Performance Evaluation of DR550 v4.5 Using the System Storage Archive Manager (SSAM) API

March 2, 2009

Author Mike Roll, Vernon Miller
Manager Kamran Amini
Modular Storage and Power Blade Performance, Tucson
IBM SYSTEMS AND TECHNOLOGY GROUP

NOTICES AND DISCLAIMERS

Product data has been reviewed for accuracy as of the date of initial publication. Product data is subject to change without notice. This information could include technical inaccuracies or typographical errors. IBM may make improvements and/or changes in the product(s) and/or programs(s) at any time without notice.

The performance data contained herein was obtained in a controlled, isolated environment. Actual results that may be obtained in other operating environments may vary significantly. While IBM has reviewed each item for accuracy in a specific situation, there is no guarantee that the same or similar results will be obtained elsewhere.

References in this document to IBM products, programs, or services does not imply that IBM intends to make such products, programs or services available in all countries in which IBM operates or does business. Any reference to an IBM Program Product in this document is not intended to state or imply that only that program product may be used. Any functionally equivalent program, that does not infringe on IBM's intellectual property rights, may be used instead. It is the user's responsibility to evaluate and verify the operation of any non-IBM product, program or service.

THE INFORMATION PROVIDED IN THIS DOCUMENT IS DISTRIBUTED "AS IS" WITHOUT ANY WARRANTY, EITHER EXPRESS OR IMPLIED. IBM EXPRESSLY DISCLAIMS ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE OR NON-INFRINGEMENT. IBM shall have no responsibility to update this information. IBM products are warranted according to the terms and conditions of the agreements (e.g., IBM Customer Agreement, Statement of Limited Warranty, International Program License Agreement, etc.) under which they are provided. IBM is not responsible for the performance or interoperability of any non-IBM products discussed herein.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products in connection with this publication and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

The provision of the information contained herein is not intended to, and does not, grant any right or license under any IBM patents or copyrights. Inquiries regarding patent or copyright licenses should be made, in writing, to:

IBM Director of Licensing
IBM Corporation
North Castle Drive
Armonk, NY 10504-1785
U.S.A.

IBM, System Storage, and AIX are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries or both.

NOTE:

The hard copy version of this document is FOR REFERENCE ONLY.

It is the responsibility of the user to ensure that they have the current version. Any outdated hard copy is invalid and must be removed from possible use. It is also the responsibility of the user to ensure the completeness of this document prior to use.

Table of Contents

DR550 v4.5 File System Gateway (FSG) Performance Evaluation

Introduction

The IBM System Storage DR550, one of the IBM Data Retention offerings, is an integrated offering for clients that need to retain and preserve electronic business records. It is designed to help store, retrieve, manage and retain regulated and non-regulated data. In addition to managing compliance data, it can also be an archiving solution for other types of data.

However, the main focus of the DR550 is to help companies manage (archive and retrieve) data assets in a non-erasable and non-rewritable storage system. It provides a secure data repository where deletion or modification of data is completely disallowed except through a well defined retention and expiration policy. New industry and government regulations are leading to greater amounts and types of data that need to be stored in this type of format.

DR550 Product Description

The DR550 brings together off-the-shelf IBM hardware and software products. The hardware comes already mounted in a secure rack, including IBM's POWER5+ servers with AIX, IBM System Storage Archive Manager (SSAM), and the IBM System Storage DS4700 Express1.

The DR550 is fully integrated. The hardware has been installed, cabled and tested in a rack, and the software has been installed and largely pre-configured. System administrators are required to define the SSAM management classes associated with the retention periods for their business needs. When combined with a content management application, data can be managed throughout it's lifecycle, including automatic deletion at the end of it's retention period.

System Storage Archive Manager (SSAM)

The primary software used to drive the DR550 is IBM's System Storage Archive Manager (SSAM). It's designed to help customers protect the integrity of data as well as automatically enforce data retention policies. Using policy-based management, data can be stored indefinitely, can be expired based on a retention event, or have a predetermined expiration date. In addition, the retention enforcement feature may be applied to data using deletion hold and release interfaces which hold data for an indefinite period of time, regardless of the expiration date or defined event.

The policy software is also designed to prevent modifications or deletions after the data is stored. With support for open standards, the new technology is designed to provide customers flexibility to use a variety of content management or archive applications.

Purpose of This Performance Evaluation

The purpose of this Performance Evaluation was to measure the performance of archiving and retrieving objects to and from the DR550 through the SSAM API. Throughput was measured in MB/sec and Objects/sec as the size of the objects changed and the number of objects per session increased. The effect of encryption on archiving and retrieving objects was also described.

Performance Workload Tools

The Tivoli Storage Manager (TSM) "FakeLoad" tool is an application program that generates a variety of workloads using the standard SSAM Application Program Interface (API). For each run, it produces data objects in the memory of the client system to be sent to the DR550, captures the performance information and generates a performance report. The tool acts as a SSAM client and eliminates all influences cause by disk IO on the client system.

Performance Measurement Approach

The Fakeload tool was used to generate archive and retrieve job streams in which:

- The size of the objects were varied (8K, 16K, 64K, 256K, 512K and 1024K).
- The number of objects per session was varied (1, 5, 10, 25, 50 & 500) to explore the performance benefit of grouping files or objects.
- Data encryption was either engaged or not engaged.

The resulting graphs show the rates at which data can be archived or retrieved in the DR500, in both MB/sec and Objects/sec.

Object Aggregation (Multi-Object Transactions)

The SSAM API supports aggregating objects (such as e-mails) either being read from or written to the DR550 into a single transaction. This greatly reduces the overhead of the API and improves throughput performance.

The performance advantage of aggregating multiple objects for archiving or retrieving is that the required overhead time (e.g. establishing a network connection, opening files, checking parameters, etc.) occurs only once for many objects rather than once for each object. The SSAM client leverages this capability by archiving or retrieving all appropriate objects or files with a single command.

Data Encryption

Encryption of data being archived on the DR550 is another option of the SSAM API running on the application server managing the data to be archived. The SSAM API can be configured to encrypt the object or file and then sends it to the DR550. Encryption is controlled within the SSAM API via an optional parameter setting.

The choice of whether or not to use encryption is a business decision based on the need to protect data during transmission to the archive repository, as well as whether data should be stored in an encrypted format.

Performance Measurement Environment

The System Storage Archive Manager was embedded on an IBM System P5 52A using POWER5+ processors which are part of a standard DR550 configuration. This entry-level server incorporates a number of the attributes of IBM's high-end servers.

The File System Gateway ran on the 2229-FSG File System Gateway node with 2 GB of RAM, powered by a 3.2GHz dual core Xeon processor using SUSE Linux™ Enterprise Server Version 10.
The 2229-FSG connected to the customer network using 1 Gbps Ethernet adapters. The 2229-FSG connected to the DR550 using a 10/100/1000 TX Ethernet adapter. As seen below, one bonded Ethernet connection was used between the customer network and the DR550's SSAM server.

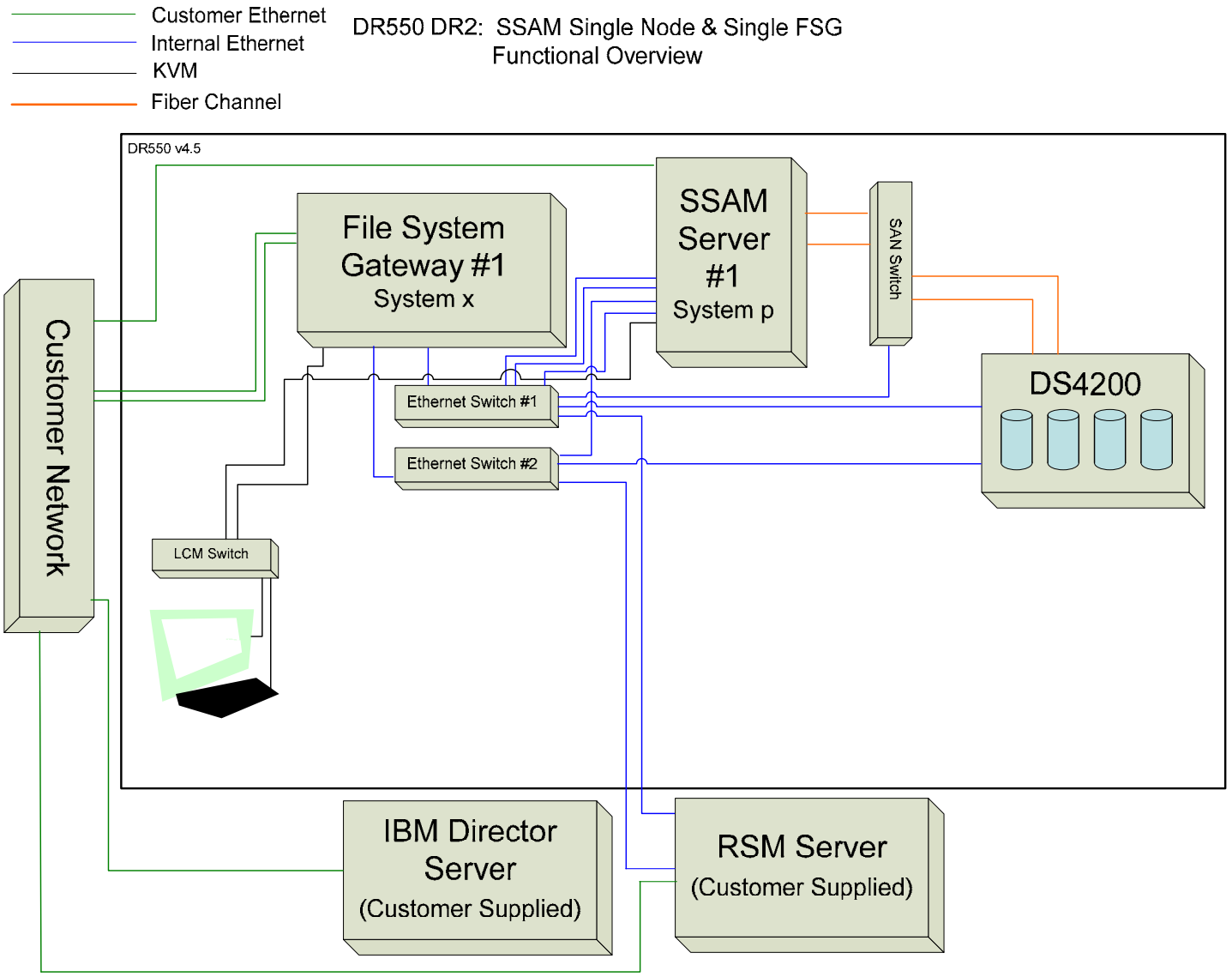
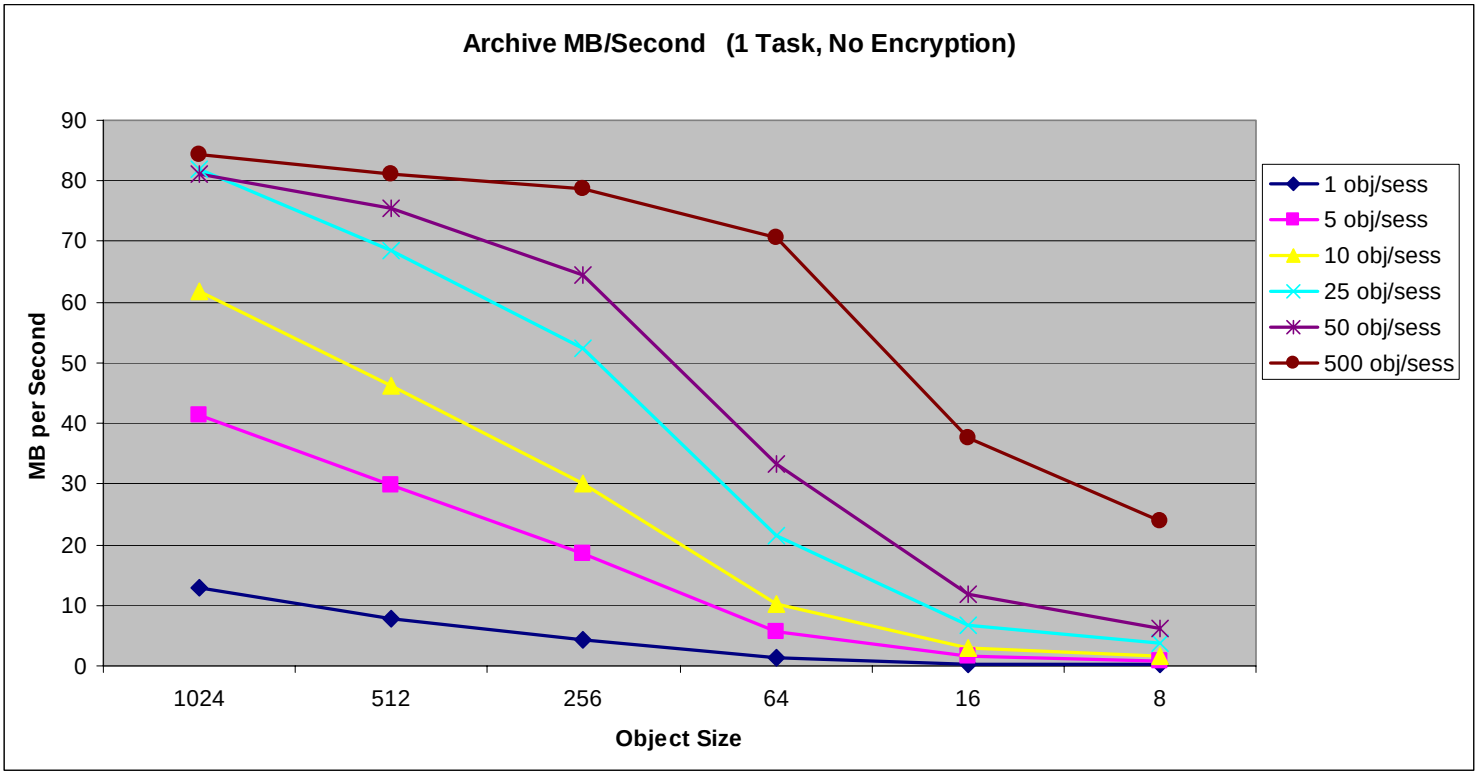


Figure #1 – Logical Connection Diagram of the DR550 DR2 used for the performance evaluation.

The storage subsystem used in the DR550 was the DS4200 with 16 750 GB SATA drives in the control unit and two EXP420 16-disk expansion drawers. There were 48 disk drives total configured with over 26 TB of storage.

The performance evaluation was performed with both write cache and write cache mirroring enabled for the logical drives within the DS4200 storage system. The DS4200 was attached to the SSAM server through a switch using two 4Gb/second Fiber Channel lines. In addition, archived data was written to a single SSAM storage pool which was spread across all disk arrays on the DS4200.

1. Archive Without Encryption – Single Task (MB/sec)

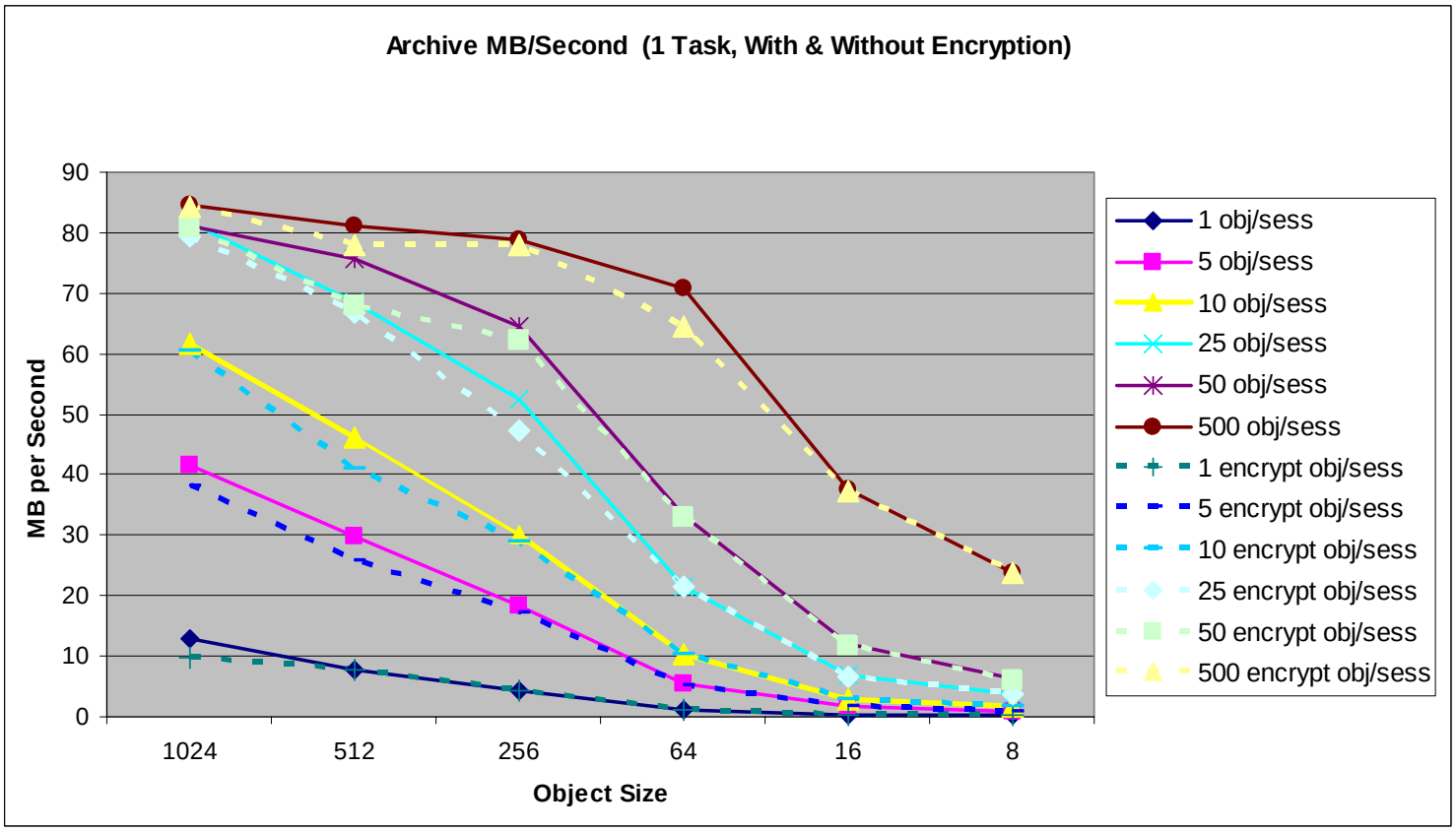


This diagram describes the rate (MB/sec) at which data can be archived in the DR550 without data encryption. The different lines show the effect of aggregating different number of objects into single transactions within the sessions. This aggregation reduces the overhead associated with the processing of each archive (or retrieve) session.

As shown above, the throughput performance improves as the number of objects per transaction in the session increases. The SSAM software becomes more efficient as the number of objects per session increases.

Also, throughput performance increases as the size of the object increases because the SSAM API overhead becomes less with larger objects. 84 MB/sec was achieved with 500 1MB objects bundled per session.

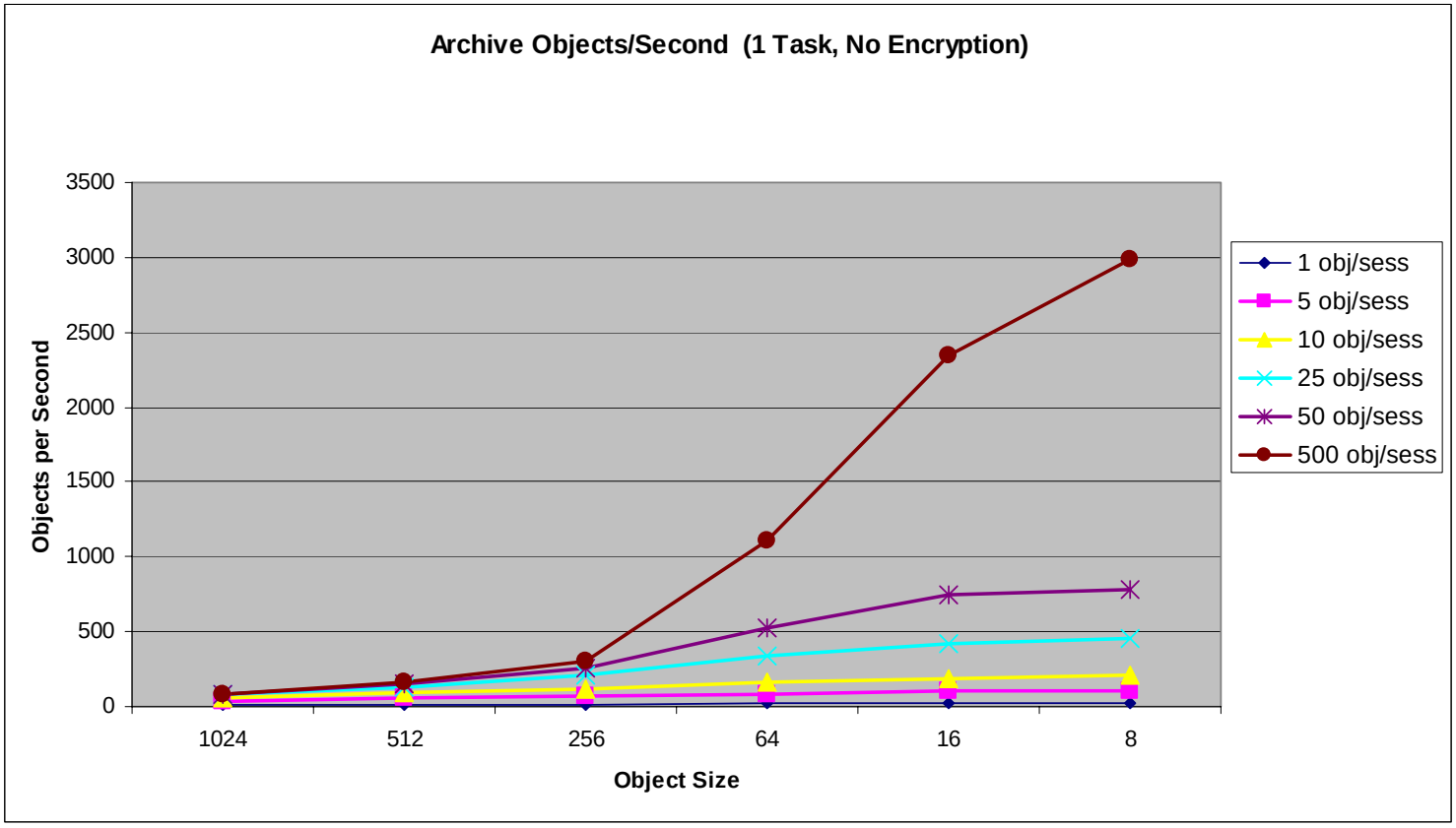
2. Archive, With and Without Encryption – Single Task (MB/sec)



This diagram shows the rate (MB/sec) at which data can be archived in the DR550 both with and without data encryption. The results indicate that enabling encryption on the SSAM API running on the application server doesn't seem to have a significant impact on DR550 performance. The application server running the SSAM API is powerful enough to encrypt data and still keep the DR550 busy.

Again, the different lines show that performance increases as greater numbers of objects are aggregated per transaction in the sessions. The throughput performance improves as the number of objects per transaction decreases and the objects get larger.

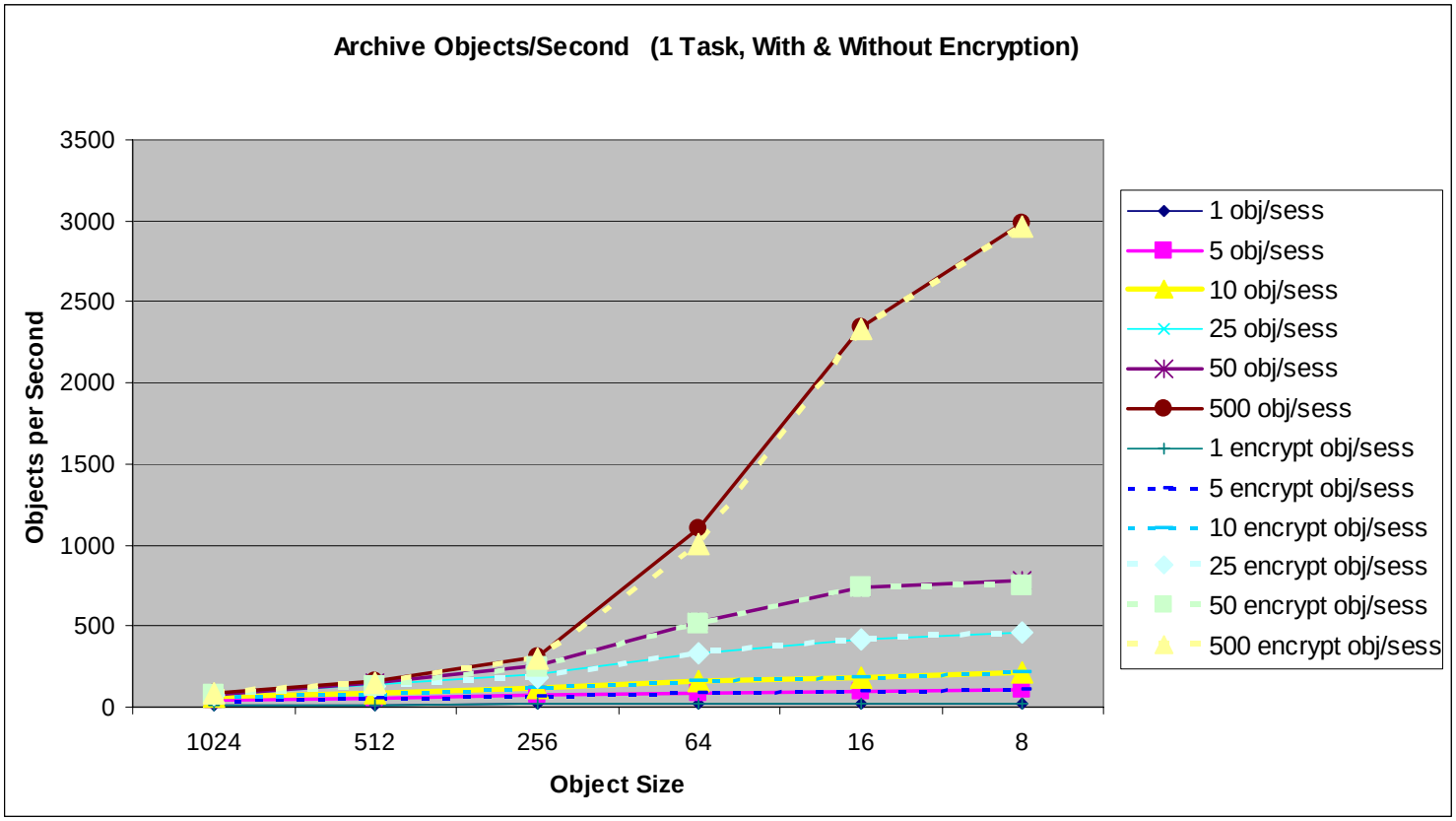
3. Archive Without Encryption – Single Task (Objects/sec)



This diagram shows the DR550 Archive performance without encryption in Objects/sec. Here, the throughput increases as the number of objects per transaction in the session increases and the object size gets smaller.

3000 Objects/sec were archived with 500 8K objects per session.

4. Archive, With and Without Encryption – Single Task (Objects/sec)

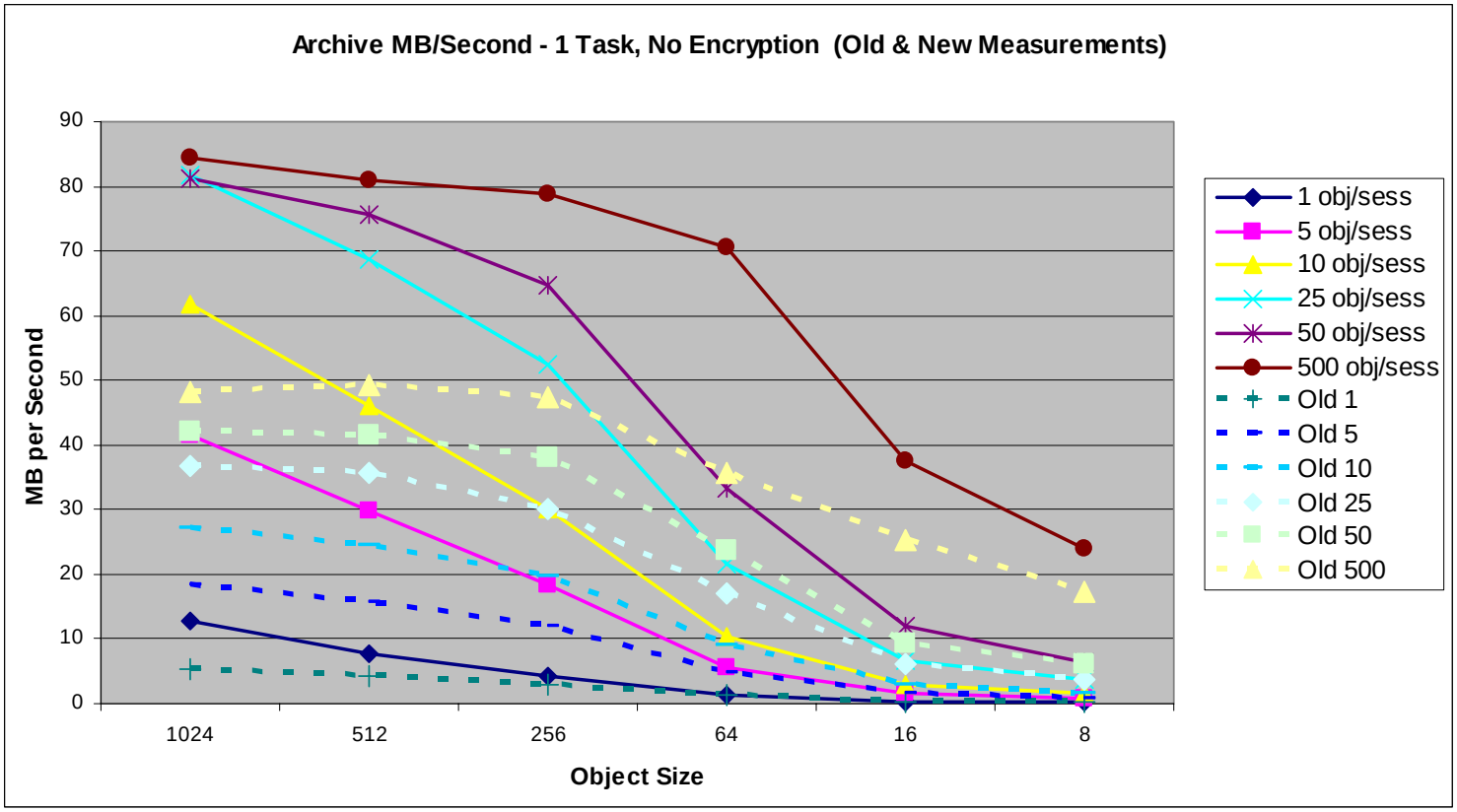


This diagram shows the DR550 Archive performance both with and without encryption in Objects/sec. The graph shows that throughput increases as the number of objects per transaction in the session increases and the object size gets smaller.

The results again indicate that enabling encryption on the SSAM API running on the application server doesn't seem to have a significant impact on DR550 performance. The SSAM application server is powerful enough to both encrypt data and still keep the DR550 fully engaged.

The performance benefits of object aggregation, (i.e. archiving multiple objects in a single transaction in a session) is also apparent. The object throughput performance improves as the number of objects per transaction in the session increases and the objects get smaller.

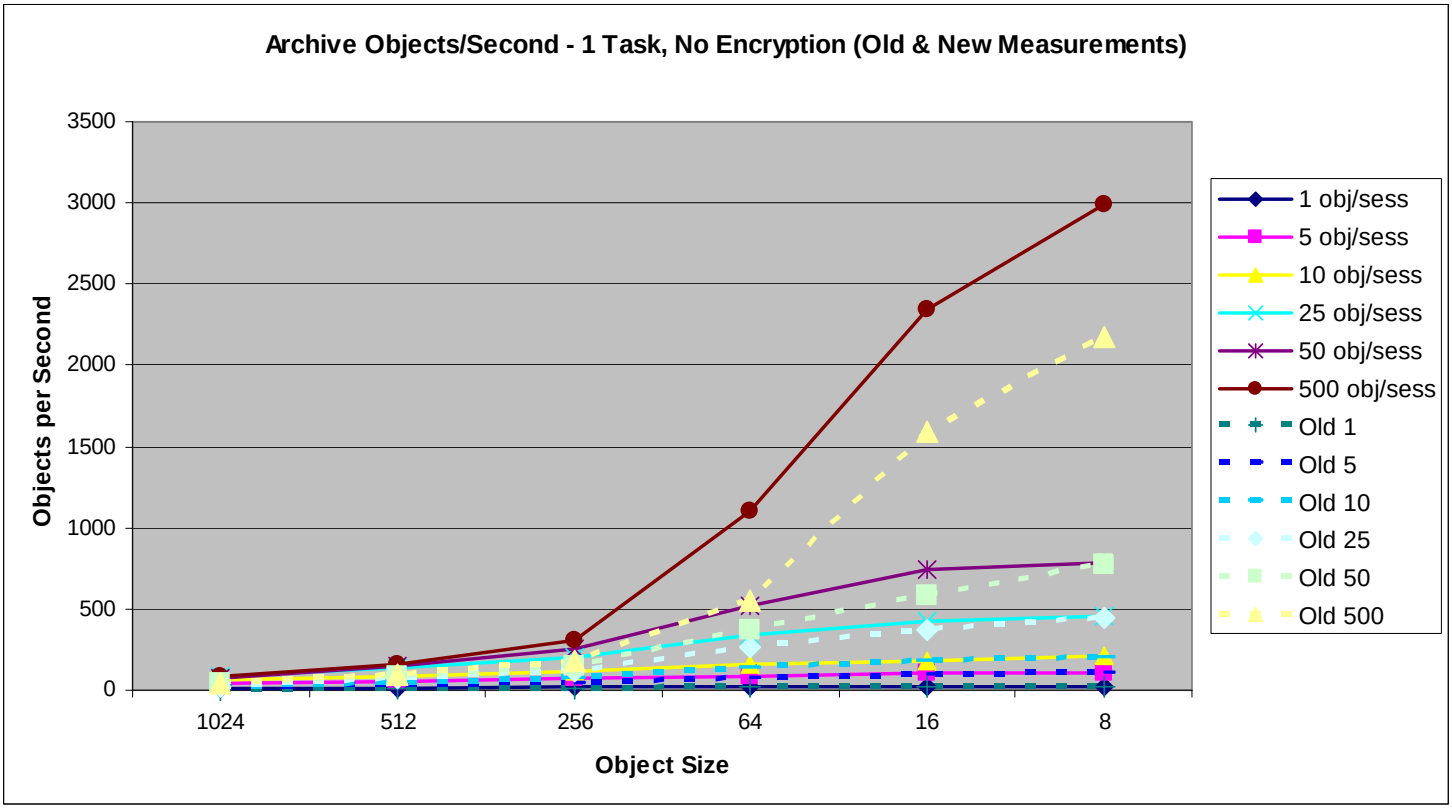
5. Archive Without Encryption – Single Task, DR550 v3.0 & v4.5 (MB/sec)



The diagram above shows the comparison of archiving speed in MB/sec for the DR550 v4.5 and DR550 v3.0 (labeled as Old 1 through Old 500). Encryption was not used. New technology in the DR550 v4.5 contributes to this improvement performance. Faster servers, storage and improved software and firmware all contribute to boost the efficiency of the DR550 v4.5.

The graphs show a distinct performance improvement between the two DR550 releases, For 1MB objects, the DR550 v4.5 with 25 objects per session is 1.95 times faster than the DR550 v3.0. The DR550 v4.5 with 500 objects per session was 1.75 times faster than DR550 v3.0 with the same workload.

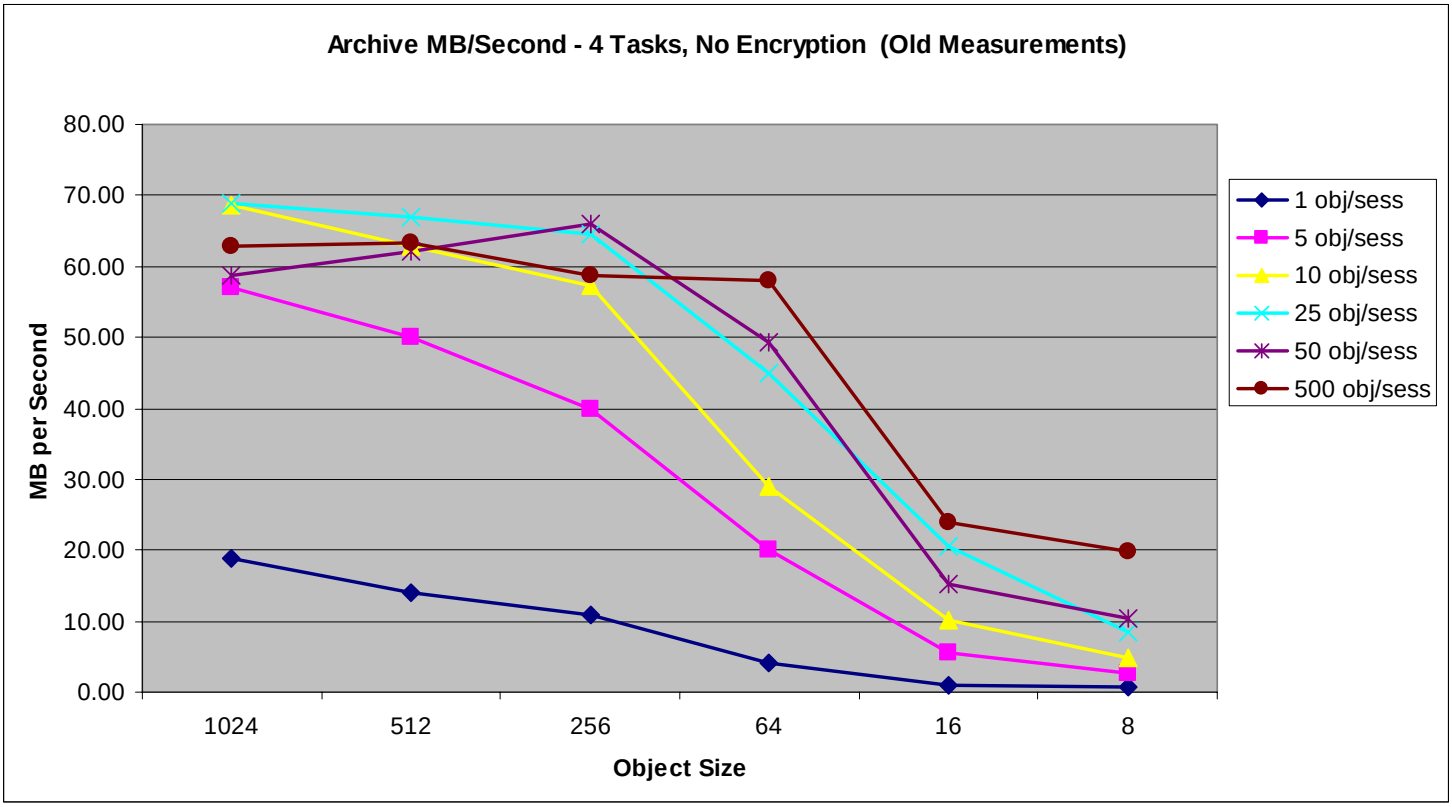
6. Archive Without Encryption – Single Task, DR550 v3.0 & v4.5 (Objects/sec)



The diagram above shows the comparison of archiving speed in objects per second for the DR550 v4.5 and DR550 v3.0 (labeled as Old 1 through Old 500). Encryption was not used. Again, the DR550 v4.5 shows improved overall performance.

For the smallest 8K objects, the DR550 v4.5 with 500 objects per session is 1.36 times faster than the DR550 v3.0. However The DR550 v4.5 with 25 objects per session shows fewer Objects/sec than DR550 v3.0 with the same workload.

7. Archive Without Encryption – 4 Concurrent Tasks, DR550 v3.0 (MB/sec)

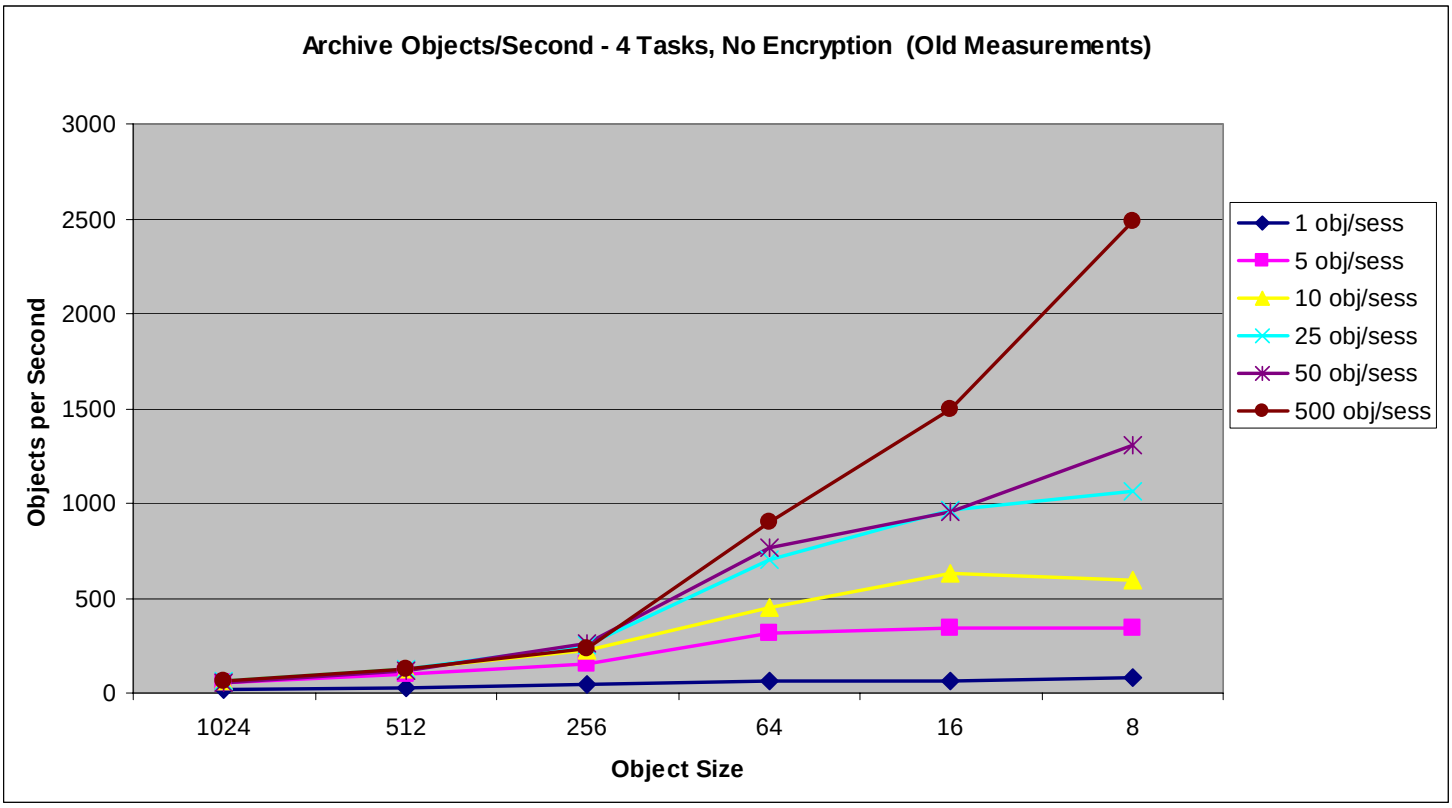


This diagram describes the rate (MB/sec) at which data was archived in the older DR550 v3.0 using 4 concurrent tasks running on the same client without data encryption. It gives a baseline to compare with a previous slide (page 11) to show how much the Archive performance throughput has improved with DR550 v4.5.

Even when taking advantage multiple concurrent processes, the older DR550 v3.0 running 4 tasks still has less Archive throughput than the DR550 v4.5 running only a single task while exercising the same workload.

For 25 objects/session, v3.0 shows 69 MB/sec vs. 82 MB/sec measured for v4.5. For 500 objects per session, v3.0 shows 63MB/sec vs.84 MB/sec measured for v4.5.

8. Archive Without Encryption – 4 Concurrent Tasks, DR550 v3.0 (Objects/sec)



This diagram displays the throughput of the older DR550 v3.0 running the same workload as shown in the previous diagram. It describes the v3.0 data archive rate in Objects per second using 4 concurrent tasks running on the same client without data encryption.

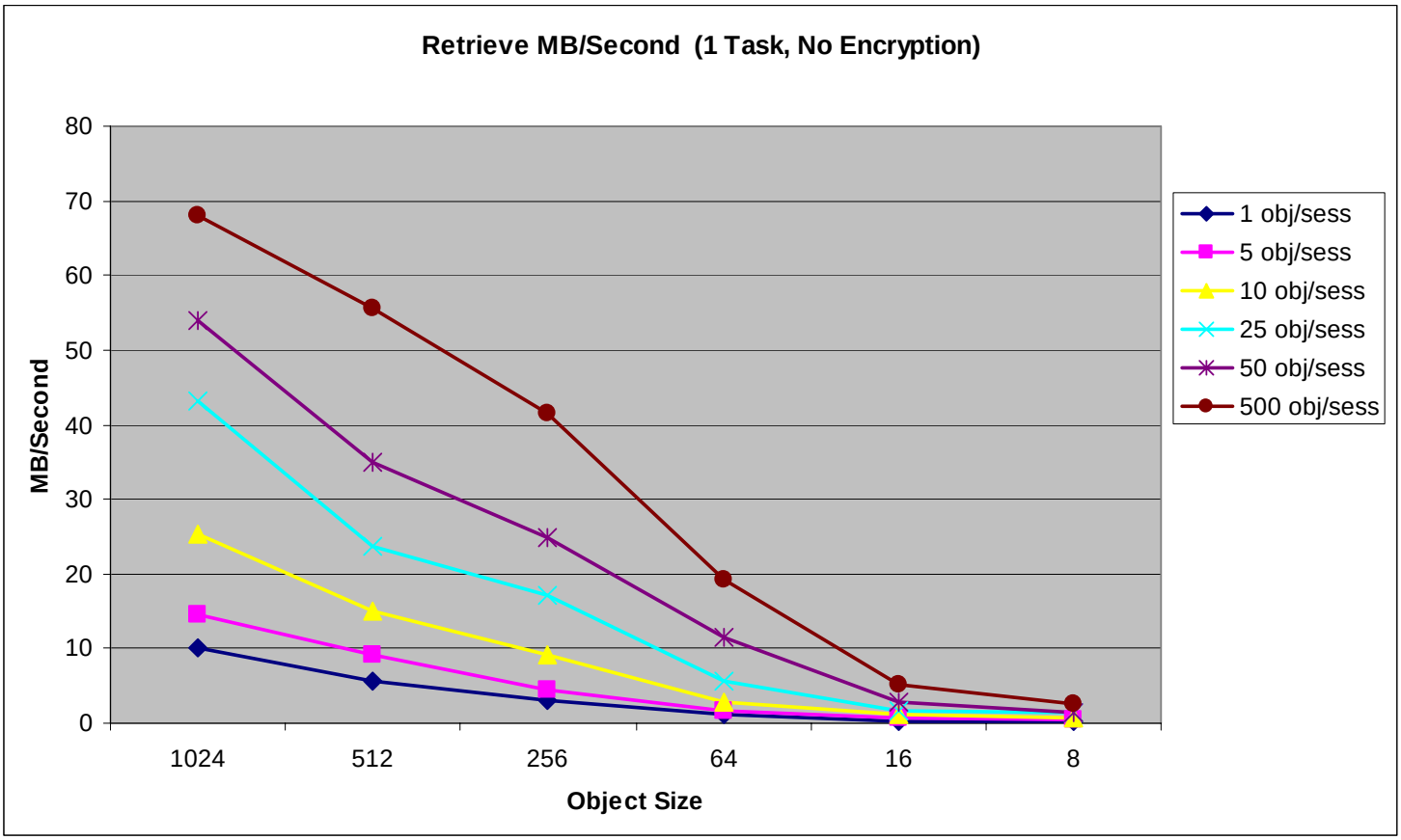
For 500 8K-sized objects per session, DR550 v4.5 still archives more objects per second running only one task than v3.0 running the same workload using 4 concurrent tasks.

However, at less the efficient bundling level of 25 8K-sized objects per session, the older v3.0 running 4 concurrent tasks exceeds the Objects/sec of DR550 v4.5 running one task.

For 500 8K-sized objects per session, v3.0 shows 2500 Objects/sec vs. 3000 8K-sized Objects per session for v4.5.

For 25 8K-sized objects per session, v3.0 shows 1030 Objects/sec vs. 495 Objects/sec for v4.5.

9. Retrieve Without Encryption – Single Task (MB/sec)

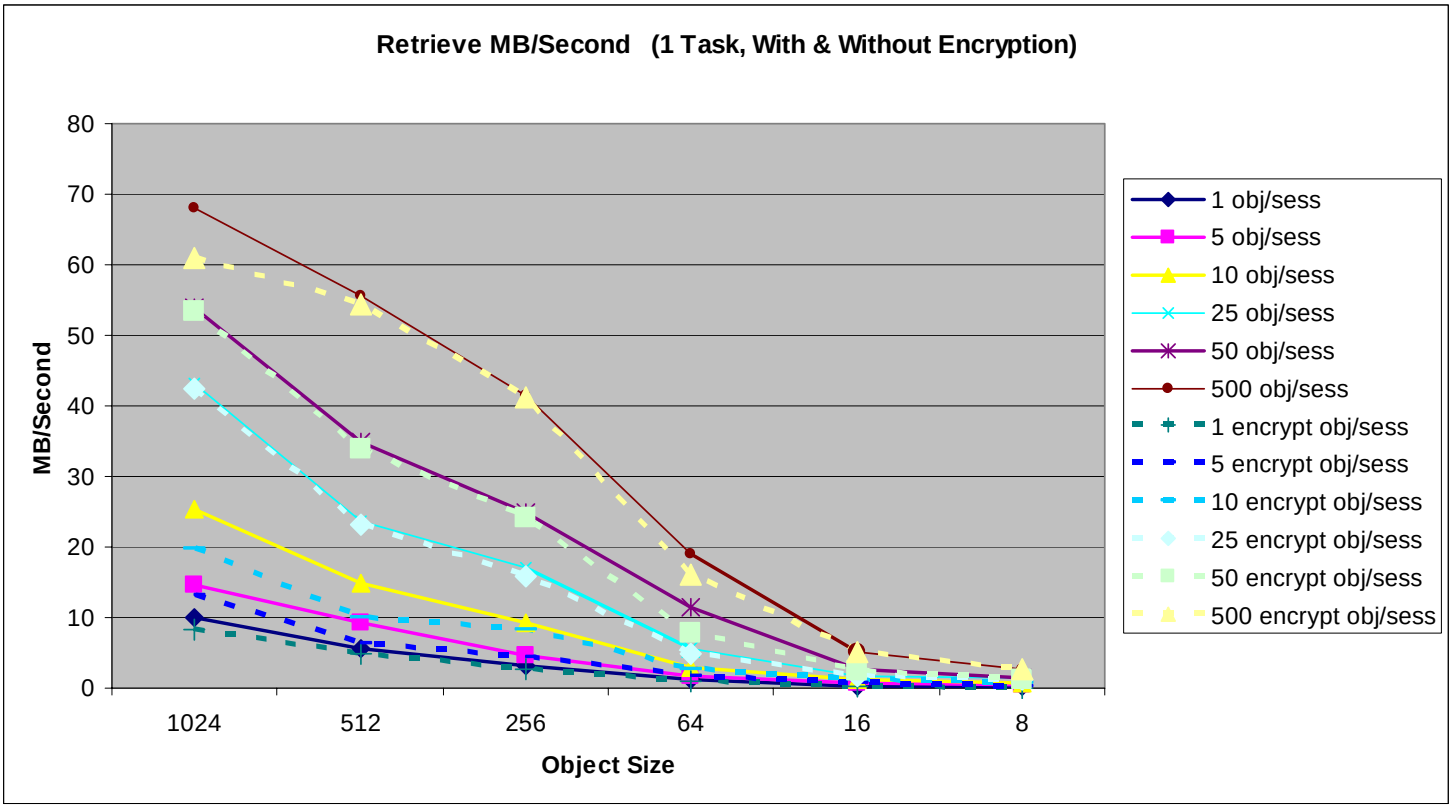


This diagram describes the rate (MB/sec) at which data can be retrieved in the DR550 without data encryption. The different lines show the effect of logically grouping different number of objects into single sessions. This bundling reduces the overhead associated with the processing of each retrieve session.

As shown above, the throughput performance improves as the number of objects per transaction in the session increases. The SSAM software becomes more efficient as the number of objects per transaction in each session increases.

Also, throughput performance increases as the size of the object being retrieved increases. 79 MB/sec was achieved with 500 1MB objects bundled per session.

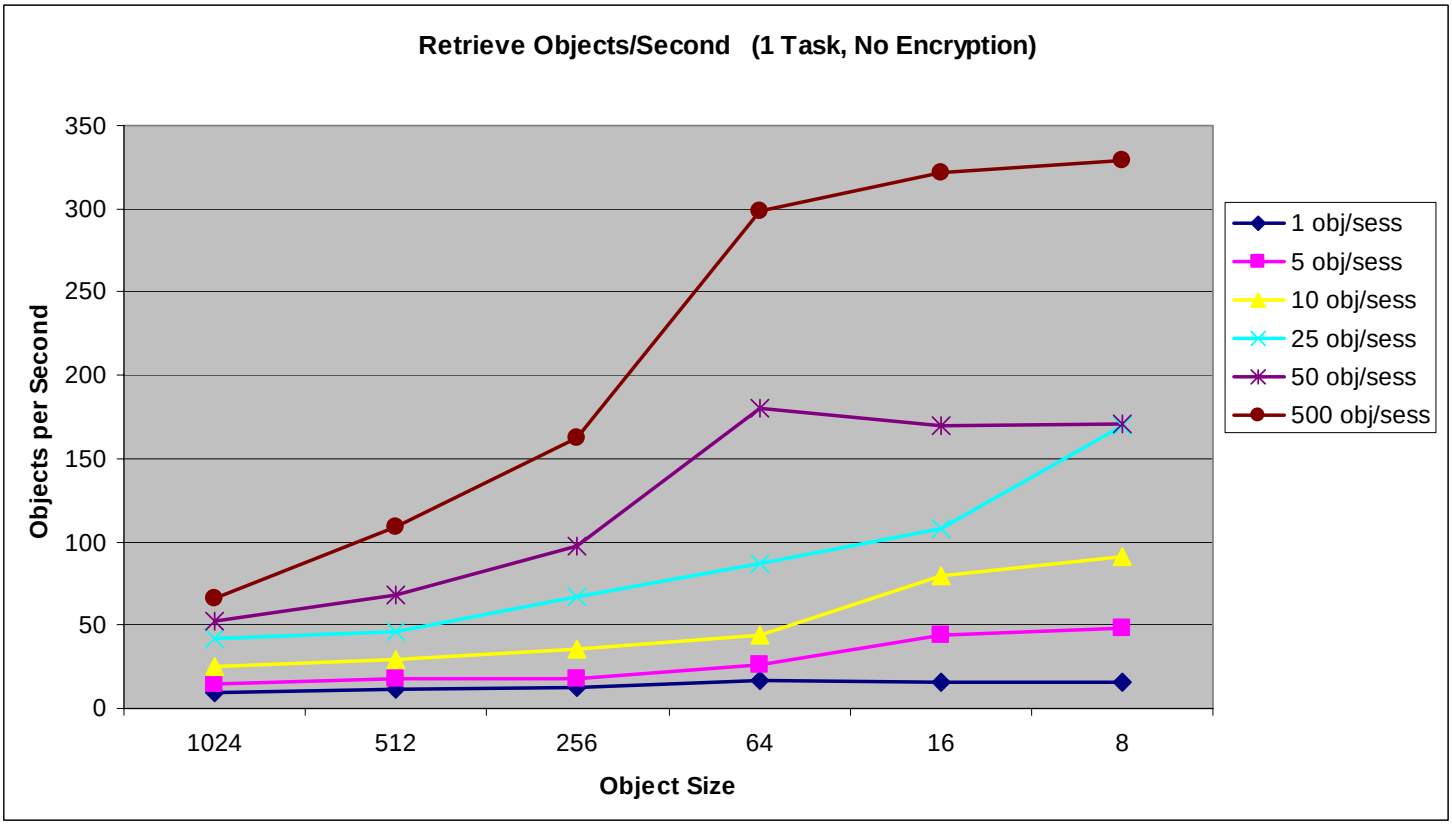
10. Retrieve, With and Without Encryption – Single Task (MB/sec)



This diagram shows the rate (MB/sec) at which data can be retrieved in the DR550, both with and without data encryption. The results indicate that enabling encryption on the SSAM API running on the application server doesn't seem to have a significant impact on DR550 performance. The application server running SSAM is powerful enough to de-encrypt data and still keep the DR550 busy.

Again, the different lines show that performance increases as greater numbers of objects are aggregated per transaction in each session. The throughput performance improves as the number of objects per transaction in the session decreases and the objects get larger.

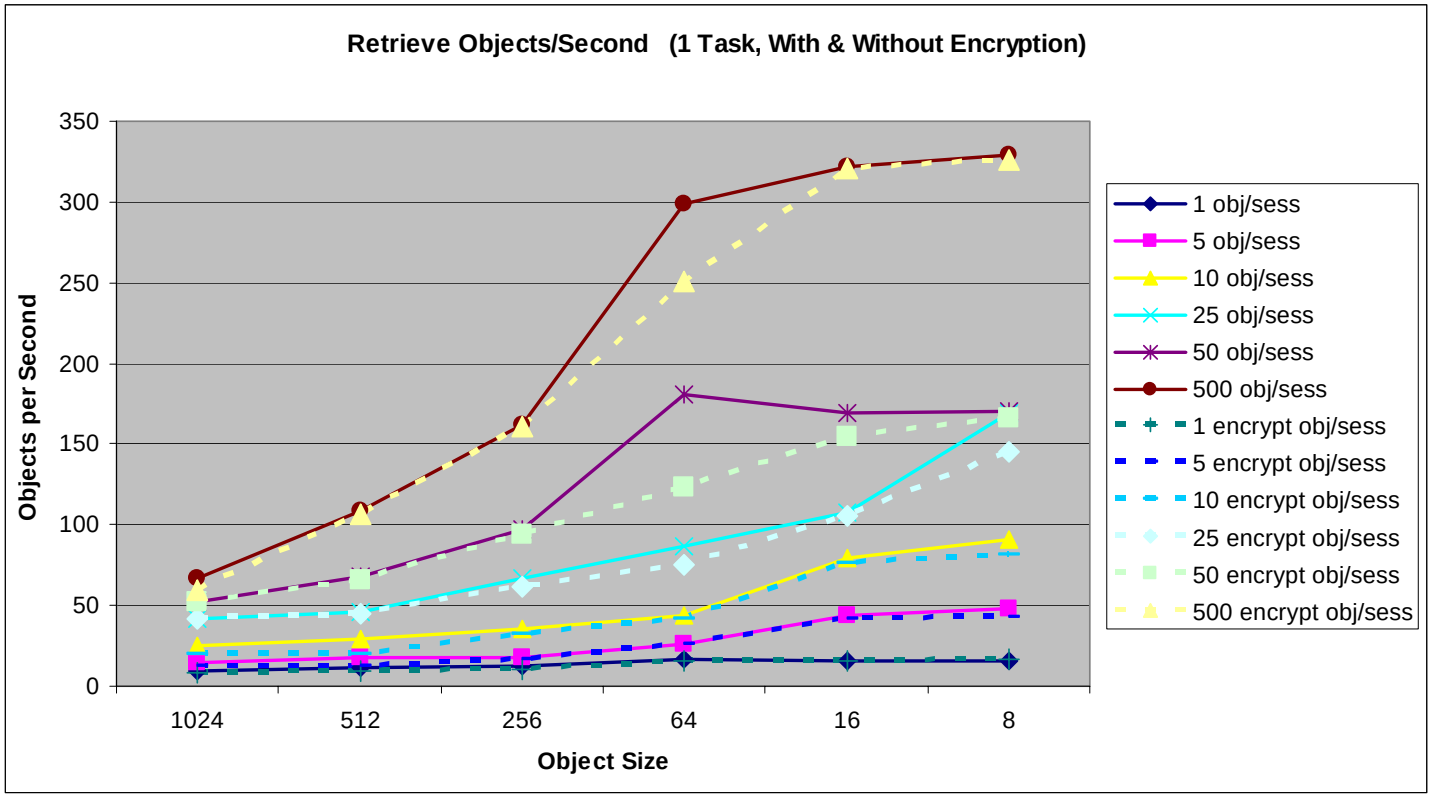
11. Retrieve Without Encryption – Single Task (Objects/sec)



This diagram shows the DR550 Retrieve performance without encryption in Objects/sec. Here, the throughput increases as the number of objects per transaction in the sessions increase and the object sizes gets smaller.

330 Objects/sec were retrieved with 500 8K objects bundled per transaction in the session.

12. Retrieve, With and Without Encryption – Single Task (Objects/sec)

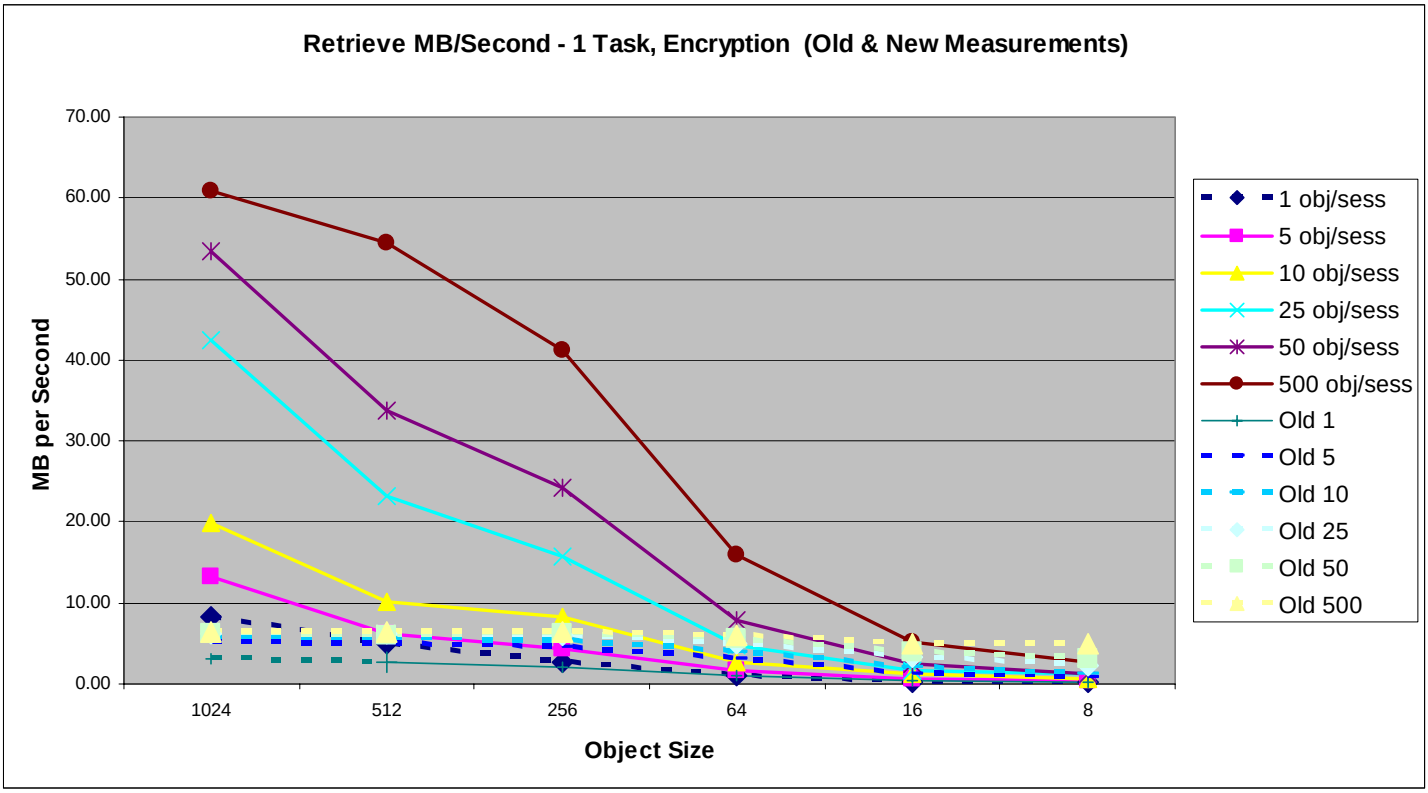


This diagram shows the DR550 Retrieve performance both with and without encryption in Objects/sec. The graph shows that throughput increases as the number of objects in the session increases and the object size gets smaller.

The results again indicate that enabling encryption on the SSAM client running on the application server doesn't seem to have a significant impact on DR550 performance. The SSAM application server is powerful enough to both de-encrypt retrieved data and still keep the DR550 fully engaged.

The performance benefits of logically grouping as many objects as possible into a session for a single retrieve are also apparent. The object throughput performance improves as the number of objects in the session increases and the objects get smaller.

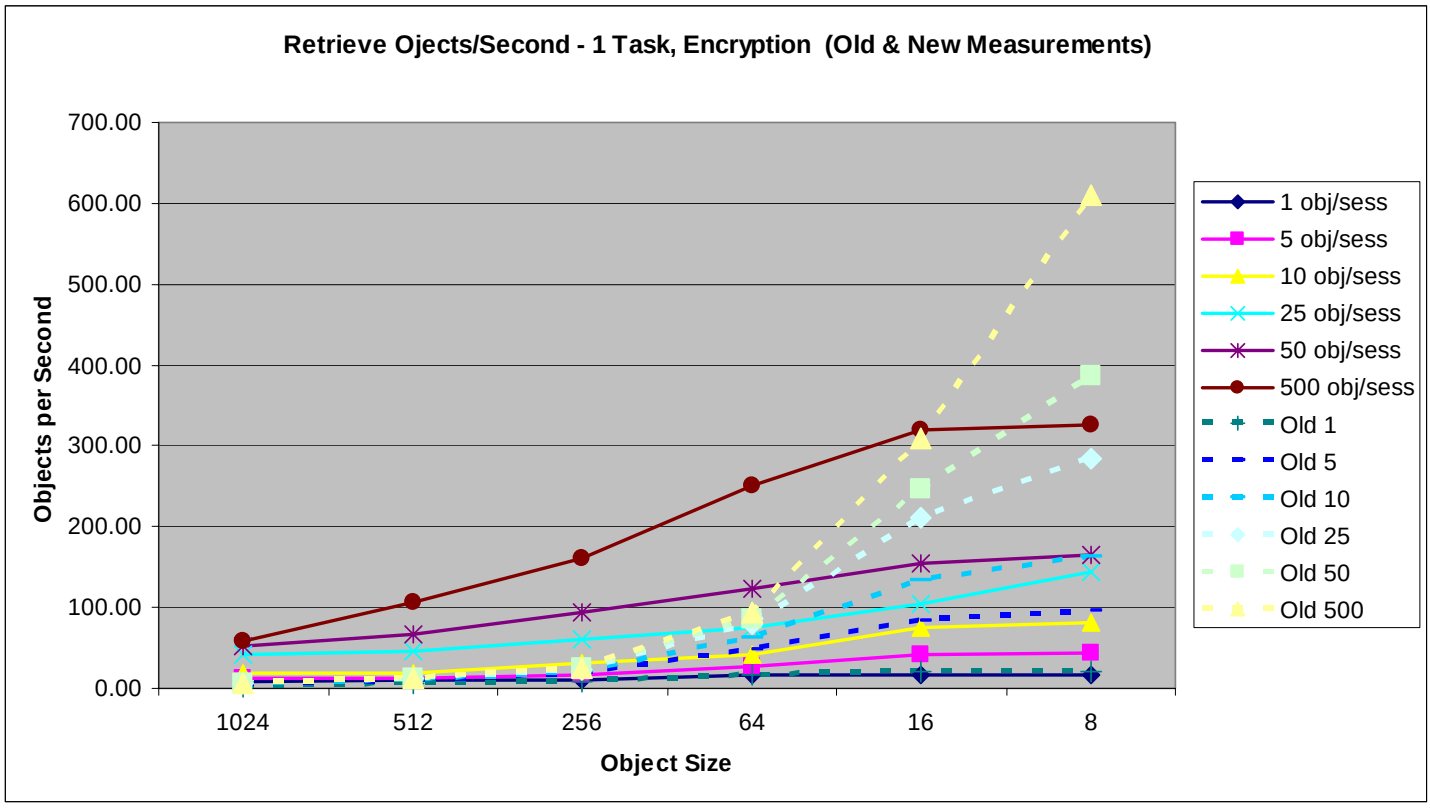
13. Retrieve With Encryption – Single Task, DR550 v3.0 & v4.5 (MB/sec)



The diagram above displays a comparison of retrieval speeds in MB/sec for the DR550 v4.5 and DR550 v3.0 (labeled as Old 1 through Old 500). Encryption was not used. New technology in the DR550 v4.5 contributes to this improvement performance. Faster servers and improved software and firmware all contribute to boost the efficiency of the DR550 v4.5.

The graphs show a distinct retrieval performance improvement between the two DR550 releases, For 1MB objects, the DR550 v4.5 with 25 objects per session is 5.2 times faster than the DR550 v3.0. The DR550 v4.5 with 500 objects per session was 7.6 times faster than DR550 v3.0 with the same workload.

14. Retrieve With Encryption – Single Task, DR550 v3.0 & v4.5 (Objects/sec)



The diagram above shows a comparison of retrieval speeds in objects per second for the DR550 v4.5 and DR550 v3.0 (labeled as Old 1 through Old 500). Encryption was not used. DR550 v4.5 shows improved retrieval performance for larger objects, but does not fare as well for smaller 16K or 8K objects.

Summary

The IBM DR550 SSAM API performance has been improved with the use of newer technology, including servers, faster storage systems and improved software and firmware. Encryption offers significant data protection and did not appear to adversely affect DR550 performance under the workloads measured in this evaluation.

The beneficial effects of aggregating multiple objects per transaction in each session were also shown. The results demonstrated that using the SSAM software to aggregate the objects improved performance throughput by reducing the overhead required to process each archive or retrieve transaction in the session.

The highest throughput (MB/sec) levels were achieved with the largest object sizes (1MB) while the greatest Object/sec rates were achieved with the smallest object sizes (8K). DR550 v4.5 showed a distinct increase in throughput performance over v3.0.

Acknowledgments

Many thanks to the SAN HW and Server Proven Test Department and their Manager, Carlos Lujan, who allowed us to use their equipment and time to generate the measurements which made this paper possible.

Thanks in particular to Dave DeHaan who provided us with valuable advice, technical consulting and DR550 hardware support, despite his own busy schedule.

Finally, we'd like to thank Mark Bayus, our DR550 DCT Program Manager, who gave encouragement and support for the Performance effort.