Daniel | First Version |